NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 1—Introduction to Matter, Energy, and Direct Current

NAVEDTRA 14173

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Sailor’s Creed

“I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country’s Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all.”

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: Subjects of matter, energy, electricity, batteries, and direct current are presented to give the student background information on topics that may be encountered in daily work.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and the occupational standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards*, NAVPERS 18068, found on line at https://buperscd.technology.navy.mil /bup_updt/upd_CD/BUPERS/enlistedManOpen.htm.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
ETCS(SW) Donnie Jones

Corrections and minor modifications prepared in
January 2003 by
ETC Jack Weatherford

NAVSUP Logistics Tracking Number
0504-LP-026-8260

TABLE OF CONTENTS

Assignments follow Appendix V.

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 6 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

(THIS PAGE IS INTENTIONALLY LEFT BLANK.)

CHAPTER 1

MATTER, ENERGY, AND ELECTRICITY

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the NRTC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completing this chapter, you will be able to:

1. State the meanings of and the relationship between matter, element, nucleus, compound, molecule, mixture, atom, electron, proton, neutron, energy, valence, valence shell, and ion.
2. State the meanings of and the relationship between kinetic energy, potential energy, photons, electron orbits, energy levels, and shells and subshells.
3. State, in terms of valence, the differences between a conductor, an insulator, and a semiconductor, and list some materials which make the best conductors and insulators.
4. State the definition of static electricity and explain how static electricity is generated.
5. State the meanings of retentivity, reluctance, permeability, ferromagnetism, natural magnet, and artificial magnet as used to describe magnetic materials.
6. State the Weber and domain theories of magnetism and list six characteristics of magnetic lines of force (magnetic flux), including their relation to magnetic induction, shielding, shape, and storage.
7. State, using the water analogy, how a difference of potential (a voltage or an electromotive force) can exist. Convert volts to microvolts, to millivolts, and to kilovolts.
8. List six methods for producing a voltage (emf) and state the operating principles of and the uses for each method.
9. State the meanings of electron current, random drift, directed drift, and ampere, and indicate the direction that an electric current flows.
10. State the relationship of current to voltage and convert amperes to milliamperes and microamperes.
11. State the definitions of and the terms and symbols for resistance and conductance, and how the temperature, contents, length and cross-sectional area of a conductor affect its resistance and conductance values.
12. List the physical and operating characteristics of and the symbols, ratings, and uses for various types of resistors; use the color code to identify resistor values.

INTRODUCTION

The origin of the modern technical and electronic Navy stretches back to the beginning of naval history, when the first navies were no more than small fleets of wooden ships, using wind-filled sails and manned oars. The need for technicians then was restricted to a navigator and semiskilled seamen who could handle the sails.

As time passed, larger ships that carried more sail were built. These ships, encouraging exploration and commerce, helped to establish world trade routes. Soon strong navies were needed to guard these sea lanes. Countries established their own navies to protect their citizens, commercial ships, and shipping lanes against pirates and warring nations. With the addition of mounted armament, gunners joined the ship's company of skilled or semiskilled technicians.

The advent of the steam engine signaled the rise of an energy source more practical than either wind and sails or manpower. With this technological advancement, the need for competent operators and technicians increased.

However, the big call for operators and technicians in the U.S. Navy came in the early part of the 20th century, when power sources, means of communication, modes of detection, and armaments moved with amazing rapidity toward involved technical development. Electric motors and generators by then had become the most widely used sources of power. Telephone systems were well established on board ship, and radio was being used more and more to relay messages from ship to ship and from ship to shore. Listening devices were employed to detect submarines. Complex optical systems were used to aim large naval rifles. Mines and torpedoes became highly developed, effective weapons, and airplanes joined the Navy team.

During the years after World War I, the Navy became more electricity and electronic minded. It was recognized that a better system of communications was needed aboard each ship, and between the ships, planes, submarines, and shore installations; and that weaponry advances were needed to keep pace with worldwide developments in that field. This growing technology carried with it the awareness that an equally skilled force of technicians was needed for maintenance and service duties.

World War II proved that all of the expense of providing equipment for the fleet and of training personnel to handle that equipment paid great dividends. The U. S. Navy had the modern equipment and highly trained personnel needed to defeat the powerful fleets of the enemy.

Today there is scarcely anyone on board a Navy ship who does not use electrical or electronic equipment. This equipment is needed in systems of electric lighting and power, intercommunications, radio, radar, sonar, loran, remote metering, weapon aiming, and certain types of mines and torpedoes. The Navy needs trained operators and technicians in this challenging field of electronics and electricity. It is to achieve this end that this module, and others like it, are published.

MATTER, ENERGY, AND ELECTRICITY

If there are roots to western science, they no doubt lie under the rubble that was once ancient Greece. With the exception of the Greeks, ancient people had little interest in the structure of materials. They accepted a solid as being just that a continuous, uninterrupted substance. One Greek school of thought believed that if a piece of matter, such as copper, were subdivided, it could be done indefinitely and still only that material would be found. Others reasoned that there must be a limit to the number of subdivisions that could be made and have the material still retain its original characteristics. They held fast to the idea that there must be a basic particle upon which all substances are built. Recent experiments have revealed that there are, indeed, several basic particles, or building blocks within all substances.

The following paragraphs explain how substances are classified as elements and compounds, and are made up of molecules and atoms. This, then, will be a learning experience about protons, electrons, valence, energy levels, and the physics of electricity.

MATTER

Matter is defined as anything that occupies space and has weight; that is, the weight and dimensions of matter can be measured. Examples of matter are air, water, automobiles, clothing, and even our own bodies. Thus, we can say that matter may be found in any one of three states: SOLID, LIQUID, and GASEOUS.

ELEMENTS AND COMPOUNDS

An ELEMENT is a substance which cannot be reduced to a simpler substance by chemical means. Examples of elements with which you are in everyday contact are iron, gold, silver, copper, and oxygen. There are now over 100 known elements. All the different substances we know about are composed of one or more of these elements.

When two or more elements are chemically combined, the resulting substance is called a COMPOUND. A compound is a chemical combination of elements which can be separated by chemical but not by physical means. Examples of common compounds are water which consists of hydrogen and oxygen, and table salt, which consists of sodium and chlorine. A MIXTURE, on the other hand, is a combination of elements and compounds, not chemically combined, that can be separated by physical means. Examples of mixtures are air, which is made up of nitrogen, oxygen, carbon dioxide, and small amounts of several rare gases, and sea water, which consists chiefly of salt and water.

Q1. What is matter, and in what three states is it found?

Q2. What is an element?

Q3. What is a compound?

Q4. What is the difference between a compound and a mixture?

MOLECULES

A MOLECULE is a chemical combination of two or more atoms, (atoms are described in the next paragraph). In a compound the molecule is the smallest particle that has all the characteristics of the compound.

Consider water, for example. Water is matter, since it occupies space and has weight. Depending on the temperature, it may exist as a liquid (water), a solid (ice), or a gas (steam). Regardless of the temperature, it will still have the same composition. If we start with a quantity of water, divide this and pour out one half, and continue this process a sufficient number of times, we will eventually end up with a quantity of water which cannot be further divided without ceasing to be water. This quantity is called a molecule of water. If this molecule of water divided, instead of two parts of water, there will be one part of oxygen and two parts of hydrogen (H_2O).

ATOMS

Molecules are made up of smaller particles called ATOMS. An atom is the smallest particle of an element that retains the characteristics of that element. The atoms of one element, however, differ from the atoms of all other elements. Since there are over 100 known elements, there must be over 100 different atoms, or a different atom for each element. Just as thousands of words can be made by combining the proper letters of the alphabet, so thousands of different materials can be made by chemically combining the proper atoms.

Any particle that is a chemical combination of two or more atoms is called a molecule. The oxygen molecule consists of two atoms of oxygen, and the hydrogen molecule consists of two atoms of hydrogen. Sugar, on the other hand, is a compound composed of atoms of carbon, hydrogen, and oxygen. These atoms are combined into sugar molecules. Since the sugar molecules can be broken down by chemical means into smaller and simpler units, we cannot have sugar atoms.

The atoms of each element are made up of electrons, protons, and, in most cases, neutrons, which are collectively called subatomic particles. Furthermore, the electrons, protons, and neutrons of one element are identical to those of any other element. The reason that there are different kinds of elements is that the number and the arrangement of electrons and protons within the atom are different for the different elements

The electron is considered to be a small negative charge of electricity. The proton has a positive charge of electricity equal and opposite to the charge of the electron. Scientists have measured the mass and size of the electron and proton, and they know how much charge each possesses. The electron and proton each have the same quantity of charge, although the mass of the proton is approximately 1837 times that of the electron. In some atoms there exists a neutral particle called a neutron. The neutron has a mass approximately equal to that of a proton, but it has no electrical charge. According to a popular theory, the electrons, protons, and neutrons of the atoms are thought to be arranged in a manner similar to a miniature solar system. The protons and neutrons form a heavy nucleus with a positive charge, around which the very light electrons revolve.

Figure 1-1 shows one hydrogen and one helium atom. Each has a relatively simple structure. The hydrogen atom has only one proton in the nucleus with one electron rotating about it. The helium atom is a little more complex. It has a nucleus made up of two protons and two neutrons, with two electrons rotating about the nucleus. Elements are classified numerically according to the complexity of their atoms. The atomic number of an atom is determined by the number of protons in its nucleus.

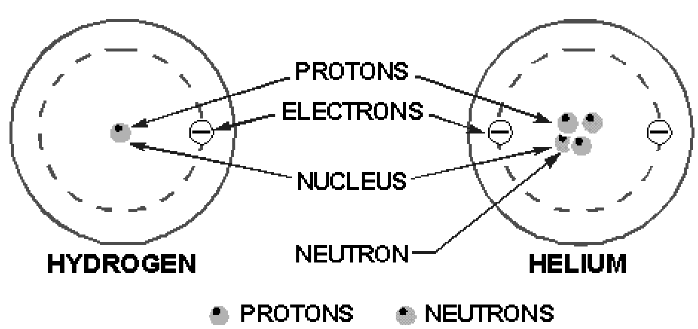


Figure 1-1.—Structures of simple atoms.

In a neutral state, an atom contains an equal number of protons and electrons. Therefore, an atom of hydrogen—which contains one proton and one electron—has an atomic number of 1; and helium, with two protons and two electrons, has an atomic number of 2. The complexity of atomic structure increases with the number of protons and electrons.

Q5. What is a molecule?

Q6. What are the three types of subatomic particles, and what are their charges?

Energy Levels

Since an electron in an atom has both mass and motion, it contains two types of energy. By virtue of its motion the electron contains KINETIC ENERGY. Due to its position it also contains POTENTIAL ENERGY. The total energy contained by an electron (kinetic plus potential) is the factor which determines the radius of the electron orbit. In order for an electron to remain in this orbit, it must neither GAIN nor LOSE energy.

It is well known that light is a form of energy, but the physical form in which this energy exists is not known.

One accepted theory proposes the existence of light as tiny packets of energy called PHOTONS. Photons can contain various quantities of energy. The amount depends upon the color of the light involved. Should a photon of sufficient energy collide with an orbital electron, the electron will absorb the photon's energy, as shown in figure 1-2. The electron, which now has a greater than normal amount of energy, will jump to a new orbit farther from the nucleus. The first new orbit to which the electron can jump has a radius four times as large as the radius of the original orbit. Had the electron received a greater amount of energy, the next possible orbit to which it could jump would have a radius nine times the original. Thus, each orbit may be considered to represent one of a large number of energy levels that the electron may attain. It must be emphasized that the electron cannot jump to just any orbit. The electron will remain in its lowest orbit until a sufficient amount of energy is available, at which time the electron will accept the energy and jump to one of a series of permissible orbits. An electron cannot exist in the space between energy levels. This indicates that the electron will not accept a photon of energy unless it contains enough energy to elevate itself to one of the higher energy levels. Heat energy and collisions with other particles can also cause the electron to jump orbits.

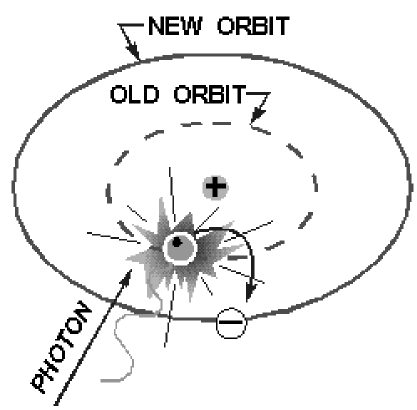


Figure 1-2.—Excitation by a photon.

Once the electron has been elevated to an energy level higher than the lowest possible energy level, the atom is said to be in an excited state. The electron will not remain in this excited condition for more than a fraction of a second before it will radiate the excess energy and return to a lower energy orbit. To illustrate this principle, assume that a normal electron has just received a photon of energy sufficient to raise it from the first to the third energy level. In a short period of time the electron may jump back to the first level emitting a new photon identical to the one it received.

A second alternative would be for the electron to return to the lower level in two jumps; from the third to the second, and then from the second to the first. In this case the electron would emit two photons, one for each jump. Each of these photons would have less energy than the original photon which excited the electron.

This principle is used in the fluorescent light where ultraviolet light photons, which are not visible to the human eye, bombard a phosphor coating on the inside of a glass tube. The phosphor electrons, in returning to their normal orbits, emit photons of light that are visible. By using the proper chemicals for the phosphor coating, any color of light may be obtained, including white. This same principle is also used in lighting up the screen of a television picture tube.

The basic principles just developed apply equally well to the atoms of more complex elements. In atoms containing two or more electrons, the electrons interact with each other and the exact path of any one electron is very difficult to predict. However, each electron lies in a specific energy band and the orbits will be considered as an average of the electron's position.

Q7. What is energy of motion called?

Q8. How is invisible light changed to visible light in a fluorescent light?

Shells and Subshells

The difference between the atoms, insofar as their chemical activity and stability are concerned, is dependent upon the number and position of the electrons included within the atom. How are these electrons positioned within the atom? In general, the electrons reside in groups of orbits called shells. These shells are elliptically shaped and are assumed to be located at fixed intervals. Thus, the shells are arranged in steps that correspond to fixed energy levels. The shells, and the number of electrons required to fill them, may be predicted by the employment of Pauli's exclusion principle. Simply stated, this principle specifies that each shell will contain a maximum of $2n^2$electrons, where n corresponds to the shell number starting with the one closest to the nucleus. By this principle, the second shell, for example, would contain $2(2)^2$ or 8 electrons when full.

In addition to being numbered, the shells are also given letter designations, as pictured in figure 1-3. Starting with the shell closest to the nucleus and progressing outward, the shells are labeled K, L, M, N, O, P, and Q, respectively. The shells are considered to be full, or complete, when they contain the following quantities of electrons: two in the K shell, eight in the L shell, 18 in the M shell, and so on, in accordance with the exclusion principle. Each of these shells is a major shell and can be divided into subshells, of which there are four, labeled s, p, d, and f. Like the major shells, the subshells are also limited as to the number of electrons which they can contain. Thus, the "s" subshell is complete when it contains two electrons, the "p" subshell when it contains 10, and the "f" subshell when it contains 14 electrons.

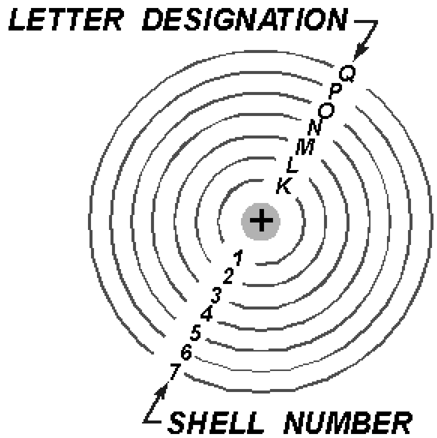


Figure 1-3.—Shell designation.

Inasmuch as the K shell can contain no more than two electrons, it must have only one subshell, the s subshell. The M shell is composed of three subshells: s, p, and d. If the electrons in the s, p, and d subshells are added, their total is found to be 18, the exact number required to fill the M shell. Notice the electron configuration for copper illustrated in figure 1-4. The copper atom contains 29 electrons, which completely fill the first three shells and subshells, leaving one electron in the "s" subshell of the N shell.

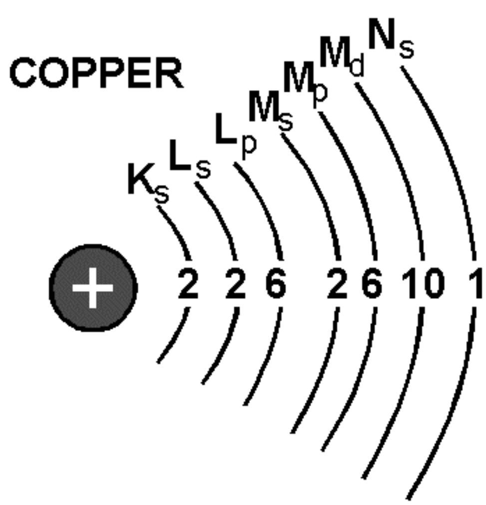


Figure 1-4.—Copper atom.

Valence

The number of electrons in the outermost shell determines the valence of an atom. For this reason, the outer shell of an atom is called the VALENCE SHELL; and the electrons contained in this shell are called VALENCE ELECTRONS. The valence of an atom determines its ability to gain or lose an electron, which in turn determines the chemical and electrical properties of the atom. An atom that is lacking only one or two electrons from its outer shell will easily gain electrons to complete its shell, but a large amount of energy is required to free any of its electrons. An atom having a relatively small number of electrons in its outer shell in comparison to the number of electrons required to fill the shell will easily lose these valence electrons. The valence shell always refers to the outermost shell.

Q9. What determines the valence of an atom?

Ionization

When the atom loses electrons or gains electrons in this process of electron exchange, it is said to be IONIZED. For ionization to take place, there must be a transfer of energy which results in a change in the internal energy of the atom. An atom having more than its normal amount of electrons acquires a negative charge, and is called a NEGATIVE ION. The atom that gives up some of its normal electrons is left with less negative charges than positive charges and is called a POSITIVE ION. Thus, ionization is the process by which an atom loses or gains electrons.

Q10. What is an ion?

CONDUCTORS, SEMICONDUCTORS, AND INSULATORS

In this study of electricity and electronics, the association of matter and electricity is important. Since every electronic device is constructed of parts made from ordinary matter, the effects of electricity on matter must be well understood. As a means of accomplishing this, all elements of which matter is made may be placed into one of three categories: CONDUCTORS, SEMICONDUCTORS, and INSULATORS, depending on their ability to conduct an electric current. CONDUCTORS are elements which conduct electricity very readily, INSULATORS have an extremely high resistance to the flow of electricity. All matter between these two extremes may be called SEMICONDUCTORS.

The electron theory states that all matter is composed of atoms and the atoms are composed of smaller particles called protons, electrons, and neutrons. The electrons orbit the nucleus which contains the protons and neutrons. It is the valence electrons that we are most concerned with in electricity. These are the electrons which are easiest to break loose from their parent atom. Normally, conductors have three or less valence electrons; insulators have five or more valence electrons; and semiconductors usually have four valence electrons.

The electrical conductivity of matter is dependent upon the atomic structure of the material from which the conductor is made. In any solid material, such as copper, the atoms which make up the molecular structure are bound firmly together. At room temperature, copper will contain a considerable amount of heat energy. Since heat energy is one method of removing electrons from their orbits, copper will contain many free electrons that can move from atom to atom. When not under the influence of an external force, these electrons move in a haphazard manner within the conductor. This movement is equal in all directions so that electrons are not lost or gained by any part of the conductor. When controlled by an external force, the electrons move generally in the same direction. The effect of this movement is felt almost instantly from one end of the conductor to the other. This electron movement is called an ELECTRIC CURRENT.

Some metals are better conductors of electricity than others. Silver, copper, gold, and aluminum are materials with many free electrons and make good conductors. Silver is the best conductor, followed by copper, gold, and aluminum. Copper is used more often than silver because of cost. Aluminum is used where weight is a major consideration, such as in high-tension power lines, with long spans between supports. Gold is used where oxidation or corrosion is a consideration and a good conductivity is required. The ability of a conductor to handle current also depends upon its physical dimensions. Conductors are usually found in the form of wire, but may be in the form of bars, tubes, or sheets.

Nonconductors have few free electrons. These materials are called INSULATORS. Some examples of these materials are rubber, plastic, enamel, glass, dry wood, and mica. Just as there is no perfect conductor, neither is there a perfect insulator.

Some materials are neither good conductors nor good insulators, since their electrical characteristics fall between those of conductors and insulators. These in-between materials are classified as SEMICONDUCTORS. Germanium and silicon are two common semiconductors used in solid-state devices.

Q11. What determines whether a substance is a conductor or an insulator?

ELECTROSTATICS

Electrostatics (electricity at rest) is a subject with which most persons entering the field of electricity and electronics are somewhat familiar. For example, the way a person's hair stands on end after a vigorous rubbing is an effect of electrostatics. While pursuing the study of electrostatics, you will gain a better understanding of this common occurrence. Of even greater significance, the study of electrostatics will provide you with the opportunity to gain important background knowledge and to develop concepts which are essential to the understanding of electricity and electronics.

Interest in the subject of static electricity can be traced back to the Greeks. Thales of Miletus, a Greek philosopher and mathematician, discovered that when an amber rod is rubbed with fur, the rod has the amazing characteristic of attracting some very light objects such as bits of paper and shavings of wood.

About 1600, William Gilbert, an English scientist, made a study of other substances which had been found to possess qualities of attraction similar to amber. Among these were glass, when rubbed with silk, and ebonite, when rubbed with fur. Gilbert classified all the substances which possessed properties similar to those of amber as electrics, a word of Greek origin meaning amber.

Because of Gilbert's work with electrics, a substance such as amber or glass when given a vigorous rubbing was recognized as being ELECTRIFIED, or CHARGED with electricity.

In the year 1733, Charles Dufay, a French scientist, made an important discovery about electrification. He found that when a glass was rubbed with fur, both the glass rod and the fur became electrified. This realization came when he systematically placed the glass rod and the fur near other electrified substances and found that certain substances which were attracted to the glass rod were repelled by the fur, and vice versa. From experiments such as this, he concluded that there must be two exactly opposite kinds of electricity.

Benjamin Franklin, American statesman, inventor, and philosopher, is credited with first using the terms POSITIVE and NEGATIVE to describe the two opposite kinds of electricity. The charge produced on a glass rod when it is rubbed with silk, Franklin labeled positive. He attached the term negative to the charge produced on the silk. Those bodies which were not electrified or charged, he called NEUTRAL.

STATIC ELECTRICITY

In a natural, or neutral state, each atom in a body of matter will have the proper number of electrons in orbit around it. Consequently, the whole body of matter composed of the neutral atoms will also be electrically neutral. In this state, it is said to have a "zero charge." Electrons will neither leave nor enter the neutrally charged body should it come in contact with other neutral bodies. If, however, any number of electrons are removed from the atoms of a body of matter, there will remain more protons than electrons and the whole body of matter will become ELECTRICALLY POSITIVE. Should the positively charged body come in contact with another body having a normal charge, or having a NEGATIVE (too many electrons) charge, an electric current will flow between them. Electrons will leave the more negative body and enter the positive body. This electron flow will continue until both bodies have equal charges. When two bodies of matter have unequal charges and are near one another, an electric force is exerted between them because of their unequal charges. However, since they are not in contact, their charges cannot equalize. The existence of such an electric force, where current cannot flow, is referred to as static electricity. ("Static" in this instance means "not moving.") It is also referred to as an electrostatic force.

One of the easiest ways to create a static charge is by friction. When two pieces of matter are rubbed together, electrons can be "wiped off" one material onto the other. If the materials used are good conductors, it is quite difficult to obtain a detectable charge on either, since equalizing currents can flow easily between the conducting materials. These currents equalize the charges almost as fast as they are created. A static charge is more easily created between nonconducting materials. When a hard rubber rod is rubbed with fur, the rod will accumulate electrons given up by the fur, as shown in figure 1-5. Since both materials are poor conductors, very little equalizing current can flow, and an electrostatic charge builds up. When the charge becomes great enough, current will flow regardless of the poor conductivity of the materials. These currents will cause visible sparks and produce a crackling sound.

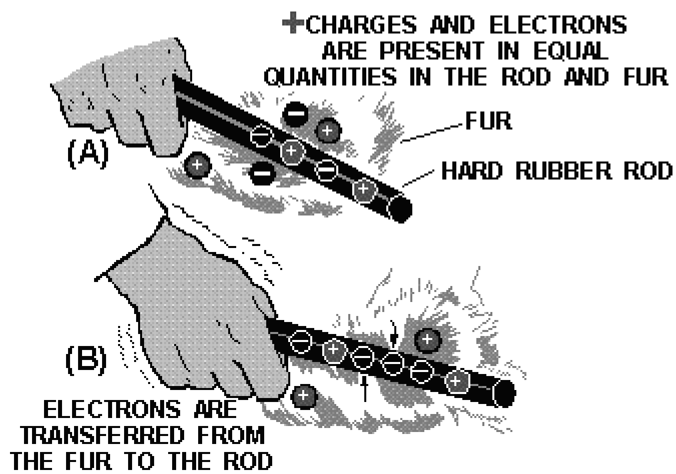


Figure 1-5.—Producing static electricity by friction.

Q12. How is a negative charge created in a neutral body?

Q13. How are static charges created?

Nature of Charges

When in a natural, or neutral state, an atom has an equal number of electrons and protons. Because of this balance, the net negative charge of the electrons in orbit is exactly balanced by the net positive charge of the protons in the nucleus, making the atom electrically neutral.

An atom becomes a positive ion whenever it loses an electron, and has an overall positive charge. Conversely, whenever an atom acquires an extra electron, it becomes a negative ion and has a negative charge.

Due to normal molecular activity, there are always ions present in any material. If the number of positive ions and negative ions is equal, the material is electrically neutral. When the number of positive ions exceeds the number of negative ions, the material is positively charged. The material is negatively charged whenever the negative ions outnumber the positive ions.

Since ions are actually atoms without their normal number of electrons, it is the excess or the lack of electrons in a substance that determines its charge. In most solids, the transfer of charges is by movement of electrons rather than ions. The transfer of charges by ions will become more significant when we consider electrical activity in liquids and gases. At this time, we will discuss electrical behavior in terms of electron movement.

Q14. What is the electrical charge of an atom which contains 8 protons and 11 electrons?

Charged Bodies

One of the fundamental laws of electricity is that LIKE CHARGES REPEL EACH OTHER and UNLIKE CHARGES ATTRACT EACH OTHER. A positive charge and negative charge, being unlike, tend to move toward each other. In the atom, the negative electrons are drawn toward the positive protons in the nucleus. This attractive force is balanced by the electron's centrifugal force caused by its rotation about the nucleus. As a result, the electrons remain in orbit and are not drawn into the nucleus. Electrons repel each other because of their like negative charges, and protons repel each other because of their like positive charges.

The law of charged bodies may be demonstrated by a simple experiment. Two pith (paper pulp) balls are suspended near one another by threads, as shown in figure 1-6.

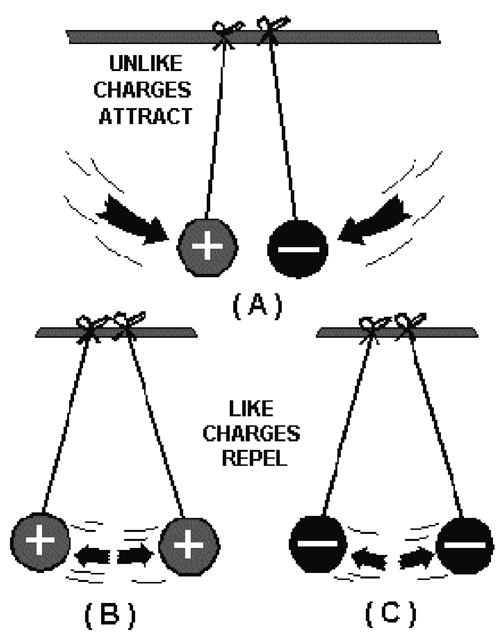


Figure 1-6.—Reaction between charged bodies.

If a hard rubber rod is rubbed with fur to give it a negative charge and is then held against the right-hand ball in part (A), the rod will give off a negative charge to the ball. The right-hand ball will have a negative charge with respect to the left-hand ball. When released, the two balls will be drawn together, as shown in figure 1-6(A). They will touch and remain in contact until the left-hand ball gains a portion of the negative charge of the right-hand ball, at which time they will swing apart as shown in figure 1-6(C). If a positive or a negative charge is placed on both balls (fig. 1-6(B)), the balls will repel each other.

Coulomb's Law of Charges

The relationship between attracting or repelling charged bodies was first discovered and written about by a French scientist named Charles A. Coulomb. Coulomb's Law states that CHARGED BODIES ATTRACT OR REPEL EACH OTHER WITH A FORCE THAT IS DIRECTLY PROPORTIONAL TO THE PRODUCT OF THEIR INDIVIDUAL CHARGES, AND IS INVERSELY PROPORTIONAL TO THE SQUARE OF THE DISTANCE BETWEEN THEM.

The amount of attracting or repelling force which acts between two electrically charged bodies in free space depends on two things—(1) their charges and (2) the distance between them.

Electric Fields

The space between and around charged bodies in which their influence is felt is called an ELECTRIC FIELD OF FORCE. It can exist in air, glass, paper, or a vacuum. ELECTROSTATIC FIELDS and DIELECTRIC FIELDS are other names used to refer to this region of force.

Fields of force spread out in the space surrounding their point of origin and, in general, DIMINISH IN PROPORTION TO THE SQUARE OF THE DISTANCE FROM THEIR SOURCE.

The field about a charged body is generally represented by lines which are referred to as ELECTROSTATIC LINES OF FORCE. These lines are imaginary and are used merely to represent the direction and strength of the field. To avoid confusion, the lines of force exerted by a <u>positive charge</u> are always shown <u>leaving</u> the charge, and for a <u>negative charge</u> they are shown <u>entering</u>. Figure 1-7 illustrates the use of lines to represent the field about charged bodies.

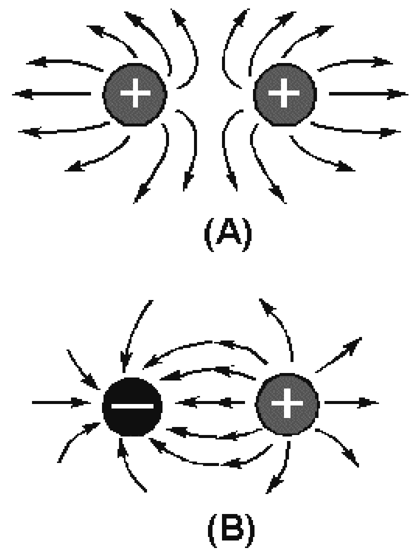


Figure 1-7.—Electrostatic lines of force.

Figure 1-7(A) represents the repulsion of like-charged bodies and their associated fields. Part (B) represents the attraction of unlike-charged bodies and their associated fields.

Q15. What is the relationship between charged bodies?

Q16. What is an electrostatic field?

Q17. In what direction are electrostatic lines of force drawn?

MAGNETISM

In order to properly understand the principles of electricity, it is necessary to study magnetism and the effects of magnetism on electrical equipment. Magnetism and electricity are so closely related that the study of either subject would be incomplete without at least a basic knowledge of the other.

Much of today's modern electrical and electronic equipment could not function without magnetism. Modern computers, tape recorders, and video reproduction equipment use magnetized tape. High-fidelity speakers use magnets to convert amplifier outputs into audible sound. Electrical motors use magnets to convert electrical energy into mechanical motion; generators use magnets to convert mechanical motion into electrical energy.

Q18. What are some examples of electrical equipment which use magnetism?

MAGNETIC MATERIALS

Magnetism is generally defined as that property of a material which enables it to attract pieces of iron. A material possessing this property is known as a MAGNET. The word originated with the ancient Greeks, who found stones possessing this characteristic. Materials that are attracted by a magnet, such as iron, steel, nickel, and cobalt, have the ability to become magnetized. These are called magnetic materials.

Materials, such as paper, wood, glass, or tin, which are not attracted by magnets, are considered nonmagnetic. Nonmagnetic materials are not able to become magnetized.

Q19. What are magnetic materials?

Ferromagnetic Materials

The most important group of materials connected with electricity and electronics are the ferromagnetic materials. Ferromagnetic materials are those which are relatively easy to magnetize, such as iron, steel, cobalt, and the alloys Alnico and Permalloy. (An alloy is made from combining two or more elements, one of which must be a metal). These new alloys can be very strongly magnetized, and are capable of obtaining a magnetic strength great enough to lift 500 times their own weight.

Natural Magnets

Magnetic stones such as those found by the ancient Greeks are considered to be NATURAL MAGNETS. These stones had the ability to attract small pieces of iron in a manner similar to the magnets which are common today. However, the magnetic properties attributed to the stones were products of nature and not the result of the efforts of man. The Greeks called these substances magnetite.

The Chinese are said to have been aware of some of the effects of magnetism as early as 2600 B.C. They observed that stones similar to magnetite, when freely suspended, had a tendency to assume a nearly north and south direction. Because of the directional quality of these stones, they were later referred to as lodestones or leading stones.

Natural magnets, which presently can be found in the United States, Norway, and Sweden, no longer have any practical use, for it is now possible to easily produce more powerful magnets.

Q20. What characteristics do all ferromagnetic materials have in common?

Artificial Magnets

Magnets produced from magnetic materials are called ARTIFICIAL MAGNETS. They can be made in a variety of shapes and sizes and are used extensively in electrical apparatus. Artificial magnets are generally made from special iron or steel alloys which are usually magnetized electrically. The material to be magnetized is inserted into a coil of insulated wire and a heavy flow of electrons is passed through the wire. Magnets can also be produced by stroking a magnetic material with magnetite or with another artificial magnet. The forces causing magnetization are represented by magnetic lines of force, very similar in nature to electrostatic lines of force.

Artificial magnets are usually classified as PERMANENT or TEMPORARY, depending on their ability to retain their magnetic properties after the magnetizing force has been removed. Magnets made from substances, such as hardened steel and certain alloys which retain a great deal of their magnetism, are called PERMANENT MAGNETS. These materials are relatively difficult to magnetize because of the opposition offered to the magnetic lines of force as the lines of force try to distribute themselves throughout the material. The opposition that a material offers to the magnetic lines of force is called RELUCTANCE. All permanent magnets are produced from materials having a high reluctance.

A material with a low reluctance, such as soft iron or annealed silicon steel, is relatively easy to magnetize but will retain only a small part of its magnetism once the magnetizing force is removed. Materials of this type that easily lose most of their magnetic strength are called TEMPORARY MAGNETS. The amount of magnetism which remains in a temporary magnet is referred to as its

RESIDUAL MAGNETISM. The ability of a material to retain an amount of residual magnetism is called the RETENTIVITY of the material.

The difference between a permanent and a temporary magnet has been indicated in terms of RELUCTANCE, a permanent magnet having a high reluctance and a temporary magnet having a low reluctance. Magnets are also described in terms of the PERMEABILITY of their materials, or the ease with which magnetic lines of force distribute themselves throughout the material. A permanent magnet, which is produced from a material with a high reluctance, has a low permeability. A temporary magnet, produced from a material with a low reluctance, would have a high permeability.

Q21. What type of magnetic material should be used to make a temporary magnet?

Q22. What is retentivity?

MAGNETIC POLES

The magnetic force surrounding a magnet is not uniform. There exists a great concentration of force at each end of the magnet and a very weak force at the center. Proof of this fact can be obtained by dipping a magnet into iron filings (fig. 1-8). It is found that many filings will cling to the ends of the magnet while very few adhere to the center. The two ends, which are the regions of concentrated lines of force, are called the POLES of the magnet. Magnets have two magnetic poles and both poles have equal magnetic strength.

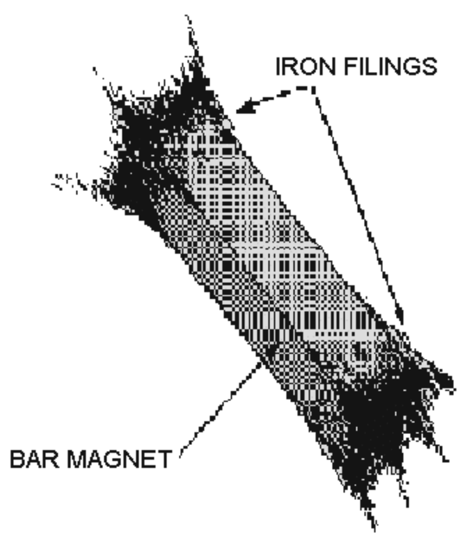


Figure 1-8.—Iron filings cling to the poles of a magnet.

Law of Magnetic Poles

If a bar magnet is suspended freely on a string, as shown in figure 1-9, it will align itself in a north and south direction. When this experiment is repeated, it is found that the same pole of the magnet will always swing toward the north magnetic pole of the earth. Therefore, it is called the north-seeking pole or simply the NORTH POLE. The other pole of the magnet is the south-seeking pole or the SOUTH POLE.

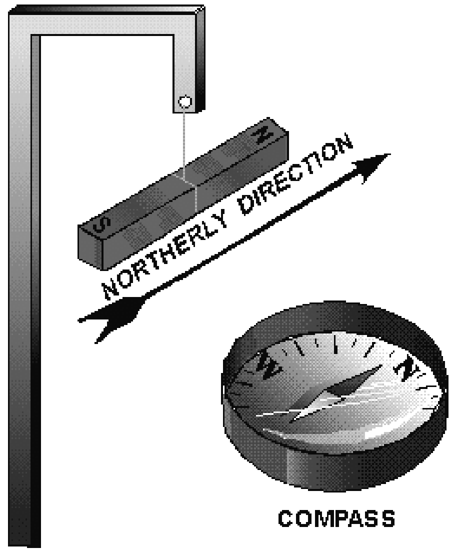


Figure 1-9.—A bar magnet acts as a compass.

A practical use of the directional characteristic of the magnet is the compass, a device in which a freely rotating magnetized needle indicator points toward the North Pole. The realization that the poles of a suspended magnet always move to a definite position gives an indication that the opposite poles of a magnet have opposite magnetic polarity.

The law previously stated regarding the attraction and repulsion of charged bodies may also be applied to magnetism if the pole is considered as a charge. The north pole of a magnet will always be attracted to the south pole of another magnet and will show a repulsion to a north pole. The law for magnetic poles is:

Like poles repel, unlike poles attract.

Q23. How does the law of magnetic poles relate to the law of electric charges?

The Earth's Magnetic Poles

The fact that a compass needle always aligns itself in a particular direction, regardless of its location on earth, indicates that the earth is a huge natural magnet. The distribution of the magnetic force about the earth is the same as that which might be produced by a giant bar magnet running through the center of the earth (fig. 1-10). The magnetic axis of the earth is located about 15º from its geographical axis thereby locating the magnetic poles some distance from the geographical poles. The ability of the north pole of the compass needle to point toward the north geographical pole is due to the presence of the magnetic pole nearby. This magnetic pole is named the magnetic North Pole. However, in actuality, it must have the polarity of a south magnetic pole since it attracts the north pole of a compass needle. The reason for this conflict in terminology can be traced to the early users of the compass. Knowing little about magnetic effects, they called the end of the compass needle that pointed towards the north geographical pole, the north pole of a compass. With our present knowledge of magnetism, we know the north pole of a compass needle (a small bar magnet) can be attracted only by an unlike magnetic pole, that is, a pole of south magnetic polarity.

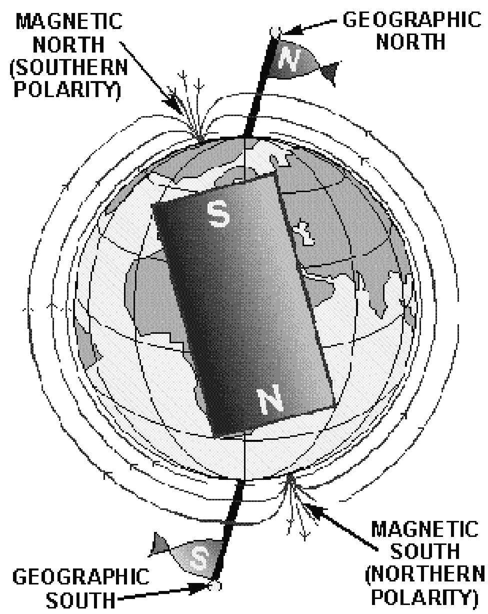


Figure 1-10.—The earth is a magnet.

Q24. A compass is located at the geographical North Pole. In which direction would its needle point?

THEORIES OF MAGNETISM

Weber's Theory

A popular theory of magnetism considers the molecular alignment of the material. This is known as Weber's theory. This theory assumes that all magnetic substances are composed of tiny molecular magnets. Any unmagnetized material has the magnetic forces of its molecular magnets neutralized by adjacent molecular magnets, thereby eliminating any magnetic effect. A magnetized material will have most of its molecular magnets lined up so that the north pole of each molecule points in one direction, and the south pole faces the opposite direction. A material with its molecules thus aligned will then have one effective north pole, and one effective south pole. An illustration of Weber's Theory is shown in figure 1-11, where a steel bar is magnetized by stroking. When a steel bar is stroked several times in the same direction by a magnet, the magnetic force from the north pole of the magnet causes the molecules to align themselves.

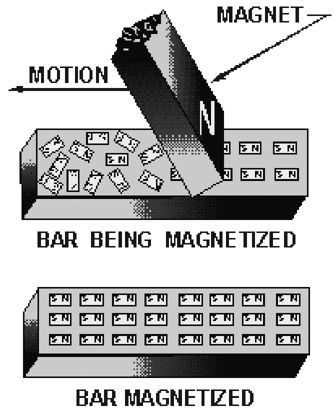


Figure 1-11.—Weber's molecular theory of magnetism.

Q25. Using Weber's molecular theory of magnetism, describe the polarity of the magnetic poles produced by stroking a magnetic material from right to left with the south pole of a magnet.

Domain Theory

A more modern theory of magnetism is based on the electron spin principle. From the study of atomic structure it is known that all matter is composed of vast quantities of atoms, each atom containing one or more orbital electrons. The electrons are considered to orbit in various shells and subshells depending upon their distance from the nucleus. The structure of the atom has previously been compared to the solar system, wherein the electrons orbiting the nucleus correspond to the planets orbiting the sun. Along with its orbital motion about the sun, each planet also revolves on its axis. It is believed that the electron also revolves on its axis as it orbits the nucleus of an atom.

It has been experimentally proven that an electron has a magnetic field about it along with an electric field. The effectiveness of the magnetic field of an atom is determined by the number of electrons spinning in each direction. If an atom has equal numbers of electrons spinning in opposite directions, the magnetic fields surrounding the electrons cancel one another, and the atom is unmagnetized. However, if more electrons spin in one direction than another, the atom is magnetized. An atom with an atomic number of 26, such as iron, has 26 protons in the nucleus and 26 revolving electrons orbiting its nucleus. If 13 electrons are spinning in a clockwise direction and 13 electrons are spinning in a counterclockwise direction, the opposing magnetic fields will be neutralized. When more than 13 electrons spin in either direction, the atom is magnetized. An example of a magnetized atom of iron is shown in figure 1-12.

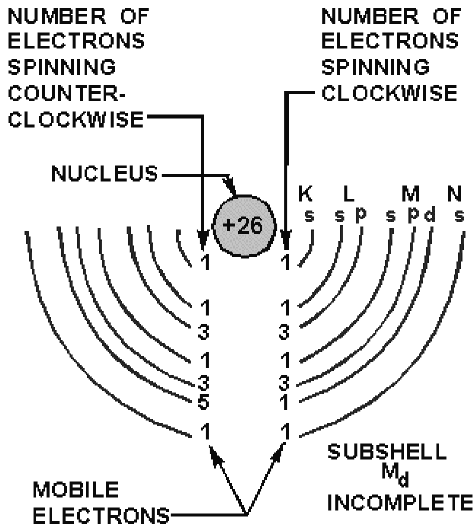


Figure 1-12.—Iron atom.

Q26. What is the difference between the domain theory and Weber's theory of magnetism?

MAGNETIC FIELDS

The space surrounding a magnet where magnetic forces act is known as the magnetic field.

A pattern of this directional force can be obtained by performing an experiment with iron filings. A piece of glass is placed over a bar magnet and the iron filings are then sprinkled on the surface of the glass. The magnetizing force of the magnet will be felt through the glass and each iron filing becomes a temporary magnet. If the glass is now tapped gently, the iron particles will align themselves with the magnetic field surrounding the magnet just as the compass needle did previously. The filings form a definite pattern, which is a visible representation of the forces comprising the magnetic field. Examination of the arrangements of iron filings in figure 1-13 will indicate that the magnetic field is very strong at the poles and weakens as the distance from the poles increases. It is also apparent that the magnetic field extends from one pole to the other, constituting a loop about the magnet.

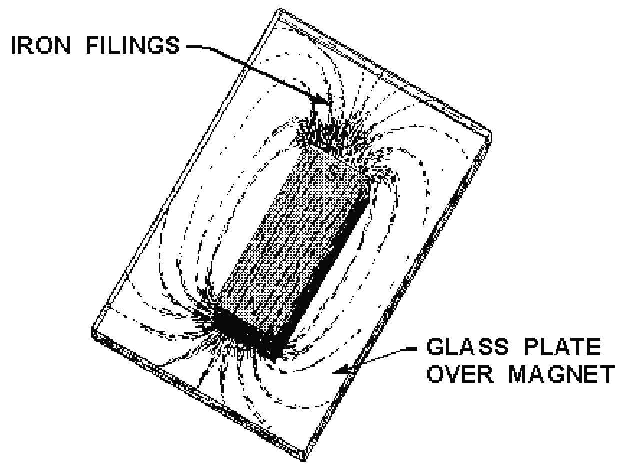


Figure 1-13.—Pattern formed by iron filings.

Q27. Refer to figure 1-13. For what purpose would you sprinkle iron filings on the glass plate?

Q28. Refer to figure 1-13. What pattern would be formed if sawdust was sprinkled on the glass instead of iron filings?

Lines of Force

To further describe and work with magnet phenomena, lines are used to represent the force existing in the area surrounding a magnet (refer to fig. 1-14). These lines, called MAGNETIC LINES OF FORCE, do not actually exist but are imaginary lines used to illustrate and describe the pattern of the magnetic field. The magnetic lines of force are assumed to emanate from the north pole of a magnet, pass through surrounding space, and enter the south pole. The lines of force then travel inside the magnet from the south pole to the north pole, thus completing a closed loop.

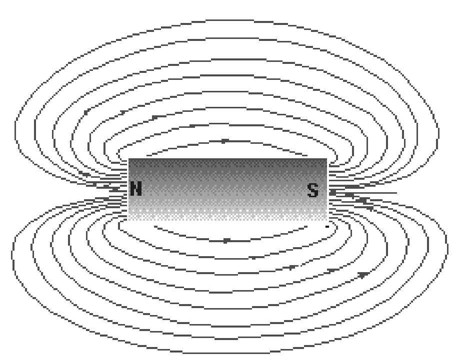


Figure 1-14.—Bar magnet showing lines of force.

When two magnetic poles are brought close together, the mutual attraction or repulsion of the poles produces a more complicated pattern than that of a single magnet. These magnetic lines of force can be plotted by placing a compass at various points throughout the magnetic field, or they can be roughly illustrated by the use of iron filings as before. A diagram of magnetic poles placed close together is shown in figure 1-15.

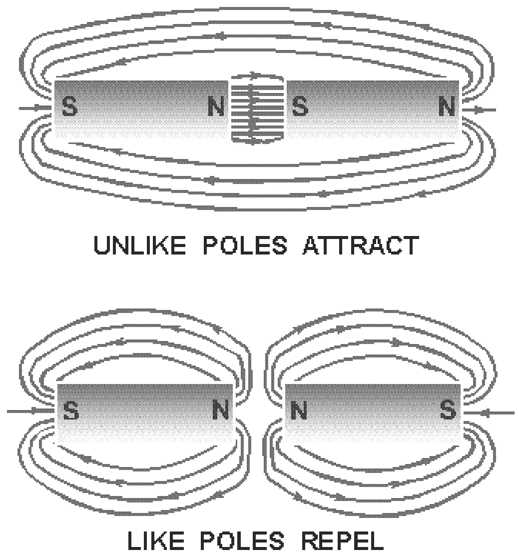


Figure 1-15.—Magnetic poles in close proximity.

Although magnetic lines of force are imaginary, a simplified version of many magnetic phenomena can be explained by assuming the magnetic lines to have certain real properties. The lines of force can be compared to rubber bands which stretch outward when a force is exerted upon them and contract when the force is removed. The characteristics of magnetic lines of force can be described as follows:

1. Magnetic lines of force are continuous and will always form closed loops.
2. Magnetic lines of force will never cross one another.
3. Parallel magnetic lines of force traveling in the same direction repel one another. Parallel magnetic lines of force traveling in opposite directions tend to unite with each other and form into single lines traveling in a direction determined by the magnetic poles creating the lines of force.
4. Magnetic lines of force tend to shorten themselves. Therefore, the magnetic lines of force existing between two unlike poles cause the poles to be pulled together.
5. Magnetic lines of force pass through all materials, both magnetic and nonmagnetic.
6. Magnetic lines of force always enter or leave a magnetic material at right angles to the surface.

Q29. What is a magnetic line of force?

Q30. In what way do magnetic lines of force differ from electrostatic lines of force?

MAGNETIC EFFECTS

MAGNETIC FLUX. The total number of magnetic lines of force leaving or entering the pole of a magnet is called MAGNETIC FLUX. The number of flux lines per unit area is known as FLUX DENSITY.

FIELD INTENSITY. The intensity of a magnetic field is directly related to the magnetic force exerted by the field.

ATTRACTION/REPULSION. The intensity of attraction or repulsion between magnetic poles may be described by a law almost identical to Coulomb's Law of Charged Bodies. The force between two poles is directly proportional to the product of the pole strengths and inversely proportional to the square of the distance between the poles.

Magnetic Induction

It has been previously stated that all substances that are attracted by a magnet are capable of becoming magnetized. The fact that a material is attracted by a magnet indicates the material must itself be a magnet at the time of attraction.

With the knowledge of magnetic fields and magnetic lines of force developed up to this point, it is simple to understand the manner in which a material becomes magnetized when brought near a magnet. As an iron nail is brought close to a bar magnet (fig. 1-16), some flux lines emanating from the north pole of the magnet pass through the iron nail in completing their magnetic path. Since magnetic lines of force travel inside a magnet from the south pole to the north pole, the nail will be magnetized in such a polarity that its south pole will be adjacent to the north pole of the bar magnet. There is now an attraction between the two magnets.

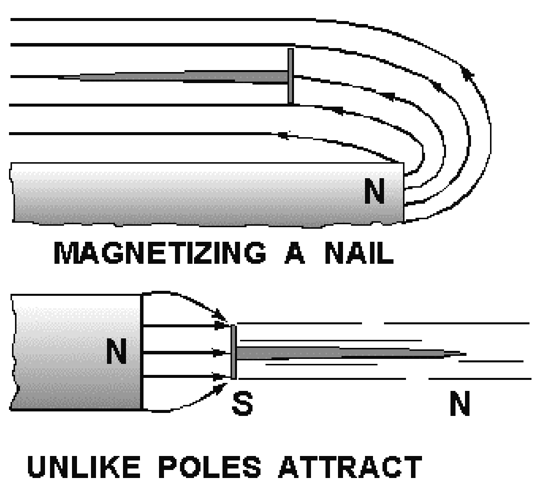


Figure 1-16.—Magnetized nail.

If another nail is brought in contact with the end of the first nail, it would be magnetized by induction. This process could be repeated until the strength of the magnetic flux weakens as distance from the bar magnet increases. However, as soon as the first iron nail is pulled away from the bar magnet, all the nails will fall. The reason being that each nail becomes a temporary magnet, and as soon as the magnetizing force is removed, their domains once again assume a random distribution.

Magnetic induction will always produce a pole polarity on the material being magnetized opposite that of the adjacent pole of the magnetizing force. It is sometimes possible to bring a weak north pole of a magnet near a strong magnet north pole and note attraction between the poles. The weak magnet, when placed within the magnetic field of the strong magnet, has its magnetic polarity reversed by the field of the stronger magnet. Therefore, it is attracted to the opposite pole. For this reason, you must keep a very weak magnet, such as a compass needle, away from a strong magnet.

Magnetism can be induced in a magnetic material by several means. The magnetic material may be placed in the magnetic field, brought into contact with a magnet, or stroked by a magnet. Stroking and contact both indicate actual contact with the material but are considered in magnetic studies as magnetizing by INDUCTION.

Magnetic Shielding

There is no known INSULATOR for magnetic flux. If a nonmagnetic material is placed in a magnetic field, there is no appreciable change in flux—that is, the flux penetrates the nonmagnetic material. For example, a glass plate placed between the poles of a horseshoe magnet will have no appreciable effect on the field although glass itself is a good insulator in an electric circuit. If a magnetic material (for example, soft iron) is placed in a magnetic field, the flux may be redirected to take advantage of the greater permeability of the magnetic material, as shown in figure 1-17. Permeability, as discussed earlier, is the quality of a substance which determines the ease with which it can be magnetized.

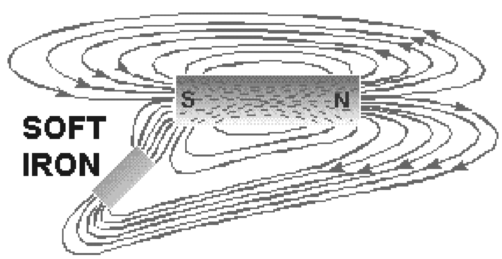


Figure 1-17.—Effects of a magnetic substance in a magnetic field.

The sensitive mechanisms of electric instruments and meters can be influenced by stray magnetic fields which will cause errors in their readings. Because instrument mechanisms cannot be insulated against magnetic flux, it is necessary to employ some means of directing the flux around the instrument. This is accomplished by placing a soft-iron case, called a MAGNETIC SCREEN or SHIELD, about the instrument. Because the flux is established more readily through the iron (even though the path is longer) than through the air inside the case, the instrument is effectively shielded, as shown by the watch and soft-iron shield in figure 1-18.

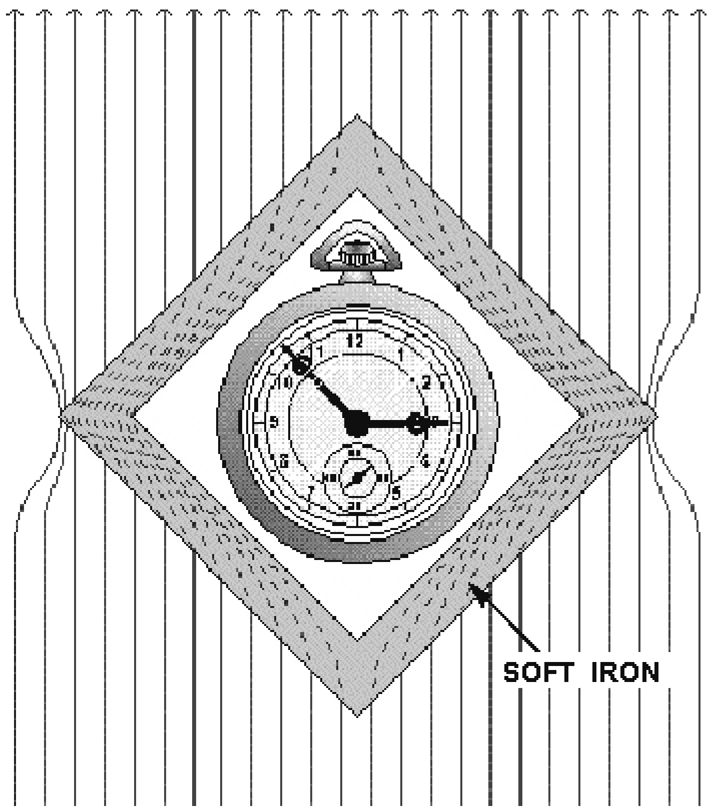


Figure 1-18.—Magnetic shield.

MAGNETIC SHAPES

Because of the many uses of magnets, they are found in various shapes and sizes. However, magnets usually come under one of three general classifications: bar magnets, horseshoe magnets, or ring magnets.

The bar magnet is most often used in schools and laboratories for studying the properties and effects of magnetism. In the preceding material, the bar magnet proved very helpful in demonstrating magnetic effects.

Another type of magnet is the ring magnet, which is used for computer memory cores. A common application for a temporary ring magnet would be the shielding of electrical instruments.

The shape of the magnet most frequently used in electrical and electronic equipment is called the horseshoe magnet. A horseshoe magnet is similar to a bar magnet but is bent in the shape of a horseshoe. The horseshoe magnet provides much more magnetic strength than a bar magnet of the same size and material because of the closeness of the magnetic poles. The magnetic strength from one pole to the other is greatly increased due to the concentration of the magnetic field in a smaller area. Electrical measuring devices quite frequently use horseshoe-type magnets.

CARE OF MAGNETS

A piece of steel that has been magnetized can lose much of its magnetism by improper handling. If it is jarred or heated, there will be a disalignment of its domains resulting in the loss of some of its effective magnetism. Had this piece of steel formed the horseshoe magnet of a meter, the meter would no longer be operable or would give inaccurate readings. Therefore, care must be exercised when handling instruments containing magnets. Severe jarring or subjecting the instrument to high temperatures will damage the device.

A magnet may also become weakened from loss of flux. Thus when storing magnets, one should always try to avoid excess leakage of magnetic flux. A horseshoe magnet should always be stored with a keeper, a soft iron bar used to join the magnetic poles. By using the keeper while the magnet is being stored, the magnetic flux will continuously circulate through the magnet and not leak off into space.

When bar magnets are stored, the same principle must be remembered. Therefore, bar magnets should always be stored in pairs with a north pole and a south pole placed together. This provides a complete path for the magnetic flux without any flux leakage.

Q31. How should a delicate instrument be protected from a magnetic field?

Q32. How should bar magnets be stored?

ELECTRICAL ENERGY

In the field of physical science, work must be defined as the PRODUCT OF FORCE AND DISPLACEMENT. That is, the force applied to move an object and the distance the object is moved are the factors of work performed.

It is important to notice that no work is accomplished unless the force applied causes a change in the position of a stationary object, or a change in the velocity of a moving object. A worker may tire by pushing against a heavy wooden crate, but unless the crate moves, no work will be accomplished.

ENERGY

In our study of energy and work, we must define energy as THE ABILITY TO DO WORK. In order to perform any kind of work, energy must be expended (converted from one form to another). Energy supplies the required force, or power, whenever any work is accomplished.

One form of energy is that which is contained by an object in motion. When a hammer is set in motion in the direction of a nail, it possesses energy of motion. As the hammer strikes the nail, the energy of motion is converted into work as the nail is driven into the wood. The distance the nail is driven into the wood depends on the velocity of the hammer at the time it strikes the nail. Energy contained by an object due to its motion is called KINETIC ENERGY. Assume that the hammer is suspended by a string in a position one meter above a nail. As a result of gravitational attraction, the hammer will experience a force pulling it downward. If the string is suddenly cut, the force of gravity will pull the hammer downward against the nail, driving it into the wood. While the hammer is suspended above the nail it has ability to do work because of its elevated position in the earth’s gravitational field. Since energy is the ability to do work, the hammer contains energy.

Energy contained by an object due to its position is called POTENTIAL ENERGY. The amount of potential energy available is equal to the product of the force required to elevate the hammer and the height to which it is elevated.

Another example of potential energy is that contained in a tightly coiled spring. The amount of energy released when the spring unwinds depends on the amount of force required to wind the spring initially.

Q33. What is the definition of energy?

Q34. What type of energy does a rolling stone have?

Q35. What kind of energy does the stone have if it is at rest at the top of a hill?

Electrical Charges

From the previous study of electrostatics, you learned that a field of force exists in the space surrounding any electrical charge. The strength of the field is directly dependent on the force of the charge.

The charge of one electron might be used as a unit of electrical charge, since charges are created by displacement of electrons; but the charge of one electron is so small that it is impractical to use. The practical unit adopted for measuring charges is the COULOMB, named after the scientist Charles Coulomb. One coulomb is equal to the charge of 6,280,000,000,000,000,000 (six quintillion two hundred and eighty quadrillion) or (6.28×10^{18}) electrons.

When a charge of one coulomb exists between two bodies, one unit of electrical potential energy exists, which is called the <u>difference of potential</u> between the two bodies. This is referred to as ELECTROMOTIVE FORCE, or VOLTAGE, and the unit of measure is the VOLT.

Electrical charges are created by the displacement of electrons, so that there exists an excess of electrons at one point, and a deficiency at another point. Consequently, a charge must always have either a negative or positive polarity. A body with an excess of electrons is considered to be negative, whereas a body with a deficiency of electrons is positive.

A difference of potential can exist between two points, or bodies, only if they have different charges. In other words, there is no difference in potential between two bodies if both have a deficiency of electrons to the same degree. If, however, one body is deficient of 6 coulombs (representing 6 volts), and the other is deficient by 12 coulombs (representing 12 volts), there is a difference of potential of 6 volts. The body with the greater deficiency is positive with respect to the other.

In most electrical circuits only the difference of potential between two points is of importance and the absolute potentials of the points are of little concern. Very often it is convenient to use one standard reference for all of the various potentials throughout a piece of equipment. For this reason, the potentials at various points in a circuit are generally measured with respect to the metal chassis on which all parts of the circuit are mounted. The chassis is considered to be at zero potential and all other potentials are either positive or negative with respect to the chassis. When used as the reference point, the chassis is said to be at GROUND POTENTIAL.

Occasionally, rather large values of voltage may be encountered, in which case the volt becomes too small a unit for convenience. In a situation of this nature, the kilovolt (kV), meaning 1,000 volts, is frequently used. As an example, 20,000 volts would be written as 20 kV. In other cases, the volt may be too large a unit, as when dealing with very small voltages. For this purpose the millivolt (mV), meaning one-thousandth of a volt, and the microvolt (μV), meaning one-millionth of a volt, are used. For example, 0.001 volt would be written as 1 mV, and 0.000025 volt would be written as 25 μV. See Appendix II for exponential symbology.

When a difference in potential exists between two charged bodies that are connected by a conductor, electrons will flow along the conductor. This flow is from the negatively charged body to the positively charged body, until the two charges are equalized and the potential difference no longer exists.

An analogy of this action is shown in the two water tanks connected by a pipe and valve in figure 1-19. At first the valve is closed and all the water is in tank A. Thus, the water pressure across the valve is at maximum. When the valve is opened, the water flows through the pipe from A to B until the water level becomes the same in both tanks. The water then stops flowing in the pipe, because there is no longer a difference in water pressure between the two tanks.

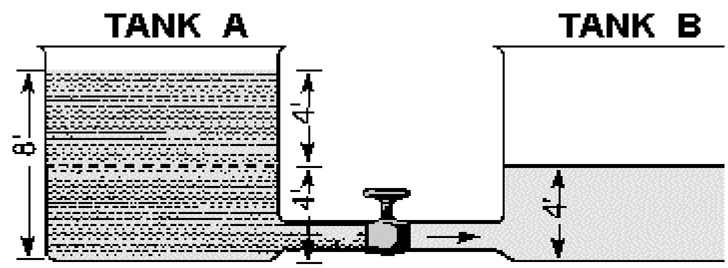


Figure 1-19.—Water analogy of electric differences of potential.

Electron movement through an electric circuit is directly proportional to the difference in potential or electromotive force (emf), across the circuit, just as the flow of water through the pipe in figure 1-19 is directly proportional to the difference in water level in the two tanks.

A fundamental law of electricity is that the ELECTRON FLOW IS DIRECTLY PROPORTIONAL TO THE APPLIED VOLTAGE. If the voltage is increased, the flow is increased. If the voltage is decreased, the flow is decreased.

Q36. What term describes voltage or emf?

Q37. Convert 2.1 kV to volts.

Q38. Express the following in more simple terms. (a) 250,000 volts, (b) 25,000,000 microvolts, (c) 0.001 millivolt.

HOW VOLTAGE IS PRODUCED

It has been demonstrated that a charge can be produced by rubbing a rubber rod with fur. Because of the friction involved, the rod acquires electrons from the fur, making it negative; the fur becomes positive due to the loss of electrons. These quantities of charge constitute a difference of potential between the rod and the fur. The electrons which make up this difference of potential are capable of doing work if a discharge is allowed to occur.

To be a practical source of voltage, the potential difference must not be allowed to dissipate, but must be maintained continuously. As one electron leaves the concentration of negative charge, another must be immediately provided to take its place or the charge will eventually diminish to the point where no further work can be accomplished. A VOLTAGE SOURCE, therefore, is a device which is capable of supplying and maintaining voltage while some type of electrical apparatus is connected to its terminals. The internal action of the source is such that electrons are continuously removed from one terminal, keeping it positive, and simultaneously supplied to the second terminal which maintains a negative charge.

Presently, there are six known methods for producing a voltage or electromotive force (emf). Some of these methods are more widely used than others, and some are used mostly for specific applications. Following is a list of the six known methods of producing a voltage.

1. FRICTION—Voltage produced by rubbing certain materials together.
2. PRESSURE (piezoelectricity)—Voltage produced by squeezing crystals of certain substances.
3. HEAT (thermoelectricity)—Voltage produced by heating the joint (junction) where two unlike metals are joined.
4. LIGHT (photoelectricity)—Voltage produced by light striking photosensitive (light sensitive) substances.
5. CHEMICAL ACTION—Voltage produced by chemical reaction in a battery cell.
6. MAGNETISM—Voltage produced in a conductor when the conductor moves through a magnetic field, or a magnetic field moves through the conductor in such a manner as to cut the magnetic lines of force of the field.

Voltage Produced by Friction

The first method discovered for creating a voltage was that of generation by friction. The development of charges by rubbing a rod with fur is a prime example of the way in which a voltage is generated by friction. Because of the nature of the materials with which this voltage is generated, it cannot be conveniently used or maintained. For this reason, very little practical use has been found for voltages generated by this method.

In the search for methods to produce a voltage of a larger amplitude and of a more practical nature, machines were developed in which charges were transferred from one terminal to another by means of rotating glass discs or moving belts. The most notable of these machines is the Van de Graaff generator. It is used today to produce potentials in the order of millions of volts for nuclear research. As these machines have little value outside the field of research, their theory of operation will not be described here.

Q39. A device which supplies a voltage is commonly referred to by what name?

Voltage Produced by Pressure

One specialized method of generating an emf utilizes the characteristics of certain ionic crystals such as quartz, Rochelle salts, and tourmaline. These crystals have the remarkable ability to generate a voltage whenever stresses are applied to their surfaces. Thus, if a crystal of quartz is squeezed, charges of opposite polarity will appear on two opposite surfaces of the crystal. If the force is reversed and the crystal is stretched, charges will again appear, but will be of the opposite polarity from those produced by squeezing. If a crystal of this type is given a vibratory motion, it will produce a voltage of reversing polarity between two of its sides. Quartz or similar crystals can thus be used to convert mechanical energy into electrical energy. This phenomenon, called the PIEZOELECTRIC EFFECT, is shown in figure 1-20. Some of the common devices that make use of piezoelectric crystals are microphones, phonograph cartridges, and oscillators used in radio transmitters, radio receivers, and sonar equipment. This method of generating an emf is not suitable for applications having large voltage or power requirements, but is widely used in sound and communications systems where small signal voltages can be effectively used.

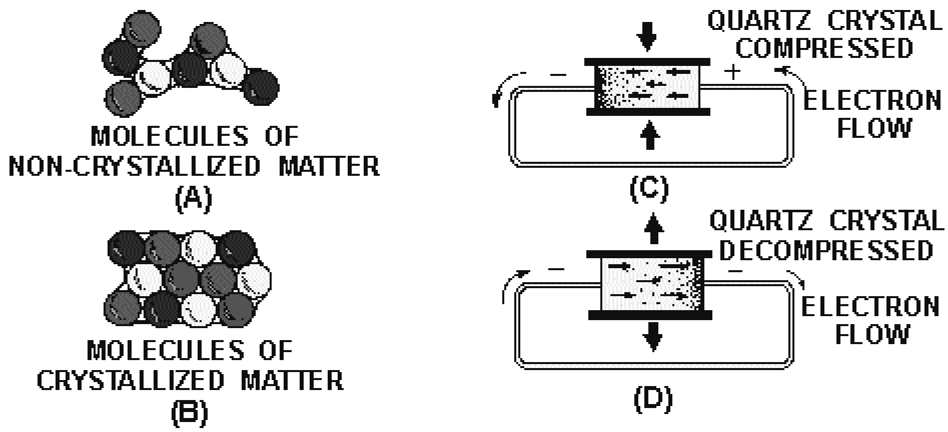


Figure 1-20.—(A) Noncrystallized structure; (B) crystallized structure; (C) compression of a crystal; (D) decompression of a crystal.

Crystals of this type also possess another interesting property, the "converse piezoelectric effect." That is, they have the ability to convert electrical energy into mechanical energy. A voltage impressed across the proper surfaces of the crystal will cause it to expand or contract its surfaces in response to the voltage applied.

Voltage Produced by Heat

When a length of metal, such as copper, is heated at one end, electrons tend to move away from the hot end toward the cooler end. This is true of most metals. However, in some metals, such as iron, the opposite takes place and electrons tend to move TOWARD the hot end. These characteristics are illustrated in figure 1-21. The negative charges (electrons) are moving through the copper away from the heat and through the iron toward the heat. They cross from the iron to the copper through the current meter to the iron at the cold junction. This device is generally referred to as a THERMOCOUPLE.

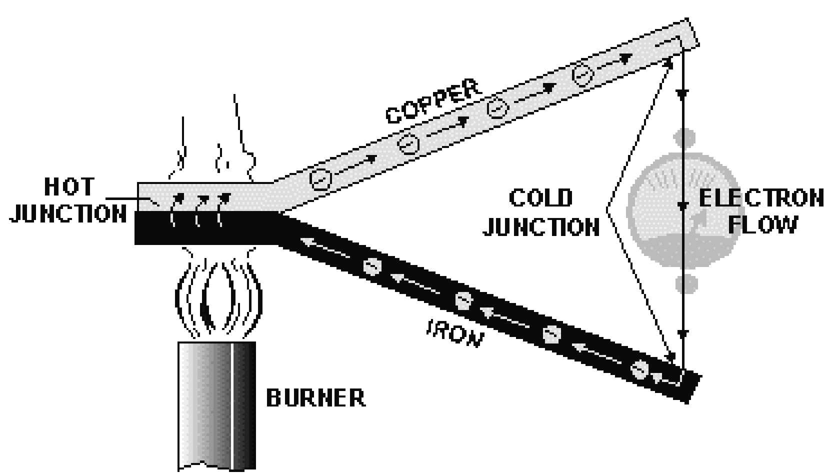


Figure 1-21.—Voltage produced by heat.

Thermocouples have somewhat greater power capacities than crystals, but their capacity is still very small if compared to some other sources. The thermoelectric voltage in a thermocouple depends mainly on the difference in temperature between the hot and cold junctions. Consequently, they are widely used to measure temperature, and as heat-sensing devices in automatic temperature control equipment. Thermocouples generally can be subjected to much greater temperatures than ordinary thermometers, such as the mercury or alcohol types.

Voltage Produced by Light

When light strikes the surface of a substance, it may dislodge electrons from their orbits around the surface atoms of the substance. This occurs because light has energy, the same as any moving force.

Some substances, mostly metallic ones, are far more sensitive to light than others. That is, more electrons will be dislodged and emitted from the surface of a highly sensitive metal, with a given amount of light, than will be emitted from a less sensitive substance. Upon losing electrons, the photosensitive (light-sensitive) metal becomes positively charged, and an electric force is created. Voltage produced in this manner is referred to as a PHOTOELECTRIC VOLTAGE.

The photosensitive materials most commonly used to produce a photoelectric voltage are various compounds of silver oxide or copper oxide. A complete device which operates on the photoelectric principle is referred to as a "photoelectric cell." There are many different sizes and types of photoelectric cells in use, and each serves the special purpose for which it is designed. Nearly all, however, have some of the basic features of the photoelectric cells shown in figure 1-22.

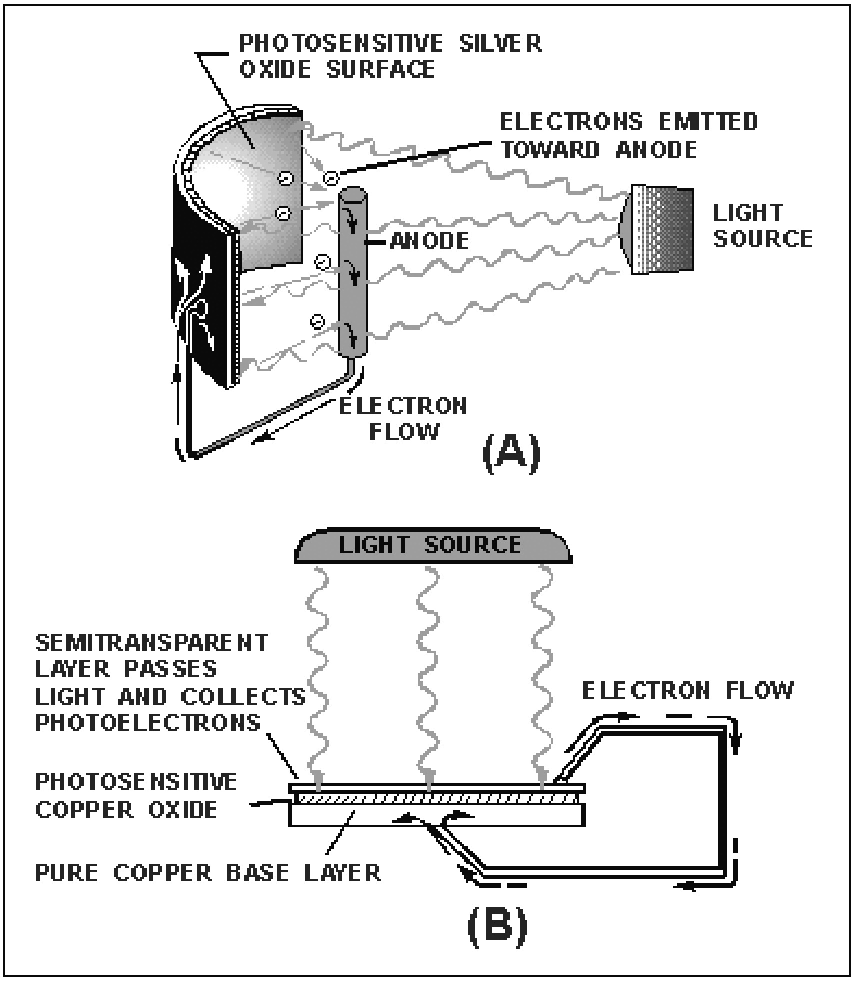


Figure 1-22.—Voltage produced by light.

The cell (fig. 1-22 view A) has a curved light-sensitive surface focused on the central anode. When light from the direction shown strikes the sensitive surface, it emits electrons toward the anode. The more intense the light, the greater the number of electrons emitted. When a wire is connected between the filament and the back, or dark side of the cell, the accumulated electrons will flow to the dark side. These electrons will eventually pass through the metal of the reflector and replace the electrons leaving the light-sensitive surface. Thus, light energy is converted to a flow of electrons, and a usable current is developed.

The cell (fig. 1-22 view B) is constructed in layers. A base plate of pure copper is coated with light-sensitive copper oxide. An extremely thin semitransparent layer of metal is placed over the copper oxide. This additional layer serves two purposes:

1. It permits the penetration of light to the copper oxide.
2. It collects the electrons emitted by the copper oxide.

An externally connected wire completes the electron path, the same as in the reflector-type cell. The photocell's voltage is used as needed by connecting the external wires to some other device, which amplifies (enlarges) it to a usable level.

The power capacity of a photocell is very small. However, it reacts to light-intensity variations in an extremely short time. This characteristic makes the photocell very useful in detecting or accurately controlling a great number of operations. For instance, the photoelectric cell, or some form of the photoelectric principle, is used in television cameras, automatic manufacturing process controls, door openers, burglar alarms, and so forth.

Voltage Produced by Chemical Action

Voltage may be produced chemically when certain substances are exposed to chemical action.

If two dissimilar substances (usually metals or metallic materials) are immersed in a solution that produces a greater chemical action on one substance than on the other, a difference of potential will exist between the two. If a conductor is then connected between them, electrons will flow through the conductor to equalize the charge. This arrangement is called a primary cell. The two metallic pieces are called electrodes and the solution is called the electrolyte. The voltaic cell illustrated in figure 1-23 is a simple example of a primary cell. The difference of potential results from the fact that material from one or both of the electrodes goes into solution in the electrolyte, and in the process, ions form in the vicinity of the electrodes. Due to the electric field associated with the charged ions, the electrodes acquire charges.

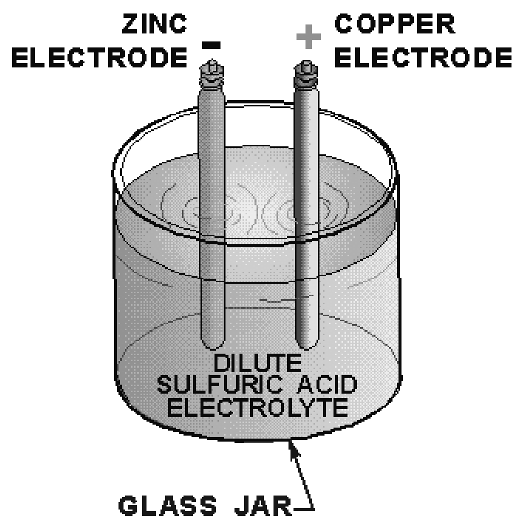


Figure 1-23.—Voltaic cell.

The amount of difference in potential between the electrodes depends principally on the metals used. The type of electrolyte and the size of the cell have little or no effect on the potential difference produced.

There are two types of primary cells, the wet cell and the dry cell. In a wet cell the electrolyte is a liquid. A cell with a liquid electrolyte must remain in an upright position and is not readily transportable. The dry cell, much more commonly used than the wet cell, is not actually dry, but contains an electrolyte mixed with other materials to form a paste. Flashlights and portable radios are commonly powered by dry cells.

Batteries are formed when several cells are connected together to increase electrical output.

Voltage Produced by Magnetism

Magnets or magnetic devices are used for thousands of different jobs. One of the most useful and widely employed applications of magnets is in the production of vast quantities of electric power from mechanical sources. The mechanical power may be provided by a number of different sources, such as gasoline or diesel engines, and water or steam turbines. However, the final conversion of these source energies to electricity is done by generators employing the principle of electromagnetic induction. These generators, of many types and sizes, are discussed in other modules in this series. The important subject to be discussed here is the fundamental operating principle of ALL such electromagnetic-induction generators.

To begin with, there are three fundamental conditions which must exist before a voltage can be produced by magnetism.

1. There must be a CONDUCTOR in which the voltage will be produced.
2. There must be a MAGNETIC FIELD in the conductor's vicinity.
3. There must be relative motion between the field and conductor. The conductor must be moved so as to cut across the magnetic lines of force, or the field must be moved so that the lines of force are cut by the conductor.

In accordance with these conditions, when a conductor or conductors MOVE ACROSS a magnetic field so as to cut the lines of force, electrons WITHIN THE CONDUCTOR are propelled in one direction or another. Thus, an electric force, or voltage, is created.

In figure 1-24, note the presence of the three conditions needed for creating an induced voltage.

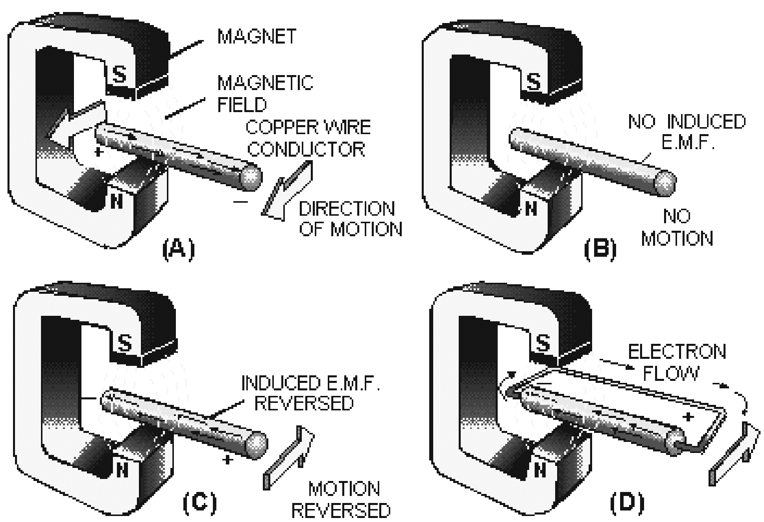


Figure 1-24.—Voltage produced by magnetism.

1. A magnetic field exists between the poles of the C-shaped magnet.
2. There is a conductor (copper wire).
3. There is a relative motion. The wire is moved back and forth ACROSS the magnetic field.

In figure 1-24 view A, the conductor is moving TOWARD the front of the page and the electrons move from left to right. The movement of the electrons occurs because of the magnetically induced emf acting on the electrons in the copper. The right-hand end becomes negative, and the left-hand end positive. The conductor is stopped at view B, motion is eliminated (one of the three required conditions), and there is no longer an induced emf. Consequently, there is no longer any difference in potential between the two ends of the wire. The conductor at view C is moving away from the front of the page. An induced emf is again created. However, note carefully that the REVERSAL OF MOTION has caused a REVERSAL OF DIRECTION in the induced emf.

If a path for electron flow is provided between the ends of the conductor, electrons will leave the negative end and flow to the positive end. This condition is shown in part view D. Electron flow will continue as long as the emf exists. In studying figure 1-24, it should be noted that the induced emf could also have been created by holding the conductor stationary and moving the magnetic field back and forth.

The more complex aspects of power generation by use of mechanical motion and magnetism are discussed later in this series, under the heading "Generators and Motors."

Q40. Name the six methods of producing a voltage.

Q41. The piezoelectric effect is an example of a voltage being produced by what method?

Q42. A thermocouple is a device that produces voltage by what method?

Q43. A battery uses what method to produce a voltage?

Q44. A generator uses what method to produce a voltage?

ELECTRIC CURRENT

It has been proven that electrons (negative charges) move through a conductor in response to an electric field. ELECTRON CURRENT FLOW will be used throughout this explanation. Electron current is defined as the directed flow of electrons. The direction of electron movement is from a region of negative potential to a region of positive potential. Therefore electric current can be said to flow from negative to positive. The direction of current flow in a material is determined by the polarity of the applied voltage. NOTE: In some electrical/electronic communities, the direction of current flow is recognized as being from positive to negative.

Q45. According to electron theory, an electric current flows from what potential to what potential?

Random Drift

All materials are composed of atoms, each of which is capable of being ionized. If some form of energy, such as heat, is applied to a material, some electrons acquire sufficient energy to move to a higher energy level. As a result, some electrons are freed from their parent atom's which then becomes ions. Other forms of energy, particularly light or an electric field, will cause ionization to occur.

The number of free electrons resulting from ionization is dependent upon the quantity of energy applied to a material, as well as the atomic structure of the material. At room temperature some materials, classified as conductors, have an abundance of free electrons. Under a similar condition, materials classified as insulators have relatively few free electrons.

In a study of electric current, conductors are of major concern. Conductors are made up of atoms that contain loosely bound electrons in their outer orbits. Due to the effects of increased energy, these outermost electrons frequently break away from their atoms and freely drift throughout the material. The free electrons, also called mobile electrons, take a path that is not predictable and drift about the material in a haphazard manner. Consequently such a movement is termed RANDOM DRIFT.

It is important to emphasize that the random drift of electrons occurs in all materials. The degree of random drift is greater in a conductor than in an insulator.

Directed Drift

Associated with every charged body there is an electrostatic field. Bodies that are charged alike repel one another and bodies with unlike charges attract each other. An electron will be affected by an electrostatic field in exactly the same manner as any negatively charged body. It is repelled by a negative charge and attracted by a positive charge. If a conductor has a difference in potential impressed across it, as shown in figure 1-25, a direction is imparted to the random drift. This causes the mobile electrons to be repelled away from the negative terminal and attracted toward the positive terminal. This constitutes a general migration of electrons from one end of the conductor to the other. The directed migration of mobile electrons due to the potential difference is called DIRECTED DRIFT.

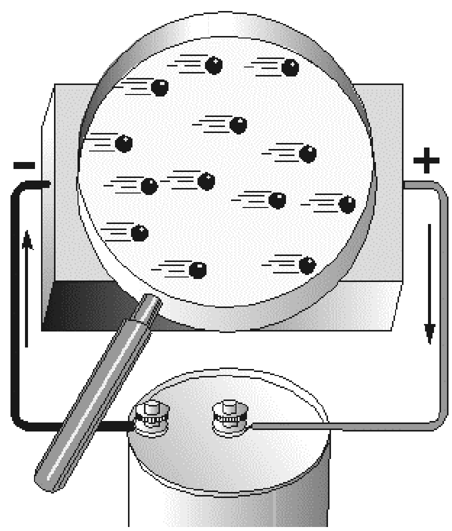


Figure 1-25.—Directed drift.

The directed movement of the electrons occurs at a relatively low VELOCITY (rate of motion in a particular direction). The effect of this directed movement, however, is felt almost instantaneously, as explained by the use of figure 1-26. As a difference in potential is impressed across the conductor, the positive terminal of the battery attracts electrons from point A. Point A now has a deficiency of electrons. As a result, electrons are attracted from point B to point A. Point B has now developed an electron deficiency, therefore, it will attract electrons. This same effect occurs throughout the conductor and repeats itself from points D to C. At the same instant the positive battery terminal attracted electrons from point A, the negative terminal repelled electrons toward point D. These electrons are attracted to point D as it gives up electrons to point C. This process is continuous for as long as a difference of potential exists across the conductor. Though an individual electron moves quite slowly through the conductor, the effect of a directed drift occurs almost instantaneously. As an electron moves into the conductor at point D, an electron is leaving at point A. This action takes place at approximately the speed of light (186,000 miles per second).

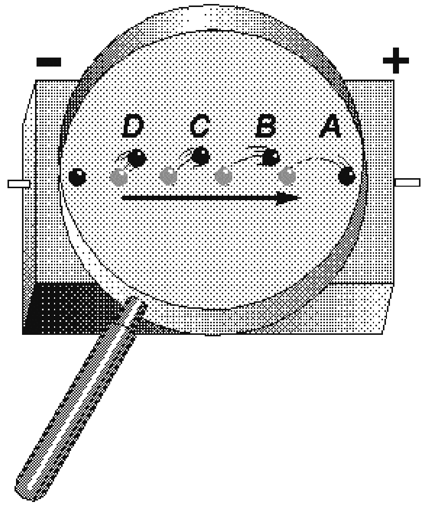


Figure 1-26.—Effect of directed drift.

Q46. The effects of directed drift take place at what rate of speed?

Magnitude of Current Flow

Electric current has been defined as the directed movement of electrons. Directed drift, therefore, is current and the terms can be used interchangeably. The expression directed drift is particularly helpful in differentiating between the random and directed motion of electrons. However, CURRENT FLOW is the terminology most commonly used in indicating a directed movement of electrons.

The magnitude of current flow is directly related to the amount of energy that passes through a conductor as a result of the drift action. An increase in the number of energy carriers (the mobile electrons) or an increase in the energy of the existing mobile electrons would provide an increase in current flow. When an electric potential is impressed across a conductor, there is an increase in the velocity of the mobile electrons causing an increase in the energy of the carriers. There is also the generation of an increased number of electrons providing added carriers of energy. The additional number of free electrons is relatively small, hence the magnitude of current flow is primarily dependent on the velocity of the existing mobile electrons.

The magnitude of current flow is affected by the difference of potential in the following manner. Initially, mobile electrons are given additional energy because of the repelling and attracting electrostatic field. If the potential difference is increased, the electric field will be stronger, the amount of energy imparted to a mobile electron will be greater, and the current will be increased. If the potential difference is decreased, the strength of the field is reduced, the energy supplied to the electron is diminished, and the current is decreased.

Q47. What is the relationship of current to voltage in a circuit?

Measurement of Current

The magnitude of current is measured in AMPERES. A current of one ampere is said to flow when one coulomb of charge passes a point in one second. Remember, one coulomb is equal to the charge of 6.28×10^{18} electrons.

Frequently, the ampere is much too large a unit for measuring current. Therefore, the MILLIAMPERE (mA), one-thousandth of an ampere, or the MICROAMPERE (µA), one-millionth of an ampere, is used. The device used to measure current is called an AMMETER and will be discussed in detail in a later module.

Q48. Convert 350 mA to amperes.

ELECTRICAL RESISTANCE

It is known that the directed movement of electrons constitutes a current flow. It is also known that the electrons do not move freely through a conductor's crystalline structure. Some materials offer little opposition to current flow, while others greatly oppose current flow. This opposition to current flow is known as RESISTANCE (R), and the unit of measure is the OHM. The standard of measure for one ohm is the resistance provided at zero degrees Celsius by a column of mercury having a cross-sectional area of one square millimeter and a length of 106.3 centimeters. A conductor has one ohm of resistance when an applied potential of one volt produces a current of one ampere. The symbol used to represent the ohm is the Greek letter omega (Ω).

Resistance, although an electrical property, is determined by the physical structure of a material. The resistance of a material is governed by many of the same factors that control current flow. Therefore, in a subsequent discussion, the factors that affect current flow will be used to assist in the explanation of the factors affecting resistance.

Q49. What is the symbol for ohm?

Factors That Affect Resistance

The magnitude of resistance is determined in part by the "number of free electrons" available within the material. Since a decrease in the number of free electrons will decrease the current flow, it can be said that the opposition to current flow (resistance) is greater in a material with fewer free electrons. Thus, the resistance of a material is determined by the number of free electrons available in a material.

A knowledge of the conditions that limit current flow and, therefore, affect resistance can now be used to consider how the type of material, physical dimensions, and temperature will affect the resistance of a conductor.

TYPE OF MATERIAL.—Depending upon their atomic structure, different materials will have different quantities of free electrons. Therefore, the various conductors used in electrical applications have different values of resistance.

Consider a simple metallic substance. Most metals are crystalline in structure and consist of atoms that are tightly bound in the lattice network. The atoms of such elements are so close together that the electrons in the outer shell of the atom are associated with one atom as much as with its neighbor. (See fig. 1-27 view A). As a result, the force of attachment of an outer electron with an individual atom is practically zero. Depending on the metal, at least one electron, sometimes two, and in a few cases, three electrons per atom exist in this state. In such a case, a relatively small amount of additional electron energy would free the outer electrons from the attraction of the nucleus. At normal room temperature materials of this type have many free electrons and are good conductors. Good conductors will have a <u>low resistance</u>.

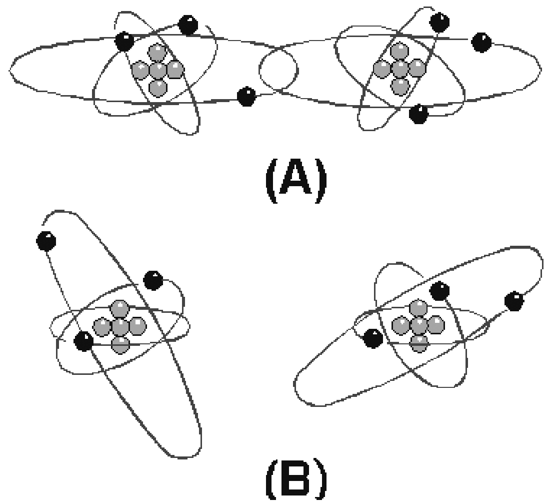


Figure 1-27.—Atomic spacing in conductors.

If the atoms of a material are farther apart, as illustrated in figure 1-27 view B, the electrons in the outer shells will not be equally attached to several atoms as they orbit the nucleus. They will be attracted by the nucleus of the parent atom only. Therefore, a greater amount of energy is required to free any of these electrons. Materials of this type are poor conductors and therefore have a high resistance.

Silver, gold, and aluminum are good conductors. Therefore, materials composed of their atoms would have a low resistance.

The element copper is the conductor most widely used throughout electrical applications. Silver has a lower resistance than copper but its cost limits usage to circuits where a high conductivity is demanded.

Aluminum, which is considerably lighter than copper, is used as a conductor when weight is a major factor.

Q50. When would silver be used as a conductor in preference to copper?

EFFECT OF CROSS-SECTIONAL AREA.—Cross-sectional area greatly affects the magnitude of resistance. If the cross-sectional area of a conductor is increased, a greater quantity of electrons are available for movement through the conductor. Therefore, a larger current will flow for a given amount of applied voltage. An increase in current indicates that when the cross-sectional area of a conductor is increased, the resistance must have decreased. If the cross-sectional area of a conductor is decreased, the number of available electrons decreases and, for a given applied voltage, the current through the conductor decreases. A decrease in current flow indicates that when the cross-sectional area of a conductor is decreased, the resistance must have increased. Thus, the RESISTANCE OF A CONDUCTOR IS INVERSELY PROPORTIONAL TO ITS CROSS-SECTIONAL AREA.

The diameter of conductors used in electronics is often only a fraction of an inch, therefore, the diameter is expressed in mils (thousandths of an inch). It is also standard practice to assign the unit circular mil to the cross-sectional area of the conductor. The circular mil is found by squaring the diameter when the diameter is expressed in mils. Thus, if the diameter is 35 mils (0.035 inch), the circular mil area is equal to $(35)^2$ or 1225 circular mils. A comparison between a square mil and a circular mil is illustrated in figure 1-28.

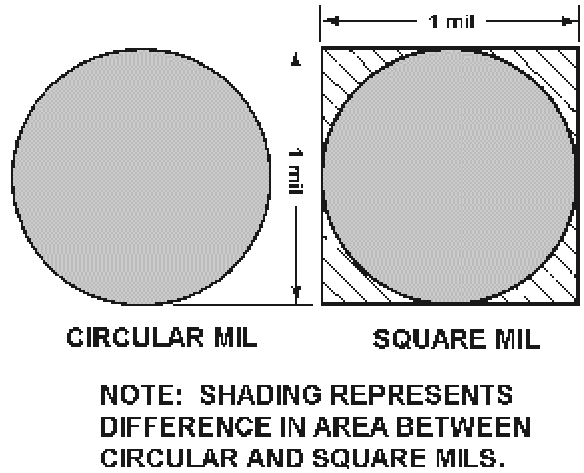


Figure 1-28.—Square and circular mils.

EFFECT OF CONDUCTOR LENGTH.—The length of a conductor is also a factor which determines the resistance of a conductor. If the length of a conductor is increased, the amount of energy given up increases. As free electrons move from atom to atom some energy is given off as heat. The longer a conductor is, the more energy is lost to heat. The additional energy loss subtracts from the energy being transferred through the conductor, resulting in a decrease in current flow for a given applied voltage. A decrease in current flow indicates an increase in resistance, since voltage was held constant. Therefore, if the length of a conductor is increased, the resistance increases. THE RESISTANCE OF A CONDUCTOR IS DIRECTLY PROPORTIONAL TO ITS LENGTH.

Q51. Which wire has the least resistance? Wire A-copper, 1000 circular mils, 6 inches long. Wire B-copper, 2000 circular mils, 11 inches long.

EFFECT OF TEMPERATURE.—Temperature changes affect the resistance of materials in different ways. In some materials an increase in temperature causes an increase in resistance, whereas in others, an increase in temperature causes a decrease in resistance. The amount of change of resistance per unit change in temperature is known as the TEMPERATURE COEFFICIENT. If for an increase in temperature the resistance of a material increases, it is said to have a POSITIVE TEMPERATURE COEFFICIENT. A material whose resistance decreases with an increase in temperature has a NEGATIVE TEMPERATURE COEFFICIENT. Most conductors used in electronic applications have a positive temperature coefficient. However, carbon, a frequently used material, is a substance having a negative temperature coefficient. Several materials, such as the alloys constantan and manganin, are considered to have a ZERO TEMPERATURE COEFFICIENT because their resistance remains relatively constant for changes in temperature.

Q52. Which temperature coefficient indicates a material whose resistance increases as temperature increases?

Q53. What term describes a material whose resistance remains relatively constant with changes in temperature?

CONDUCTANCE

Electricity is a study that is frequently explained in terms of opposites. The term that is the opposite of resistance is CONDUCTANCE. Conductance is the ability of a material to pass electrons. The factors that affect the magnitude of resistance are exactly the same for conductance, but they affect conductance in the opposite manner. Therefore, conductance is directly proportional to area, and inversely proportional to the length of the material. The temperature of the material is definitely a factor, but assuming a constant temperature, the conductance of a material can be calculated.

The unit of conductance is the MHO (G), which is ohm spelled backwards. Recently the term mho has been redesignated SIEMENS (S). Whereas the symbol used to represent resistance (R) is the Greek letter omega (Ω), the symbol used to represent conductance (G) is (S). The relationship that exists between resistance (R) and conductance (G) or (S) is a reciprocal one. A reciprocal of a number is one divided by that number. In terms of resistance and conductance:

$$R = \frac{1}{G}, \quad G = \frac{1}{R}$$

Q54. What is the unit of conductance and what other term is sometimes used?

Q55. What is the relationship between conductance and resistance?

ELECTRICAL RESISTORS

Resistance is a property of every electrical component. At times, its effects will be undesirable. However, resistance is used in many varied ways. RESISTORS are components manufactured to possess <u>specific values</u> of resistance. They are manufactured in many types and sizes. When drawn using its schematic representation, a resistor is shown as a series of jagged lines, as illustrated in figure 1-29.

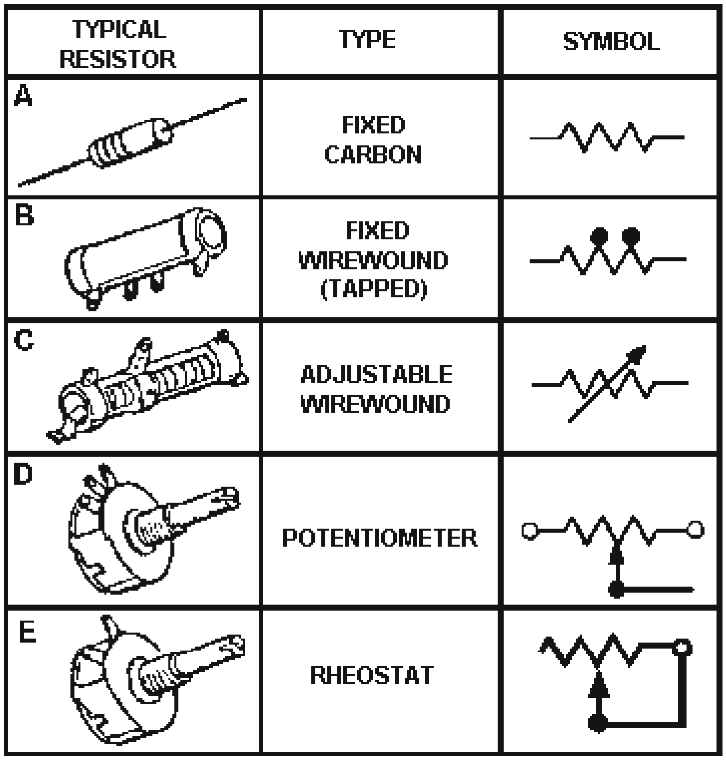


Figure 1-29.—Types of resistors.

Q56. What is schematic symbol for a resistor?

Composition of Resistors

One of the most common types of resistors is the molded composition, usually referred to as the carbon resistor. These resistors are manufactured in a variety of sizes and shapes. The chemical composition of the resistor determines its ohmic value and is accurately controlled by the manufacturer in the development process. They are made in ohmic values that range from one ohm to millions of ohms. The physical size of the resistor is related to its wattage rating, which is the ability of resistor to dissipate heat caused by the resistance.

Carbon resistors, as you might suspect, have as their principal ingredient the element carbon. In the manufacturer of carbon resistors, fillers or binders are added to the carbon to obtain various resistor values. Examples of these fillers are clay, bakelite, rubber, and talc. These fillers are doping agents and cause the overall conduction characteristics to change.

Carbon resistors are the most common resistors found because they are easy to manufacturer, inexpensive, and have a tolerance that is adequate for most electrical and electronic applications. Their prime disadvantage is that they have a tendency to change value as they age. One other disadvantage of carbon resistors is their limited power handling capacity.

The disadvantage of carbon resistors can be overcome by the use of WIREWOUND resistors (fig. 1-29 (B) and (C)). Wirewound resistors have very accurate values and possess a higher current handling capability than carbon resistors. The material that is frequently used to manufacture wirewound resistors is German silver which is composed of copper, nickel, and zinc. The qualities and quantities of these elements present in the wire determine the resistivity of the wire. (The resistivity of the wire is the measure or ability of the wire to resist current. Usually the percent of nickel in the wire determines the resistivity.) One disadvantage of the wirewound resistor is that it takes a large amount of wire to manufacture a resistor of high ohmic value, thereby increasing the cost. A variation of the wirewound resistor provides an exposed surface to the resistance wire on one side. An adjustable tap is attached to this side. Such resistors, sometimes with two or more adjustable taps, are used as voltage dividers in power supplies and other applications where a specific voltage is desired to be "tapped" off.

Q57. What does the wattage rating of a resistor indicate?

Q58. What are the two disadvantages of carbon-type resistors?

Q59. What type resistor should be used to overcome the disadvantages of the carbon resistor?

Fixed and Variable Resistors

There are two kinds of resistors, FIXED and VARIABLE. The fixed resistor will have one value and will never change (other than through temperature, age, etc.). The resistors shown in A and B of figure 1-29 are classed as fixed resistors. The tapped resistor illustrated in B has several fixed taps and makes more than one resistance value available. The sliding contact resistor shown in C has an adjustable collar that can be moved to tap off any resistance within the ohmic value range of the resistor.

There are two types of variable resistors, one called a POTENTIOMETER and the other a RHEOSTAT (see views D and E of fig. 1-29.) An example of the potentiometer is the volume control on your radio, and an example of the rheostat is the dimmer control for the dash lights in an automobile. There is a slight difference between them. Rheostats usually have two connections, one fixed and the other moveable. Any variable resistor can properly be called a rheostat. The potentiometer always has three connections, two fixed and one moveable. Generally, the rheostat has a limited range of values and a high current-handling capability. The potentiometer has a wide range of values, but it usually has a limited current-handling capability. Potentiometers are always connected as voltage dividers. (Voltage dividers are discussed in Chapter 3.)

Q60. Describe the differences between the rheostat connections and those of the potentiometer.

Q61. Which type of variable resistor should you select for controlling a large amount of current?

Wattage Rating

When a current is passed through a resistor, heat is developed within the resistor. The resistor must be capable of dissipating this heat into the surrounding air; otherwise, the temperature of the resistor rises causing a change in resistance, or possibly causing the resistor to burn out.

The ability of the resistor to dissipate heat depends upon the design of the resistor itself. This ability to dissipate heat depends on the amount of surface area which is exposed to the air. A resistor designed to dissipate a large amount of heat must therefore have a large physical size. The heat dissipating capability of a resistor is measured in WATTS (this unit will be explained later in chapter 3). Some of the more common wattage ratings of carbon resistors are: one-eighth watt, one-fourth watt, one-half watt, one watt, and two watts. In some of the newer state-of-the-art circuits of today, much smaller wattage resistors are used. Generally, the type that you will be able to physically work with are of the values given. The higher the wattage rating of the resistor the larger is the physical size. Resistors that dissipate very large amounts of power (watts) are usually wirewound resistors. Wirewound resistors with wattage ratings up to 50 watts are not uncommon. Figure 1-30 shows some resistors which have different wattage ratings. Notice the relative sizes of the resistors.

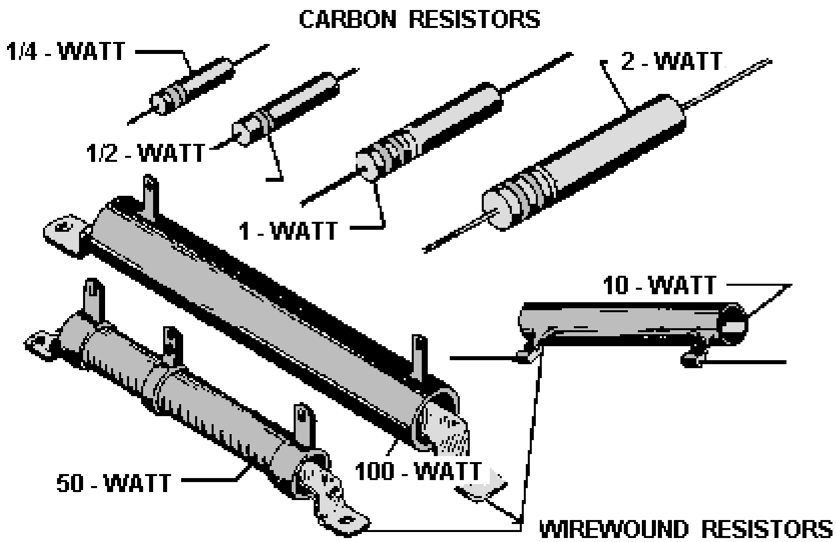


Figure 1-30.—Resistors of different wattage ratings.

Standard Color Code System

In the standard color code system, four bands are painted on the resistor, as shown in figure 1-31.

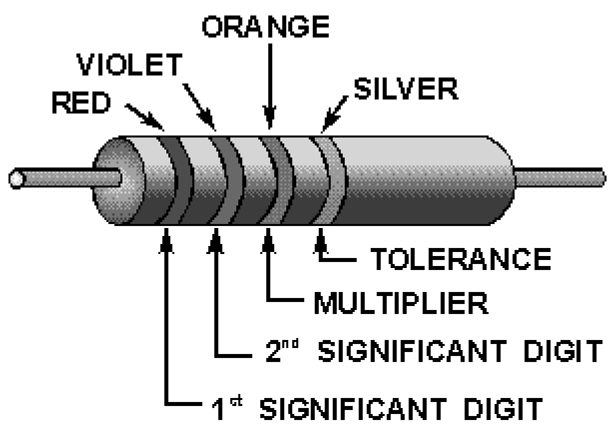
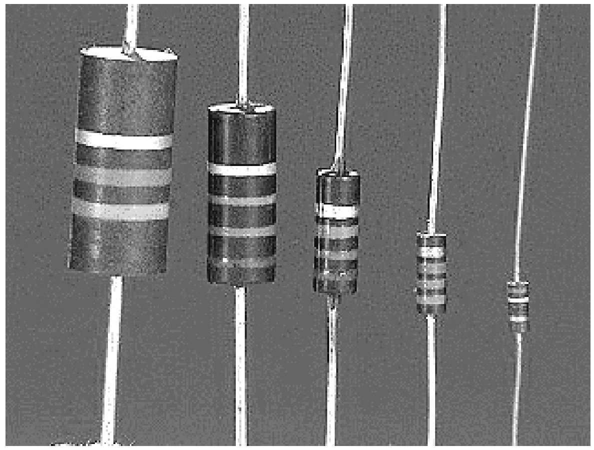


Figure 1-31.—Resistor color codes.



Examples of resistor color codes.

The color of the first band indicates the value of the first significant digit. The color of the second band indicates the value of the second significant digit. The third color band represents a decimal multiplier by which the first two digits must be multiplied to obtain the resistance value of the resistor. The colors for the bands and their corresponding values are shown in Table 1-1.

Table 1-1.—Standard Color Code for Resistors

COLOR	SIGNIFICANT FIGURE	DECIMAL MULTIPLIER	TOLERANCE PERCENT	RELIABILITY LEVEL PER 1,000 HRS.
BLACK	0	1	PERCENT	—
BROWN	1	10	1	1.0%
RED	2	100	2	0.1%
ORANGE	3	1,000	—	0.01%
YELLOW	4	10,000	—	0.001%
GREEN	5	100,000	—	—
BLUE	6	1,000,000	—	—
VIOLET	7	10,000,000	—	—
GRAY	8	100,000,000	—	—
WHITE	9	1,000,000,000	—	—
GOLD	—	.1	5	—
SILVER	—	.01	10	—
NO COLOR	—	—	20	—

Use the example colors shown in figure 1-31. Since red is the color of the first band, the first significant digit is 2. The second band is violet, therefore the second significant digit is 7. The third band is orange, which indicates that the number formed as a result of reading the first two bands is multiplied by 1000. In this case 27 x 1000 = 27,000 ohms. The last band on the resistor indicates the tolerance; that is, the manufacturer's allowable ohmic deviation above and below the numerical value indicated by the resistor's color code. In this example, the color silver indicates a tolerance of 10 percent. In other words, the actual value of the resistor may fall somewhere within 10 percent above and 10 percent below the value indicated by the color code. This resistor has an indicated value of 27,000 ohms. Its tolerance is 10 percent x 27,000 ohms, or 2,700 ohms. Therefore, the resistor's actual value is somewhere between 24,300 ohms and 29,700 ohms.

When measuring resistors, you will find situations in which the quantities to be measured may be extremely large, and the resulting number using the basic unit, the ohm, may prove too cumbersome. Therefore, a metric system prefix is usually attached to the basic unit of measurement to provide a more manageable unit. Two of the most commonly used prefixes are kilo and mega. Kilo is the prefix used to represent thousand and is abbreviated k. Mega is the prefix used to represent million and is abbreviated M.

In the example given above, the 27,000-ohm resistor could have been written as 27 kilohms or 27 kΩ. Other examples are: 1,000 ohms = 1 kΩ; 10,000 ohms = 10 kΩ; 100,000 ohms = 100 kΩ. Likewise, 1,000,000 ohms is written as 1 megohm or 1 MΩ and 10,000,000 ohms = 10 MΩ.

Q62. A carbon resistor has a resistance of 50 ohms, and a tolerance of 5 percent. What are the colors of bands one, two, three, and four, respectively?

SIMPLIFYING THE COLOR CODE.—Resistors are the most common components used in electronics. The technician must identify, select, check, remove, and replace resistors. Resistors and resistor circuits are usually the easiest branches of electronics to understand.

The resistor color code sometimes presents problems to a technician. It really should not, because once the resistor color code is learned, you should remember it for the rest of your life.

Black, brown, red, orange, yellow, green, blue, violet, gray, white—this is the order of colors you should know automatically. There is a memory aid that will help you remember the code in its proper order. Each word starts with the first letter of the colors. If you match it up with the color code, you will not forget the code.

Bad Boys Run Over Yellow Gardenias Behind Victory Garden Walls,

or:

Black	—	Bad
Brown	—	Boys
Red	—	Run
Orange	—	Over
Yellow	—	Yellow
Green	—	Gardenias
Blue	—	Behind
Violet	—	Victory
Gray	—	Garden
White	—	Walls

There are many other memory aid sentences that you might want to ask about from experienced technicians. You might find one of the other sentences easier to remember.

There is still a good chance that you will make a mistake on a resistor's color band. Most technicians do at one time or another. If you make a mistake on the first two significant colors, it usually is not too serious. If you make a miscue on the third band, you are in trouble, because the value is going to be at least 10 times too high or too low. Some important points to remember about the third band are:

When the third band is . . . .

Black, the resistor's value is less than 100 ohms.

Brown, the resistor's value is in hundreds of ohms.

Red, the resistor's value is in thousands of ohms.

Orange, the resistor's value is in tens of thousands of ohms.

Yellow, the resistor's value is in hundreds of thousands of ohms.

Green, the resistor's value is in megohms.

Blue, the resistor's value is in tens of megohms or more.

Although you may find any of the above colors in the third band, red, orange, and yellow are the most common. In some cases, the third band will be silver or gold. You multiply the first two bands by 0.01 if it is silver, and 0.1 if it is gold.

The fourth band, which is the tolerance band, usually does not present too much of a problem. If there is no fourth band, the resistor has a 20-percent tolerance; a silver fourth band indicates a 10-percent tolerance; and a gold fourth band indicates a 5-percent tolerance. Resistors that conform to military specifications have a fifth band. The fifth band indicates the reliability level per 1,000 hours of operation as follows:

Fifth band color	Level
Brown	1.0%
Red	0.1%
Orange	0.01%
Yellow	0.001%

For a resistor whose the fifth band is color coded brown, the resistor's chance of failure will not exceed 1 percent for every 1,000 hours of operation.

In equipment such as the Navy's complex computers, the reliability level is very significant. For example, in a piece of equipment containing 10,000 orange fifth-band resistors, no more than one resistor will fail during 1,000 hours of operation. This is very good reliability. More information on resistors is contained in *NEETS Module 19*.

Q63. A carbon resistor has the following color bands: The first band is yellow, followed by violet, yellow, and silver. What is the ohmic value of the resistor?

Q64. The same resistor mentioned in question 63 has a yellow fifth band. What does this signify?

Q65. A resistor is handed to you for identification with the following color code: the first band is blue, followed by gray, green, gold, and brown. What is the resistor's value?

Some resistors, both wirewound and composition, will not use the resistor color code. These resistors will have the ohmic value and tolerance imprinted on the resistor itself.

SUMMARY

With the completion of this chapter, you now have gained the necessary information which is the foundation for the further study of electricity. The following is a summary of the important parts in the chapter.

In describing the composition of matter, the following terms are important for you to remember:

MATTER is defined as anything that occupies space and has weight.

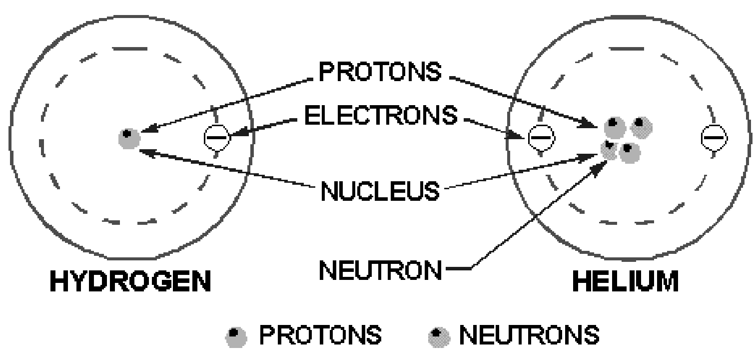
AN ELEMENT is a substance which cannot be reduced to a simpler substance by chemical means.

A COMPOUND is a chemical combination of elements which can be separated by chemical means, but not by physical means. It is created by chemically combining two or more elements.

A MIXTURE is a combination of elements or compounds that can be separated by physical means.

A MOLECULE is the chemical combination of two or more atoms. In a compound, the molecule is the smallest particle that has all the characteristics of the compound.

AN ATOM is the smallest particle of an element that retains the characteristics of that element. An atom is made up of electrons, protons, and neutrons. The number and arrangement of these subatomic particles determine the kind of element.



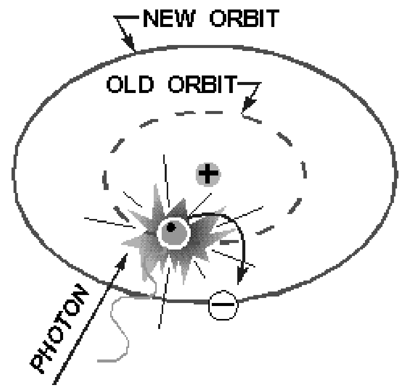
AN ELECTRON is considered to be a negative charge of electricity.

A PROTON is considered to be a positive charge of electricity.

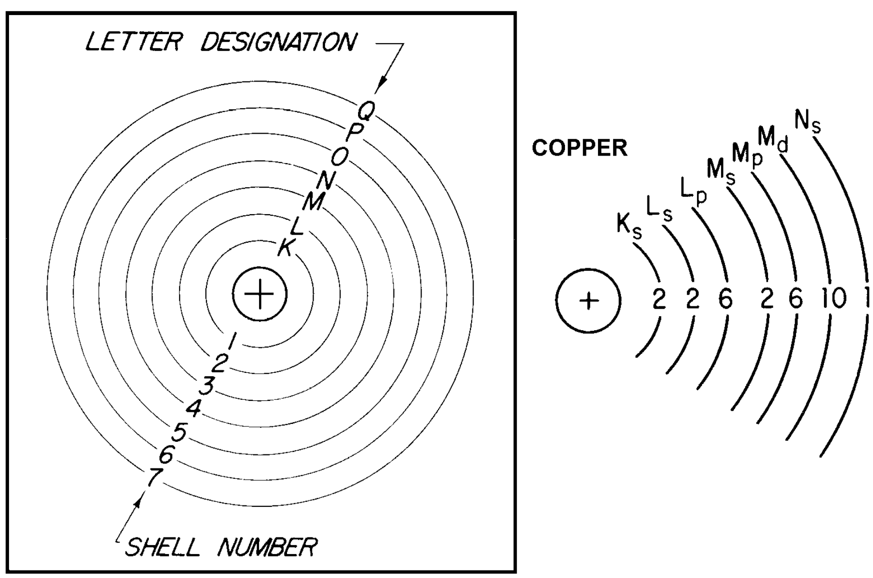
A NEUTRON is a neutral particle in that it has no electrical charge.

ENERGY in an electron is of two types—kinetic (energy of motion) and potential (energy of position).

ENERGY LEVELS of the electron exist because the electron has mass and motion. The motion gives it kinetic energy and its position gives it potential energy. Energy balance keeps the electron in orbit and should it gain energy it will assume an orbit further from the center of the atom. It will remain at that level for only a fraction of a second before it radiates the excess energy and goes back to a lower orbit.



SHELLS AND SUBSHELLS of electrons are the orbits of the electrons in the atom. Each shell contains a maximum of 2 times its number squared electrons. Shells are lettered K through Q, starting with K, which is the closest to the nucleus. The shell can be split into 4 subshells labeled s, p, d, and f, which can contain 2, 6, 10, and 14 electrons, respectively.

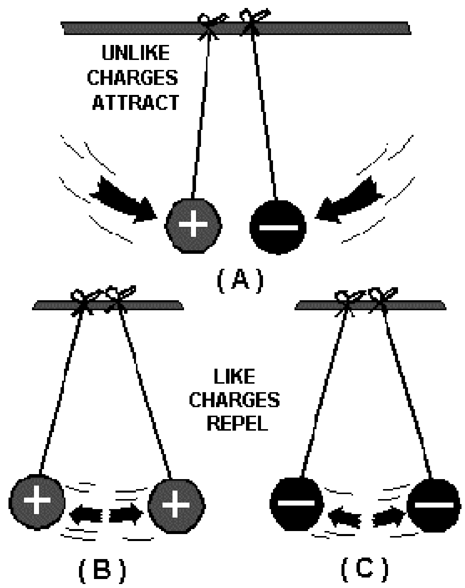


VALENCE OF AN ATOM is determined by the number of electrons in the outermost shell. The shell is referred to as the valence shell, and the electrons within it are valence electrons. An atom with few valence electrons requires little energy to free the valence electrons.

IONIZATION refers to the electrons contained in an atom. An atom with a positive charge has lost some of its electrons, and is called a positive ion. A negatively charged atom is a negative ion.

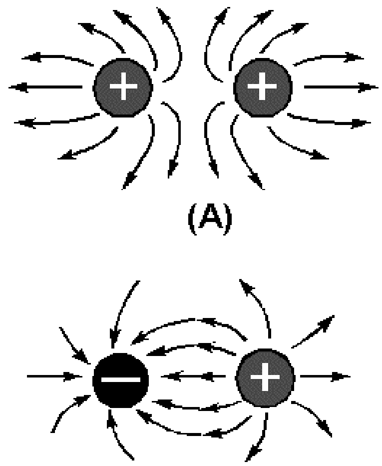
CONDUCTORS, SEMICONDUCTORS, AND INSULATORS are categorized as such by the number of valence electrons in their atoms. The conductor normally has 3 or less valence electrons and offers little opposition to the flow of electrons (electric current). The insulator contains 5 or more valence electrons and offers high opposition to electron flow. The semiconductor usually has four valence electrons of conductivity and is in the midrange. The best conductors in order of conductance are silver, copper, gold, and aluminum.

CHARGED BODIES affect each other as follows: When two bodies having unequal charges are brought close to each other, they will tend to attract each other in an attempt to equalize their respective charges. When two bodies, both having either positive or negative charges, are brought close together, they tend to repel each other as no equalization can occur. When the charge on one body is high enough with respect to the charge on an adjacent body, an equalizing current will flow between the bodies regardless of the conductivity of the material containing the bodies.



A NEUTRAL BODY may be attracted to either a positively or negatively charged body due to the relative difference in their respective charges.

CHARGED BODIES will attract or repel each other with a force that is directly proportional to the product of their individual charges and inversely proportional to the square of the distance between the bodies.



(A)

(B)

<u>ELECTROSTATIC LINES</u> of force are a graphic representation of the field around a charged body. These lines are imaginary. Lines from a positively charged body are indicated as flowing out from the body, while lines from a negatively charged body are indicated as flowing into the body.

<u>MAGNETISM</u> is that property of a material which enables it to attract pieces of iron. A material with this property is called a magnet. Any material that is attracted to a magnet can be made into a magnet itself.

<u>FERROMAGNETIC MATERIALS</u> are materials that are easy to magnetize; e.g., iron, steel, and cobalt.

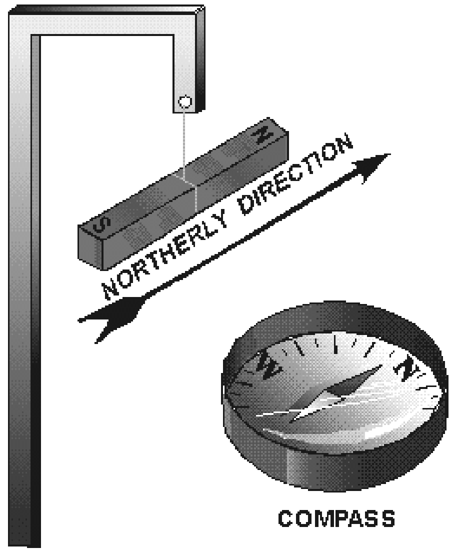
<u>NATURAL MAGNETS</u>, called magnetite, lodestones, or leading stones, were the first magnets to be studied. Most magnets in practical use are artificial or man-made magnets, and are made either by electrical means or by stroking a magnetic material with a magnet.

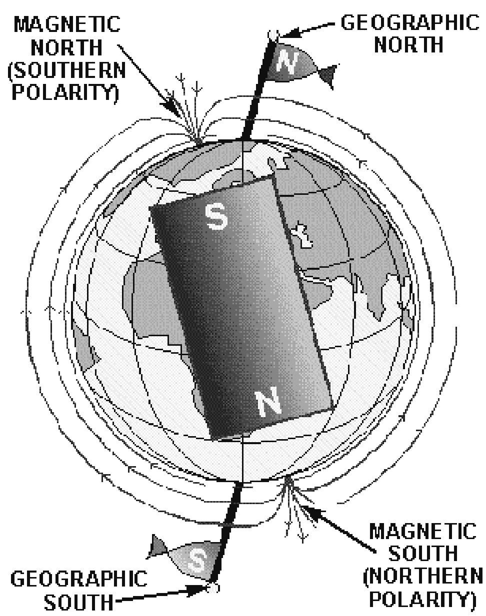
<u>RELUCTANCE</u> is defined as the opposition of a material to being magnetized.

<u>PERMEABILITY</u> is defined as the ease with which a material accepts magnetism. A material which is easy to magnetize does not hold its magnetism very long, and vice versa.

<u>RETENTIVITY</u> is defined as the ability of a material to retain magnetism.

<u>A MAGNETIC POLE</u> is located at each end of a magnet. The majority of the magnetic force is concentrated at these poles and is approximately equal at both poles.

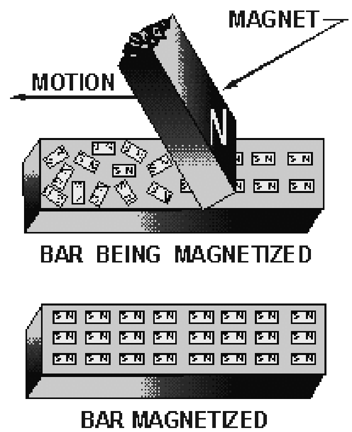




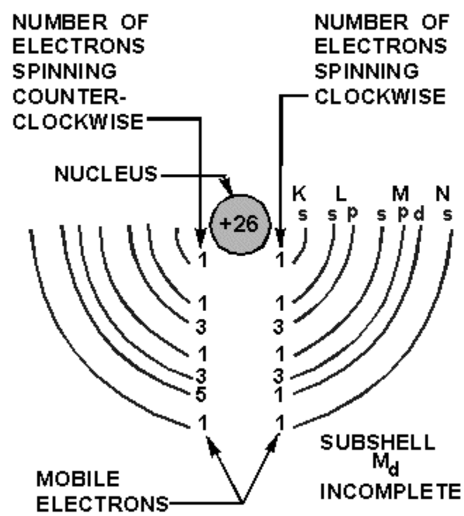
THE NORTH POLE, or north seeking pole, of a magnet freely suspended on a string always points toward the north geographical pole.

THE LIKE POLES of magnets repel each other, while the unlike poles attract each other.

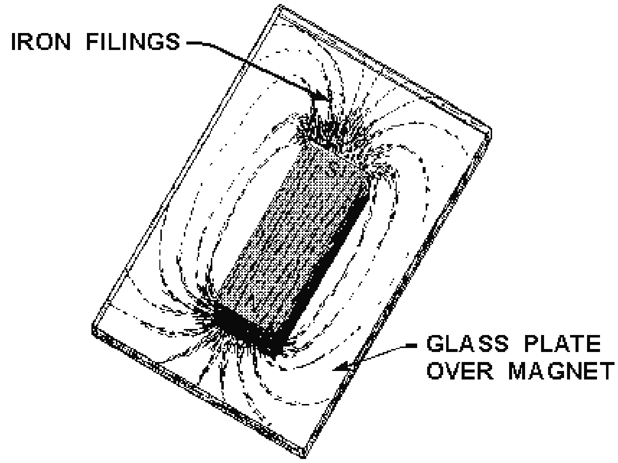
WEBER'S THEORY OF MAGNETISM assumes that all magnetic material is made up of magnetic molecules which, if lined up in north to south pole order, will be a magnet. If not lined up, the magnetic fields about the molecules will neutralize each other and no magnetic effect will be noted.



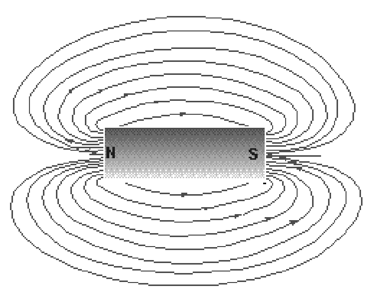
THE DOMAIN THEORY OF MAGNETISM states that if the electrons of the atoms in a material spin more in one direction than in the other, the material will become magnetized.



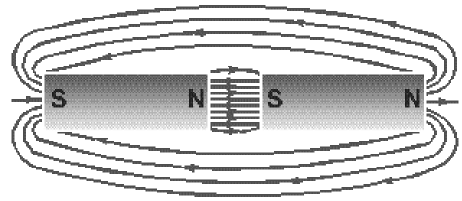
A MAGNETIC FIELD is said to exist in the space surrounding a magnet.



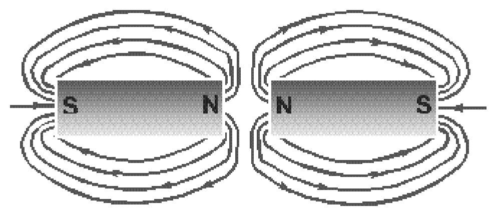
MAGNETIC LINES OF FORCE are imaginary lines used to describe the patterns of the magnetic field about a magnet. These lines are assumed to flow externally from the north pole and into the south pole.



MAGNETIC FLUX is the total number of magnetic lines of force leaving or entering the pole of a magnet.



UNLIKE POLES ATTRACT



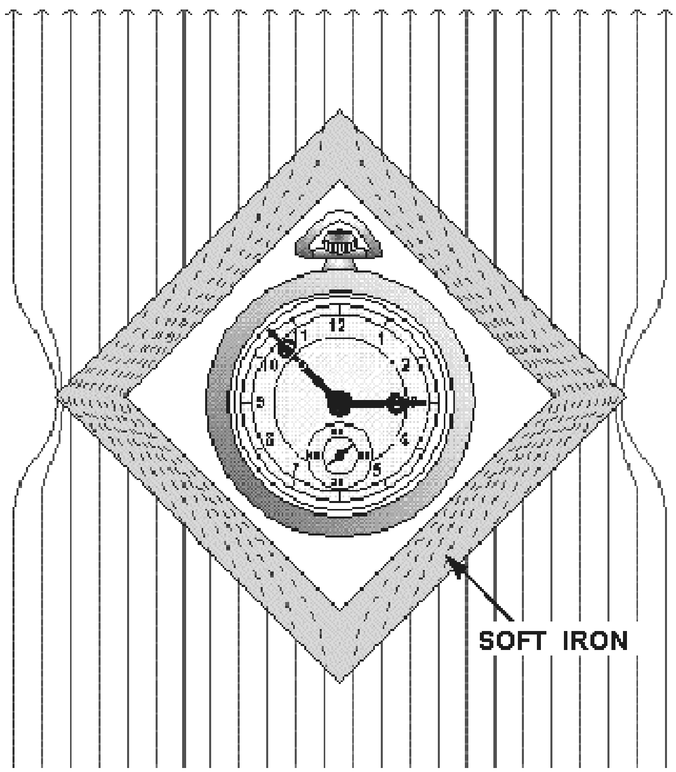
LIKE POLES REPEL

FLUX DENSITY is the number of flux lines per unit area.

FIELD INTENSITY or the intensity of a magnetic field is directly related to the magnetic force exerted by the field.

THE INTENSITY OF ATTRACTION/REPULSION between magnetic poles may be described by a law almost identical to Coulomb's Law of Charged Bodies, that is, the force between two poles is directly proportional to the product of the pole strengths and inversely proportional to the square of the distance between the poles.

MAGNETIC SHIELDING can be accomplished by placing a soft iron shield around the object to be protected, thus directing the lines of force around the object.



MAGNETS ARE CLASSIFIED BY SHAPE and include the bar magnet, the horseshoe magnet, and the ring magnet. The ring magnet is used in computer memory circuits; the horseshoe magnet in some meter circuits.

ENERGY may be defined as the ability to do work.

THE COULOMB (C) is the basic unit used to indicate an electrical charge. One coulomb is equal to a charge of 6.28×10^{18} electrons. When one coulomb of charge exists between two bodies, the electromotive force (or voltage) is one volt.

VOLTAGE is measured as the difference of potential of two charges of interest.

VOLTAGE MEASUREMENTS may be expressed in the following units: volts (V), kilovolts (kV), millivolts (mV), or microvolts (μV).

For example:

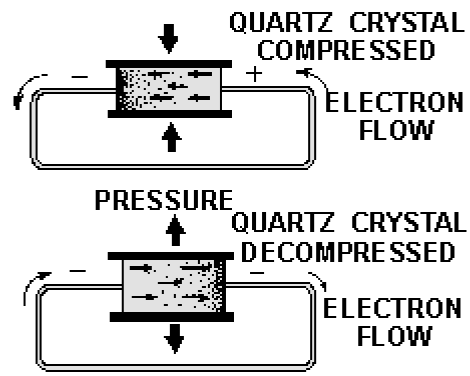
1 kV = 1,000 V

1 mV = 0.001 V

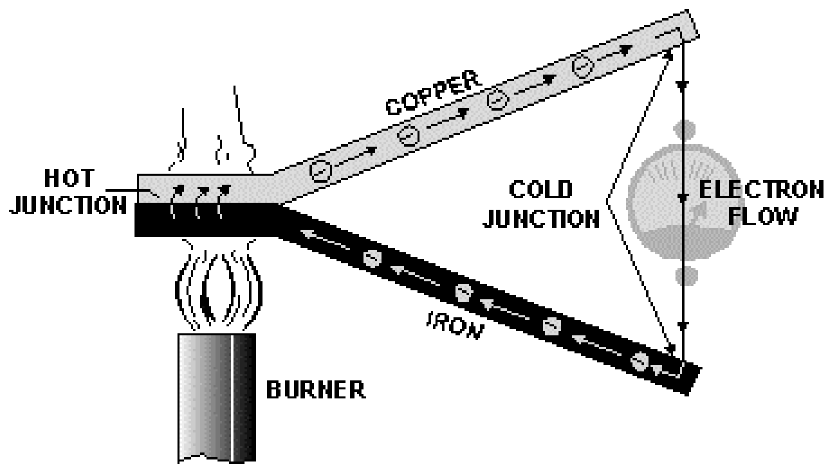
1 μV = 0.000001 V

METHODS OF PRODUCING A VOLTAGE include:

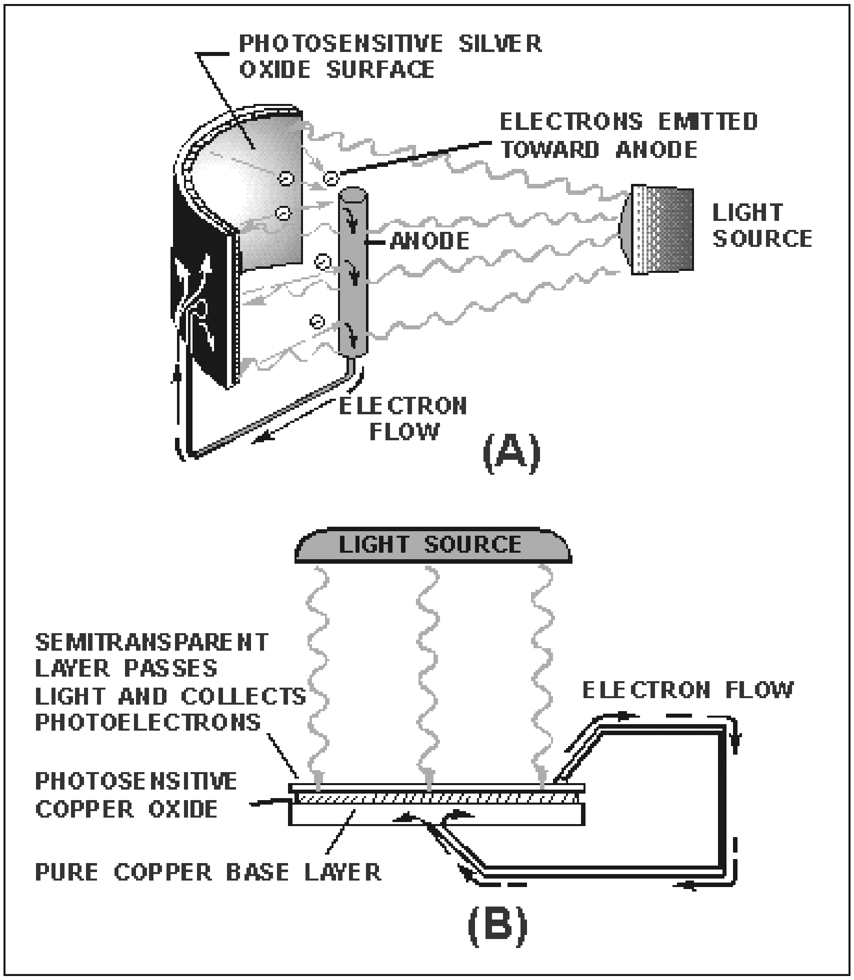
1. Friction
2. Pressure (piezoelectricity)



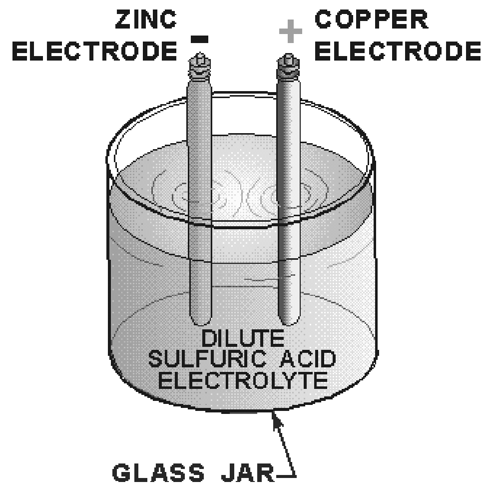
3. Heat (thermoelectricity)



4. Light (photoelectricity)

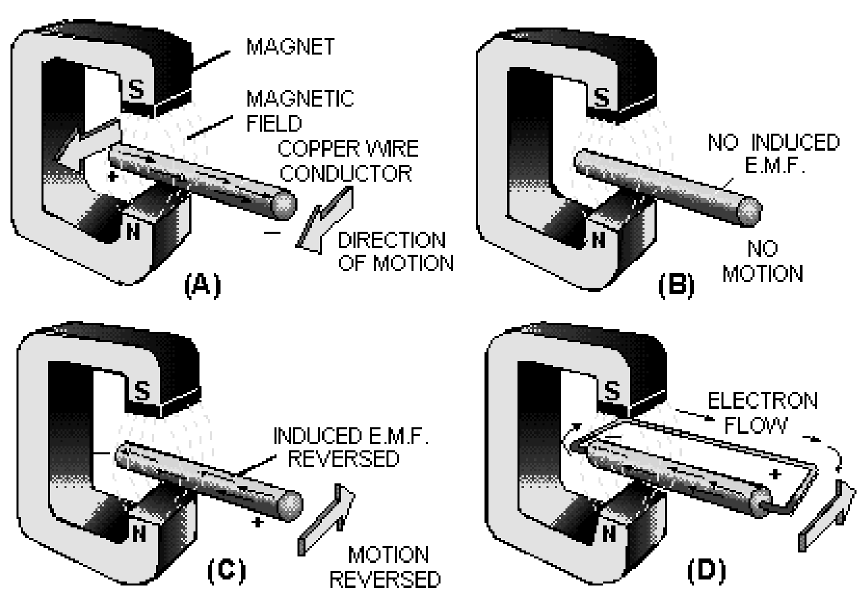


5. Chemical action (battery)



6. Magnetism (electromagnetic induction generator)

ELECTROMAGNETIC INDUCTION GENERATOR To produce voltage by use of magnetism, three conditions must be met: There must be a CONDUCTOR in which the voltage will be produced; there must be a MAGNETIC FIELD in the conductor's vicinity; and there must be relative motion between the field and conductor. When these conditions are met, electrons WITHIN THE CONDUCTOR are propelled in one direction or another, creating an electromotive force, or voltage.



ELECTRON CURRENT is based on the assumption that electron current flow is from negative to positive through a circuit.

AN ELECTRIC CURRENT is a directed movement of electrons in a conductor or circuit.

THE AMPERE is the basic unit used to indicate an electric current. A current of one ampere is said to flow when one coulomb of charge (6.28×10^{18} electrons) passes a given point in one second of time. Current measurements may be expressed in the following units: ampere (A), milliampere (mA), and microampere (µA). Current in a circuit increases in direct proportion to the voltage (emf) applied across the circuit.

RESISTANCE is the opposition to current. The ohm is the basic unit of resistance and is represented by the Greek letter omega (Ω). A conductor is said to have one ohm of resistance when an emf of one volt causes one ampere of current to flow in the conductor. Resistance may be expressed in the following units: ohm (Ω), kilohm (kΩ), and megohms (MΩ). For example, 1,000,000Ω = 1,000 kΩ = 1 MΩ.

THE RESISTANCE OF A MATERIAL is determined by the type, the physical dimensions, and the temperature of the material that is,

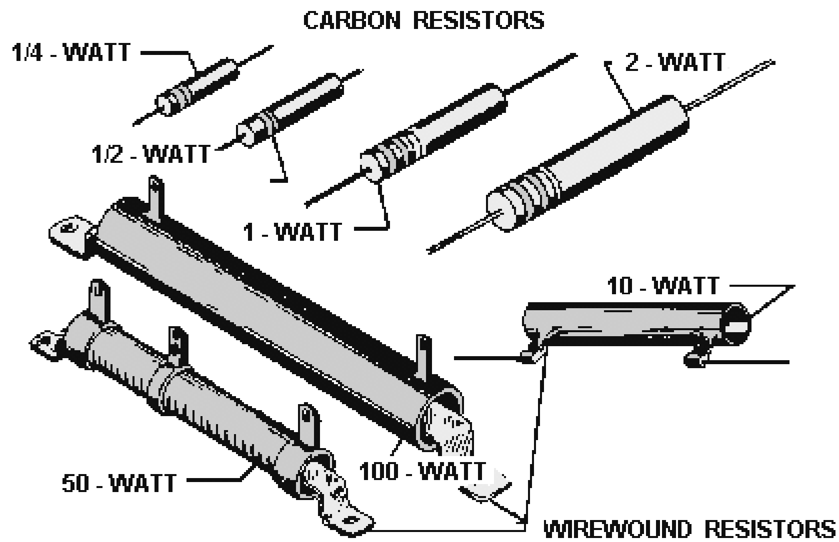
1. A good conductor contains an abundance of free electrons.
2. As the cross-sectional area of a given conductor is increased, the resistance will decrease.
3. As the length of a conductor is increased, the resistance will increase.
4. In a material having a positive temperature coefficient, the resistance will increase as the temperature is increased.

THE CONDUCTANCE OF A MATERIAL is the reciprocal of resistance.

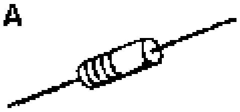
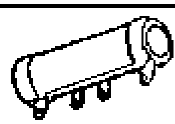
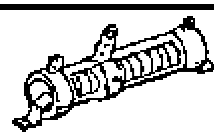
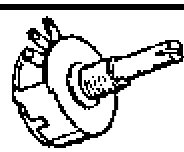
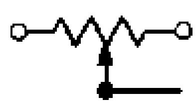
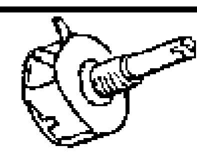
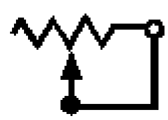
THE UNIT OF CONDUCTANCE is the mho and the symbol is V. G or S.

THE RESISTOR is manufactured to provide a specific value of resistance.

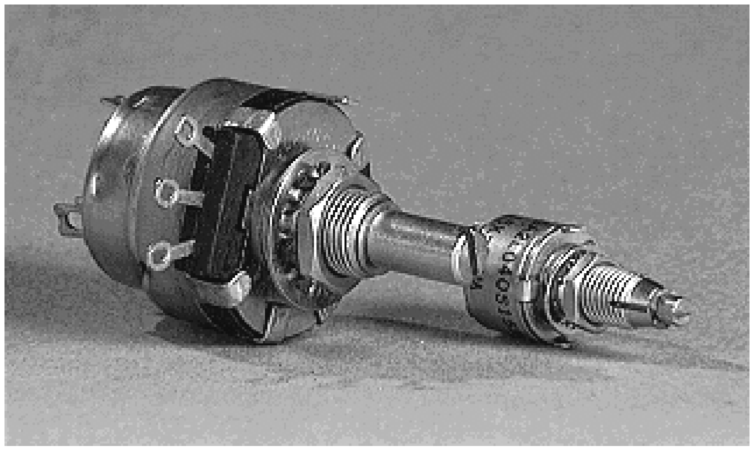
THE CARBON RESISTOR is made of carbon, with fillers and binders blended in to control the ohmic value.



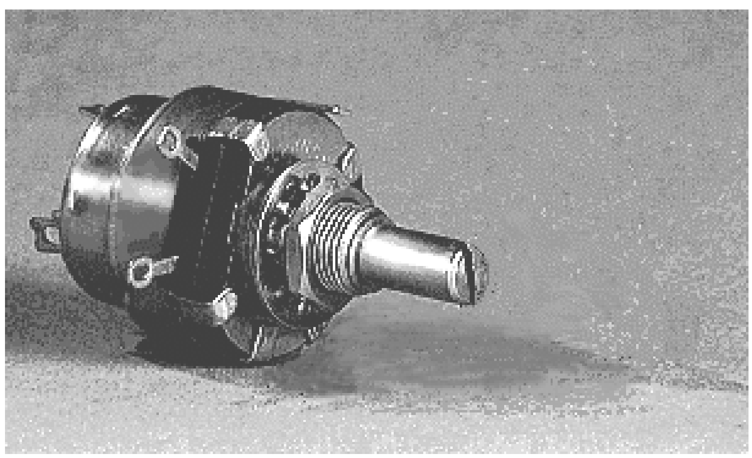
<u>THE RESISTANCE OF A WIREWOUND RESISTOR</u> is determined by the metal content of the wire and the wire's length. Wirewound resistors may be tapped so two or more different voltage values may be taken off the same resistor.

TYPICAL RESISTOR	TYPE	SYMBOL
A 	FIXED CARBON	
B 	FIXED WIREWOUND (TAPPED)	
C 	ADJUSTABLE WIREWOUND	
D 	POTENTIOMETER	
E 	RHEOSTAT	

<u>THE POTENTIOMETER AND THE RHEOSTAT</u> are variable resistors and can be adjusted to any resistance value within their ohmic range. The rheostat is usually used for relatively high current applications and has two connections; the potentiometer has 3 connections and is a relatively high-resistance, low-current device.



Two examples of potentiometers.



Example of a rheostat.

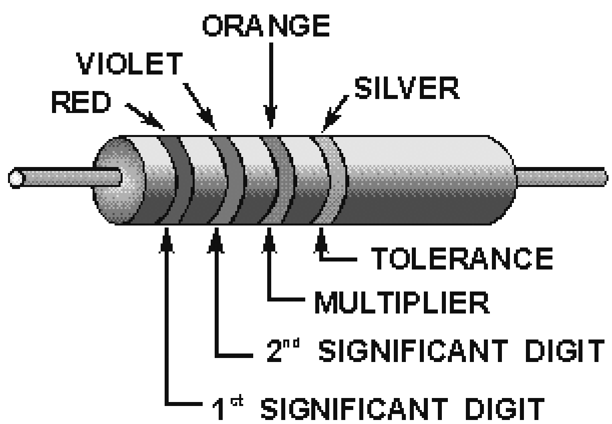
COLOR	SIGNIFICANT FIGURE	DECIMAL MULTIPLIER	TOLERANCE PERCENT	RELIABILITY LEVEL PER 1,000 HRS.
BLACK	0	1	PERCENT	—
BROWN	1	10	1	1.0%
RED	2	100	2	0.1%
ORANGE	3	1,000	—	0.01%
YELLOW	4	10,000	—	0.001%
GREEN	5	100,000	—	—
BLUE	6	1,000,000	—	—
VIOLET	7	10,000,000	—	—
GRAY	8	100,000,000	—	—
WHITE	9	1,000,000,000	—	—
GOLD	—	.1	5	—
SILVER	—	.01	10	—
NO COLOR	—	—	20	—

Table 1-1.—Standard Color Code for Resistors

THE WATTAGE RATING OF A RESISTOR is related to the resistor's physical size, that is, the greater the surface area exposed to the air, the larger the rating.

THE STANDARD COLOR CODE for resistors is used to determine the following:

1. Ohmic value
2. Tolerance
3. Reliability level (on some resistors)



ANSWERS TO QUESTIONS Q1. THROUGH Q65.

A1. *Anything that occupies space and has weight. Solids, liquids, gases.*

A2. *A substance which cannot be reduced to a simpler substance by chemical means.*

A3. *A substance consisting of two or more elements.*

A4. *A compound is a chemical combination of elements that cannot be separated by physical means. A mixture is a physical combination of elements and compounds that are not chemically combined.*

A5. *A chemical combination of two or more atoms.*

A6. *Electrons-negative, protons-positive, and neutrons-neutral.*

A7. *Kinetic energy.*

A8. *Invisible light photons (ultraviolet) bombard the phosphor atom in the light tube. The phosphor atoms emit visible light photons.*

A9. *The number of electrons in the outer shell.*

A10. *An atom with more or less than its normal number of electrons.*

A11. *The number of valence electrons.*

A12. *Through the accumulation of excess electrons.*

A13. *By friction.*

A14. *Negative.*

A15. *Like charges repel, and unlike charges attract with a force directly proportional to the product of their charges and inversely proportional to the square of the distance between them.*

A16. *The space between and around charged bodies.*

A17. *Leaving positive, entering negative.*

A18. *Motors, generators, speakers, computers, televisions, tape recorders, and many others.*

A19. *Those materials that are attracted by magnets and have the ability to become magnetized.*

A20. *The relative ease with which they are magnetized.*

A21. *A material that exhibits low reluctance and high permeability, such as iron or soft steel.*

A22. *The ability of a material to retain magnetism.*

A23. *They are very similar; like charges repel, unlike charges attract, like poles repel—unlike poles attract.*

A24. *To the magnetic north pole.*

A25. *South pole at the right, north pole at the left.*

A26. The domain theory is based upon the electron spin principle; Weber's theory uses the concept of tiny molecular magnets.

A27. To enable you to "see" the magnetic field.

A28. No specific pattern, sawdust is a nonmagnetic material.

A29. An imaginary line used to illustrate magnetic effects.

A30. Electrostatic lines of force do not form closed loops.

A31. By shielding or surrounding the instrument with a soft iron case, called a magnetic shield or screen.

A32. In pairs, with opposite poles together to provide a complete path for magnetic flux.

A33. The ability to do work.

A34. Kinetic energy.

A35. Potential energy.

A36. Difference of potential.

A37. 2100 volts.

A38. (a) 250 kV, (b) 25 V, (c) 1 μV.

A39. A voltage source.

A40. Friction, pressure, heat, light, chemical action, and magnetism.

A41. Pressure.

A42. Heat.

A43. Chemical.

A44. Magnetic.

A45. Electron theory assumes that electron flow is from negative to positive.

A46. The speed of light (186,000 miles per second, 300,000,000 meters per second).

A47. Current increases as voltage increases.

A48. 0.35 amperes.

A49. Ω

A50. When the need for conductivity is great enough to justify the additional expense.

A51. Wire B.

A52. Positive.

A53. Zero temperature coefficient.

A54. The mho (v), siemens.

A55. They are reciprocals of each other.

A56.

A57. Its ability to dissipate heat.

A58. 1. Change value with age. 2. Limited power capacity.

A59. The wirewound resistor.

A60. The rheostat may have two connections, one fixed and one moveable; the potentiometer always has three connections, one moveable and two fixed.

A61. The rheostat.

A62. The bands are green, black, black, and gold.

A63. 470,000 ohms (470 kilohms).

A64. The resistor's chance of failure is 0.001 percent for 1000 hours of operation.

A65. 6,800,000 ohms (6.8 megohms), with 5% tolerance, and a 1% reliability level.

CHAPTER 2

BATTERIES

LEARNING OBJECTIVES

Upon completing this chapter, you will be able to:

1. State the purpose of a cell.
2. State the purpose of the three parts of a cell.
3. State the difference between the two types of cells.
4. Explain the chemical process that takes place in the primary and secondary cells.
5. Recognize and define the terms electrochemical action, anode, cathode, and electrolyte.
6. State the causes of polarization and local action and describe methods of preventing these effects.
7. Identify the parts of a dry cell.
8. Identify the various dry cells in use today and some of their capabilities and limitations.
9. Identify the four basic secondary cells, their construction, capabilities, and limitations.
10. Define a battery, and identify the three ways of combining cells to form a battery.
11. Describe general maintenance procedures for batteries including the use of the hydrometer, battery capacity, and rating and battery charging.
12. Identify the five types of battery charges.
13. Observe the safety precautions for working with and around batteries.

INTRODUCTION

The purpose of this chapter is to introduce and explain the basic theory and characteristics of batteries. The batteries which are discussed and illustrated have been selected as representative of many models and types which are used in the Navy today. No attempt has been made to cover every type of battery in use, however, after completing this chapter you will have a good working knowledge of the batteries which are in general use.

First, you will learn about the building block of all batteries, the CELL. The explanation will explore the physical makeup of the cell and the methods used to combine cells to provide useful voltage, current, and power. The chemistry of the cell and how chemical action is used to convert chemical energy to electrical energy are also discussed.

In addition, the care, maintenance, and operation of batteries, as well as some of the safety precautions that should be followed while working with and around batteries are discussed.

Batteries are widely used as sources of direct-current electrical energy in automobiles, boats, aircraft, ships, portable electric/electronic equipment, and lighting equipment. In some instances, they are used as the only source of power; while in others, they are used as a secondary or standby power source.

A battery consists of a number of cells assembled in a common container and connected together to function as a source of electrical power.

THE CELL

A cell is a device that transforms chemical energy into electrical energy. The simplest cell, known as either a galvanic or voltaic cell, is shown in figure 2-1. It consists of a piece of carbon (C) and a piece of zinc (Zn) suspended in a jar that contains a solution of water (H_20) and sulfuric acid (H_2S0_4) called the electrolyte.

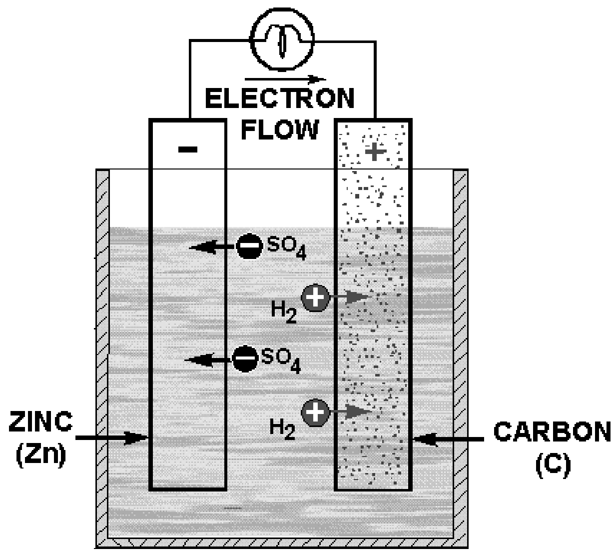


Figure 2-1.—Simple voltaic or galvanic cell.

The cell is the fundamental unit of the battery. A simple cell consists of two electrodes placed in a container that holds the electrolyte.

In some cells the container acts as one of the electrodes and, in this case, is acted upon by the electrolyte. This will be covered in more detail later.

ELECTRODES

The electrodes are the conductors by which the current leaves or returns to the electrolyte. In the simple cell, they are carbon and zinc strips that are placed in the electrolyte; while in the dry cell (fig. 2-2), they are the carbon rod in the center and zinc container in which the cell is assembled.

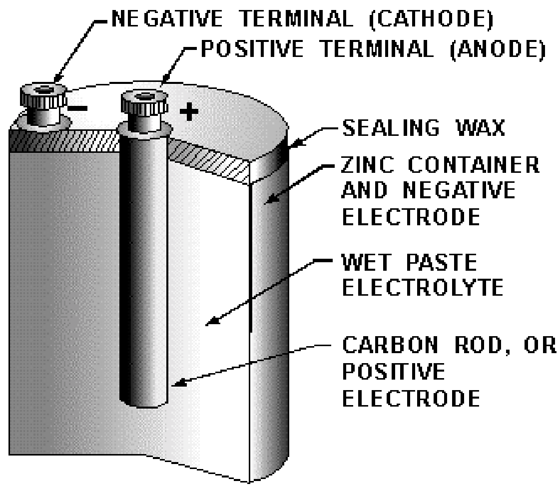


Figure 2-2.—Dry cell, cross-sectional view.

ELECTROLYTE

The electrolyte is the solution that acts upon the electrodes. The electrolyte, which provides a path for electron flow, may be a salt, an acid, or an alkaline solution. In the simple galvanic cell, the electrolyte is in a liquid form. In the dry cell, the electrolyte is a paste.

CONTAINER

The container which may be constructed of one of many different materials provides a means of holding (containing) the electrolyte. The container is also used to mount the electrodes. In the voltaic cell the container must be constructed of a material that will not be acted upon by the electrolyte.

Q1. What is the purpose of a cell?

Q2. What are the three parts of a cell?

Q3. What is the purpose of each of the three parts of a cell?

PRIMARY CELL

A primary cell is one in which the chemical action eats away one of the electrodes, usually the negative electrode. When this happens, the electrode must be replaced or the cell must be discarded. In the galvanic-type cell, the zinc electrode and the liquid electrolyte are usually replaced when this happens. In the case of the dry cell, it is usually cheaper to buy a new cell.

SECONDARY CELL

A secondary cell is one in which the electrodes and the electrolyte are altered by the chemical action that takes place when the cell delivers current. These cells may be restored to their original condition by forcing an electric current through them in the direction opposite to that of discharge. The automobile storage battery is a common example of the secondary cell.

Q4. What are the two types of cells?

Q5. What is the main difference between the two types of cells?

ELECTROCHEMICAL ACTION

If a load (a device that consumes electrical power) is connected externally to the electrodes of a cell, electrons will flow under the influence of a difference in potential across the electrodes from the CATHODE (negative electrode), through the external conductor to the ANODE (positive electrode).

A cell is a device in which chemical energy is converted to electrical energy. This process is called ELECTROCHEMICAL action.

The voltage across the electrodes depends upon the materials from which the electrodes are made and the composition of the electrolyte. The current that a cell delivers depends upon the resistance of the entire circuit, including that of the cell itself. The internal resistance of the cell depends upon the size of the electrodes, the distance between them in the electrolyte, and the resistance of the electrolyte. The larger the electrodes and the closer together they are in the electrolyte (without touching), the lower the internal resistance of the cell and the more current the cell is capable of supplying to the load.

Q6. What is electrochemical action?

Q7. What is another name for the (a) positive electrode, and the (b) negative electrode?

PRIMARY CELL CHEMISTRY

When a current flows through a primary cell having carbon and zinc electrodes and a diluted solution of sulfuric acid and water (combined to form the electrolyte), the following chemical reaction takes place.

The current flow through the load is the movement of electrons from the negative electrode of the cell (zinc) and to the positive electrode (carbon). This causes fewer electrons in the zinc and an excess of electrons in the carbon. Figure 2-1 shows the hydrogen ions (H_2) from the sulfuric acid being attracted to the carbon electrode. Since the hydrogen ions are positively charged, they are attracted to the negative charge on the carbon electrode. This negative charge is caused by the excess of electrons. The zinc electrode has a positive charge because it has lost electrons to the carbon electrode. This positive charge attracts the negative ions ($S0_4$) from the sulfuric acid. The negative ions combine with the zinc to form zinc sulfate. This action causes the zinc electrode to be eaten away. Zinc sulfate is a grayish-white substance that is sometimes seen on the battery post of an automobile battery.

The process of the zinc being eaten away and the sulfuric acid changing to hydrogen and zinc sulfate is the cause of the cell discharging. When the zinc is used up, the voltage of the cell is reduced to zero.

In figure 2-1 you will notice that the zinc electrode is labeled negative and the carbon electrode is labeled positive. This represents the current flow <u>outside</u> the cell from negative to positive.

The zinc combines with the sulfuric acid to form zinc sulfate and hydrogen. The zinc sulfate dissolves in the electrolyte (sulfuric acid and water) and the hydrogen appears as gas bubbles around the carbon electrode. As current continues to flow, the zinc gradually dissolves and the solution changes to zinc sulfate and water. The carbon electrode does not enter into the chemical changes taking place, but simply provides a return path for the current.

Q8. In the primary cell, why are negative ions attracted to the negative terminal of the cell?

Q9. How do electrons get from the negative electrode to the positive electrode?

Q10. What causes the negative electrode to be eaten away?

SECONDARY CELL CHEMISTRY

As stated before, the differences between primary and secondary cells are, the secondary cell can be recharged and the electrodes are made of different materials. The secondary cell shown in figure 2-3 uses sponge lead as the cathode and lead peroxide as the anode. This is the lead-acid type cell and will be used to explain the general chemistry of the secondary cell. Later in the chapter when other types of secondary cells are discussed, you will see that the materials which make up the parts of a cell are different, but that the chemical action is essentially the same.

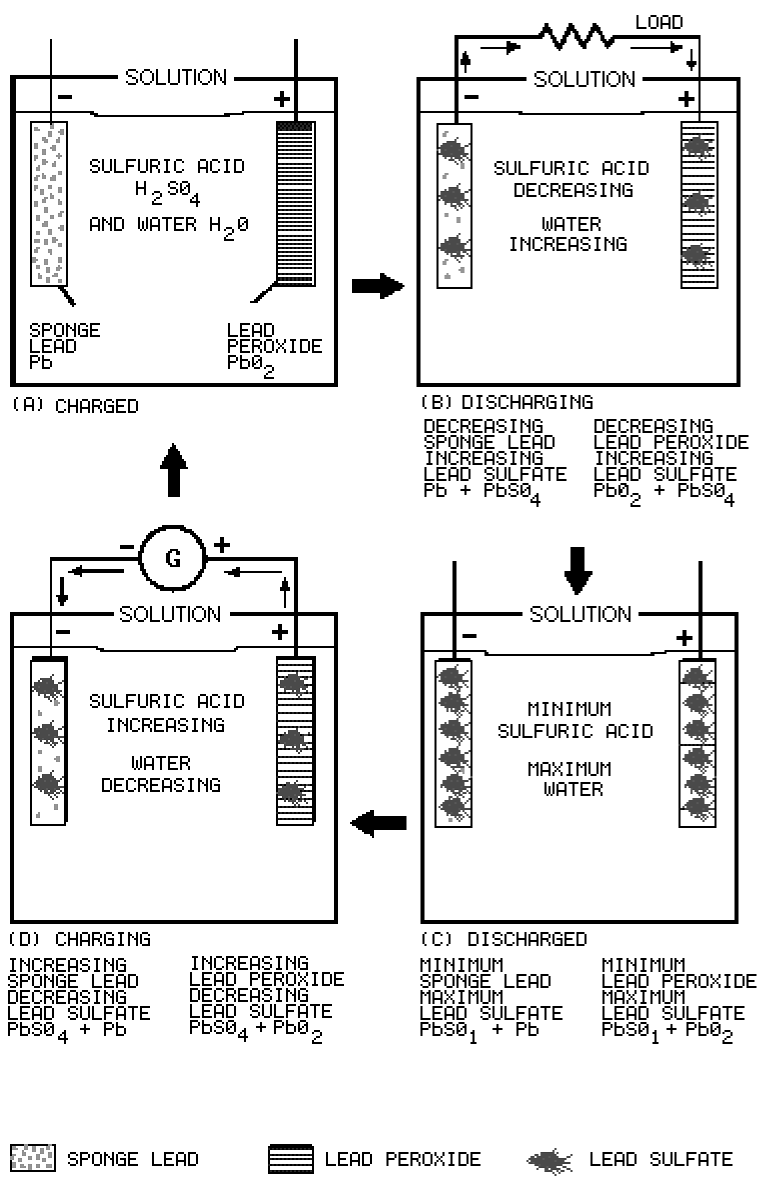


Figure 2-3.—Secondary cell.

Figure 2-3 view A shows a lead-acid secondary cell that is fully charged. The cathode is pure sponge lead, the anode is pure lead peroxide, and the electrolyte is a mixture of sulfuric acid and water.

Figure 2-3 view B shows the secondary cell discharging. A load is connected between the cathode and anode; current flows negative to positive as shown. This current flow creates the same process as was explained for the primary cell with the following exceptions.

In the primary cell the zinc cathode was eaten away by the sulfuric acid. In the secondary cell the sponge-like construction of the cathode retains the lead sulfate formed by the chemical action of the sulfuric acid and the lead. In the primary cell the carbon anode was not chemically acted upon by the sulfuric acid. In the secondary cell the lead peroxide anode is chemically changed to lead sulfate by the sulfuric acid.

When the cell is fully discharged it will be as shown in figure 2-3 view C. The anode and cathode retain some lead peroxide and sponge lead but the amounts of lead sulfate in each is maximum. The electrolyte has a minimum amount of sulfuric acid. With this condition no further chemical action can take place within the cell.

As you know, the secondary cell can be recharged. Recharging is the process of reversing the chemical action that occurs as the cell discharges. To recharge the cell, a voltage source, such as a generator, is connected as shown in figure 2-3 view D. The negative terminal of the voltage source is connected to the cathode of the cell and the positive terminal of the voltage source is connected to the anode of the cell. With this arrangement the lead sulfate is chemically changed back to sponge lead in the cathode, lead peroxide in the anode, and sulfuric acid in the electrolyte. After all the lead sulfate is chemically changed, the cell is fully charged as shown in figure 2-3 view A. Once the cell has been charged, the discharge-charge cycle may be repeated.

Q11. Refer to figure 2-3(B). Why is the sulfuric acid decreasing?

Q12. Refer to figure 2-3(D). How is it possible for the sulfuric acid to be increasing?

Q13. Refer to figure 2-3(D). When all the lead sulfate has been converted, what is the condition of the cell?

POLARIZATION OF THE CELL

The chemical action that occurs in the cell while the current is flowing causes hydrogen bubbles to form on the surface of the anode. This action is called POLARIZATION. Some hydrogen bubbles rise to the surface of the electrolyte and escape into the air, some remain on the surface of the anode. If enough bubbles remain around the anode, the bubbles form a barrier that increases internal resistance. When the internal resistance of the cell increases, the output current is decreased and the voltage of the cell also decreases.

A cell that is heavily polarized has no useful output. There are several methods to prevent polarization or to depolarize the cell.

One method uses a vent on the cell to permit the hydrogen to escape into the air. A disadvantage of this method is that hydrogen is not available to reform into the electrolyte during recharging. This problem is solved by adding water to the electrolyte, such as in an automobile battery. A second method is to use material that is rich in oxygen, such as manganese dioxide, which supplies free oxygen to combine with the hydrogen and form water.

A third method is to use a material that will absorb the hydrogen, such as calcium. The calcium releases hydrogen during the charging process. All three methods remove enough hydrogen so that the cell is practically free from polarization.

LOCAL ACTION

When the external circuit is removed, the current ceases to flow, and, theoretically, all chemical action within the cell stops. However, commercial zinc contains many impurities, such as iron, carbon, lead, and arsenic. These impurities form many small electrical cells within the zinc electrode in which current flows between the zinc and its impurities. Thus, the chemical action continues even though the cell itself is not connected to a load.

Local action may be prevented by using pure zinc (which is not practical), by coating the zinc with mercury, or by adding a small percentage of mercury to the zinc during the manufacturing process. The treatment of the zinc with mercury is called amalgamating (mixing) the zinc. Since mercury is many times heavier than an equal volume of water, small particles of impurities weighing less than mercury will float to the surface of the mercury. The removal of these impurities from the zinc prevents local action. The mercury is not readily acted upon by the acid. When the cell is delivering current to a load, the mercury continues to act on the impurities in the zinc. This causes the impurities to leave the surface of the zinc electrode and float to the surface of the mercury. This process greatly increases the storage life of the cell.

Q14. Describe three ways to prevent polarization.

Q15. Describe local action

.

TYPES OF CELLS

The development of new and different types of cells in the past decade has been so rapid that it is virtually impossible to have a complete knowledge of all the various types. A few recent developments are the silver-zinc, nickel-zinc, nickel-cadmium, silver-cadmium, organic and inorganic lithium, and mercury cells.

PRIMARY DRY CELL

The dry cell is the most popular type of primary cell. It is ideal for simple applications where an inexpensive and noncritical source of electricity is all that is needed.

The dry cell is not actually dry. The electrolyte is not in a liquid state, but is a moist paste. If it should become totally dry, it would no longer be able to transform chemical energy to electrical energy.

Construction of a Dry Cell

The construction of a common type of dry cell is shown in figure 2-4. These dry cells are also referred to as Leclanche' cells. The internal parts of the cell are located in a cylindrical zinc container. This zinc container serves as the negative electrode (cathode) of the cell. The container is lined with a nonconducting material, such as blotting paper, to separate the zinc from the paste. A carbon electrode is located in the center, and it serves as the positive terminal (anode) of the cell. The paste is a mixture of several substances such as ammonium chloride, powdered coke, ground carbon, manganese dioxide, zinc chloride, graphite, and water.

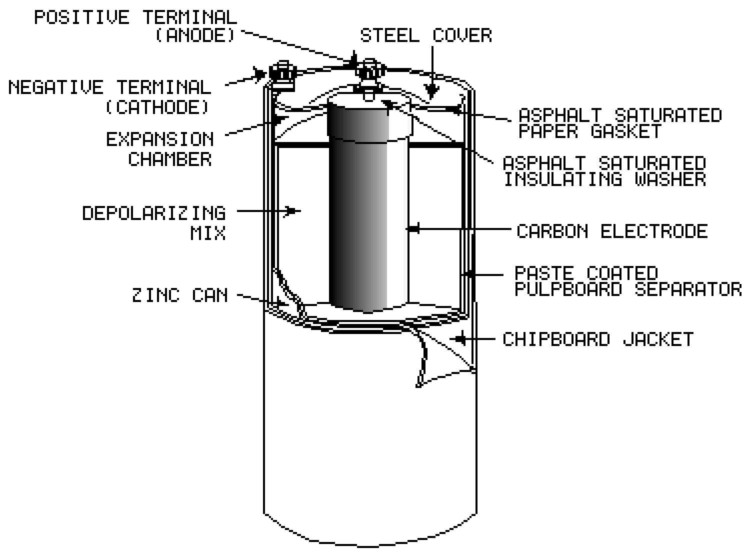


Figure 2-4.—Cutaway view of the general-purpose dry cell.

This paste, which is packed in the space between the anode and the blotting paper, also serves to hold the anode rigid in the center of the cell. When the paste is packed in the cell, a small space is left at the top for expansion of the electrolytic paste caused by the depolarization action. The cell is then sealed with a cardboard or plastic seal.

Since the zinc container is the cathode, it must be protected with some insulating material to be electrically isolated. Therefore, it is common practice for the manufacturer to enclose the cells in cardboard and metal containers.

The dry cell (fig. 2-4) is basically the same as the simple voltaic cell (wet cell) described earlier, as far as its internal chemical action is concerned. The action of the water and the ammonium chloride in the paste, together with the zinc and carbon electrodes, produces the voltage of the cell. Manganese dioxide is added to reduce polarization when current flows and zinc chloride reduces local action when the cell is not being used.

A cell that is not being used (sitting on the shelf) will gradually deteriorate because of slow internal chemical changes (local action). This deterioration is usually very slow if cells are properly stored. If unused cells are stored in a cool place, their shelf life will be greatly preserved. Therefore, to minimize deterioration, they should be stored in refrigerated spaces.

The blotting paper (paste-coated pulpboard separator) serves two purposes—(1) it keeps the paste from making actual contact with the zinc container and (2) it permits the electrolyte from the paste to filter through to the zinc slowly. The cell is sealed at the top to keep air from entering and drying the electrolyte. Care should be taken to prevent breaking this seal.

Q16. What serves as the cathode of a dry cell?

Q17. Why is a dry cell called a DRY cell?

Q18. What does the term "shelf life" mean?

Mercuric-Oxide Zinc Cell

The mercuric-oxide zinc cell (mercury cell) is a primary cell that was developed during World War II. Two important assets of the mercury cell are its ability to produce current for a long period of time and a long shelf life when compared to the dry cell shown in figure 2-4. The mercury cell also has a very stable output voltage and is a power source that can be made in a small physical size.

With the birth of the space program and the development of small transceivers and miniaturized equipment, a power source of small size was needed. Such equipment requires a small cell which is capable of delivering maximum electrical energy at a constant discharge voltage. The mercury cell, which is one of the smallest cells, meets these requirements.

Present mercury cells are manufactured in three basic types as shown in figure 2-5. The wound-anode type, shown in figure 2-5 view A, has an anode composed of a corrugated zinc strip with a paper absorbent. The zinc is mixed with mercury, and the paper is soaked in the electrolyte which causes it to swell and press against the zinc and make positive contact. This process ensures that the electrolyte makes contact with the anode.

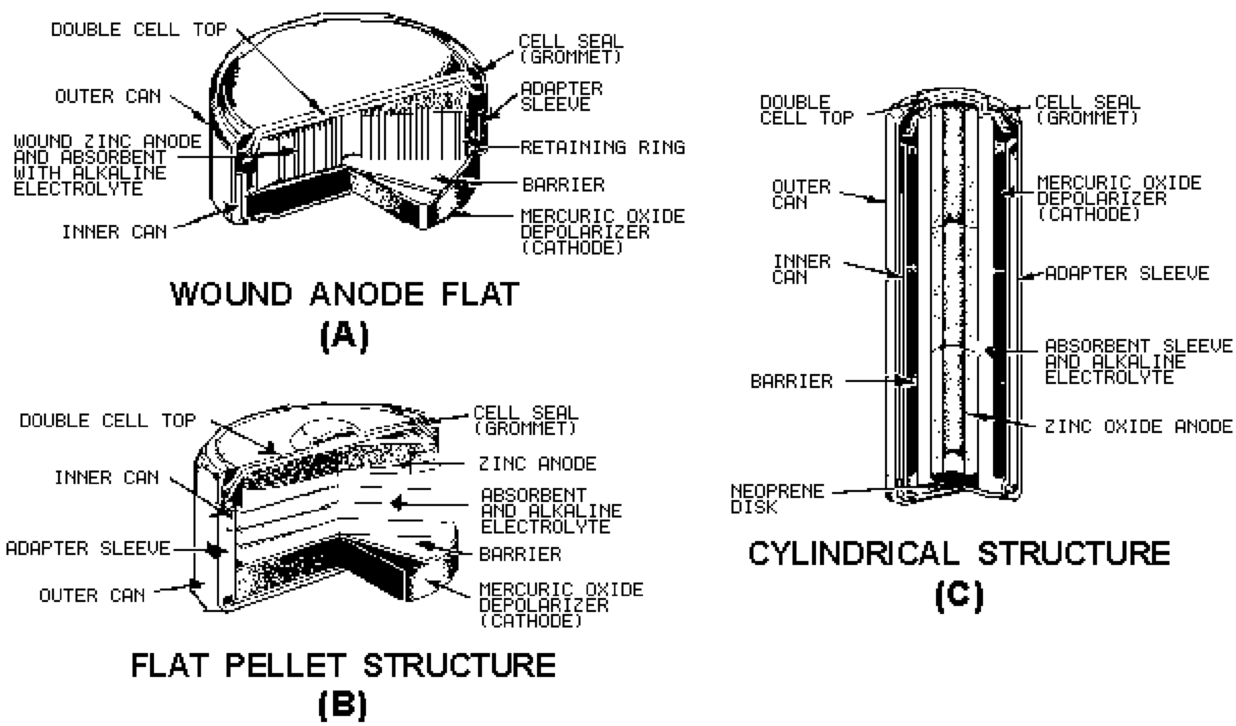


Figure 2-5.—Mercury cells.

In the pressed-powder cells, shown in figure 2-5 views B and C, the zinc powder for the anode is mixed prior to being pressed into shape. The absorbent shown in the figure is paper soaked in the electrolyte. The space between the inner and outer containers provides passage for any gas generated by an improper chemical balance or impurities present within the cell.

If the anode and cathode of a cell are connected together without a load, a SHORT CIRCUIT condition exists. Short circuits (shorts) can be very dangerous. They cause excessive heat, pressure, and current flow which may cause serious damage to the cell or be a safety hazard to personnel.

WARNING

Do not short the mercury cell. Shorted mercury cells have exploded with considerable force.

Other Types of Cells

There are many different types of primary cells. Because of such factors as cost, size, ease of replacement, and voltage or current needs, many types of primary cells have been developed. The following is a brief description of some of the primary cells in use today.

The Manganese Dioxide-Alkaline-Zinc Cell is similar to the zinc-carbon cell except for the electrolyte used. This type of cell offers better voltage stability and longer life than the zinc-carbon type. It also has a longer shelf life and can operate over a wide temperature range. The manganese dioxide-alkaline-zinc cell has a voltage of 1.5 volts and is available in a wide range of sizes. This cell is commonly referred to as the alkaline cell.

The Magnesium-Manganese Dioxide Cell uses magnesium as the anode material. This allows a higher output capacity over an extended period of time compared to the zinc-carbon cell. This cell produces a voltage of approximately 2 volts. The disadvantage of this type of cell is the production of hydrogen during its operation.

The Lithium-Organic Cell and the Lithium-Inorganic Cell are recent developments of a new line of high-energy cells. The main advantages of these types of cells are very high power, operation over a wide temperature range, they are lighter than most cells, and have a remarkably long shelf life of up to 20 years.

CAUTION

Lithium cells contain toxic materials under pressure. Do not puncture, recharge, short-circuit, expose to excessively high temperatures, or incinerate. Use these batteries/cells only in approved equipment. Do not throw in trash.

Q19. Why should a mercury cell NOT be shorted?

Q20. What factors should be considered when selecting a primary cell for a power source?

SECONDARY WET CELLS

Secondary cells are sometimes known as wet cells. There are four basic type of wet cells, the lead-acid, nickel-cadmium, silver-zinc, and silver-cadmium.

Lead-Acid Cell

The lead-acid cell is the most widely used secondary cell. The previous explanation of the secondary cell describes exactly the manner in which the lead-acid cell provides electrical power. The discharging and charging action presented in electrochemical action describes the lead-acid cell.

You should recall that the lead-acid cell has an anode of lead peroxide, a cathode of sponge lead, and the electrolyte is sulfuric acid and water.

Nickel-Cadmium Cell

The nickel-cadmium cell (NICAD) is far superior to the lead-acid cell. In comparison to lead-acid cells, these cells generally require less maintenance throughout their service life in regard to the adding of electrolyte or water. The major difference between the nickel-cadmium cell and the lead-acid cell is the material used in the cathode, anode, and electrolyte. In the nickel-cadmium cell the cathode is cadmium hydroxide, the anode is nickel hydroxide, and the electrolyte is potassium hydroxide and water.

The nickel-cadmium and lead-acid cells have capacities that are comparable at normal discharge rates, but at high discharge rates the nickel-cadmium cell can deliver a larger amount of power. In addition the nickel-cadmium cell can:

1. Be charged in a shorter time,

2. Stay idle longer in any state of charge and keep a full charge when stored for a longer period of time, and

3. Be charged and discharged any number of times without any appreciable damage.

Due to their superior capabilities, nickel-cadmium cells are being used extensively in many military applications that require a cell with a high discharge rate. A good example is in the aircraft storage battery.

Silver-Zinc Cells

The silver-zinc cell is used extensively to power emergency equipment. This type of cell is relatively expensive and can be charged and discharged fewer times than other types of cells. When compared to the lead-acid or nickel-cadmium cells, these disadvantages are overweighed by the light weight, small size, and good electrical capacity of the silver-zinc cell.

The silver-zinc cell uses the same electrolyte as the nickel-cadmium cell (potassium hydroxide and water), but the anode and cathode differ from the nickel-cadmium cell. The anode is composed of silver oxide and the cathode is made of zinc.

Silver-Cadmium Cell

The silver-cadmium cell is a fairly recent development for use in storage batteries. The silver-cadmium cell combines some of the better features of the nickel-cadmium and silver-zinc cells. It has more than twice the shelf life of the silver-zinc cell and can be recharged many more times. The disadvantages of the silver-cadmium cell are high cost and low voltage production.

The electrolyte of the silver-cadmium cell is potassium hydroxide and water as in the nickel-cadmium and silver-zinc cells. The anode is silver oxide as in the silver-zinc cell and the cathode is cadmium hydroxide as in the nicad cell. You may notice that different combinations of materials are used to form the electrolyte, cathode, and anode of different cells. These combinations provide the cells with different qualities for many varied applications.

Q21. What are the four basic types of secondary (wet) cells?

Q22. What are the advantages of a nicad cell over a lead-acid cell?

Q23. What type of cell is most commonly used for emergency systems?

Q24. What three cells use the same electrolyte?

BATTERIES

A battery is a voltage source that uses chemical action to produce a voltage. In many cases the term battery is applied to a single cell, such as the flashlight battery. In the case of a flashlight that uses a battery of 1.5 volts, the battery is a single cell. The flashlight that is operated by 6 volts uses four cells in a single case and this is a battery composed of more than one cell. There are three ways to combine cells to form a battery.

COMBINING CELLS

In many cases, a battery-powered device may require more electrical energy than one cell can provide. The device may require either a higher voltage or more current, and in some cases both. Under such conditions it is necessary to combine, or interconnect, a sufficient number of cells to meet the higher requirements. Cells connected in SERIES provide a higher voltage, while cells connected in PARALLEL provide a higher current capacity. To provide adequate power when both voltage and current requirements are greater than the capacity of one cell, a combination SERIES-PARALLEL network of cells must be used.

Series-Connected Cells

Assume that a load requires a power supply of 6 volts and a current capacity of 1/8 ampere. Since a single cell normally supplies a voltage of only 1.5 volts, more than one cell is needed. To obtain the higher voltage, the cells are connected in series as shown in figure 2-6.

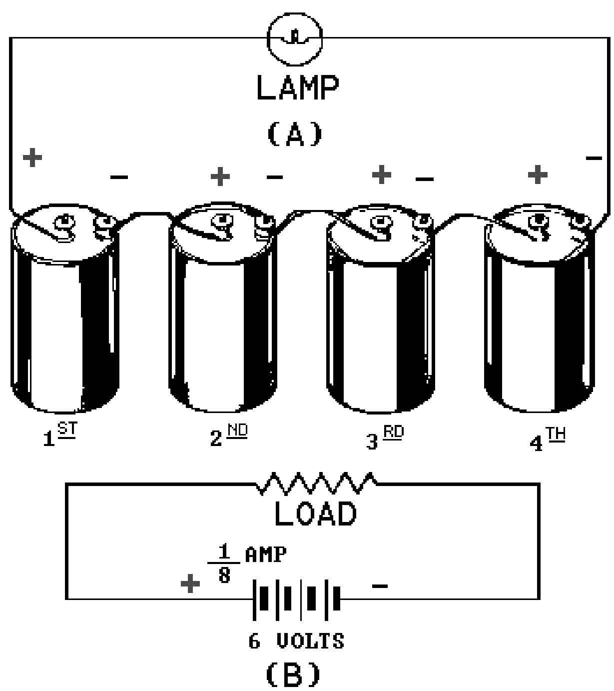


Figure 2-6.—(A) Pictorial view of series-connected cells; (B) Schematic of series connection.

Figure 2-6 view B is a schematic representation of the circuit shown in figure 2-6 view A. The load is shown by the resistance symbol and the battery is indicated by one long and one short line per cell.

In a series hookup, the negative electrode (cathode) of the first cell is connected to the positive electrode (anode) of the second cell, the negative electrode of the second to the positive of the third, etc.

The positive electrode of the first cell and negative electrode of the last cell then serve as the terminals of the battery. In this way, the voltage is 1.5 volts for each cell in the series line. There are four cells, so the output terminal voltage is 1.5 x 4, or 6 volts. When connected to the load, 1/8 ampere flows through the load and each cell of the battery. This is within the capacity of each cell. Therefore, only four series-connected cells are needed to supply this particular load.

CAUTION

When connecting cells in series, connect alternate terminals together (− to +, − to +, etc.) Always have two remaining terminals that are used for connection to the load only. Do not connect the two remaining terminals together as this is a short across the battery and would not only quickly discharge the cells but could cause some types of cells to explode.

Parallel-Connected Cells

In this case, assume an electrical load requires only 1.5 volts, but will require 1/2 ampere of current. (Assume that a single cell will supply only 1/8 ampere.) To meet this requirement, the cells are connected in parallel, as shown in figure 2-7 view A and schematically represented in 2-7 view B. In a parallel connection, all positive cell electrodes are connected to one line, and all negative electrodes are connected to the other. No more than one cell is connected between the lines at any one point; so the voltage between the lines is the same as that of one cell, or 1.5 volts. However, each cell may contribute its maximum allowable current of 1/8 ampere to the line. There are four cells, so the total line current is 1/8 x 4, or 1/2 ampere. In this case four cells in parallel have enough capacity to supply a load requiring 1/2 ampere at 1.5 volts.

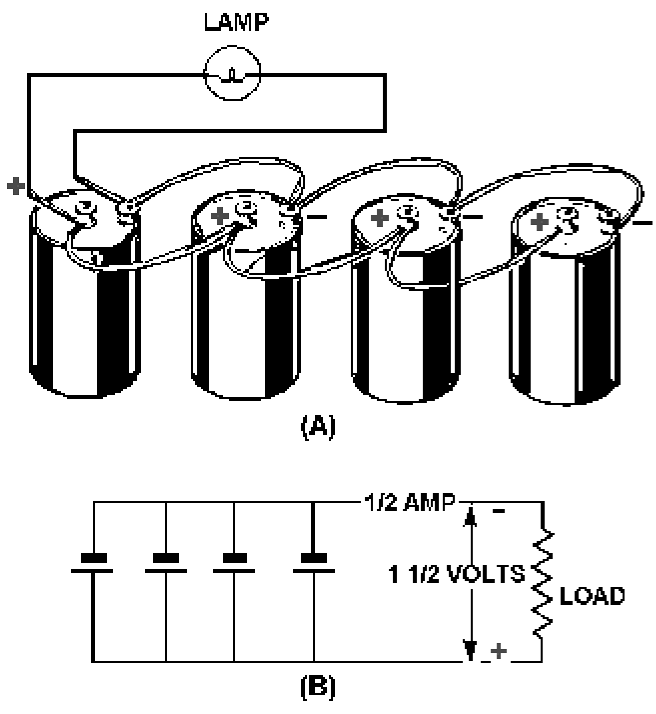


Figure 2-7.—(A) Pictorial view of parallel-connected cells; (B) Schematic of parallel connection.

Series-Parallel-Connected Cells

Figure 2-8 depicts a battery network supplying power to a load requiring both a voltage and a current greater than one cell can provide. To provide the required 4.5 volts, groups of three 1.5-volt cells are connected in series. To provide the required 1/2 ampere of current, four series groups are connected in parallel, each supplying 1/8 ampere of current.

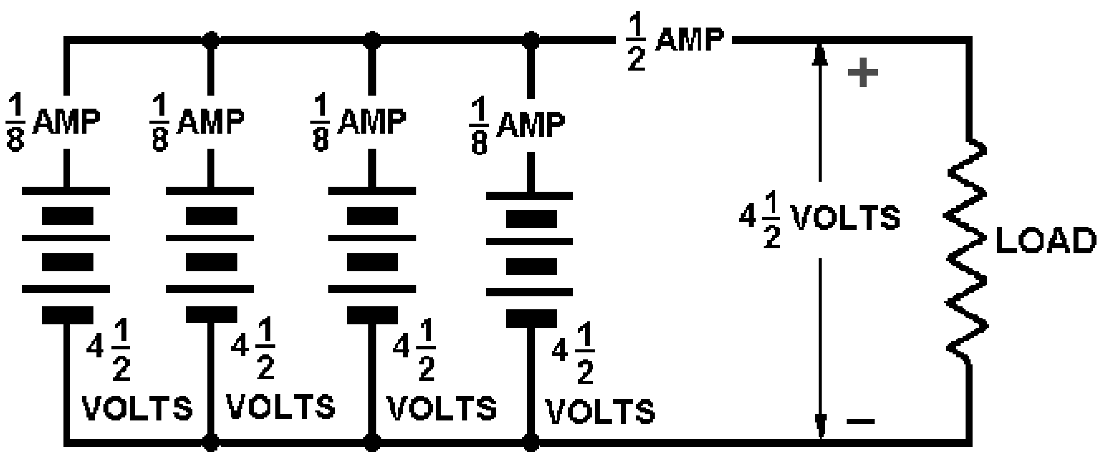


Figure 2-8.—Schematic of series-parallel connected cells.

The connections shown have been used to illustrate the various methods of combining cells to form a battery. Series, parallel, and series-parallel circuits will be covered in detail in the next chapter, "Direct Current."

Some batteries are made from primary cells. When a primary-cell battery is completely discharged, the entire battery must be replaced. Because there is nothing else that can be done to primary cell batteries, the rest of the discussion on batteries will be concerned with batteries made of secondary cells.

Q25. What does the term battery normally refer to?

Q26. What are the three ways of combining cells, and what is each used for?

BATTERY CONSTRUCTION

Secondary cell batteries are constructed using the various secondary cells already described. The lead-acid battery is one of the most common batteries in use today and will be used to explain battery construction. The nickel-cadmium battery is being used with increasing frequency and will also be discussed.

Figure 2-9 shows the makeup of a lead-acid battery. The container houses the separate cells. Most containers are hard rubber, plastic, or some other material that is resistant to the electrolyte and mechanical shock and will withstand extreme temperatures. The container (battery case) is vented through vent plugs to allow the gases that form within the cells to escape. The plates in the battery are the cathodes and anodes that were discussed earlier. In figure 2-10 the negative plate group is the cathode of the individual cells and the positive plate group is the anode. As shown in the figure, the plates are interlaced with a terminal attached to each plate group. The terminals of the individual cells are connected together by link connectors as shown in figure 2-9. The cells are connected in series in the battery and the positive terminal of one end cell becomes the positive terminal of the battery. The negative terminal of the opposite end cell becomes the negative terminal of the battery.

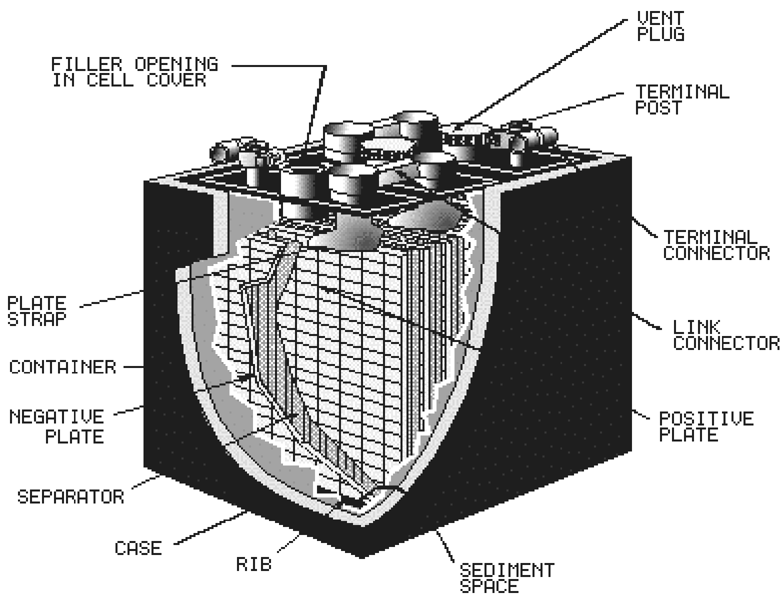


Figure 2-9.—Lead-acid battery construction.

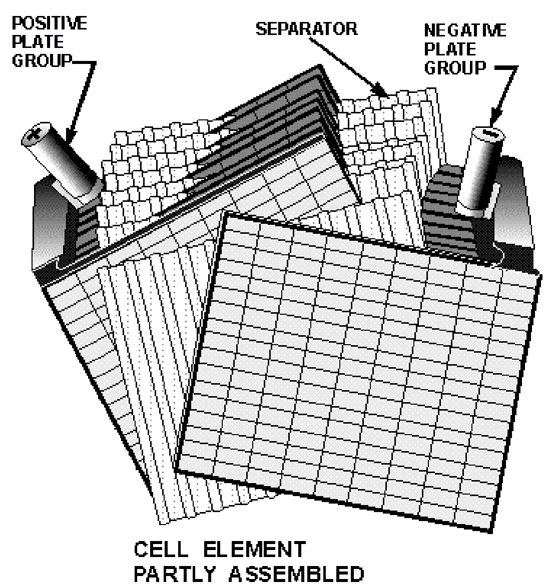


Figure 2-10.—Lead-acid battery plate arrangement.

The terminals of a lead-acid battery are usually identified from one another by their size and markings. The positive terminal, marked (+) is sometimes colored red and is physically larger than the negative terminal, marked (–).

The individual cells of the lead-acid battery are not replaceable, so in the event one cell fails the battery must be replaced.

The nickel-cadmium battery is similar in construction to the lead-acid battery with the exception that it has individual cells which can be replaced. The cell of the nicad battery is shown in figure 2-11.

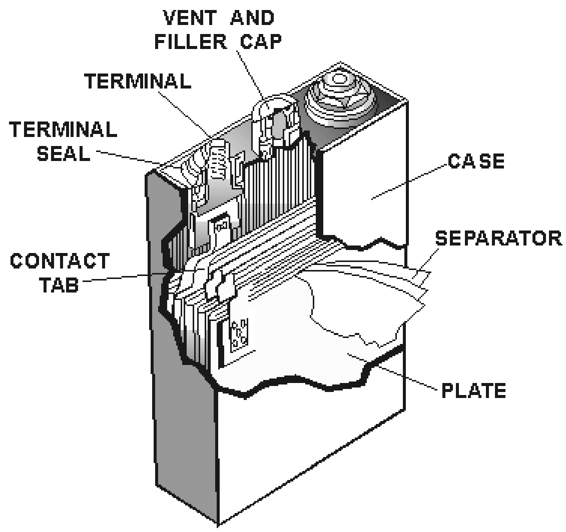


Figure 2-11.—Nickel-cadmium cell.

The construction of secondary cell batteries is so similar, that it is difficult to distinguish the type of battery by simply looking at it. The type of battery must be known to properly check or recharge the battery. Each battery should have a nameplate that gives a description of its type and electrical characteristics.

Q27. Other than the type of cell used, what is the major difference between the construction of the lead-acid and nicad battery?

Q28. How is the type of battery most easily determined?

BATTERY MAINTENANCE

The following information concerns the maintenance of secondary-cell batteries and is of a general nature. <u>You must check the appropriate technical manuals for the specific type of battery prior to performing maintenance on any battery</u>.

Specific Gravity

For a battery to work properly, its electrolyte (water plus active ingredient) must contain a certain amount of active ingredient. Since the active ingredient is dissolved in the water, the *amount* of active ingredient cannot be measured directly. An indirect way to determine whether or not the electrolyte contains the proper amount of active ingredient is to measure the electrolyte's **specific gravity**. Specific gravity is the ratio of the weight of a certain amount of a given substance compared to the weight of the same amount of pure water. The specific gravity of pure water is 1.0. Any substance that floats has a specific gravity less than 1.0. Any substance that sinks has a specific gravity greater than 1.0.

The active ingredient in electrolyte (sulfuric acid, potassium hydroxide, etc.) is heavier than water. Therefore, the electrolyte has a specific gravity greater than 1.0. The acceptable range of specific gravity for a given battery is provided by the battery's manufacturer. To measure a battery's specific gravity, use an instrument called a HYDROMETER.

The Hydrometer

A hydrometer, shown in figure 2-12, is a glass syringe with a float inside it. The float is a hollow glass tube sealed at both ends and weighted at the bottom end, with a scale calibrated in specific gravity marked on its side. To test an electrolyte, draw it into the hydrometer using the suction bulb. Draw enough electrolyte into the hydrometer to make the float rise. Do not draw in so much electrolyte that the float rises into the suction bulb. The float will rise to a point determined by the specific gravity of the electrolyte. If the electrolyte contains a large amount of active ingredient, its specific gravity will be relatively high. The float will rise higher than it would if the electrolyte contained only a small amount of active ingredient.

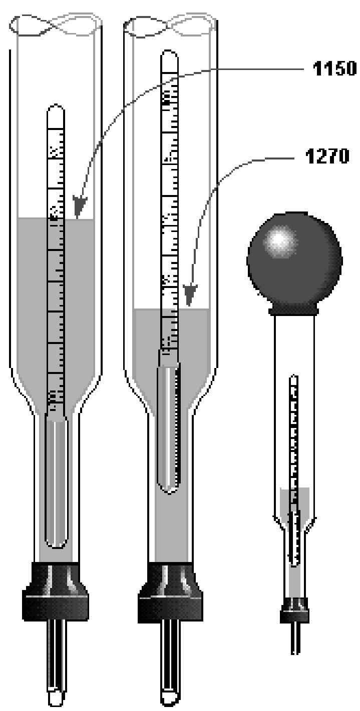


Figure 2-12.—Hydrometer.

To read the hydrometer, hold it in a vertical position and read the scale at the point that surface of the electrolyte touches the float. Refer to the manufacturer's technical manual to determine whether or not the battery's specific gravity is within specifications.

Note: Hydrometers should be flushed with fresh water after each use to prevent inaccurate readings. Storage battery hydrometers must not be used for any other purpose.

Q29. What is the purpose of the hydrometer?

Q30. Which electrolyte has more active ingredient? Electrolyte A, specific gravity 1.015? Electrolyte B, specific gravity 1.125?

Other Maintenance

The routine maintenance of a battery is very simple. Terminals should be checked periodically for cleanliness and good electrical connection. The battery case should be inspected for cleanliness and evidence of damage. The level of electrolyte should be checked and if the electrolyte is low, distilled water should be added to bring the electrolyte to the proper level. Maintenance procedures for batteries are normally determined by higher authority and each command will have detailed procedures for battery care and maintenance.

Safety Precautions With Batteries

All types of batteries should be handled with care:

1. NEVER SHORT THE TERMINALS OF A BATTERY.
2. CARRYING STRAPS SHOULD BE USED WHEN TRANSPORTING BATTERIES.
3. PROTECTIVE CLOTHING, SUCH AS RUBBER APRON, RUBBER GLOVES, AND A FACE SHIELD SHOULD BE WORN WHEN WORKING WITH BATTERIES.
4. NO SMOKING, ELECTRIC SPARKS, OR OPEN FLAMES SHOULD BE PERMITTED NEAR CHARGING BATTERIES.
5. CARE SHOULD BE TAKEN TO PREVENT SPILLING OF THE ELECTROLYTE.

In the event electrolyte is splashed or spilled on a surface, such as the floor or table, it should be diluted with large quantities of water and cleaned up immediately.

If the electrolyte is spilled or splashed on the skin or eyes, IMMEDIATELY flush the skin or eyes with large quantities of fresh water for a minimum of 15 minutes. If the electrolyte is in the eyes, be sure the upper and lower eyelids are pulled out sufficiently to allow the fresh water to flush under the eyelids. The medical department should be notified as soon as possible and informed of the type of electrolyte and the location of the accident.

CAPACITY AND RATING OF BATTERIES

The CAPACITY of a battery is measured in ampere-hours. The ampere-hour capacity is equal to the product of the current in amperes and the time in hours during which the battery will supply this current. The ampere-hour capacity varies inversely with the discharge current. For example, a 400 ampere-hour battery will deliver 400 amperes for 1 hour or 100 amperes for 4 hours.

Storage batteries are RATED according to their rate of discharge and ampere-hour capacity. Most batteries are rated according to a 20-hour rate of discharge. That is, if a fully charged battery is completely discharged during a 20-hour period, it is discharged at the 20-hour rate. Thus, if a battery can deliver 20 amperes continuously for 20 hours, the battery has a rating of 20 amperes x 20 hours, or 400 ampere-hours. Therefore, the 20-hour rating is equal to the average current that a battery is capable of supplying without interruption for an interval of 20 hours. (Note: Aircraft batteries are rated according to a 1-hour rate of discharge.)

All standard batteries deliver 100 percent of their available capacity if discharged in 20 hours or more, but they will deliver less than their available capacity if discharged at a faster rate. The faster they discharge, the less ampere-hour capacity they have.

The low-voltage limit, as specified by the manufacturer, is the limit beyond which very little useful energy can be obtained from a battery. This low-voltage limit is normally a test used in battery shops to determine the condition of a battery.

Q31. When should safety precautions pertaining to batteries be observed?

Q32. How long should a 200 ampere-hour battery be able to deliver 5 amperes?

BATTERY CHARGING

It should be remembered that adding the active ingredient to the electrolyte of a discharged battery does not recharge the battery. Adding the active ingredient only increases the specific gravity of the electrolyte and does not convert the plates back to active material, and so does not bring the battery back to a charged condition. A charging current must be passed through the battery to recharge it.

Batteries are usually charged in battery shops. Each shop will have specific charging procedures for the types of batteries to be charged. The following discussion will introduce you to the types of battery charges.

The following types of charges may be given to a storage battery, depending upon the condition of the battery:

1. Initial charge
2. Normal charge
3. Equalizing charge
4. Floating charge
5. Fast charge

Initial Charge

When a new battery is shipped dry, the plates are in an uncharged condition. After the electrolyte has been added, it is necessary to charge the battery. This is accomplished by giving the battery a long low-rate initial charge. The charge is given in accordance with the manufacturer's instructions, which are shipped with each battery. If the manufacturer's instructions are not available, reference should be made to the detailed instructions for charging batteries found in current Navy directives.

Normal Charge

A normal charge is a routine charge that is given in accordance with the nameplate data during the ordinary cycle of operation to restore the battery to its charged condition.

Equalizing Charge

An equalizing charge is a special extended normal charge that is given periodically to batteries as part of a maintenance routine. It ensures that all the sulfate is driven from the plates and that all the cells are restored to a maximum specific gravity. The equalizing charge is continued until the specific gravity of all cells, corrected for temperature, shows no change for a 4-hour period.

Floating Charge

In a floating charge, the charging rate is determined by the battery voltage rather than by a definite current value. The floating charge is used to keep a battery at full charge while the battery is idle or in light duty. It is sometimes referred to as a trickle charge and is accomplished with low current.

Fast Charge

A fast charge is used when a battery must be recharged in the shortest possible time. The charge starts at a much higher rate than is normally used for charging. It should be used only in an emergency, as this type charge may be harmful to the battery.

Charging Rate

Normally, the charging rate of Navy storage batteries is given on the battery nameplate. If the available charging equipment does not have the desired charging rates, the nearest available rates should be used. However, the rate should never be so high that violent gassing (explained later in this text) occurs.

Charging Time

The charge must be continued until the battery is fully charged. Frequent readings of specific gravity should be taken during the charge and compared with the reading taken before the battery was placed on charge.

Gassing

When a battery is being charged, a portion of the energy breaks down the water in the electrolyte. Hydrogen is released at the negative plates and oxygen at the positive plates. These gases bubble up through the electrolyte and collect in the air space at the top of the cell. If violent gassing occurs when the battery is first placed on charge, the charging rate is too high. If the rate is not too high, steady gassing develops as the charging proceeds, indicating that the battery is nearing a fully charged condition.

WARNING

A mixture of hydrogen and air can be dangerously explosive. No smoking, electric sparks, or open flames should be permitted near charging batteries.

Q33. Can a battery be recharged by adding more electrolyte?

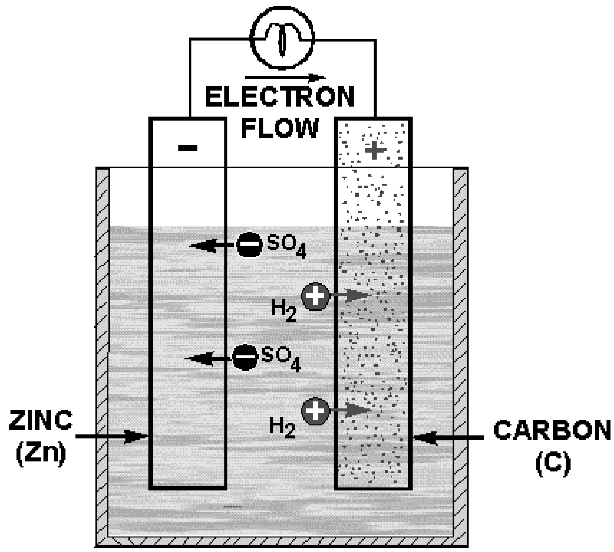
Q34. If violent gassing occurs during a battery charge, what action should be taken?

SUMMARY

In this chapter you learned that batteries are widely used as sources of direct-current. You were introduced to electrochemical action and the way it works in a cell, the cell itself, the type and parts of a cell, and how cells are connected together to form batteries. You learned the construction and maintenance of batteries and some of the safety precautions in handling and working with batteries.

Several new terms were introduced in this chapter. The following is a summary of the chapter on batteries.

A CELL is a device that transforms chemical energy into electrical energy. The cell has three parts; the electrodes, the electrolyte, and the container. There are two basic cells: primary and secondary.



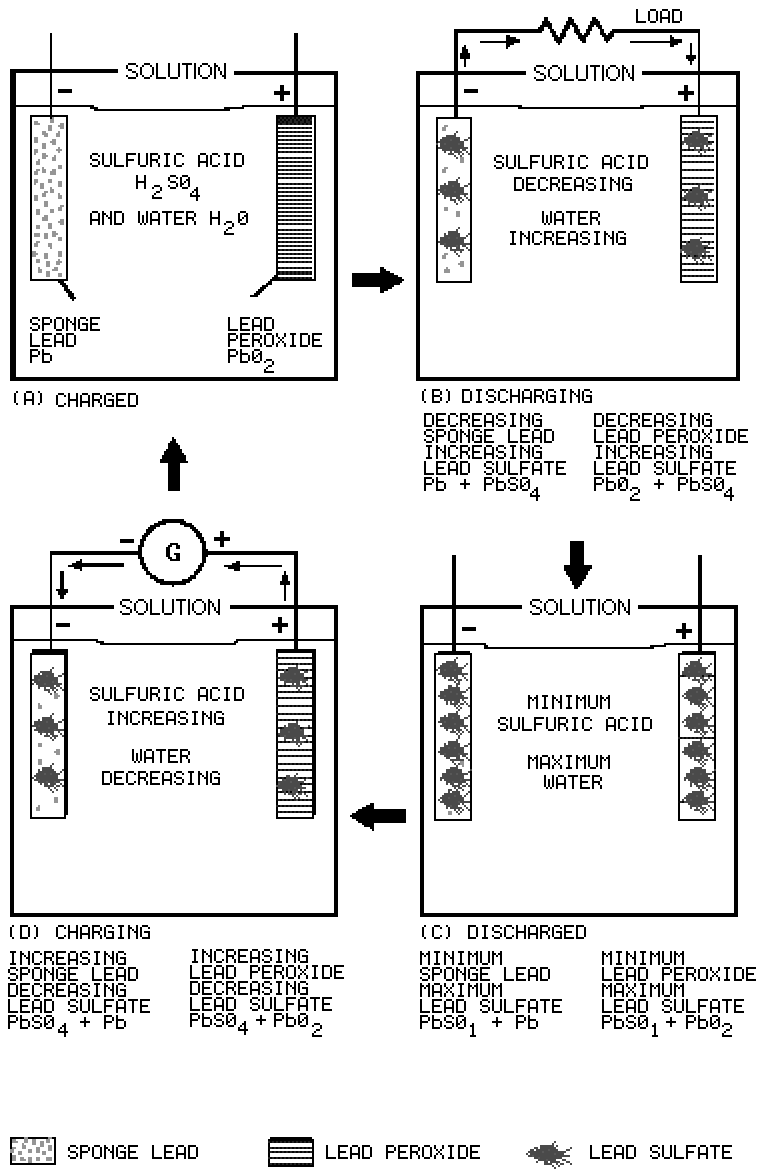
THE ELECTRODES are the current conductors of the cell.

THE ELECTROLYTE is the solution that acts upon the electrodes.

THE CONTAINER holds the electrolyte and provides a means of mounting the electrodes.

THE PRIMARY CELL is a cell in which the chemical action finally destroys one of the electrodes, usually the negative. The primary cell cannot be recharged.

THE SECONDARY CELL is a cell in which the chemical action alters the electrodes and electrolyte. The electrodes and electrolyte can be restored to their original condition by recharging the cell.

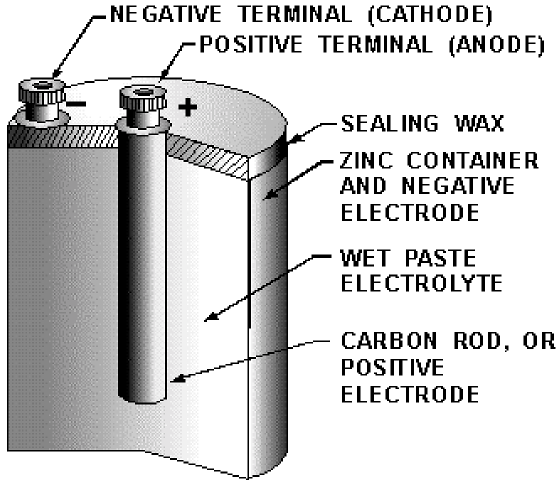


ELECTROCHEMICAL ACTION is the process of converting chemical energy into electrical energy.

THE ANODE is the positive electrode of a cell.

THE CATHODE is the negative electrode of a cell.

PRIMARY CELL CHEMISTRY is the process in which electrons leaving the cathode to the load cause a positive charge which attracts negative ions from the electrolyte. The negative ions combine with the material of the cathode and form a substance such as lead-sulfate. Electrons from the load to the anode create a negative charge which attracts positive ions (hydrogen) from the electrolyte.

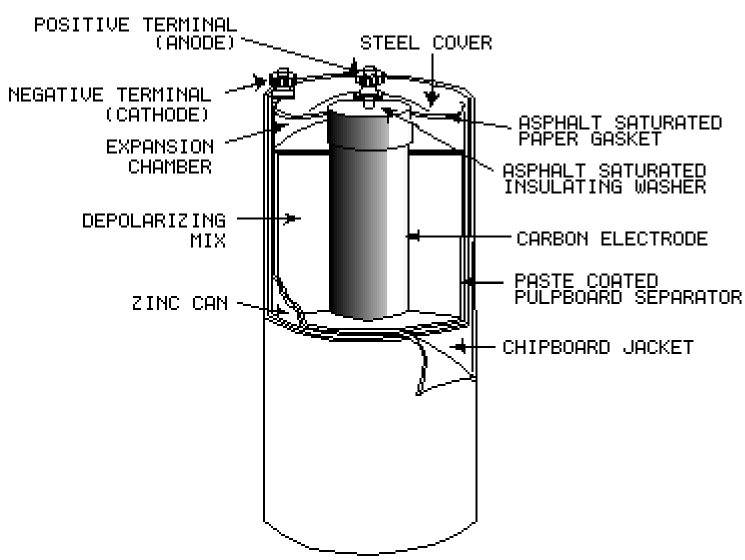


SECONDARY CELL CHEMISTRY is the process in which the electrolyte acts upon and chemically changes both electrodes. This process also depletes the amount of active material in the electrolyte. A charging current applied to the cell reverses the process and restores the cell to its original condition.

POLARIZATION is the effect of hydrogen surrounding the anode of a cell which increases the internal resistance of the cell. Polarization can be prevented by venting the cell, adding a material rich in oxygen, or adding a material that will absorb hydrogen.

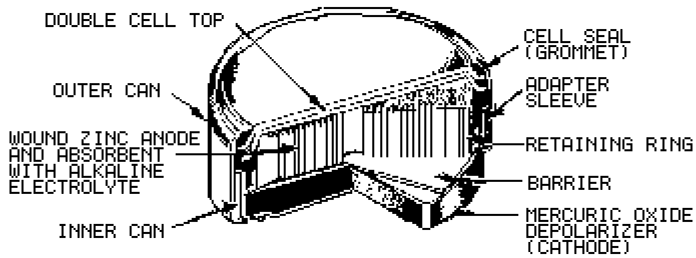
LOCAL ACTION is the continuation of current flow within the cell when there is no external load. It is caused by impurities in the electrode and can be prevented by the use of mercury amalgamated with the material of the electrode.

DRY CELL is the type commonly referred to as the "flashlight battery." Since the electrolyte is not in liquid form, but is a paste, the term dry cell is used. In most dry cells the case is the cathode.



SHELF LIFE is the period the cell may be stored and still be usable.

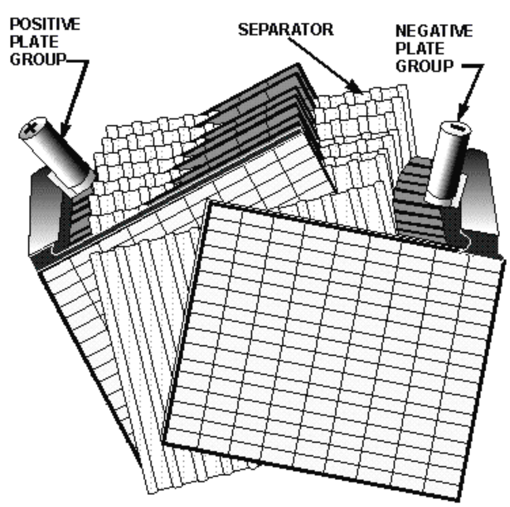
MERCURY CELLS should never be shorted because of the danger of explosion.



WOUND ANODE FLAT

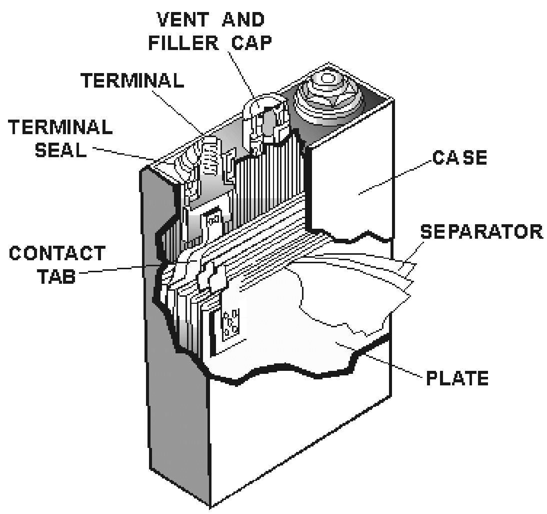
DRY CELLS are of many types, each having advantages and disadvantages. The type selected for use depends on such factors as cost, size, ease of replacement, and voltage or current needs.

THE LEAD-ACID CELL is the most widely used secondary cell. The lead-acid cell produces electricity by electrochemical action. The anode is lead peroxide, the cathode is sponge lead, and the electrolyte is sulfuric acid and water.



CELL ELEMENT
PARTLY ASSEMBLED

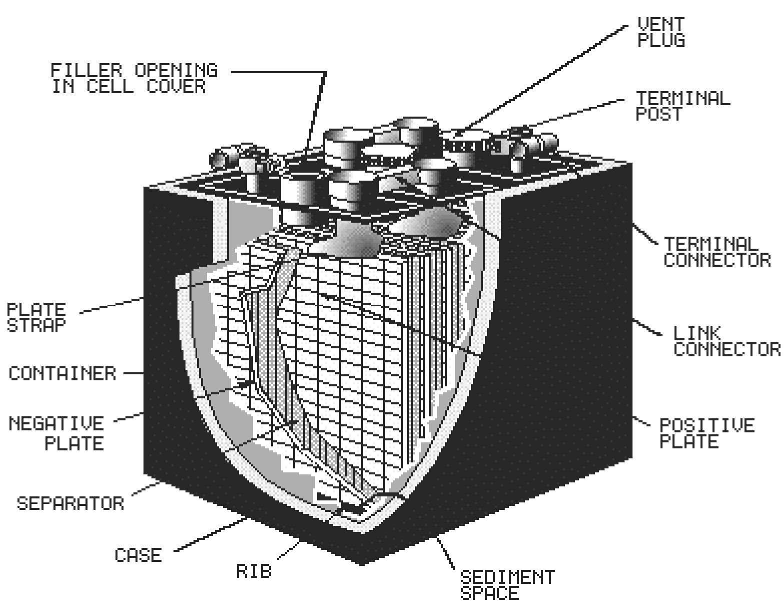
THE NICKEL-CADMIUM CELL, commonly called the NICAD, has the following advantages over the lead-acid cell; charges in a shorter period of time, delivers a larger amount of power, stays idle longer, and can be charged and discharged many times. The anode is nickel hydroxide, the cathode is cadmium hydroxide, and the electrolyte is potassium hydroxide and water.



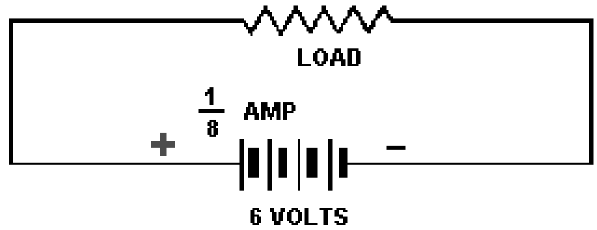
THE SILVER-ZINC CELL is used mostly for emergency equipment. It is light, small, and has a large power capacity for its size. The anode is silver oxide, the cathode is zinc, and the electrolyte is potassium hydroxide and water.

THE SILVER-CADMIUM CELL combines the better features of the nickel-cadmium and silver-zinc cells. The anode is silver-oxide, the cathode is cadmium hydroxide, and the electrolyte is potassium hydroxide.

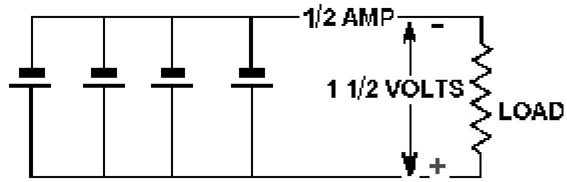
A BATTERY is a voltage source in a single container made from one or more cells. The cells can be combined in series, parallel, or series-parallel.



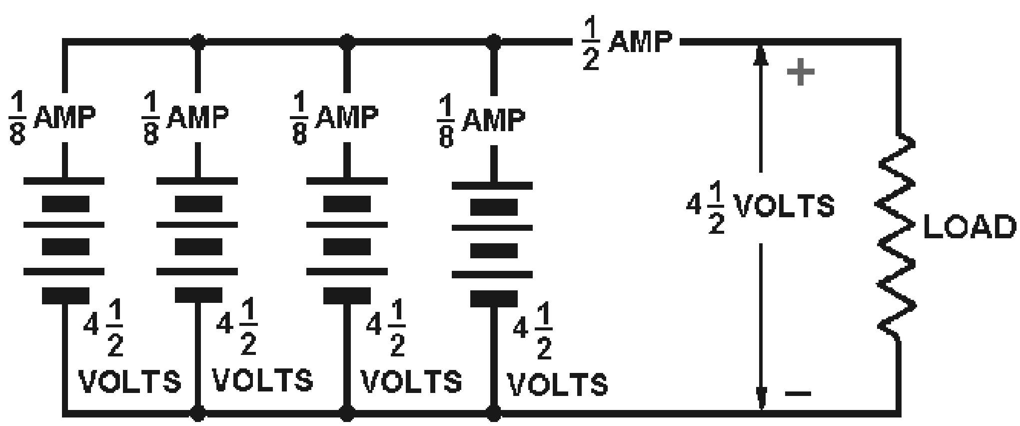
SERIES CONNECTED CELLS provide a higher voltage than a single cell, with no increase in current.



PARALLEL CONNECTED CELLS provide a higher current than a single cell, with no increase in voltage.

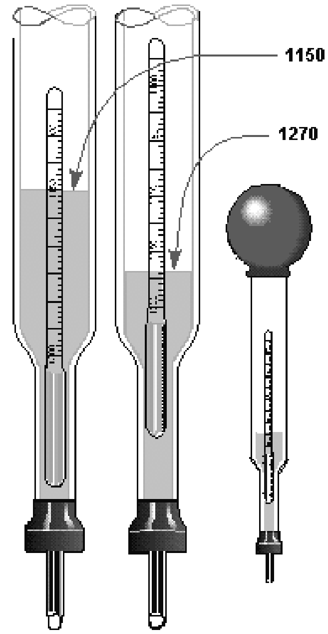


SERIES-PARALLEL CONNECTED CELLS provide a higher voltage and a higher current than a single cell.



TYPES OF BATTERIES can be determined from nameplate data.

HYDROMETER provides the means to check the specific gravity of the electrolyte.



SAFETY PRECAUTIONS should always be followed when working with or around batteries.

CAPACITY is an indication of the current-supplying capability of the battery for a specific period of time; e.g., 400 ampere-hour.

RATING is the capacity of the battery for a specific rate of discharge. In most batteries the rating is given for a 20 hour discharge cycle; e.g., 20 amperes for 20 hours.

BATTERY CHARGE is the process of reversing the current flow through the battery to restore the battery to its original condition. The addition of active ingredient to the electrolyte will not recharge the battery. There are five types of charges:

1. Initial charge
2. Normal charge
3. Equalizing charge
4. Floating charge
5. Fast charge

GASSING is the production of hydrogen gas caused by a portion of the charge current breaking down the water in the electrolyte. Steady gassing is normal during the charging process. Violent gassing indicates that the charge rate is too high.

ANSWERS TO QUESTIONS Q1. THROUGH Q34.

A1. A cell is a device that converts chemical energy to electrical energy.

A2. The electrodes, the electrolyte, and the container.

A3. The electrodes are the current conductors of the cell. The electrolyte is the solution that acts upon the electrodes. The container holds the electrolyte and provides a means of mounting the electrodes.

A4. Primary and secondary.

A5. The secondary cell can be restored to its original condition by an electric current. The primary cell cannot.

A6. The process of converting chemical energy into electrical energy.

A7. (a) The anode, (b) the cathode.

A8. The positive charge caused by electrons leaving the negative electrode attracts the negative ions.

A9. By current flow through the load.

A10. The chemical action between the negative electrode and the electrolyte.

A11. The sulfuric acid is chemically acting upon the anode and cathode which creates a current flow through the load.

A12. The charging currents causes the lead sulfate in the anode and cathode to be changed back to lead peroxide, sponge lead, and sulfuric acid.

A13. Fully charged.

A14. Vent the cell, add a material rich in oxygen, and use a material that will absorb hydrogen.

A15. Current flow in a cell with no external load.

A16. The zinc container.

A17. The electrolyte is not a liquid but is in the form of a paste.

A18. The period that a cell can be stored and still be useable.

A19. The danger of explosion.

A20. Cost, size, ease of replacement, and voltage or current needs.

A21. Lead-acid, nickel-cadmium (NICAD), silver-zinc, and silver-cadmium.

A22. Can be charged in a shorter time, can deliver a larger amount of power, and stays idle longer.

A23. Silver-zinc cell.

A24. Silver-cadmium, silver-zinc, and nickel-cadmium.

A25. A voltage source in a single container made from one or more cells.

A26. Series, to increase voltage but not current. Parallel, to increase current but not voltage. Series-Parallel, to increase both current and voltage.

A27. The cells in the nicad battery can be replaced.

A28. By looking at the nameplate data.

A29. To measure the amount of active ingredient in the electrolyte.

A30. Electrolyte B. It is heavier per unit volume.

A31. At all times.

A32. Forty hours.

A33. No, a current must be passed through the battery.

A34. Reduce the charging rate.

CHAPTER 3

DIRECT CURRENT

LEARNING OBJECTIVES

Upon completing this chapter, you will be able to:

1. Identify the term schematic diagram and identify the components in a circuit from a simple schematic diagram.
2. State the equation for Ohm's law and describe the effects on current caused by changes in a circuit.
3. Given simple graphs of current versus power and voltage versus power, determine the value of circuit power for a given current and voltage.
4. Identify the term power, and state three formulas for computing power.
5. Compute circuit and component power in series, parallel, and combination circuits.
6. Compute the efficiency of an electrical device.
7. Solve for unknown quantities of resistance, current, and voltage in a series circuit.
8. Describe how voltage polarities are assigned to the voltage drops across resistors when Kirchhoff's voltage law is used.
9. State the voltage at the reference point in a circuit.
10. Define open and short circuits and describe their effects on a circuit.
11. State the meaning of the term source resistance and describe its effect on a circuit.
12. Describe in terms of circuit values the circuit condition needed for maximum power transfer.
13. Compute efficiency of power transfer in a circuit.
14. Solve for unknown quantities of resistance, current, and voltage in a parallel circuit.
15. State the significance of the polarity assigned to a current when using Kirchhoff's current law.
16. State the meaning of the term equivalent resistance.
17. Compute resistance, current, voltage, and power in voltage dividers.
18. Describe the method by which a single voltage divider can provide both positive and negative voltages.
19. Recognize the safety precautions associated with the hazard of electrical shock.
20. Identify the first aid procedures for a victim of electrical shock.

INTRODUCTION

The material covered in this chapter contains many new terms that are explained as you progress through the material. The basic dc circuit is the easiest to understand, so the chapter begins with the basic circuit and from there works into the basic schematic diagram of that circuit. The schematic diagram is used in all your future work in electricity and electronics. It is very important that you become familiar with the symbols that are used.

This chapter also explains how to determine the total resistance, current, voltage, and power in a series, parallel, or combination circuit through the use of Ohm's and Kirchhoff's laws. The voltage divider network, series, parallel, and series-parallel practice problem circuits will be used for practical examples of what you have learned.

THE BASIC ELECTRIC CIRCUIT

The flashlight is an example of a basic electric circuit. It contains a source of electrical energy (the dry cells in the flashlight), a load (the bulb) which changes the electrical energy into a more useful form of energy (light), and a switch to control the energy delivered to the load.

Before you study a schematic representation of the flashlight, it is necessary to define certain terms. The LOAD is any device through which an electrical current flows and which changes this electrical energy into a more useful form. Some common examples of loads are a lightbulb, which changes electrical energy to light energy; an electric motor, which changes electrical energy into mechanical energy; and the speaker in a radio, which changes electrical energy into sound. The SOURCE is the device which furnishes the electrical energy used by the load. It may consist of a simple dry cell (as in a flashlight), a storage battery (as in an automobile), or a power supply (such as a battery charger). The SWITCH, which permits control of the electrical device, interrupts the current delivered to the load.

SCHEMATIC REPRESENTATION

The technician's main aid in troubleshooting a circuit in a piece of equipment is the SCHEMATIC DIAGRAM. The schematic diagram is a "picture" of the circuit that uses symbols to represent the various circuit components; physically large or complex circuits can be shown on a relatively small diagram. Before studying the basic schematic, look at figure 3-1. This figure shows the symbols that are used in this chapter. These, and others like them, are referred to and used throughout the study of electricity and electronics.

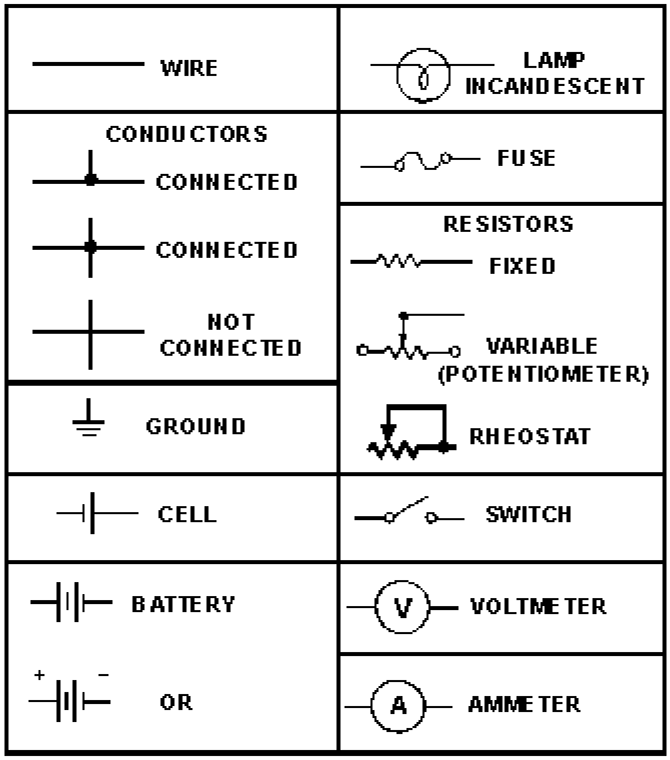


Figure 3-1.—Symbols commonly used in electricity.

The schematic in figure 3-2 represents a flashlight. View A of the figure shows the flashlight in the off or deenergized state. The switch (S1) is open. There is no complete path for current (I) through the circuit, and the bulb (DS1) does not light. In figure 3-2 view B, switch S1 is closed. Current flows in the direction of the arrows from the negative terminal of the battery (BAT), through the switch (S1), through the lamp (DS1), and back to the positive terminal of the battery. With the switch closed the path for current is complete. Current will continue to flow until the switch (S1) is moved to the open position or the battery is completely discharged.

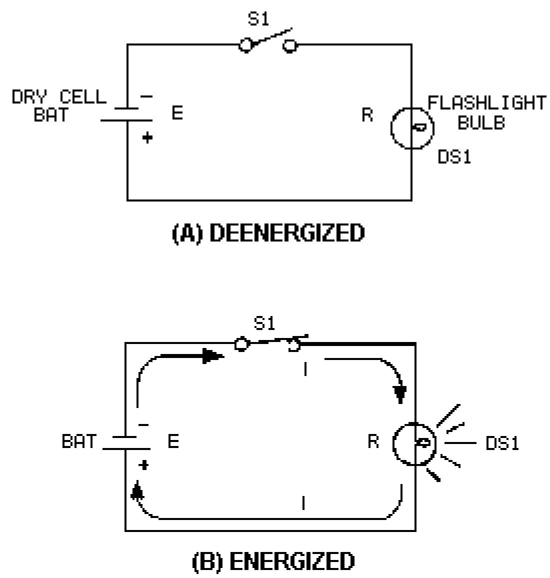


Figure 3-2.—Basic flashlight schematic.

Q1. In figure 3-2, what part of the circuit is the (a) load and (b) source?

Q2. What happens to the path for current when S1 is open as shown in figure 3-2(A)?

Q3. What is the name given to the "picture" of a circuit such as the one shown in figure 3-2?

OHM'S LAW

In the early part of the 19th century, George Simon Ohm proved by experiment that a precise relationship exists between current, voltage, and resistance. This relationship is called Ohm's law and is stated as follows:

The current in a circuit is DIRECTLY proportional to the applied voltage and INVERSELY proportional to the circuit resistance. Ohm's law may be expressed as an equation:

$$I = \frac{E}{R}$$

Where: I = current in amperes
E = voltage in volts
R = resistance in ohms

As stated in Ohm's law, current is inversely proportional to resistance. <u>This means, as the resistance in a circuit increases, the current decreases proportionately</u>.

In the equation

$$I = \frac{E}{R}$$

if any two quantities are known, the third one can be determined. Refer to figure 3-2(B), the schematic of the flashlight. If the battery (BAT) supplies a voltage of 1.5 volts and the lamp (DS1) has a resistance of 5 ohms, then the current in the circuit can be determined. Using this equation and substituting values:

$$I = \frac{E}{R} = \frac{1.5\text{ volts}}{5\text{ ohms}} = .3\text{ ampere}$$

If the flashlight were a two-cell flashlight, we would have twice the voltage, or 3.0 volts, applied to the circuit. Using this voltage in the equation:

$$I = \frac{E}{R} = \frac{3.0\text{ volts}}{5\text{ ohms}} = .6\text{ ampere}$$

You can see that the current has doubled as the voltage has doubled. This demonstrates that the <u>current is directly proportional to the applied voltage.</u>

If the value of resistance of the lamp is doubled, the equation will be:

$$I = \frac{E}{R} = \frac{3.0\text{ volts}}{10\text{ ohms}} = .3\text{ ampere}$$

The current has been reduced to one half of the value of the previous equation, or .3 ampere. This demonstrates that the <u>current is inversely proportional to the resistance</u>. Doubling the value of the resistance of the load reduces circuit current value to one half of its former value.

APPLICATION OF OHM'S LAW

By using Ohm's law, you are able to find the resistance of a circuit, knowing only the <u>voltage</u> and the <u>current</u> in the circuit.

In any equation, if all the variables (parameters) are known except one, that unknown can be found. For example, using Ohm's law, if current (I) and voltage (E) are known, resistance (R) the only parameter not known, can be determined:

1. Basic formula:

$$I = \frac{E}{R}$$

2. Remove the divisor by multiplying both sides by R:

$$R \times I = \frac{E}{\cancel{R}} \times \frac{\cancel{R}}{1}$$

3. Result of step 2: R x I = E

4. To get R alone (on one side of the equation) divide both sides by I:

$$\frac{R\cancel{I}}{\cancel{I}} = \frac{E}{I}$$

5. The basic formula, transposed for R, is:

$$R = \frac{E}{I}$$

Refer to figure 3-3 where E equals 10 volts and I equals 1 ampere. Solve for R, using the equation just explained.

Given: E = 10 volts

I = 1 ampere

Solution:

$$R = \frac{E}{I}$$

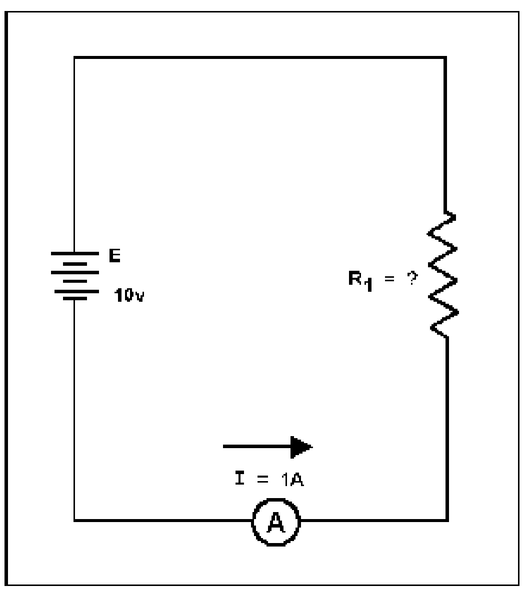


Figure 3-3.—Determining resistance in a basic circuit.

Insert the values of the known quantities:

$$R = \frac{10 \text{ volts}}{1 \text{ ampere}}$$

$$R = 10 \text{ ohms}$$

The basic formula can also be used to solve for E:

Take the basic formula: $I = \frac{E}{R}$

multiply both sides by R:

$$I \times R = \frac{E}{R} \times \frac{R}{1}$$

Results: $E = I \times R$

This equation can be used to find the voltage for the circuit shown in figure 3-4.

Given: $I = .5$ ampere
$R = 45$ ohms

Solution: $E = I \times R$
$E = .5 \text{ ampere} \times 45 \text{ ohms}$
$E = 22.5 \text{ volts}$

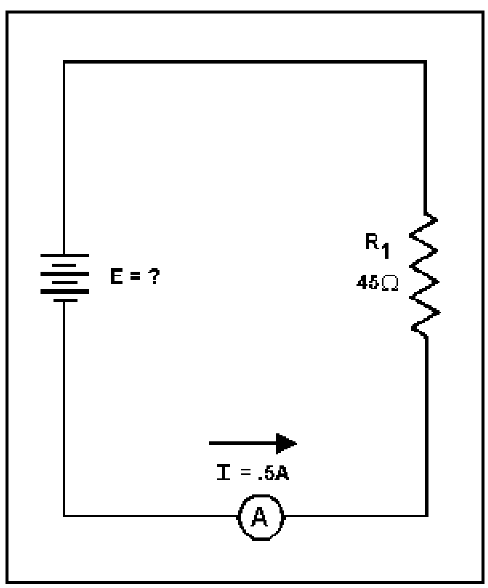


Figure 3-4.—Determining voltage in a basic circuit.

The Ohm's law equation and its various forms may be obtained readily with the aid of figure 3-5. The circle containing E, I, and R is divided into two parts, with E above the line and with I and R below the line. To determine the unknown quantity, first cover that quantity with a finger. The position of the uncovered letters in the circle will indicate the mathematical operation to be performed. For example, to find I, cover I with a finger. The uncovered letters indicate that E is to be divided by R, or

$$I = \frac{E}{R}$$

To find the formula for E, cover E with your finger. The result indicates that I is to be multiplied by R, or E = IR. To find the formula for R, cover R. The result indicates that E is to be divided by I, or

$$R = \frac{E}{I}$$

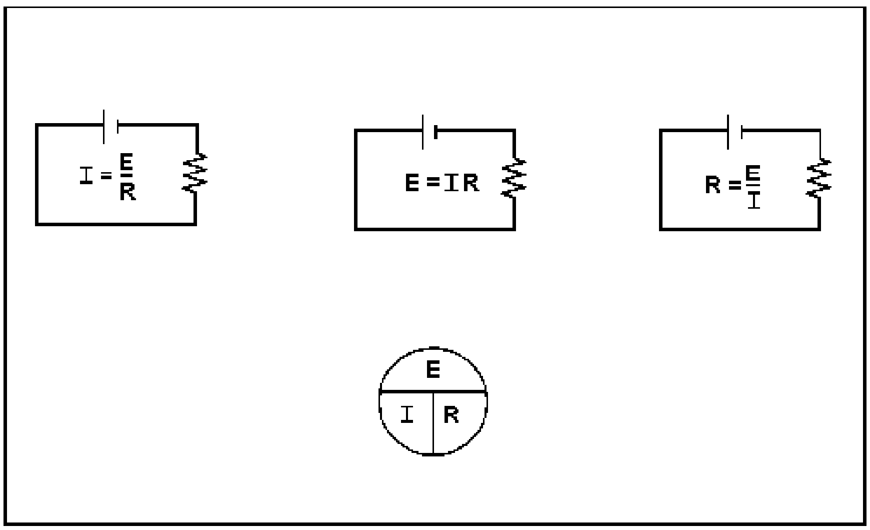


Figure 3-5.—Ohm's law in diagram form.

You are cautioned not to rely wholly on the use of this diagram when you transpose the Ohm's law formulas. The diagram should be used to supplement your knowledge of the algebraic method. Algebra is a basic tool in the solution of electrical problems.

Q4. According to Ohm's law, what happens to circuit current if the applied voltage (a) increases, (b) decreases?

Q5. According to Ohm's law, what happens to circuit current if circuit resistance (a) increases, (b) decreases?

Q6. What is the equation used to find circuit resistance if voltage and current values are known?

GRAPHICAL ANALYSIS OF THE BASIC CIRCUIT

One of the most valuable methods of analyzing a circuit is by constructing a graph. No other method provides a more convenient or more rapid way to observe the characteristics of an electrical device.

The first step in constructing a graph is to obtain a table of data. The information in the table can be obtained by taking measurements on the circuit under examination, or can be obtained theoretically through a series of Ohm's law computations. The latter method is used here.

Since there are three variables (E, I, and R) to be analyzed, there are three distinct graphs that may be constructed.

To construct any graph of electrical quantities, it is standard practice to vary one quantity in a specified way and note the changes which occur in a second quantity. The quantity which is intentionally varied is called the <u>independent variable</u> and is plotted on the horizontal axis. The horizontal axis is known as the X-AXIS. The second quantity, which varies as a result of changes in the first quantity, is called the <u>dependent variable</u> and is plotted on the vertical, or Y-AXIS. Any other quantities involved are held constant.

For example, in the circuit shown in figure 3-6, if the resistance was held at 10 ohms and the voltage was varied, the resulting changes in current could then be graphed. The resistance is the constant, the voltage is the independent variable, and the current is the dependent variable.

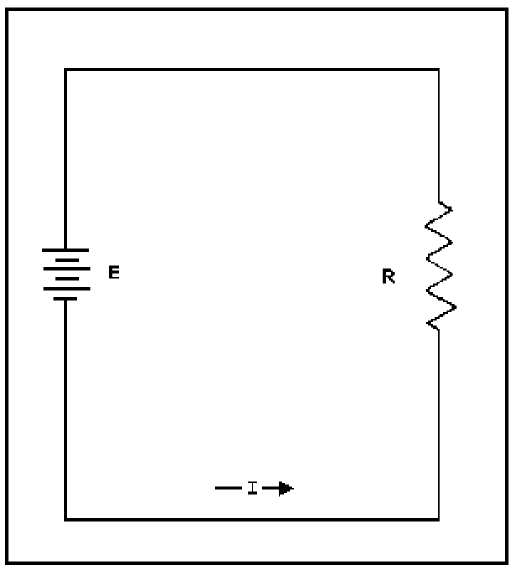
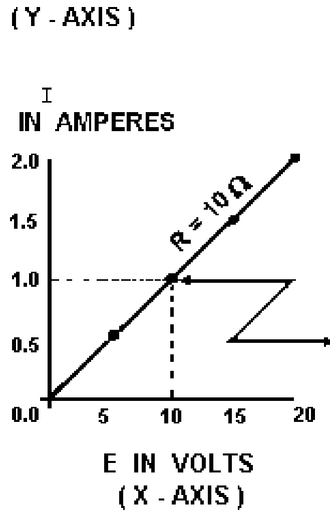


Figure 3-6.—Three variables in a basic circuit.

Figure 3-7 shows the graph and a table of values. This table shows R held constant at 10 ohms as E is varied from 0 to 20 volts in 5-volt steps. Through the use of Ohm's law, you can calculate the value of current for each value of voltage shown in the table. When the table is complete, the information it contains can be used to construct the graph shown in figure 3-7. For example, when the voltage applied to the 10-ohm resistor is 10 volts, the current is 1 ampere. These values of current and voltage determine a point on the graph. When all five points have been plotted, a smooth curve is drawn through the points.



R = 10 Ω	
E	I
0	0.0
5	0.5
10	1.0
15	1.5
20	2.0

Figure 3-7.—Volt-ampere characteristic.

Through the use of this curve, the value of current through the resistor can be quickly determined for any value of voltage between 0 and 20 volts.

Since the curve is a straight line, it shows that equal changes of voltage across the resistor produce equal changes in current through the resistor. This fact illustrates an important characteristic of the basic law—the current varies directly with the applied voltage when the resistance is held constant.

When the voltage across a load is held constant, the current depends solely upon the resistance of the load. For example, figure 3-8 shows a graph with the voltage held constant at 12 volts. The independent variable is the resistance which is varied from 2 ohms to 12 ohms. The current is the dependent variable. Values for current can be calculated as:

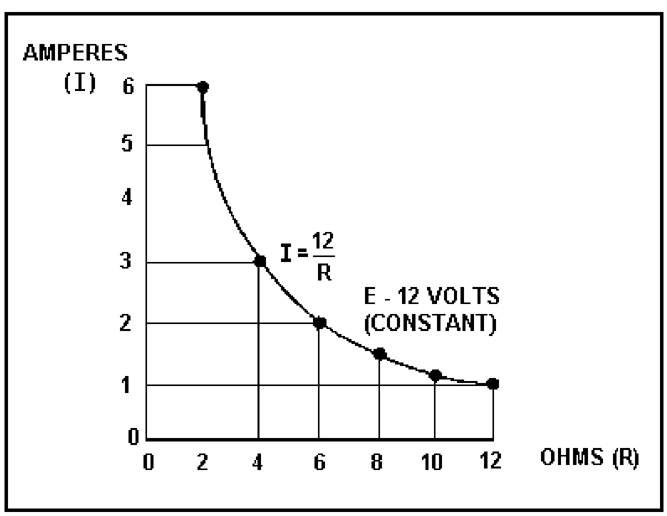


Figure 3-8.—Relationship between current and resistance.

Given: $E = 12 \text{ volts}$

$R = 2 \text{ ohms to } 12 \text{ ohms}$

Soultion: $I = \dfrac{E}{R}$

$$I = \frac{12 \text{ volts}}{12 \text{ ohms}} = 1 \text{ ampere}$$

$$I = \frac{12 \text{ volts}}{10 \text{ ohms}} = 12 \text{ ampere}$$

$$I = \frac{12 \text{ volts}}{8 \text{ ohms}} = 15 \text{ ampere}$$

$$I = \frac{12 \text{ volts}}{6 \text{ ohms}} = 2 \text{ ampere}$$

This process can be continued for any value of resistance. You can see that as the resistance is halved, the current is doubled; when the resistance is doubled, the current is halved.

This illustrates another important characteristic of Ohm's law—current varies inversely with resistance when the applied voltage is held constant.

Q7. Using the graph in figure 3-7, what is the approximate value of current when the voltage is 12.5 volts?

Q8. Using the graph in figure 3-8, what is the approximate value of current when the resistance is 3 ohms?

POWER

Power, whether electrical or mechanical, pertains to the rate at which work is being done. Work is done whenever a force causes motion. When a mechanical force is used to lift or move a weight, work is done. However, force exerted WITHOUT causing motion, such as the force of a compressed spring acting between two fixed objects, does not constitute work.

Previously, it was shown that voltage is an electrical force, and that voltage forces current to flow in a closed circuit. However, when voltage exists but current does not flow because the circuit is open, no work is done. This is similar to the spring under tension that produced no motion. When voltage causes electrons to move, work is done. The instantaneous RATE at which this work is done is called the electric power rate, and is measured in WATTS.

A total amount of work may be done in different lengths of time. For example, a given number of electrons may be moved from one point to another in 1 second or in 1 hour, depending on the RATE at which they are moved. In both cases, total work done is the same. However, when the work is done in a short time, the wattage, or INSTANTANEOUS POWER RATE, is greater than when the same amount of work is done over a longer period of time.

As stated, the basic unit of power is the watt. Power in watts is equal to the voltage across a circuit multiplied by current through the circuit. This represents the rate at any given instant at which work is being done. The symbol P indicates electrical power. Thus, the basic power formula is P = E x I, where E is voltage and I is current in the circuit. The amount of power changes when either voltage or current, or both voltage and current, are caused to change.

In practice, the ONLY factors that can be changed are voltage and resistance. In explaining the different forms that formulas may take, current is sometimes presented as a quantity that is changed. Remember, if current is changed, it is because either voltage or resistance has been changed.

Figure 3-9 shows a basic circuit using a source of power that can be varied from 0 to 8 volts and a graph that indicates the relationship between voltage and power.

The resistance of this circuit is 2 ohms; this value does not change. Voltage (E) is increased (by increasing the voltage source), in steps of 1 volt, from 0 volts to 8 volts. By applying Ohm's law, the current (I) is determined for each step of voltage. For instance, when E is 1 volt, the current is:

$$I = \frac{E}{R}$$

$$I = \frac{1 \text{ volt}}{2 \text{ ohms}}$$

$$I = 0.5 \text{ ampere}$$

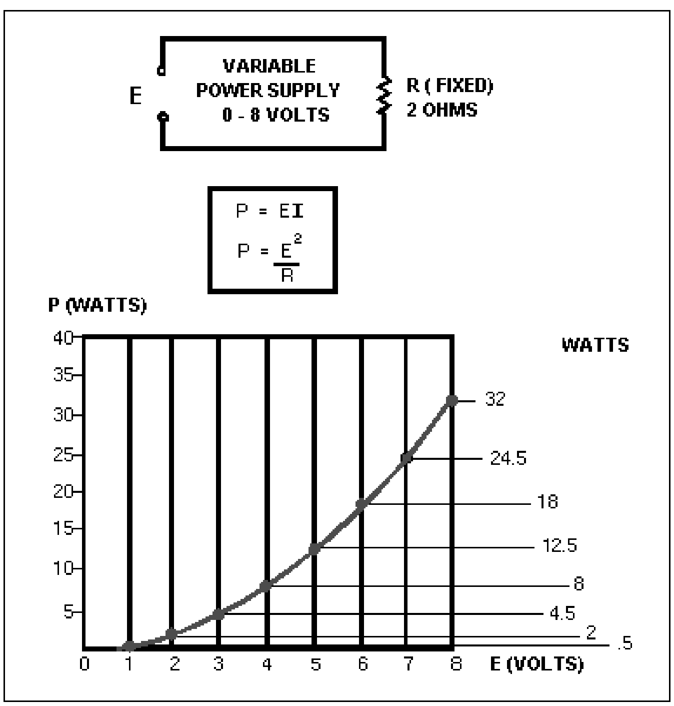


Figure 3-9.—Graph of power related to changing voltage.

Power (P), in watts, is determined by applying the basic power formula:

$$P = E \times I$$
$$P = 1 \text{ volt} \times 0.5 \text{ ampere}$$
$$P = 0.5 \text{ watt}$$

When E is increased to 2 volts:

$$I = \frac{E}{R}$$
$$I = \frac{2 \text{ volts}}{2 \text{ ohms}}$$
$$I = 1 \text{ ampere}$$

and

P = E x I

P = 2 volts x 1 ampere

P = 2 watts

When E is increased to 3 volts:

$$I = \frac{E}{R}$$
$$I = \frac{3 \text{ volts}}{2 \text{ ohms}}$$
$$I = 1.5 \text{ amperes}$$

and

$$P = E \times I$$
$$P = 3 \text{ volts} \times 1.5 \text{ ampere}$$
$$P = 4.5 \text{ watts}$$

You should notice that when the voltage was increased to 2 volts, the power increased from .5 watts to 2 watts or 4 times. When the voltage increased to 3 volts, the power increased to 4.5 watts or 9 times. This shows that if the resistance in a circuit is held constant, the power varies directly with the SQUARE OF THE VOLTAGE.

Another way of proving that power varies as the square of the voltage when resistance is held constant is:

Since: $I = \dfrac{E}{R}$

By substitution in: $P = E \times I$

You get: $P = E \times \dfrac{E}{R}$

Or: $P = \dfrac{E \times E}{R}$

Therefore: $P = \dfrac{E^2}{R}$

Another important relationship may be seen by studying figure 3-10. Thus far, power has been calculated with voltage and current (P = E x I), and with voltage and resistance

$$P = \frac{E^2}{R}$$

Referring to figure 3-10, note that power also varies as the square of current just as it does with voltage. Thus, another formula for power, with current and resistance as its factors, is $P = I^2R$. This can be proved by:

Since: $E = I \times R$

By substituition in: $P = E \times I$

You get: $P = I \times R \times I$

Or: $P = I \times I \times R$

Therefore: $P = I^2 \times R$

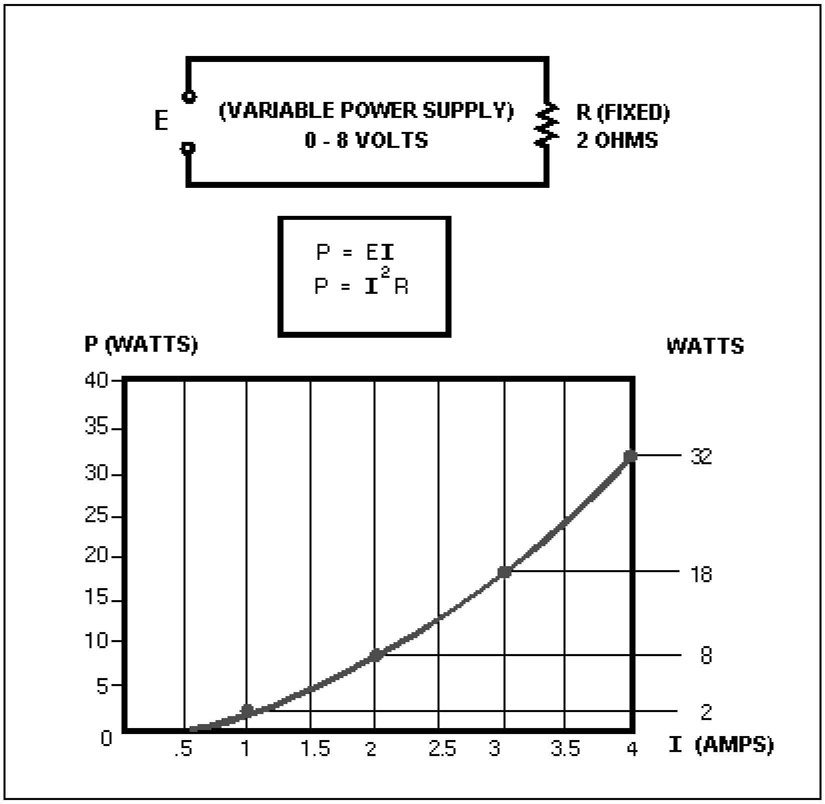


Figure 3-10.—Graph of power related to changing current.

Up to this point, four of the most important electrical quantities have been discussed. These are voltage (E), current (I), resistance (R), and power (P). You must understand the relationships which exist among these quantities because they are used throughout your study of electricity. In the preceding paragraphs, P was expressed in terms of alternate pairs of the other three basic quantities E, I, and R. In practice, you should be able to express any one of these quantities in terms of any two of the others.

Figure 3-11 is a summary of 12 basic formulas you should know. The four quantities E, I, R, and P are at the center of the figure. Adjacent to each quantity are three segments. Note that in each segment, the basic quantity is expressed in terms of two other basic quantities, and no two segments are alike.

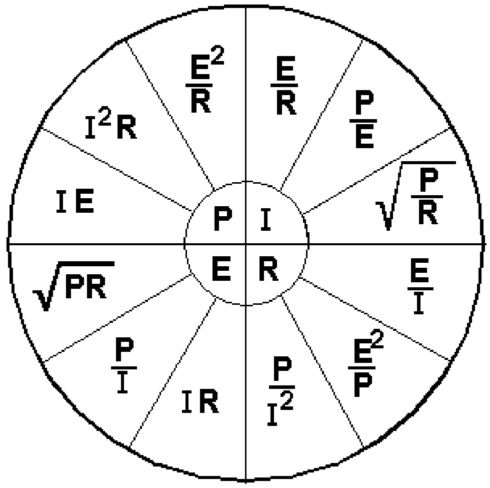


Figure 3-11.—Summary of basic formulas.

For example, the formula wheel in figure 3-11 could be used to find the formula to solve the following problem:

A circuit has a voltage source that delivers 6 volts and the circuit uses 3 watts of power. What is the resistance of the load?

Since R is the quantity you have been asked to find, look in the section of the wheel that has R in the center. The segment

$$\frac{E^2}{P}$$

contains the quantities you have been given. The formula you would use is

$$R = \frac{E^2}{P}$$

The problem can now be solved.

Given: $E = 6$ volts
$P = 3$ watts

Soultion: $R = \dfrac{E^2}{P}$

$$\frac{(6 \text{ volts})^2}{3 \text{ watts}}$$

$$R = \frac{36}{3} = 12 \text{ ohms}$$

Q9. What is the term applied to the rate at which a mechanical or electrical force causes motion?

Q10. How can the amount of current be changed in a circuit?

Q11. What are the three formulas for electrical power?

POWER RATING

Electrical components are often given a power rating. The power rating, in watts, indicates the rate at which the device converts electrical energy into another form of energy, such as light, heat, or motion. An example of such a rating is noted when comparing a 150-watt lamp to a 100-watt lamp. The higher wattage rating of the 150-watt lamp indicates it is capable of converting more electrical energy into light energy than the lamp of the lower rating. Other common examples of devices with power ratings are soldering irons and small electric motors.

In some electrical devices the wattage rating indicates the maximum power the device is designed to use rather than the normal operating power. A 150-watt lamp, for example, uses 150 watts <u>when operated at the specified voltage printed on the bulb</u>. In contrast, a device such as a resistor is not normally given a voltage or a current rating. A resistor is given a power rating in watts and can be operated at <u>any combination of voltage and current</u> as long as the power rating is not exceeded. In most circuits, the actual power used by a resistor is considerably less than the power rating of the resistor because a 50% safety factor is used. For example, if a resistor normally used 2 watts of power, a resistor with a power rating of 3 watts would be used.

Resistors of the same resistance value are available in different wattage values. Carbon resistors, for example, are commonly made in wattage ratings of 1/8, 1/4, 1/2, 1, and 2 watts. The larger the physical size of a carbon resistor the higher the wattage rating. This is true because a larger surface area of material radiates a greater amount of heat more easily.

When resistors with wattage ratings greater than 5 watts are needed, wirewound resistors are used. Wirewound resistors are made in values between 5 and 200 watts. Special types of wirewound resistors are used for power in excess of 200 watts.

As with other electrical quantities, prefixes may be attached to the word watt when expressing very large or very small amounts of power. Some of the more common of these are the kilowatt (1,000 watts), the megawatt (1,000,000 watts), and the milliwatt (1/1,000 of a watt).

Q12. What is the current in a circuit with 5 ohms of resistance that uses 180 watts of power? (refer to figure 3-12)

Q13. What type of resistor should be used in the circuit described in question 12?

Q14. What is the power used in a circuit that has 10 amperes of current through a 10-ohm resistor?

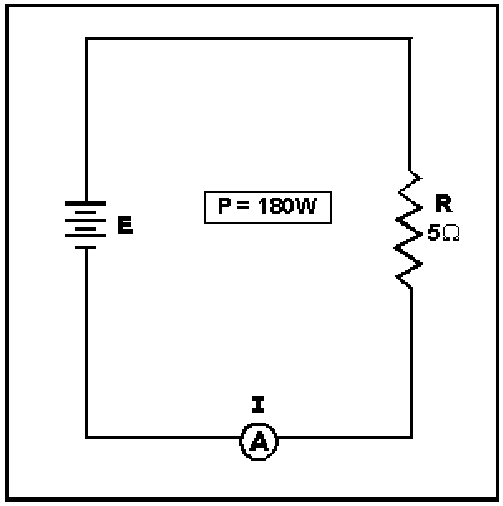


Figure 3-12.—Circuit for computing electrical quantities.

POWER CONVERSION AND EFFICIENCY

The term power consumption is common in the electrical field. It is applied to the use of power in the same sense that gasoline consumption is applied to the use of fuel in an automobile.

Another common term is <u>power conversion</u>. Power is used by electrical devices and is converted from one form of energy to another. An electrical motor converts electrical energy to mechanical energy. An electric light bulb converts electrical energy into light energy and an electric range converts electrical energy into heat energy. Power used by electrical devices is measured in energy. This practical unit of electrical energy is equal to 1 watt of power used continuously for 1 hour. The term kilowatt hour (kWh) is used more extensively on a daily basis and is equal to 1,000 watt-hours.

The EFFICIENCY of an electrical device is the ratio of power converted to useful energy divided by the power consumed by the device. This number will always be less than one (1.00) because of the losses in any electrical device. If a device has an efficiency rating of .95, it effectively transforms 95 watts into useful energy for every 100 watts of input power. The other 5 watts are lost to heat, or other losses which cannot be used.

Calculating the amount of power converted by an electrical device is a simple matter. You need to know the length of time the device is operated and the input power or horsepower rating. Horsepower, a unit of work, is often found as a rating on electrical motors. <u>One horsepower is equal to 746 watts</u>. Example: A 3/4-hp motor operates 8 hours a day. How much power is converted by the motor per month? How many kWh does this represent?

Given: $t = 8$ hrs x 30 days

$P = 3/4$ hp

Solution: Convert horsepower to watts

$P = \text{hp} \times 746$ watts

$P = 3/4 \times 746$ watts

$P = 559$ watts

Convert watts to watt-hours

$$P = \text{work x time}$$

$$P = 559 \text{ watts x } 8 \text{ x } 30$$

$$P = 134{,}000 \text{ watt-hours per month}$$

(NOTE: These figures are rounded to the nearest 1000.)

To convert to kWh

$$P = \frac{\text{Power in watt-hours}}{1000}$$

$$P = \frac{134{,}000 \text{ in watt-hours}}{1000}$$

$$P = 134 \text{ kWh}$$

If the motor actually uses 137 kWh per month, what is the efficiency of the motor?

Given: Power converted = 134 kWh per month

Power used = 137 kWh per month

Solution:

$$EFF = \frac{\text{Power converted}}{\text{Power used}}$$

$$EFF = \frac{134 \text{ kWh per month}}{137 \text{ kWh per month}}$$

$$EFF = .978 \text{ (Rounded to three figures)}$$

Q15. How much power is converted by a 1-horsepower motor in 12 hours?

Q16. What is the efficiency of the motor if it actually uses 9.5 kWh in 12 hours?

SERIES DC CIRCUITS

When two unequal charges are connected by a conductor, a complete pathway for current exists. An electric circuit is a complete conducting pathway. It consists not only of the conductor, but also includes the path through the voltage source. <u>Inside the voltage source</u> current flows from the positive terminal, through the source, emerging at the negative terminal.

SERIES CIRCUIT CHARACTERISTICS

A SERIES CIRCUIT is defined as a circuit that contains only ONE PATH for current flow. To compare the basic circuit that has been discussed and a more complex series circuit, figure 3-13 shows two circuits. The basic circuit has only one lamp and the series circuit has three lamps connected in series.

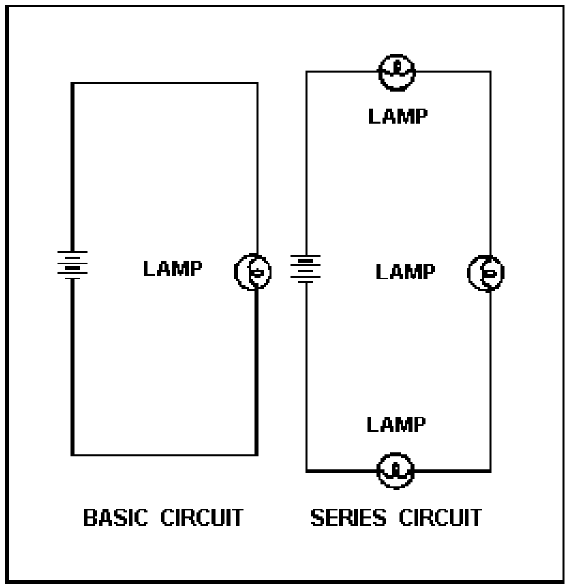


Figure 3-13.—Comparison of basic and series circuits.

Resistance in a Series Circuit

Referring to figure 3-13, the current in a series circuit must flow through each lamp to complete the electrical path in the circuit. Each additional lamp offers added resistance. In a series circuit, THE TOTAL CIRCUIT RESISTANCE (R_T) IS EQUAL TO THE SUM OF THE INDIVIDUAL RESISTANCES.

As an equation: $R_T = R_1 + R_2 + R_3 + \ldots R_n$

NOTE: The subscript n denotes any number of additional resistances that might be in the equation.

Example: In figure 3-14 a series circuit consisting of three resistors: one of 10 ohms, one of 15 ohms, and one of 30 ohms, is shown. A voltage source provides 110 volts. What is the total resistance?

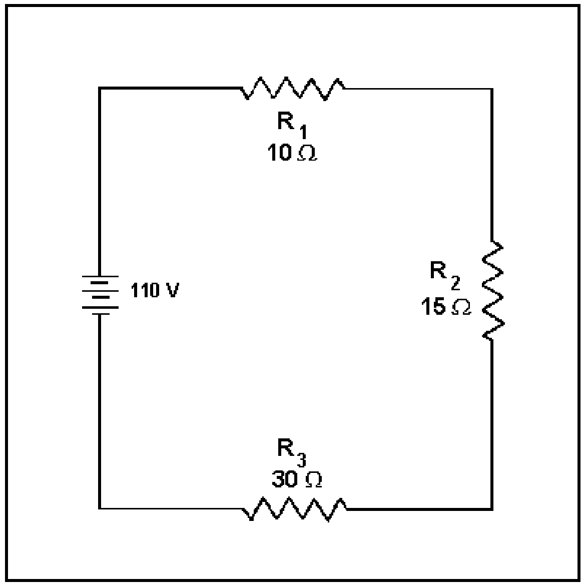


Figure 3-14.—Solving for total resistance in a series circuit.

Given: $R_1 = 10$ ohms
$R_2 = 15$ ohms
$R_3 = 30$ ohms

Soulution: $R_T = R_1 + R_2 + R_3$
$R_T = 10$ ohms $+ 15$ ohms $+ 30$ ohms
$R_T = 55$ ohms

In some circuit applications, the total resistance is known and the value of one of the circuit resistors has to be determined. The equation $R_T = R_1 + R_2 + R_3$ can be transposed to solve for the value of the unknown resistance.

Example: In figure 3-15 the total resistance of a circuit containing three resistors is 40 ohms. Two of the circuit resistors are 10 ohms each. Calculate the value of the third resistor (R_3).

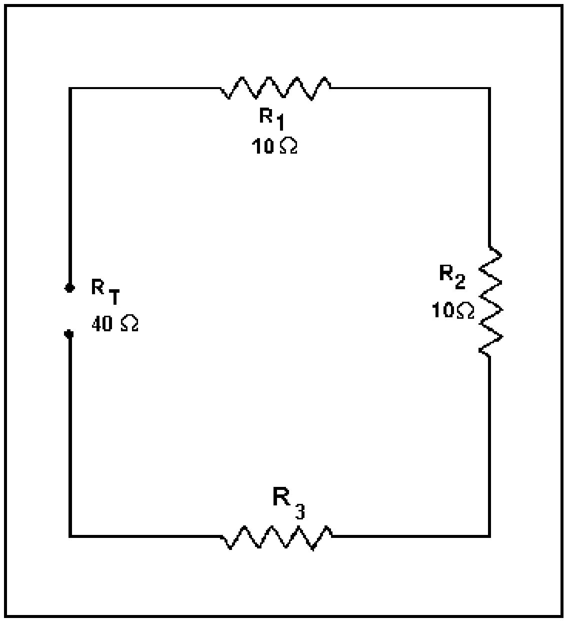


Figure 3-15.—Calculating the value of one resistance in a series circuit.

Given:

$R_1 = 40$ ohms
$R_2 = 10$ ohms
$R_3 = 10$ ohms

Solution:

$$R_T = R_1 + R_2 + R_3$$

(Subtract $R_1 + R_2$ from both sides of the equation.)

$$R_T - R_1 - R_2 = R_3$$
$$R_3 = R_T - R_1 - R_2$$
$$R_3 = 40 \text{ ohms} - 10 \text{ ohms} - 10 \text{ ohms}$$
$$R_3 = 40 \text{ ohms} - 20 \text{ ohms}$$
$$R_3 = 20 \text{ ohms}$$

Current in a Series Circuit

Since there is only one path for current in a series circuit, the same current must flow through each component of the circuit. To determine the current in a series circuit, only the current through one of the components need be known.

The fact that the same current flows through each component of a series circuit can be verified by inserting meters into the circuit at various points, as shown in figure 3-16. If this were done, each meter would be found to indicate the same value of current.

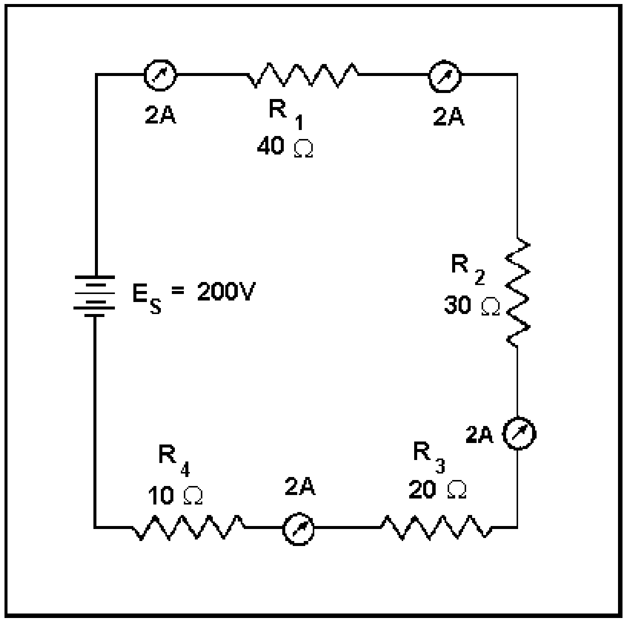


Figure 3-16.—Current in a series circuit.

Voltage in a Series Circuit

The voltage dropped across the resistor in a circuit consisting of a single resistor and a voltage source is the total voltage across the circuit and is equal to the applied voltage. The total voltage across a series circuit that consists of more than one resistor is also equal to the applied voltage, but consists of the sum of the individual resistor voltage drops. In any series circuit, the SUM of the resistor voltage drops must equal the source voltage. This statement can be proven by an examination of the circuit shown in figure 3-17. In this circuit a source potential (E_T) of 20 volts is dropped across a series circuit consisting of two 5-ohm resistors. The total resistance of the circuit (R_T) is equal to the sum of the two individual resistances, or 10 ohms. Using Ohm's law the circuit current may be calculated as follows:

Given: $E_T = 20 \text{ volts}$
$R_T = 10 \text{ ohms}$

Solution: $$I_T = \frac{E_T}{R_T}$$

$$I_T = \frac{20 \text{ volts}}{10 \text{ ohms}}$$

$$I_T = 2 \text{ amps}$$

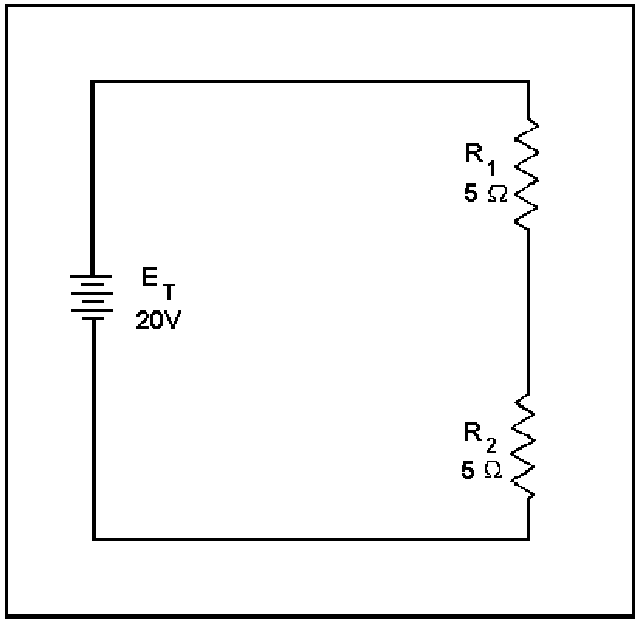


Figure 3-17.—Calculating individual voltage drops in a series circuit.

Since the value of the resistors is known to be 5 ohms each, and the current through the resistors is known to be 2 amperes, the voltage drops across the resistors can be calculated. The voltage (E_1) across R_1 is therefore:

Given: $I_1 = 2$ amperes
$R_1 = 5$ ohms

Solution: $E_1 = I_1 \times R_1$
$E_1 = 2 \text{ amperes} \times 5 \text{ ohms}$
$E_1 = 10 \text{ volts}$

By inspecting the circuit, you can see that R_2 is the same ohmic value as R_1 and carries the same current. The voltage drop across R_2 is therefore also equal to 10 volts. Adding these two 10-volts drops together gives a total drop of 20 volts, exactly equal to the applied voltage. For a series circuit then:

$$E_T = E_1 + E_2 + E_3 = \ldots E_n$$

Example: A series circuit consists of three resistors having values of 20 ohms, 30 ohms, and 50 ohms, respectively. Find the applied voltage if the current through the 30 ohm resistor is 2 amps. (The abbreviation amp is commonly used for ampere.)

To solve the problem, a circuit diagram is first drawn and labeled (fig 3-18).

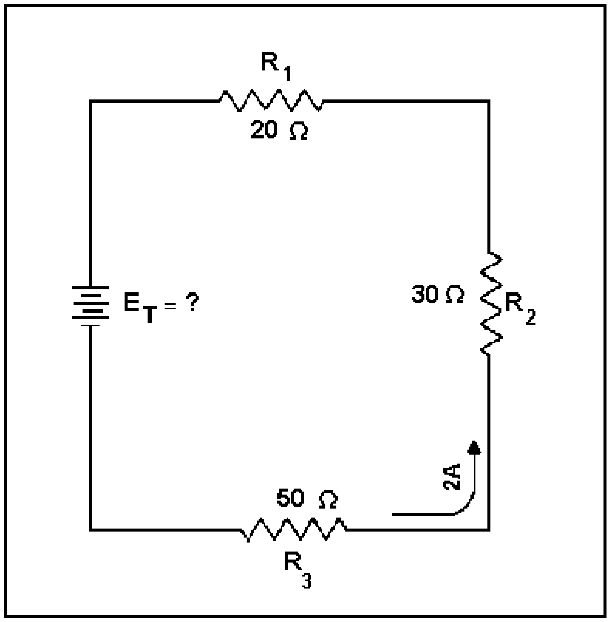


Figure 3-18.—Solving for applied voltage in a series circuit.

Given:

$R_1 = 20$ ohms
$R_2 = 30$ ohms
$R_3 = 50$ ohms
$I = 2$ amps

Solution:

$E_T = E_1 + E_2 + E_3$
$E_1 = R_1 \times I_1$ (I_1 = The current through resistor R_1)
$E_2 = R_2 \times I_2$
$E_3 = R_3 \times I_3$

Substituting:

$$E_T = (R_1 \times I_1) + (R_2 \times I_2) + (R_3 \times I_3)$$

$$E_T = (20 \text{ ohms} \times 2 \text{ amps}) + (30 \text{ ohms} \times 2 \text{ amps}) + (50 \text{ ohms} \times 2 \text{ amps})$$

$$E_T = 40 \text{ volts} + 60 \text{ volts} + 100 \text{ volts}$$

$$E_T = 200 \text{ volts}$$

NOTE: When you use Ohm's law, the quantities for the equation MUST be taken from the SAME part of the circuit. In the above example the voltage across R_2 was computed using the current through R_2 and the resistance of R_2.

The value of the voltage dropped by a resistor is determined by the applied voltage and is in proportion to the circuit resistances. The voltage drops that occur in a series circuit are in direct proportion to the resistances. This is the result of having the same current flow through each resistor—the larger the ohmic value of the resistor, the larger the voltage drop across it.

Q17. A series circuit consisting of three resistors has a current of 3 amps. If R_1 = 20 ohms, R_2= 60 ohms, and R_3 = 80 ohms, what is the (a) total resistance and (b) source voltage of the circuit?

Q18. What is the voltage dropped by each resistor of the circuit described in question 17?

Q19. If the current was increased to 4 amps, what would be the voltage drop across each resistor in the circuit described in question 17?

Q20. What would have to be done to the circuit described in question 17 to increase the current to 4 amps?

Power in a Series Circuit

Each of the resistors in a series circuit consumes power which is dissipated in the form of heat. Since this power must come from the source, the total power must be equal to the power consumed by the circuit resistances. In a series circuit the total power is equal to the SUM of the power dissipated by the individual resistors. Total power (P_T) is equal to:

$$P_T = P_1 + P_2 + P_3 \ldots P_n$$

Example: A series circuit consists of three resistors having values of 5 ohms, 10 ohms, and 15 ohms. Find the total power when 120 volts is applied to the circuit. (See fig. 3-19.)

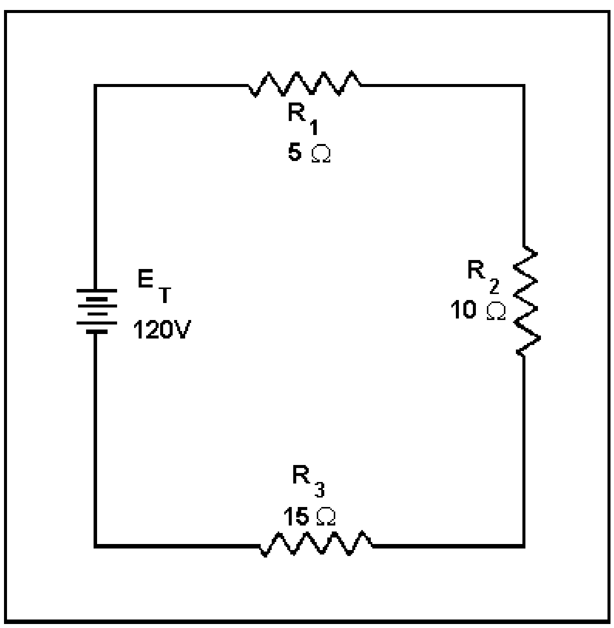


Figure 3-19.—Solving for total power in a series circuit.

Given:

$$R_1 = 5 \text{ ohms}$$
$$R_2 = 10 \text{ ohms}$$
$$R_3 = 15 \text{ ohms}$$
$$E = 120 \text{ volts}$$

Solution: The total resistance is found first.

$$R_T = R_1 + R_2 + R_3$$
$$R_T = 5 \text{ ohms} + 10 \text{ ohms} + 15 \text{ ohms}$$
$$R_T = 30 \text{ ohms}$$

By using the total resistance and the applied voltage, the circuit current is calculated.

$$I = \frac{E_T}{R_T}$$

$$I = \frac{120 \text{ volts}}{30 \text{ ohms}}$$

$$I = 4 \text{ amps}$$

By means of the power formulas, the power can be calculated for each resistor:

For R_1: $P_1 = I^2 \times R_1$
$$P_1 = (4 \text{ amps})^2 \times 5 \text{ ohms}$$
$$P_1 = 80 \text{ watts}$$

For R_2: $P_2 = I^2 \times R_2$
$$P_2 = (4 \text{ amps})^2 \times 10 \text{ ohms}$$
$$P_2 = 160 \text{ watts}$$

For R_3: $P_3 = I^2 \times R_3$
$$P_3 = (4 \text{ amps})^2 \times 15 \text{ ohms}$$
$$P_3 = 240 \text{ watts}$$

For total power:
$$P_T = P_1 + P_2 + P_3$$
$$P_T = 80 \text{ watts} + 160 \text{ watts} + 240 \text{ watts}$$
$$P_T = 480 \text{ watts}$$

To check the answer, the total power delivered by the source can be calculated:

$$P_{source} = I_{source} \times E_{source}$$
$$P_{source} = 4 \text{ amps} \times 120 \text{ volts}$$
$$P_{source} = 480 \text{ watts}$$

The total power is equal to the sum of the power used by the individual resistors.

SUMMARY OF CHARACTERISTICS

The important factors governing the operation of a series circuit are listed below. These factors have been set up as a group of rules so that they may be easily studied. These rules must be completely understood before the study of more advanced circuit theory is undertaken.

Rules for Series DC Circuits

1. The same current flows through each part of a series circuit.
2. The total resistance of a series circuit is equal to the sum of the individual resistances.
3. The total voltage across a series circuit is equal to the sum of the individual voltage drops.
4. The voltage drop across a resistor in a series circuit is proportional to the ohmic value of the resistor.
5. The total power in a series circuit is equal to the sum of the individual powers used by each circuit component.

SERIES CIRCUIT ANALYSIS

To establish a procedure for solving series circuits, the following sample problems will be solved.

Example: Three resistors of 5 ohms, 10 ohms, and 15 ohms are connected in series with a power source of 90 volts as shown in figure 3-20. Find the total resistance, circuit current, voltage drop of each resistor, power of each resistor, and total power of the circuit.

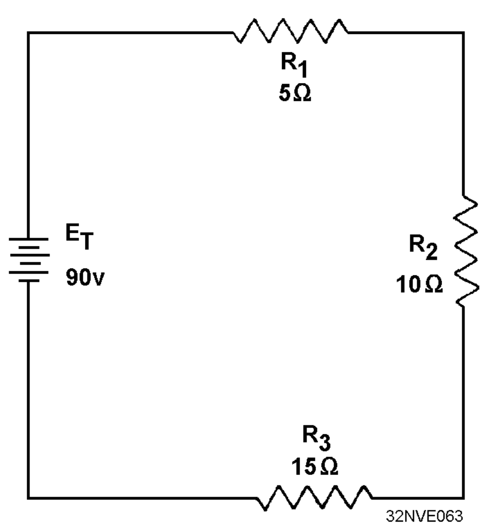


Figure 3-20.—Solving for various values in a series circuit.

In solving the circuit the total resistance will be found first. Next, the circuit current will be calculated. Once the current is known, the voltage drops and power dissipations can be calculated.

Given:

$R_1 = 5$ ohms
$R_2 = 10$ ohms
$R_3 = 15$ ohms
$E = 90$ volts

Solution:

$R_T = R_1 + R_2 + R_3$
$R_T = 5 \text{ ohms} + 10 \text{ ohms} + 15 \text{ ohms}$
$R_T = 30$ ohms

$$I = \frac{E_T}{R_T}$$

$$I = \frac{90 \text{ volts}}{30 \text{ ohms}}$$

$I = 3$ amps

$E_1 = IR_1$
$E_1 = 3 \text{ amperes} \times 5 \text{ ohms}$
$E_1 = 15$ volts

$E_2 = IR_2$
$E_2 = 3 \text{ amperes} \times 10 \text{ ohms}$
$E_2 = 30$ volts

$E_3 = IR_3$
$E_3 = 3 \text{ amperes} \times 15 \text{ ohms}$
$E_3 = 45$ volts

$P_1 = I \times E_1$
$P_1 = 3 \text{ amperes} \times 15 \text{ volts}$
$P_1 = 45$ watts

$P_2 = I \times E_2$
$P_2 = 3 \text{ amperes} \times 30 \text{ volts}$
$P_2 = 90$ watts

$$P_3 = I \times E_3$$
$$P_3 = 3 \text{ amperes} \times 45 \text{ volts}$$
$$P_3 = 135 \text{ watts}$$

$$P_T = E_1 \times I$$
$$P_T = 90 \text{ volts} \times 3 \text{ amps}$$
$$P_T = 270 \text{ watts}$$

or

$$P_T = P_1 + P_2 + P_3$$
$$P_T = 45 \text{ watts} + 90 \text{ watts} + 135 \text{ watts}$$
$$P_T = 270 \text{ watts}$$

Example: Four resistors, $R_1 = 10$ ohms, $R_2 = 10$ ohms, $R_3 = 50$ ohms, and $R_4 = 30$ ohms, are connected in series with a power source as shown in figure 3-21. The current through the circuit is 1/2 ampere.

a. What is the battery voltage?

b. What is the voltage across each resistor?

c. What is the power expended in each resistor?

d. What is the total power?

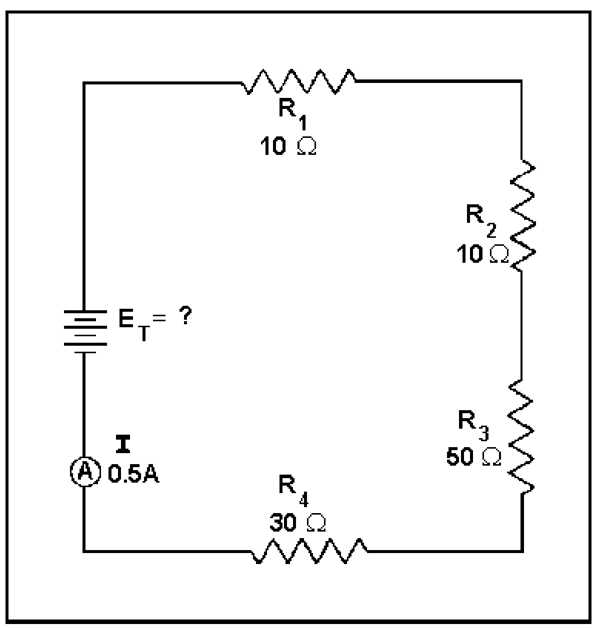


Figure 3-21.—Computing series circuit values.

Given:

$$R_1 = 10 \text{ ohms}$$
$$R_2 = 10 \text{ ohms}$$
$$R_3 = 50 \text{ ohms}$$
$$R_4 = 30 \text{ ohms}$$
$$I = 0.5 \text{ amps}$$

Solution (a):

$$E_T = IR_T$$
$$R_T = R_1 + R_2 + R_3 + R_4$$
$$R_T = 10 \text{ ohms} + 10 \text{ ohms} + 50 \text{ ohms} + 30 \text{ ohms}$$
$$R_T = 100 \text{ ohms}$$
$$E_T = 0.5 \text{ amps} \times 100 \text{ ohms}$$
$$E_T = 50 \text{ volts}$$

Solution (b):

$$E_1 = IR_1$$
$$E_1 = 0.5 \text{ amperes} \times 10 \text{ ohms}$$
$$E_1 = 5 \text{ volts}$$

$$E_2 = IR_2$$
$$E_2 = 0.5 \text{ amperes} \times 10 \text{ ohms}$$
$$E_2 = 5 \text{ volts}$$

$$E_3 = IR_3$$
$$E_3 = 0.5 \text{ amperes} \times 50 \text{ ohms}$$
$$E_3 = 25 \text{ volts}$$

$$E_4 = IR_4$$
$$E_4 = 0.5 \text{ amperes} \times 30 \text{ ohms}$$
$$E_4 = 15 \text{ volts}$$

Solution (c):

$$P_1 = IE_1$$
$$P_1 = 0.5 \text{ amperes} \times 5 \text{ volts}$$
$$P_1 = 2.5 \text{ watts}$$

$$P_2 = IE_2$$
$$P_2 = 0.5 \text{ amperes} \times 5 \text{ volts}$$
$$P_2 = 2.5 \text{ watts}$$

$$P_3 = IE_3$$
$$P_3 = 0.5 \text{ amperes} \times 25 \text{ volts}$$
$$P_3 = 12.5 \text{ watts}$$

$$P_4 = IE_4$$
$$P_4 = 0.5 \text{ amperes} \times 15 \text{ volts}$$
$$P_4 = 7.5 \text{ watts}$$

Solution (d):

$$P_T = P_1 + P_2 + P_3 + P_4$$
$$P_T = 2.5 \text{ watts} + 2.5 \text{ watts} + 12.5 \text{ watts} + 7.5 \text{ watts}$$
$$P_T = 25 \text{ watts}$$

or

$$P_T = IE_T$$
$$P_T = 0.5 \text{ amperes} \times 50 \text{ volts}$$
$$P_T = 25 \text{ watts}$$

or

$$P_T = \frac{E_T^2}{R_T}$$

$$P_T = \frac{(50 \text{ volts})^2}{100 \text{ ohms}}$$

$$P_T = \frac{2500 \text{ volts}}{100 \text{ ohms}}$$

$$P_T = 25 \text{ watts}$$

An important fact to keep in mind when applying Ohm's law to a series circuit is to consider whether the values used are component values or total values. When the information available enables the use of Ohm's law to find total resistance, total voltage, and total current, total values must be inserted into the formula. To find total resistance:

$$R_T = \frac{E_T}{I_T}$$

To find total voltage:

$$E_T = I_T \times R_T$$

To find total current:

$$I_T = \frac{E_T}{R_T}$$

NOTE: I_T is equal to I in a series circuit. However, the distinction between I_T and I in the formula should be noted. The reason for this is that future circuits may have several currents, and it will be necessary to differentiate between I_T and other currents.

To compute any quantity (E, I, R, or P) associated with a single given resistor, the values used in the formula must be obtained from that particular resistor. For example, to find the value of an unknown resistance, the voltage across and the current through that particular resistor must be used.

To find the value of a resistor:

$$R = \frac{E_R}{I_R}$$

To find the voltage drop across a resistor:

$$E_R = I_R \times R$$

To find current through a resistor:

$$I_R = \frac{E_R}{R}$$

Q21. A series circuit consists of two resistors in series. R_1 = 25 ohms and R_2 = 30 ohms. The circuit current is 6 amps. What is the (a) source voltage, (b) voltage dropped by each resistor, (c) total power, and (d) power used by each resistor?

KIRCHHOFF'S VOLTAGE LAW

In 1847, G. R. Kirchhoff extended the use of Ohm's law by developing a simple concept concerning the voltages contained in a series circuit loop. Kirchhoff's voltage law states:

"The algebraic sum of the voltage drops in any closed path in a circuit and the electromotive forces in that path is equal to zero."

To state Kirchhoff's law another way, the voltage drops and voltage sources in a circuit are equal at any given moment in time. If the voltage sources are assumed to have one sign (positive or negative) at that instant and the voltage drops are assumed to have the opposite sign, the result of adding the voltage sources and voltage drops will be zero.

NOTE: The terms electromotive force and emf are used when explaining Kirchhoff's law because Kirchhoff's law is used in alternating current circuits (covered in Module 2). In applying Kirchhoff's law to direct current circuits, the terms electromotive force and emf apply to voltage sources such as batteries or power supplies.

Through the use of Kirchhoff's law, circuit problems can be solved which would be difficult, and often impossible, with knowledge of Ohm's law alone. When Kirchhoff's law is properly applied, an equation can be set up for a closed loop and the unknown circuit values can be calculated.

POLARITY OF VOLTAGE

To apply Kirchhoff's voltage law, the meaning of voltage polarity must be understood.

In the circuit shown in figure 3-22, the current is shown flowing in a counterclockwise direction. Notice that the end of resistor R_1, into which the current flows, is marked NEGATIVE ( ). The end of R_1 at which the current leaves is marked POSITIVE (+). These polarity markings are used to show that the end of R_1 into which the current flows is at a higher negative potential than the end of the resistor at which the current leaves. Point A is more negative than point B.

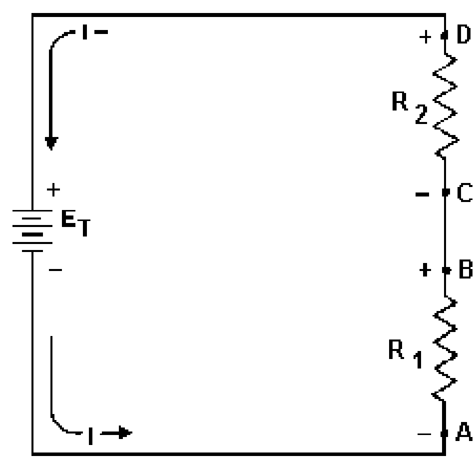


Figure 3-22.—Voltage polarities.

Point C, which is at the same potential as point B, is labeled negative. This is to indicate that point C is more negative than point D. To say a point is positive (or negative) without stating what the polarity is based upon has no meaning. In working with Kirchhoff's law, positive and negative polarities are assigned in the direction of current flow.

APPLICATION OF KIRCHHOFF'S VOLTAGE LAW

Kirchhoff's voltage law can be written as an equation, as shown below:

$$E_a + E_b + E_c + \ldots E_n = 0$$

where E_a, E_b, etc., are the voltage drops or emf's around any closed circuit loop. To set up the equation for an actual circuit, the following procedure is used.

1. Assume a direction of current through the circuit. (The correct direction is desirable but not necessary.)

2. Using the assumed direction of current, assign polarities to all resistors through which the current flows.

3. Place the correct polarities on any sources included in the circuit.

4. Starting at any point in the circuit, trace around the circuit, writing down the amount and polarity of the voltage across each component in succession. The polarity used is the sign AFTER the assumed current has passed through the component. Stop when the point at which the trace was started is reached.

5. Place these voltages, with their polarities, into the equation and solve for the desired quantity.

Example: Three resistors are connected across a 50-volt source. What is the voltage across the third resistor if the voltage drops across the first two resistors are 25 volts and 15 volts?

Solution: First, a diagram, such as the one shown in figure 3-23, is drawn. Next, a direction of current is assumed (as shown). Using this current, the polarity markings are placed at each end of each resistor and also on the terminals of the source. Starting at point A, trace around the circuit in the direction of current flow, recording the voltage and polarity of each component. Starting at point A and using the components from the circuit:

$$(+E_x) + (+E_2) + (+E_1) + (-E_A) = 0$$

Substituting values from the circuit:

$$E_x + 15 \text{ volts} + 25 \text{ volts} - 50 \text{ volts} = 0$$
$$E_x - 10 \text{ volts} = 0$$
$$E_x = 10 \text{ volts}$$

The unknown voltage (E_x) is found to be 10 volts

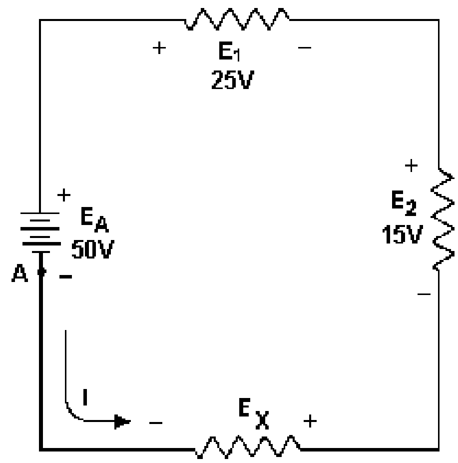


Figure 3-23.—Determining unknown voltage in a series circuit.

Using the same idea as above, you can solve a problem in which the current is the unknown quantity.

Example: A circuit having a source voltage of 60 volts contains three resistors of 5 ohms, 10 ohms, and 15 ohms. Find the circuit current.

Solution: Draw and label the circuit (fig. 3-24). Establish a direction of current flow and assign polarities. Next, starting at any point—point A will be used in this example—write out the loop equation.

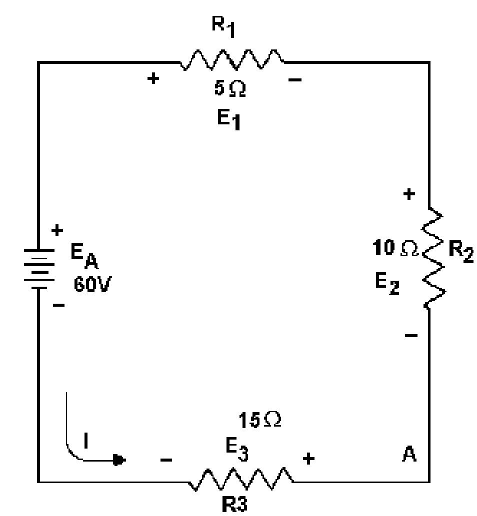


Figure 3-24.—Correct direction of assumed current.

Basic equation:

$$E_2 + E_1 + E_A + E_3 = 0$$

Since E=IR, by substitution:

$$(I \times R_2) + (I \times R_1) + E_A + (I \times R_3) = 0$$

Substituting Values:

$$(I \times 10 \text{ ohms}) + (I \times 5 \text{ ohms}) + (-60 \text{ volts}) + (I \times 15 \text{ ohms}) = 0$$

Combining like terms:

$$(I \times 30 \text{ ohms}) + (-60 \text{ volts}) = 0$$

$$(I \times 30 \text{ ohms}) = 60 \text{ volts}$$

$$I = \frac{60 \text{ volts}}{30 \text{ ohms}}$$

$$I = 2 \text{ amps}$$

Since the current obtained in the above calculations is a positive 2 amps, the assumed direction of current was correct. To show what happens if the incorrect direction of current is assumed, the problem will be solved as before, but with the opposite direction of current. The circuit is redrawn showing the new direction of current and new polarities in figure 3-25. Starting at point A the loop equation is:

$$E_3 + E_A + E_1 + E_2 = 0$$

$$(I \times R_3) + E_A + (I \times R_1) + (I \times R_2) = 0$$

Substituting Values:

$$(I \times 15 \text{ ohms}) + 60 \text{ volts} + (I \times 5 \text{ ohms}) + (I \times 10 \text{ ohms}) = 0$$

Combining like terms:

$$(I \times 30 \text{ ohms}) + 60 \text{ volts} = 0$$

$$I \times 30 \text{ ohms} = -60 \text{ volts}$$

$$I = \frac{-60 \text{ volts}}{30 \text{ ohms}}$$

$$I = -2 \text{ amps}$$

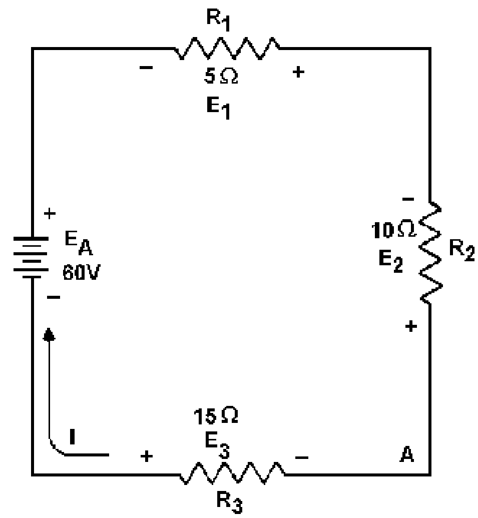


Figure 3-25.—Incorrect direction of assumed current.

Notice that the AMOUNT of current is the same as before. The polarity, however, is NEGATIVE. The negative polarity simply indicates the wrong direction of current was assumed. Should it be necessary to use this current in further calculations on the circuit using Kirchhoff's law, the negative polarity should be retained in the calculations.

Series Aiding and Opposing Sources

In many practical applications a circuit may contain more than one source of emf. Sources of emf that cause current to flow in the same direction are considered to be SERIES AIDING and the voltages are added. Sources of emf that would tend to force current in opposite directions are said to be SERIES OPPOSING, and the effective source voltage is the difference between the opposing voltages. When two opposing sources are inserted into a circuit current flow would be in a direction determined by the larger source. Examples of series aiding and opposing sources are shown in figure 3-26.

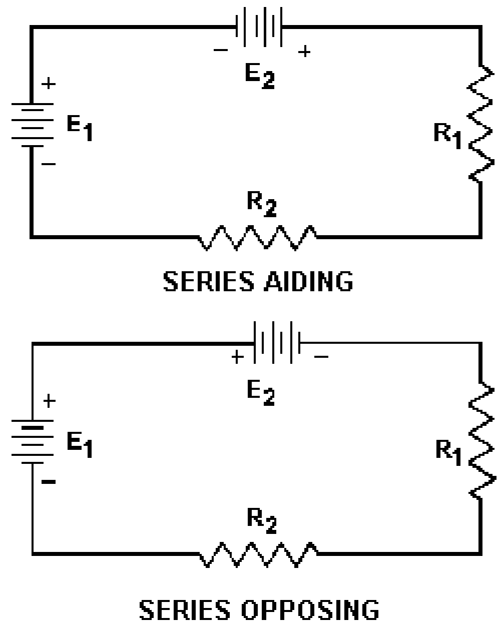


Figure 3-26.—Aiding and opposing sources.

A simple solution may be obtained for a multiple-source circuit through the use of Kirchhoff's voltage law. In applying this method, the same procedure is used for the multiple-source circuit as was used above for the single-source circuit. This is demonstrated by the following example.

Example: Using Kirchhoff's voltage equation, find the amount of current in the circuit shown in fig 3-27.

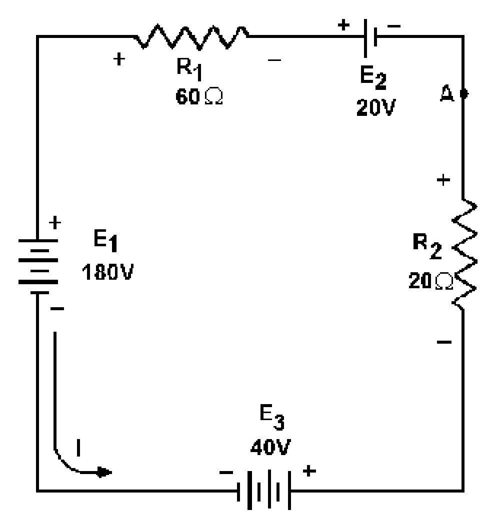


Figure 3-27.—Solving for circuit current using Kirchhoff's voltage equation.

Solution: As before, a direction of current flow is assumed and polarity signs are placed on the drawing. The loop equation will be started at point A.

$$E_2 + E_{R1} + E_1 + E_3 + E_{R2} = 0$$

$$20\text{ volts} + (I \times 60\text{ ohms}) + (-180\text{ volts}) + 40\text{ volts} + (I \times 20\text{ ohms}) = 0$$

$$20\text{ volts} - 180\text{ volts} + 40\text{ volts} + (I \times 60\text{ ohms}) + (I \times 20\text{ ohms}) = 0$$

$$-120\text{ volts} + (I \times 80\text{ ohms}) = 0$$

$$I \times 80\text{ ohms} = 120\text{ volts}$$

$$I = \frac{120\text{ volts}}{80\text{ ohms}}$$

$$I = 1.5\text{ amps}$$

Q22. When using Kirchhoff's voltage law, how are voltage polarities assigned to the voltage drops across resistors?

Q23. Refer to figure 3-27, if R_1 was changed to a 40-ohm resistor, what would be the value of circuit current (I_T)?

Q24. Refer to figure 3-27. What is the effective source voltage of the circuit using the 40-ohm resistor?

CIRCUIT TERMS AND CHARACTERISTICS

Before you learn about the types of circuits other than the series circuit, you should become familiar with some of the terms and characteristics used in electrical circuits. These terms and characteristics will be used throughout your study of electricity and electronics.

REFERENCE POINT

A reference point is an arbitrarily chosen point to which all other points in the circuit are compared. In series circuits, any point can be chosen as a reference and the electrical potential at all other points can be determined in reference to that point. In figure 3-28 point A shall be considered the reference point. Each series resistor in the illustrated circuit is of equal value. The applied voltage is equally distributed across each resistor. The potential at point B is 25 volts more positive than at point A. Points C and D are 50 volts and 75 volts more positive than point A respectively.

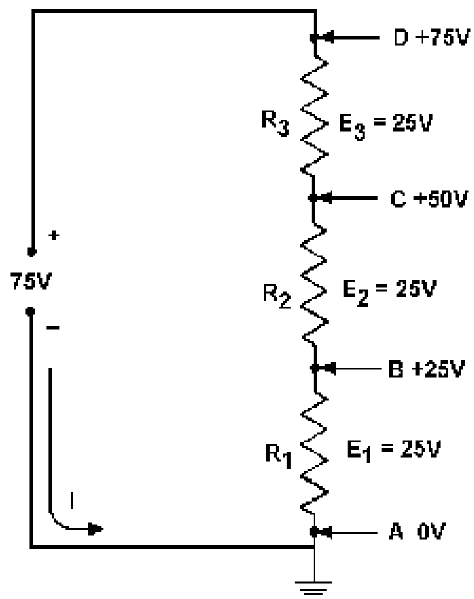


Figure 3-28.—Reference points in a series circuit.

When point B is used as the reference, as in figure 3-29, point D would be positive 50 volts in respect to the new reference point. The former reference point, A, is 25 volts negative in respect to point B.

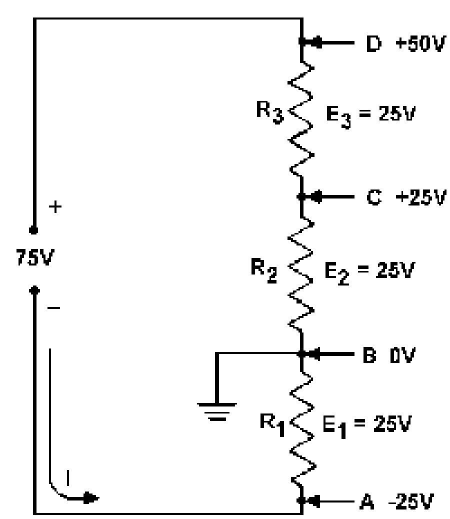


Figure 3-29.—Determining potentials with respect to a reference point.

As in the previous circuit illustration, the reference point of a circuit is always considered to be at zero potential. Since the earth (ground) is said to be at a zero potential, the term GROUND is used to denote a common electrical point of zero potential. In figure 3-30, point A is the zero reference, or ground, and the symbol for ground is shown connected to point A. Point C is 75 volts positive in respect to ground.

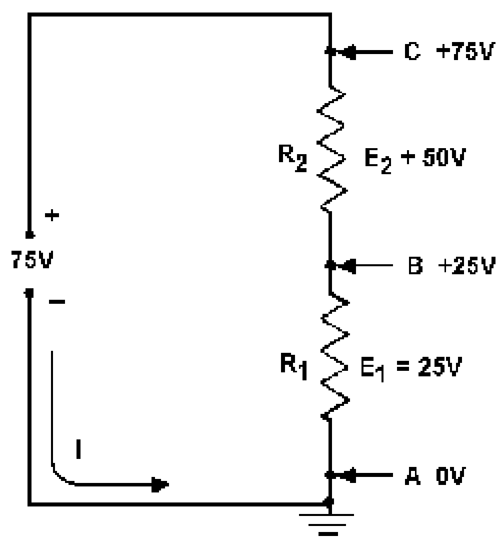


Figure 3-30.—Use of ground symbols.

In most electrical equipment, the metal chassis is the common ground for the many electrical circuits. When each electrical circuit is completed, common points of a circuit at zero potential are connected directly to the metal chassis, thereby eliminating a large amount of connecting wire. The electrons pass through the metal chassis (a conductor) to reach other points of the circuit. An example of a chassis grounded circuit is illustrated in figure 3-31.

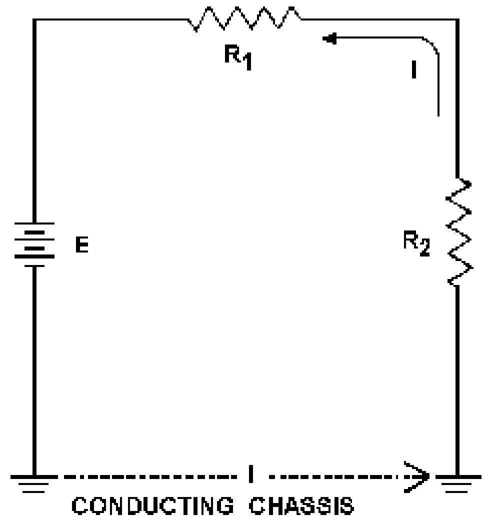


Figure 3-31.—Ground used as a conductor.

Most voltage measurements used to check proper circuit operation in electrical equipment are taken in respect to ground. One meter lead is attached to a grounded point and the other meter lead is moved to various test points. Circuit measurement is explained in more detail in NEETS Module 3.

OPEN CIRCUIT

A circuit is said to be OPEN when a break exists in a complete conducting pathway. Although an open occurs when a switch is used to deenergize a circuit, an open may also develop accidentally. To restore a circuit to proper operation, the open must be located, its cause determined, and repairs made.

Sometimes an open can be located visually by a close inspection of the circuit components. Defective components, such as burned out resistors, can usually be discovered by this method. Others, such as a break in wire covered by insulation or the melted element of an enclosed fuse, are not visible to the eye. Under such conditions, the understanding of the effect an open has on circuit conditions enables a technician to make use of test equipment to locate the open component.

In figure 3-32, the series circuit consists of two resistors and a fuse. Notice the effects on circuit conditions when the fuse opens.

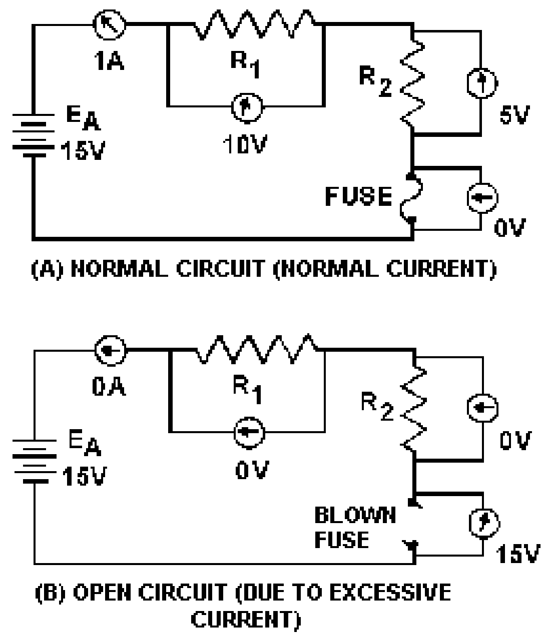


Figure 3-32.—Normal and open circuit conditions. (A) Normal current; (B) Excessive current.

Current ceases to flow; therefore, there is no longer a voltage drop across the resistors. Each end of the open conducting path becomes an extension of the battery terminals and the voltage felt across the open is equal to the applied voltage (E_A).

An open circuit has INFINITE resistance. INFINITY represents a quantity so large it cannot be measured. The symbol for infinity is ∞. In an open circuit, $R_T = \infty$.

SHORT CIRCUIT

A short circuit is an accidental path of low resistance which passes an abnormally high amount of current. A short circuit exists whenever the resistance of a circuit or the resistance of a part of a circuit drops in value to almost zero ohms. A short often occurs as a result of improper wiring or broken insulation.

In figure 3-33, a short is caused by improper wiring. Note the effect on current flow. Since the resistor has in effect been replaced with a piece of wire, practically all the current flows through the short and very little current flows through the resistor. Electrons flow through the short (a path of almost zero resistance) and the remainder of the circuit by passing through the 10-ohm resistor and the battery. The amount of current flow increases greatly because its resistive path has decreased from 10,010 ohms to 10 ohms. Due to the excessive current flow the 10-ohm resistor becomes heated. As it attempts to dissipate this heat, the resistor will probably be destroyed. Figure 3-34 shows a pictorial wiring diagram, rather than a schematic diagram, to indicate how broken insulation might cause a short circuit.

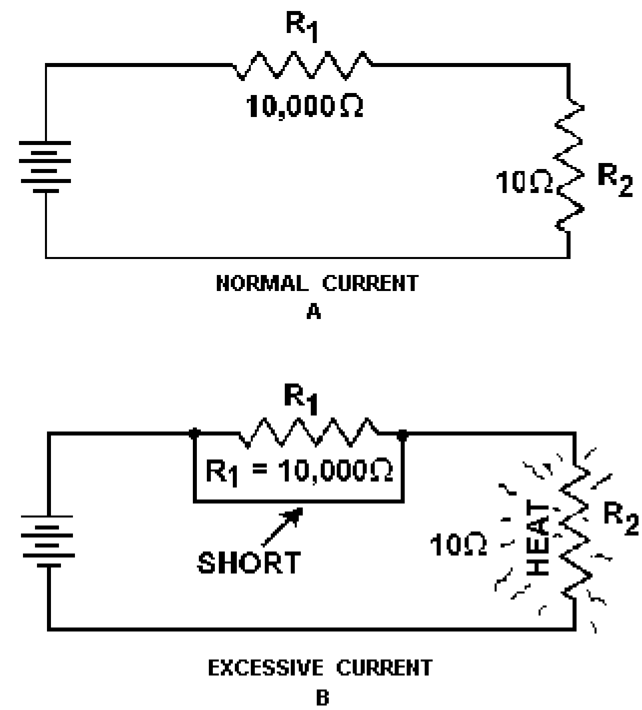


Figure 3-33.—Normal and short circuit conditions.

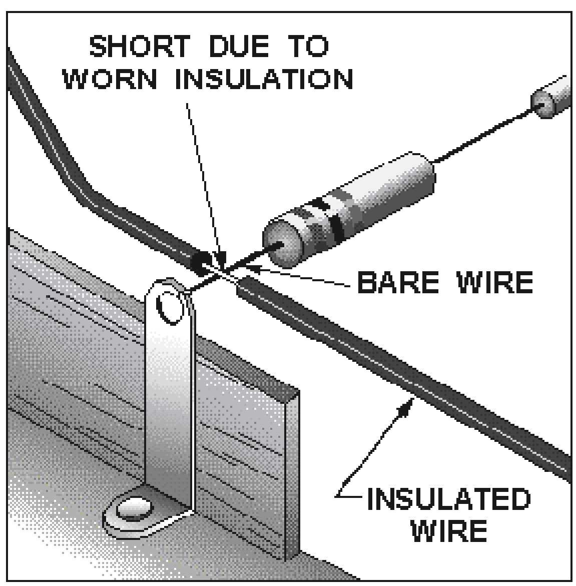


Figure 3-34.—Short due to broken insulation.

SOURCE RESISTANCE

A meter connected across the terminals of a good 1.5-volt battery reads about 1.5 volts. When the same battery is inserted into a complete circuit, the meter reading decreases to something less than 1.5 volts. This difference in terminal voltage is caused by the INTERNAL RESISTANCE of the battery (the opposition to current offered by the electrolyte in the battery). All sources of electromotive force have some form of internal resistance which causes a drop in terminal voltage as current flows through the source.

This principle is illustrated in figure 3-35, where the internal resistance of a battery is shown as R_i. In the schematic, the internal resistance is indicated by an additional resistor in series with the battery. The battery, with its internal resistance, is enclosed within the dotted lines of the schematic diagram. With the switch open, the voltage across the battery terminals reads 15 volts. When the switch is closed, current flow causes voltage drops around the circuit. The circuit current of 2 amperes causes a voltage drop of 2 volts across R_i. The 1-ohm internal battery resistance thereby drops the battery terminal voltage to 13 volts. Internal resistance cannot be measured directly with a meter. An attempt to do this would damage the meter.

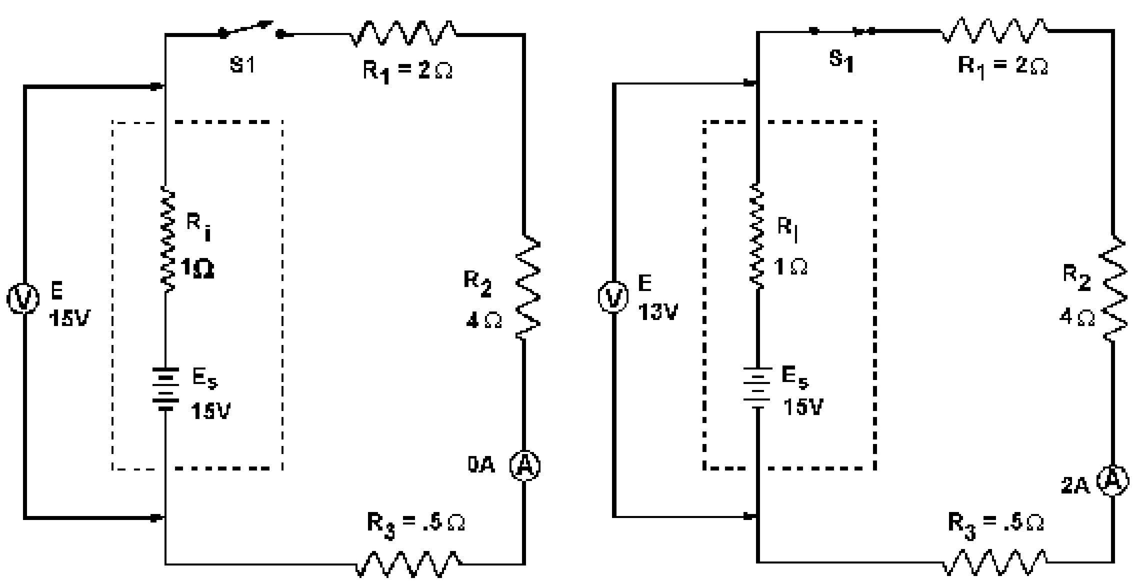
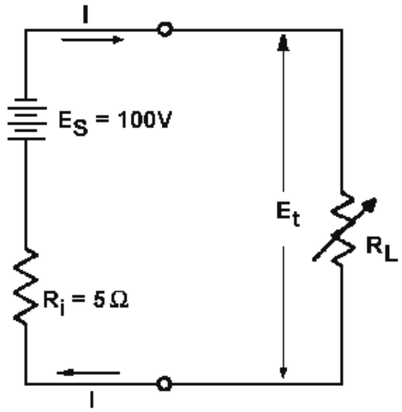


Figure 3-35.—Effect of internal resistance.

The effect of the source resistance on the power output of a dc source may be shown by an analysis of the circuit in figure 3-36. When the variable load resistor (R_L) is set at the zero-ohm position (equivalent to a short circuit), current (I) is calculated using the following formula:

$$I = \frac{E_S}{R_i} = \frac{100 \text{ volts}}{5 \text{ ohms}} = 20 \text{ amperes}$$

This is the maximum current that may be drawn from the source. The terminal voltage across the short circuit is zero volts and all the voltage is across the resistance within the source.

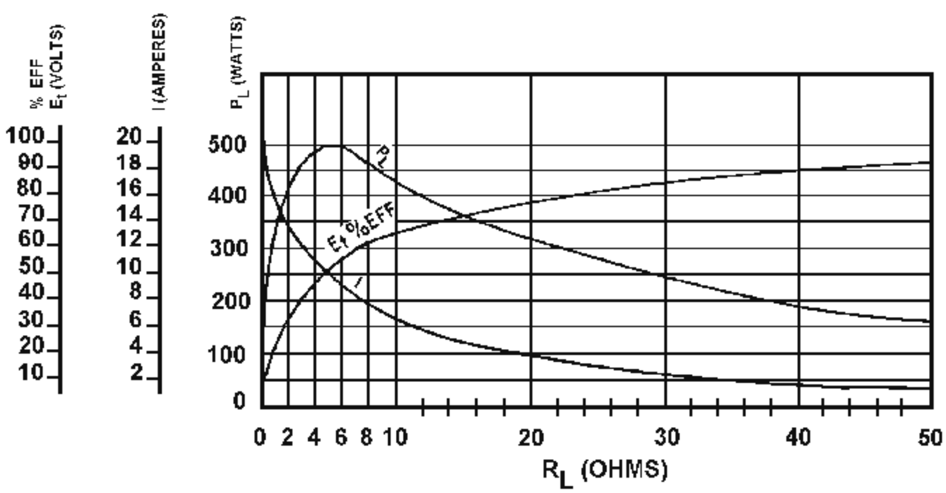


E_S = OPEN - CIRCUIT VOLTAGE OF SOURCE
R_i = INTERNAL RESISTANCE OF SOURCE
E_t = TERMINAL VOLTAGE
R_L = RESISTANCE OF LOAD
P_L = POWER USED IN LOAD
I = CURRENT FROM SOURCE
% EFF. = PERCENTAGE OF EFFICIENCY

(A)
CIRCUIT AND SYMBOL DESIGNATION

R_L	E_t	I	P_L	%EFF.
0	0	20	0	0
1	16.7	16.7	278.9	16.7
2	28.6	14.3	409	28.6
3	37.5	12.5	468.8	37.5
4	44.4	11.1	492.8	44.4
5	50	10	500	50
6	54.5	9.1	496.0	54.5
7	58.3	8.3	483.9	58.3
8	61.6	7.7	474.3	61.6
9	64.3	7.1	456.5	64.3
10	66.7	6.7	446.9	66.7
20	80	4	320	80
30	85.7	2.9	248.5	85.7
40	88.9	2.2	195.6	88.9
50	90.9	1.9	172.7	90.9

(B)
CHART



(C)
GRAPH

Figure 3-36.—Effect of source resistance on power output.

If the load resistance (R_L) were increased (the internal resistance remaining the same), the current drawn from the source would decrease. Consequently, the voltage drop across the internal resistance would decrease. At the same time, the terminal voltage applied across the load would increase and approach a maximum as the current approaches zero amps.

POWER TRANSFER AND EFFICIENCY

Maximum power is transferred from the source to the load when the resistance of the load is equal to the internal resistance of the source. This theory is illustrated in the table and the graph of figure 3-36. When the load resistance is 5 ohms, matching the source resistance, the maximum power of 500 watts is developed in the load.

The efficiency of power transfer (ratio of output power to input power) from the source to the load increases as the load resistance is increased. The efficiency approaches 100 percent as the load resistance approaches a relatively large value compared with that of the source, since less power is lost in the source. The efficiency of power transfer is only 50 percent at the maximum power transfer point (when the load resistance equals the internal resistance of the source). The efficiency of power transfer approaches zero efficiency when the load resistance is relatively small compared with the internal resistance of the source. This is also shown on the chart of figure 3-36.

The problem of a desire for both high efficiency and maximum power transfer is resolved by a compromise between maximum power transfer and high efficiency. Where the amounts of power involved are large and the efficiency is important, the load resistance is made large relative to the source resistance so that the losses are kept small. In this case, the efficiency is high. Where the problem of matching a source to a load is important, as in communications circuits, a strong signal may be more important than a high percentage of efficiency. In such cases, the efficiency of power transfer should be only about 50 percent; however, the power transfer would be the maximum which the source is capable of supplying.

You should now understand the basic concepts of series circuits. The principles which have been presented are of lasting importance. Once equipped with a firm understanding of series circuits, you hold the key to an understanding of the parallel circuits to be presented next.

Q25. A circuit has a source voltage of 100 volts and two 50-ohm resistors connected in series. If the reference point for this circuit is placed between the two resistors, what would be the voltage at the reference point?

Q26. If the reference point in question 25 were connected to ground, what would be the voltage level of the reference point?

Q27. What is an open circuit?

Q28. What is a short circuit?

Q29. Why will a meter indicate more voltage at the battery terminal when the battery is out of a circuit than when the battery is in a circuit?

Q30. What condition gives maximum power transfer from the source to the load?

Q31. What is the efficiency of power transfer in question 30?

Q32. A circuit has a source voltage of 25 volts. The source resistance is 1 ohm and the load resistance is 49 ohms. What is the efficiency of power transfer?

PARALLEL DC CIRCUITS

The discussion of electrical circuits presented up to this point has been concerned with series circuits in which there is only one path for current. There is another basic type of circuit known as the PARALLEL CIRCUIT with which you must become familiar. Where the series circuit has only one path for current, the parallel circuit has more than one path for current.

Ohm's law and Kirchhoff's law apply to all electrical circuits, but the characteristics of a parallel dc circuit are different than those of a series dc circuit.

PARALLEL CIRCUIT CHARACTERISTICS

A PARALLEL CIRCUIT is defined as one having more than one current path connected to a common voltage source. Parallel circuits, therefore, must contain two or more resistances which are not connected in series. An example of a basic parallel circuit is shown in figure 3-37.

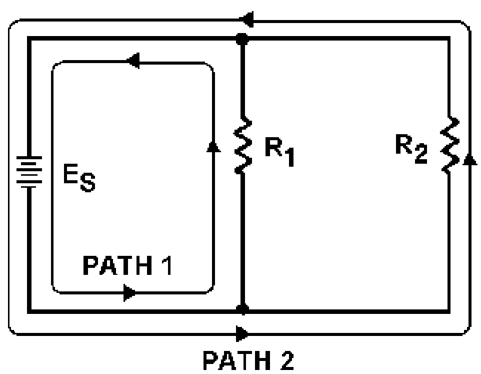


Figure 3-37.—Example of a basic parallel circuit.

Start at the voltage source (E_s) and trace counterclockwise around the circuit. Two complete and separate paths can be identified in which current can flow. One path is traced from the source, through resistance R_1, and back to the source. The other path is from the source, through resistance R_2, and back to the source.

Voltage in a Parallel Circuit

You have seen that the source voltage in a series circuit divides proportionately across each resistor in the circuit. IN A PARALLEL CIRCUIT, THE SAME VOLTAGE IS PRESENT IN EACH BRANCH. (A branch is a section of a circuit that has a complete path for current.) In figure 3-37 this voltage is equal to the applied voltage (E_s). This can be expressed in equation form as:

$$E_S = E_{R1} = E_{R2}$$

Voltage measurements taken across the resistors of a parallel circuit, as illustrated by figure 3-38 verify this equation. Each meter indicates the same amount of voltage. Notice that the voltage across each resistor is the same as the applied voltage.

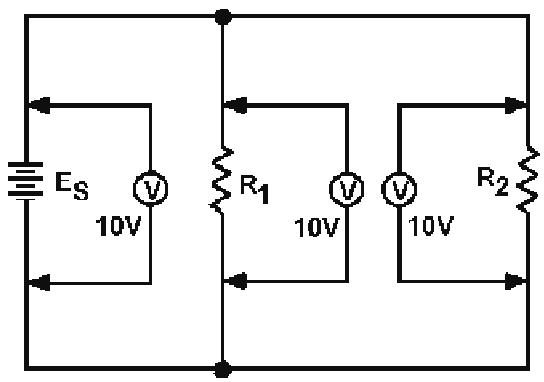


Figure 3-38.—Voltage comparison in a parallel circuit.

Example: Assume that the current through a resistor of a parallel circuit is known to be 4.5 milliamperes (4.5 mA) and the value of the resistor is 30,000 ohms (30 kΩ). Determine the source voltage. The circuit is shown in figure 3-39.

Given:

$$R_2 = 30{,}000 \text{ ohms } (30\text{k}\Omega)$$
$$I_{R2} = 4.5 \text{ milliamps } (4.5\text{mA or } .0045 \text{ amps})$$

Solution:

$$E = IR$$
$$E_{R2} = .0045 \text{ amp} \times 30{,}000 \text{ ohms}$$
$$E_{R2} = 135 \text{ volts}$$

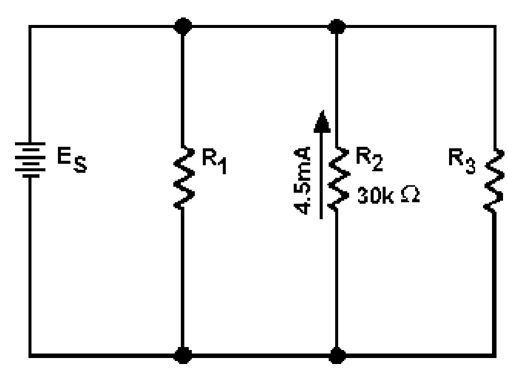


Figure 3-39.—Example problem parallel circuit.

Since the source voltage is equal to the voltage of a branch:

$$E_S = E_{R2}$$
$$E_S = 135 \text{ volts}$$

To simplify the math operation, the values can be expressed in powers of ten as follows:

$$30{,}000 \text{ ohms} = 30 \times 10^3 \text{ ohms}$$
$$4.5\text{mA} = 4.5 \times 10^{-3} \text{ amps}$$
$$E_{R2} = (4.5 \times 10^{-3}) \text{ amps} \times (30 \times 10^3) \text{ ohms}$$
$$E_{R2} = (4.5 \times 30 \times 10^{-3} \times 10^3) \text{ volts}$$
$$(10^{-3} \times 10^3 = 10^{-3+3} = 10^0 = 1)$$
$$E_{R2} = (4.5 \times 30 \times 1) \text{ volts}$$
$$E_{R2} = 135 \text{ volts}$$
$$E_S = E_{R2}$$
$$E_S = 135 \text{ volts}$$

If you are not familiar with the use of the powers of 10 or would like to brush up on it, Mathematics, Vol. 1, NAVEDTRA 10069-C, will be of great help to you.

Q33. What would the source voltage (E_S) in figure 3-39 be if the current through R_2 were 2 milliamps?

Current in a Parallel Circuit

Ohm's law states that the current in a circuit is inversely proportional to the circuit resistance. This fact is true in both series and parallel circuits.

There is a single path for current in a series circuit. The amount of current is determined by the total resistance of the circuit and the applied voltage. <u>In a parallel circuit the source current divides among the available paths.</u>

The behavior of current in parallel circuits will be shown by a series of illustrations using example circuits with different values of resistance for a given value of applied voltage.

Part (A) of figure 3-40 shows a basic series circuit. Here, the total current must pass through the single resistor. The amount of current can be determined.

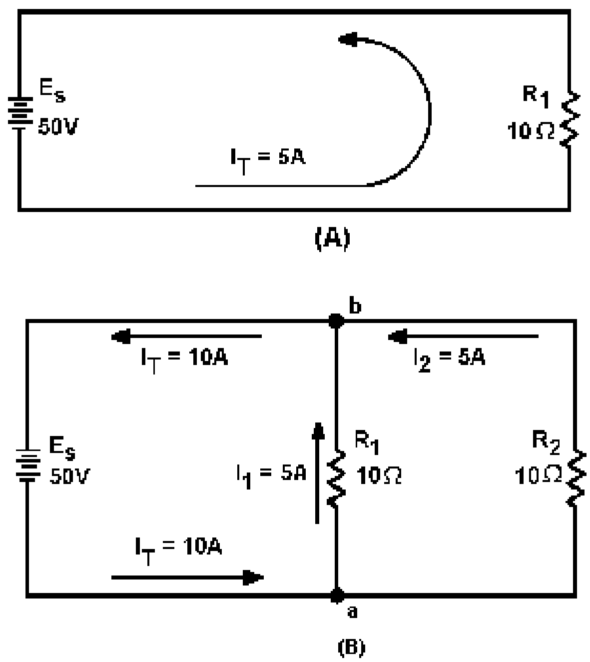


Figure 3-40.—Analysis of current in parallel circuit.

Given:

$$E_s = 50 \text{ volts}$$
$$R_1 = 10 \text{ ohms}$$

Solution:

$$I = \frac{E}{R}$$
$$I_T = \frac{E_s}{R_1}$$
$$I_T = \frac{50 \text{ volts}}{10 \text{ ohms}}$$
$$I_T = 5 \text{ amps}$$

Part (B) of figure 3-40 shows the same resistor (R_1) with a second resistor (R_2) of equal value connected in parallel across the voltage source. When Ohm's law is applied, the current flow through each resistor is found to be the same as the current through the single resistor in part (A).

Given:

$$E_S = 50 \text{ volts}$$
$$R_1 = 10 \text{ ohms}$$
$$R_2 = 10 \text{ ohms}$$

Solution:

$$I = \frac{E}{R}$$

$$E_S = E_{R1} = E_{R2}$$

$$I_{R1} = \frac{E_{R1}}{R_1}$$

$$I_{R1} = \frac{50 \text{ volts}}{10 \text{ ohms}}$$

$$I_{R1} = 5 \text{ amps}$$

$$I_{R2} = \frac{E_{R2}}{R_2}$$

$$I_{R2} = \frac{50 \text{ volts}}{10 \text{ ohms}}$$

$$I_{R2} = 5 \text{ amps}$$

It is apparent that if there is 5 amperes of current through each of the two resistors, there must be a TOTAL CURRENT of 10 amperes drawn from the source.

The total current of 10 amperes, as illustrated in figure 3-40(B), leaves the negative terminal of the battery and flows to point a. Since point a is a connecting point for the two resistors, it is called a JUNCTION. At junction a, the total current divides into two currents of 5 amperes each. These two currents flow through their respective resistors and rejoin at junction b. The total current then flows from junction b back to the positive terminal of the source. The source supplies a total current of 10 amperes and each of the two equal resistors carries one-half the total current.

Each individual current path in the circuit of figure 3-40(B) is referred to as a BRANCH. Each branch carries a current that is a portion of the total current. Two or more branches form a NETWORK.

From the previous explanation, the characteristics of current in a parallel circuit can be expressed in terms of the following general equation:

$$I_T = I_1 + I_2 + \ldots I_n$$

Compare part (A) of figure 3-41 with part (B) of the circuit in figure 3-40. Notice that doubling the value of the second branch resistor (R_2) has no effect on the current in the first branch (I_{R1}), but does reduce the second branch current (I_{R2}) to one-half its original value. The total circuit current drops to a value equal to the sum of the branch currents. These facts are verified by the following equations.

Given:

$$E_S = 50 \text{ volts}$$
$$R_1 = 10 \text{ ohms}$$
$$R_2 = 20 \text{ ohms}$$

Solution:

$$I = \frac{E}{R}$$

$$E_S = E_{R1} = E_{R2}$$

$$I = \frac{E_{R1}}{R_1}$$

$$I = \frac{50 \text{ volts}}{10 \text{ ohms}}$$

$$I_{R1} = 5 \text{ amps}$$

$$I_{R2} = \frac{E_{R2}}{R_2}$$

$$I_{R2} = \frac{50 \text{ volts}}{20 \text{ ohms}}$$

$$I_{R2} = 2.5 \text{ amps}$$

$$I_T = I_{R1} + I_{R2}$$

$$I_T = 5 \text{ amps} + 2.5 \text{ amps}$$

$$I_T = 7.5 \text{ amps}$$

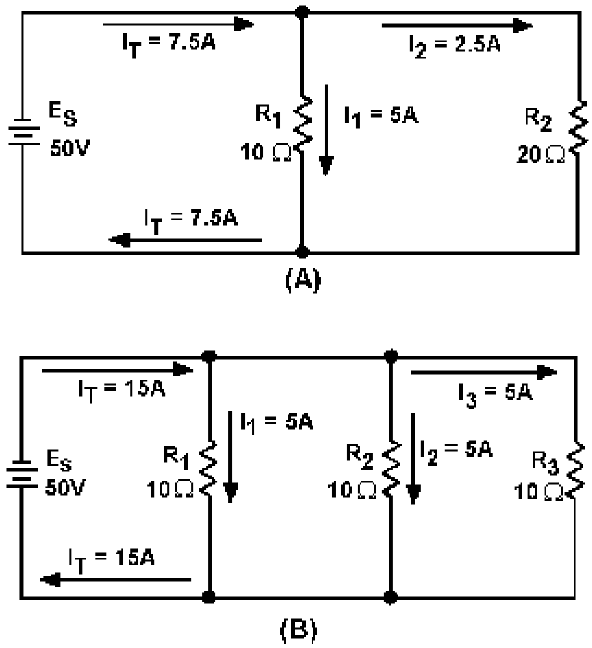


Figure 3-41.—Current behavior in parallel circuits.

The amount of current flow in the branch circuits and the total current in the circuit shown in figure 3-41(B) are determined by the following computations.

Given:

$$E_S = 50 \text{ volts}$$
$$R_1 = 10 \text{ ohms}$$
$$R_2 = 10 \text{ ohms}$$
$$R_3 = 10 \text{ ohms}$$

Solution:

$$I = \frac{E}{R}$$

$$E_S = E_{R1} = E_{R2} = E_{R3}$$

$$I_{R1} = \frac{E_{R1}}{R_1}$$

$$I_{R1} = \frac{50 \text{ volts}}{10 \text{ ohms}}$$

$$I_{R1} = 5 \text{ amps}$$

$$I_{R2} = \frac{E_{R2}}{R_2}$$

$$I_{R2} = \frac{50 \text{ volts}}{10 \text{ ohms}}$$

$$I_{R2} = 5 \text{ amps}$$

$$I_{R3} = \frac{E_{R3}}{R_3}$$

$$I_{R3} = \frac{50 \text{ volts}}{10 \text{ ohms}}$$

$$I_{R3} = 5 \text{ amps}$$

$$I_T = I_{R1} + I_{R2} + I_{R3}$$

$$I_T = 5 \text{ amps} + 5 \text{ amps} + 5 \text{ amps}$$

$$I_T = 15 \text{ amps}$$

Notice that the sum of the ohmic values in each circuit shown in figure 3-41 is equal (30 ohms), and that the applied voltage is the same (50 volts). However, the total current in 3-41(B) (15 amps) is twice the amount in 3-41(A) (7.5 amps). It is apparent, therefore, that the manner in which resistors are connected in a circuit, as well as their actual ohmic values, affect the total current.

The division of current in a parallel network follows a definite pattern. This pattern is described by KIRCHHOFF'S CURRENT LAW which states:

"The algebraic sum of the currents entering and leaving any junction of conductors is equal to zero."

This law can be stated mathematically as:

$$I_a + l_b + \ldots I_n = 0$$

where: I_a, I_b, etc., are the currents entering and leaving the junction. Currents ENTERING the junction are considered to be POSITIVE and currents LEAVING the junction are considered to be NEGATIVE. When solving a problem using Kirchhoff's current law, the currents must be placed into the equation WITH THE PROPER POLARITY SIGNS ATTACHED.

Example: Solve for the value of I_3 in figure 3-42.

Given:

$$I_1 = 10 \text{ amps}$$
$$I_2 = 3 \text{ amps}$$
$$I_4 = 5 \text{ amps}$$

$$I_a + I_b + \ldots I_n = 0$$

Solution:

$$I_a + l_b + \ldots I_a = 0$$

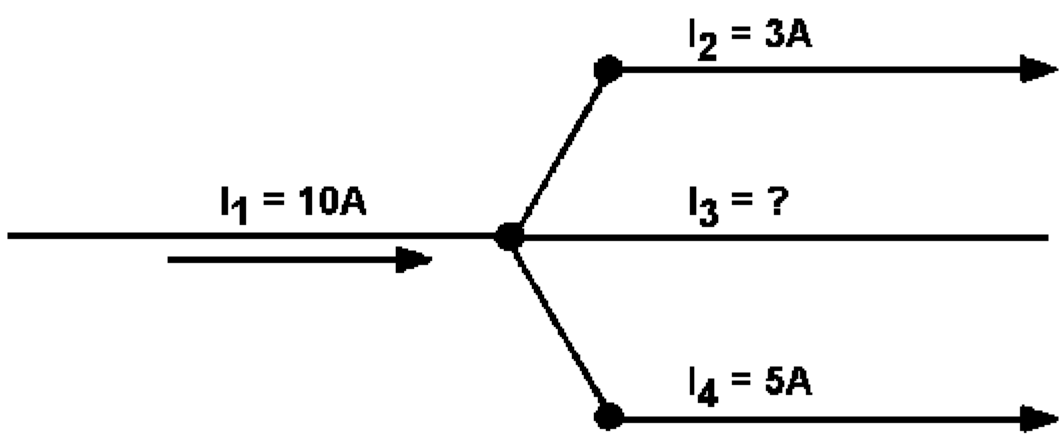


Figure 3-42.—Circuit for example problem.

The currents are placed into the equation with the proper signs.

$$I_1 + I_2 + I_3 + I_4 = 0$$
$$10 \text{ amps} + (-3 \text{ amps}) + I_3 + (-5 \text{ amps}) = 0$$
$$I_3 + 2 \text{ amps} = 0$$
$$I_3 = -2 \text{ amps}$$

I_3 has a value of 2 amperes, and the negative sign shows it to be a current LEAVING the junction.

Example. Using figure 3-43, solve for the magnitude and direction of I_3.

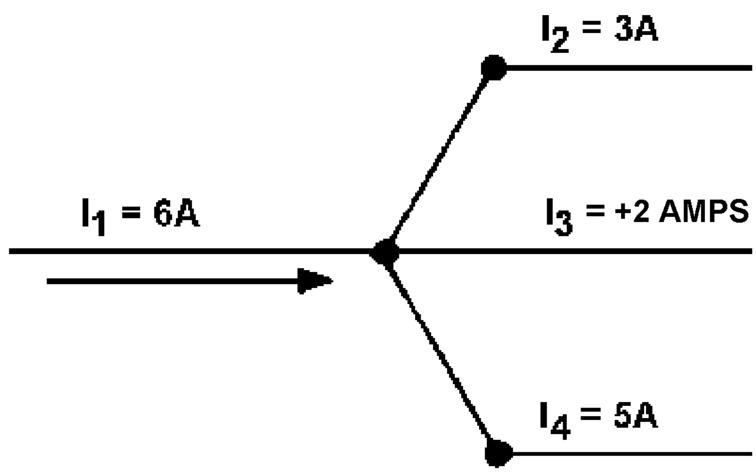


Figure 3-43.—Circuit for example problem.

Given:

$$I_1 = 6 \text{ amps}$$
$$I_2 = 3 \text{ amps}$$
$$I_4 = 5 \text{ amps}$$

Solution:

$$I_a + I_b + \dots I_n = 0$$

$$I_1 + I_2 + I_3 + I_4 = 0$$

$$6 \text{ amps} + (-3 \text{ amps}) + I_3 + (-5 \text{ amps}) = 0$$

$$I_3 + (+2 \text{ amps}) = 0$$

$$I_3 = +2 \text{ amps}$$

I_3 is 2 amperes and its positive sign shows it to be a current entering the junction.

Q34. There is a relationship between total current and current through the individual components in a circuit. What is this relationship in a series circuit and a parallel circuit?

Q35. In applying Kirchhoff's current law, what does the polarity of the current indicate?

Resistance in a Parallel Circuit

In the example diagram, figure 3-44, there are two resistors connected in parallel across a 5-volt battery. Each has a resistance value of 10 ohms. A complete circuit consisting of two parallel paths is formed and current flows as shown.

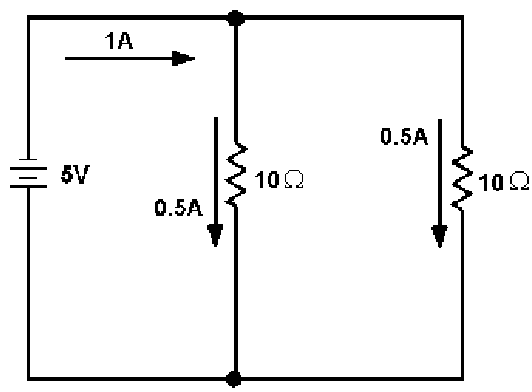


Figure 3-44.—Two equal resistors connected in parallel.

Computing the individual currents shows that there is one-half of an ampere of current through each resistance. The total current flowing from the battery to the junction of the resistors, and returning from the resistors to the battery, is equal to 1 ampere.

The total resistance of the circuit can be calculated by using the values of total voltage (E_T) and total current (I_T).

NOTE: From this point on the abbreviations and symbology for electrical quantities will be used in example problems.

Given:

$$E_T = 5\ V$$
$$I_T = 1\ A$$

Solution:

$$R = \frac{E}{I}$$
$$R_T = \frac{E_T}{I_T}$$
$$R_T = \frac{5\ V}{1\ A}$$
$$R_T = 5\ \Omega$$

This computation shows the total resistance to be 5 ohms; one-half the value of either of the two resistors.

Since the total resistance of a parallel circuit is smaller than any of the individual resistors, total resistance of a parallel circuit is not the sum of the individual resistor values as was the case in a series circuit. The total resistance of resistors in parallel is also referred to as EQUIVALENT RESISTANCE (R_{eq}). The terms total resistance and equivalent resistance are used interchangeably.

There are several methods used to determine the equivalent resistance of parallel circuits. The best method for a given circuit depends on the number and value of the resistors. For the circuit described above, where all resistors have the same value, the following simple equation is used:

$$R_{eq} = \frac{R}{N}$$

R_{eq} = equivalent parallel resistance

R = ohmic value of one resistor

N = number of resistors

This equation is valid for any number of parallel resistors of EQUAL VALUE.

Example: Four 40-ohm resistors are connected in parallel. What is their equivalent resistance?

Given:

$$R_1 + R_2 + R_3 + R_4$$
$$R_1 = 40\ \Omega$$

Solution:

$$R_{eq} = \frac{R}{N}$$
$$R_{eq} = \frac{40\Omega}{4}$$
$$R_{eq} = 10\ \Omega$$

Figure 3-45 shows two resistors of unequal value in parallel. Since the total current is shown, the equivalent resistance can be calculated.

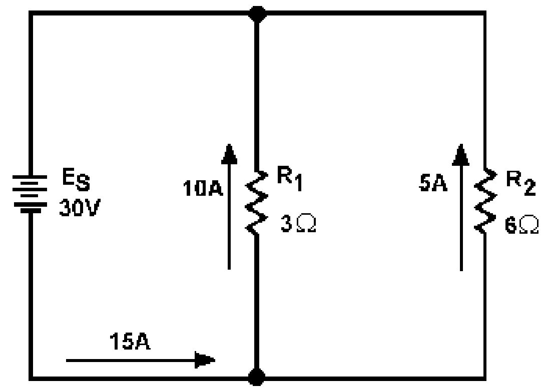


Figure 3-45.—Example circuit with unequal parallel resistors.

Given:

$$E_S = 30\text{ V}$$
$$I_T = 15\text{ A}$$

Solution:

$$R_{eq} = \frac{E_S}{I_T}$$
$$R_{eq} = \frac{30\text{ V}}{15\text{ A}}$$
$$R_{eq} = 2\,\Omega$$

The equivalent resistance of the circuit shown in figure 3-45 is smaller than either of the two resistors (R_1, R_2). An important point to remember is that the equivalent resistance of a parallel circuit is always less than the resistance of any branch.

Equivalent resistance can be found if you know the individual resistance values and the source voltage. By calculating each branch current, adding the branch currents to calculate total current, and dividing the source voltage by the total current, the total can be found. This method, while effective, is somewhat lengthy. A quicker method of finding equivalent resistance is to use the general formula for resistors in parallel:

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots \frac{1}{R_n}$$

If you apply the general formula to the circuit shown in figure 3-45 you will get the same value for equivalent resistance (2Ω) as was obtained in the previous calculation that used source voltage and total current.

Given:

$$R_1 = 3\Omega$$
$$R_2 = 6\Omega$$

Solution:

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2}$$
$$\frac{1}{R_{eq}} = \frac{1}{3\Omega} + \frac{1}{6\Omega}$$

Convert the fractions to a common denominator.

$$\frac{1}{R_{eq}} = \frac{2}{6\Omega} + \frac{1}{6\Omega}$$
$$\frac{1}{R_{eq}} = \frac{3}{6\Omega}$$
$$\frac{1}{R_{eq}} = \frac{1}{2\Omega}$$

Since both sides are reciprocals (divided into one), disregard the reciprocal function.

$$R_{eq} = 2\Omega$$

The formula you were given for equal resistors in parallel

$$(R_{eq} = \frac{R}{N})$$

is a simplification of the general formula for resistors in parallel

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots \frac{1}{R_n}$$

There are other simplifications of the general formula for resistors in parallel which can be used to calculate the total or equivalent resistance in a parallel circuit.

RECIPROCAL METHOD.—This method is based upon taking the reciprocal of each side of the equation. This presents the general formula for resistors in parallel as:

$$R_{eq} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \ldots \frac{1}{R_n}}$$

This formula is used to solve for the equivalent resistance of a number of unequal parallel resistors. You must find the lowest common denominator in solving these problems. If you are a little hazy on finding the lowest common denominator, brush up on it in *Mathematics Volume 1*, NAVEDTRA 10069 (Series).

Example: Three resistors are connected in parallel as shown in figure 3-46. The resistor values are: R_1 = 20 ohms, R_2 = 30 ohms, R_3 = 40 ohms. What is the equivalent resistance? (Use the reciprocal method.)

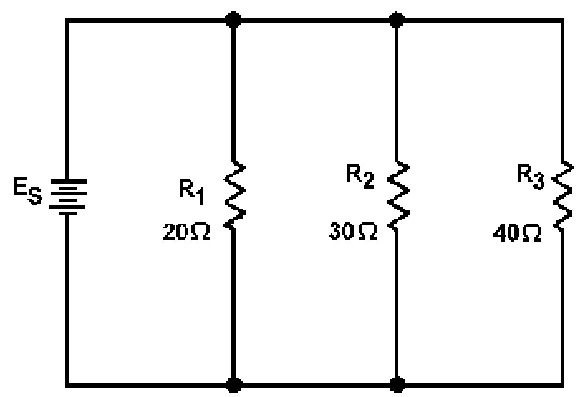


Figure 3-46.—Example parallel circuit with unequal branch resistors.

Given:

$$R_1 = 20\Omega$$
$$R_2 = 30\Omega$$
$$R_3 = 40\Omega$$

Solution:

$$R_{eq} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R3}}$$

$$R_{eq} = \frac{1}{\frac{1}{20\Omega} + \frac{1}{30\Omega} + \frac{1}{40\Omega}}$$

$$R_{eq} = \frac{1}{\frac{6}{120\Omega} + \frac{4}{120\Omega} + \frac{3}{120\Omega}}$$

$$R_{eq} = \frac{1}{\frac{13}{120}\Omega}$$

$$R_{eq} = \frac{120}{13}\Omega$$

$$R_{eq} = 9.23\Omega$$

PRODUCT OVER THE SUM METHOD.—A convenient method for finding the equivalent, or total, resistance of two parallel resistors is by using the following formula.

$$R_{eq} = \frac{R_1 \times R_2}{R_1 + R_2}$$

This equation, called the product over the sum formula, is used so frequently it should be committed to memory.

Example: What is the equivalent resistance of a 20-ohm and a 30-ohm resistor connected in parallel, as in figure 3-47?

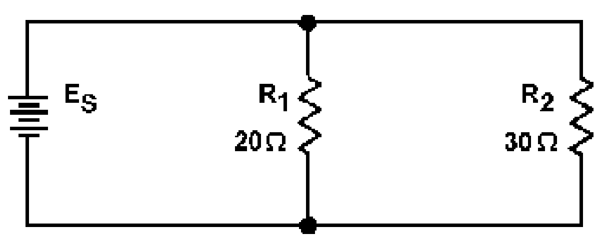


Figure 3-47.—Parallel circuit with two unequal resistors.

Given:

$$R_1 = 20\Omega$$
$$R_2 = 30\Omega$$

Solution:

$$R_{eq} = \frac{R_1 \times R_2}{R_1 + R_2}$$
$$R_{eq} = \frac{20\Omega \times 30\Omega}{20\Omega + 30\Omega}$$
$$R_{eq} = \frac{600}{50}\Omega$$
$$R_{eq} = 12\Omega$$

Q36. Four equal resistors are connected in parallel, each resistor has an ohmic value of 100 ohms, what is the equivalent resistance?

Q37. Three resistors connected in parallel have values of 12 kΩ, 20 kΩ, and 30 kΩ. What is the equivalent resistance?

Q38. Two resistors connected in parallel have values of 10 kΩ and 30 kΩ. What is the equivalent resistance?

Power in a Parallel Circuit

Power computations in a parallel circuit are essentially the same as those used for the series circuit. Since power dissipation in resistors consists of a heat loss, power dissipations are additive regardless of how the resistors are connected in the circuit. The total power is equal to the sum of the power dissipated by the individual resistors. Like the series circuit, the total power consumed by the parallel circuit is:

$$P_T = P_1 + P_2 + \ldots P_n$$

Example: Find the total power consumed by the circuit in figure 3-48.

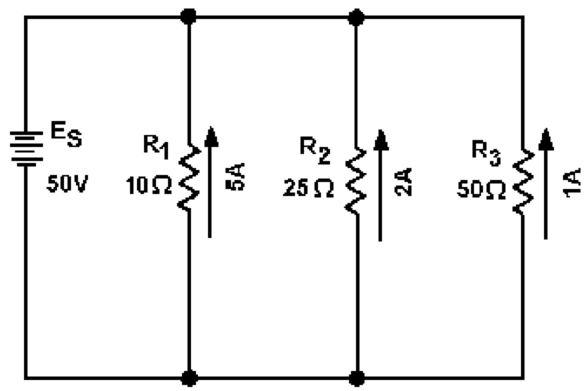


Figure 3-48.—Example parallel circuit.

Given:

$$R_1 = 10\ \Omega$$
$$I_{R1} = 5A$$

$$R_2 = 25\ \Omega$$
$$I_{R2} = 2A$$

$$R_3 = 50\ \Omega$$
$$I_{R3} = 1A$$

Solution:

$$P = I^2R$$
$$P_{R1} = (I_{R1})^2 \times R_1$$
$$P_{R1} = (5\ A)^2 \times 10\Omega$$
$$P_{R1} = 250W$$

$$P_{R2} = (I_{R2})^2 \times R_2$$
$$P_{R2} = (2\ A)^2 \times 25\Omega$$
$$P_{R2} = 100W$$

$$P_{R3} = (I_{R3})^2 \times R_3$$
$$P_{R3} = (1\ A)^2 \times 50\Omega$$
$$P_{R3} = 50W$$

$$P_T = P_{R1} + P_{R2} + P_{R3}$$
$$P_T = 250W + 100W + 50W$$
$$P_T = 400W$$

Since the total current and source voltage are known, the total power can also be computed by:

Given:

$$E_S = 50\ V$$
$$I_T = 8\ A$$

Solution:

$$P_T = E_S \times I_T$$
$$P_T = 50\ V \times 8\ A$$
$$P_T = 400W$$

Equivalent Circuits

In the study of electricity, it is often necessary to reduce a complex circuit into a simpler form. Any complex circuit consisting of resistances can be redrawn (reduced) to a basic equivalent circuit containing the voltage source and a single resistor representing total resistance. This process is called reduction to an EQUIVALENT CIRCUIT.

Figure 3-49 shows a parallel circuit with three resistors of equal value and the redrawn equivalent circuit. The parallel circuit shown in part A shows the original circuit. To create the equivalent circuit, you must first calculate the equivalent resistance.

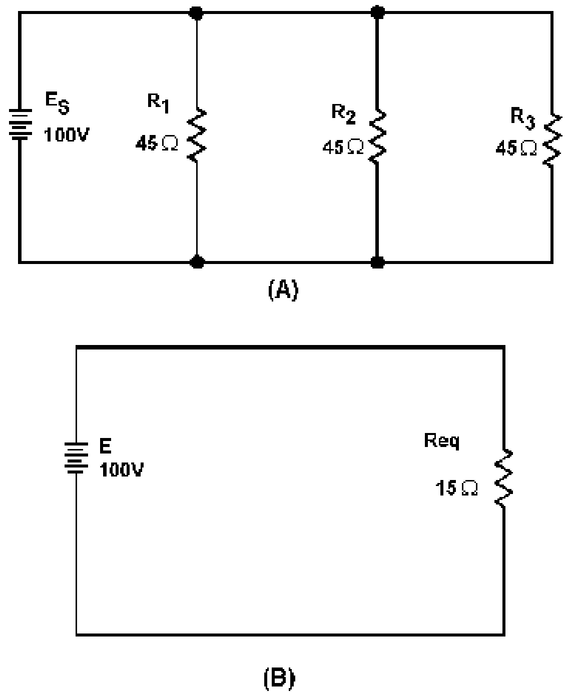


Figure 3-49.—Parallel circuit with equivalent circuit.

Given:

$$R_1 = 45\Omega$$
$$R_2 = 45\Omega$$
$$R3 = 45\Omega$$

Solution:

$$R_{eq} = \frac{R}{N}$$
$$R_{eq} = \frac{45\Omega}{3}$$
$$R_{eq} = 15\ \Omega$$

Once the equivalent resistance is known, a new circuit is drawn consisting of a single resistor (to represent the equivalent resistance) and the voltage source, as shown in part B.

Rules for Parallel DC Circuits

1. The same voltage exists across each branch of a parallel circuit and is equal to the source voltage.

2. The current through a branch of a parallel network is inversely proportional to the amount of resistance of the branch.

3. The total current of a parallel circuit is equal to the sum of the individual branch currents of the circuit.

4. The total resistance of a parallel circuit is found by the general formula:

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots \frac{1}{R_n}$$

or one of the formulas derived from this general formula.

5. The total power consumed in a parallel circuit is equal to the sum of the power consumptions of the individual resistances.

SOLVING PARALLEL CIRCUIT PROBLEMS

Problems involving the determination of resistance, voltage, current, and power in a parallel circuit are solved as simply as in a series circuit. The procedure is the same — (1) draw the circuit diagram, (2) state the values given and the values to be found, (3) select the equations to be used in solving for the unknown quantities based upon the known quantities, and (4) substitute the known values in the equation you have selected and solve for the unknown value.

Example: A parallel circuit consists of five resistors. The value of each resistor is known and the current through R_1 is known. You are asked to calculate the value for total resistance, total power, total current, source voltage, the power used by each resistor, and the current through resistors R_2, R_3, R_4, and R_5.

Given:

$$\begin{aligned} R_1 &= 20\Omega \\ R_2 &= 30\Omega \\ R_3 &= 18\Omega \\ R_4 &= 18\Omega \\ R_5 &= 18\Omega \\ I_{R1} &= 9A \end{aligned}$$

Find:

$$R_T, E_S, I_T, P_T, I_{R2}, I_{R3}, I_{R4},$$
$$I_{R5}, P_{R1}, P_{R2}, P_{R3}, P_{R4}, P_{R5}$$

This may appear to be a large amount of mathematical manipulation. However, if you use the step-by-step approach, the circuit will fall apart quite easily.

The first step in solving this problem is for you to draw the circuit and indicate the known values as shown in figure 3-50.

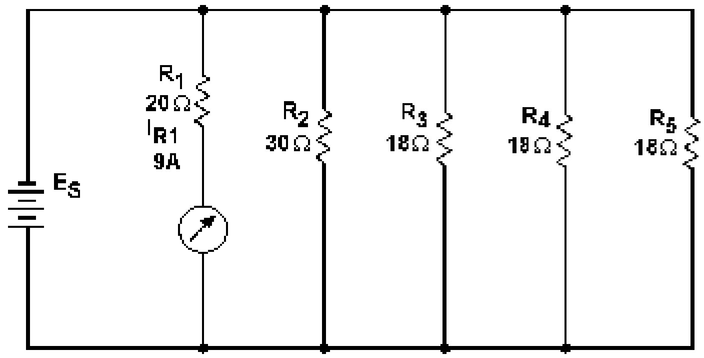


Figure 3-50.—Parallel circuit problem.

There are several ways to approach this problem. With the values you have been given, you could first solve for R_T, the power used by R_1, or the voltage across R_1, which you know is equal to the source voltage and the voltage across each of the other resistors. Solving for R_T or the power used by R_1 will not help in solving for the other unknown values.

Once the voltage across R_1 is known, this value will help you calculate other unknowns. Therefore the logical unknown to solve for is the source voltage (the voltage across R_1).

Given:

$$R_1 = 20\Omega$$
$$I_{R1} = 9A$$
$$E_{R1} = E_S$$

Solution:

$$E_S = R_1 \times I_{R1}$$
$$E_S = 9A \times 20\Omega$$
$$E_S = 180V$$

Now that source voltage is known, you can solve for current in each branch.

Given:

$$E_S = 180V$$
$$R_2 = 30\Omega$$
$$R_3 = 18\Omega$$
$$R_4 = 18\Omega$$
$$R_5 = 18\Omega$$

Solution:

$$I_{R2} = \frac{E_S}{R_2}$$

$$I_{R2} = \frac{180\ \text{V}}{30\ \Omega}$$

$$I_{R2} = 6\ \text{A}$$

$$I_{R3} = \frac{E_S}{R_3}$$

$$I_{R3} = \frac{180\ \text{V}}{18\ \Omega}$$

$$I_{R3} = 10\ \text{A}$$

Since $R_3 = R_4 = R_5$ and the voltage across each branch is the same:

$$I_{R4} = 10\text{A}$$
$$I_{R5} = 10\text{A}$$

Solving for total resistance.

Given:

$$R_1 = 20\Omega$$
$$R_2 = 30\Omega$$
$$R_3 = 18\Omega$$
$$R_4 = 18\Omega$$
$$R_5 = 18\Omega$$

Solution:

$$R_T = R_{eq}$$

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \frac{1}{R_4} + \frac{1}{R_5}$$

$$\frac{1}{R_T} = \frac{1}{20\Omega} + \frac{1}{30\Omega} + \frac{1}{18\Omega} + \frac{1}{18\Omega} + \frac{1}{18\Omega}$$

$$\frac{1}{R_T} = \frac{9 + 6 + 10 + 10 + 10\Omega}{180 \quad (LCD)}$$

$$R_T = \frac{45\Omega}{180}$$

$$R_T = \frac{180}{45\Omega}$$

$$R_T = 4\Omega$$

An alternate method for solving for R_T can be used. By observation, you can see that R_3, R_4, and R_5 are of equal ohmic value. Therefore an equivalent resistor can be substituted for these three resistors in solving for total resistance.

Given:

$$R_3 = R_4 = R_5 = 18\Omega$$

Solution:

$$R_{eq1} = \frac{R}{N}$$

$$R_{eq1} = \frac{18\Omega}{3}$$

$$R_{eq1} = 6\Omega$$

The circuit can now be redrawn using a resistor labeled R_{eq1} in place of R_3, R_4, and R_5 as shown in figure 3-51.

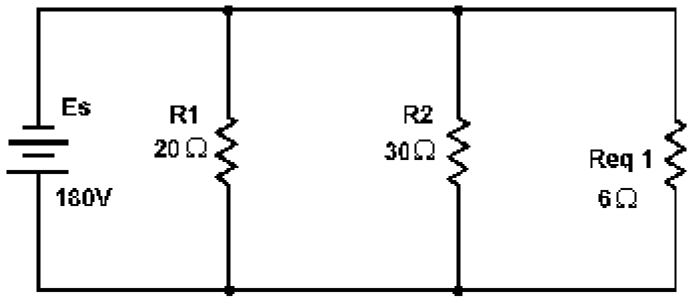


Figure 3-51.—First equivalent parallel circuit.

An equivalent resistor can be calculated and substituted for R_1 and R_2 by use of the product over the sum formula.

Given:

$$R_1 = 20\Omega$$
$$R_2 = 30\Omega$$

Solution:

$$R_{eq} = \frac{R_1 \times R_2}{R_1 + R_2}$$
$$R_{eq\,2} = \frac{20\Omega \times 30\Omega}{20\Omega + 30\Omega}$$
$$R_{eq\,2} = \frac{600}{50}\Omega$$
$$R_{eq\,2} = 12\Omega$$

The circuit is now redrawn again using a resistor labeled R_{eq2} in place of R_1 and R_2 as shown in figure 3-52.

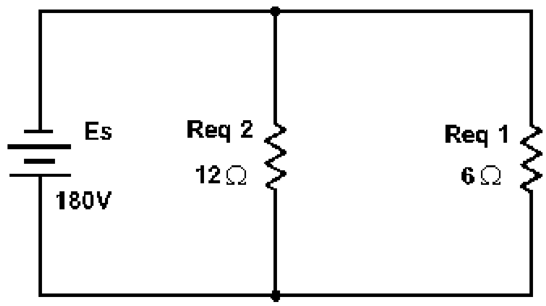


Figure 3-52.—Second equivalent parallel circuit.

You are now left with two resistors in parallel. The product over the sum method can now be used to solve for total resistance.

Given:

$$R_{eq1} = 6\Omega$$
$$R_{eq1} = 12\Omega$$
$$R_T = R_{eq}$$

Solution:

$$R_{eq} = \frac{R_1 \times R_2}{R_1 + R_2}$$

$$R_T = \frac{R_{eq1} \times R_{eq2}}{R_{eq1} + R_{eq2}}$$

$$R_T = \frac{6\Omega \times 12\Omega}{6\Omega + 12\Omega}$$

$$R_T = \frac{72}{18}\Omega$$

$$R_T = 4\Omega$$

This agrees with the solution found by using the general formula for solving for resistors in parallel.

The circuit can now be redrawn as shown in figure 3-53 and total current can be calculated.

Figure 3-53.—Parallel circuit redrawn to final equivalent circuit.

Given:

$$E_S = 180V$$
$$R_T = 4\Omega$$

Solution:

$$I_T = \frac{E_S}{R_T}$$
$$I_T = \frac{180V}{4\Omega}$$
$$I_T = 45A$$

This solution can be checked by using the values already calculated for the branch currents.

Given:

$$I_{R1} = 9A$$
$$I_{R2} = 6A$$
$$I_{R3} = 10A$$
$$I_{R4} = 10A$$
$$I_{R5} = 10A$$

Solution:

$$I_T = I_{R1} + I_{R2} + \ldots I_{Rn}$$
$$I_T = 9A + 6A + 10A + 10A + 10A$$
$$I_T = 45A$$

Now that total current is known, the next logical step is to find total power.

Given:

$$E_S = 180V$$
$$I_T = 45A$$

Solution:

$$P = EI$$
$$P_T = E_S \times I_T$$
$$P_T = 180V \times 45A$$
$$P_T = 8100 \text{ watts} = 8.1 \text{ kW}$$

Solving for the power in each branch.

Given:

$$E_S = 180V$$
$$I_{R1} = 9A$$
$$I_{R2} = 6A$$
$$I_{R3} = 10A$$
$$I_{R4} = 10A$$
$$I_{R5} = 10A$$

Solution:

$$P = EI$$
$$P_{R1} = E_S \times I_{R1}$$
$$P_{R1} = 180V \times 9A$$
$$P_{R1} = 1620W$$

$$P_{R2} = E_S \times I_{R2}$$
$$P_{R2} = 180V \times 6A$$
$$P_{R2} = 1080W$$

$$P_{R3} = E_S \times I_{R3}$$
$$P_{R3} = 180V \times 10A$$
$$P_{R3} = 1800W$$

Since $I_{R3} = I_{R4} = I_{R5}$ then, $P_{R3} = P_{R4} = P_{R5} = 1800$ W. The previous calculation for total power can now be checked.

Given:

$$P_{R1} = 1620W$$
$$P_{R2} = 1080W$$
$$P_{R3} = 1800W$$
$$P_{R4} = 1800W$$
$$P_{R5} = 1800W$$

Solution:

$$P_T = P_{R1} + P_{R2} + P_{R3} + P_{R4} + P_{R5}$$
$$P_T = 1620W + 1080W + 1800W + 1800W + 1800W$$
$$P_T = 8100W$$
$$P_T = 8.1kW$$

Q39. What term identifies a single resistor that represents total resistance of a complex circuit?

Q40. The total power in both series and parallel circuits is computed with the formula: $P_T = P_1 + P_2 + P_3 + ...P_n$. Why can this formula be used for both series and parallel circuits?

Q41. A circuit consists of three resistors connected in parallel across a voltage source. $R_1 = 40\Omega$, $R_2 = 30\Omega$, $R_3 = 40\Omega$, and $P_{R3} = 360$ watts. Solve for R_T, E_S and I_{R2}. (Hint: Draw and label the circuit first.)

SERIES-PARALLEL DC CIRCUITS

In the preceding discussions, series and parallel dc circuits have been considered separately. The technician will encounter circuits consisting of both series and parallel elements. A circuit of this type is referred to as a COMBINATION CIRCUIT. Solving for the quantities and elements in a combination circuit is simply a matter of applying the laws and rules discussed up to this point.

SOLVING COMBINATION-CIRCUIT PROBLEMS

The basic technique used for solving dc combination-circuit problems is the use of equivalent circuits. To simplify a complex circuit to a simple circuit containing only one load, equivalent circuits are substituted (on paper) for the complex circuit they represent. To demonstrate the method used to solve combination circuit problems, the network shown in figure 3-54(A) will be used to calculate various circuit quantities, such as resistance, current, voltage, and power.

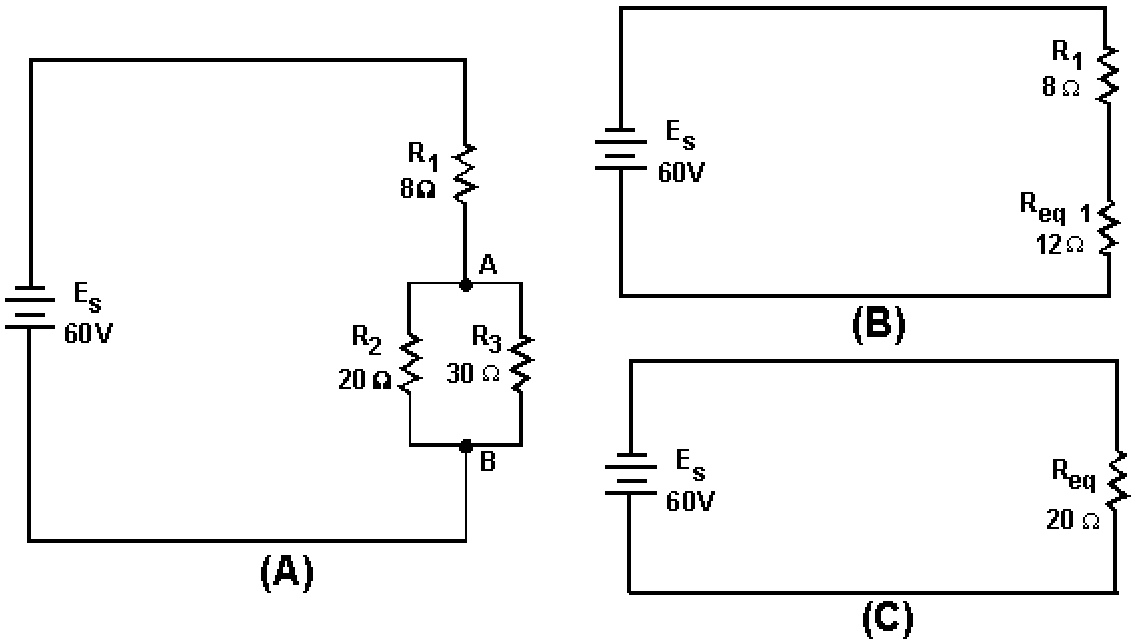


Figure 3-54.—Example combination circuit.

Examination of the circuit shows that the only quantity that can be computed with the given information is the equivalent resistance of R_2 and R_3.

Given:

$$R_2 = 20\Omega$$
$$R_3 = 30\Omega$$

Solution:

$$R_{eq1} = \frac{R_2 \times R_3}{R_2 + R_3} \quad \text{(Product over the sum)}$$

$$R_{eq1} = \frac{20\Omega \times 30\Omega}{20\Omega + 30\Omega}$$

$$R_{eq1} = \frac{600}{50}\Omega$$

$$R_{eq1} = 12\Omega$$

Now that the equivalent resistance for R_2 and R_3 has been calculated, the circuit can be redrawn as a series circuit as shown in figure 3-54(B).

The equivalent resistance of this circuit (total resistance) can now be calculated.

Given:

$$R_1 = 8\Omega \quad \text{(Resistors in series)}$$
$$R_{eq1} = 12\Omega$$

Solution:

$$R_{eq} = R_1 + R_{eq1}$$
$$R_{eq} = 8\Omega + 12\Omega$$
$$R_{eq} = 20\Omega$$
$$\text{or}$$
$$R_T = 20\Omega$$

The original circuit can be redrawn with a single resistor that represents the equivalent resistance of the entire circuit as shown in figure 3-54(C).

To find total current in the circuit:

Given:

$$E_S = 60V$$
$$R_T = 20\Omega$$

Solution:

$$I_T = \frac{E_S}{R_T}$$

$$I_T = \frac{60V}{20\Omega} \quad \text{(Ohm's Law)}$$

$$I_T = 3A$$

To find total power in the circuit:

Given:

$$E_S = 60V$$
$$I_T = 3A$$

Solution:

$$P_T = E_S \times I_T$$
$$P_T = 60V \times 3A$$
$$P_T = 180W$$

To find the voltage dropped across R_1, R_2, and R_3, refer to figure 3-54(B). R_{eq1} represents the parallel network of R_2 and R_3. Since the voltage across each branch of a parallel circuit is equal, the voltage across R_{eq1} (E_{eq1}) will be equal to the voltage across R_2 (E_{R2}) and also equal to the voltage across R_3 (E_{R3}).

Given:

$$I_T = 3A \quad \text{(Current through each part of a series circuit is equal to total current)}$$
$$R_1 = 8\Omega$$
$$R_{eq1} = 12\Omega$$

Solution:

$$E_{R1} = I_1 \times R_1$$
$$E_{R1} = 3A \times 8\Omega$$
$$E_{R1} = 24V$$
$$E_{R2} = E_{R3} = E_{eq1}$$
$$E_{eq1} = I_T \times R_{eq1}$$
$$E_{eq1} = 3A \times 12\Omega$$
$$E_{eq1} = 36V$$
$$E_{R2} = 36V$$
$$E_{R3} = 36V$$

To find power used by R_1:

Given:

$$E_{R1} = 24V$$
$$I_T = 3A$$

Solution:

$$P_{R1} = E_{R1} \times I_T$$
$$P_{R1} = 24V \times 3A$$
$$P_{R1} = 72W$$

To find the current through R_2 and R_3, refer to the original circuit, figure 3-54(A). You know E_{R2} and E_{R3} from previous calculation.

Given:

$$E_{R2} = 36V$$
$$E_{R3} = 36V$$
$$R_2 = 20\Omega$$
$$R_2 = 30\Omega$$

Solution:

$$I_{R2} = \frac{E_{R2}}{R_2} \quad \text{(Ohm's Law)}$$

$$I_{R2} = \frac{36V}{20\Omega}$$

$$I_{R2} = 1.8A$$

$$I_{R3} = \frac{E_{R3}}{R_3}$$

$$I_{R3} = \frac{36V}{30\Omega}$$

$$I_{R3} = 1.2A$$

To find power used by R_2 and R_3, using values from previous calculations:

Given:

$$E_{R2} = 36V$$
$$E_{R3} = 36V$$
$$I_{R2} = 1.8A$$
$$I_{R2} = 1.2A$$

Solution:

$$P_{R2} = E_{R2} \times I_{R2}$$
$$P_{R2} = 36V \times 1.8A$$
$$P_{R2} = 64.8W$$

$$P_{R3} = E_{R3} \times I_{R3}$$
$$P_{R3} = 36V \times 1.2A$$
$$P_{R3} = 43.2W$$

Now that you have solved for the unknown quantities in this circuit, you can apply what you have learned to any series, parallel, or combination circuit. It is important to remember to first look at the circuit and from observation make your determination of the type of circuit, what is known, and what you are looking for. A minute spent in this manner may save you many unnecessary calculations.

Having computed all the currents and voltages of figure 3-54, a complete description of the operation of the circuit can be made. The total current of 3 amps leaves the negative terminal of the battery and flows through the 8-ohm resistor (R_1). In so doing, a voltage drop of 24 volts occurs across resistor R_1. At point A, this 3-ampere current divides into two currents. Of the total current, 1.8 amps flows through the 20-ohm resistor. The remaining current of 1.2 amps flows from point A, down through the 30-ohm resistor to point B. This current produces a voltage drop of 36 volts across the 30-ohm resistor. (Notice that the voltage drops across the 20- and 30-ohm resistors are the same.) The two branch currents of 1.8 and 1.2 amps combine at junction B and the total current of 3 amps flows back to the source. The action of the circuit has been completely described with the exception of power consumed, which could be described using the values previously computed.

It should be pointed out that the combination circuit is not difficult to solve. The key to its solution lies in knowing the order in which the steps of the solution must be accomplished.

Practice Circuit Problem

Figure 3-55 is a typical combination circuit. To make sure you understand the techniques of solving for the unknown quantities, solve for E_{R1}.

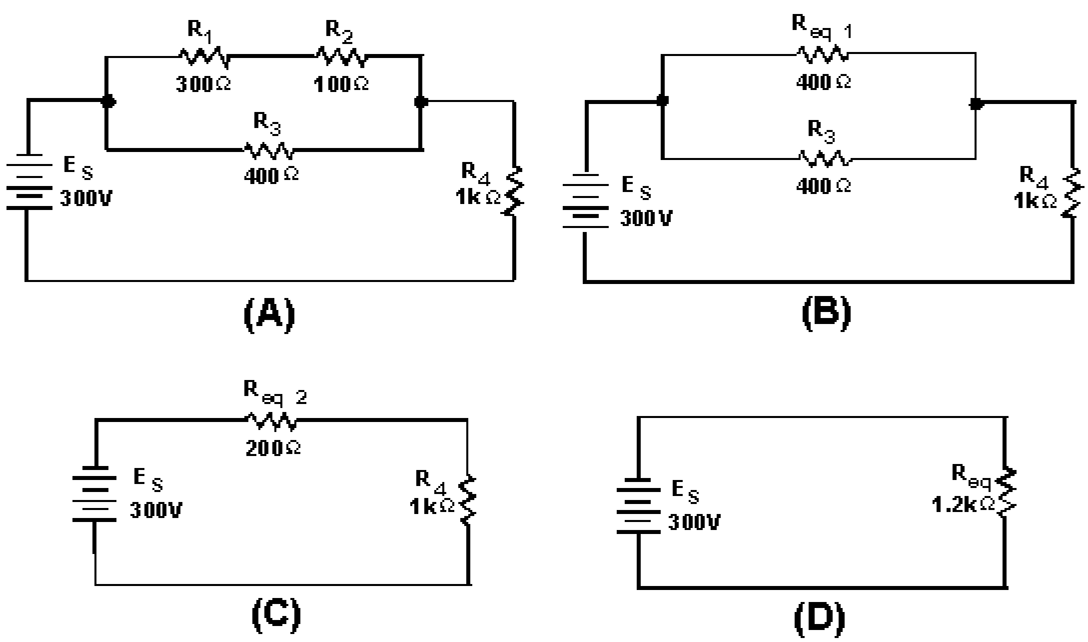


Figure 3-55.—Combination practice circuit.

It is not necessary to solve for all the values in the circuit to compute the voltage drop across resistor R_1 (E_{R1}). First look at the circuit and determine that the values given do not provide enough information to solve for E_{R1} directly.

If the current through R_1 (I_{R1}) is known, then E_{R1} can be computed by applying the formula:

$$E_{R1} = R_1 \times I_{R1}$$

The following steps will be used to solve the problem.

1. The total resistance (R_T) is calculated by the use of equivalent resistance.

Given:

$$R_1 = 300\Omega$$
$$R_2 = 100\Omega$$

Solution:

$$R_{eq1} = R_1 + R_2$$
$$R_{eq1} = 300\Omega + 100\Omega$$
$$R_{eq1} = 400\Omega$$

Redraw the circuit as shown in figure 3-55(B).

Given:

$$R_{eq1} = 400\Omega$$
$$R_3 = 400\Omega$$

Solution:

$$R_{eq2} = \frac{R}{N} \quad \text{(Equal resistors in parallel)}$$
$$R_{eq2} = \frac{400\Omega}{2}$$
$$R_{eq2} = 200\Omega$$

Solution:

$$R_{eq2} = \frac{R}{N}$$
$$R_{eq2} = \frac{400\Omega}{2}$$
$$R_{eq2} = 200\Omega$$

Redraw the circuit as shown in figure 3-55(C).

Given:

$$R_{eq2} = 200\Omega$$
$$R_4 = 1k\Omega$$

Solution:

$$R_{eq} = R_{eq2} + R_4$$
$$R_{eq} = 200\Omega + 1k\Omega$$
$$R_{eq} = 1.2k\Omega$$

2. The total current (I_T) is now computed.

Given:

$$E_S = 300V$$
$$R_{eq} = 1.2k\Omega$$

Solution:

$$I_T = \frac{E_S}{R_{eq}}$$

$$I_T = \frac{300V}{1.2k\Omega}$$

$$I_T = 250mA$$

3. Solve for the voltage dropped across R_{eq2}. This represents the voltage dropped across the network R_1, R_2, and R_3 in the original circuit.

Given:

$$R_{eq} = 200\Omega$$
$$I_T = 250mA$$

Solution:

$$E_{Req2} = R_{eq2} \times I_T$$
$$E_{Req2} = 200\Omega \times 250mA$$
$$E_{Req2} = 50V$$

4. Solve for the current through R_{eq1}. (R_{eq1} represents the network R_1 and R_2 in the original circuit.) Since the voltage across each branch of a parallel circuit is equal to the voltage across the equivalent resistor representing the circuit:

Given:

$$E_{Req2} = E_{Req1}$$
$$E_{Req1} = 50V$$
$$R_{eq1} = 400\Omega$$

Solution:

$$I_{Req1} = \frac{E_{Req1}}{R_{eq1}}$$

$$I_{Req1} = \frac{50V}{400\Omega}$$

$$I_{Req1} = 125mA$$

5. Solve for the voltage dropped across R_1 (the quantity you were asked to find). Since R_{eq1} represents the series network of R_1 and R_2 and total current flows through each resistor in a series circuit, I_{R1} must equal I_{Req1}.

Given:

$$I_{R1} = 125mA$$
$$R_1 = 300\Omega$$

Solution:

$$E_{R1} = I_{R1} \times R_1$$
$$E_{R1} = 125mA \times 300\Omega$$
$$E_{R1} = 37.5V$$

Q42. Refer to figure 3-55(A). If the following resistors were replaced with the values indicated: $R_1 = 900\Omega$, $R_3 = lk\Omega$, what is the total power in the circuit? What is E_{R2}?

REDRAWING CIRCUITS FOR CLARITY

You will notice that the schematic diagrams you have been working with have shown parallel circuits drawn as neat square figures, with each branch easily identified.

In actual practice the wired circuits and more complex schematics are rarely laid out in this simple form. For this reason, it is important for you to recognize that circuits can be drawn in a variety of ways, and to learn some of the techniques for redrawing them into their simplified form. When a circuit is redrawn for clarity or to its simplest form, the following steps are used.

1. Trace the current paths in the circuit.

2. Label the junctions in the circuit.

3. Recognize points which are at the same potential.

4. Visualize a rearrangement, "stretching" or "shrinking," of connecting wires.

5. Redraw the circuit into simpler form (through stages if necessary).

To redraw any circuit, start at the source, and trace the path of current flow through the circuit. At points where the current divides, called JUNCTIONS, parallel branches begin. These junctions are key points of reference in any circuit and should be labeled as you find them. The wires in circuit schematics are assumed to have NO RESISTANCE and there is NO VOLTAGE drop along any wire. This means that any unbroken wire is at the same voltage all along its length, until it is interrupted by a resistor, battery, or some other circuit component. In redrawing a circuit, a wire can be "stretched" or "shrunk" as much as you like without changing any electrical characteristic of the circuit.

Figure 3-56(A) is a schematic of a circuit that is not drawn in the box-like fashion used in previous illustrations. To redraw this circuit, start at the voltage source and trace the path for current to the junction marked (a). At this junction the current divides into three paths. If you were to stretch the wire to show the three current paths, the circuit would appear as shown in figure 3-56(B).

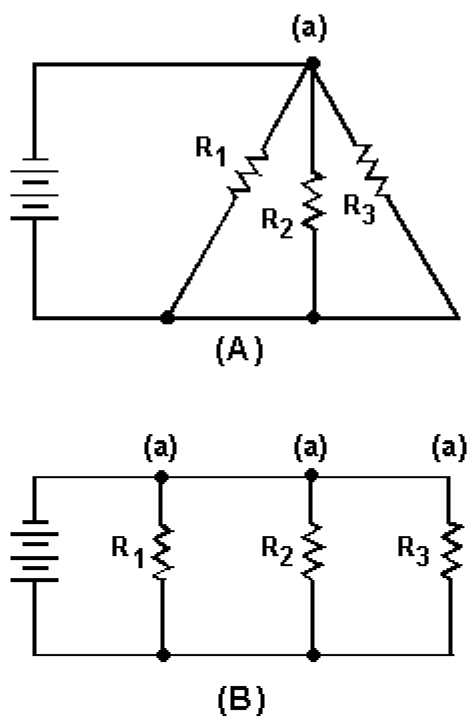


Figure 3-56.—Redrawing a simple parallel circuit.

While these circuits may appear to be different, the two drawings actually represent the same circuit. The drawing in figure 3-56(B) is the familiar box-like structure and may be easier to work with. Figure 3-57(A) is a schematic of a circuit shown in a box-like structure, but may be misleading. This circuit in reality is a series-parallel circuit that may be redrawn as shown in figure 3-57(B). The drawing in part (B) of the figure is a simpler representation of the original circuit and could be reduced to just two resistors in parallel.

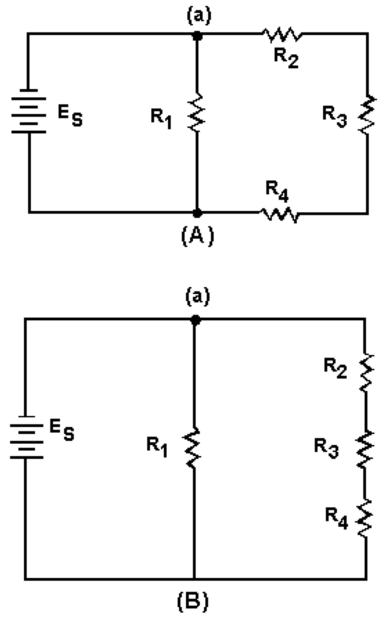


Figure 3-57.—Redrawing a simple series-parallel circuit.

Redrawing a Complex Circuit

Figure 3-58(A) shows a complex circuit that may be redrawn for clarification in the following steps.

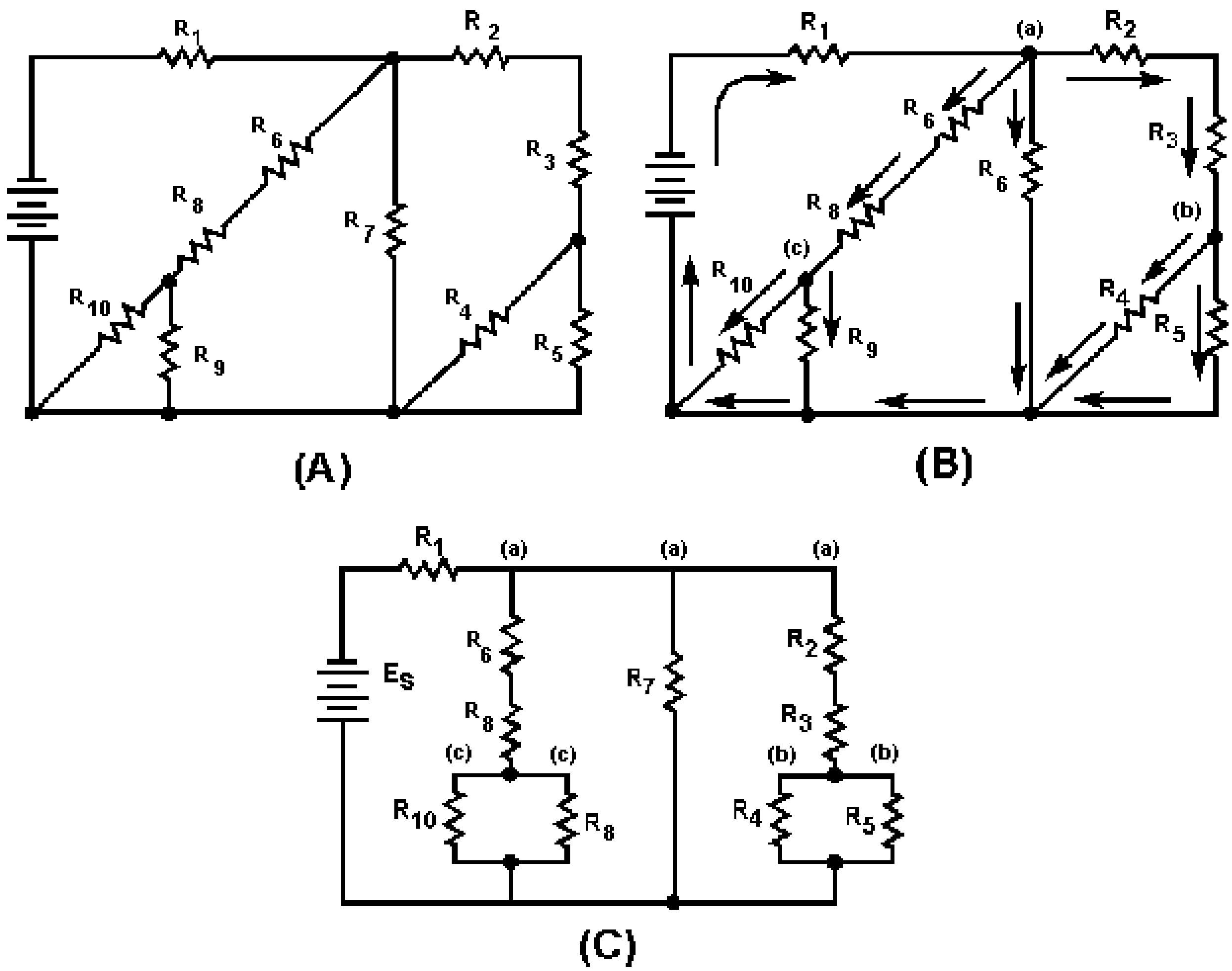


Figure 3-58.—Redrawing a complex circuit.

NOTE: As you redraw the circuit, draw it in simple box-like form. Each time you reach a junction, a new branch is created by stretching or shrinking the wires.

Start at the negative terminal of the voltage source. Current flows through R_1 to a junction and divides into three paths; label this junction (a). Follow one of the paths of current through R_2 and R_3 to a junction where the current divides into two more paths. This junction is labeled (b).

The current through one branch of this junction goes through R_5 and back to the source. (The most direct path.) Now that you have completed a path for current to the source, return to the last junction, (b). Follow current through the other branch from this junction. Current flows from junction (b) through R_4 to the source. All the paths from junction (b) have been traced. Only one path from junction (a) has been completed. You must now return to junction (a) to complete the other two paths. From junction (a) the current flows through R_7 back to the source. (There are no additional branches on this path.) Return to junction (a) to trace the third path from this junction. Current flows through R_6 and R_8 and comes to a junction. Label this junction (c). From junction (c) one path for current is through R_9 to the source. The other path for current from junction (c) is through R_{10} to the source. All the junctions in this circuit have now been labeled. The circuit and the junction can be redrawn as shown in figure 3-58(C). It is much easier to recognize the series and parallel paths in the redrawn circuit.

Q43. What is the total resistance of the circuit shown in figure 3-59? (Hint: Redraw the circuit to simplify and then use equivalent resistances to compute for R_T.)

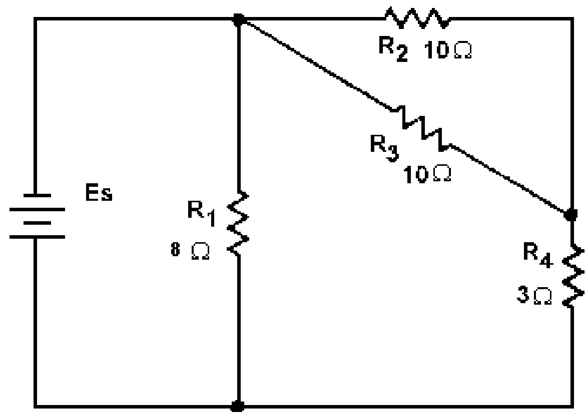


Figure 3-59.—Simplification circuit problem.

Q44. What is the total resistance of the circuit shown in figure 3-60?

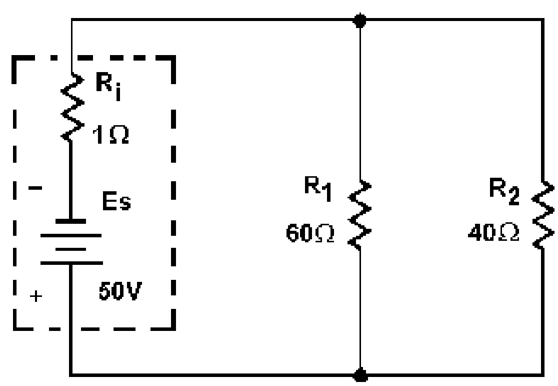


Figure 3-60.—Source resistance in a parallel circuit.

Q45. What effect does the internal resistance have on the rest of the circuit shown in figure 3-60?

EFFECTS OF OPEN AND SHORT CIRCUITS

Earlier in this chapter the terms open and short circuits were discussed. The following discussion deals with the effects on a circuit when an open or a short occurs.

The major difference between an open in a parallel circuit and an open in a series circuit is that in the parallel circuit the open would not necessarily disable the circuit. If the open condition occurs in a series portion of the circuit, there will be no current because there is no complete path for current flow. If, on the other hand, the open occurs in a parallel path, some current will still flow in the circuit. The parallel branch where the open occurs will be effectively disabled, total resistance of the circuit will INCREASE, and total current will DECREASE.

To clarify these points, figure 3-61 illustrates a series parallel circuit. First the effect of an open in the series portion of this circuit will be examined. Figure 3-61(A) shows the normal circuit, R_T = 40 ohms and I_T = 3 amps. In figure 3-61(B) an open is shown in the series portion of the circuit, there is no complete path for current and the resistance of the circuit is considered to be infinite.

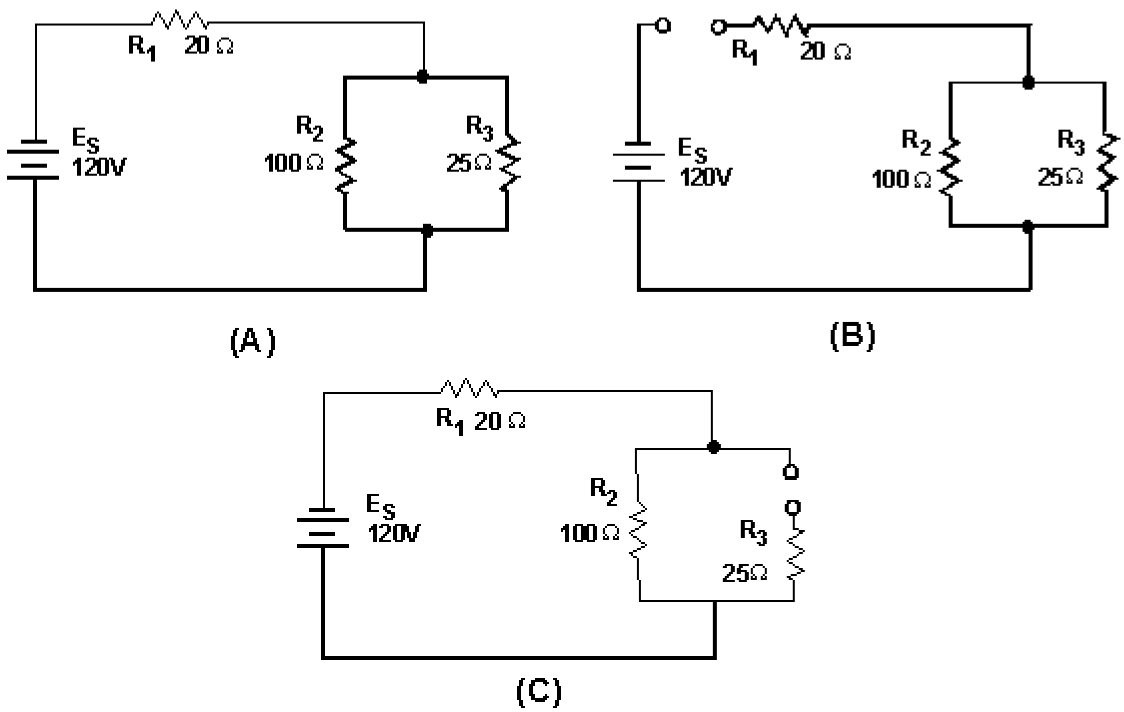


Figure 3-61.—Series-parallel circuit with opens.

In figure 3-61(C) an open is shown in the parallel branch of R_3. There is no path for current through R_3. In the circuit, current flows through R_1 and R_2 only. Since there is only one path for current flow, R_1 and R_2 are effectively in series.

Under these conditions $R_T = 120\Omega$ and I_T = 1 amp. As you can see, when an open occurs in a parallel branch, total circuit resistance increases and total circuit current decreases.

A short circuit in a parallel network has an effect similar to a short in a series circuit. In general, the short will cause an increase in current and the possibility of component damage regardless of the type of circuit involved. To illustrate this point, figure 3-62 shows a series-parallel network in which shorts are developed. In figure 3-62 (A) the normal circuit is shown. $R_T = 40$ ohms and $I_T = 3$ amps.

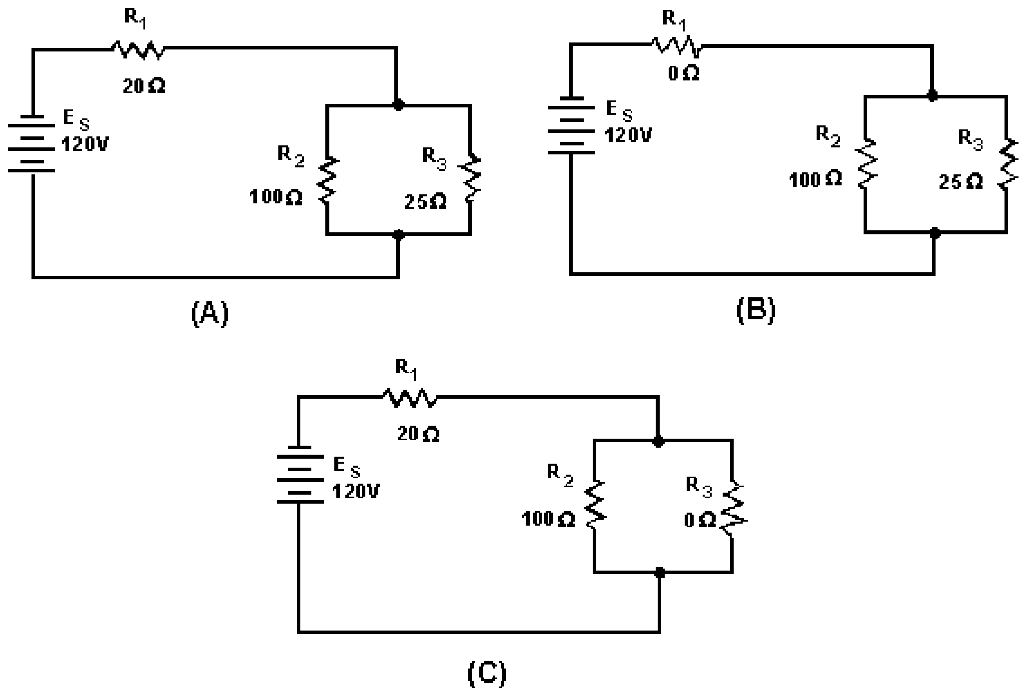


Figure 3-62.—Series-parallel circuit with shorts.

In figure 3-62 (B), R_1 has shorted. R_1 now has zero ohms of resistance. The total of the resistance of the circuit is now equal to the resistance of the parallel network of R_2 and R_3, or 20 ohms. Circuit current has increased to 6 amps. All of this current goes through the parallel network (R_2, R_3) and this increase in current would most likely damage the components.

In figure 3-62 (C), R_3 has shorted. With R_3 shorted there is a short circuit in parallel with R_2. The short circuit routes the current around R_2, effectively removing R_2 from the circuit. Total circuit resistance is now equal to the resistance of R_1, or 20 ohms.

As you know, R_2 and R_3 form a parallel network. Resistance of the network can be calculated as follows:

Given:

$$R_2 = 100\Omega$$
$$R_3 = 0\Omega$$

Solution:

$$R_{eq} = \frac{R_2 \times R_3}{R_2 + R_3}$$

$$R_{eq} = \frac{100\Omega \times 0\Omega}{100\Omega + 0\Omega}$$

$$R_{eq} = 0\Omega$$

The total circuit current with R_3 shorted is 6 amps. All of this current flows through R_1 and would most likely damage R_1. Notice that even though only one portion of the parallel network was shorted, the entire paralleled network was disabled.

Opens and shorts alike, if occurring in a circuit, result in an overall change in the equivalent resistance. This can cause undesirable effects in other parts of the circuit due to the corresponding change in the total current flow. A short usually causes components to fail in a circuit which is not properly fused or otherwise protected. The failure may take the form of a burned-out resistor, damaged source, or a fire in the circuit components and wiring.

Fuses and other circuit protection devices are installed in equipment circuits to prevent damage caused by increases in current. These circuit protection devices are designed to open if current increases to a predetermined value. Circuit protection devices are connected in series with the circuit or portion of the circuit that the device is protecting. When the circuit protection device opens, current flow ceases in the circuit.

A more thorough explanation of fuses and other circuit protection devices is presented in Module 3, *Introduction to Circuit Protection, Control, and Measurement*.

Q46. What is the effect on total resistance and total current in a circuit if an open occurs in (a) a parallel branch, and (b) in a series portion?

Q47. What is the effect on total resistance and total current in a circuit if a short occurs in (a) a parallel branch, and (b) in a series portion?

Q48. If one branch of a parallel network is shorted, what portion of circuit current flows through the remaining branches?

VOLTAGE DIVIDERS

Most electrical and electronics equipment use voltages of various levels throughout their circuitry.

One circuit may require a 90-volt supply, another a 150-volt supply, and still another a 180-volt supply. These voltage requirements could be supplied by three individual power sources. This method is expensive and requires a considerable amount of room. The most common method of supplying these voltages is to use a single voltage source and a VOLTAGE DIVIDER. Before voltage dividers are explained, a review of what was discussed earlier concerning voltage references may be of help.

As you know, some circuits are designed to supply both positive and negative voltages. Perhaps now you wonder if a negative voltage has any less potential than a positive voltage. The answer is that 100 volts is 100 volts. Whether it is negative or positive does not affect the feeling you get when you are shocked.

Voltage polarities are considered as being positive or negative in respect to a reference point, usually ground. Figure 3-63 will help to illustrate this point.

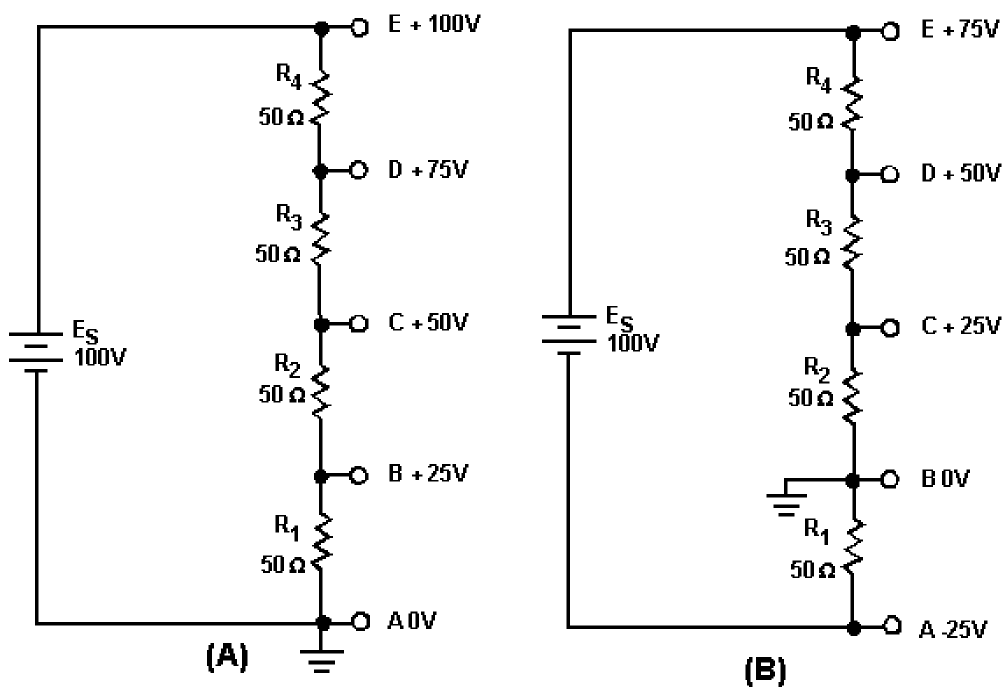


Figure 3-63.—Voltage polarities.

Figure 3-63(A) shows a series circuit with a voltage source of 100 volts and four 50-ohm resistors connected in series. The ground, or reference point, is connected to one end of resistor R_1. The current in this circuit determined by Ohm's law is .5 amp. Each resistor develops (drops) 25 volts. The five tap-off points indicated in the schematic are points at which the voltage can be measured. As indicated on the schematic, the voltage measured at each of the points from point A to point E starts at zero volts and becomes <u>more positive</u> in 25 volt steps to a value of positive 100 volts.

In figure 3-63(B), the ground, or reference point has been moved to point B. The current in the circuit is still .5 amp and each resistor still develops 25 volts. The total voltage developed in the circuit remains at 100 volts, but because the <u>reference point</u> has been changed, the voltage at point A is <u>negative</u> 25 volts. Point E, which was at positive 100 volts in figure 3-63(A), now has a voltage of positive 75 volts. As you can see the voltage at any point in the circuit is dependent on three factors; the current through the resistor, the ohmic value of the resistor, <u>and the reference point in the circuit.</u>

A typical voltage divider consists of two or more resistors connected in series across a source voltage (E_s). The source voltage must be as high or higher than any voltage developed by the voltage divider. As the source voltage is dropped in successive steps through the series resistors, any desired portion of the source voltage may be "tapped off" to supply individual voltage requirements. The values of the series resistors used in the voltage divider are determined by the voltage and current requirements of the loads.

Figure 3-64 is used to illustrate the development of a simple voltage divider. The requirement for this voltage divider is to provide a voltage of 25 volts and a current of 910 milliamps to the load from a source voltage of 100 volts. Figure 3-64(A) provides a circuit in which 25 volts is available at point B. If the load was connected between point B and ground, you might think that the load would be supplied with 25 volts. This is not true since the load connected between point B and ground forms a parallel network of the load and resistor R_1. (Remember that the value of resistance of a parallel network is always less than the value of the smallest resistor in the network.)

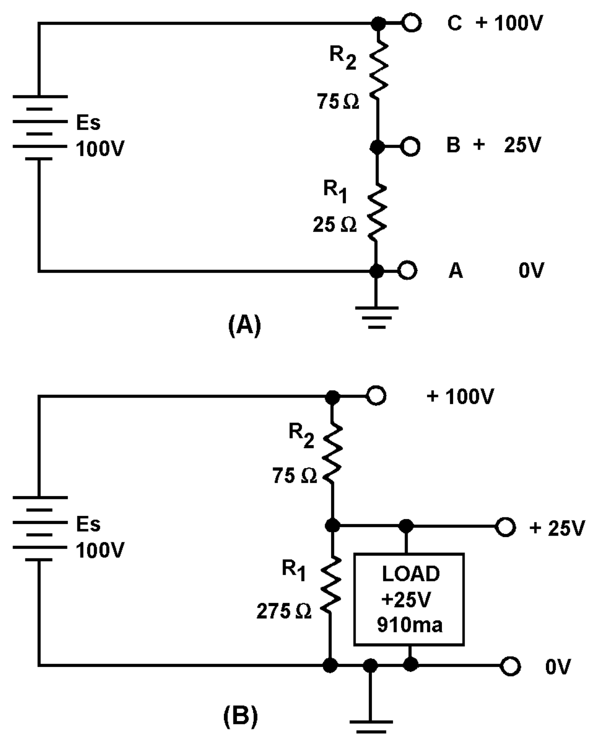


Figure 3-64.—Simple voltage divider.

Since the resistance of the network would now be less than 25 ohms, the voltage at point B would be less than 25 volts. This would not satisfy the requirement of the load.

To determine the size of resistor used in the voltage divider, a rule-of-thumb is used. The current in the divider resistor should equal approximately 10 percent of the load current. This current, which does not flow through any of the load devices, is called bleeder current.

Given this information, the voltage divider can be designed using the following steps.

1. Determine the load requirement and the available voltage source.

$$E_S = 100V$$
$$E_{load} = 25V$$
$$I_{load} = 910mA$$

2. Select bleeder current by applying the 10% rule-of-thumb.

$$I_{R1} = 10\% \times I_{load}$$
$$I_{R1} = .1 \times 910mA$$
$$I_{R1} = 91mA$$

3. Calculate bleeder resistance.

$$R_1 = \frac{E_{R1}}{I_{R1}}$$

$$R_1 = \frac{25V}{91mA}$$

$$R1 = 274.73\Omega$$

The value of R_1 may be rounded off to 275 ohms:

$$R_1 = 275\Omega$$

4. Calculate the total current (load plus bleeder).

$$I_T = I_{load} + I_{R1}$$
$$I_T = 910mA + 91mA$$
$$I_T = 1A \text{ (rounded off)}$$

5. Calculate the resistance of the other divider resistor(s).

$$E_{R2} = E_S - E_{R1}$$
$$E_{R2} = 100V - 25V$$
$$E_{R2} = 75V$$
$$R_2 = \frac{E_{R2}}{I_T}$$
$$R_2 = \frac{75V}{1A}$$
$$R_2 = 75\Omega$$

The voltage divider circuit can now be drawn as shown in figure 3-64(B).

Q49. What information must be known to determine the component values for a voltage divider?

Q50. If a voltage divider is required for a load that will use 450 mA of current, what should be the value of bleeder current?

Q51. If the load in question 50 requires a voltage of +90 V, what should be the value of the bleeder resistor?

Q52. If the source voltage for the voltage divider in question 50 supplies 150 volts, what is the total current through the voltage divider?

MULTIPLE-LOAD VOLTAGE DIVIDERS

A multiple-load voltage divider is shown in figure 3-65. An important point that was not emphasized before is that when using the 10% rule-of-thumb to calculate the bleeder current, you must take 10% of the total load current.

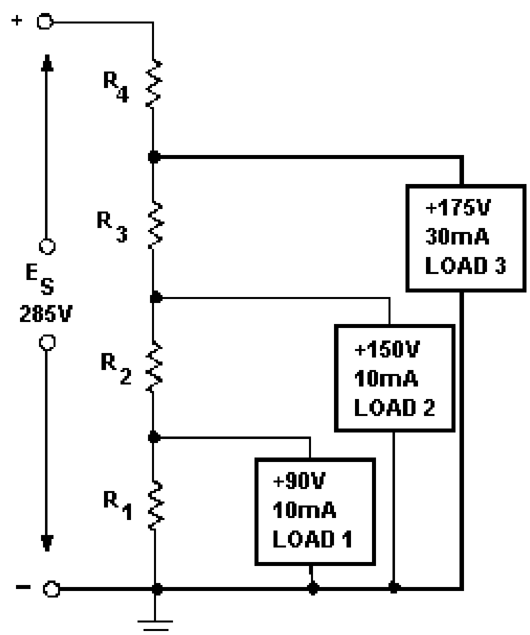


Figure 3-65.—Multiple-load voltage divider.

Given the information shown in figure 3-65, you can calculate the values for the resistors needed in the voltage-divider circuits. The same steps will be followed as in the previous voltage divider problem.

Given:

$$\text{Load 1: } E = 90V$$
$$I = 10mA$$

$$\text{Load 2: } E = 150V$$
$$I = 10mA$$

$$\text{Load 3: } E = 175V$$
$$I = 30mA$$

$$E_S = 285V$$

The bleeder current should be 10% of the total load current.

Solution:

$$I_{R1} = 10\% \times I(\text{load total})$$
$$I_{R1} = 10\% \times (10mA + 10mA + 30mA)$$
$$I_{R1} = 5mA$$

Since the voltage across R_1 (E_{R1}) is equal to the voltage requirement for load 1, Ohm's law can be used to calculate the value for R_1.

Solution:

$$R_1 = \frac{E_{R1}}{I_{R1}}$$
$$R_1 = \frac{90V}{5mA}$$
$$R_1 = 18k\Omega$$

The current through R_2 (I_{R2}) is equal to the current through R_1 plus the current through load 1.

Solution:

$$I_{R2} = I_{R1} + I_{load1}$$
$$I_{R2} = 5mA + 10mA$$
$$I_{R2} = 15mA$$

The voltage across R_2 (E_{R2}) is equal to the <u>difference</u> between the voltage requirements of load 1 and load 2.

$$E_{R2} = E_{load2} - E_{load1}$$
$$E_{R2} = 150V - 90V$$
$$E_{R2} = 60V$$

Ohm's law can now be used to solve for the value of R_2.

Solution:

$$R_2 = \frac{E_{R2}}{I_{R2}}$$
$$R_2 = \frac{60V}{15mA}$$
$$R_2 = 4k\Omega$$

The current through R_3 (I_{R3}) is equal to the current through R_2 plus the current through load 2.

$$I_{R3} = I_{R2} + I_{load2}$$
$$I_{R3} = 15mA + 10mA$$
$$I_{R3} = 25mA$$

The voltage across R_3 (E_{R3}) equals the difference between the voltage requirement of load 3 and load 2.

$$E_{R3} = E_{load3} - E_{load2}$$
$$E_{R3} = 175V - 150V$$
$$E_{R3} = 25V$$

Ohm's law can now be used to solve for the value of R_3.

Solution:

$$R_3 = \frac{E_{R3}}{I_{R3}}$$
$$R_3 = \frac{25V}{25mA}$$
$$R_3 = 1k\Omega$$

The current through R_4 (I_{R4}) is equal to the current through R_3 plus the current through load 3. I_{R4} is equal to total circuit current (I_T).

$$I_{R4} = I_{R3} + I_{load3}$$
$$I_{R4} = 25mA + 30mA$$
$$I_{R4} = 55mA$$

The voltage across R_4 (E_{R4}) equals the difference between the source voltage and the voltage requirement of load 3.

$$E_{R4} = E_S - E_{load3}$$
$$E_{R4} = 285V - 175V$$
$$E_{R4} = 110V$$

Ohm's law can now be used to solve for the value of R_4.

Solution:

$$R_4 = \frac{E_{R4}}{I_{R4}}$$
$$R_4 = \frac{110V}{55mA}$$
$$R_4 = 2k\Omega$$

With the calculations just explained, the values of the resistors used in the voltage divider are as follows:

$$R_1 = 18k\Omega$$
$$R_2 = 4k\Omega$$
$$R_3 = 1k\Omega$$
$$R_4 = 2k\Omega$$

POWER IN THE VOLTAGE DIVIDER

Power in the voltage divider is an extremely important quantity. The power dissipated by the resistors in the voltage divider should be calculated to determine the power handling requirements of the resistors. Total power of the circuit is needed to determine the power requirement of the source.

The power for the circuit shown in figure 3-65 is calculated as follows:

Given:

$$E_{R1} = 90V$$
$$I_{R1} = 5mA$$

Solution:

$$P_{R1} = E_{R1} \times I_{R1}$$
$$P_{R1} = 90V \times 5mA$$
$$P_{R1} = .45W$$

The power in each resistor is calculated just as for R_1. When the calculations are performed, the following results are obtained:

$$P_{R2} = .9W$$
$$P_{R3} = .625W$$
$$P_{R4} = 6.05W$$

To calculate the power for load 1:

Given:

$$E_{load1} = 90V$$
$$I_{load1} = 10mA$$

Solution:

$$P_{load1} = E_{load1} \times I_{load1}$$
$$P_{load1} = 90V \times 10mA$$
$$P_{load1} = .9W$$

The power in each load is calculated just as for load 1. When the calculations are performed, the following results are obtained.

$$P_{load2} = 1.5W$$
$$P_{load3} = 5.25W$$

Total power is calculated by summing the power consumed by the loads and the power dissipated by the divider resistors. The total power in the circuit is 15.675 watts.

The power used by the loads and divider resistors is supplied by the source. This applies to all electrical circuits; power for all components is supplied by the source.

Since power is the product of voltage and current, the power supplied by the source is equal to the source voltage multiplied by the total circuit current (E_s x I_T).

In the circuit of figure 3-65, the total power can be calculated by:

Given:

$$E_S = 285V$$
$$I_T = 55mA\ (I_{R4})$$

Solution:

$$P_T = E_S \times I_T$$
$$P_T = 285V \times 55mA$$
$$P_T = 15.675W$$

VOLTAGE DIVIDER WITH POSITIVE AND NEGATIVE VOLTAGE REQUIREMENTS

In many cases the load for a voltage divider requires both positive and negative voltages. Positive and negative voltages can be supplied from a single source voltage by connecting the ground (reference point) between two of the divider resistors. The exact point in the circuit at which the reference point is placed depends upon the voltages required by the loads.

For example, a voltage divider can be designed to provide the voltage and current to three loads from a given source voltage.

Given:

$$\begin{aligned} &\text{Load 1:} && E = -25V \\ & && I = 300mA \\ &\text{Load 2:} && E = +50V \\ & && I = 50mA \\ &\text{Load 3:} && E = +250V \\ & && I = 100mA \end{aligned}$$

$$E_S = 310V$$

The circuit is drawn as shown in figure 3-66. Notice the placement of the ground reference point. The values for resistors R_1, R_3, and R_4 are computed exactly as was done in the last example. I_{R1} is the bleeder current and can be calculated as follows:

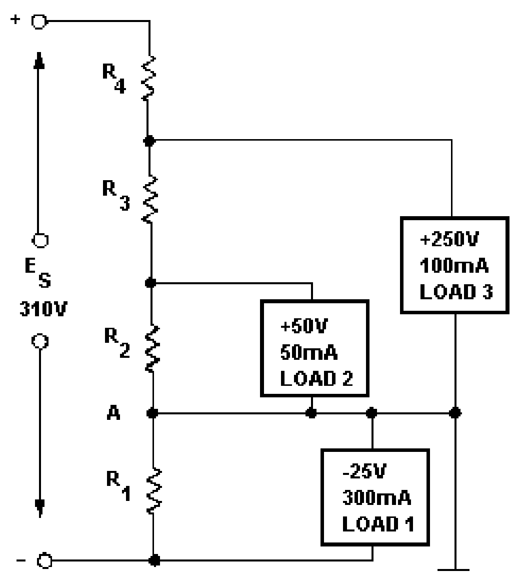


Figure 3-66.—Voltage divider providing both positive and negative voltages.

Solution:

$$\begin{aligned} I_{R1} &= 10\% \times I(\text{load total}) \\ I_{R1} &= 10\% \times (300mA) \\ I_{R1} &= 30mA \end{aligned}$$

Calculate the value of R_1.

Solution:

$$R_1 = \frac{E_{R1}}{I_{R1}}$$

$$R_1 = \frac{25V}{45mA}$$

$$R_1 = 556\Omega$$

Calculate the current through R_2 using Kirchhoff's current law.

At point A:

$$I_{R1} + I_{load1} + I_{R2} + I_{load2} + I_{load3} = 0$$
$$45mA + 300mA + I_{R2} - 50mA - 100mA = 0$$
$$345mA + I_{R2} - 150mA = 0$$
$$195mA + I_{R2} = 0$$
$$I_{R2} = -195mA$$

(or 195mA leaving point A)

Since E_{R2} = E load 2, you can calculate the value of R_2.

Solution:

$$R_2 = \frac{E_{R2}}{I_{R2}}$$

$$R_2 = \frac{50V}{195mA}$$

$$R_2 = 256\Omega$$

Calculate the current through R_3.

$$I_{R3} = I_{R2} + I_{load2}$$
$$I_{R3} = 195mA + 50mA$$
$$I_{R3} = 245mA$$

The voltage across R_3 (E_{R3}) equals the difference between the voltage requirements of loads 3 and 2.

Solution:

$$E_{R3} = E_{load3} - E_{load2}$$
$$E_{R3} = 250V - 50V$$
$$E_{R3} = 200V$$

Calculate the value of R_3.

Solution:

$$R_3 = \frac{E_{R3}}{I_{R3}}$$

$$R_3 = \frac{200V}{245mA}$$

$$R_3 = 816\Omega$$

Calculate the current through R_4.

$$I_{R4} = I_{R3} + I_{load3}$$
$$I_{R4} = 245mA + 100mA$$
$$I_{R4} = 345mA$$

The voltage across E_{R4} equals the source voltage (E_s) minus the voltage requirement of load 3 and the voltage requirement of load 1. Remember Kirchhoff's voltage law which states that the sum of the voltage drops and emfs around any closed loop is equal to zero.

Solution:

$$E_{R4} = E_S - E_{load3} - E_{load1}$$
$$E_{R4} = 310V - 250V - 25v$$
$$E_{R4} = 35V$$

Calculate the value of R_4.

Solution:

$$R_4 = \frac{E_{R4}}{I_{R4}}$$

$$R_4 = \frac{35V}{345mA}$$

$$R_4 = 101.4\Omega$$

With the calculations just explained, the values of the resistors used in the voltage/divider are as follows:

$$R_1 = 556\Omega$$
$$R_2 = 256\Omega$$
$$R_3 = 816\Omega$$
$$R_4 = 101\Omega$$

From the information just calculated, any other circuit quantity, such as power, total current, or resistance of the load, could be calculated.

PRACTICAL APPLICATION OF VOLTAGE DIVIDERS

In actual practice the computed value of the bleeder resistor does not always come out to an even value. Since the rule-of-thumb for bleeder current is only an estimated value, the bleeder resistor can be of a value close to the computed value. (If the computed value of the resistance were 510 ohms, a 500-ohm resistor could be used.) Once the actual value of the bleeder resistor is selected, the bleeder current must be recomputed. The voltage developed by the bleeder resistor must be equal to the voltage requirement of the load in parallel with the bleeder resistor.

The value of the remaining resistors in the voltage divider is computed from the current through the remaining resistors and the voltage across them. These values must be used to provide the required voltage and current to the loads.

If the computed values for the divider resistors are not even values; series, parallel, or series-parallel networks can be used to provide the required resistance.

Example: A voltage divider is required to supply two loads from a 190.5 volts source. Load 1 requires +45 volts and 210 milliamps; load 2 requires +165 volts and 100 milliamps.

Calculate the bleeder current using the rule-of-thumb.

Given:

$$I_{load1} = 210mA$$
$$I_{load2} = 100mA$$

Solution:

$$I_{R1} = 10\% \times (210mA + 100mA)$$
$$I_{R1} = 31mA$$

Calculate the ohmic value of the bleeder resistor.

Given:

$$E_{R1} = 45V \quad (E_{load1})$$
$$I_{R1} = 31mA$$

Solution:

$$R_1 = \frac{E_{R1}}{I_{R1}}$$
$$R_1 = \frac{45V}{31mA}$$
$$R_1 = 1451.6\Omega$$

Since it would be difficult to find a resistor of 1451.6 ohms, a practical choice for R_1 is 1500 ohms.

Calculate the actual bleeder current using the selected value for R_1.

Given:

$$E_{R1} = 45V$$
$$R_1 = 1.5k\Omega$$

Solution:

$$I_{R1} = \frac{E_{R1}}{R_1}$$

$$I_{R1} = \frac{45V}{1.5k\Omega}$$

$$I_{R1} = 30mA$$

Using this value for I_{R1}, calculate the resistance needed for the next divider resistor. The current (I_{R2}) is equal to the bleeder current plus the current used by load 1.

Given:

$$I_{R1} = 30mA$$
$$I_{load1} = 210mA$$

Solution:

$$I_{R2} = I_{R1} + I_{load1}$$
$$I_{R2} = 30mA + 210mA$$
$$I_{R4} = 240mA$$

The voltage across R_2 (E_{R2}) is equal to the difference between the voltage requirements of loads 2 and 1, or 120 volts.

Calculate the value of R_2.

Given:

$$E_{R2} = 120V$$
$$I_{R2} = 240mA$$

Solution:

$$R_2 = \frac{E_{R2}}{I_{R2}}$$

$$R_2 = \frac{120V}{240mA}$$

$$R_2 = 500\Omega$$

The value of the final divider resistor is calculated with I_{R3} (I_{R2} + I load 2) equal to 340 mA and E_{R3} (E_s – E load 2) equal to 25.5V.

Given:

$$E_{R3} = 25.5V$$
$$I_{R3} = 340mA$$

Solution:

$$R_3 = \frac{E_{R3}}{I_{R3}}$$

$$R_3 = \frac{25.5V}{340mA}$$

$$R_3 = 75\Omega$$

A 75-ohm resistor may not be easily obtainable, so a network of resistors equal to 75 ohms can be used in place of R_3.

Any combination of resistor values adding up to 75 ohms could be placed in series to develop the required network. For example, if you had two 37.5-ohm resistors, you could connect them in series to get a network of 75 ohms. One 50-ohm and one 25-ohm resistor or seven 10-ohm and one 5-ohm resistor could also be used.

A parallel network could be constructed from two 150-ohm resistors or three 225-ohm resistors. Either of these parallel networks would also be a network of 75 ohms.

The network used in this example will be a series-parallel network using three 50-ohm resistors.

With the information given, you should be able to draw this voltage divider network.

Once the values for the various divider resistors have been selected, you can compute the power used by each resistor using the methods previously explained. When the power used by each resistor is known, the wattage rating required of each resistor determines the physical size and type needed for the circuit. This circuit is shown in figure 3-67.

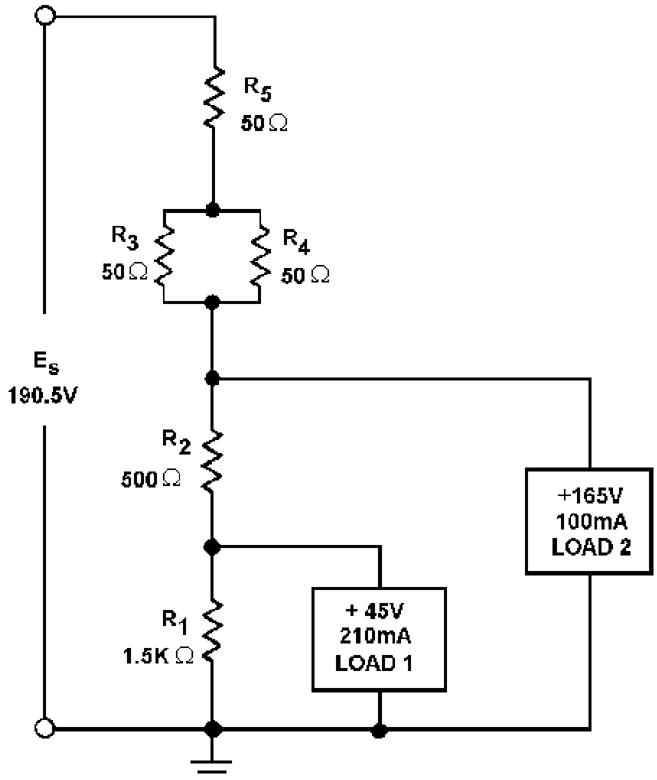


Figure 3-67.—Practical example of a voltage divider.

Q53. In figure 3-67, why is the value of R_1 calculated first?

Q54. In figure 3-67, how is (a) the current through R_2 and (b) the voltage drop across R_2 computed?

Q55. In figure 3-67, what is the power dissipated in R_1?

Q56. In figure 3-67, what is the purpose of the series-parallel network R_3, R_4, and R_5?

Q57. In figure 3-67, what should be the minimum wattage ratings of R_3 and R_5?

Q58. If the load requirement consists of both positive and negative voltages, what technique is used in the voltage divider to supply the loads from a single voltage source?

EQUIVALENT CIRCUIT TECHNIQUES

The circuit solutions that you have studied up to this point have been obtained mainly through the use of formulas derived from Ohm's law. As in many other fields of science, electricity has its share of special shortcut methods. Some of the special circuit analysis techniques are: THEVENIN'S THEOREM, which uses a process of circuit reduction to Thevenin's equivalent circuit; and NORTON'S THEOREM, which is reduction of a circuit to Norton's equivalent. Another method is called LOOP ANALYSIS. This uses Kirchhoff's voltage law to simultaneously solve problems in parallel branches of a circuit. The use of these methods should be reserved until you have become thoroughly familiar with the methods covered thus far in this chapter. You may want to explore some of the special techniques later in your career.

ELECTRICAL SAFETY

Safety precautions must always be observed by persons working around electric circuits and equipment to avoid injury from electric shock. Detailed safety precautions are contained in NAVMAT P-5l00, *Safety Precautions for Shore Activities* and OPNAVINST 5l00-19, *Navy Safety Precautions for Forces Afloat*.

The danger of shock from a 450-volt ac electrical service system is well recognized by operating personnel. This is shown by the relatively low number of reports of serious shock received from this voltage, despite its widespread use. On the other hand, a number of fatalities have been reported due to contact with low-voltage circuits. Despite a fairly widespread, but totally unfounded, popular belief to the contrary, low-voltage circuits (115 volts and below) are very dangerous and can cause death when the resistance of the body is lowered. Fundamentally, current, rather than voltage, is the measure of shock intensity. The passage of even a very small current through a vital part of the human body can cause DEATH. The voltage necessary to produce the fatal current is dependent upon the resistance of the body, contact conditions, the path through the body, etc. For example, when a 60-hertz alternating current, is passed through a human body from hand to hand or from hand to foot, and the current is gradually increased, it will cause the following effects: At about 1 milliampere (0.001 ampere), the shock can be felt; at about 10 milliamperes (0.01 ampere), the shock is of sufficient intensity to prevent voluntary control of the muscles; and at about 100 milliamperes (0.1 ampere) the shock is fatal if it lasts for 1 second or more. The above figures are the results of numerous investigations and are approximate because individuals differ in their resistance to electrical shock. It is most important to recognize that the resistance of the human body cannot be relied upon to prevent a fatal shock from 115 volts or less— FATALITIES FROM VOLTAGES AS LOW AS 30 VOLTS HAVE BEEN RECORDED. Tests have shown that body resistance under unfavorable conditions may be as low as 300 ohms, and possibly as low as 100 ohms from temple to temple if the skin is broken.

Conditions aboard ship add to the chance of receiving an electrical shock. Aboard ship the body is likely to be in contact with the metal structure of the ship and the body resistance may be low because of perspiration or damp clothing. Extra care and awareness of electrical hazards aboard ship are needed.

Short circuits can be caused by accidentally placing or dropping a metal tool, rule, flashlight case, or other conducting article across an energized line. The arc and fire which result, even on relatively low-voltage circuits, may cause extensive damage to equipment and serious injury to personnel.

Since ship service power distribution systems are designed to be ungrounded, many persons believe it is safe to touch one conductor, since no electrical current would flow. This is not true, since the distribution system is not totally isolated from the hull of the ship. If one conductor of an ungrounded electrical system is touched while the body is in contact with the hull of the ship or other metal equipment enclosure, a fatal electric current may pass through the body. ALL LIVE ELECTRIC CIRCUITS SHALL BE TREATED AS POTENTIAL HAZARDS AT ALL TIMES.

DANGER SIGNALS

Personnel should constantly be on the alert for any signs which might indicate a malfunction of electric equipment. Besides the more obvious visual signs, the reaction of other senses, such as hearing, smell, and touch, should also make you aware of possible electrical malfunctions. Examples of signs which you must be alert for are: fire, smoke, sparks, arcing, or an unusual sound from an electric motor.

Frayed and damaged cords or plugs; receptacles, plugs, and cords which feel warm to the touch; slight shocks felt when handling electrical equipment; unusually hot running electric motors and other electrical equipment; an odor of burning or overheated insulation; electrical equipment which either fails to operate or operates irregularly; and electrical equipment which produces excessive vibrations are also indications of malfunctions. When any of the above signs are noted, they are to be reported immediately to a qualified technician. DO NOT DELAY. Do not operate faulty equipment. Above all, do not attempt to make any repairs yourself if you are not qualified to do so. Stand clear of any suspected hazard and instruct others to do likewise.

- Warning Signs—They have been placed for your protection. To disregard them is to invite personal injury as well as possible damage to equipment. Switches and receptacles with a temporary warning tag, indicating work is being performed, are not to be touched.

- Working Near Electrical Equipment—When work must be performed in the immediate vicinity of electrical equipment, check with the technician responsible for the maintenance of the equipment so you can avoid any potential hazards of which you may not be immediately aware.

- Authorized Personnel Only—Because of the danger of fire, damage to equipment, and injury to personnel, all repair and maintenance work on electrical equipment shall be done only by authorized persons. Keep your hands off of all equipment which you have not been specifically authorized to handle. Particularly stay clear of electrical equipment opened for inspection, testing, or servicing.

- Circuit Breakers and Fuses—Covers for all fuse boxes, junction boxes, switch boxes, and wiring accessories should be kept closed. Any cover which is not closed or is missing should be reported to the technician responsible for its maintenance. Failure to do so may result in injury to personnel or damage to equipment in the event accidental contact is made with exposed live circuits.

ELECTRICAL FIRES

Carbon dioxide (CO_2) is used in fighting electrical fires. It is nonconductive and, therefore, the safest to use in terms of electrical safety. It also offers the least likelihood of damaging equipment. However, if the discharge horn of a CO_2 extinguisher is allowed to accidentally touch an energized circuit, the horn may transmit a shock to the person handling the extinguisher.

The very qualities which cause CO_2 to be a valuable extinguishing agent also make it dangerous to life. When it replaces oxygen in the air to the extent that combustion cannot be sustained, respiration also cannot be sustained. Exposure of a person to an atmosphere of high concentration of CO_2 will cause suffocation.

FIRST AID FOR ELECTRIC SHOCK

A person who has stopped breathing is not necessarily dead, but is in immediate danger. Life is dependent upon oxygen, which is breathed into the lungs and then carried by the blood to every body cell. Since body cells cannot store oxygen, and since the blood can hold only a limited amount (and that only for a short time), death will surely result from continued lack of breathing.

However, the heart may continue to beat for some time after breathing has stopped, and the blood may still be circulated to the body cells. Since the blood will, for a short time, contain a small supply of oxygen, the body cells will not die immediately. For a very few minutes, there is some chance that the person's life may be saved.

The process by which a person who has stopped breathing can be saved is called ARTIFICIAL VENTILATION (RESPIRATION).

The purpose of artificial ventilation is to force air out of the lungs and into the lungs, in rhythmic alternation, until natural breathing is reestablished. Artificial ventilation should be given only when natural breathing has stopped; it should NOT be given to any person who is breathing naturally. You should not assume that an individual who is unconscious due to electrical shock has stopped breathing. To tell if someone suffering from an electrical shock is breathing, place your hands on the person's sides, at the level of the lowest ribs. If the victim is breathing, you will usually be able to feel the movement. Remember: DO NOT GIVE ARTIFICIAL VENTILATION TO A PERSON WHO IS BREATHING NATURALLY.

Records show that seven out of ten victims of electric shock were revived when artificial respiration was started in <u>less than 3 minutes</u>. After 3 minutes, the chances of revival decrease rapidly.

Once it has been determined that breathing has stopped, the person nearest the victim should start the artificial ventilation without delay and send others for assistance and medical aid. The only logical, permissible delay is that required to free the victim from contact with the electricity in the quickest, safest way. This step, while it must be taken quickly, must be done with great care; otherwise, there may be two victims instead of one. In the case of portable electric tools, lights, appliances, equipment, or portable outlet extensions, this should be done by turning off the supply switch or by removing the plug from its receptacle. If the switch or receptacle cannot be quickly located, the suspected electrical device may be pulled free of the victim. Other persons arriving on the scene must be clearly warned not to touch the suspected equipment until it is deenergized. Aid should be enlisted to unplug the device as soon as possible. The injured person should be pulled free of contact with stationary equipment (such as a bus bar) if the equipment cannot be quickly deenergized, or if considerations of military operation or unit survival prevent immediate shutdown of the circuits.

This can be done quickly and safely by carefully applying the following procedures:

1. Protect yourself with dry insulating material.

2. Use a dry board, belt, clothing, or other available nonconductive material to free the victim from electrical contact. DO NOT TOUCH THE VICTIM UNTIL THE SOURCE OF ELECTRICITY HAS BEEN REMOVED.

Once the victim has been removed from the electrical source, it should be determined, if the person is breathing. If the person is not breathing, a method of artificial ventilation is used.

Sometimes victims of electrical shock suffer cardiac arrest (heart stoppage) as well as loss of breathing. Artificial ventilation alone is not enough in cases where the heart has stopped. A technique known as Cardiopulmonary Resuscitation (CPR) has been developed to provide aid to a person who has stopped breathing and suffered a cardiac arrest. Because you most likely will be working in the field of electricity, the risk of electrical shock is higher than most other Navy occupations. You should, at your earliest opportunity, learn the technique of CPR.

CPR is relatively easy to learn and is taught in courses available from the American Red Cross, some Navy Medical Departments, and in the *Standard First Aid Training Course* (NAVEDTRA 12081).

Q59. Is it considered safe for a person to touch any energized low-voltage conductor with the bare hand?

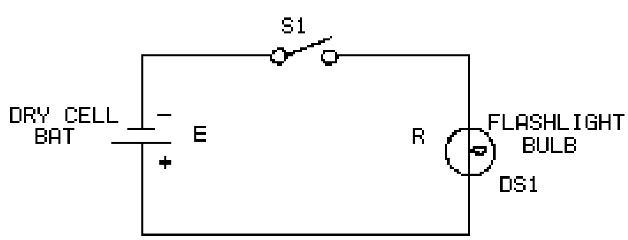
Q60. What should you do if you become aware of a possible malfunction in a piece of electrical equipment?

Q61. Who should perform CPR?

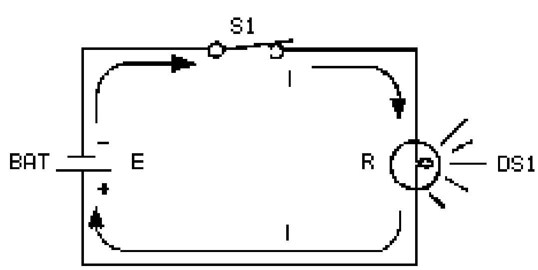
SUMMARY

With the completion of this chapter you have gained a basic understanding of dc circuits. The information you have learned will provide you with a firm foundation for continuing your study of electricity. The following is a summary of the important points in the chapter.

A BASIC ELECTRIC CIRCUIT consists of a source of electrical energy connected to a load. The load uses the energy and changes it to a useful form.

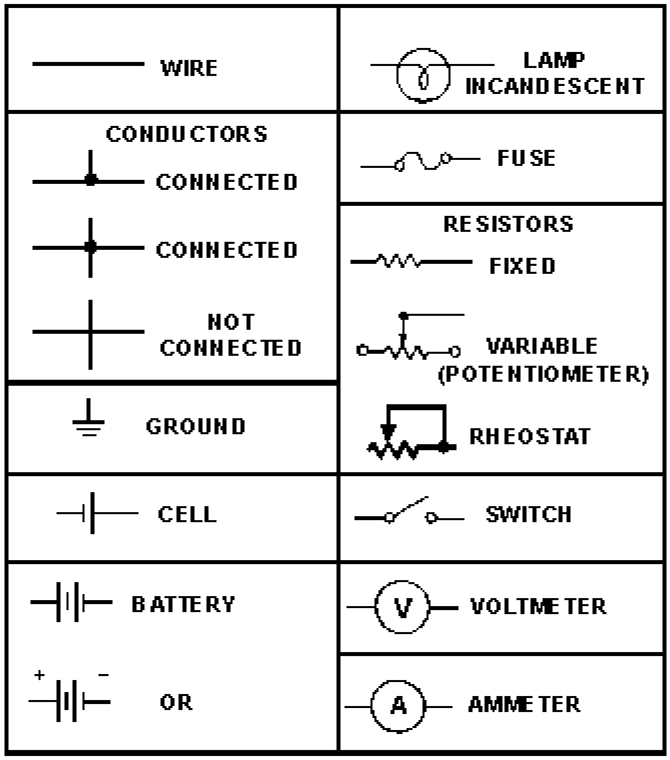


(A) DEENERGIZED



(B) ENERGIZED

A SCHEMATIC DIAGRAM is a "picture" of a circuit, which uses symbols to represent components. The space required to depict an electrical or electronic circuit is greatly reduced by the use of a schematic.



VOLTAGE (E) is the electrical force or pressure operating in a circuit.

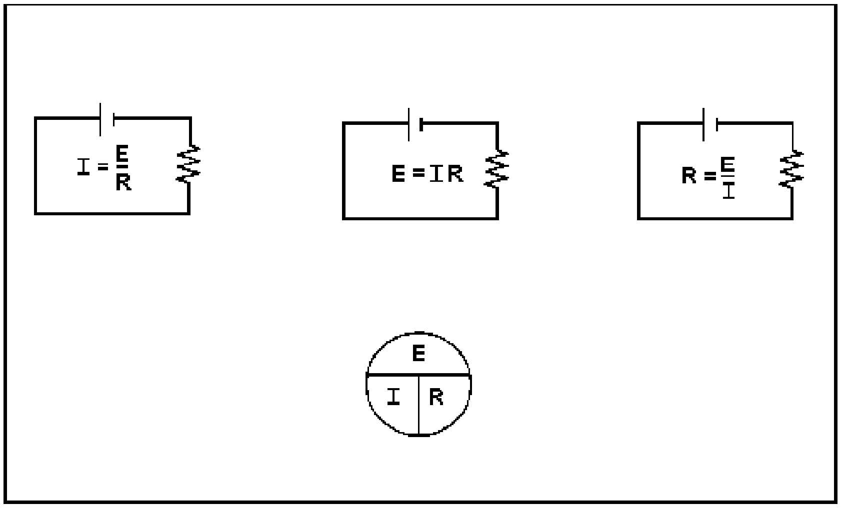
AN AMPERE (A) represents the current flow produced by one volt working across one ohm of resistance.

RESISTANCE (R) is the opposition to current. It is measured in ohms (Ω). One ohm of resistance will limit the current produced by one volt to a level of one ampere.

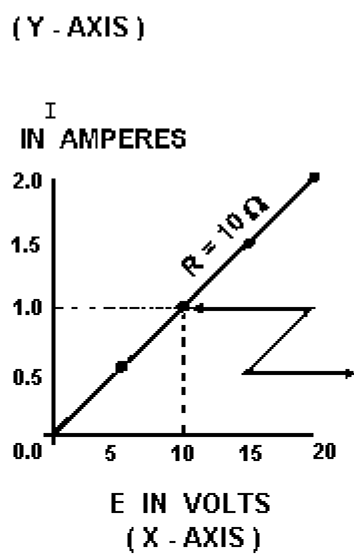
THE OHM'S FORMULA can be transposed to find one of the values in a circuit if the other two values are known. You can transpose the Ohm's law formula

$$I = \frac{E}{R}$$

mathematically, or you can use the Ohm's law figure to determine the mathematical relationship between R, E, and I.



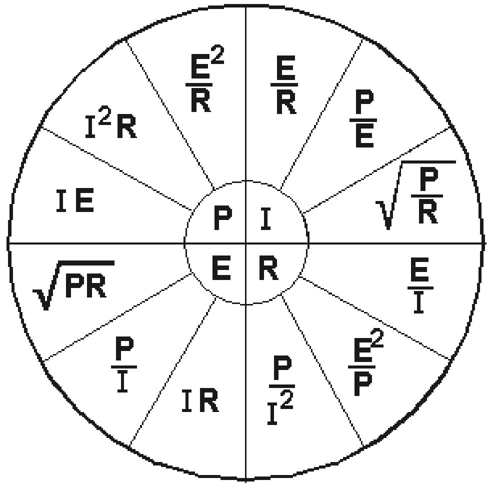
GRAPHICAL ANALYSIS of the relationship between R, E, and I can be studied by plotting these quantities on a graph. Such a graph is useful for observing the characteristics of an electrical device.



R = 10 Ω	
E	I
0	0.0
5	0.5
10	1.0
15	1.5
20	2.0

POWER is the rate of doing work per unit of time. The time required to perform a given amount of work will determine the power expended. As a formula, P = E x I, where P = power in watts, E = voltage in volts, and I = current in amperes.

THE FOUR BASIC ELECTRICAL QUANTITIES are P, I, E, R. Any single unknown quantity can be expressed in terms of any two of the other known quantities. The formula wheel is a simple representation of these relationships.



POWER RATING in watts indicates the rate at which a device converts electrical energy into another form of energy. The power rating of a resistor indicates the maximum power the resistor can withstand without being destroyed.

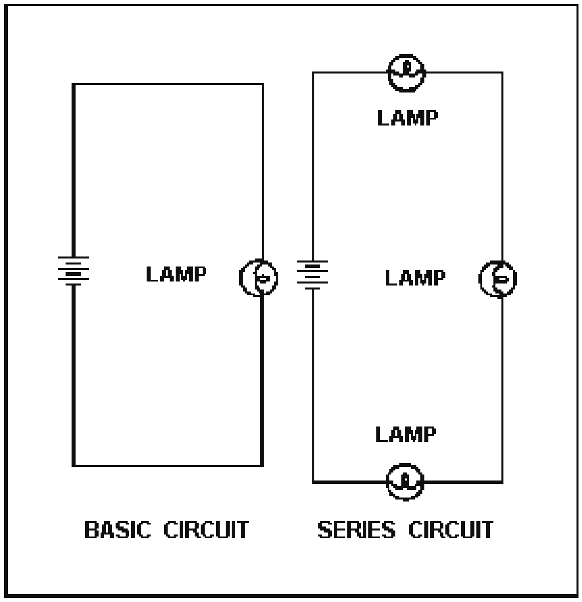
POWER USED by an electrical device is measured in watt-hours. One watt-hour is equal to one watt used continuously for one hour.

THE EFFICIENCY of an electrical device is equal to the electrical power converted into useful energy divided by the electrical power supplied to the device.

$$\text{EFF} = \frac{\text{Power converted}}{\text{Power used}}$$

HORSEPOWER is a unit of measurement often used to rate electrical motors. It is a unit of work. One horsepower is equal to 746 watts.

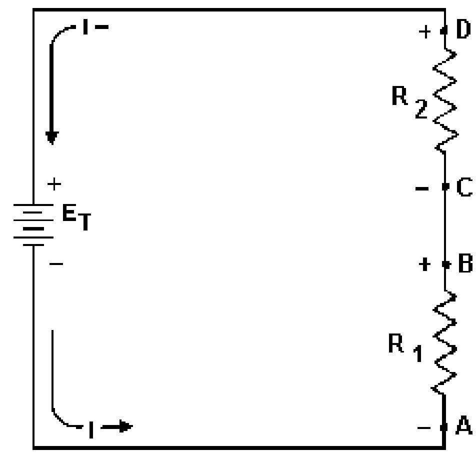
A SERIES CIRCUIT is defined as a circuit that has only one path for current flow.



RULES FOR SERIES DC CIRCUITS:

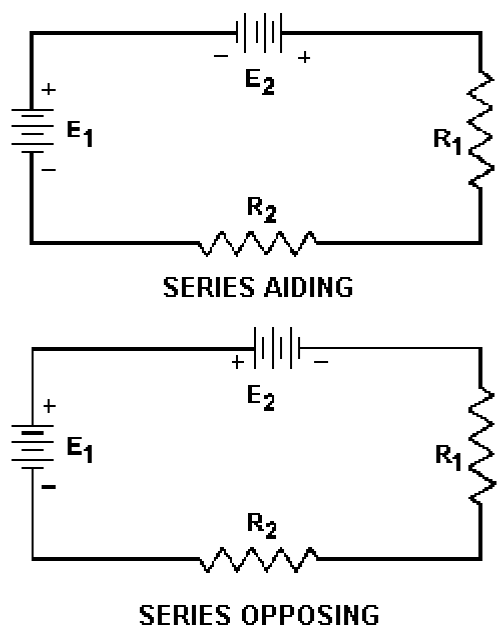
- The same current flows through each part of a series circuit.
- The total resistance of a series circuit is equal to the sum of the individual resistances.
- The total voltage across a series circuit is equal to the sum of the individual voltage drops.
- The voltage drop across a resistor in a series circuit is proportional to the ohmic value of the resistor.
- The total power in a series circuit is equal to the sum of the individual power used by each circuit component.

KIRCHHOFF'S VOLTAGE LAW states: The algebraic sum of the voltage drops in any closed path in a circuit and the electromotive forces in that path is equal to zero, or $E_a + E_b + E_c + ... E_n = 0$.



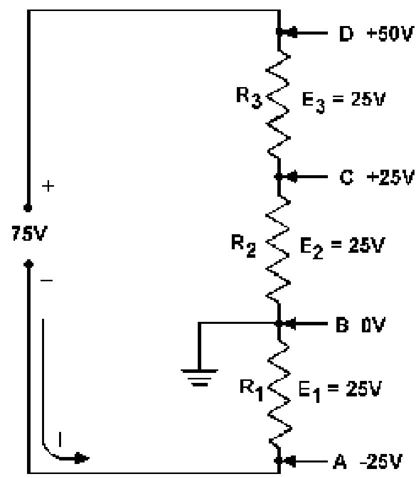
VOLTAGE POLARITIES must be used when applying Kirchhoff's voltage law. The point at which current enters a load (resistor) is considered negative with respect to the point at which current leaves the load.

SERIES AIDING VOLTAGES cause current to flow in the same direction; thus the voltages are added.



SERIES OPPOSING VOLTAGES tend to force current to flow in opposite directions; thus the equivalent voltage is the difference between the opposing voltages.

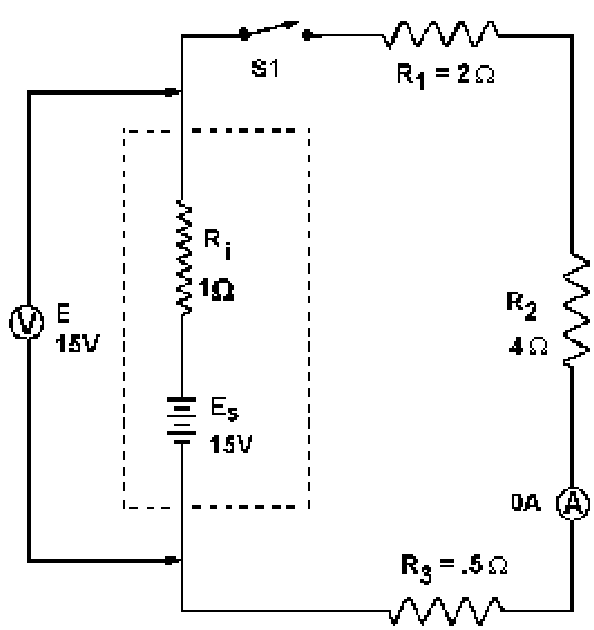
A REFERENCE POINT is a chosen point in a circuit to which all other points are compared.

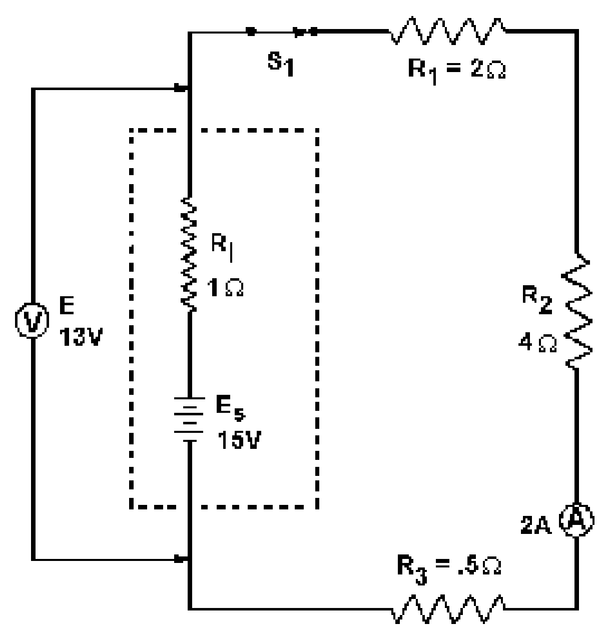


AN OPEN CIRCUIT is one in which a break exists in the complete conducting pathway.

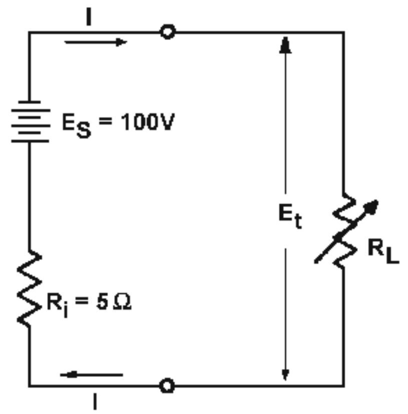
A SHORT CIRCUIT is an accidental path of low resistance which passes an abnormally high amount of current.

INTERNAL RESISTANCE causes a drop in the terminal voltage of a source as current flows through the source. The decrease in terminal voltage is caused by the voltage drop across the internal resistance. All sources of electromotive force have some form of internal resistance.





HIGH EFFICIENCY in a circuit is achieved when the resistance of the load is high with respect to the resistance of the source.

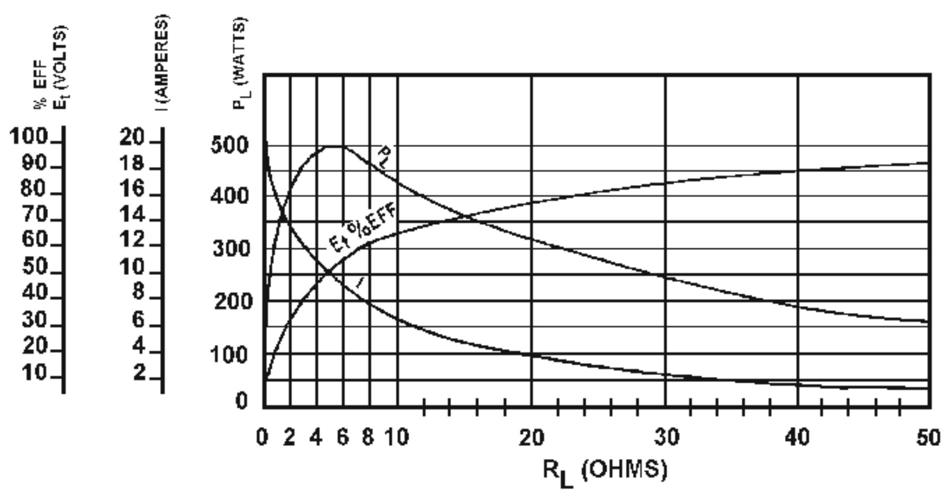


E_S = OPEN - CIRCUIT VOLTAGE OF SOURCE
R_i = INTERNAL RESISTANCE OF SOURCE
E_t = TERMINAL VOLTAGE
R_L = RESISTANCE OF LOAD
P_L = POWER USED IN LOAD
I = CURRENT FROM SOURCE
% EFF. = PERCENTAGE OF EFFICIENCY

(A)
CIRCUIT AND SYMBOL DESIGNATION

R_L	E_t	I	P_L	%EFF.
0	0	20	0	0
1	16.7	16.7	278.9	16.7
2	28.6	14.3	409	28.6
3	37.5	12.5	468.8	37.5
4	44.4	11.1	492.8	44.4
5	50	10	500	50
6	54.5	9.1	496.0	54.5
7	58.3	8.3	483.9	58.3
8	61.6	7.7	474.3	61.6
9	64.3	7.1	456.5	64.3
10	66.7	6.7	446.9	66.7
20	80	4	320	80
30	85.7	2.9	248.5	85.7
40	88.9	2.2	195.6	88.9
50	90.9	1.9	172.7	90.9

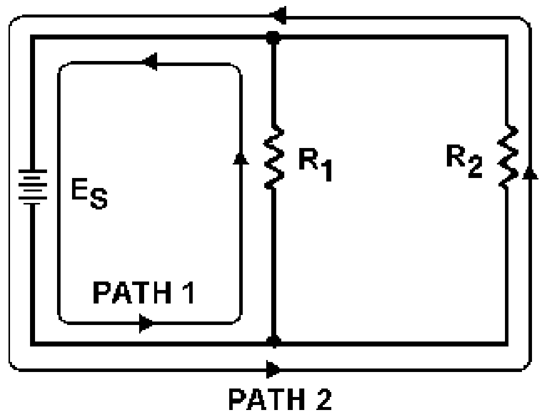
(B)
CHART



(C)
GRAPH

POWER TRANSFER in a circuit is highest when the resistance of the load equals the resistance of the source.

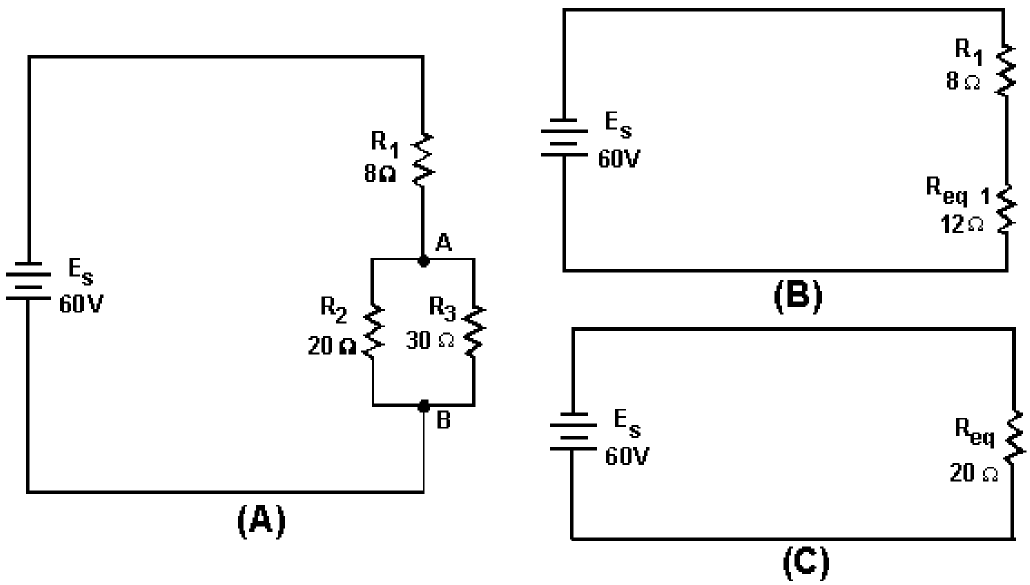
A PARALLEL CIRCUIT is a circuit having more than one current path connected to a common voltage source.



RULES FOR PARALLEL DC CIRCUITS:

- The same voltage exists across each branch of a parallel circuit and is equal to the source voltage.
- The current through a branch of a parallel network is inversely proportional to the amount of resistance of the branch.
- The total current of a parallel circuit is equal to the sum of the currents of the individual branches of the circuit.
- The total resistance of a parallel circuit is equal to the reciprocal of the sum of the reciprocals of the individual resistances of the circuit.
- The total power consumed in a parallel circuit is equal to the sum of the power consumptions of the individual resistances.

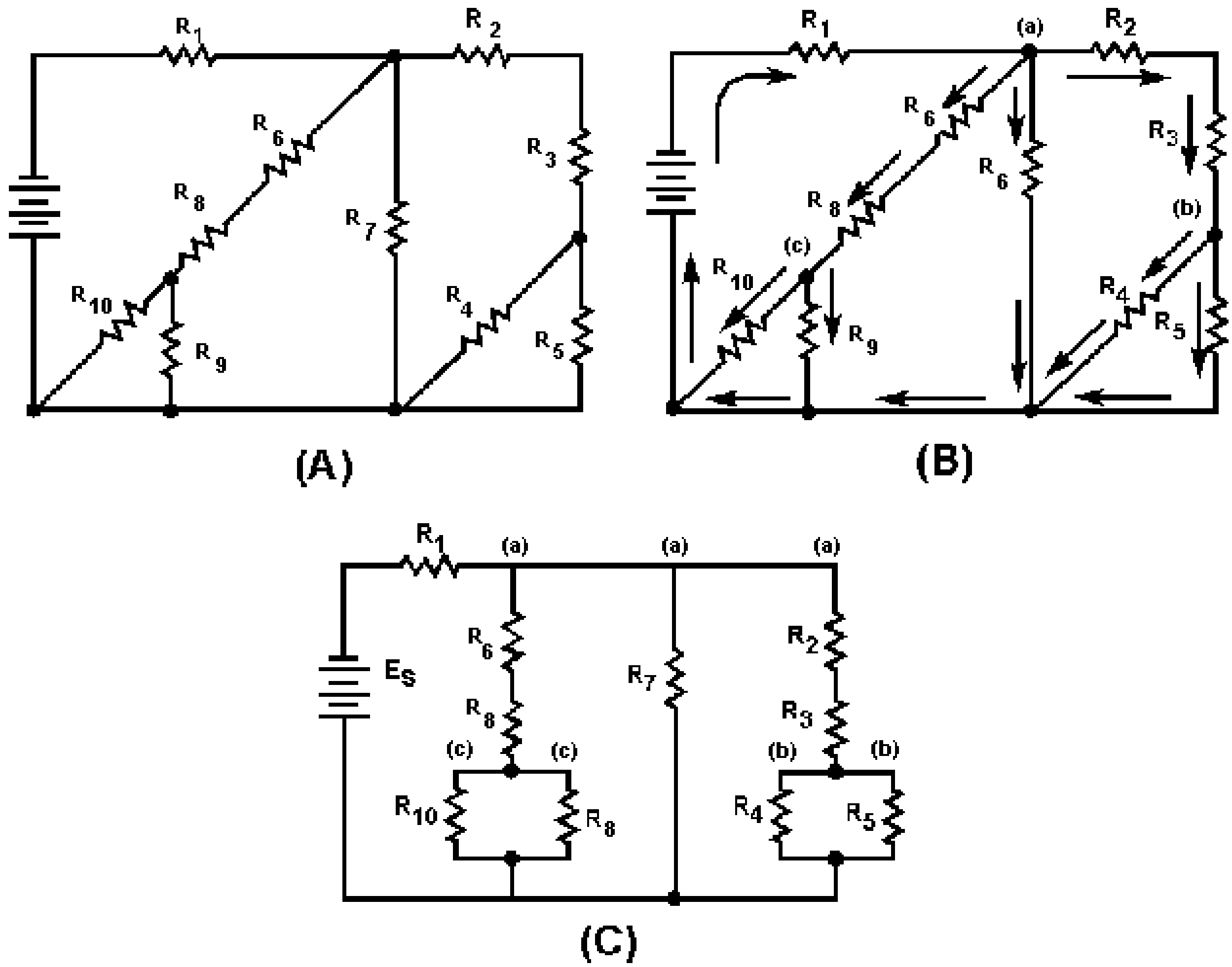
THE SOLUTION OF A COMBINATION CIRCUIT is a matter of applying the laws and rules for series and parallel circuits as applicable.



<u>ALL PARALLEL CIRCUITS ARE COMBINATION CIRCUITS</u> when the internal resistance of the source is taken into account.

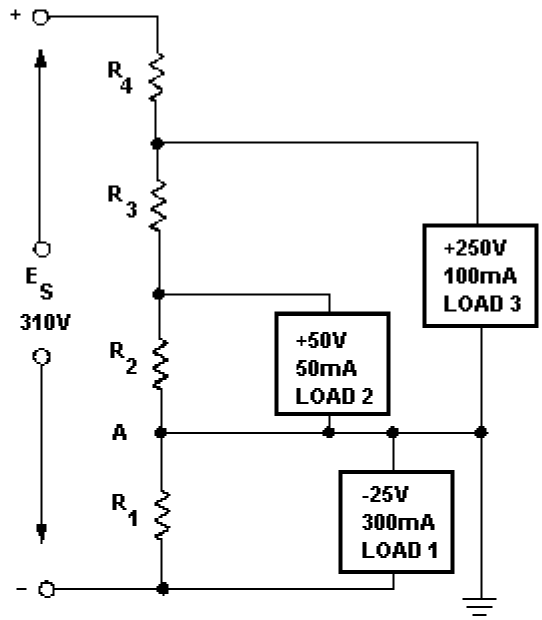
<u>REDRAWING CIRCUITS FOR CLARITY</u> is accomplished in the following steps:

1. Trace the current paths in the circuit.
2. Label the junctions in the circuit.
3. Recognize points which are at the same potential.
4. Visualize rearrangements, "stretching" or "shrinking," of connecting wires.
5. Redraw the circuit into simpler form (through stages if necessary).



EQUIPMENT PROTECTION from short-circuit current is accomplished by use of fuses and other circuit protection devices.

A VOLTAGE DIVIDER is a series circuit in which desired portions of the source voltage may be tapped off for use in equipment. Both negative and positive voltage can be provided to the loads by the proper selection of the reference point (ground).



ELECTRICAL SAFETY PRECAUTIONS must be observed. A fatal shock can occur from 0.1 ampere of current. Voltages as low as 30 volts have been recorded as causing sufficient current to be fatal.

ALL LIVE ELECTRICAL CIRCUITS shall be treated as potential hazards at all times.

ELECTRONIC OR ELECTRICAL EQUIPMENT discovered to be faulty or unsafe should be reported immediately to proper authority.

ELECTRICAL OR ELECTRONIC EQUIPMENT should be used and repaired by authorized personnel only.

A CO_2 EXTINGUISHER should be used to extinguish electrical fires.

FIRST AID FOR ELECTRICAL SHOCK includes the following actions:

- Remove the victim from the source of the shock.
- Check the victim to see if the person is breathing.
- If the victim is not breathing, give artificial ventilation. The preferred method is mouth-to-mouth.
- CPR may be necessary if the heartbeat has stopped, but do not attempt this unless you have been trained in its use. OBTAIN MEDICAL ASSISTANCE AS SOON AS POSSIBLE.

ANSWERS TO QUESTIONS Q1. THROUGH Q61.

A1. (a) DS1, the flashlight bulb (b) BAT, the dry cell

A2. The path for current is incomplete; or, there is no path for current with S1 open.

A3. A schematic diagram.

A4. (a) Current increases (b) Current decreases

A5. (a) Current decreases (b) Current increases

A6.

$$R = \frac{E}{I}$$

A7. 1.25 amperes.

A8. 4 amperes.

A9. Power.

A10. By changing the circuit resistance or the voltage of the power source.

A11.

$$P = E \times I, \quad P = \frac{E^2}{R}, \quad P = I^2 \times R$$

A12. 6 amperes.

A13. A wirewound resistor.

A14. 1 kilowatt.

A15. 8,952 watt hours or 8.952 kWh.

A16. 942 (rounded to 3 places).

A17.

(a). 160 ohms
(b). 480 volts

A18.

E_1 = 60 volts
E_2 = 180 volts
E_3 = 240 volts

A19.

E_1 = 80 volts
E_2 = 240 volts
E_3 = 320 volts

A20. The source voltage would have to be increased to 640 volts.

A21.

(a) 330 volts

(b) E_1 = 150 volts
E_2 = 180 volts

(c) 1.98 kilowatts

(d) P_1 = 900 watts
P_2 = 1.08 kilowatts

A22. The point at which current enters the resistor is assigned a negative polarity and the point at which current leaves the resistor is assigned a positive polarity.

A23. 2 amperes.

A24. 120 volts.

A25. 50 volts.

A26. Zero volts.

A27. A circuit where there is no longer a complete path for current flow.

A28. An accidental path of low resistance which passes an abnormally high amount of current.

A29. The internal (source) resistance of the battery will drop some of the voltage.

A30. When the load resistance equals the source resistance.

A31. 50 percent.

A32.

$$98 \text{ percent } \left(\frac{12.25 \text{ watts}}{12.5 \text{ watts}} \times 100 \right)$$

A33. *60 volts.*

A34. *Total current in a series circuit flows through every circuit component but in a parallel circuit total current divides among the available paths.*

A35. *Whether the current is entering the junction (+) or leaving the junction (-).*

A36.

$$25 \text{ ohms } \left(R_{eq} = \frac{R}{N} \right)$$

A37.

$$6\ k\Omega \left(R_{eq} = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}} \right) \quad \text{(use powers of tens)}$$

A38.

$$7.5\ k\Omega \ \left(R_{eq} = \frac{R_1 \times R_2}{R_1 + R_2} \right)$$

A39. *Equivalent resistor or R_{eq}.*

A40. *In both cases all the power used in the circuit must come from the source.*

A41.

$$R_T = 12\Omega \quad \left(\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)$$

$$E_S = 120V \ \left(E_S = E_{R3} = \sqrt{P_{R3} \times R_3} \right)$$

$$I_{R2} = 4A \quad \left(I_{R2} = \frac{E_S}{R_2} \right)$$

A42. *$P_T = 60\ W$, $E_{R2} = 10\ V$.*

A43. *4Ω.*

A44. *25Ω.*

A45. Because of the 2-volt drop across the internal resistance, only 48 volts is available for the rest of the circuit.

A46. (a) Total resistance increases, total current decreases (b) Total resistance becomes infinite, total current is equal to zero

A47. (a) Total resistance decreases, total current increases (b) Total resistance decreases, total current increases

A48. None.

A49. The source voltage and load requirements (voltage and current).

A50. 45 mA rule-of-thumb.

A51. 2 kΩ.

A52. 495 mA.

A53. R_1 is the bleeder resistor. Bleeder current must be known before any of the remaining divider resistor ohmic values can be computed.

A54. (a) By adding the bleeder current (I_{R1}) and the current through load 1(b) By subtracting the voltage of load 1 from the voltage of load 2.

A55. 1.35 watts.

A56. The series-parallel network drops the remaining source voltage and is used to take the place of a single resistor (75 ohms) when the required ohmic value is not available in a single resistor.

A57. R_3 = 2 watts; R_5 = 6 watts.

A58. The ground (reference point) is placed in the proper point in the voltage divider so that positive and negative voltages are supplied.

A59. NEVER! All energized electric circuits should be considered potentially dangerous.

A60. You should immediately report this condition to a qualified technician.

A61. Only trained, qualified personnel.

APPENDIX I

GLOSSARY

AMMETER—An instrument for measuring the amount of electron flow in amperes.

AMPERE—The basic unit of electrical current.

ANODE—A positive electrode of an electrochemical device (such as a primary or secondary electric cell) toward which the negative ions are drawn.

ATTRACTION—The force that tends to make two objects approach each other. Attraction exists between two unlike magnetic poles (North and South) or between two unlike static charges (plus and minus).

BATTERY—A device for converting chemical energy into electrical energy.

BATTERY CAPACITY—The amount of energy available from a battery. Battery capacity is expressed in ampere-hours.

BLEEDER CURRENT—The current through a bleeder resistor. In a voltage divider, bleeder current is usually determined by the 10 percent rule of thumb.

BLEEDER RESISTOR—A resistor which is used to draw a fixed current.

BRANCH—An individual current path in a parallel circuit.

CATHODE—The general name for any negative electrode.

CELL—A single unit that transforms chemical energy into electrical energy. Batteries are made up of cells.

CHARGE—Represents electrical energy. A material having an excess of electrons is said to have a negative charge. A material having a deficiency of electrons is said to have a positive charge.

CIRCUIT—The complete path of an electric current.

CIRCULAR MIL—An area equal to that of a circle with a diameter of 0.001 inch. It is used for measuring the cross-sectional area of wires.

COMBINATION CIRCUIT—A series-parallel circuit.

CONDUCTANCE—The ability of a material to conduct or carry an electric current. It is the reciprocal of the resistance of the material, and is expressed in mhos or siemans.

CONDUCTIVITY—Ease with which a substance transmits electricity.

CONDUCTOR—(1) A material with a large number of free electrons. (2) A material which easily permits electric current to flow.

COULOMB—A measure of the quantity of electricity. One coulomb is equal to 6.28×10^{18} electrons.

COULOMB'S LAW—Also called the law of electric charges or the law of electrostatic attraction. Coulomb's Law states that charged bodies attract or repel each other with a force that is directly proportional to the product of their individual charges and inversely proportional to the square of the distance between them.

CPR—Cardio-Pulminary Resuscitation.

CROSS-SECTIONAL AREA—The area of a "slice" of an object. When applied to electrical conductors, it is usually expressed in circular mils.

CURRENT—The flow of electrons past a reference point. The passage of electrons through a conductor. Measured in amperes.

DEAD SHORT—A short circuit having minimum resistance.

DIELECTRIC FIELD—The space between and around charged bodies in which their influence is felt. Also called Electric Field of Force or an Electrostatic Field.

DIRECT CURRENT—An electric current that flows in one direction only.

DOMAIN THEORY—A theory of magnetism based upon the electron-spin principle. Spinning electrons have a magnetic field. If more electrons spin in one direction than another, the atom is magnetized.

DRY CELL—An electrical cell in which the electrolyte is not a liquid. In most dry cells the electrolyte is in the form of a paste.

EFFICIENCY—The ratio of output power to input power, generally expressed as a percentage.

ELECTRIC CURRENT—The flow of electrons.

ELECTRICAL CHARGE—Symbol Q, q. Electric energy stored on or in an object. The negative charge is caused by an excess of electrons; the positive charge is caused by a deficiency of electrons.

ELECTROCHEMICAL—The action of converting chemical energy into electrical energy.

ELECTRODE—The terminal at which electricity passes from one medium into another, such as in an electrical cell where the current leaves or returns to the electrolyte.

ELECTROLYTE—A solution of a substance which is capable of conducting electricity. An electrolyte may be in the form of either a liquid or a paste.

ELECTROMAGNET—An electrically excited magnet capable of exerting mechanical force, or of performing mechanical work.

ELECTROMAGNETIC—The term describing the relationship between electricity and magnetism. Having both magnetic and electric properties.

ELECTROMAGNETIC INDUCTION—The production of a voltage in a coil due to a change in the number of magnetic lines of force (flux linkages) passing through the coil.

ELECTRON—The elementary negative charge that revolves around the nucleus of an atom.

ELECTRON SHELL—A group of electrons which have a common energy level that forms part of the outer structure (shell) of an atom.

ELECTROSTATIC—Pertaining to electricity at rest, such as charges on an object (static electricity).

ELEMENT—A substance, in chemistry, that cannot be divided into simpler substances by any means ordinarily available.

EMF—(Electromotive Force) The force which causes electricity to flow between two points with different electrical charges or when there is a difference of potential between the two points. The unit of measurement in volts.

ENERGY—The ability or capacity to do work.

EQUIVALENT RESISTANCE—(R_{eq}) A resistance that represents the total ohmic values of a circuit component or group of circuit components. Usually drawn as a single resistor when simplifying complex circuits.

FERROMAGNETIC MATERIAL—A highly magnetic material, such as iron, cobalt, nickel, or alloys, make up these materials.

FIELD OF FORCE—A term used to describe the total force exerted by an action-at-a-distance phenomenon such as gravity upon matter, electric charges acting upon electric charges, magnetic forces acting upon other magnets or magnetic materials.

FIXED RESISTOR—A resistor having a definite resistance value that cannot be adjusted.

FLUX—In electrical or electromagnetic devices, a general term used to designate collectively all the electric or magnetic lines of force in a region.

FLUX DENSITY—The number of magnetic lines of force passing through a given area.

GAS—One of the three states of matter having no fixed form or volume. (Steam is a gas.)

GRAPH—A pictorial presentation of the relation between two or more variable quantities, such as between an applied voltage and the current it produces in a circuit.

GROUND POTENTIAL—Zero potential with respect to the ground or earth.

HORSEPOWER—The English unit of power, equal to work done at the rate of 550 foot-pounds per second. Equal to 746 watts of electrical power.

HORSESHOE MAGNET—A permanent magnet or electromagnet bent into the shape of a horseshoe or having a U-shape to bring the two poles near each other.

HYDROMETER—An instrument used to measure specific gravity. In batteries hydrometers are used to indicate the state of charge by the specific gravity of the electrolyte.

INDUCED CHARGE—An electrostatic charge produced on an object by the electric field that surrounds a nearby object.

INDUCED CURRENT—Current due to the relative motion between a conductor and a magnetic field.

INDUCED ELECTROMOTIVE FORCE—The electromotive force induced in a conductor due to the relative motion between a conductor and a magnetic field.

INDUCED VOLTAGE—See Induced Electromotive Force.

INDUCTION—The act or process of producing voltage by the relative motion of a magnetic field across a conductor.

INFINITE—(1) Extending indefinitely, endless. (2) Boundless having no limits. (3) An incalculable number.

INSULATION—(1) A material used to prevent the leakage of electricity from a conductor and to provide mechanical spacing or support to protect against accidental contact. (2) Use of material in which current flow is negligible to surround or separate a conductor to prevent loss of current.

INSULATOR—(1) Material of such low conductivity that the flow of current through it can usually be neglected. (2) Device having high-electric resistance, used for supporting or separating conductors so as to prevent undesired flow of current from the conductors to other objects.

INVERSELY—Inverted or reversed in position or relationship.

ION—An electrically charged atom or group of atoms. Negative ions have an excess of electrons; positive ions have a deficiency of electrons.

IONIZE—To make an atom or molecule of an element lose an electron, as by X-ray bombardment, and thus be converted into a positive ion. The freed electron may attach itself to a neutral atom or molecule to form a negative ion.

JUNCTION—(1) The connection between two or more conductors. (2) The contact between two dissimilar metals or materials, as is in a thermocouple.

KILO—A prefix meaning one thousand.

KINETIC ENERGY—Energy which a body possesses by virtue of its motion.

KIRCHHOFF'S LAWS—(1) The algebraic sum of the currents flowing toward any point and the current flowing from that point in an electric network is zero. (2) The algebraic sum of the products of the current and resistance in each of the conductors in any closed path in a network is equal to the algebraic sum of the electromotive forces in the path.

LAW OF MAGNETISM—Like poles repel; unlike poles attract.

LEAD-ACID CELL—A cell in an ordinary storage battery, in which electrodes are grids of lead containing an active material consisting of certain lead oxides that change in composition during charging and discharging. The electrodes or plates are immersed in an electrolyte of diluted sulfuric acid.

LINE OF FORCE—A line in an electric or magnetic field that shows the direction of the force.

LIQUID—One of the three states of matter which has a definite volume but no definite form. (Water is a liquid.)

LOAD—(1) A device through which an electric current flows and which changes electrical energy into another form. (2) Power consumed by a device or circuit in performing its function.

LOCAL ACTION—A continuation of current flow within an electrical cell when there is no external load. Caused by impurities in the electrode.

MAGNETIC FIELD—The space in which a magnetic force exists.

MAGNETIC POLES—The section of a magnet where the flux lines are concentrated; also where they enter and leave the magnet.

MAGNETISM—The property possessed by certain materials by which these materials can exert mechanical force on neighboring masses of magnetic materials; and can cause currents to be induced in conducting bodies moving relative to the magnetized bodies.

MATTER—Any physical entity which possesses mass.

MEGA—A prefix meaning one million, also Meg.

MHO—Unit of conductance: the reciprocal of the ohm. Replaced by siemens.

MICRO—A prefix meaning one-millionth.

MILLI—A prefix meaning one-thousandth.

NEGATIVE ELECTRODE—A terminal or electrode having more electrons than normal. Electrons flow out of the negative terminal of a voltage source.

NEGATIVE TEMPERATURE COEFFICIENT—The temperature coefficient expressing the amount of reduction in the value of a quantity, such as resistance for each degree of increase in temperature.

NETWORK—A combination of electrical components. In a parallel circuit it is composed of two or more branches.

NEUTRAL—In a normal condition, hence neither positive nor negative. A neutral object has a normal number of electrons.

OHM—The unit of electrical resistance. It is that value of electrical resistance through which a constant potential difference of 1 volt across the resistance will maintain a current flow of 1 ampere through the resistance.

OHMIC VALUE—Resistance in ohms.

OHM'S LAW—The current in an electric circuit is directly proportional to the electromotive force in the circuit. The most common form of the law is E = IR, where E is the electromotive force or voltage across the circuit, I is the current flowing in the circuit, and R is the resistance of the circuit.

OPEN CIRCUIT—(1) The condition of an electrical circuit caused by the breaking of continuity of one or more conductors of the circuit; usually an undesired condition. (2) A circuit which does not provide a complete path for the flow of current.

PARALLEL CIRCUIT—Two or more electrical devices connected to the same pair of terminals so separate currents flow through each; electrons have more than one path to travel from the negative to the positive terminal.

PERMEABILITY—The measure of the ability of a material to act as a path for magnetic lines of force.

PHOTOELECTRIC VOLTAGE—A voltage produced by light.

PICO—A prefix adopted by the National Bureau of Standards meaning 10^{-12}.

PIEZOELECTRIC EFFECT—The effect of producing a voltage by placing a stress, either by compression, expansion, or twisting, on a crystal and, conversely, producing a stress in a crystal by applying a voltage to it.

PLATE—One of the electrodes in a storage battery.

POLARITY—(1) The condition in an electrical circuit by which the direction of the flow of current can be determined. Usually applied to batteries and other direct voltage sources. (2) Two opposite charges, one positive and one negative. (3) A quality of having two opposite magnetic poles, one north and the other south.

POLARIZATION—The effect of hydrogen surrounding the anode of a cell which increases the internal resistance of the cell.

POTENTIAL ENERGY—Energy due to the position of one body with respect to another body or to the relative parts of the same body.

POTENTIOMETER—A 3-terminal resistor with one or more sliding contacts, which functions as an adjustable voltage divider.

POWER—The rate of doing work or the rate of expending energy. The unit of electrical power is the watt.

PRIMARY CELL—An electrochemical cell in which the chemical action eats away one of the electrodes, usually the negative electrode.

RECIPROCAL—The value obtained by dividing the number 1 by any quantity.

REFERENCE POINT—A point in a circuit to which all other points in the circuit are compared.

RELUCTANCE—A measure of the opposition that a material offers to magnetic lines of force.

REPULSION—The mechanical force tending to separate bodies having like electrical charges or like magnetic polarity.

RESIDUAL MAGNETISM—Magnetism remaining in a substance after removal of the magnetizing force.

RESISTANCE—(1) The property of a conductor which determines the amount of current that will flow as the result of the application of a given electromotive force. All conductors possess some resistance, but when a device is made especially for the purpose of limiting current flow, it is called a resistor. A resistance of 1 ohm will allow a current of 1 ampere to flow through it when a potential of 1 volt is applied. (2) The opposition which a device or material offers to the flow of current. The effect of resistance is to raise the temperature of the material or device carrying the current. (3) A circuit element designed to offer a predetermined resistance to current flow.

RESISTOR—The electrical component which offers resistance to the flow of current. It may be a coil of fine wire or a composition rod.

RETENTIVITY—The ability of a material to retain its magnetism.

RHEOSTAT—(1) A resistor whose value can be varied. (2) A variable resistor which is used for the purpose of adjusting the current in a circuit.

SCHEMATIC CIRCUIT DIAGRAM—A circuit diagram in which component parts are represented by simple, easily drawn symbols. May be called schematic.

SCHEMATIC SYMBOLS—A letter, abbreviation, or design used to represent specific characteristics or components on a schematic diagram.

SECONDARY CELL—A cell that can be recharged by passing a current through the cell in a direction opposite to the discharge current.

SERIES CIRCUIT—An arrangement where electrical devices are connected so that the total current must flow through all the devices; electrons have one path to travel from the negative terminal to the positive terminal.

SERIES-PARALLEL CIRCUIT—A circuit that consists of both series and parallel networks.

SHELF LIFE—The period of time that a cell or battery may be stored and still be useful.

SHIELDING—A metallic covering used to prevent magnetic or electromagnetic fields from effecting an object.

SHORT CIRCUIT—A low resistance connection between two points of different potential in a circuit, usually accidental and usually resulting in excessive current flow that may cause damage.

SIEMANS—The new and preferred term for mho.

SOLID—One of the three states of matter which has definite volume and shape. (Ice is a solid.)

SOURCE VOLTAGE—The device which furnishes the electrical energy used by a load.

SPECIFIC GRAVITY—The ratio between the density of a substance and that of pure water at a given temperature.

STATIC ELECTRICITY—Stationary electricity that is in the form of a charge. The accumulated electric charge on an object.

SWITCH—A device to connect, disconnect, or change the connections in an electrical circuit.

TAPPED RESISTOR—A wire-wound, fixed resistor having one or more additional terminals along its length, generally for voltage-divider applications.

TEMPERATURE COEFFICIENT—The amount of change of resistance in a material per unit change in temperature.

TERMINAL—An electrical connection.

THERMOCOUPLE—A junction of two dissimilar metals that produces a voltage when heated.

TOLERANCE—(1) The maximum error or variation from the standard permissible in a measuring instrument. (2) A maximum electrical or mechanical variation from specifications which can be tolerated without impairing the operation of a device.

TOTAL RESISTANCE—(R_T) The equivalent resistance of an entire circuit. For a series circuit: $R_T = R_1 + R_2 + R_3 \ldots R_n$. For parallel circuits:

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots \frac{1}{R_n}$$

UNIDIRECTIONAL—In one direction only.

VALENCE—The measure of the extent to which an atom is able to combine directly with other atoms. It is believed to depend on the number and arrangement of the electrons in the outermost shell of the atom.

VALENCE SHELL—The electrons that form the outermost shell of an atom.

VARIABLE RESISTOR—A wire-wound or composition resistor, the value of which may be changed.

VOLT—The unit of electromotive force or electrical pressure. One volt is the pressure required to send 1 ampere of current through a resistance of 1 ohm.

VOLTAGE—(1) The term used to signify electrical pressure. Voltage is a force which causes current to flow through an electrical conductor. (2) The voltage of a circuit is the greatest effective difference of potential between any two conductors of the circuit

VOLTAGE DIVIDER—A series circuit in which desired portions of the source voltage may be tapped off for use in equipment.

VOLTAGE DROP—The difference in voltage between two points. It is the result of the loss of electrical pressure as a current flows through a resistance.

WATT—The practical unit of electrical power. It is the amount of power used when one ampere of dc flows through a resistance of one ohm.

WATTAGE RATING—A rating expressing the maximum power that a device can safely handle.

WATT-HOUR—A practical unit of electrical energy equal to one watt of power for one hour.

WEBER'S THEORY—A theory of magnetism which assumes that all magnetic material is composed of many tiny magnets. A piece of magnetic material that is magnetized has all of the tiny magnets aligned so that the north pole of each magnet points in one direction.

WIRE—A solid or stranded group of solid, cylindrical conductors having low resistance to current flow, with any associated insulation.

WORK—The product of force and motion.

APPENDIX II

LAWS OF EXPONENTS

The International Symbols Committee has adopted prefixes for denoting decimal multiples of units. The National Bureau of Standards has followed the recommendations of this committee, and has adopted the following list of prefixes:

Numbers	Powers of ten	Prefixes	Symbols
1,000,000,000,000	10^{12}	tera	T
1,000,000,000	10^{9}	giga	G
1,000,000	10^{6}	mega	M
1,000	10^{3}	kilo	k
100	10^{2}	hecto	h
10	10	deka	da
.1	10^{-1}	deci	d
.01	10^{-2}	centi	c
.001	10^{-3}	milli	m
.000001	10^{-6}	micro	u
.000000001	10^{-9}	nano	n
.000000000001	10^{-12}	Pico	p
.000000000000001	10^{-15}	femto	F
.000000000000000001	10^{-18}	atto	a

To multiply like (with same base) exponential quantities, add the exponents. In the language of algebra the rule is $a^m \times a^n = a^{m+n}$

$$10^4 \times 10^2 = 10^{4+2} = 10^6$$

$$0.003 \times 825.2 = 3 \times 10^{-3} \times 8.252 \times 10^2$$

$$= 24.756 \times 10^{-1} = 2.4756$$

To divide exponential quantities, subtract the exponents. In the language of algebra the rule is

$$\frac{a^m}{a^n} = a^{m-n}$$

or

$$10^8 \div 10^2 = 10^6$$

*Generally used with electrical quantities.

$$3{,}000 \div 0.015 = (3 \times 10^3) \div (1.5 \times 10^{-2})$$
$$= 2 \times 10^5 = 200{,}000$$

To raise an exponential quantity to a power, multiply the exponents. In the language of algebra $(x^m)^n = x^{mn}$.

$$(10^3)^4 = 10^{3 \times 4} = 10^{12}$$
$$2{,}500^2 = (2.5 \times 10^3)^2 = 6.25 \times 10^6 = 6{,}250{,}000$$

Any number (except zero) raised to the zero power is one. In the language of algebra $x^0 = 1$

$$x^3 \div x^3 = 1$$
$$10^4 \div 10^4 = 1$$

Any base with a negative exponent is equal to 1 divided by the base with an equal positive exponent. In the language of algebra $x^{-a} = 1/x^a$

$$10^{-2} = \frac{1}{10^2} = \frac{1}{100}$$
$$5a^{-3} = \frac{5}{a^3}$$
$$(6a)^{-1} = \frac{1}{6a}$$

To raise a product to a power, raise each factor of the product to that power.

$$(2 \times 10)^2 = 2^2 \times 10^2$$
$$3{,}000^3 = (3 \times 10^3)^3 = 27 \times 10^9$$

To find the nth root of an exponential quantity, divide the exponent by the index of the root. Thus, the nth root of $a^m = a^{m/n}$.

$$\sqrt{x^6} = x^{6/2} = x^3$$

$$3\sqrt{64 \times 10^3} = 4 \times 10 = 40$$

APPENDIX III

SQUARE AND SQUARE ROOTS

N	N^2	$\sqrt{N}$	N	N^2	$\sqrt{N}$	N	N^2	$\sqrt{N}$
1	1	1.000	41	1681	6.4031	81	6561	9.0000
2	4	1.414	42	1764	6.4807	82	6724	9.0554
3	9	1.732	43	1849	6.5574	83	6889	9.1104
4	16	2.000	44	1936	6.6332	84	7056	9.1652
5	25	2.236	45	2025	6.7082	85	7225	9.2195
6	36	2.449	46	2116	6.7823	86	7396	9.2736
7	49	2.646	47	2209	6.8557	87	7569	9.3274
8	64	2.828	48	2304	6.9282	88	7744	9.3808
9	81	3.000	49	2401	7.0000	89	7921	9.4340
10	100	3.162	50	2500	7.0711	90	8100	9.4868
11	121	3.3166	51	2601	7.1414	91	8281	9.5394
12	144	3.4641	52	2704	7.2111	92	8464	9.5917
13	169	3.6056	53	2809	7.2801	93	8649	9.6437
14	196	3.7417	54	2916	7.3485	94	8836	9.6954
15	225	3.8730	55	3025	7.4162	95	9025	9.7468
16	256	4.0000	56	3136	7.4833	96	9216	9.7980
17	289	4.1231	57	3249	7.5498	97	9409	9.8489
18	324	4.2426	58	3364	7.6158	98	9604	9.8995
19	361	4.3589	59	3481	7.6811	99	9801	9.9499
20	400	4.4721	60	3600	7.7460	100	10000	10.0000
21	441	4.5826	61	3721	7.8102	101	10201	10.0499
22	484	4.6904	62	3844	7.8740	102	10404	10.0995
23	529	4.7958	63	3969	7.9373	103	10609	10.1489
24	576	4.8990	64	4096	8.0000	104	10816	10.1980
25	625	5.0000	65	4225	8.0623	105	11025	10.2470
26	676	5.0990	66	4356	8.1240	106	11236	10.2956
27	729	5.1962	67	4489	8.1854	107	11449	10.3441
28	784	5.2915	68	4624	8.2462	108	11664	10.3923
29	841	5.3852	69	4761	8.3066	109	11881	10.4403
30	900	5.4772	70	4900	8.3666	110	12100	10.4881
31	961	5.5678	71	5041	8.4261	111	12321	10.5357
32	1024	5.6569	72	5184	8.4853	112	12544	10.5830
33	1089	5.7447	73	5329	8.5440	113	12769	10.6301
34	1156	5.8310	74	5476	8.6023	114	12996	10.6771
35	1225	5.9161	75	5625	8.6603	115	13225	10.7238
36	1296	6.0000	76	5776	8.7178	116	13456	10.7703
37	1369	6.0828	77	5929	8.7750	117	13689	10.8167
38	1444	6.1644	78	6084	8.8318	118	13924	10.8628
39	1521	6.2450	79	6241	8.8882	119	14161	10.9087
40	1600	6.3246	80	6400	8.9443	120	14400	10.9545

For numbers up to 120. For larger numbers divide into factors smaller than 120.

Examples: $\sqrt{225}$ and $\sqrt{16200}$

$$225 = 5 \times 45$$
$$\sqrt{225} = \sqrt{5} \times \sqrt{45}$$
$$\sqrt{225} = 2.236 \times 6.7082$$
$$\sqrt{225} = 15$$

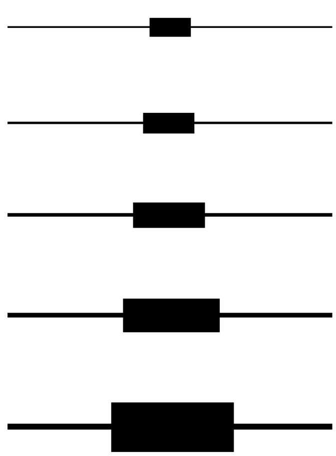
$$16200 = 100 \times 81 \times 2$$
$$\sqrt{16200} = \sqrt{100} \times \sqrt{81} \times \sqrt{2}$$
$$\sqrt{16200} = 10 \times 9 \times 1.414$$
$$\sqrt{16200} = 127.26$$

APPENDIX IV

COMPARISON OF UNITS IN ELECTRIC AND MAGNETIC CIRCUITS; AND CARBON RESISTOR SIZE COMPARISON BY WATTAGE RATING

	Electric circuit	Magnetic circuit
Force.	Volt, E, or e.m.f.	Gilberts, F, or m.m.f.
Flow	Ampere, I	Flux, Φ, in maxwells
Opposition.	Ohms, R	Reluctance, $\mathcal{R}$
Law.	Ohm's law, $I = \frac{E}{R}$	Rowland's law, $\Phi = \frac{F}{\mathcal{R}}$
Intensity of force	Volts per cm. of length.	$H = \frac{1.257IN}{L}$, gilberts per centimeter of length.
Density.	Current density—for example, amperes per cm^2.	Flux density—for example, lines per cm^2, or gausses.

Carbon Resistor Size Comparison by Wattage Rating



1/8 WATT

1/4 WATT

1/2 WATT

1 WATT

2 WATT

APPENDIX V

USEFUL FORMULAS FOR DC CIRCUITS

Ohm's Law for D.C. Circuits

$$I = \frac{E}{R} = \frac{P}{E} = \sqrt{\frac{P}{R}}$$

$$R = \frac{E}{I} = \frac{P}{I^2} = \frac{E^2}{P}$$

$$E = IR = \frac{P}{I} = \sqrt{PR}$$

$$P = EI = \frac{E^2}{R} = I^2 R$$

Resistors in Series

$$R_T = R_1 + R_2 + \ldots R_n$$

Resistors in Parallel

Two resistors

$$R_T = \frac{R_1 R_2}{R_1 + R_2}$$

More than two

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots \frac{1}{R_n}$$

Where R = resistance in ohms,
I = current in amperes,
E = potential across R in volts,
P = power in watts

OHM'S LAW FORMULAS FOR D.C. CIRCUITS

Known Values	Formulas for Determining Unknown Values of . . .			
	R	I	E	P
I&R			IR	$I^2 R$
I&E	$\frac{E}{I}$			EI
I&P	$\frac{P}{I^2}$		$\frac{P}{I}$	
R&E		$\frac{E}{R}$		$\frac{E^2}{R}$
R&P		$\sqrt{\frac{P}{R}}$	$\sqrt{PR}$	
E&P	$\frac{E^2}{P}$	$\frac{P}{E}$		

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Turning to Electricity," pages 1-1 through 1-65.

1-1. Matter can be found in which of the following forms?

1. Solid
2. Liquid
3. Gaseous
4. Each of the above

1-2. A substance that CANNOT be reduced to a simpler substance by chemical means is called a/an

1. element
2. mixture
3. compound
4. solution

1-3. A molecule is the smallest possible particle that retains the characteristic of which of the following substances?

1. An element
2. A mixture
3. A compound
4. A solution

1-4. An atom is the smallest possible particle that retains the characteristic of which of the following substances?

1. An element
2. A mixture
3. A compound
4. A solution

1-5. What subatomic particle has a negative charge and a small mass?

1. Proton
2. Electron
3. Positron
4. Neutron

1-6. What subatomic particle has a positive charge and a large mass?

1. Proton
2. Electron
3. Positron
4. Neutron

1-7. What subatomic particle has no charge?

1. Proton
2. Electron
3. Positron
4. Neutron

1-8. When light is represented as a tiny packet of energy, what are these packets of energy called?

1. Angstroms
2. Photons
3. Wavelengths
4. Frequencies

1-9. If light energy collides with an orbiting electron, what happens to the electron?

1. The electron will move around the same orbit faster
2. The electron will jump to an orbit further from the nucleus
3. The electron will jump to an orbit closer to the nucleus
4. The electron will merge with the nucleus

1-10. After the action described in question 1-9 occurs, the electron will return to the condition it had before being acted upon by the light. When the electron returns to this condition, which of the following actions occurs?

1. The nucleus becomes lighter
2. The atom becomes an ion
3. Light energy is emitted
4. The valence of the atom changes

1-11. The number of electrons in the outermost shell of an atom determines which of the following characteristics of the atom?

1. Valence
2. Atomic weight
3. Atomic number
4. Number of shells

1-12. When an atom gains or loses an electron, which of the following terms applies?

1. Unbalanced
2. Lightened
3. Neutral
4. Ionized

1-13. What is the main difference between conductors, semiconductors, and insulators?

1. The temperature differences
2. The physical state of their mass
3. The number of free electrons
4. The designations of the outer shells

1-14. A substance with an excess of electrons is considered to be in what electrical state?

1. Neutral
2. Positive
3. Negative
4. Discharged

1-15. Which of following actions describes the easiest way to accumulate a static electric charge?

1. Friction between two conductors
2. Friction between two insulators
3. Pressure between two conductors
4. Pressure between two insulators

1-16. An atom that contains 6 protons and 5 electrons has what electrical charge?

1. Positive
2. Negative
3. Neutral
4. Intermediate

1-17. How do "like" and "unlike" charges react to one another?

1. Unlike charges repel each other, like charges repel each other
2. Unlike charges attract each other, like charges attract each other
3. Unlike charges repel each other, like charges attract each other
4. Unlike charges attract each other, like charges repel each other

1-18. What is/are the term(s) applied to the space between and around charged bodies in which their influence is felt?

1. Electric field of force
2. Electrostatic field
3. Dielectric field
4. Each of the above

1-19. Electrostatic lines of force are drawn in which of the following manners?

1. Entering negative charge, entering positive charge
2. Entering negative charge, leaving positive charge
3. Leaving negative charge, leaving positive charge
4. Leaving negative charge, entering positive charge

1-20. Which of the following devices use magnetism?

1. Batteries
2. Light bulbs
3. High-fidelity speakers
4. Each of the above

1-21. Magnetic materials have which of the following qualities?

1. They are attracted by magnets
2. They can be magnetized
3. Both 1 and 2 above
4. They are electrical insulators

1-22. Ferromagnetic materials have which of the following qualities?

1. They are all alloys
2. They all contain nickel
3. They make very weak magnets
4. They are relatively easy to magnetize

1-23. A material with low reluctance and high permeability such as iron or soft steel is used to make what type of magnet?

1. Temporary
2. Permanent
3. Residual
4. Natural

1-24. The ability of a material to retain magnetism is called

1. permeability
2. retentivity
3. reluctance
4. ionization

1-25. The law of magnetic poles states which of the following relationships?

1. Like poles attract, unlike poles attract
2. Like poles attract, unlike poles repel
3. Like poles repel, unlike poles repel
4. Like poles repel, unlike poles attract

1-26. The north indicating pole of a compass needle is attracted to which of the following poles of the earth?

1. The geographic north pole
2. The magnetic north pole
3. The geographic south pole
4. The magnetic south pole

1-27. Weber's theory of magnetism assumes that magnetic material is composed of

1. tiny molecular magnets
2. domains of magnetic influence
3. large blocks of material acting as magnets
4. atoms with electrons spinning different directions

1-28. According to the domain theory, if an atom with 26 electrons has 20 electrons spinning counterclock-wise, the atom is considered to be

1. charged
2. insulated
3. neutralized
4. magnetized

1-29. If a glass plate is placed over a magnet and iron filings are sprinkled over the glass, a pattern will be visible. What does this pattern indicate?

1. The magnetic field
2. The electrostatic field
3. The piezoelectric effect
4. The chemical reaction of the magnet and the filings

1-30. An imaginary line used to illustrate a magnetic effect is known as a/an

1. magnetic pole
2. force field pole
3. magnetic line of force
4. electrostatic line of force

1-31. Which of the following is NOT a property of magnetic lines of force?

1. They form closed loops around the magnet
2. They leave the magnetic material at right angles to the surface
3. They cross each other at right angles
4. They leave the north pole and enter the south pole of the magnet

1-32. A magnetic shield or screen used to protect a delicate instrument should be made of which of the following materials?

1. Plastic
2. Copper
3. Soft iron
4. Aluminum

1-33. Bar magnets should be stored in which of the following manners?

1. Separately
2. In pairs at 90 degree angles
3. In pairs with north poles together
4. In pairs with a north pole and a south pole together

1-34. What is the term applied to the ability to do work?

1. Power
2. Energy
3. Voltage
4. Current

1-35. An object that is in motion has what type of energy?

1. Kinetic
2. Magnetic
3. Newtonian
4. Potential

1-36. A book sitting on a shelf has what kind of energy?

1. Kinetic
2. Potential
3. Newtonian
4. Magnetic

1-37. Which of the following term(s) apply(ies) to the difference of potential between two bodies?

1. Voltage
2. Electromotive force
3. Both 1 and 2 above
4. Current

1-38. Which of the following terms is equal to "2.1 kV?"

1. 210 V
2. 2100 V
3. 21,000 V
4. 2.1×10^6 V

1-39. 250µV is equal to which of the following terms?

1. .25 mV
2. .00025 V
3. 250×10^{-6} V
4. All of the above

1-40. What is the general term that describes a device which supplies a voltage?

1. A voltage source
2. A voltage supply
3. A voltage generator
4. A voltage producer

1-41. In addition to friction, magnetism, and chemical action, which of the following methods can be used to produce a voltage?

1. Pressure
2. Heat
3. Light
4. Each of the above

IN ANSWERING QUESTIONS 1-42 THROUGH 1-46, MATCH THE VOLTAGE PRODUCING METHOD LISTED IN COLUMN B TO THE DEVICE LISTED IN COLUMN A.

COLUMN A	COLUMN B
1-42. Radio receiver's oscillator	1. Heat
1-43. Thermocouple	2. Pressure
1-44. Automobile battery	3. Magnetism
1-45. Automobile generator	4. Chemical action
1-46. Flashlight cell	

1-47. Current in an electric circuit is caused by which of the following actions?

1. Electrons moving from negative to positive
2. Electrons moving from positive to negative
3. Protons moving from negative to positive
4. Protons moving from positive to negative

1-48. When directed drift takes place, at what speed does the effect take place?

1. 100,000 miles per hour
2. 186,000 miles per second
3. 300,000 meters per hour
4. 500,000 meters per second

1-49. If the voltage in a circuit increases, what happens to the current?

1. Current increases
2. Current decreases
3. Current remains the same
4. Current fluctuates rapidly

1-50. Which of the following values is equal to 100mA?

1. 1.0 ampere
2. 10.0 amperes
3. 0.10 ampere
4. 0.01 ampere

1-51. What symbol is used to represent the ohm?

1. A
2. O
3. μ
4. Ω

1-52. If low weight is the major factor, which of the following materials should be used as a conductor?

1. Aluminum
2. Copper
3. Silver
4. Gold

1-53. What material is MOST widely used as a conductor in electrical equipment?

1. Aluminum
2. Copper
3. Silver
4. Gold

1-54. Resistance of a conductor will increase with which of the following changes to the cross-sectional area and length of the conductor?

1. Cross-sectional area is increased, length is increased
2. Cross-sectional area is increased, length is decreased
3. Cross-sectional area is decreased, length is increased
4. Cross-sectional area is decreased, length is decreased

1-55. A material whose resistance decreases as the temperature increases has what temperature coefficient?

1. Positive
2. Negative
3. Zero
4. Neutral

1-56. A material whose resistance remains constant as the temperature increases has what temperature coefficient?

1. Positive
2. Negative
3. Zero
4. Neutral

1-57. Which of the following units is NOT a unit of conductance?

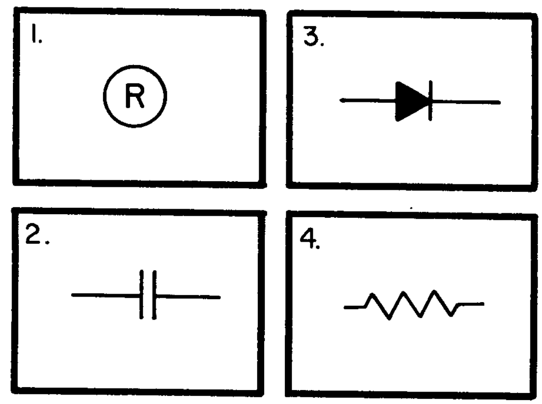
1. Siemens
2. S
3. G
4. Ohm

1-58. Resistance bears which, if any, of the following relationships to conductance?

1. A direct relationship
2. A reciprocal relationship
3. An inverse square relationship
4. None

THIS SPACE LEFT BLANK INTENTIONALLY.

1-59. Which of the following schematic symbols is used to represent a resistor?



1-60. How is the ability of a resistor to dissipate heat indicated?

1. By the wattage rating
2. By the voltage rating
3. By the resistance rating
4. By the tolerance

1-61. Carbon resistors have which of the following disadvantages?

1. A high cost factor
2. An extremely large physical size
3. The resistance value changes with age
4. A limited range of resistance values

1-62. Which of the following types of resistors will overcome the disadvantages of a carbon resistor?

1. Rheostat
2. Potentiometer
3. Molded composition
4. Wirewound resistor

1-63. What is the total number of connections on (a) a rheostat and (b) a potentiometer?

1. (a) Two (b) two
2. (a) Two (b) three
3. (a) Three (b) two
4. (a) Three (b) three

1-64. Which, if any, of the following types of variable resistors is used to control a large amount of current?

1. Rheostat
2. Potentiometer
3. Wirewound potentiometer
4. None of the above

1-65. A carbon resistor is color-coded orange, orange, orange. What is the resistance value of this resistor?

1. 2.2 kΩ
2. 3.3 kΩ
3. 33.0 kΩ
4. 440.0 kΩ

1-66. What are the allowable limits of ohmic value in a resistor color coded blue, green, yellow, gold?

1. 682.5 kΩ to 617.5 kΩ
2. 715.0 kΩ to 585.0 kΩ
3. 7.98 MΩ to 7.22 MΩ
4. 8.36 MΩ to 6.84 MΩ

1-67. Of the following, which color of the fifth band on a resistor indicates the LEAST chance of failure?

1. Red
2. Brown
3. Yellow
4. Orange

THIS SPACE LEFT BLANK INTENTIONALLY.

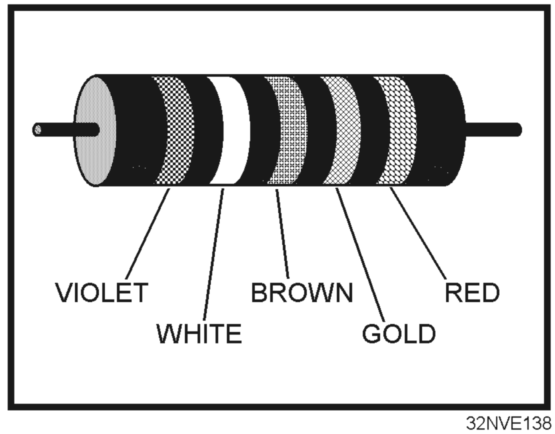


Figure 1A.—Resistor with color coding.

IN ANSWERING QUESTIONS 1-68 THROUGH 1-70, REFER TO FIGURE 1A.

1-68. What is the ohmic value of the resistor?

1. 8Ω
2. 79Ω
3. 790Ω
4. 800Ω

1-69. What is the specified tolerance of the resistor?

1. 1%
2. 5%
3. 10%
4. 20%

1-70. What is the specified reliability of the resistor?

1. 1.0%
2. 0.1%
3. 0.01%
4. 0.001%

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Batteries," pages 2-1 through 2-29.

2-1. Which of the following is the purpose of an electrical cell?

1. To change mechanical energy to electrical energy
2. To change chemical energy to electrical energy
3. To change electrical energy to mechanical energy
4. To change electrical energy to chemical energy

2-2. What are the three basic parts of a cell?

1. Electrodes, electrolyte, container
2. Electrodes, acid, water
3. Anode, cathode, ions
4. Anode, load, depolarizer

IN ANSWERING QUESTIONS 2-3 THROUGH 2-6, SELECT THE PHRASE FROM THE FOLLOWING LIST THAT DESCRIBES THE PART OF A CELL IN THE QUESTION.

A. PARTS OF A CELL	B. DESCRIPTIVE PHRASE
2-3. Electrolyte	1. negative electrode
2-4. Container	2. positive electrode
2-5. Anode	3. solution acting upon the electrode
2-6. Cathode	4. mounting for the electrode

2-7. What term is given to the process that takes place inside a cell?

1. Electromagnetic action
2. Piezoelectric action
3. Electromechanical action
4. Electrochemical action

2-8. With respect to recharging a primary or secondary cell, of the following statements, which one is correct?

1. The secondary cell can be recharged by passing current through it in the proper direction
2. The primary cell can be recharged by passing current through it in the proper direction
3. The secondary cell can only be recharged by changing the electrodes
4. The primary cell can only be recharged by changing the electrolyte

2-9. What determines the amount of current that a cell can deliver to the external circuit?

1. The internal resistance of the cell only
2. The resistance of the external load only
3. The circuit resistance and the internal resistance of the cell
4. The circuit capacitance and number of free electrons in the load

2-10. Which of the following actions will lower the internal resistance of a cell?

1. Decreasing the size of the electrodes
2. Increasing the size of the electrodes
3. Increasing the spacing between the electrodes
4. Increasing the resistance of the electrolyte

2-11. What causes negative ions to be attracted to the cathode of a primary cell while the cell is discharging?

1. A negative charge caused by a loss of electrons
2. A negative charge caused by an excess of electrons
3. A positive charge caused by a loss of electrons
4. A positive charge caused by an excess of electrons

2-12. What causes hydrogen to be attracted to the anode of a primary cell when the cell is discharging?

1. A negative charge caused by a loss of electrons
2. A negative charge caused by an excess of electrons
3. A positive charge caused by a loss of electrons
4. A positive charge caused by an excess of electrons

2-13. What causes the cathode to be "eaten away" in the primary cell while the cell is discharging?

1. The material of the cathode combines with the negative ions to form a new substance.
2. The material of the cathode dissolves in the electrolyte.
3. The material of the cathode leaves the negative terminal of the cell and goes through the load to the anode.
4. Bacteria in the electrolyte erodes the material in the cathode.

2-14. The primary cell is completely discharged when which of the following conditions exists?

1. The cathode is completely eaten away
2. The active ingredient in the electrolyte is used up
3. The voltage of the cell is reduced to zero
4. Each of the above

2-15. In a zinc-carbon primary cell, what is the function of the carbon electrode?

1. To generate electrons
2. To supply a return path for current
3. To speed electrolysis
4. To collect hydrogen

2-16. The lead-acid cell is an example of which of the following types of cells?

1. The dry cell
2. The voltaic cell
3. The primary cell
4. The secondary cell

2-17. In a fully charged lead-acid cell, what is the composition of the anode, cathode, and electrolyte respectively?

1. Zinc, carbon, and water
2. Carbon, lead, sulfuric acid and water
3. Lead peroxide, sponge lead, sulfuric acid, and water
4. Nickel, cadmium, potassium hydroxide, and water

2-18. Which of the following actions will recharge a secondary cell?

1. Adding more water to the electrolyte
2. Adding more active ingredient to the electrolyte
3. Connecting the negative terminal of a voltage source to the cathode of the cell and the positive terminal of the voltage source to the anode of the cell
4. Connecting the negative terminal of a voltage source to the anode of the cell and the positive terminal of the voltage source to the cathode of the cell

A. Sulfuric acid decreasing
B. Sulfuric acid increasing
C. Sponge lead decreasing
D. Sponge lead increasing
E. Lead peroxide decreasing
F. Lead peroxide increasing
G. Lead sulfate decreasing
H. Lead sulfate increasing

Figure 2A.—Lead acid chemical actions.

IN ANSWERING QUESTIONS 2-19 AND 2-20, REFER TO FIGURE 2A. SELECT THE CORRECT CHEMICAL ACTIONS WITHIN A LEAD-ACID CELL FOR THE CONDITION STATED IN EACH QUESTION.

2-19. The cell is discharging.

1. A, C, E, H
2. A, D, E, G
3. B, C, F, G
4. B, D, F, H

2-20. The cell is charging.

1. A, C, F, H
2. B, C, F, H
3. A, D, F, G
4. B, D, F, G

2-21. When all the lead sulfate in a lead-acid cell is converted to sulfuric acid, lead peroxide, and sponge lead, what is the condition of the cell?

1. Fully charged
2. Discharged
3. Sulfated
4. Unusable

2-22. Polarization has what effects on an electrical cell?

1. Decreases internal resistance, thereby increasing the output voltage
2. Decreases internal resistance, thereby decreasing the output voltage
3. Increases internal resistance, thereby increasing the output voltage
4. Increases internal resistance, thereby decreasing the output voltage

2-23. Which of the following methods is used to control polarization in a cell?

1. Venting the cell
2. Heating the electrolyte
3. Adding mercury to the electrode material
4. Using an electrolyte that absorbs oxygen

2-24. Which of the following is caused by local action in a cell?

1. Shelf life is reduced
2. Hydrogen is generated in large quantities
3. Impurities rise to the surface of the electrolyte
4. Mercury coating of the zinc electrode is worn away

2-25. In a dry cell, what is the consistency of the electrolyte?

1. Solid
2. Liquid
3. Paste
4. Powder

2-26. What serves as the cathode in a common type of dry cell?

1. Carbon electrode
2. Zinc container
3. Steel cover
4. Nickel terminal

2-27. How should the dry cell be stored to obtain maximum shelf life?

1. In a dark container
2. In a heated cabinet
3. In a ventilated area
4. In a refrigerated space

2-28. The blotting paper in a dry cell serves which of the following purposes?

1. Separates the paste from the zinc
2. Permits the electrolyte from the paste to filter through to the zinc slowly
3. Both 1 and 2 above
4. Keeps the electrolyte dry

2-29. Of the following characteristics, which one describes the mercury cell?

1. It is physically one of the largest cells
2. It has a very stable output voltage
3. It is designed to be rechargeable
4. It produces a large amount of current but has a short shelf life

2-30. Which of the following describes the shorting of a cell?

1. Decreasing the length of a cell
2. Connecting the anode and cathode together without a load
3. Using the cell below its full potential
4. Providing a recharge voltage that is not sufficient to recharge the cell

2-31. What is/are the advantages(s) of using a manganese-dioxide-alkaline- zinc cell over the zinc-carbon cell?

1. Better voltage stability
2. Longer storage life
3. Operates over a wide temperature range
4. All the above

2-32. What is the common name for manganese-dioxide-alkaline-zinc cell?

1. Alkaline cell
2. Long-life cell
3. Moz cell
4. Manganese-dioxide cell

2-33. Which of the following factors should be considered when selecting a primary cell as a power source?

1. Power requirement
2. Type of electrolyte used
3. Container material
4. All of the above

2-34. Of the following types of cells, which one is a primary cell?

1. Nickel cadmium
2. Silver zinc
3. Lithium organic
4. Silver cadmium

2-35. Which of the following is/are the difference(s) in the construction of a NICAD cell as compared to a lead-acid cell?

1. The electrolyte used
2. The material of the anode
3. The material of the cathode
4. All of the above

2-36. What is the most common use of a silver-zinc cell?

1. Flashlight batteries
2. Automobile batteries
3. Aircraft storage batteries
4. Emergency equipment batteries

2-37. In addition to the nickel-cadmium and silver-zinc cells, which of the following cells uses potassium hydroxide as the active ingredient in the electrolyte?

1. Lead-acid cell
2. Silver-cadmium
3. Lithium-inorganic cell
4. Magnesium-manganese dioxide cell

2-38. What is the minimum number of cells necessary to form a battery?

1. One
2. Two
3. Three
4. Four

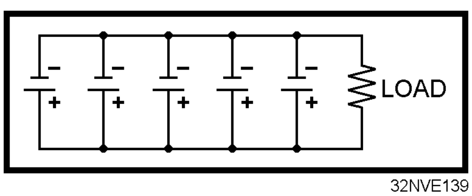


Figure 2B.—Battery consisting of five cells.

IN ANSWERING QUESTIONS 2-39 AND 2-40, REFER TO FIGURE 2B. EACH CELL IS 1.5 VOLTS AND HAS A CAPACITY OF 1/8 AMPERE.

2-39. What type of connection is used to combine the cells?

1. Series
2. Parallel
3. Series-parallel

2-40. What is the (a) voltage output and (b) current capacity of the circuit?

1. (a) 1.5 volts (b) 1/8 ampere
2. (a) 1.5 volts (b) 5/8 ampere
3. (a) 7.5 volts (b) 1/8 ampere
4. (a) 7.5 volts (b) 5/8 ampere

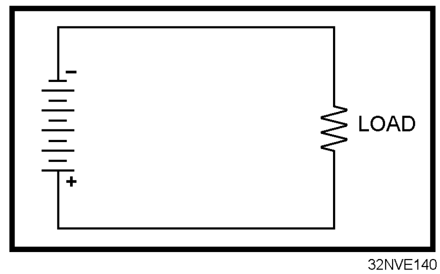


Figure 2C.—Five cells connected to form a battery.

IN ANSWERING QUESTIONS 2-41 AND 2-42, REFER TO FIGURE 2C. EACH CELL IS 1.5 VOLTS AND HAS A CAPACITY OF 1/8 AMPERE.

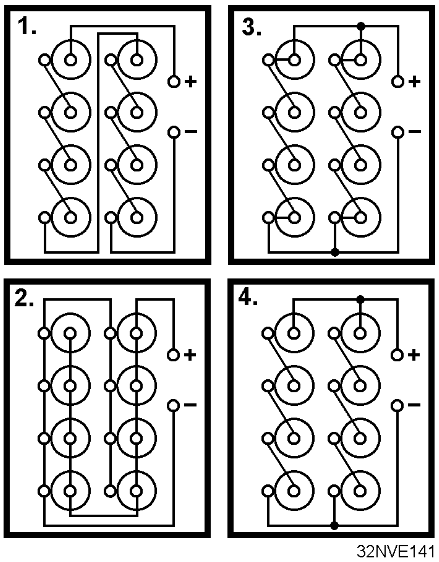
2-41. What type of connection is used to combine the cells?

1. Series
2. Parallel
3. Series-parallel

2-42. What is the (a) voltage output and (b) current capacity of the circuit?

1. (a) 1.5 volts (b) 1/8 ampere
2. (a) 1.5 volts (b) 5/8 ampere
3. (a) 7.5 volts (b) 1/8 ampere
4. (a) 7.5 volts (b) 5/8 ampere

2-43. Which of the following diagrams shows the proper connections for obtaining 6 volts at 1/4 ampere? (Each cell is 1.5 volts and has a capacity of 1/8 amp.)



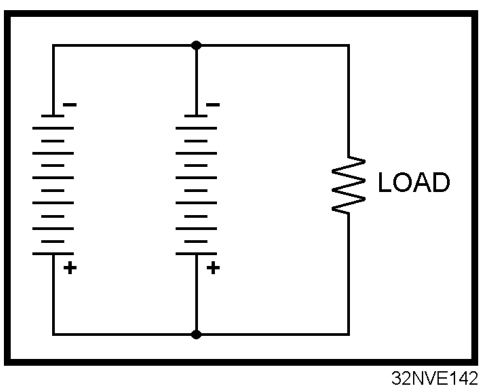


Figure 2D.—Battery consisting of 12 cells.

IN ANSWERING QUESTIONS 2-44 AND 2-45, REFER TO FIGURE 2D. EACH CELL EQUALS 1.5 VOLTS AND HAS A CAPACITY OF 1/8 AMPERE.

2-44. What type of connection is used to combine the cells?

1. Series
2. Parallel
3. Series-parallel

2-45. What is the (a) voltage output and (b) current capacity of the circuit?

1. (a) 1.5 volts (b) 1.5 amperes
2. (a) 4.5 volts (b) 1/2 ampere
3. (a) 9 volts (b) 1/4 ampere
4. (a) 18 volts (b) 1/8 ampere

2-46. What is the first step in performing maintenance on a secondary-cell battery?

1. Check the level of the electrolyte
2. Check the technical manual for information on the specific type of battery
3. Check the terminals for cleanliness and good electrical connection
4. Check the battery case for cleanliness and evidence of damage

2-47. When a hydrometer is used to check the specific gravity of the electrolyte in a battery, to what level should the electrolyte be drawn?

1. Enough to just wet the float
2. Enough so the float will rise without entering the suction bulb
3. Enough so the top one-third of the float will rise into the suction bulb
4. Enough so the float is completely covered by the electrolyte

2-48. To flush a hydrometer, which of the following liquids should be used?

1. Sulfuric acid
2. Salt water
3. Fresh water
4. A solution of baking soda and water

2-49. If the electrolyte level in a battery is low, what should be added to the electrolyte to bring it to the proper level?

1. Tap water
2. Sulfuric acid
3. Potassium hydroxide
4. Distilled water

2-50. Which one of the following safety precautions for batteries is NOT correct?

1. Terminals should be electrically connected together before transporting a battery
2. Care should be taken to prevent the spilling of electrolyte
3. Smoking, open flames, and electrical sparks are prohibited around charging batteries
4. Protective clothing, such as rubber apron, rubber gloves, and face shield, should be worn when working on batteries

2-51. If electrolyte comes in contact with the skin, what first aid treatment should be given immediately to the affected area?

1. Cover with petroleum jelly
2. Wrap with a sterile bandage
3. Apply an antiseptic lotion
4. Flush with fresh water

2-52. A battery with a capacity of 600 ampere-hours should provide 3 amperes for a maximum of how many hours?

1. 100 hr
2. 200 hr
3. 300 hr
4. 600 hr

2-53. A battery is rated according to a 20-hour rate of discharge at 300 ampere-hours. Which of the following currents is the maximum current that will allow the battery to deliver its rated capacity?

1. 15 amperes
2. 20 amperes
3. 25 amperes
4. 30 amperes

2-54. Which of the following types of routine charges follows the nameplate data in restoring a battery to its charged condition during the ordinary cycle of operation?

1. Initial
2. Floating
3. Normal
4. Fast

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 2-55 THROUGH 2-58, MATCH THE DESCRIPTION GIVEN IN THE FOLLOWING LIST WITH THE TYPE OF BATTERY CHARGE IN THE QUESTION.

A. TYPE OF CHARGE	B. DESCRIPTION
2-55. Initial charge	1. Used in emergency only
2-56. Equalizing charge	2. Used periodically as part of a maintenance routine
2-57. Floating charge	3. Used to keep a battery at full charge while the battery is idle
2-58. Fast charge	4. Used after electrolyte is added to a dry-shipped battery

2-59. If violent gassing occurs during the charging of a battery, which of the following actions should be taken?

1. Increase the room ventilation
2. Decrease the room temperature
3. Increase the charging rate
4. Decrease the charging rate

2-60. If a battery is being charged at the proper rate, which, if any of the following types of gassing should occur?

1. Steady gassing
2. Intermittent gassing
3. Violent gassing
4. None

ASSIGNMENT 3

Textbook assignment: Chapter 3, "Direct Current," pages 3-1 through 3-126.

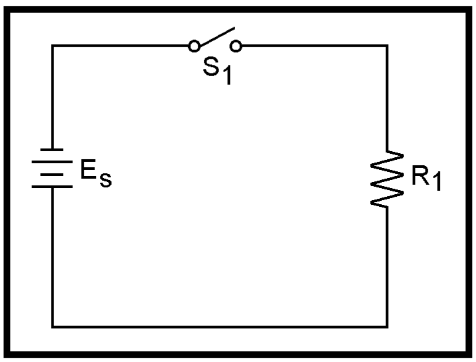


Figure 3A.—Basic circuit.

IN ANSWERING QUESTIONS 3-1 THROUGH 3-3, REFER TO FIGURE 3A.

3-1. What parts of the circuit represent the (a) source and (b) load?

1. (a) E_s (b) S_1
2. (a) E_s (b) R_1
3. (a) S_1 (b) R_1
4. (a) S_1 (b) E_s

3-2. Which of the following terms describes the circuit condition?

1. Partially shorted
2. Partially open
3. Shorted
4. Open

3-3. Which of the following terms describes the figure 3A?

1. Parts layout
2. Exploded view
3. Wiring diagram
4. Schematic diagram

3-4. If circuit voltage is held constant, circuit current will react in what manner as the resistance (a) increases, and (b) decreases?

1. (a) Increase (b) decrease
2. (a) Increase (b) increase
3. (a) Decrease (b) decrease
4. (a) Decrease (b) increase

3-5. If circuit resistance is held constant, circuit current will react in what manner as the voltage (a) increases, and (b) decreases?

1. (a) Increase (b) decrease
2. (a) Increase (b) increase
3. (a) Decrease (b) decrease
4. (a) Decrease (b) increase

3-6. According to Ohm's law, what formula should be used to calculate circuit voltage if resistance and current value are known?

1. $E = \frac{R}{I}$
2. $E = \frac{I}{R}$
3. $E = IR$
4. $E = \frac{I}{IR}$

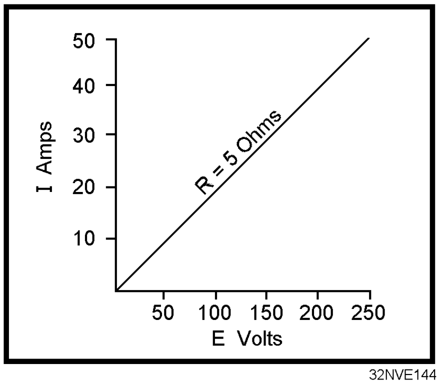


Figure 3B.—Graph of current and voltage.

IN ANSWERING QUESTIONS 3-7 AND 3-8, REFER TO FIGURE 3B.

3-7. If the current is 15 amperes, what is the value of the voltage?

1. 50 V
2. 75 V
3. 100 V
4. 150 V

3-8. If the voltage is 200 volts, what is the value of the current?

1. 10 A
2. 20 A
3. 30 A
4. 40 A

3-9. Which of the following terms applies to the rate at which an electrical force causes motion?

1. Power
2. Energy
3. Inertia
4. Each of the above

3-10. Which of the following circuit quantities can be varied ONLY by varying one of the other circuit quantities?

1. Voltage
2. Current
3. Resistance
4. Each of the above

3-11. Which of the following is a correct formula for determining power in an electrical circuit?

1. $P = EI$
2. $P = I^2R$
3. $P = \frac{E^2}{R}$
4. Each of the above

3-12. What is the current in a circuit with 15 ohms of resistance that uses 135 watts of power?

1. 10 A
2. 15 A
3. 3 A
4. 9 A

3-13. What is the total power used by a 15-ohm resistor with 4 amps of current?

1. 60 W
2. 240 W
3. 360 W
4. 900 W

3-14. What type of resistor should be used in question 3-13?

1. Carbon
2. Wirewound
3. Precision
4. Composition

3-15. How much total energy is converted by a 1-horsepower motor in 10 hours?

1. 7.46 kWh
2. 8.32 kWh
3. 8.59 kWh
4. 9.32 kWh

3-16. If the energy used by the motor in question 3-15 is 9.5 kWh, what is the efficiency of the motor?

1. .981
2. .904
3. .876
4. .785

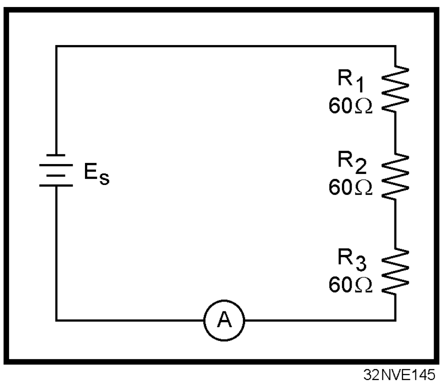


Figure 3C.—Series circuit.

IN ANSWERING QUESTIONS 3-17 THROUGH 3-23, REFER TO FIGURE 3C.

3-17. What is the total circuit resistance (R)?

1. 20Ω
2. 60Ω
3. 180Ω
4. 240Ω

3-18. If the circuit current is 3 amps, what is the source voltage (E_s)?

1. 60 V
2. 180 V
3. 540 V
4. 720 V

3-19. What is the total voltage dropped by each resistor in question 3-18?

1. 20 V
2. 60 V
3. 180 V
4. 540 V

3-20. If the current decreases to 2 amps, what is the total voltage drop across each resistor?

1. 120 V
2. 230 V
3. 310 V
4. 400 V

3-21. What would have to be done to the circuit to cause the current to decrease to 2 amps?

1. The source voltage would have to be increased
2. The source voltage would have to be decreased
3. The resistance of R_1 would have to be decreased
4. One of the resistors would have to be removed from the circuit

3-22. If the circuit current is 2 amps, what is the total power used by each resistor?

1. 240 W
2. 460 W
3. 620 W
4. 800 W

3-23. What is the total power used in the circuit if E_s = 360 V?

1. 720 W
2. 1380 W
3. 1860 W
4. 2400 W

3-24. When Kirchoff's voltage law is used to assign polarities to the voltage drop across a resistor, which of the following references is used to indicate the end of the resistor that the current enters?

1. Ground
2. Neutral
3. Negative
4. Positive

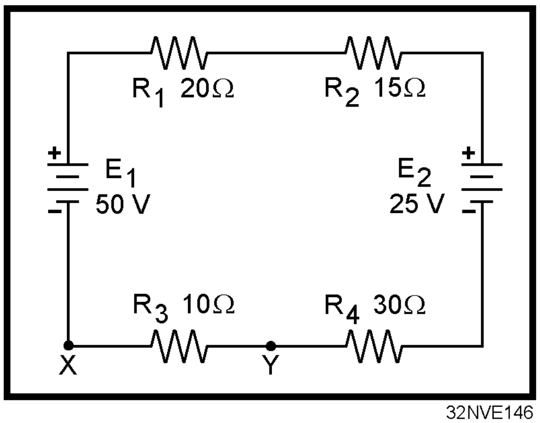


Figure 3D.—Multiple source circuit.

IN ANSWERING QUESTIONS 3-25 AND 3-26, REFER TO FIGURE 3D.

3-25. What is the effective source voltage?

1. 15 V
2. 25 V
3. 50 V
4. 75 V

3-26. What is the total amount and direction of current through R_3?

1. 1.0 A from Y to X
2. 1.0 A from X to Y
3. .33 A from Y to X
4. .33 A from X to Y

3-27. Which of the following terms applies to a circuit in which there is NO complete path for current?

1. Open
2. Short
3. Closed
4. Grounded

3-28. A circuit in which the resistance is almost zero ohms is referred to by which of the following terms?

1. Open
2. Short
3. Closed
4. Broken

THIS SPACE LEFT BLANK INTENTIONALLY.

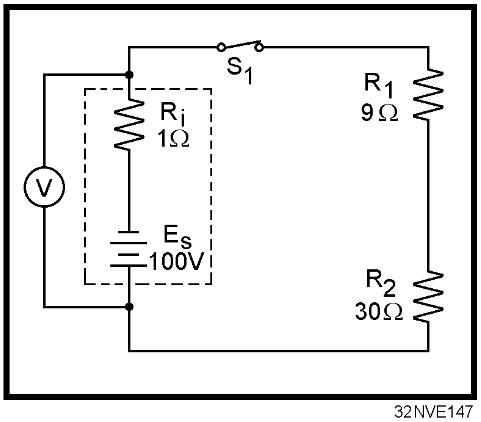


Figure 3E.—Series circuit and source resistance.

IN ANSWERING QUESTIONS 3-29 THROUGH 3-32, REFER TO FIGURE 3E.

3-29. If R_2 has a short circuit, what will most likely happen to the circuit?

1. R_1 will be destroyed
2. E_s will increase
3. V will indicate O volts
4. S_1 will automatically open

3-30. What is the total voltage drop across R_i when the switch is closed?

1. 2.5 V
2. 6.5 V
3. 97.5 V
4. 100.0 V

3-31. What will the meter indicate with (a) S_1 open, and (b) S_1 closed?

1. (a) 100 V (b) 100 V
2. (a) 97.5 V (b) 100 V
3. (a) 100 V (b) 97.5 V
4. (a) 97.5 V (b) 97.5 V

3-32. To achieve maximum power transfer in the circuit, which of the following conditions must be met?

1. $R_i = R_L$
2. $I_s = I_L$
3. $E_s = E_L$
4. $K_s = K_L$

3-33. Maximum power is transferred from a source to a load when the value of the load resistance is of what value when compared to the source resistance?

1. Equal
2. Twice
3. One-half
4. Several times

3-34. When maximum power is transferred from a source to a load, what is the efficiency of power transfer?

1. 5%
2. 25%
3. 50%
4. 95%

3-35. A circuit consists of three resistors connected in parallel. R_1 = 30 ohms, R_2 = 15 ohms, and R_3 = 10 ohms. If the current through R_2 = 4 amperes, what is the total source voltage?

1. 20 V
2. 60 V
3. 120 V
4. 220 V

3-36. What is the relationship of total current to the current through a component in (a) a series circuit, and (b) a parallel circuit?

1. (a) Divides (b) divides
2. (a) Divides (b) equals
3. (a) Equals (b) equals
4. (a) Equals (b) divides

3-37. If a current has a negative polarity when Kirchoff's current law is applied, which of the following, statements is true of the current?

1. It is from a battery
2. It is from a generator
3. It is entering a junction
4. It is leaving a junction

3-38. Three equal resistors are connected in parallel and each resistor has an ohmic value of 300 ohms. What is the equivalent resistance of the circuit?

1. 100Ω
2. 150Ω
3. 600Ω
4. 900Ω

3-39. Three resistors with ohmic values of 120 ohms, 60 ohms, and 40 ohms are connected in parallel. What is the equivalent resistance of the circuit?

1. 10Ω
2. 20Ω
3. 30Ω
4. 40Ω

3-40. Two resistors with ohmic values of 90 ohms and 45 ohms are connected in parallel. What is the equivalent resistance of the circuit?

1. 10Ω
2. 20Ω
3. 30Ω
4. 40Ω

3-41. Which of the following terms describes a single resistor that represents a complex circuit?

1. Equal resistor
2. Phantom resistor
3. Schematic resistor
4. Equivalent resistor

THIS SPACE LEFT BLANK INTENTIONALLY.

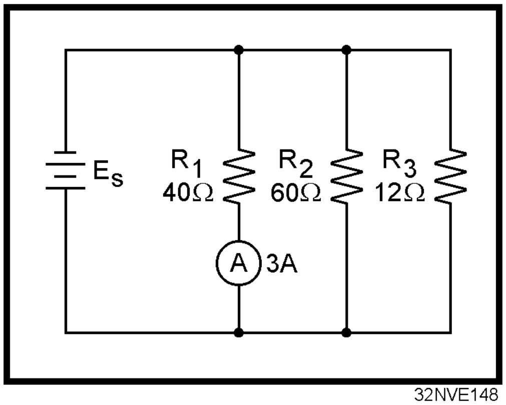


Figure 3F.—Parallel circuit.

IN ANSWERING QUESTIONS 3-42 THROUGH 3-46, REFER TO FIGURE 3F.

3-42. What is the value of E_s?

1. 336 V
2. 300 V
3. 240 V
4. 120 V

3-43. What is the value of current through R_2?

1. 1 A
2. 2 A
3. 3 A
4. 4 A

3-44. What is the approximate value of total resistance?

1. 8Ω
2. 37Ω
3. 112Ω
4. 257Ω

3-45. What is the value of total power?

1. 1.2 kW
2. 1.5 kW
3. 1.8 kW
4. 2.0 kW

3-46. What is the total power consumed by R_3?

1. 108 W
2. 240 W
3. 360 W
4. 1200 W

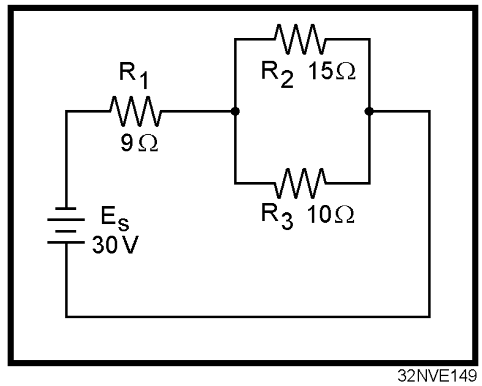


Figure 3G.—Series-parallel circuit.

IN ANSWERING QUESTIONS 3-47 THROUGH 3-49, REFER TO FIGURE 3G.

3-47. What is the value of the total resistance?

1. 3.6Ω
2. 15Ω
3. 34Ω
4. 40Ω

3-48. What is the total power used in the circuit?

1. 22.5 W
2. 26.5 W
3. 60.0 W
4. 250.0 W

3-49. What is the total voltage drop across R_3?

1. 8 V
2. 12 V
3. 18 V
4. 30 V

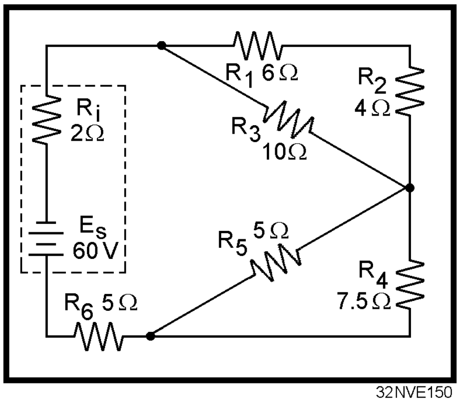


Figure 3H.—Complex circuit.

IN ANSWERING QUESTIONS 3-50 AND 3-51, REFER TO FIGURE 3H.

3-50. What is the value of total resistance?

1. 5Ω
2. 8Ω
3. 13Ω
4. 15Ω

3-51. If an equivalent resistor is used to represent the network of R_1, R_2, R_3, R_4, R_5, and R_6, what is the total voltage drop across this resistor?

1. 8V
2. 26V
3. 52V
4. 60V

3-52. If an open occurs in a series portion of a circuit, what is the effect on (a) total resistance, and (b) total current?

1. (a) Decreases to zero
   (b) Becomes infinite
2. (a) Decreases to zero
   (b) Decreases to zero
3. (a) Becomes infinite
   (b) Becomes infinite
4. (a) Becomes infinite
   (b) Decreases to zero

3-53. If an open occurs in a parallel branch of a circuit, what is the effect on (a) total resistance, and (b) total current?

1. (a) Increases (b) decreases
2. (a) Increases (b) increases
3. (a) Decreases (b) decreases
4. (a) Decreases (b) increases

3-54. If a short circuit occurs in a series portion of a circuit, what is the effect on (a) total resistance, and (b) total current?

1. (a) Increases (b) decreases
2. (a) Increases (b) increases
3. (a) Decreases (b) decreases
4. (a) Decreases (b) increases

3-55. If a short circuit occurs in a parallel branch of a circuit, what is the effect in (a) total resistance, and (b) total current?

1. (a) Increases (b) decreases
2. (a) Increases (b) increases
3. (a) Decreases (b) decreases
4. (a) Decreases (b) increases

3-56. If one branch of a parallel network shorts, what portion of the circuit current, if any, will flow through the remaining branches?

1. An amount determined by the combined resistance of the remaining branches
2. All
3. One-half
4. None

3-57. Which of the following circuit quantities need NOT be known before designing a voltage divider?

1. The current of the source
2. The voltage of the source
3. The current requirement of the load
4. The voltage requirement of the load

THE FOLLOWING INFORMATION IS TO BE USED IN ANSWERING QUESTIONS 3-58 THROUGH 3-60: A VOLTAGE DIVIDER IS REQUIRED TO SUPPLY A SINGLE LOAD WITH +150 VOLTS AND 300 MILLIAMPS OF CURRENT. THE SOURCE VOLTAGE IS 250 VOLTS. (HINT: DRAW THE CIRCUIT.)

3-58. What should be the value of the bleeder current?

1. 3 A
2. 300 mA
3. 30 mA
4. 3 mA

3-59. What should be the ohmic value of the bleeder resistor?

1. 50
2. 500
3. 5 k
4. 50 k

3-60. What is the value of total current?

1. 303 mA
2. 330 mA
3. 600 mA
4. 3300 mA

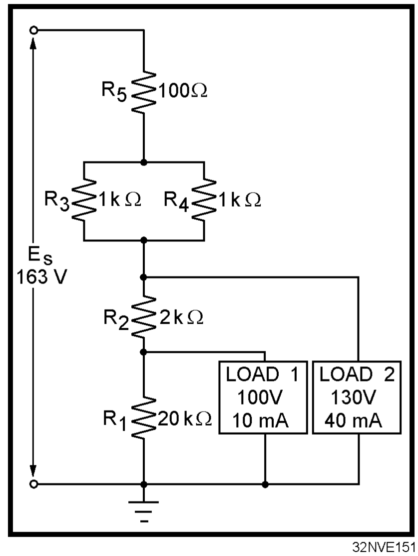


Figure 3I.—Voltage divider.

IN ANSWERING QUESTIONS 3-61 THROUGH 3-66, REFER TO FIGURE 3I.

3-61. Why must the value of R_1 be calculated first?

1. For convenience
2. The current through R_2 depends on the value of R_1
3. The voltage drop across R_1depends on the value of load 1
4. In any circuit, values for resistors labeled R_1 are calculated first

3-62. How is the current through R_2 calculated?

1. By adding I_{R1} and the current requirement of load 1
2. By adding the current requirements of load 1 and load 2
3. By subtracting the current requirement of load 1 from the current requirement of load 2
4. By subtracting the current requirement of load 2 from the current requirement of load 1

3-63. How is the voltage drop across R_2 calculated?

1. By adding the voltage requirements of load 1 and load 2
2. By subtracting the voltage drops across R_5 and R_3 from the source voltage
3. By subtracting the voltage requirement of load 1 from the voltage requirement of load 2
4. By subtracting the voltage requirements of load 1 and load 2 from the source voltage

3-64. What is the minimum wattage rating required for R_5?

1. 1 W
2. 2 W
3. 1/2 W
4. 1/4 W

3-65. What is the total power supplied by the source?

1. 3.765 W
2. 7.965 W
3. 8.209 W
4. 8.965 W

3-66. What is the purpose of using the series-parallel network consisting of R_3, R_4, and R_5 in place of a single resistor?

1. It provides the desired resistance with resistor values that are easily obtainable
2. It provides the close tolerance required for the circuit
3. It is more reliable than the use of a single resistor
4. It costs less by using three resistors of lower wattage rating than a single, large power resistor

3-67. A single voltage divider provides both negative and positive voltages from a single source voltage through the use of a

1. ground between two of the dividing resistors
2. ground to the positive terminal of the source
3. ground to the negative terminal of the source
4. ground to the input of all loads requiring a negative voltage

3-68. Which of the following voltages are considered dangerous?

1. Voltages above 115 volts only
2. Voltages above 230 volts only
3. Voltages above 450 volts only
4. All voltages

3-69. If you discover a possible malfunction in an electric circuit, which of the following actions should be taken?

1. Attempt repairs yourself
2. Report the malfunction to a qualified technician
3. Ignore the malfunction unless you were assigned to repair it
4. Secure the circuit immediately by removing power at the nearest switch

3-70. If a person has stopped breathing and there is NO detectable heartbeat, who should perform CPR?

1. Medical personnel only
2. The first person on the scene
3. Emergency Medical Technicians only
4. Trained, qualified personnel only

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 2—Introduction to Alternating Current and Transformers

NAVEDTRA 14174

Sailor's Creed

"I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country's Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all."

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: Subjects of alternating current, inductance, capacitance, inductive and capacitive reactance, and transformers are presented to give the student background information on topics that may be encountered in daily work.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and the occupational standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards*, NAVPERS 18068, found on line at https://buperscd.technology.navy.mil /bup_updt/upd_CD/BUPERS/enlistedManOpen.htm.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
DSC Ray A. Jackson

Corrections and minor modifications prepared in
January 2003 by
ETC Craig Reidl

NAVSUP Logistics Tracking Number
0504-LP-026-8270

TABLE OF CONTENTS

Assignments follow Appendix VI.

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 10 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

(THIS PAGE IS INTENTIONALLY LEFT BLANK.)

CHAPTER 1

CONCEPTS OF ALTERNATING CURRENT

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. State the differences between ac and dc voltage and current.
2. State the advantages of ac power transmission over dc power transmission.
3. State the "left-hand rule" for a conductor.
4. State the relationship between current and magnetism.
5. State the methods by which ac power can be generated.
6. State the relationship between frequency, period, time, and wavelength.
7. Compute peak-to-peak, instantaneous, effective, and average values of voltage and current.
8. Compute the phase difference between sine waves.

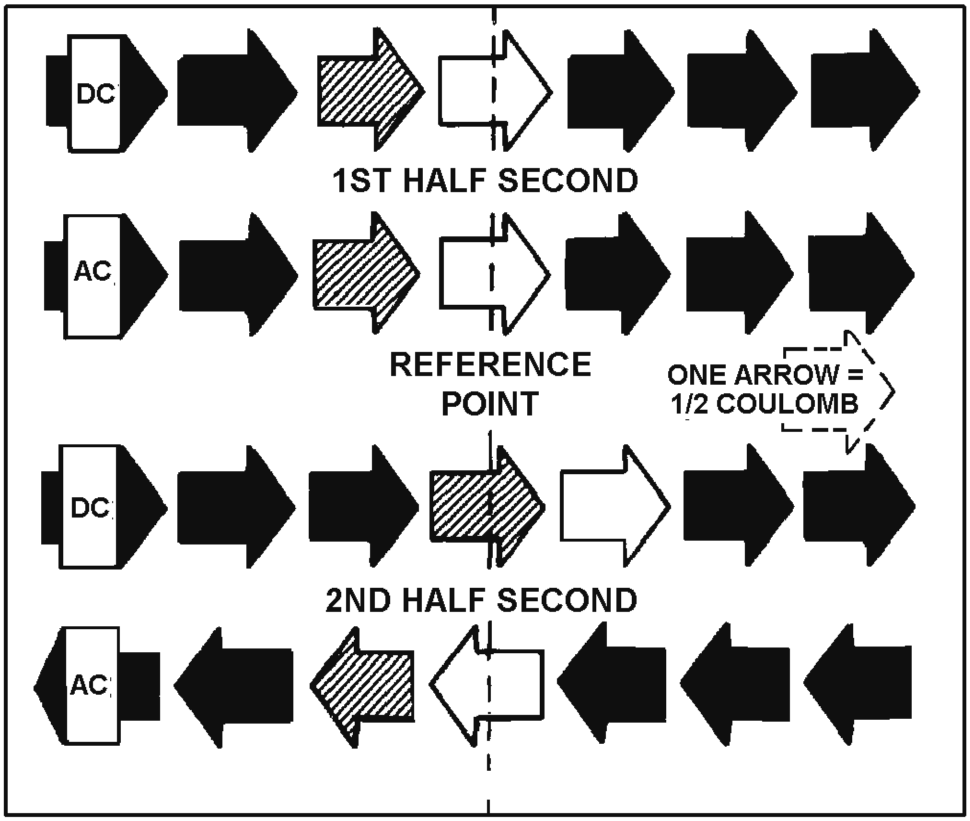
CONCEPTS OF ALTERNATING CURRENT

All of your study thus far has been with direct current (dc), that is, current which does not change direction. However, as you saw in module 1 and will see later in this module, a coil rotating in a magnetic field actually generates a current which regularly changes direction. This current is called ALTERNATING CURRENT or ac.

AC AND DC

Alternating current is current which constantly changes in amplitude, and which reverses direction at regular intervals. You learned previously that direct current flows only in one direction, and that the amplitude of current is determined by the number of electrons flowing past a point in a circuit in one second. If, for example, a coulomb of electrons moves past a point in a wire in one second and all of the electrons are moving in the same direction, the amplitude of direct current in the wire is one ampere. Similarly, if half a coulomb of electrons moves in one direction past a point in the wire in half a second, then reverses direction and moves past the same point in the opposite direction during the next half-second, a total of one coulomb of electrons passes the point in one second. The amplitude of the alternating current is one ampere. The preceding comparison of dc and ac as illustrated. Notice that one white arrow plus one striped arrow comprise one coulomb.

COMPARING DC & AC CURRENT FLOW IN A WIRE



Q1. Define direct current.

Q2. Define alternating current.

DISADVANTAGES OF DC COMPARED TO AC

When commercial use of electricity became wide-spread in the United States, certain disadvantages in using direct current in the home became apparent. If a commercial direct-current system is used, the voltage must be generated at the level (amplitude or value) required by the load. To properly light a 240-volt lamp, for example, the dc generator must deliver 240 volts. If a 120-volt lamp is to be supplied power from the 240-volt generator, a resistor or another 120-volt lamp must be placed in series with the 120-volt lamp to drop the extra 120 volts. When the resistor is used to reduce the voltage, an amount of power equal to that consumed by the lamp is wasted.

Another disadvantage of the direct-current system becomes evident when the direct current (I) from the generating station must be transmitted a long distance over wires to the consumer. When this happens, a large amount of power is lost due to the resistance (R) of the wire. The power loss is equal to I^2R. However, this loss can be greatly reduced if the power is transmitted over the lines at a very high voltage level and a low current level. This is not a practical solution to the power loss in the dc system since the load would then have to be operated at a dangerously high voltage. Because of the disadvantages related to transmitting and using direct current, practically all modern commercial electric power companies generate and distribute alternating current (ac).

Unlike direct voltages, alternating voltages can be stepped up or down in amplitude by a device called a TRANSFORMER. (The transformer will be explained later in this module.) Use of the transformer permits efficient transmission of electrical power over long-distance lines. At the electrical power station, the transformer output power is at high voltage and low current levels. At the consumer end of the transmission lines, the voltage is stepped down by a transformer to the value required by the load. Due to its inherent advantages and versatility, alternating current has replaced direct current in all but a few commercial power distribution systems.

Q3. What is a disadvantage of a direct-current system with respect to supply voltage?

Q4. What disadvantage of a direct current is due to the resistance of the transmission wires?

Q5. What kind of electrical current is used in most modern power distribution systems?

VOLTAGE WAVEFORMS

You now know that there are two types of current and voltage, that is, direct current and voltage and alternating current and voltage. If a graph is constructed showing the amplitude of a dc voltage across the terminals of a battery with respect to time, it will appear in figure 1-1 view A. The dc voltage is shown to have a constant amplitude. Some voltages go through periodic changes in amplitude like those shown in figure 1-1 view B. The pattern which results when these changes in amplitude with respect to time are plotted on graph paper is known as a WAVEFORM. Figure 1-1 view B shows some of the common electrical waveforms. Of those illustrated, the sine wave will be dealt with most often.

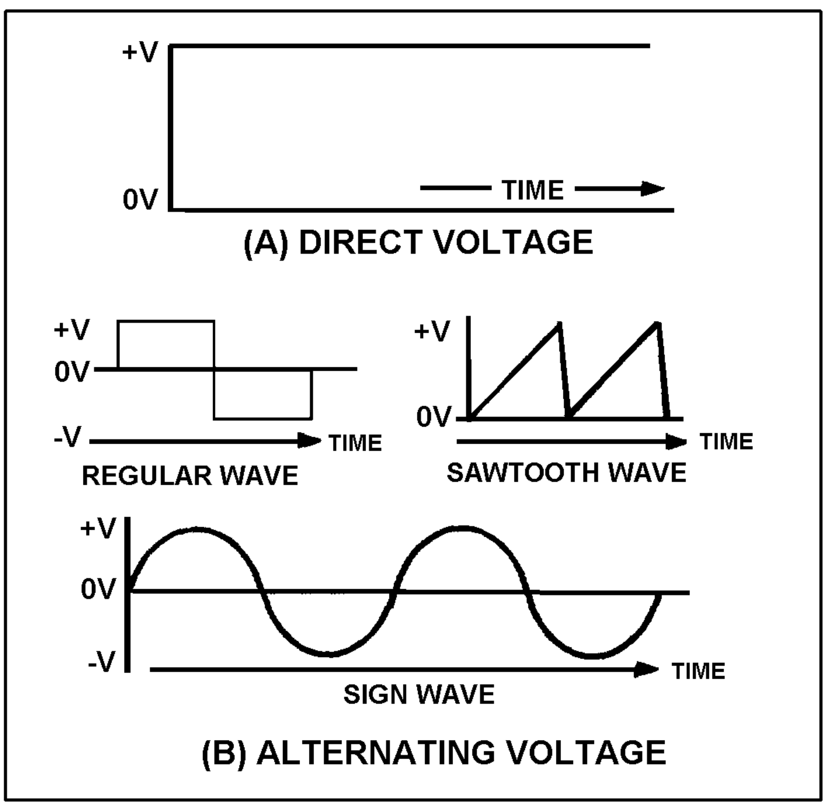


Figure 1-1.—Voltage waveforms: (A) Direct voltage; (B) Alternating voltage.

ELECTROMAGNETISM

The sine wave illustrated in figure 1-1 view B is a plot of a current which changes amplitude and direction. Although there are several ways of producing this current, the method based on the principles of electromagnetic induction is by far the easiest and most common method in use.

The fundamental theories concerning simple magnets and magnetism were discussed in Module 1, but how magnetism can be used to produce electricity was only briefly mentioned. This module will give you a more in-depth study of magnetism. The main points that will be explained are how magnetism is affected by an electric current and, conversely, how electricity is affected by magnetism. This general subject area is most often referred to as ELECTROMAGNETISM. To properly understand electricity you must first become familiar with the relationships between magnetism and electricity. For example, you must know that:

- An electric current always produces some form of magnetism.
- The most commonly used means for producing or using electricity involves magnetism.
- The peculiar behavior of electricity under certain conditions is caused by magnetic influences.

MAGNETIC FIELDS

In 1819 Hans Christian Oersted, a Danish physicist, found that a definite relationship exists between magnetism and electricity. He discovered that an electric current is always accompanied by certain magnetic effects and that these effects obey definite laws.

MAGNETIC FIELD AROUND A CURRENT-CARRYING CONDUCTOR

If a compass is placed in the vicinity of a current-carrying conductor, the compass needle will align itself at right angles to the conductor, thus indicating the presence of a magnetic force. You can demonstrate the presence of this force by using the arrangement illustrated in figure 1-2. In both (A) and (B) of the figure, current flows in a vertical conductor through a horizontal piece of cardboard. You can determine the direction of the magnetic force produced by the current by placing a compass at various points on the cardboard and noting the compass needle deflection. The direction of the magnetic force is assumed to be the direction in which the north pole of the compass points.

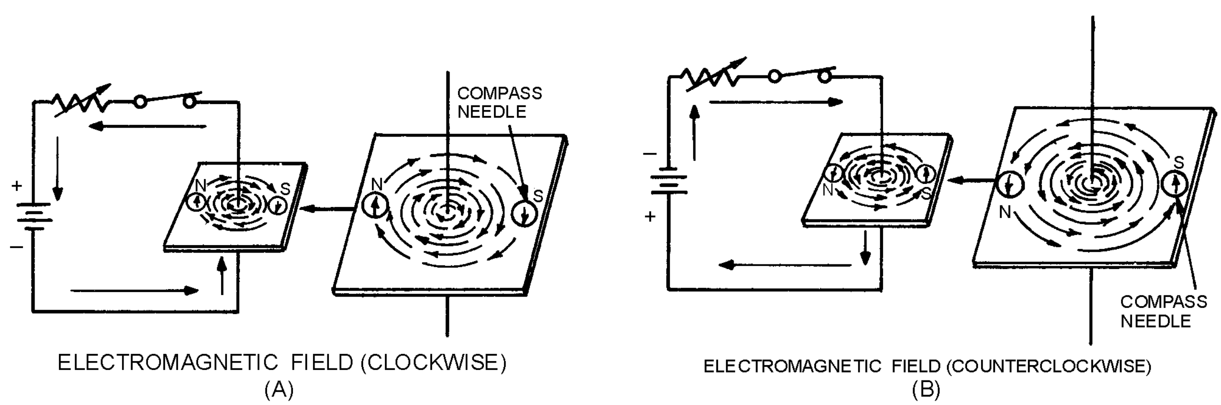


Figure 1-2.—Magnetic field around a current-carrying conductor.

In figure 1-2 (A), the needle deflections show that a magnetic field exists in circular form around the conductor. When the current flows upward (see figure 1-2(A)), the direction of the field is clockwise, as viewed from the top. However, if you reverse the polarity of the battery so that the current flows downward (see figure 1-2(B)), the direction of the field is counterclockwise.

The relation between the direction of the magnetic lines of force around a conductor and the direction of electron current flow in the conductor may be determined by means of the LEFT-HAND RULE FOR A CONDUCTOR: if you grasp the conductor in your left hand with the thumb extended in the direction of the electron flow (current) (– to +), your fingers will point in the direction of the magnetic lines of force. Now apply this rule to figure 1-2. Note that your fingers point in the direction that the north pole of the compass points when it is placed in the magnetic field surrounding the wire.

An arrow is generally used in electrical diagrams to denote the direction of current in a length of wire (see figure 1-3(A)). Where a cross section of a wire is shown, an end view of the arrow is used. A cross-sectional view of a conductor that is carrying current toward the observer is illustrated in figure 1-3(B). Notice that the direction of current is indicated by a dot, representing the head of the arrow. A conductor that is carrying current away from the observer is illustrated in figure 1-3(C). Note that the direction of current is indicated by a cross, representing the tail of the arrow. Also note that the magnetic field around a current-carrying conductor is perpendicular to the conductor, and that the magnetic lines of force are equal along all parts of the conductor.

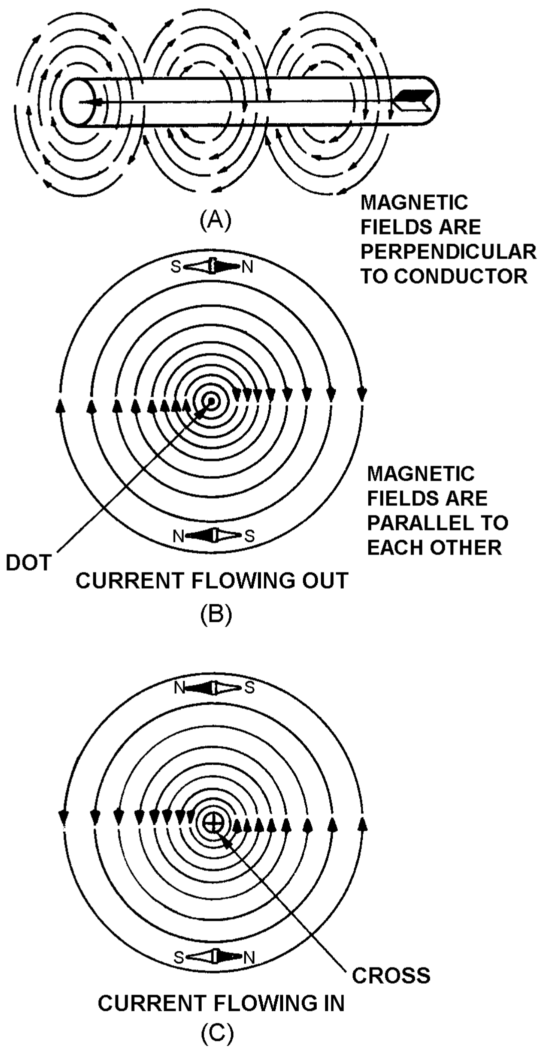
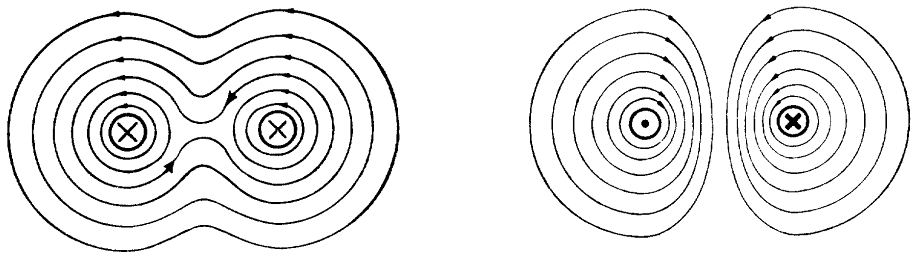


Figure 1-3.—Magnetic field around a current-carrying conductor, detailed view.

When two adjacent parallel conductors are carrying current in the same direction, the magnetic lines of force combine and increase the strength of the field around the conductors, as shown in figure 1-4(A). Two parallel conductors carrying currents in opposite directions are shown in figure 1-4(B). Note that the field around one conductor is opposite in direction to the field around the other conductor. The resulting lines of force oppose each other in the space between the wires, thus deforming the field around each conductor. This means that if two parallel and adjacent conductors are carrying currents in the same direction, the fields about the two conductors aid each other. Conversely, if the two conductors are carrying currents in opposite directions, the fields about the conductors repel each other.



CURRENT FLOWING IN
THE SAME DIRECTION
(A)

CURRENT FLOWING IN
OPPOSITE DIRECTION
(B)

Figure 1-4.—Magnetic field around two parallel conductors.

Q6. When placed in the vicinity of a current-carrying conductor, the needle of a compass becomes aligned at what angle to the conductor?

Q7. What is the direction of the magnetic field around a vertical conductor when (a) the current flows upward and (b) the current flows downward.

Q8. The "left-hand rule" for a conductor is used for what purpose

Q9. In what direction will the compass needle point when the compass is placed in the magnetic field surrounding a wire?

Q10. When two adjacent parallel wires carry current in the same direction, the magnetic field about one wire has what effect on the magnetic field about the other conductor?

Q11. When two adjacent parallel conductors carry current in opposite directions, the magnetic field about one conductor has what effect on the magnetic field about the other conductor?

MAGNETIC FIELD OF A COIL

Figure 1-3(A) illustrates that the magnetic field around a current-carrying wire exists at all points along the wire. Figure 1-5 illustrates that when a straight wire is wound around a core, it forms a coil and that the magnetic field about the core assumes a different shape. Figure 1-5(A) is actually a partial cutaway view showing the construction of a simple coil. Figure 1-5(B) shows a cross-sectional view of the same coil. Notice that the two ends of the coil are identified as X and Y.

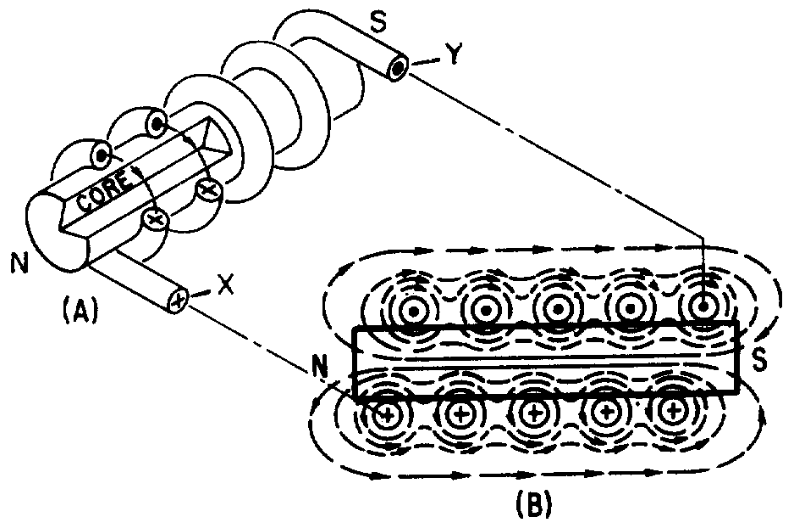


Figure 1-5.—Magnetic field produced by a current-carrying coil.

When current is passed through the coil, the magnetic field about each turn of wire links with the fields of the adjacent turns. (See figure 1-4(A)). The combined influence of all the turns produces a two-pole field similar to that of a simple bar magnet. One end of the coil is a north pole and the other end is a south pole.

Polarity of an Electromagnetic Coil

Figure 1-2 shows that the direction of the magnetic field around a straight wire depends on the direction of current in that wire. Thus, a reversal of current in a wire causes a reversal in the direction of the magnetic field that is produced. It follows that a reversal of the current in a coil also causes a reversal of the two-pole magnetic field about the coil.

When the direction of the current in a coil is known, you can determine the magnetic polarity of the coil by using the LEFT-HAND RULE FOR COILS. This rule, illustrated in figure 1-6, is stated as follows:

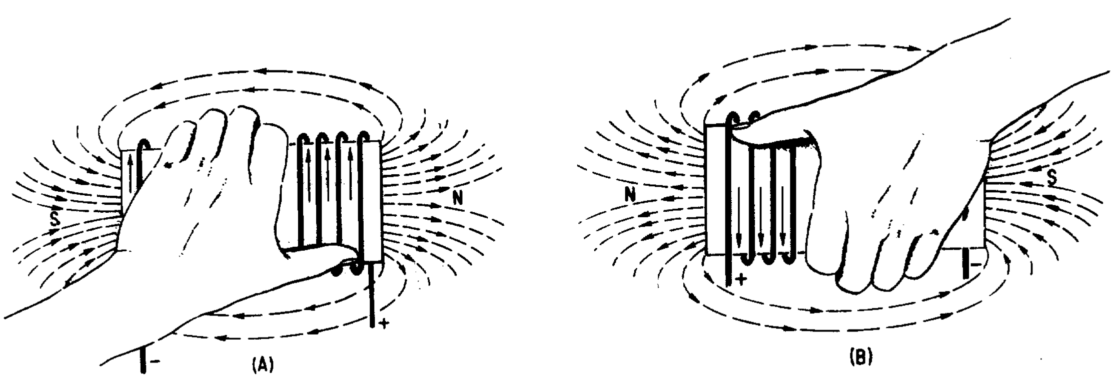


Figure 1-6.—Left-hand rule for coils.

Grasp the coil in your left hand, with your fingers "wrapped around" in the direction of the electron current flow. Your thumb will then point toward the north pole of the coil.

Strength of an Electromagnetic Field

The strength or intensity of a coil's magnetic field depends on a number of factors. The main ones are listed below and will be discussed again later.

- The number of turns of wire in the coil.
- The amount of current flowing in the coil.
- The ratio of the coil length to the coil width.
- The type of material in the core.

Losses in an Electromagnetic Field

When current flows in a conductor, the atoms in the conductor all line up in a definite direction, producing a magnetic field. When the direction of the current changes, the direction of the atoms' alignment also changes, causing the magnetic field to change direction. To reverse all the atoms requires that power be expended, and this power is lost. This loss of power (in the form of heat) is called HYSTERESIS LOSS. Hysteresis loss is common to all ac equipment; however, it causes few problems except in motors, generators, and transformers. When these devices are discussed later in this module, hysteresis loss will be covered in more detail.

Q12. What is the shape of the magnetic field that exists around (a) a straight conductor and (b) a coil?

Q13. What happens to the two-pole field of a coil when the current through the coil is reversed?

Q14. What rule is used to determine the polarity of a coil when the direction of the electron current flow in the coil is known?

Q15. State the rule whose purpose is described in Q14.

BASIC AC GENERATION

From the previous discussion you learned that a current-carrying conductor produces a magnetic field around itself. In module 1, under producing a voltage (emf) using magnetism, you learned how a changing magnetic field produces an emf in a conductor. That is, if a conductor is placed in a magnetic field, and either the field or the conductor moves, an emf is induced in the conductor. This effect is called electromagnetic induction.

CYCLE

Figures 1-7 and 1-8 show a suspended loop of wire (conductor) being rotated (moved) in a clockwise direction through the magnetic field between the poles of a permanent magnet. For ease of explanation, the loop has been divided into a dark half and light half. Notice in (A) of the figure that the dark half is moving along (parallel to) the lines of force. Consequently, it is cutting NO lines of force. The same is true of the light half, which is moving in the opposite direction. Since the conductors are cutting no lines of force, no emf is induced. As the loop rotates toward the position shown in (B), it cuts more and more lines of force per second (inducing an ever-increasing voltage) because it is cutting more directly across the field (lines of force). At (B), the conductor is shown completing one-quarter of a complete revolution, or 90º , of a complete circle. Because the conductor is now cutting directly across the field, the voltage induced in the conductor is maximum. When the value of induced voltage at various points during the rotation from (A) to (B) is plotted on a graph (and the points connected), a curve appears as shown below.

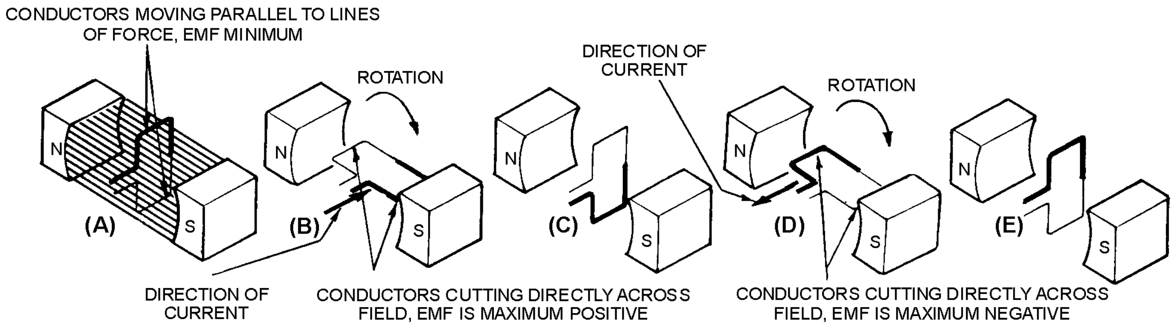


Figure 1-7.—Simple alternating-current generator.

As the loop continues to be rotated toward the position shown below in (C), it cuts fewer and fewer lines of force. The induced voltage decreases from its peak value. Eventually, the loop is once again moving in a plane parallel to the magnetic field, and no emf is induced in the conductor.

The loop has now been rotated through half a circle (one alternation or 180º). If the preceding quarter-cycle is plotted, it appears as shown below.

When the same procedure is applied to the second half of rotation (180º through 360º), the curve appears as shown below. Notice the only difference is in the polarity of the induced voltage. Where previously the polarity was positive, it is now negative.

The sine curve shows the value of induced voltage at each instant of time during rotation of the loop. Notice that this curve contains 360º , or two alternations. TWO ALTERNATIONS represent ONE complete CYCLE of rotation.

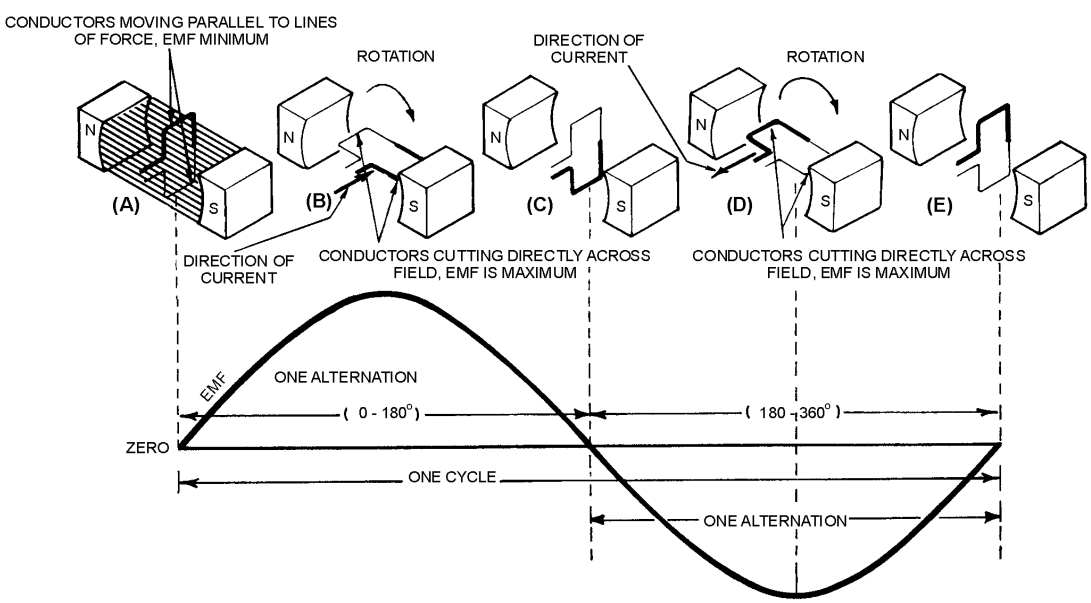


Figure 1-8.—Basic alternating-current generator.

Assuming a closed path is provided across the ends of the conductor loop, you can determine the direction of current in the loop by using the LEFT-HAND RULE FOR GENERATORS. Refer to figure 1-9. The left-hand rule is applied as follows: First, place your left hand on the illustration with the fingers as shown. Your THUMB will now point in the direction of rotation (relative movement of the wire to the magnetic field); your FOREFINGER will point in the direction of magnetic flux (north to south); and your MIDDLE FINGER (pointing out of the paper) will point in the direction of electron current flow.

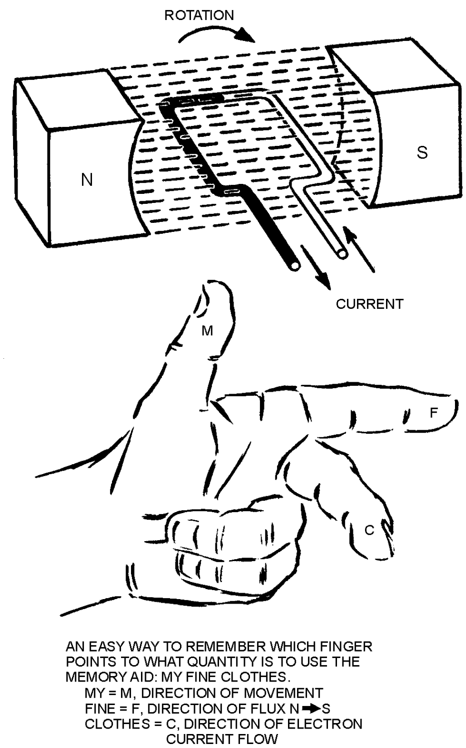


Figure 1-9.—Left-hand rule for generators.

By applying the left-hand rule to the dark half of the loop in (B) in figure 1-8, you will find that the current flows in the direction indicated by the heavy arrow. Similarly, by using the left-hand rule on the light half of the loop, you will find that current therein flows in the opposite direction. The two induced voltages in the loop add together to form one total emf. It is this emf which causes the current in the loop.

When the loop rotates to the position shown in (D) of figure 1-8, the action reverses. The dark half is moving up instead of down, and the light half is moving down instead of up. By applying the left-hand rule once again, you will see that the total induced emf and its resulting current have reversed direction. The voltage builds up to maximum in this new direction, as shown by the sine curve in figure 1-8. The loop finally returns to its original position (E), at which point voltage is again zero. The sine curve represents one complete cycle of voltage generated by the rotating loop. All the illustrations used in this chapter show the wire loop moving in a clockwise direction. In actual practice, the loop can be moved clockwise or counterclockwise. Regardless of the direction of movement, the left-hand rule applies.

If the loop is rotated through 360º at a steady rate, and if the strength of the magnetic field is uniform, the voltage produced is a sine wave of voltage, as indicated in figure 1-9. Continuous rotation of the loop will produce a series of sine-wave voltage cycles or, in other words, an ac voltage.

As mentioned previously, the cycle consists of two complete alternations in a period of time. Recently the HERTZ (Hz) has been designated to indicate one cycle per second. If ONE CYCLE PER SECOND is ONE HERTZ, then 100 cycles per second are equal to 100 hertz, and so on. Throughout the NEETS, the term cycle is used when no specific time element is involved, and the term hertz (Hz) is used when the time element is measured in seconds.

Q16. When a conductor is rotated in a magnetic field, at what points in the cycle is emf (a) at maximum amplitude and (b) at minimum amplitude?

Q17. One cycle is equal to how many degrees of rotation of a conductor in a magnetic field?

Q18. State the left-hand rule used to determine the direction of current in a generator.

Q19. How is an ac voltage produced by an ac generator?

FREQUENCY

If the loop in the figure 1-8 (A) makes one complete revolution each second, the generator produces one complete cycle of ac during each second (1 Hz). Increasing the number of revolutions to two per second will produce two complete cycles of ac per second (2 Hz). The number of complete cycles of alternating current or voltage completed each second is referred to as the FREQUENCY. Frequency is always measured and expressed in hertz.

Alternating-current frequency is an important term to understand since most ac electrical equipments require a specific frequency for proper operation.

Q20. Define Frequency.

PERIOD

An individual cycle of any sine wave represents a definite amount of TIME. Notice that figure 1-10 shows 2 cycles of a sine wave which has a frequency of 2 hertz (Hz). Since 2 cycles occur each second, 1 cycle must require one-half second of time. The time required to complete one cycle of a waveform is called the PERIOD of the wave. In figure 1-10, the period is one-half second. The relationship between time (t) and frequency (f) is indicated by the formulas

$$t = \frac{1}{f} \quad \text{and} \quad f = \frac{1}{t}$$

where t = period in seconds and
f = frequency in hertz

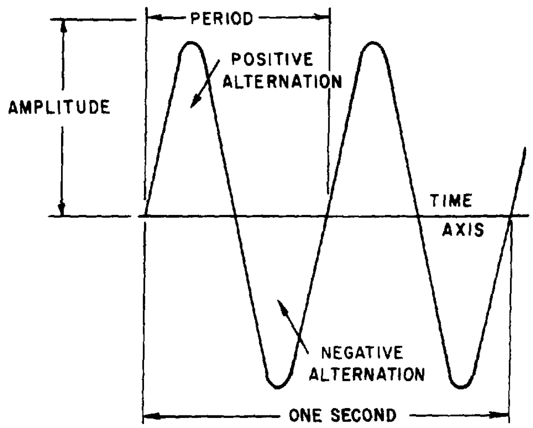


Figure 1-10.—Period of a sine wave.

Each cycle of the sine wave shown in figure 1-10 consists of two identically shaped variations in voltage. The variation which occurs during the time the voltage is positive is called the POSITIVE ALTERNATION. The variation which occurs during the time the voltage is negative is called the NEGATIVE ALTERNATION. In a sine wave, these two alternations are identical in size and shape, but opposite in polarity.

The distance from zero to the maximum value of each alternation is called the AMPLITUDE. The amplitude of the positive alternation and the amplitude of the negative alternation are the same.

WAVELENGTH

The time it takes for a sine wave to complete one cycle is defined as the period of the waveform. The distance traveled by the sine wave during this period is referred to as WAVELENGTH. Wavelength, indicated by the symbol λ (Greek lambda), is the distance along the waveform from one point to the same point on the next cycle. You can observe this relationship by examining figure 1-11. The point on the waveform that measurement of wavelength begins is not important as long as the distance is measured to the same point on the next cycle (see figure 1-12).

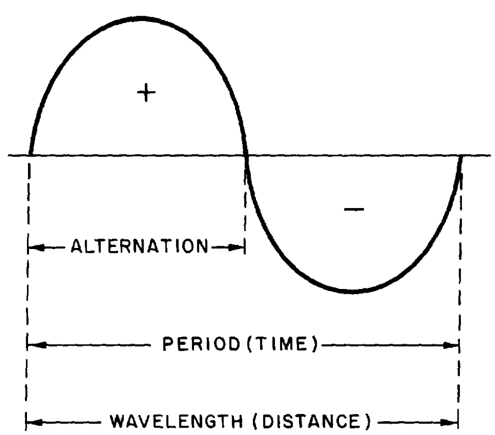


Figure 1-11.—Wavelength.

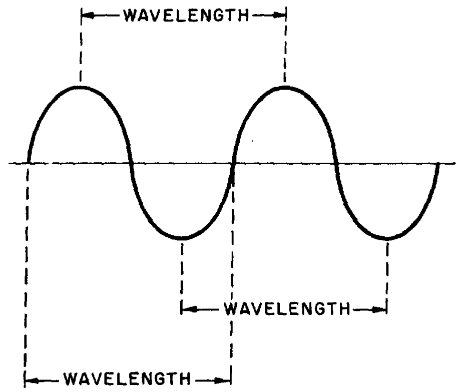


Figure 1-12.—Wavelength measurement.

Q21. What term is used to indicate the time of one complete cycle of a waveform?

Q22. What is a positive alternation?

Q23. What do the period and the wavelength of a sine wave measure, respectively?

ALTERNATING CURRENT VALUES

In discussing alternating current and voltage, you will often find it necessary to express the current and voltage in terms of MAXIMUM or PEAK values, PEAK-to-PEAK values, EFFECTIVE values, AVERAGE values, or INSTANTANEOUS values. Each of these values has a different meaning and is used to describe a different amount of current or voltage.

PEAK AND PEAK-TO-PEAK VALUES

Refer to figure 1-13. Notice it shows the positive alternation of a sine wave (a half-cycle of ac) and a dc waveform that occur simultaneously. Note that the dc starts and stops at the same moment as does the positive alternation, and that both waveforms rise to the same maximum value. However, the dc values are greater than the corresponding ac values at all points except the point at which the positive alternation passes through its maximum value. <u>At this point the dc and ac values are equal. This point on the sine wave is referred to as the maximum or peak value.</u>

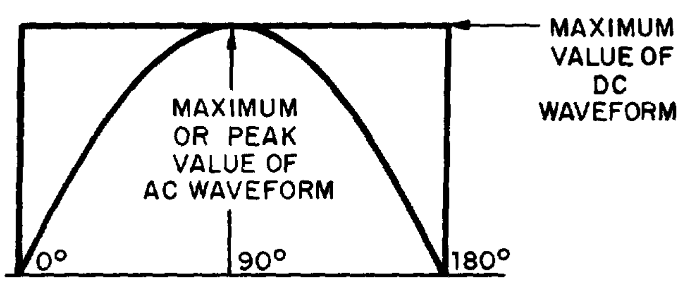


Figure 1-13.—Maximum or peak value.

During each complete cycle of ac there are always two maximum or peak values, one for the positive half-cycle and the other for the negative half-cycle. The difference between the peak positive value and the peak negative value is called the peak-to-peak value of the sine wave. This value is twice the maximum or peak value of the sine wave and is sometimes used for measurement of ac voltages. Note the difference between peak and peak-to-peak values in figure 1-14. Usually alternating voltage and current are expressed in EFFECTIVE VALUES (a term you will study later) rather than in peak-to-peak values.

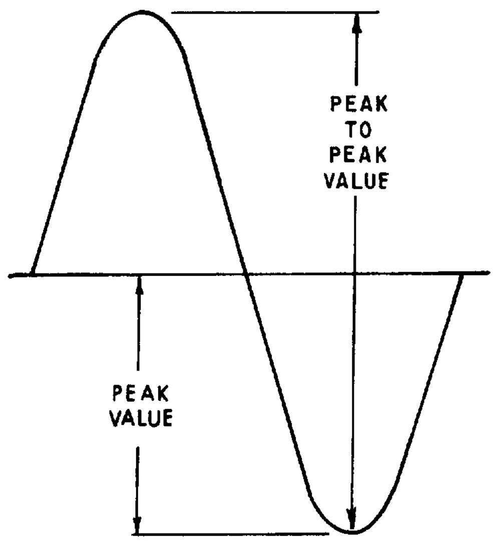


Figure 1-14.—Peak and peak-to-peak values.

Q24. What is meant by peak and peak-to-peak values of ac?

Q25. How many times is the maximum or peak value of emf or current reached during one cycle of ac?

INSTANTANEOUS VALUE

The INSTANTANEOUS value of an alternating voltage or current is the value of voltage or current at one particular instant. The value may be zero if the particular instant is the time in the cycle at which the polarity of the voltage is changing. It may also be the same as the peak value, if the selected instant is the time in the cycle at which the voltage or current stops increasing and starts decreasing. There are actually an infinite number of instantaneous values between zero and the peak value.

AVERAGE VALUE

The AVERAGE value of an alternating current or voltage is the average of ALL the INSTANTANEOUS values during ONE alternation. Since the voltage increases from zero to peak value and decreases back to zero during one alternation, the average value must be some value between those two limits. You could determine the average value by adding together a series of instantaneous values of the alternation (between 0º and 180º), and then dividing the sum by the number of instantaneous values used. The computation would show that one alternation of a sine wave has an average value equal to 0.636 times the peak value. The formula for average voltage is

$$E_{avg} = 0.636 \times E_{max}$$

where E_{avg} is the average voltage of one alternation, and E_{max} is the maximum or peak voltage. Similarly, the formula for average current is

$$I_{avg} = 0.636 \times I_{max}$$

where I_{avg} is the average current in one alternation, and I_{max} is the maximum or peak current.

Do not confuse the above definition of an average value with that of the average value of a complete cycle. Because the voltage is positive during one alternation and negative during the other alternation, the average value of the voltage values occurring during the complete cycle is zero.

Q26. If any point on a sine wave is selected at random and the value of the current or voltage is measured at that one particular moment, what value is being measured?

Q27. What value of current or voltage is computed by averaging all of the instantaneous values during the negative alternation of a sine wave?

Q28. What is the average value of all of the instantaneous currents or voltages occurring during one complete cycle of a sine wave?

Q29. What mathematical formulas are used to find the average value of current and average value of voltage of a sine wave?

Q30. If E_{max} is 115 volts, what is E_{avg}?

Q31. If I_{avg} is 1.272 ampere, what is I_{max}?

EFFECTIVE VALUE OF A SINE WAVE

E_{max}, E_{avg}, I_{max}, and I_{avg} are values used in ac measurements. Another value used is the EFFECTIVE value of ac This is the value of alternating voltage or current that will have the same effect on a resistance as a comparable value of direct voltage or current will have on the same resistance.

In an earlier discussion you were told that when current flows in a resistance, heat is produced. When direct current flows in a resistance, the amount of electrical power converted into heat equals I^2R watts. However, since an alternating current having a maximum value of 1 ampere does not maintain a constant value, the alternating current will not produce as much heat in the resistance as will a direct current of 1 ampere.

Figure 1-15 compares the heating effect of 1 ampere of dc to the heating effect of 1 ampere of ac.

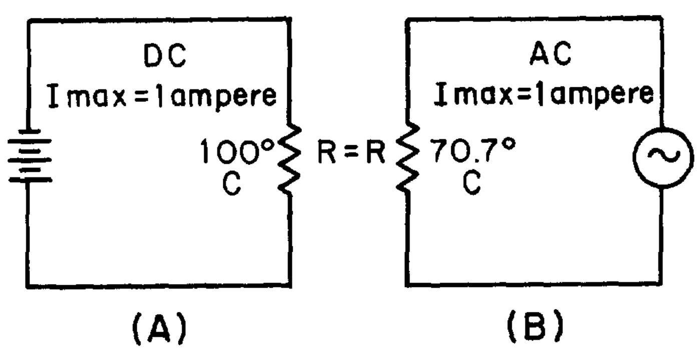


Figure 1-15.—Heating effect of ac and dc.

Examine views A and B of figure 1-15 and notice that the heat (70.7º C) produced by 1 ampere of alternating current (that is, an ac with a maximum value of 1 ampere) is only 70.7 percent of the heat (100º C) produced by 1 ampere of direct current. Mathematically,

$$\frac{\text{The heating effect of 1 maximum a.c. ampere}}{\text{The heating effect of 1 maximum d.c. ampere}} = \frac{70.7^\circ\text{C}}{100^\circ\text{C}} = 0.707$$

Therefore, for effective value of ac (I_{eff}) = $0.707 \times I_{max}$.

The rate at which heat is produced in a resistance forms a convenient basis for establishing an effective value of alternating current, and is known as the "heating effect" method. An alternating current is said to have an effective value of one ampere when it produces heat in a given resistance at the same rate as does one ampere of direct current.

You can compute the effective value of a sine wave of current to a fair degree of accuracy by taking equally-spaced instantaneous values of current along the curve and extracting the square root of the average of the sum of the squared values.

For this reason, the effective value is often called the "root-mean-square" (rms) value. Thus,

$$I_{eff} = \sqrt{\text{Average of the sum of the squares of } I_{inst}}.$$

Stated another way, the effective or rms value (I_{eff}) of a sine wave of current is 0.707 times the maximum value of current (I_{max}). Thus, $I_{eff} = 0.707 \times I_{max}$. When I_{eff} is known, you can find I_{max} by using the formula $I_{max} = 1.414 \times I_{eff}$. You might wonder where the constant 1.414 comes from. To find out, examine figure 1-15 again and read the following explanation. Assume that the dc in figure 1-15(A) is maintained at 1 ampere and the resistor temperature at 100º C. Also assume that the ac in figure 1-15(B) is increased until the temperature of the resistor is 100º C. At this point it is found that a maximum ac value of 1.414 amperes is required in order to have the same heating effect as direct current. Therefore, in the ac circuit the maximum current required is 1.414 times the effective current. It is important for you to remember the above relationship and that the effective value (I_{eff}) of any sine wave of current is always 0.707 times the maximum value (I_{max}).

Since alternating current is caused by an alternating voltage, the ratio of the effective value of voltage to the maximum value of voltage is the same as the ratio of the effective value of current to the maximum value of current. Stated another way, the effective or rms value (E_{eff}) of a sine-wave of voltage is 0.707 times the maximum value of voltage (E_{max}), Thus,

$$E_{eff} = \sqrt{\text{Average of the sum of the squares of } E_{inst}}$$

or,

$$E_{eff} = 0.707 \times E_{max}$$

and,

$$E_{max} = 1.414 \times E_{eff}$$

When an alternating current or voltage value is specified in a book or on a diagram, the value is an effective value unless there is a definite statement to the contrary. Remember that all meters, unless marked to the contrary, are calibrated to indicate effective values of current and voltage.

Problem: A circuit is known to have an alternating voltage of 120 volts and a peak or maximum current of 30 amperes. What are the peak voltage and effective current values?

Given:

$$E_s = 120\text{ V}$$
$$E_{max} = 30\text{ A}$$

Solution:

$$E_{max} = 1.414 \times E_{eff}$$
$$E_{max} = 1.414 \times 120 \text{ volts}$$
$$E_{max} = 169.68 \text{ volts}$$
$$I_{eff} = 0.707 \times I_{max}$$
$$I_{eff} = 0.707 \times 30 \text{ amperes}$$
$$I_{eff} = 21.21 \text{ amperes}$$

Figure 1-16 shows the relationship between the various values used to indicate sine-wave amplitude. Review the values in the figure to ensure you understand what each value indicates.

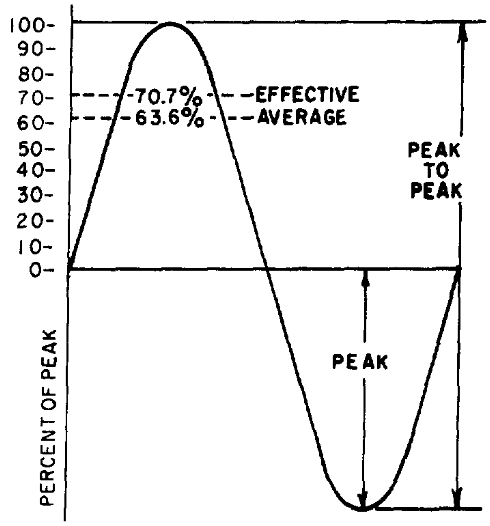


Figure 1-16.—Various values used to indicate sine-wave amplitude.

Q32. What is the most convenient basis for comparing alternating and direct voltages and currents?

Q33. What value of ac is used as a comparison to dc?

Q34. What is the formula for finding the effective value of an alternating current?

Q35. If the peak value of a sine wave is 1,000 volts, what is the effective (E_{eff}) value?

Q36. If I_{eff} = 4.25 ampere, what is I_{max}?

SINE WAVES IN PHASE

When a sine wave of voltage is applied to a resistance, the resulting current is also a sine wave. This follows Ohm's law which states that current is directly proportional to the applied voltage. Now examine figure 1-17. Notice that the sine wave of voltage and the resulting sine wave of current are superimposed on the same time axis. Notice also that as the voltage increases in a positive direction, the current increases along with it, and that when the voltage reverses direction, the current also reverses direction. When two sine waves, such as those represented by figure 1-17, are precisely in step with one another, they are said to be IN PHASE. To be in phase, the two sine waves must go through their maximum and minimum points at the same time and in the same direction.

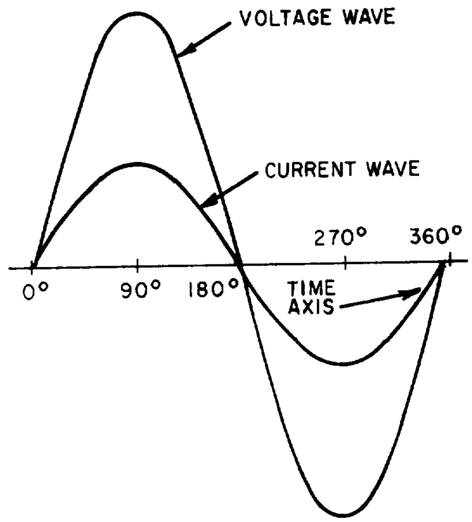


Figure 1-17.—Voltage and current waves in phase.

In some circuits, several sine waves can be in phase with each other. Thus, it is possible to have two or more voltage drops in phase with each other and also be in phase with the circuit current.

SINE WAVES OUT OF PHASE

Figure 1-18 shows voltage wave E_1 which is considered to start at 0° (time one). As voltage wave E_1 reaches its positive peak, voltage wave E_2 starts its rise (time two). Since these voltage waves do not go through their maximum and minimum points at the same instant of time, a PHASE DIFFERENCE exists between the two waves. The two waves are said to be OUT OF PHASE. For the two waves in figure 1-18 the phase difference is 90° .

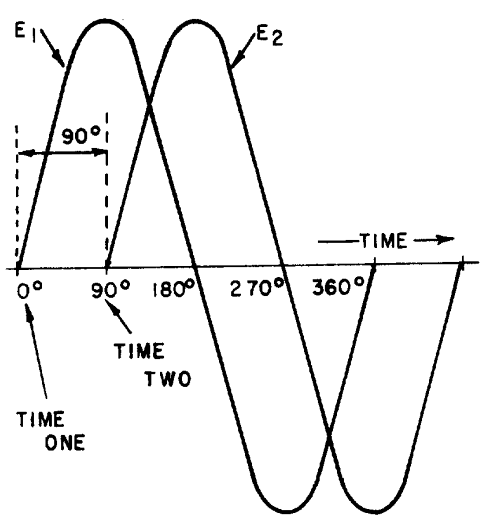


Figure 1-18.—Voltage waves 90° out of phase.

To further describe the phase relationship between two sine waves, the terms LEAD and LAG are used. The amount by which one sine wave leads or lags another sine wave is measured in degrees. Refer again to figure 1-18. Observe that wave E_2 starts 90º later in time than does wave E_1. You can also describe this relationship by saying that wave E_1 leads wave E_2 by 90º, or that wave E_2 lags wave E_1 by 90º. (Either statement is correct; it is the phase relationship between the two sine waves that is important.)

It is possible for one sine wave to lead or lag another sine wave by any number of degrees, except 0º or 360º. When the latter condition exists, the two waves are said to be in phase. Thus, two sine waves that differ in phase by 45º are actually out of phase with each other, whereas two sine waves that differ in phase by 360º are considered to be in phase with each other.

A phase relationship that is quite common is shown in figure 1-19. Notice that the two waves illustrated differ in phase by 180º. Notice also that although the waves pass through their maximum and minimum values at the same time, their instantaneous voltages are always of opposite polarity. If two such waves exist across the same component, and the waves are of equal amplitude, they cancel each other. When they have different amplitudes, the resultant wave has the same polarity as the larger wave and has an amplitude equal to the difference between the amplitudes of the two waves.

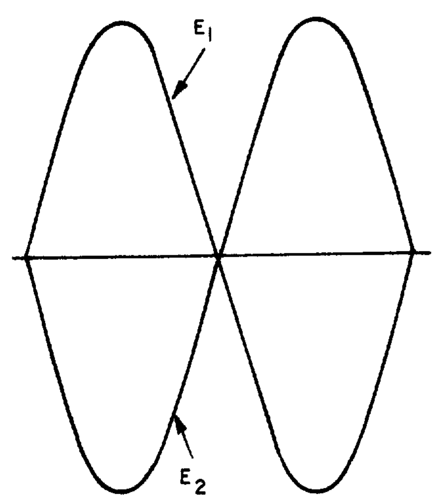


Figure 1-19.—Voltage waves 180º out of phase.

To determine the phase difference between two sine waves, locate the points on the time axis where the two waves cross the time axis traveling in the same direction. The number of degrees between the crossing points is the phase difference. The wave that crosses the axis at the later time (to the right on the time axis) is said to lag the other wave.

Q37. When are the voltage wave and the current wave in a circuit considered to be in phase?

Q38. When are two voltage waves considered to be out of phase?

Q39. What is the phase relationship between two voltage waves that differ in phase by 360° ?

Q40. How do you determine the phase difference between two sine waves that are plotted on the same graph?

OHM'S LAW IN AC CIRCUITS

Many ac circuits contain resistance only. The rules for these circuits are the same rules that apply to dc circuits. Resistors, lamps, and heating elements are examples of resistive elements. When an ac circuit contains only resistance, Ohm's Law, Kirchhoff's Law, and the various rules that apply to voltage, current, and power in a dc circuit also apply to the ac circuit. The Ohm's Law formula for an ac circuit can be stated as

$$I_{eff} = \frac{E_{eff}}{R} \text{ or } I = \frac{E}{R}$$

Remember, unless otherwise stated, all ac voltage and current values are given as effective values. The formula for Ohm's Law can also be stated as

$$I_{avg} = \frac{E_{avg}}{R} \text{ or } I_{max} = \frac{E_{max}}{R}$$

$$I_{peak-to-peak} = \frac{E_{peak-to-peak}}{R}$$

The important thing to keep in mind is: <u>Do Not mix ac values</u>. When you solve for effective values, all values you use in the formula <u>must be effective values</u>. Similarly, when you solve for average values, all values you use <u>must be average values</u>. This point should be clearer after you work the following problem: A series circuit consists of two resistors (R1 = 5 ohms and R2 = 15 ohms) and an alternating voltage source of 120 volts. What is I_{avg}?

Given: $R1 = 5$ ohms

$R2 = 15$ ohms

$E_s = 120$ ohms

Solution: First solve for total resistance R_T.

$$R_T = R1 + R2$$
$$R_T = 5 \text{ ohms} + 15 \text{ ohms}$$
$$R_T = 20 \text{ ohms}$$

The alternating voltage is assumed to be an effective value (since it is not specified to be otherwise). Apply the Ohm's Law formula.

$$I_{eff} = \frac{E_{eff}}{R}$$

$$I_{eff} = \frac{120 \text{ volts}}{20 \text{ ohms}}$$

$$I_{eff} = 6 \text{ amperes}$$

The problem, however, asked for the average value of current (I_{avg}). To convert the effective value of current to the average value of current, you must first determine the peak or maximum value of current, I_{max}.

$$I_{max} = 1.414 \times I_{eff}$$

$$I_{max} = 1.414 \times 6 \text{ amperes}$$

$$I_{max} = 8.484 \text{ amperes}$$

You can now find I_{avg}. Just substitute 8.484 amperes in the I_{avg} formula and solve for I_{avg}.

$$I_{avg} = 0.636 \times I_{max}$$

$$I_{avg} = 0.636 \times 8.484 \text{ amperes}$$

$$I_{avg} = 5.4 \text{ amperes (rounded off to one decimal place)}$$

Remember, you can use the Ohm's Law formulas to solve any purely resistive ac circuit problem. Use the formulas in the same manner as you would to solve a dc circuit problem.

Q41. A series circuit consists of three resistors (R1 = 10Ω, R2 = 20Ω, R3 = 15Ω) and an alternating voltage source of 100 volts. What is the effective value of current in the circuit?

Q42. If the alternating source in Q41 is changed to 200 volts peak-to-peak, what is I_{avg}?

Q43. If E_{eff} is 130 volts and I_{eff} is 3 amperes, what is the total resistance (R_T) in the circuit?

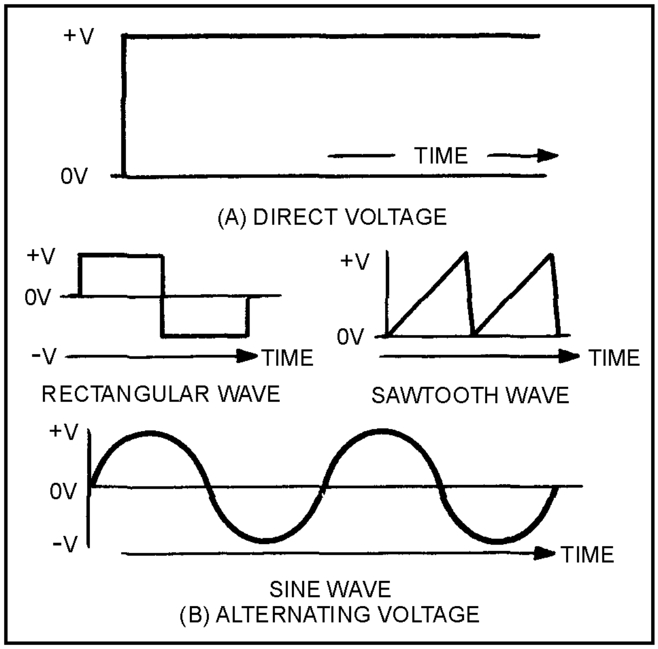
SUMMARY

Before going on to chapter 2, read the following summary of the material in chapter 1. This summary will reinforce what you have already learned.

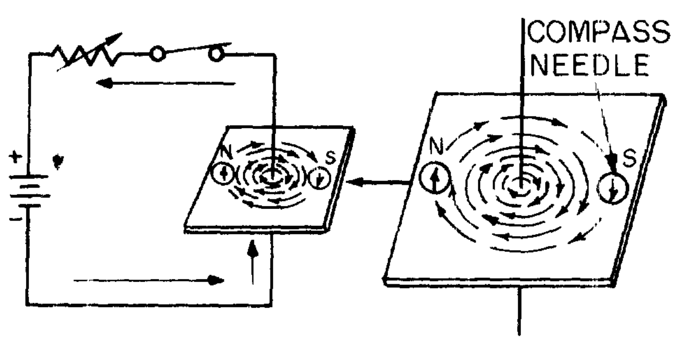
DC AND AC—Direct current flows in one direction only, while alternating current is constantly changing in amplitude and direction.

ADVANTAGES AND DISADVANTAGES OF AC AND DC—Direct current has several disadvantages compared to alternating current. Direct current, for example, must be generated at the voltage level required by the load. Alternating current, however, can be generated at a high level and stepped down at the consumer end (through the use of a transformer) to whatever voltage level is required by the load. Since power in a dc system must be transmitted at low voltage and high current levels, the I^2R power loss becomes a problem in the dc system. Since power in an ac system can be transmitted at a high voltage level and a low current level, the I^2R power loss in the ac system is much less than that in the dc system.

VOLTAGE WAVEFORMS—The waveform of voltage or current is a graphical picture of changes in voltage or current values over a period of time.

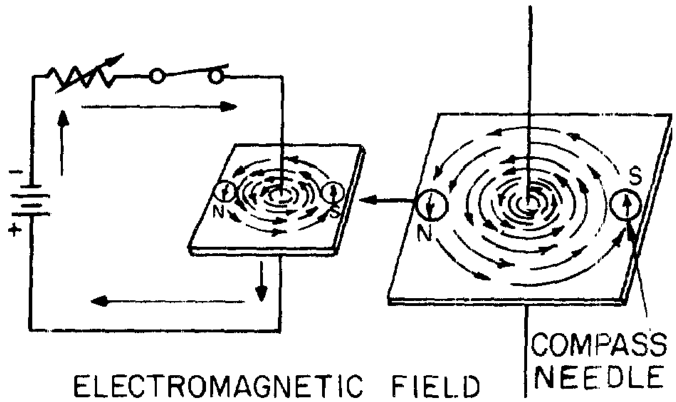


ELECTROMAGNETISM—When a compass is placed in the vicinity of a current-carrying conductor, the needle aligns itself at right angles to the conductor. The north pole of the compass indicates the direction of the magnetic field produced by the current. By knowing the direction of current, you can use the left-hand rule for conductors to determine the direction of the magnetic lines of force.



ELECTROMAGNETIC FIELD (CLOCKWISE)

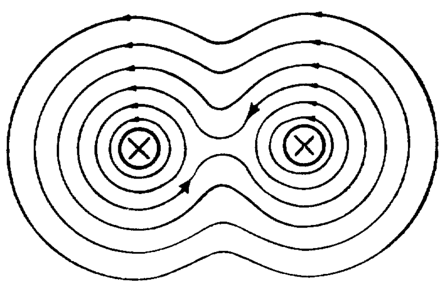
(A)



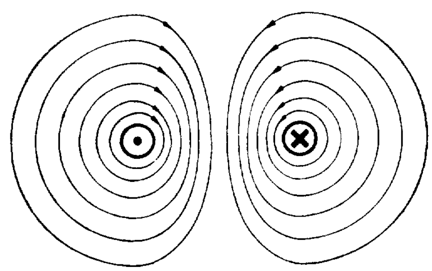
ELECTROMAGNETIC FIELD
(COUNTERCLOCKWISE)
(B)

Arrows are generally used in electrical diagrams to indicate the direction of current in a wire. A cross (+) on the end of a cross-sectional view of a wire indicates that current is flowing away from you, while a dot (·) indicates that current is flowing toward you.

When two adjacent parallel conductors carry current in the same direction, the magnetic fields around the conductors aid each other. When the currents in the two conductors flow in opposite directions, the fields around the conductors oppose each other.



CURRENT FLOWING IN THE SAME DIRECTION (A)



CURRENT FLOWING IN OPPOSITE DIRECTION (B)

MAGNETIC FIELD OF A COIL—When wire is wound around a core, it forms a COIL. The magnetic fields produced when current flows in the coil combine. The combined influence of all of the fields around the turns produce a two-pole field similar to that of a simple bar magnet.

When the direction of current in the coil is reversed, the polarity of the two-pole field of the coil is reversed.

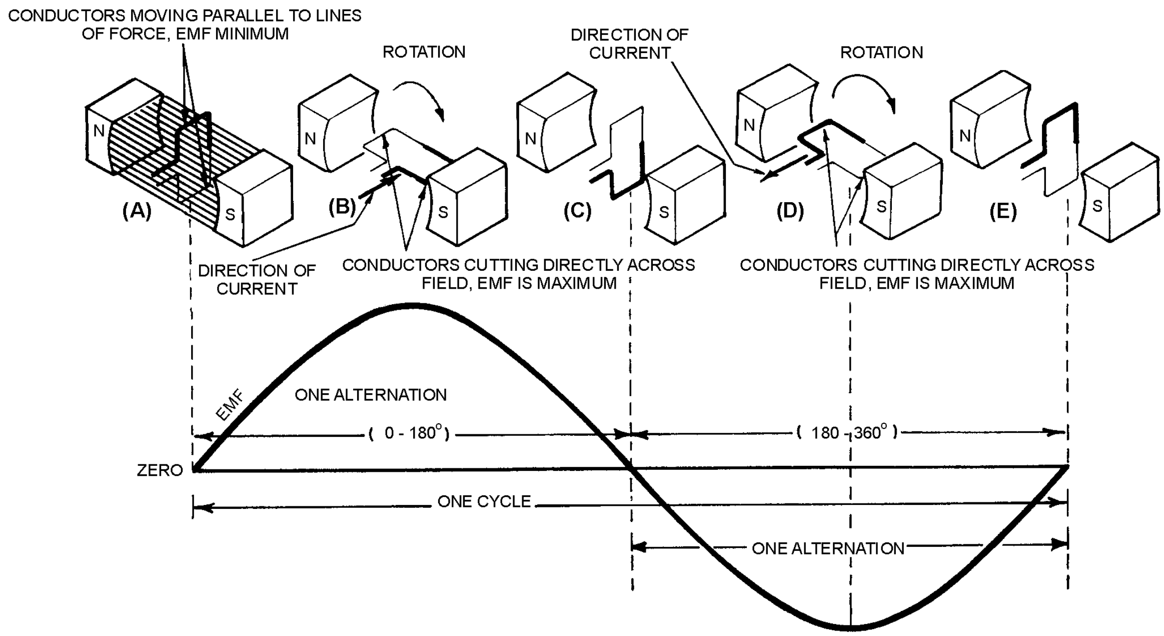
The strength of the magnetic field of the coil is dependent upon:

- The number of turns of the wire in the coil.
- The amount of current in the coil.
- The ratio of the coil length to the coil width.

- The type of material in the core.

BASIC AC GENERATION—When a conductor is in a magnetic field and either the field or the conductor moves, an emf (voltage) is induced in the conductor. This effect is called electromagnetic induction.

A loop of wire rotating in a magnetic field produces a voltage which constantly changes in amplitude and direction. The waveform produced is called a sine wave and is a graphical picture of alternating current (ac). One complete revolution (360º) of the conductor produces one cycle of ac. The cycle is composed of two alternations: a positive alternation and a negative alternation. One cycle of ac in one second is equal to 1 hertz (1 Hz).



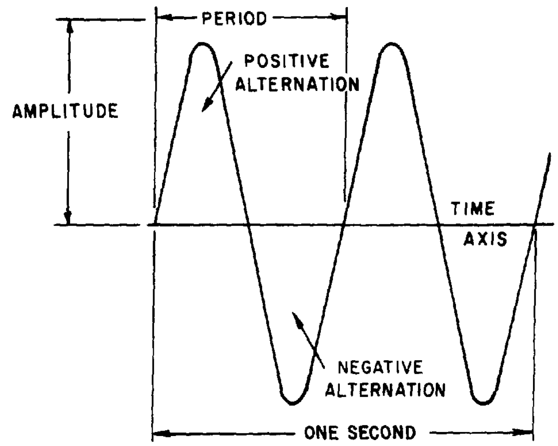
FREQUENCY—The number of cycles of ac per second is referred to as the FREQUENCY. AC frequency is measured in hertz. Most ac equipment is rated by frequency as well as by voltage and current.

PERIOD—The time required to complete one cycle of a waveform is called the PERIOD OF THE WAVE.

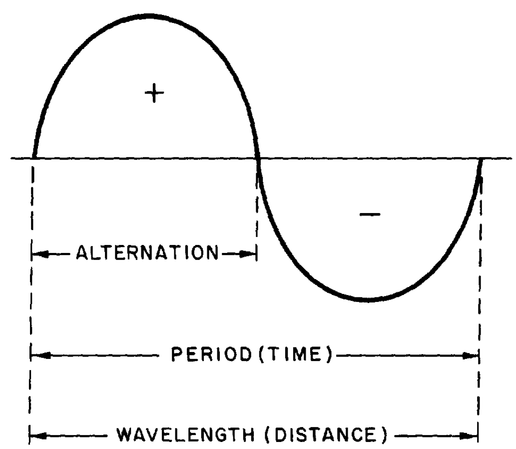
Each ac sine wave is composed of two alternations. The alternation which occurs during the time the sine wave is positive is called the positive alternation. The alternation which occurs during the time the sine wave is negative is called the negative alternation. In each cycle of sine wave, the two alternations are identical in size and shape, but opposite in polarity.

The period of a sine wave is inversely proportional to the frequency; e.g., the higher the frequency, the shorter the period. The mathematical relationships between time and frequency are

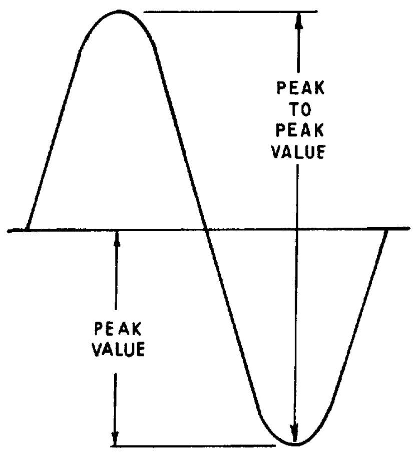
$$t = \frac{1}{f} \text{ and } f = \frac{1}{t}$$



WAVELENGTH—The period of a sine wave is defined as the time it takes to complete one cycle. The distance the waveform covers during this period is referred to as the wavelength. Wavelength is indicated by lambda (λ) and is measured from a point on a given waveform (sine wave) to the corresponding point on the next waveform.



PEAK AND PEAK-TO-PEAK VALUES—The maximum value reached during one alternation of a sine wave is the peak value. The maximum reached during the positive alternation to the maximum value reached during the negative alternation is the peak-to-peak value. The peak-to-peak value is twice the peak value.



INSTANTANEOUS VALUE—The instantaneous value of a sine wave of alternating voltage or current is the value of voltage or current at one particular instant of time. There are an infinite number of instantaneous values between zero and the peak value.

AVERAGE VALUE—The average value of a sine wave of voltage or current is the average of all the instantaneous values during one alternation. The average value is equal to 0.636 of the peak value. The formulas for average voltage and average current are:

$$E_{avg} = 0.636 \times E_{max}$$
$$I_{avg} = 0.636 \times I_{max}$$

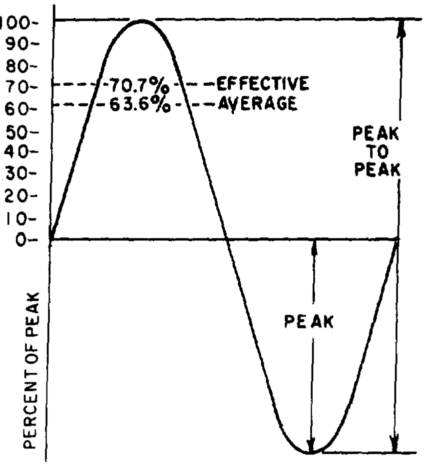
Remember: The average value (E_{avg} or I_{avg}) is for one alternation only. The average value of a complete sine wave is zero.

EFFECTIVE VALUE—The effective value of an alternating current or voltage is the value of alternating current or voltage that produces the same amount of heat in a resistive component that would be produced in the same component by a direct current or voltage of the same value. The effective value of a sine wave is equal to 0.707 times the peak value. The effective value is also called the root mean square or rms value.

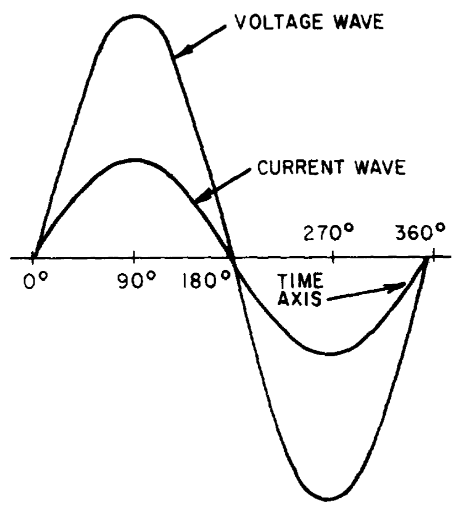
The term rms value is used to describe the process of determining the effective value of a sine wave by using the instantaneous value of voltage or current. You can find the rms value of a current or voltage by taking equally spaced instantaneous values on the sine wave and extracting the square root of the average of the sum of the instantaneous values. This is where the term "Root-Mean-Square" (rms) value comes from.

The formulas for effective and maximum values of voltage and current are:

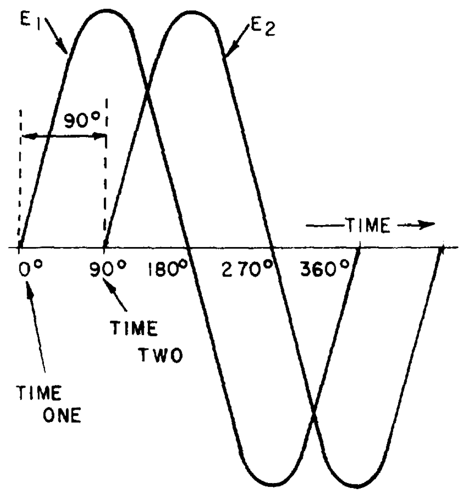
$$E_{eff} = 0.707 \times E_{max}$$
$$E_{max} = 1.414 \times E_{eff}$$
$$I_{eff} = 0.707 \times I_{max}$$
$$I_{max} = 1.414 \times I_{eff}$$



SINE WAVES IN PHASE—When two sine waves are exactly in step with each other, they are said to be in phase. To be in phase, both sine waves must go through their minimum and maximum points at the same time and in the same direction.



SINE WAVES OUT OF PHASE—When two sine waves go through their minimum and maximum points at different times, a phase difference exists between them. The two waves are said to be out of phase with each other. To describe this phase difference, the terms lead and lag are used. The wave that reaches its minimum (or maximum) value first is said to lead the other wave. The term lag is used to describe the wave that reaches its minimum (or maximum) value some time after the first wave does. When a sine wave is described as leading or lagging, the difference in degrees is usually stated. For example, wave E_1 leads wave E_2 by 90° , or wave E_2 lags wave E_1 by 90° . Remember: Two sine waves can differ by any number of degrees except 0° and 360° . Two sine waves that differ by 0° or by 360° are considered to be in phase. Two sine waves that are opposite in polarity and that differ by 180° are said to be out of phase, even though they go through their minimum and maximum points at the same time.



<u>OHM'S LAW IN AC CIRCUIT</u>—All dc rules and laws apply to an ac circuit that contains <u>only resistance</u>. The important point to remember is: Do not <u>mix ac values</u>. Ohm's Law formulas for ac circuits are given below:

$$I = \frac{E}{R}$$

$$I_{eff} = \frac{E_{eff}}{R}$$

$$I_{avg} = \frac{E_{avg}}{R}$$

$$I_{max} = \frac{E_{max}}{R}$$

$$I_{Peak-to-Peak} = \frac{E_{Peak-to-Peak}}{R}$$

ANSWERS TO QUESTIONS Q1. THROUGH Q43.

A1. An electrical current which flows in one direction only.

A2. An electrical current which is constantly varying in amplitude, and which changes direction at regular intervals.

A3. The dc voltage must be generated at the level required by the load.

A4. The I^2R power loss is excessive.

A5. Alternating current (ac).

A6. The needle aligns itself at right angles to the conductor.

A7. (a) clockwise (b) counterclockwise.

A8. It is used to determine the relation between the direction of the magnetic lines of force around a conductor and the direction of current through the conductor.

A9. The north pole of the compass will point in the direction of the magnetic lines of force.

A10. It combines with the other field.

A11. It deforms the other field.

A12. (a) The field consists of concentric circles in a plane perpendicular to the wire (b) the field of each turn of wire links with the fields of adjacent turns producing a two-pole field similar in shape to that of a simple bar magnet.

A13. The polarity of the two-pole field reverses.

A14. Use the left-hand rule for coils.

A15. Grasp the coil in your left hand, with your fingers "wrapped around" in the direction of electron flow. The thumb will point toward the north pole.

A16. (a) When the conductors are cutting directly across the magnetic lines of force (at the 90º and 270º points). (b) When the conductors are moving parallel to the magnetic lines of force (at the 0°, 180°, and 360° points).

A17. 360°.

A18. Extend your left hand so that your thumb points in the direction of conductor movement, and your forefinger points in the direction of the magnetic flux (north to south). Now point your middle finger 90° from the forefinger and it will point in the direction of electron current flow in the conductor.

A19. Continuous rotation of the conductor through magnetic fines of force produces a series of cycles of alternating voltage or, in other words, an alternating voltage or a sine wave of voltage.

A20. Frequency is the number of complete cycles of alternating voltage or current completed each second.

A21. Period.

A22. A positive alternation is the positive variation in the voltage or current of a sine curve.

A23. The period measures time and the wavelength measures distance.

A24. The peak value is the maximum value of one alternation; the peak-to-peak value is twice the maximum or peak value.

A25. Twice.

A26. The instantaneous value (E_{inst} or I_{inst})

A27. Average value (E_{avg} or I_{avg})

A28. Zero

A29.

$$I_{avg} = 0.636 \times I_{max}$$

$$E_{avg} = 0.636 \times E_{max}$$

A30.

$$E_{avg} = 0.636 \times 115 \text{ volts}$$

$$E_{avg} = 73.14 \text{ volts}$$

A31.

$$\text{If } I_{avg} = I_{max} \times 0.636, \text{ then } I_{max} = \frac{I_{avg}}{0.636}$$

Thus,

$$I_{max} = \frac{1.272}{0.636} \text{ ampere} = 2 \text{ amperes}$$

A32. The power (heat) produced in a resistance by a dc voltage is compared to that produced in the same resistance by an ac voltage of the same peak amplitude.

A33. The effective value.

A34.

$$I_{eff} = 0.707 \times I_{max}$$

A35.

$$
\begin{aligned}
E_{eff} &= 0.707 \times E_{max} \\
&= 0.707 \times E_{max} \\
&= 0.707 \times 1{,}000 \text{ volts} \\
E_{eff} &= 707 \text{ volts}
\end{aligned}
$$

A36.

$$
\begin{aligned}
I_{max} &= 1.414 \times I_{eff} \\
&= 1.414 \times 4.25 \text{ amperes} \\
&= 6 \text{ amperes.}
\end{aligned}
$$

(Remember: Unless specified otherwise, the voltage or current value is always considered to be the effective value.)

A37. *When the two waves go through their maximum and minimum points at the same time and in the same direction.*

A38. *When the waves do not go through their maximum and minimum points at the same time, a PHASE DIFFERENCE exists, and the two waves are said to be out of phase. (Two waves are also considered to be out of phase if they differ in phase by 180° and their instantaneous voltages are always of opposite polarity, even though both waves go through their maximum and minimum points at the same time).*

A39. *They are in phase with each other.*

A40. *Locate the points on the time axis where the two waves cross traveling in the same direction. The number of degrees between these two points is the phase difference.*

A41.

$$
I_{eff} = \frac{100}{45} = 2.22 \text{ ampers}
$$

A42. $I_{avg} = 0.636 \times I_{max} = 1.41$ *amperes.*

A43. *43.3 ohms.*

CHAPTER 2

INDUCTANCE

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Write the basic unit of and the symbol for inductance.
2. State the type of moving field used to generate an emf in a conductor.
3. Define the term "inductance."
4. State the meanings of the terms "induced emf" and "counter emf."
5. State Lenz's law.
6. State the effect that inductance has on steady direct current, and direct current that is changing in magnitude.
7. List five factors that affect the inductance of a coil, and state how various physical changes in these factors affect inductance.
8. State the principles and sequences involved in the buildup and decay of current in an LR series circuit.
9. Write the formula for computing one time constant in an LR series circuit.
10. Solve L/R time constant problems.
11. State the three types of power loss in an inductor.
12. Define the term "mutual inductance."
13. State the meaning of the term "coupled circuits."
14. State the meaning of the term "coefficient of coupling."
15. Given the inductance values of and the coefficient of coupling between two series-connected inductors, solve for mutual inductance, M.
16. Write the formula for the "total inductance" of two inductors connected in series-opposing.
17. Given the inductance values of and the mutual inductance value between two coils connected in series-aiding, solve for their combined inductance, L_T.

INDUCTANCE

The study of inductance presents a very challenging but rewarding segment of electricity. It is challenging in the sense that, at first, it will seem that new concepts are being introduced. You will realize as this chapter progresses that these "new concepts" are merely extensions and enlargements of fundamental principles that you learned previously in the study of magnetism and electron physics. The study of inductance is rewarding in the sense that a thorough understanding of it will enable you to acquire a working knowledge of electrical circuits more rapidly.

CHARACTERISTICS OF INDUCTANCE

Inductance is the characteristic of an electrical circuit that opposes the starting, stopping, or a change in value of current. The above statement is of such importance to the study of inductance that it bears repeating. Inductance is the characteristic of an electrical conductor that OPPOSES CHANGE in CURRENT. The symbol for inductance is L and the basic unit of inductance is the HENRY (H). One henry is equal to the inductance required to induce one volt in an inductor by a change of current of one ampere per second.

You do not have to look far to find a physical analogy of inductance. Anyone who has ever had to push a heavy load (wheelbarrow, car, etc.) is aware that it takes more work to start the load moving than it does to keep it moving. Once the load is moving, it is easier to keep the load moving than to stop it again. This is because the load possesses the property of INERTIA. Inertia is the characteristic of mass which opposes a CHANGE in velocity. Inductance has the same effect on current in an electrical circuit as inertia has on the movement of a mechanical object. It requires more energy to start or stop current than it does to keep it flowing.

Q1. What is the basic unit of inductance and the abbreviation for this unit?

ELECTROMOTIVE FORCE (EMF)

You have learned that an electromotive force is developed whenever there is relative motion between a magnetic field and a conductor.

Electromotive force is a difference of potential or voltage which exists between two points in an electrical circuit. In generators and inductors the emf is developed by the action between the magnetic field and the electrons in a conductor. This is shown in figure 2-1.

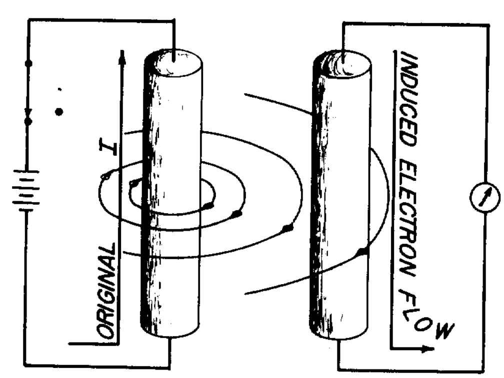
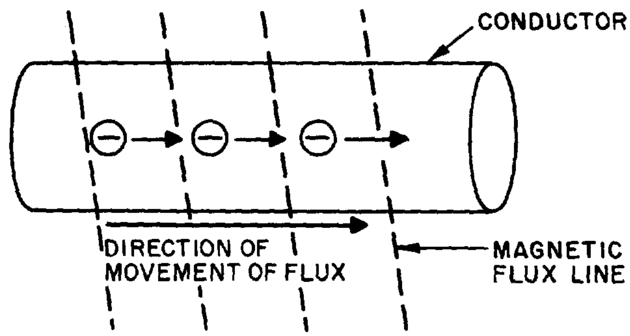
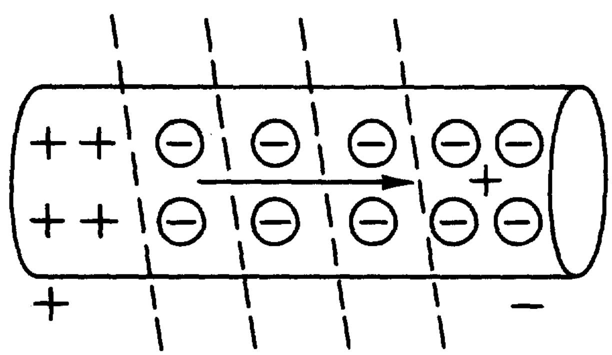


Figure 2-1.—Generation of an emf in an electrical conductor.

When a magnetic field moves through a stationary metallic conductor, electrons are dislodged from their orbits. The electrons move in a direction determined by the movement of the magnetic lines of flux. This is shown below:

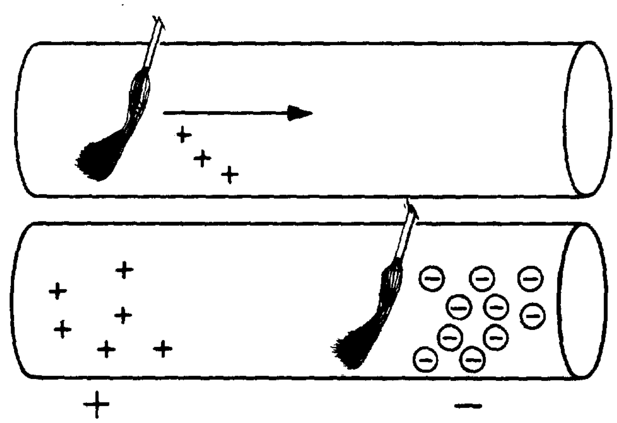


The electrons move from one area of the conductor into another area. The area that the electrons moved from has fewer negative charges (electrons) and becomes positively charged. The area the electrons move into becomes negatively charged. This is shown below:



The difference between the charges in the conductor is equal to a difference of potential (or voltage). This voltage caused by the moving magnetic field is called electromotive force (emf).

In simple terms, the action of a moving magnetic field on a conductor can be compared to the action of a broom. Consider the moving magnetic field to be a moving broom. As the magnetic broom moves along (through) the conductor, it gathers up and pushes electrons before it, as shown below:



The area from which electrons are moved becomes positively charged, while the area into which electrons are moved becomes negatively charged. The potential difference between these two areas is the electromotive force or emf.

Q2. An emf is generated in a conductor when the conductor is cut by what type of field?

SELF-INDUCTANCE

Even a perfectly straight length of conductor has some inductance. As you know, current in a conductor produces a magnetic field surrounding the conductor. When the current changes, the magnetic field changes. This causes relative motion between the magnetic field and the conductor, and an electromotive force (emf) is induced in the conductor. This emf is called a SELF-INDUCED EMF because it is induced in the conductor carrying the current. The emf produced by this moving magnetic field is also referred to as COUNTER ELECTROMOTIVE FORCE (cemf). The polarity of the counter electromotive force is in the opposite direction to the applied voltage of the conductor. The overall effect will be to oppose a change in current magnitude. This effect is summarized by Lenz's law which states that: THE INDUCED EMF IN ANY CIRCUIT IS ALWAYS IN A DIRECTION TO OPPOSE THE EFFECT THAT PRODUCED IT.

If the shape of the conductor is changed to form a loop, then the electromagnetic field around each portion of the conductor cuts across some other portion of the same conductor. This is shown in its simplest form in figure 2-2. A length of conductor is looped so that two portions of the conductor lie next to each other. These portions are labeled conductor 1 and conductor 2. When the switch is closed, current (electron flow) in the conductor produces a magnetic field around ALL portions of the conductor. For simplicity, the magnetic field (expanding lines of flux) is shown in a single plane that is perpendicular to both conductors. Although the expanding field of flux originates at the same time in both conductors, it is considered as originating in conductor 1 and its effect on conductor 2 will be explained. With increasing current, the flux field expands outward from conductor 1, cutting across a portion of conductor 2. This results in an induced emf in conductor 2 as shown by the dashed arrow. Note that the induced emf is in the opposite direction to (in OPPOSITION to) the battery current and voltage, as stated in Lenz's law.

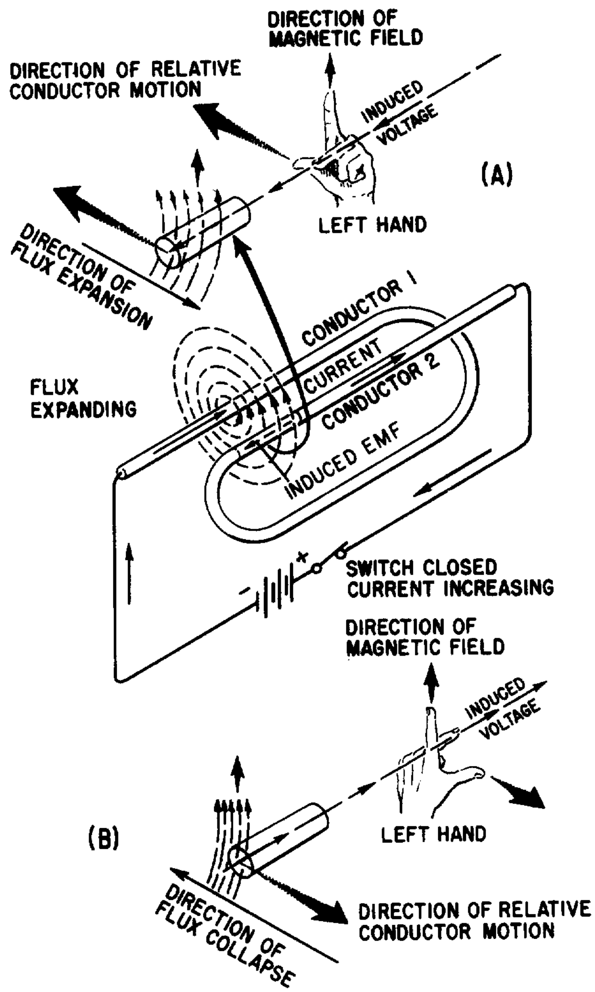


Figure 2-2.—Self-inductance.

The direction of this induced voltage may be determined by applying the LEFT-HAND RULE FOR GENERATORS. This rule is applied to a portion of conductor 2 that is "lifted" and enlarged for this purpose in figure 2-2(A). This rule states that if you point the thumb of your left hand in the direction of relative motion of the conductor and your index finger in the direction of the magnetic field, your middle finger, extended as shown, will now indicate the direction of the induced current which will generate the induced voltage (cemf) as shown.

In figure 2-2(B), the same section of conductor 2 is shown after the switch has been opened. The flux field is collapsing. Applying the left-hand rule in this case shows that the reversal of flux MOVEMENT has caused a reversal in the direction of the induced voltage. The induced voltage is now in the same direction as the battery voltage. The most important thing for you to note is that the self-induced voltage opposes BOTH changes in current. That is, when the switch is closed, this voltage delays the initial buildup of current by opposing the battery voltage. When the switch is opened, it keeps the current flowing in the same direction by aiding the battery voltage.

Thus, from the above explanation, you can see that when a current is building up it produces an expanding magnetic field. This field induces an emf in the direction opposite to the actual flow of current.

This induced emf opposes the growth of the current and the growth of the magnetic field. If the increasing current had not set up a magnetic field, there would have been no opposition to its growth. The whole reaction, or opposition, is caused by the creation or collapse of the magnetic field, the lines of which as they expand or contract cut across the conductor and develop the counter emf.

Since all circuits have conductors in them, you can assume that all circuits have inductance. However, inductance has its greatest effect only when there is a change in current. Inductance does NOT oppose current, only a CHANGE in current. Where current is constantly changing as in an ac circuit, inductance has more effect.

Q3. Define inductance.

Q4. What is meant by induced emf? By counter emf?

Q5. State Lenz's law.

Q6. What effect does inductance have (a) on steady direct current and (b) on direct current while it is changing in amplitude?

To increase the property of inductance, the conductor can be formed into a loop or coil. A coil is also called an inductor. Figure 2-3 shows a conductor formed into a coil. Current through one loop produces a magnetic field that encircles the loop in the direction as shown in figure 2-3(A). As current increases, the magnetic field expands and cuts all the loops as shown in figure 2-3(B). The current in each loop affects all other loops. The field cutting the other loop has the effect of increasing the opposition to a current change.

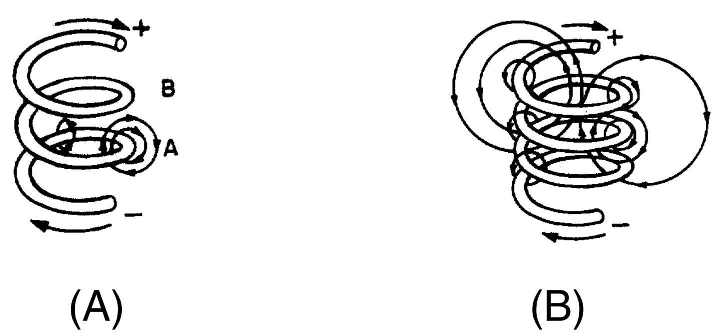


Figure 2-3.—Inductance.

Inductors are classified according to core type. The core is the center of the inductor just as the core of an apple is the center of an apple. The inductor is made by forming a coil of wire around a core. The core material is normally one of two basic types: soft-iron or air. An iron-core inductor and its schematic symbol (which is represented with lines across the top of it to indicate the presence of an iron core) are shown, in figure 2-4(A). The air-core inductor may be nothing more than a coil of wire, but it is usually a coil formed around a hollow form of some nonmagnetic material such as cardboard. This material serves no purpose other than to hold the shape of the coil. An air-core inductor and its schematic symbol are shown in figure 2-4(B).

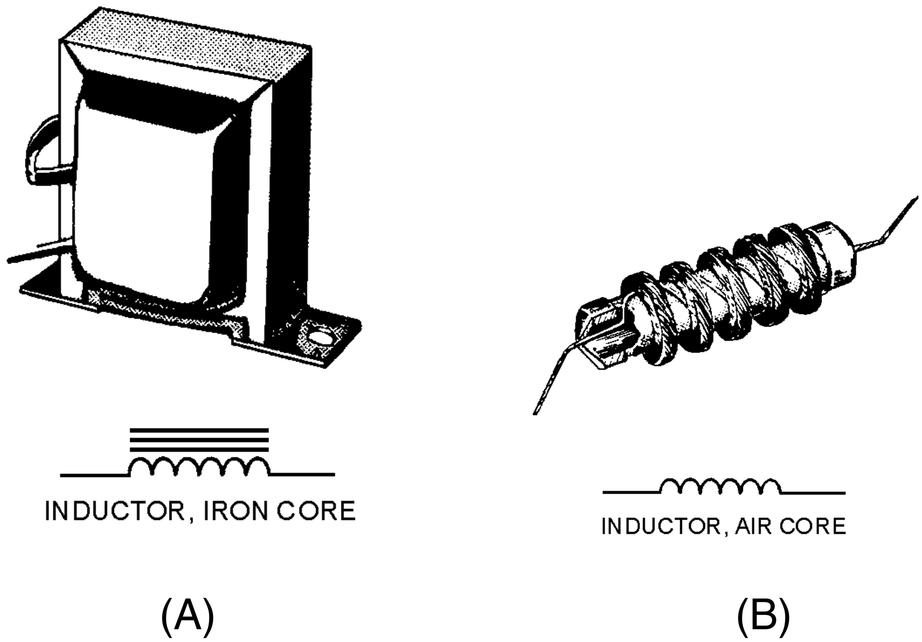


Figure 2-4.—Inductor types and schematic symbols.

Factors Affecting Coil Inductance

There are several physical factors which affect the inductance of a coil. They include the number of turns in the coil, the diameter of the coil, the coil length, the type of material used in the core, and the number of layers of winding in the coils.

Inductance depends entirely upon the physical construction of the circuit, and can only be measured with special laboratory instruments. Of the factors mentioned, consider first how the number of turns affects the inductance of a coil. Figure 2-5 shows two coils. Coil (A) has two turns and coil (B) has four turns. In coil (A), the flux field set up by one loop cuts one other loop. In coil (B), the flux field set up by one loop cuts three other loops. Doubling the number of turns in the coil will produce a field twice as strong, if the same current is used. A field twice as strong, cutting twice the number of turns, will induce four times the voltage. Therefore, it can be said that the <u>inductance varies as the square of the number of turns</u>.

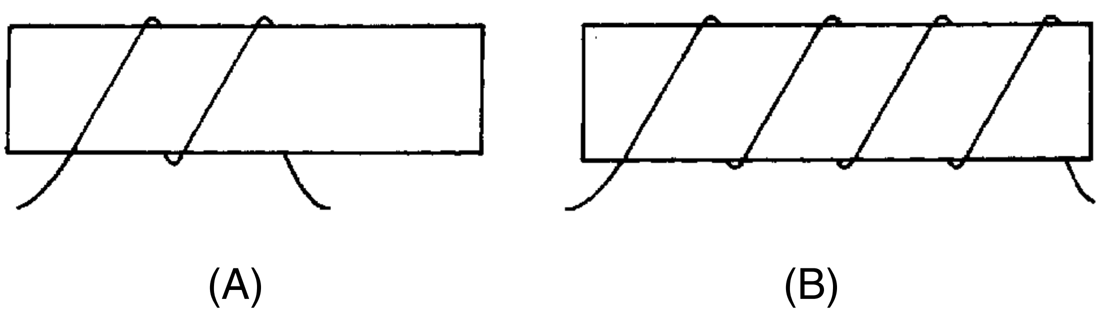


Figure 2-5.—Inductance factor (turns).

The second factor is the coil diameter. In figure 2-6 you can see that the coil in view B has twice the diameter of coil view A. Physically, it requires more wire to construct a coil of large diameter than one of small diameter with an equal number of turns. Therefore, more lines of force exist to induce a counter emf in the coil with the larger diameter. Actually, the inductance of a coil increases directly as the cross-sectional area of the core increases. Recall the formula for the area of a circle: $A = \pi r^2$. Doubling the radius of a coil increases the inductance by a factor of four.

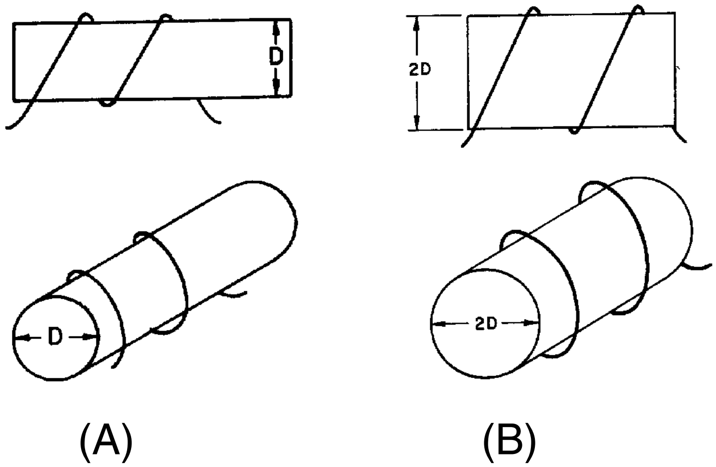


Figure 2-6.—Inductance factor (diameter).

The third factor that affects the inductance of a coil is the length of the coil. Figure 2-7 shows two examples of coil spacings. Coil (A) has three turns, rather widely spaced, making a relatively long coil. A coil of this type has few flux linkages, due to the greater distance between each turn. Therefore, coil (A) has a relatively low inductance. Coil (B) has closely spaced turns, making a relatively short coil. This close spacing increases the flux linkage, increasing the inductance of the coil. Doubling the length of a coil while keeping the same number of turns halves the value of inductance.

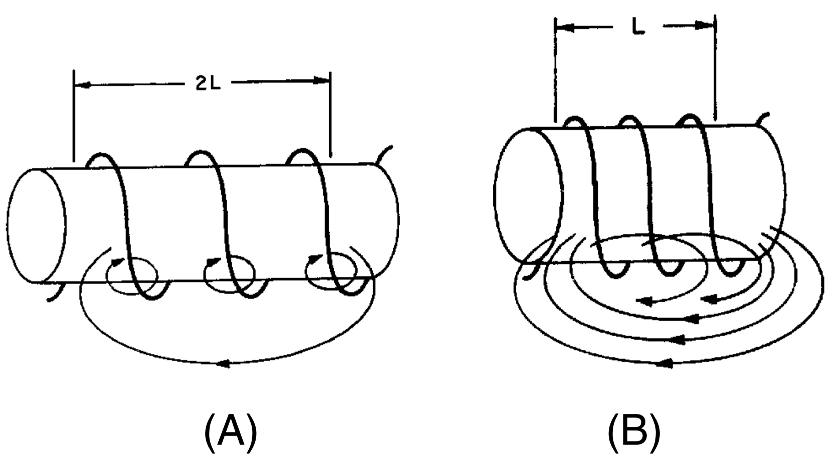


Figure 2-7.—Inductance factor (coil length). CLOSELY WOUND

The fourth physical factor is the type of core material used with the coil. Figure 2-8 shows two coils: Coil (A) with an air core, and coil (B) with a soft-iron core. The magnetic core of coil (B) is a better path for magnetic lines of force than is the nonmagnetic core of coil (A). The soft-iron magnetic core's high permeability has less reluctance to the magnetic flux, resulting in more magnetic lines of force. This increase in the magnetic lines of force increases the number of lines of force cutting each loop of the coil, thus increasing the inductance of the coil. It should now be apparent that the inductance of a coil increases directly as the permeability of the core material increases.

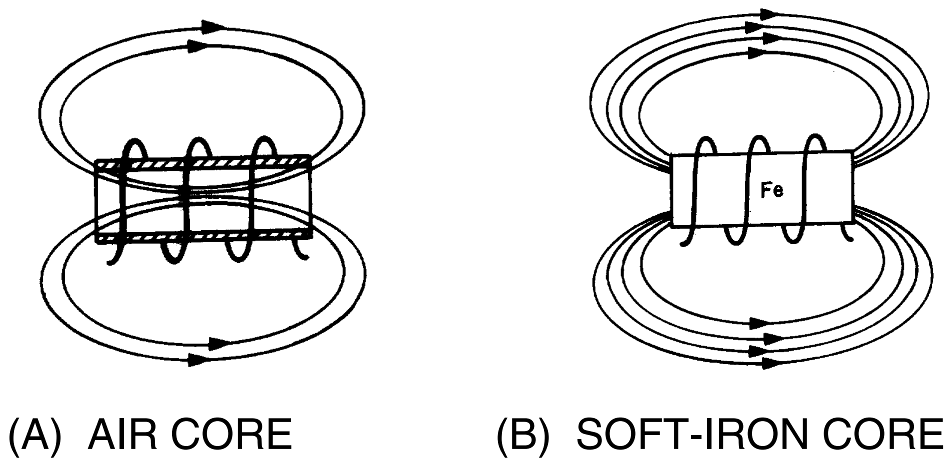


Figure 2-8.—Inductance factor (core material).

Another way of increasing the inductance is to wind the coil in layers. Figure 2-9 shows three cores with different amounts of layering. The coil in figure 2-9(A) is a poor inductor compared to the others in the figure because its turns are widely spaced and there is no layering. The flux movement, indicated by the dashed arrows, does not link effectively because there is only one layer of turns. A more inductive coil is shown in figure 2-9(B). The turns are closely spaced and the wire has been wound in two layers. The two layers link each other with a greater number of flux loops during all flux movements. Note that nearly all the turns, such as X, are next to four other turns (shaded). This causes the flux linkage to be increased.

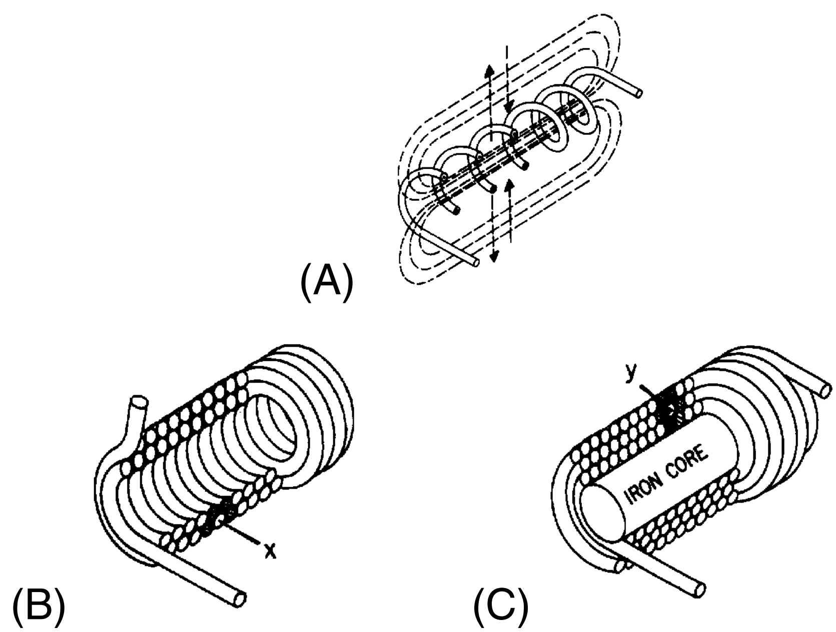


Figure 2-9.—Coils of various inductances.

A coil can be made still more inductive by winding it in three layers, as shown in figure 2-9(C). The increased number of layers (cross-sectional area) improves flux linkage even more. Note that some turns, such as Y, lie directly next to six other turns (shaded). In actual practice, layering can continue on through many more layers. The important fact to remember, however, is that the inductance of the coil increases with each layer added.

As you have seen, several factors can affect the inductance of a coil, and all of these factors are variable. Many differently constructed coils can have the same inductance. The important information to remember, however, is that inductance is dependent upon the degree of linkage between the wire conductor(s) and the electromagnetic field. In a straight length of conductor, there is very little flux linkage between one part of the conductor and another. Therefore, its inductance is extremely small. It was shown that conductors become much more inductive when they are wound into coils. This is true because there is maximum flux linkage between the conductor turns, which lie side by side in the coil.

Q7.

a. *List five factors that affect the inductance of a coil.*

b. *Bending a straight piece of wire into a loop or coil has what effect on the inductance of the wire?*

c. *Doubling the number of turns in a coil has what effect on the inductance of the coil?*

d. *Decreasing the diameter of a coil has what effect on the inductance of the coil?*

e. *Inserting a soft-iron core into a coil has what effect on the inductance of the coil?*

f. *Increasing the number of layers of windings in a coil has what effect on the inductance of the coil?*

UNIT OF INDUCTANCE

As stated before, the basic unit of inductance (L) is the HENRY (H), named after Joseph Henry, the co-discoverer with Faraday of the principle of electromagnetic induction. An inductor has an inductance of 1 henry if an emf of 1 volt is induced in the inductor when the current through the inductor is changing at the rate of 1 ampere per second. The relationship between the induced voltage, the inductance, and the rate of change of current with respect to time is stated mathematically as:

$$E_{ind} = L\frac{\Delta I}{\Delta t}$$

where E_{ind} is the induced emf in volts; L is the inductance in henrys; and ΔI is the change in current in amperes occurring in Δt seconds. The symbol Δ (Greek letter delta), means "a change in ....". The henry is a large unit of inductance and is used with relatively large inductors. With small inductors, the millihenry is used. (A millihenry is equal to 1×10^{-3} henry, and one henry is equal to 1,000 millihenrys.) For still smaller inductors the unit of inductance is the microhenry (μH). (A $\mu H = 1 \times 10^{-6} H$, and one henry is equal to 1,000,000 microhenrys.)

GROWTH AND DECAY OF CURRENT IN AN LR SERIES CIRCUIT

When a battery is connected across a "pure" inductance, the current builds up to its final value at a rate determined by the battery voltage and the internal resistance of the battery. The current buildup is gradual because of the counter emf generated by the self-inductance of the coil. When the current starts to flow, the magnetic lines of force move outward from the coil. These lines cut the turns of wire on the inductor and build up a counter emf that opposes the emf of the battery. This opposition causes a delay in the time it takes the current to build up to a steady value. When the battery is disconnected, the lines of force collapse. Again these lines cut the turns of the inductor and build up an emf that tends to prolong the flow of current.

A voltage divider containing resistance and inductance may be connected in a circuit by means of a special switch, as shown in figure 2-10(A). Such a series arrangement is called an LR series circuit.

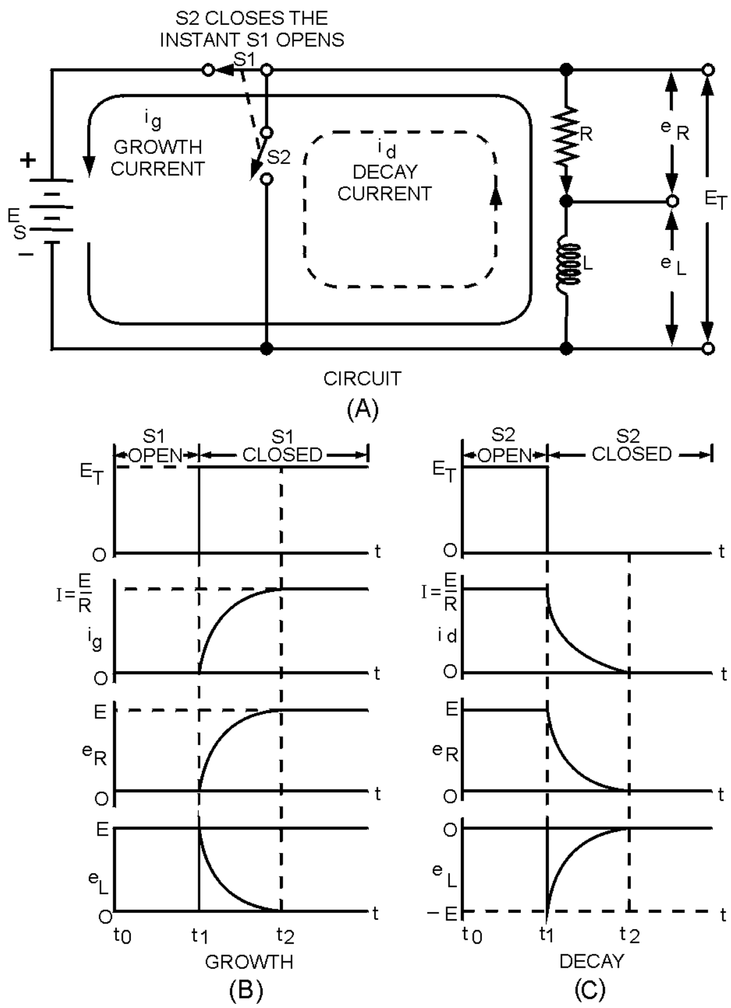


Figure 2-10.—Growth and decay of current in an LR series circuit.

When switch S_1 is closed (as shown), a voltage E_S appears across the voltage divider. At this instant the current will attempt to increase to its maximum value. However, this instantaneous current change causes coil L to produce a back EMF, which is opposite in polarity and almost equal to the EMF of the source. This back EMF opposes the rapid current change. Figure 2-10(B) shows that at the instant switch S_1 is closed, there is no measurable growth current (i_g), a minimum voltage drop is across resistor R, and maximum voltage exists across inductor L.

As current starts to flow, a voltage (e_R) appears across R, and the voltage across the inductor is reduced by the same amount. The fact that the voltage across the inductor (L) is reduced means that the growth current (i_g) is increased and consequently e_R is increased. Figure 2-10(B) shows that the voltage across the inductor (e_L) finally becomes zero when the growth current (i_g) stops increasing, while the voltage across the resistor (e_R) builds up to a value equal to the source voltage (E_S).

Electrical inductance is like mechanical inertia, and the growth of current in an inductive circuit can be likened to the acceleration of a boat on the surface of the water. The boat does not move at the instant a constant force is applied to it. At this instant all the applied force is used to overcome the inertia of the boat. Once the inertia is overcome the boat will start to move. After a while, the speed of the boat reaches its maximum value and the applied force is used up in overcoming the friction of the water against the hull.

When the battery switch (S_1) in the LR circuit of figure 2-10(A) is closed, the rate of the current increase is maximum in the inductive circuit. At this instant all the battery voltage is used in overcoming the emf of self-induction which is a maximum because the rate of change of current is maximum. Thus the battery voltage is equal to the drop across the inductor and the voltage across the resistor is zero. As time goes on more of the battery voltage appears across the resistor and less across the inductor. The rate of change of current is less and the induced emf is less. As the steady-state condition of the current is approached, the drop across the inductor approaches zero and all of the battery voltage is "dropped" across the resistance of the circuit.

Thus the voltages across the inductor and the resistor change in magnitude during the period of growth of current the same way the force applied to the boat divides itself between the effects of inertia and friction. In both examples, the force is developed first across the inertia/inductive effect and finally across the friction/resistive effect.

Figure 2-10(C) shows that when switch S_2 is closed (source voltage E_S removed from the circuit), the flux that has been established around the inductor (L) collapses through the windings. This induces a voltage e_L in the inductor that has a polarity opposite to E_S and is essentially equal to E_S in magnitude. The induced voltage causes decay current (i_d) to flow in resistor R in the same direction in which current was flowing originally (when S_1 was closed). A voltage (e_R) that is initially equal to source voltage (E_S) is developed across R. The voltage across the resistor (e_R) rapidly falls to zero as the voltage across the inductor (e_L) falls to zero due to the collapsing flux.

Just as the example of the boat was used to explain the growth of current in a circuit, it can also be used to explain the decay of current in a circuit. When the force applied to the boat is removed, the boat still continues to move through the water for a while, eventually coming to a stop. This is because energy was being stored in the inertia of the moving boat. After a period of time the friction of the water overcomes the inertia of the boat, and the boat stops moving. Just as inertia of the boat stored energy, the magnetic field of an inductor stores energy. Because of this, even when the power source is removed, the stored energy of the magnetic field of the inductor tends to keep current flowing in the circuit until the magnetic field collapse.

Q8.

a. *When voltage is first applied to a series LR circuit, how much opposition does the inductance have to the flow of current compared to that of the circuit resistance?*

b. *In a series circuit containing a resistor (R_1) and an inductor (L_1), what voltage exists across R_1 when the counter emf is at its maximum value?*

c. *What happens to the voltage across the resistance in an LR circuit during current buildup in the circuit, and during current decay in the circuit?*

L/R Time Constant

The L/R TIME CONSTANT is a valuable tool for use in determining the time required for current in an inductor to reach a specific value. As shown in figure 2-11, one L/R time constant is the time required for the current in an inductor to increase to 63 percent (actually 63.2 percent) of the maximum current. Each time constant is equal to the time required for the current to increase by 63.2 percent of the difference in value between the current flowing in the inductor and the maximum current. Maximum current flows in the inductor after five L/R time constants are completed. The following example should clear up any confusion about time constants. Assume that maximum current in an LR circuit is 10 amperes. As you know, when the circuit is energized, it takes time for the current to go from zero to 10 amperes. When the first time constant is completed, the current in the circuit is equal to 63.2% of 10 amperes. Thus the amplitude of current at the end of 1 time constant is 6.32 amperes.

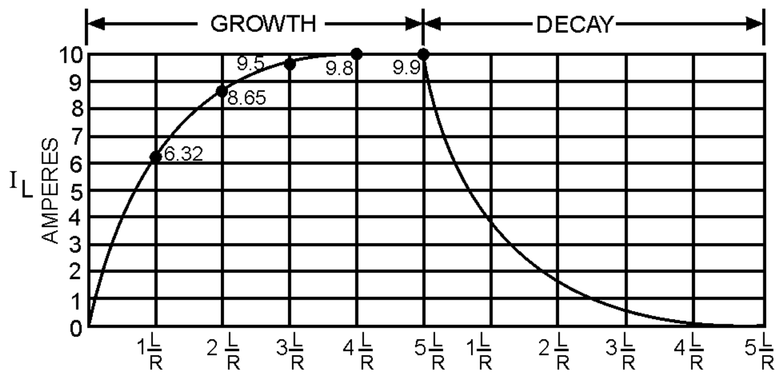


Figure 2-11.—L/R time constant.

During the second time constant, current again increases by 63.2% (.632) of the difference in value between the current flowing in the inductor and the maximum current. This difference is 10 amperes minus 6.32 amperes and equals 3.68 amperes; 63.2% of 3.68 amperes is 2.32 amperes. This increase in current during the second time constant is added to that of the first time constant. Thus, upon completion of the second time constant, the amount of current in the LR circuit is 6.32 amperes + 2.32 amperes = 8.64 amperes.

During the third constant, current again increases:

$$10 \text{ amperes} - 8.64 \text{ amperes} = 1.36 \text{ amperes}$$
$$1.36 \text{ amperes} \times .632 = 0.860 \text{ ampere}$$
$$8.64 \text{ amperes} + 0.860 \text{ ampere} = 9.50 \text{ amperes}$$

During the fourth time constant, current again increases:

$$10 \text{ amperes} - 9.50 \text{ amperes} = 0.5 \text{ ampere}$$
$$0.5 \text{ ampere} \times .632 = 0.316 \text{ ampere}$$
$$9.50 \text{ amperes} + 0.316 \text{ ampere} = 9.82 \text{ amperes}$$

During the fifth time constant, current increases as before:

$$10 \text{ amperes} - 9.82 \text{ amperes} = 0.18 \text{ ampere}$$
$$0.18 \text{ ampere} \times .632 = 0.114 \text{ ampere}$$
$$9.82 \text{ amperes} + .114 \text{ ampere} = 9.93 \text{ amperes}$$

Thus, the current at the end of the fifth time constant is almost equal to 10.0 amperes, the maximum current. For all practical purposes the slight difference in value can be ignored.

When an LR circuit is deenergized, the circuit current decreases (decays) to zero in five time constants at the same rate that it previously increased. If the growth and decay of current in an LR circuit are plotted on a graph, the curve appears as shown in figure 2-11. Notice that current increases and decays at the same rate in five time constants.

The value of the time constant in seconds is equal to the inductance in henrys divided by the circuit resistance in ohms.

The formula used to calculate one L/R time constant is:

$$\text{Time Constant (TC) in seconds} = \frac{L \text{ (in henrys)}}{R \text{ (in ohms)}}$$

Q9. What is the formula for one L/R time constant?

Q10.

a. The maximum current applied to an inductor is 1.8 amperes. How much current flowed in the inductor 3 time constants after the circuit was first energized?

b. What is the minimum number of time constants required for the current in an LR circuit to increase to its maximum value?

c. A circuit containing only an inductor and a resistor has a maximum of 12 amperes of applied current flowing in it. After 5 L/R time constants the circuit is opened. How many time constants is required for the current to decay to 1.625 amperes?

POWER LOSS IN AN INDUCTOR

Since an inductor (coil) consists of a number of turns of wire, and since all wire has some resistance, every inductor has a certain amount of resistance. Normally this resistance is small. It is usually neglected in solving various types of ac circuit problems because the reactance of the inductor (the opposition to alternating current, which will be discussed later) is so much greater than the resistance that the resistance has a negligible effect on the current.

However, since some inductors are designed to carry relatively large amounts of current, considerable power can be dissipated in the inductor even though the amount of resistance in the inductor is small. This power is wasted power and is called COPPER LOSS. The copper loss of an inductor can be calculated by multiplying the square of the current in the inductor by the resistance of the winding (I^2R).

In addition to copper loss, an iron-core coil (inductor) has two iron losses. These are called HYSTERESIS LOSS and EDDY-CURRENT LOSS. Hysteresis loss is due to power that is consumed in reversing the magnetic field of the inductor core each time the direction of current in the inductor changes.

Eddy-current loss is due to heating of the core by circulating currents that are induced in the iron core by the magnetic field around the turns of the coil. These currents are called eddy currents and circulate within the iron core only.

All these losses dissipate power in the form of heat. Since this power cannot be returned to the electrical circuit, it is lost power.

Q11. State three types of power loss in an inductor.

MUTUAL INDUCTANCE

Whenever two coils are located so that the flux from one coil links with the turns of the other coil, a change of flux in one coil causes an emf to be induced in the other coil. This allows the energy from one coil to be transferred or coupled to the other coil. The two coils are said to be coupled or linked by the property of MUTUAL INDUCTANCE (M). The amount of mutual inductance depends on the relative positions of the two coils. This is shown in figure 2-12. If the coils are separated a considerable distance, the amount of flux common to both coils is small and the mutual inductance is low. Conversely, if the coils are close together so that nearly all the flux of one coil links the turns of the other, the mutual inductance is high. The mutual inductance can be increased greatly by mounting the coils on a common iron core.

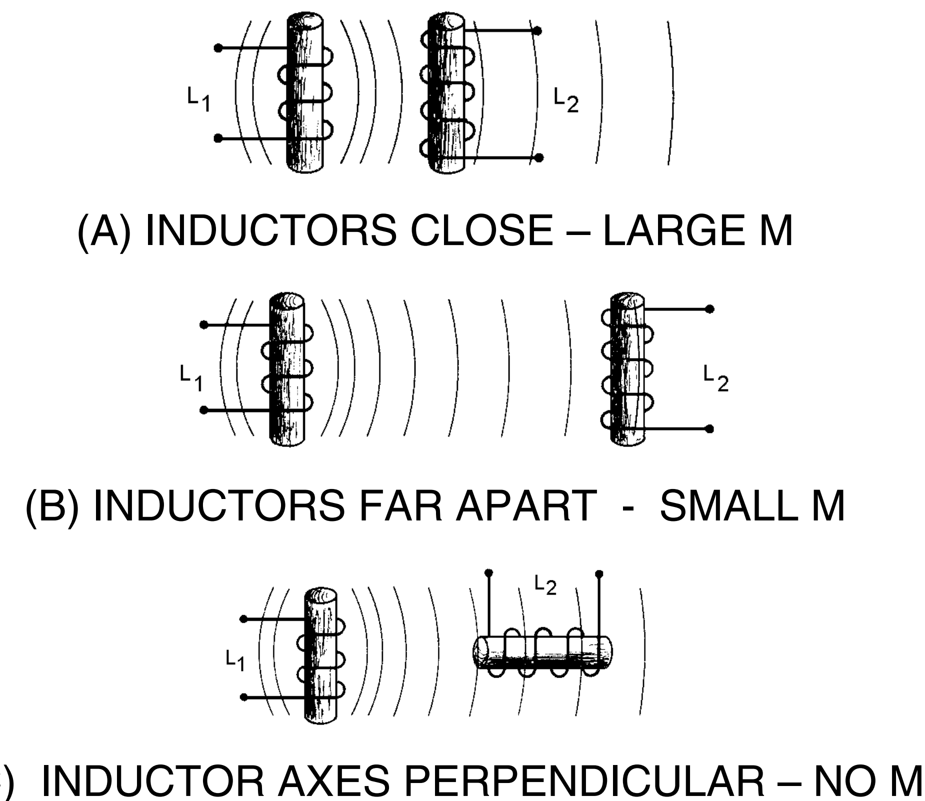


Figure 2-12.—The effect of position of coils on mutual inductance (M).

Two coils are placed close together as shown in figure 2-13. Coil 1 is connected to a battery through switch S, and coil 2 is connected to an ammeter (A). When switch S is closed as in figure 2-13(A), the current that flows in coil 1 sets up a magnetic field that links with coil 2, causing an induced voltage in coil 2 and a momentary deflection of the ammeter. When the current in coil 1 reaches a steady value, the ammeter returns to zero. If switch S is now opened as in figure 2-13(B), the ammeter (A) deflects momentarily in the opposite direction, indicating a momentary flow of current in the opposite direction in coil 2. This current in coil 2 is produced by the collapsing magnetic field of coil 1.

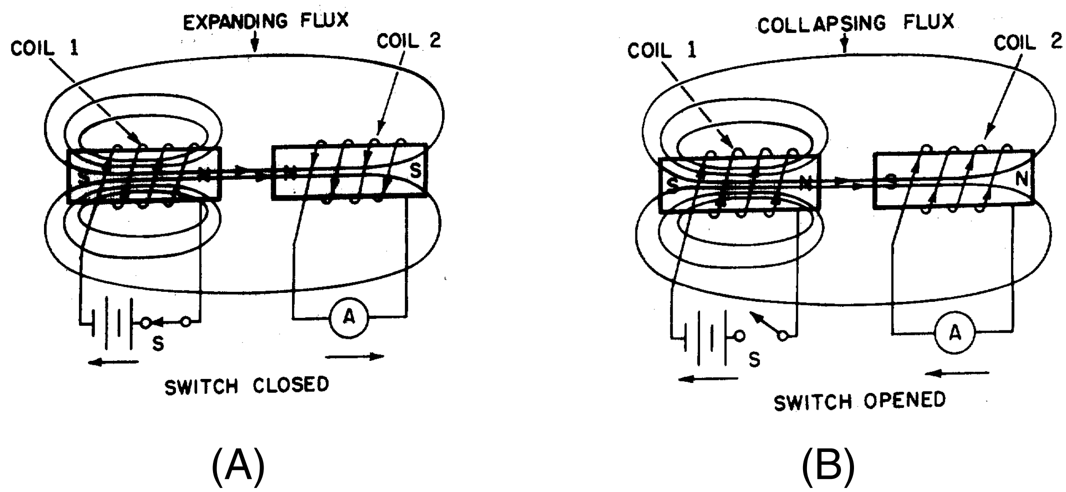


Figure 2-13.—Mutual inductance.

Q12. Define mutual inductance.

FACTORS AFFECTING MUTUAL INDUCTANCE

The mutual inductance of two adjacent coils is dependent upon the physical dimensions of the two coils, the number of turns in each coil, the distance between the two coils, the relative positions of the axes of the two coils, and the permeability of the cores.

The COEFFICIENT OF COUPLING between two coils is equal to the ratio of the flux cutting one coil to the flux originated in the other coil. If the two coils are so positioned with respect to each other so that all of the flux of one coil cuts all of the turns of the other, the coils are said to have a unity coefficient of coupling. It is never exactly equal to unity (1), but it approaches this value in certain types of coupling devices. If all of the flux produced by one coil cuts only half the turns of the other coil, the coefficient of coupling is 0.5. The coefficient of coupling is designated by the letter K.

The mutual inductance between two coils, L_1 and L_2, is expressed in terms of the inductance of each coil and the coefficient of coupling K. As a formula:

$$M = K\sqrt{L_1 L_2}$$

$$\begin{aligned} \text{where:} \quad M &= \text{Mutual inductance in henrys} \\ K &= \text{Coefficient of coupling} \\ L_1, L_2 &= \text{Inductance of coil in henrys} \end{aligned}$$

Example problem:

One 10-H coil and one 20-H coil are connected in series and are physically close enough to each other so that their coefficient of coupling is 0.5. What is the mutual inductance between the coils?

$$\begin{aligned} \text{Use the formula:} \quad M &= K\sqrt{L_1 L_2} \\ M &= 0.5\sqrt{(10H)(20H)} \\ M &= 0.5\sqrt{200H} \\ M &= 0.5 \times 14.14H \\ M &= 7.07H \end{aligned}$$

Q13. When are two circuits said to be coupled?

Q14. What is meant by the coefficient of coupling?

SERIES INDUCTORS WITHOUT MAGNETIC COUPLING

When inductors are well shielded or are located far enough apart from one another, the effect of mutual inductance is negligible. If there is no mutual inductance (magnetic coupling) and the inductors are connected in series, the total inductance is equal to the sum of the individual inductances. As a formula:

$$L_T = L_1 + L_2 + L_3 + .. L_n$$

where L_T is the total inductance; L_1, L_2, L_3 are the inductances of L_1, L_2, L_3; and L_n means that any number (n) of inductors may be used. The inductances of inductors in series are added together like the resistances of resistors in series.

SERIES INDUCTORS WITH MAGNETIC COUPLING

When two inductors in series are so arranged that the field of one links the other, the combined inductance is determined as follows:

$$L_T = L_1 + L_2 \pm 2M$$

$$\begin{aligned} \text{where:} \quad L_T &= \text{The total inductance} \\ L_1, L_2 &= \text{The inductances of } L_1, L_2 \\ M &= \text{The mutual inductance between the two inductors} \end{aligned}$$

The plus sign is used with M when the magnetic fields of the two inductors are aiding each other, as shown in figure 2-14. The minus sign is used with M when the magnetic field of the two inductors oppose each other, as shown in figure 2-15. The factor 2M accounts for the influence of L_1 on L_2 and L_2 on L_1.

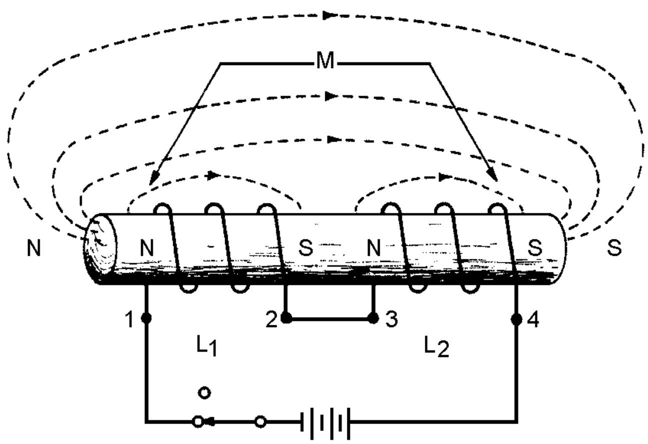


Figure 2-14.—Series inductors with aiding fields.

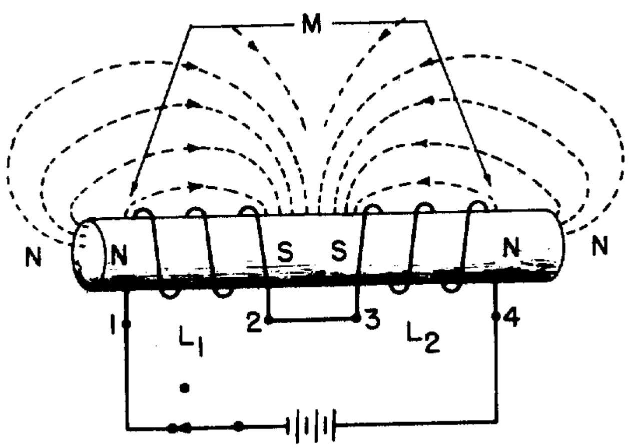


Figure 2-15.—Series inductors with opposing fields.

Example problem:

A 10-H coil is connected in series with a 5-H coil so the fields aid each other. Their mutual inductance is 7 H. What is the combined inductance of the coils?

Use the formula: $L_T = L_1 + L_2 + 2M$

$$L_T = 10\text{ H} + 5\text{ H} + 2(7\text{ H})$$

$$L_T = 29\text{ H}$$

Q15. Two series-connected 7-H inductors are adjacent to each other; their coefficient of coupling is 0.64. What is the value of M?

Q16. A circuit contains two series inductors aligned in such a way that their magnetic fields oppose each other. What formula should you use to compute total inductance in this circuit?

Q17. The magnetic fields of two coils are aiding each other. The inductance of the coils are 3 H and 5 H, respectively. The coils' mutual inductance is 5 H. What is their combined inductance?

PARALLEL INDUCTORS WITHOUT COUPLING

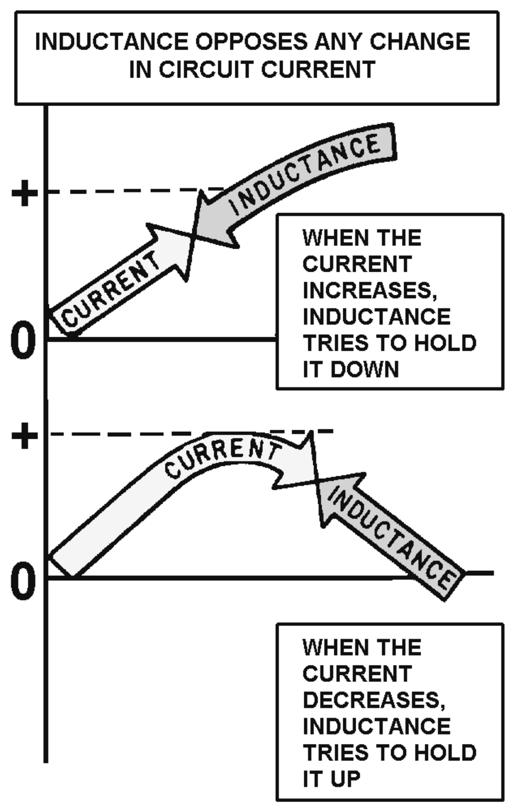
The total inductance (L_T) of inductors in parallel is calculated in the same manner that the total resistance of resistors in parallel is calculated, provided the coefficient of coupling between the coils is zero. Expressed mathematically:

$$\frac{1}{L_T} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} \dots + \frac{1}{L_N}$$

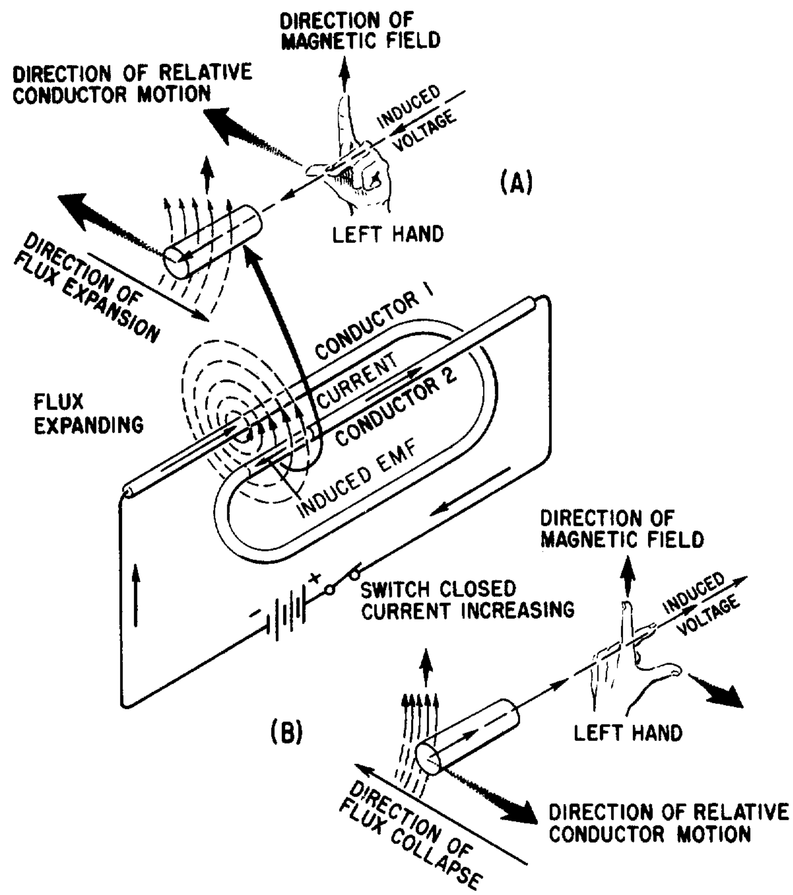
SUMMARY

The important points of this chapter are summarized below. Study this information before continuing, as this information will lay the foundation for later chapters.

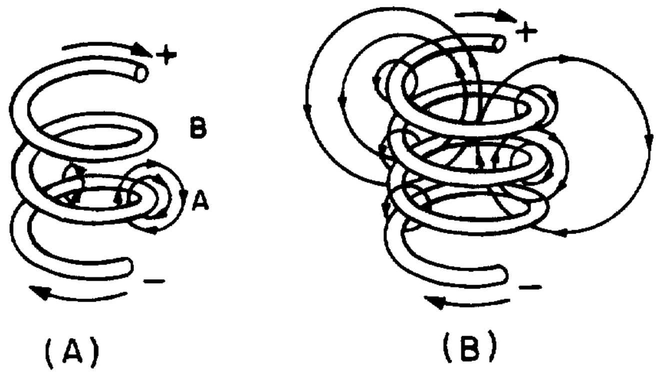
INDUCTANCE—The characteristic of an electrical circuit that opposes a change in current. The reaction (opposition) is caused by the creation or destruction of a magnetic field. When current starts to flow, magnetic lines of force are created. These lines of force cut the conductor inducing a counter emf in a direction that opposes current.



SELF-INDUCTANCE—The process by which a circuit induces an emf into itself by its own moving magnetic field. All electrical circuits possess self-inductance. This opposition (inductance), however, only takes place when there is a change in current. Inductance does NOT oppose current, only a CHANGE in current. The property of inductance can be increased by forming the conductor into a loop. In a loop, the magnetic lines of force affect more of the conductor at one time. This increases the self-induced emf.

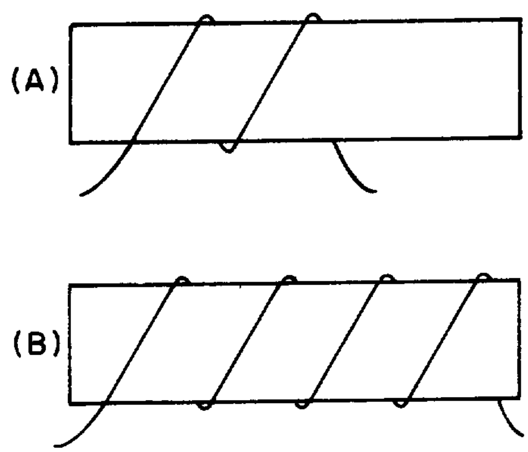


INDUCTANCE OF A COIL—The property of inductance can be further increased if the conductor is formed into a coil. Because a coil contains more loops, more of the conductor can be affected by the magnetic field. Inductors (coils) are classified according to core type. The core material is normally either air or soft iron.

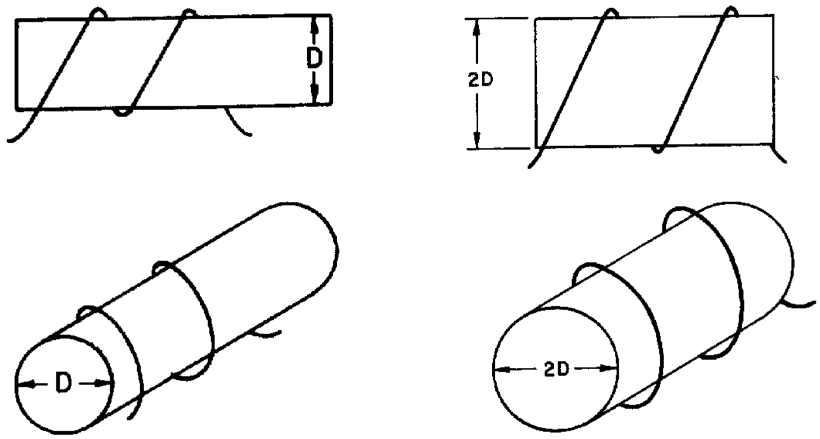


FACTORS AFFECTING COIL INDUCTANCE—The inductance of a coil is entirely dependent upon its physical construction. Some of the factors affecting the inductance are:

- The number of turns in the coil. Increasing the number of turns will increase the inductance.

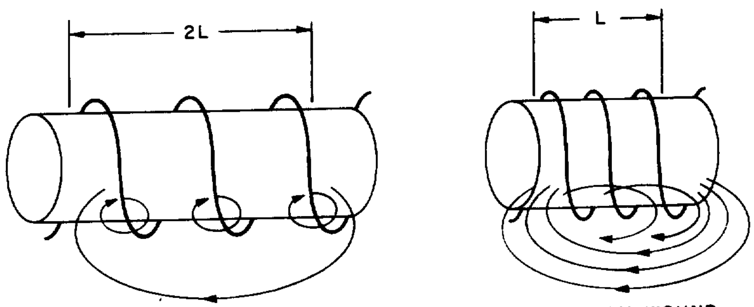


- The coil diameter. The inductance increases directly as the cross-sectional area of the coil increases.



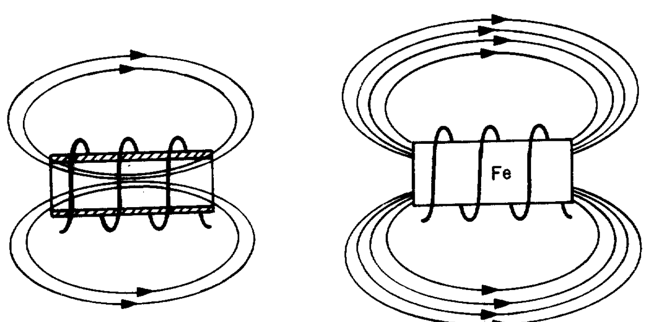
(A) (B)

- The length of the coil. When the length of the coil is increased while keeping the number of turns the same, the turn-spacing is increased. This decreases the inductance of the coil.



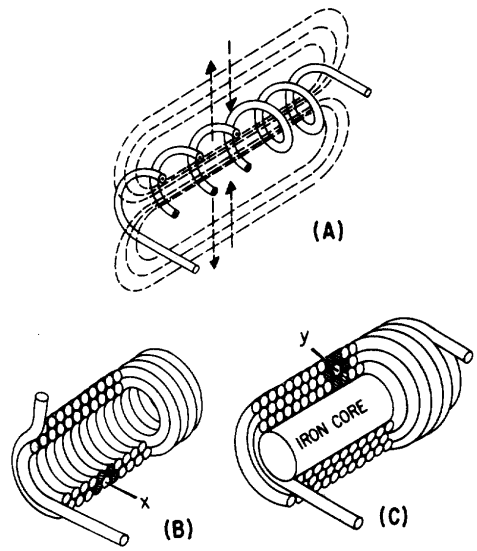
(A) WIDELY SPACED (B) CLOSELY WOUND

- The type of core material. Increasing the permeability of the core results in increasing the inductance of the coil.



(A) AIR CORE (B) SOFT-IRON CORE

- Winding the coil in layers. The more layers used to form a coil, the greater effect the magnetic field has on the conductor. By layering a coil, you can increase the inductance.



UNIT OF INDUCTANCE—Inductance (L) is measured in henrys (H). An inductor has an inductance of one henry (H) if an emf of one volt is induced in the inductor when the current through the inductor is changing at the rate of 1 ampere per second. Common units of inductance are henry (H), millihenry (mH) and the microhenry (μH).

GROWTH AND DECAY OF CURRENT IN AN LR CIRCUIT—The required for the current in an inductor to increase to 63.2 percent of the maximum current or to decrease to 36.8 percent of the maximum current is known as the time constant. The letter symbol for an LR time constant is L/R. As a formula:

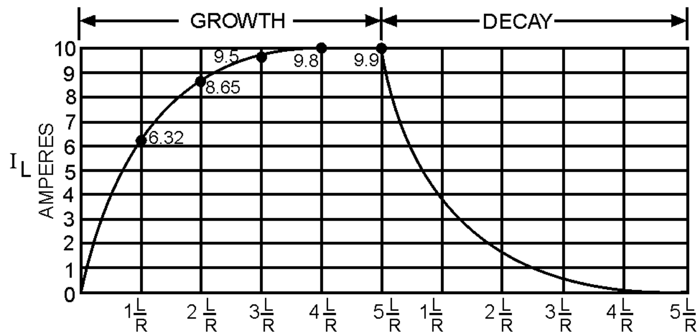
$$\frac{L}{R}.$$

As a formula:

$$t \text{ (in seconds)} = \frac{L \text{ (in henrys)}}{R \text{ (in ohms)}}$$

or

$$t \text{ (in microseconds)} = \frac{L \text{ (in microhenrys)}}{R \text{ (in ohms)}}$$

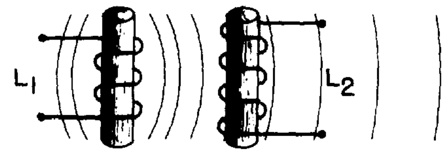


The time constant of an LR circuit may also be defined as the time required for the current in the inductor to grow or decay to its final value if it continued to grow or decay at its initial rate. For all practical purposes, the current in the inductor reaches a maximum value in 5 "Time Constants" and decreases to zero in 5 "Time Constants".

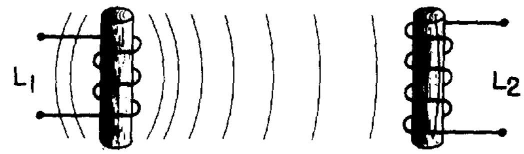
POWER LOSSES IN AN INDUCTOR—Since an inductor (coil) contains a number turns of wire, and all wire has some resistance, the inductor has a certain amount of resistance. This resistance is usually very small and has a negligible effect on current. However, there are power losses in an inductor. The main power losses in an inductor are copper loss, hysteresis loss, and eddy-current loss. Copper loss can be calculated by multiplying the square of the current by the resistance of the wire in the coil (I_2R). Hysteresis loss is due to power that is consumed in reversing the magnetic field of the core each time the current direction changes. Eddy-current loss is due to core heating caused by circulating currents induced in an iron core by the magnetic field of the coil.

MUTUAL INDUCTANCE—When two coils are located so that the flux from one coil cuts the turns of the other coil, the coils have mutual inductance. The amount of mutual inductance depends upon several factors: the relative position of the axes of the two coils; the permeability of the cores; the physical dimensions of the two coils; the number of turns in each coil, and the distance between the coils. The coefficient of coupling K specifies the amount of coupling between the coils. If all of the flux from one coil cuts all of the turns of the other coil, the coefficient of coupling K is 1 or unity. If none of the flux from one coil cuts the turns of the other coil, the coefficient of coupling is zero. The mutual inductance between two coils (L_1 and L_2) may be expressed mathematically as:

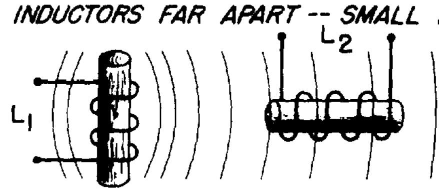
$$M = K\sqrt{L_1 L_2}$$



(A) *INDUCTORS CLOSE -- LARGE M*



(B) *INDUCTORS FAR APART -- SMALL M*



(C) *INDUCTOR AXES PERPENDICULAR -- NO M*

COMPUTING THE INDUCTANCE OF A CIRCUIT—When the total inductance of a circuit is computed, the individual inductive values are treated the same as resistance values. The inductances of inductors in series are added like the resistances of resistors in series. That is,

$$L_T = L_1 + L_2 + L_3 \ldots + L_n$$

The inductances of inductors in parallel are combined mathematically like the resistances of resistors in parallel. That is,

$$L_T = \frac{1}{\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} \ldots \frac{1}{L_n}}$$

Both of the above formulas are accurate, providing there is no mutual inductance between the inductors.

ANSWERS TO QUESTIONS Q1. THROUGH Q17.

A1. *The henry, H.*

A2. *Magnetic field.*

A3. *Inductance is the property of a coil (or circuit) which opposes any CHANGE in current.*

A4. *Induced emf is the emf which appears across a conductor when there is relative motion between the conductor and a magnetic field; counter emf is the emf induced in a conductor that opposes the applied voltage.*

A5. *The induced emf in any circuit is in a direction to oppose the effect that produced it.*

A6.

- *a.* *No effect.*
- *b.* *Inductance opposes any change in the amplitude of current.*

A7.

- a.
  1. *The numbers of turns in a coil.*
  2. *The type of material used in the core.*
  3. *The diameter of the coil.*
  4. *The coil length.*
  5. *The number of layers of windings in the coil.*
- *b.* *Increases inductance.*
- *c.* *Increases inductance.*
- *d.* *Decreases inductance.*
- *e.* *Increases inductance.*
- *f.* *Increases inductance.*

A8.

- *a.* *Inductance causes a very large opposition to the flow of current when voltage is first applied to an LR circuit; resistance causes comparatively little opposition to current at that time.*
- *b.* *Zero.*
- *c.* *During current buildup, the voltage across the resistor gradually increases to the same voltage as the source voltage; and during current decay the voltage across the resistor gradually drops to zero*

A9.

$$t = \frac{L}{R}$$

A10.

a. 1.71 amperes.

b. 5 time constants.

c. 2 time constants.

A11. Copper loss; hysteresis loss; eddy-current loss.

A12. Mutual inductance is the property existing between two coils so positioned that flux from one coil cuts the windings of the other coil.

A13. When they are arranged so that energy from one circuit is transferred to the other circuit.

A14. The ratio of the fines of force produced by one coil to the lines of force that link another coil. It is never greater than one.

A15.

$$4.48\ \text{H}\ (\text{because } M = K\sqrt{7H \times 7H} = 0.64 \times 7H = 4.48H)$$

A16.

$$L_T = L_1 + L_2 - 2M$$

A17.

$$L_T = 18\ \text{H}\ (\text{because } L_T = L_1 + L_2 + 2M$$
$$L_T = 3H + 5H + 2(5H) = 18H)$$

CHAPTER 3

CAPACITANCE

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Define the terms "capacitor" and "capacitance."
2. State four characteristics of electrostatic lines of force.
3. State the effect that an electrostatic field has on a charged particle.
4. State the basic parts of a capacitor.
5. Define the term "farad".
6. State the mathematical relationship between a farad, a microfarad, and a picofarad.
7. State three factors that affect the value of capacitance.
8. Given the dielectric constant and the area of and the distance between the plates of a capacitor, solve for capacitance.
9. State two types of power losses associated with capacitors.
10. Define the term "working voltage" of a capacitor, and compute the working voltage of a capacitor.
11. State what happens to the electrons in a capacitor when the capacitor is charging and when it is discharging.
12. State the relationship between voltage and time in an RC circuit when the circuit is charging and discharging.
13. State the relationship between the voltage drop across a resistor and the source voltage in an RC circuit.
14. Given the component values of an RC circuit, compute the RC time constant.
15. Use the universal time constant chart to determine the value of an unknown capacitor in an RC circuit.
16. Calculate the value of total capacitance in a circuit containing capacitors of known value in series.
17. Calculate the value of total capacitance in a circuit containing capacitors of known value in parallel.
18. State the difference between different types of capacitors.
19. Determine the electrical values of capacitors using the color code.

CAPACITANCE

In the previous chapter you learned that inductance is the property of a coil that causes electrical energy to be stored in a magnetic field about the coil. The energy is stored in such a way as to <u>oppose any change in current</u>. CAPACITANCE is similar to inductance because it also causes a storage of energy. A CAPACITOR is a device that stores electrical energy in an <u>ELECTROSTATIC FIELD</u>. The energy is stored in such a way as to <u>oppose any change in voltage</u>. Just how capacitance opposes a change in voltage is explained later in this chapter. However, it is first necessary to explain the principles of an electrostatic field as it is applied to capacitance.

Q1. Define the terms "capacitor" and "capacitance."

THE ELECTROSTATIC FIELD

You previously learned that opposite electrical charges attract each other while like electrical charges repel each other. The reason for this is the existence of an electrostatic field. Any charged particle is surrounded by invisible lines of force, called electrostatic lines of force. These lines of force have some interesting characteristics:

- They are polarized from positive to negative.
- They radiate from a charged particle in straight lines and do not form closed loops.
- They have the ability to pass through any known material.
- They have the ability to distort the orbits of tightly bound electrons.

Examine figure 3-1. This figure represents two unlike charges surrounded by their electrostatic field. Because an electrostatic field is polarized positive to negative, arrows are shown radiating <u>away</u> from the positive charge and <u>toward</u> the negative charge. Stated another way, the field from the positive charge is pushing, while the field from the negative charge is pulling. The effect of the field is to push and pull the unlike charges together.

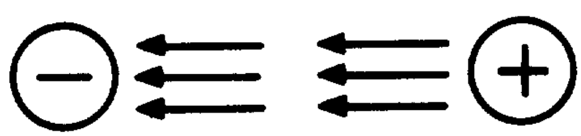


Figure 3-1.—Electrostatic field attracts two unlike charged particles.

In figure 3-2, two like charges are shown with their surrounding electrostatic field. The effect of the electrostatic field is to push the charges apart.

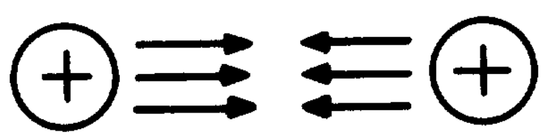


Figure 3-2.—Electrostatic field repels two like charged particles.

If two unlike charges are placed on opposite sides of an atom whose outermost electrons cannot escape their orbits, the orbits of the electrons are distorted as shown in figure 3-3. Figure 3-3(A) shows the normal orbit. Part (B) of the figure shows the same orbit in the presence of charged particles. Since the electron is a negative charge, the positive charge attracts the electrons, pulling the electrons closer to the positive charge. The negative charge repels the electrons, pushing them further from the negative charge. It is this ability of an electrostatic field to attract and to repel charges that allows the capacitor to store energy.

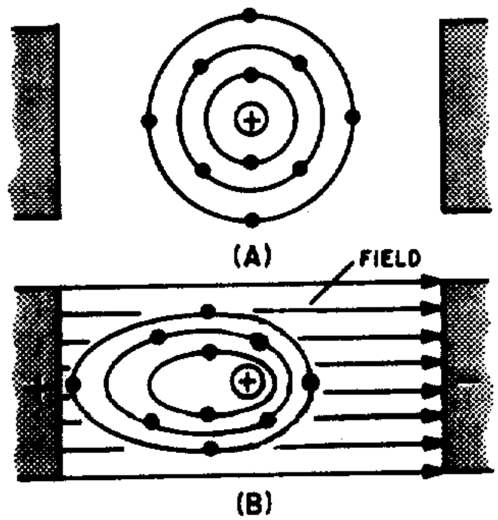


Figure 3-3.—Distortion of electron orbital paths due to electrostatic force.

Q2. State four characteristics of electrostatic lines of force.

Q3. An electron moves into the electrostatic field between a positive charge and a negative charge. Toward which charge will the electron move?

THE SIMPLE CAPACITOR

A simple capacitor consists of two metal plates separated by an insulating material called a dielectric, as illustrated in figure 3-4. Note that one plate is connected to the positive terminal of a battery; the other plate is connected through a closed switch (S1) to the negative terminal of the battery. Remember, an insulator is a material whose electrons cannot easily escape their orbits. Due to the battery voltage, plate A is charged positively and plate B is charged negatively. (How this happens is explained later in this chapter.) Thus an electrostatic field is set up between the positive and negative plates. The electrons on the negative plate (plate B) are attracted to the positive charges on the positive plate (plate A).

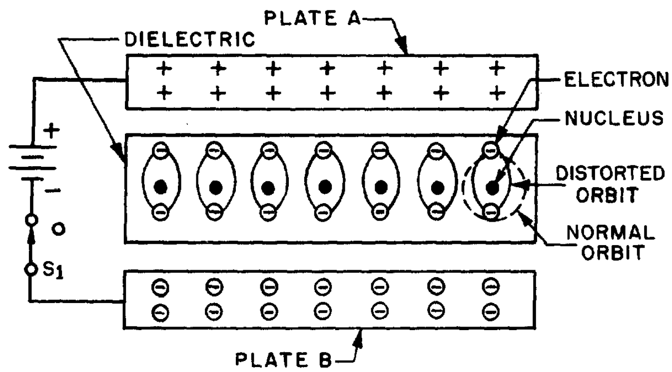


Figure 3-4.—Distortion of electron orbits in a dielectric.

Notice that the orbits of the electrons in the dielectric material are distorted by the electrostatic field. The distortion occurs because the electrons in the dielectric are attracted to the top plate while being repelled from the bottom plate. When switch S1 is opened, the battery is removed from the circuit and the charge is retained by the capacitor. This occurs because the dielectric material is an insulator, and the electrons in the bottom plate (negative charge) have no path to reach the top plate (positive charge). The distorted orbits of the atoms of the dielectric plus the electrostatic force of attraction between the two plates hold the positive and negative charges in their original position. Thus, the energy which came from the battery is now stored in the electrostatic field of the capacitor. Two slightly different symbols for representing a capacitor are shown in figure 3-5. Notice that each symbol is composed of two plates separated by a space that represents the dielectric. The curved plate in (B) of the figure indicates the plate should be connected to a negative polarity.

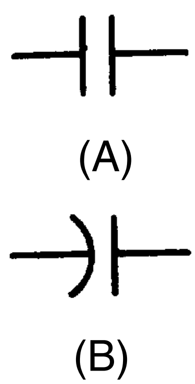


Figure 3-5.—Circuit symbols for capacitors.

Q4. What are the basic parts of a capacitor?

THE FARAD

Capacitance is measured in units called FARADS. A one-farad capacitor stores one coulomb (a unit of charge (Q) equal to 6.28×10^{18} electrons) of charge when a potential of 1 volt is applied across the terminals of the capacitor. This can be expressed by the formula:

$$C(\text{farads}) = \frac{Q\ (\text{coulombs})}{E\ (\text{volts})}$$

The farad is a very large unit of measurement of capacitance. For convenience, the microfarad (abbreviated µF) or the picofarad (abbreviated µF) is used. One (1.0) microfarad is equal to 0.000001 farad or 1×10^{-6} farad, and 1.0 picofarad is equal to 0.000000000001 farad or 1.0×10^{-12} farad. Capacitance is a physical property of the capacitor and does not depend on circuit characteristics of voltage, current, and resistance. A given capacitor always has the same value of capacitance (farads) in one circuit as in any other circuit in which it is connected.

Q5. Define the term "farad."

Q6. What is the mathematical relationship between a farad, a microfarad, and a picofarad.

FACTORS AFFECTING THE VALUE OF CAPACITANCE

The value of capacitance of a capacitor depends on three factors:

- The area of the plates.
- The distance between the plates.
- The dielectric constant of the material between the plates.

PLATE AREA affects the value of capacitance in the same manner that the size of a container affects the amount of water that can be held by the container. A capacitor with the large plate area can store more charges than a capacitor with a small plate area. Simply stated, "the larger the plate area, the larger the capacitance".

The second factor affecting capacitance is the DISTANCE BETWEEN THE PLATES. Electrostatic lines of force are strongest when the charged particles that create them are close together. When the charged particles are moved further apart, the lines of force weaken, and the ability to store a charge decreases.

The third factor affecting capacitance is the DIELECTRIC CONSTANT of the insulating material between the plates of a capacitor. The various insulating materials used as the dielectric in a capacitor differ in their ability to respond to (pass) electrostatic lines of force. A dielectric material, or insulator, is rated as to its ability to respond to electrostatic lines of force in terms of a figure called the DIELECTRIC CONSTANT. A dielectric material with a high dielectric constant is a better insulator than a dielectric material with a low dielectric constant. Dielectric constants for some common materials are given in the following list:

Material	Constant
Vacuum	1.0000
Air	1.0006
Paraffin paper	3.5
Glass	5 to 10
Mica	3 to 6
Rubber	2.5 to 35
Wood	2.5 to 8
Glycerine (15ºC)	56
Petroleum	2
Pure water	81

Notice the dielectric constant for a vacuum. Since a vacuum is the standard of reference, it is assigned a constant of one. The dielectric constants of all materials are compared to that of a vacuum. Since the dielectric constant of air has been determined to be approximately the same as that of a vacuum, the dielectric constant of AIR is also considered to be equal to one.

The formula used to compute the value of capacitance is:

$$C = 0.2249\left(\frac{KA}{d}\right)$$

Where C = capacitance in picofarads

A = area of one plate, in square inches

d = distance between the plates, in inches

K = dielectric constant of the insulating material

0.2249 = a constant resulting from conversion from Metric to English units.

Example: Find the capacitance of a parallel plate capacitor with paraffin paper as the dielectric.

Given: $K = 3.5$

$d = 0.05$ inch

$A = 12$ square inches

Solution: $C = 0.2249\left(\frac{KA}{d}\right)$

$$C = 0.2249\left(\frac{3.5 \times 12}{0.05}\right)$$

$C = 189$ picofarads

By examining the above formula you can see that capacitance varies directly as the dielectric constant and the area of the capacitor plates, and inversely as the distance between the plates.

Q7. State three factors that affect the capacitance of a capacitor.

Q8. A parallel plate capacitor has the following values: K = 81, d = .025 inches, A = 6 square inches. What is the capacitance of the capacitor?

VOLTAGE RATING OF CAPACITORS

In selecting or substituting a capacitor for use, consideration must be given to (1) the value of capacitance desired and (2) the amount of voltage to be applied across the capacitor. If the voltage applied across the capacitor is too great, the dielectric will break down and arcing will occur between the capacitor plates. When this happens the capacitor becomes a short-circuit and the flow of direct current through it can cause damage to other electronic parts. Each capacitor has a voltage rating (a working voltage) that should not be exceeded.

The working voltage of the capacitor is the maximum voltage that can be steadily applied without danger of breaking down the dielectric. The working voltage depends on the type of material used as the dielectric and on the thickness of the dialectic. (A high-voltage capacitor that has a thick dielectric must have a relatively large plate area in order to have the same capacitance as a similar low-voltage capacitor having a thin dielectric.) The working voltage also depends on the applied frequency because the losses, and the resultant heating effect, <u>increase as the frequency increases</u>.

A capacitor with a voltage rating of 500 volts dc cannot be safely subjected to an alternating voltage or a pulsating direct voltage having an effective value of 500 volts. Since an alternating voltage of 500 volts (rms) has a peak value of 707 volts, a capacitor to which it is applied should have a working voltage of at least 750 volts. In practice, a capacitor should be selected so that its working voltage is at least 50 percent greater than the highest effective voltage to be applied to it.

CAPACITOR LOSSES

Power loss in a capacitor may be attributed to dielectric hysteresis and dielectric leakage. Dielectric hysteresis may be defined as an effect in a dielectric material similar to the hysteresis found in a magnetic material. It is the result of changes in orientation of electron orbits in the dielectric because of the rapid reversals of the polarity of the line voltage. The amount of power loss due to dielectric hysteresis depends upon the type of dielectric used. A vacuum dielectric has the smallest power loss.

Dielectric leakage occurs in a capacitor as the result of LEAKAGE CURRENT through the dielectric. Normally it is assumed that the dielectric will effectively prevent the flow of current through the capacitor. Although the resistance of the dielectric is extremely high, a minute amount of current does flow. Ordinarily this current is so small that for all practical purposes it is ignored. However, if the leakage through the dielectric is abnormally high, there will be a rapid loss of charge and an overheating of the capacitor.

The power loss of a capacitor is determined by loss in the dielectric. If the loss is negligible and the capacitor returns the total charge to the circuit, it is considered to be a perfect capacitor with a power loss of zero.

Q9. Name two types of power losses associated with a capacitor.

Q10.

a. Define the term "working voltage" of a capacitor.

b. What should be the working voltage of a capacitor in a circuit that is operating at 600 volts?

CHARGING AND DISCHARGING A CAPACITOR

CHARGING

In order to better understand the action of a capacitor in conjunction with other components, the charge and discharge actions of a purely capacitive circuit are analyzed first. For ease of explanation the capacitor and voltage source shown in figure 3-6 are assumed to be perfect (no internal resistance), although this is impossible in practice.

In figure 3-6(A), an uncharged capacitor is shown connected to a four-position switch. With the switch in position 1 the circuit is open and no voltage is applied to the capacitor. Initially each plate of the capacitor is a neutral body and until a difference of potential is impressed across the capacitor, no electrostatic field can exist between the plates.

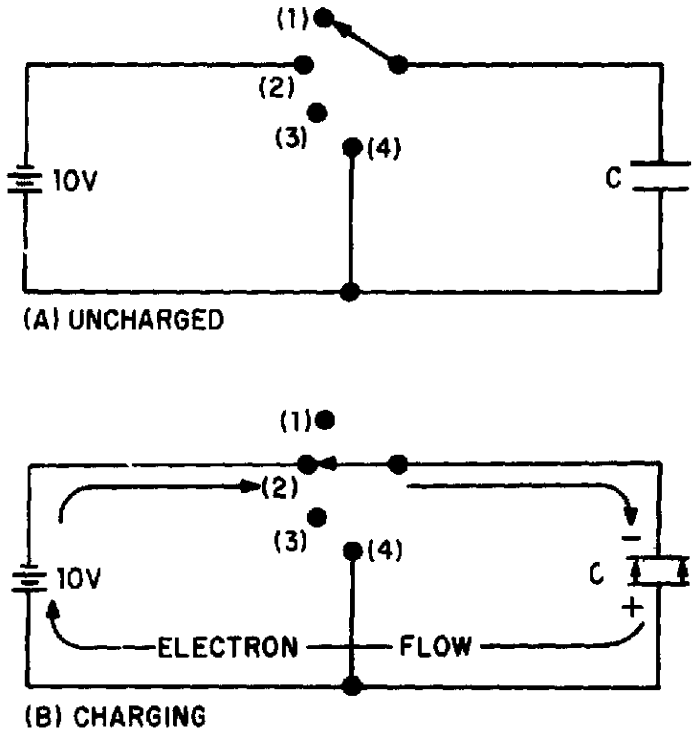


Figure 3-6.—Charging a capacitor.

To CHARGE the capacitor, the switch must be thrown to position 2, which places the capacitor across the terminals of the battery. Under the assumed perfect conditions, the capacitor would reach full charge instantaneously. However, the charging action is spread out over a period of time in the following discussion so that a step-by-step analysis can be made.

At the instant the switch is thrown to position 2 (fig. 3-6(B)), a displacement of electrons occurs simultaneously in all parts of the circuit. This electron displacement is directed away from the negative terminal and toward the positive terminal of the source (the battery). A brief surge of current will flow as the capacitor charges.

If it were possible to analyze the motion of the individual electrons in this surge of charging current, the following action would be observed. See figure 3-7.

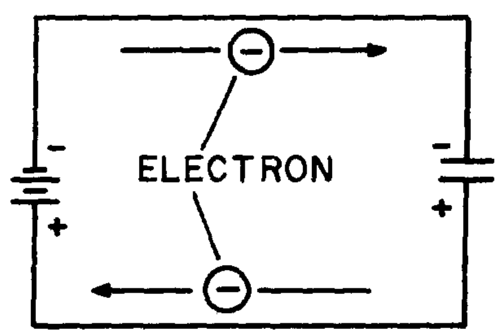


Figure 3-7.—Electron motion during charge.

At the instant the switch is closed, the positive terminal of the battery extracts an electron from the bottom conductor. The negative terminal of the battery forces an electron into the top conductor. At this same instant an electron is forced into the top plate of the capacitor and another is pulled from the bottom plate. Thus, in every part of the circuit a clockwise DISPLACEMENT of electrons occurs simultaneously.

As electrons accumulate on the top plate of the capacitor and others depart from the bottom plate, a difference of potential develops across the capacitor. Each electron forced onto the top plate makes that plate more negative, while each electron removed from the bottom causes the bottom plate to become more positive. Notice that the polarity of the voltage which builds up across the capacitor is such as to oppose the source voltage. The source voltage (emf) forces current around the circuit of figure 3-7 in a clockwise direction. The emf developed across the capacitor, however, has a tendency to force the current in a counterclockwise direction, opposing the source emf. As the capacitor continues to charge, the voltage across the capacitor rises until it is equal to the source voltage. Once the capacitor voltage equals the source voltage, the two voltages balance one another and current ceases to flow in the circuit.

In studying the charging process of a capacitor, you must be aware that NO current flows THROUGH the capacitor. The material between the plates of the capacitor must be an insulator. However, to an observer stationed at the source or along one of the circuit conductors, the action has all the appearances of a true flow of current, even though the insulating material between the plates of the capacitor prevents the current from having a complete path. The current which appears to flow through a capacitor is called DISPLACEMENT CURRENT.

When a capacitor is fully charged and the source voltage is equaled by the counter electromotive force (cemf) across the capacitor, the electrostatic field between the plates of the capacitor is maximum. (Look again at figure 3-4.) Since the electrostatic field is maximum the energy stored in the dielectric is also maximum.

If the switch is now opened as shown in figure 3-8(A), the electrons on the upper plate are isolated. The electrons on the top plate are attracted to the charged bottom plate. Because the dielectric is an insulator, the electrons can not cross the dielectric to the bottom plate. The charges on both plates will be effectively trapped by the electrostatic field and the capacitor will remain charged indefinitely. You should note at this point that the insulating dielectric material in a practical capacitor is not perfect and small leakage current will flow through the dielectric. This current will eventually dissipate the charge. However, a high quality capacitor may hold its charge for a month or more.

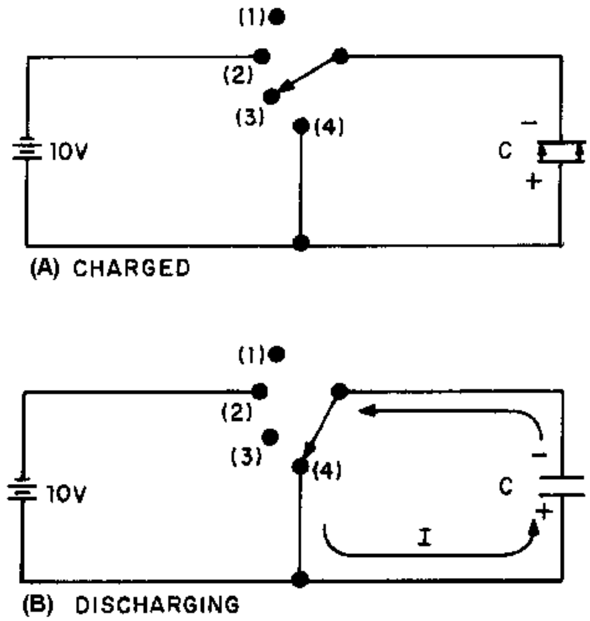


Figure 3-8.—Discharging a capacitor.

To review briefly, when a capacitor is connected across a voltage source, a surge of charging current flows. This charging current develops a cemf across the capacitor which opposes the applied voltage. When the capacitor is fully charged, the cemf is equal to the applied voltage and charging current ceases. At full charge, the electrostatic field between the plates is at maximum intensity and the energy stored in the dielectric is maximum. If the charged capacitor is disconnected from the source, the charge will be retained for some period of time. The length of time the charge is retained depends on the amount of leakage current present. Since electrical energy is stored in the capacitor, a charged capacitor can act as a source emf.

DISCHARGING

To DISCHARGE a capacitor, the charges on the two plates must be neutralized. This is accomplished by providing a conducting path between the two plates as shown in figure 3-8(B). With the switch in position (4) the excess electrons on the negative plate can flow to the positive plate and neutralize its charge. When the capacitor is discharged, the distorted orbits of the electrons in the dielectric return to their normal positions and the stored energy is returned to the circuit. It is important for you to note that a capacitor does not consume power. The energy the capacitor draws from the source is recovered when the capacitor is discharged.

Q11. State what happens to the electrons in a capacitor circuit when (a) the capacitor is charging and (b) the capacitor is discharging.

CHARGE AND DISCHARGE OF AN RC SERIES CIRCUIT

Ohm's law states that the voltage across a resistance is equal to the current through the resistance times the value of the resistance. This means that a voltage is developed across a resistance ONLY WHEN CURRENT FLOWS through the resistance.

A capacitor is capable of storing or holding a charge of electrons. When uncharged, both plates of the capacitor contain essentially the same number of free electrons. When charged, one plate contains more free electrons than the other plate. The difference in the number of electrons is a measure of the charge on the capacitor. The accumulation of this charge builds up a voltage across the terminals of the capacitor, and the charge continues to increase until this voltage equals the applied voltage. The charge in a capacitor is related to the capacitance and voltage as follows:

$$Q = CE,$$

in which Q is the charge in coulombs, C the capacitance in farads, and E the emf across the capacitor in volts.

CHARGE CYCLE

A voltage divider containing resistance and capacitance is connected in a circuit by means of a switch, as shown at the top of figure 3-9. Such a series arrangement is called an RC series circuit.

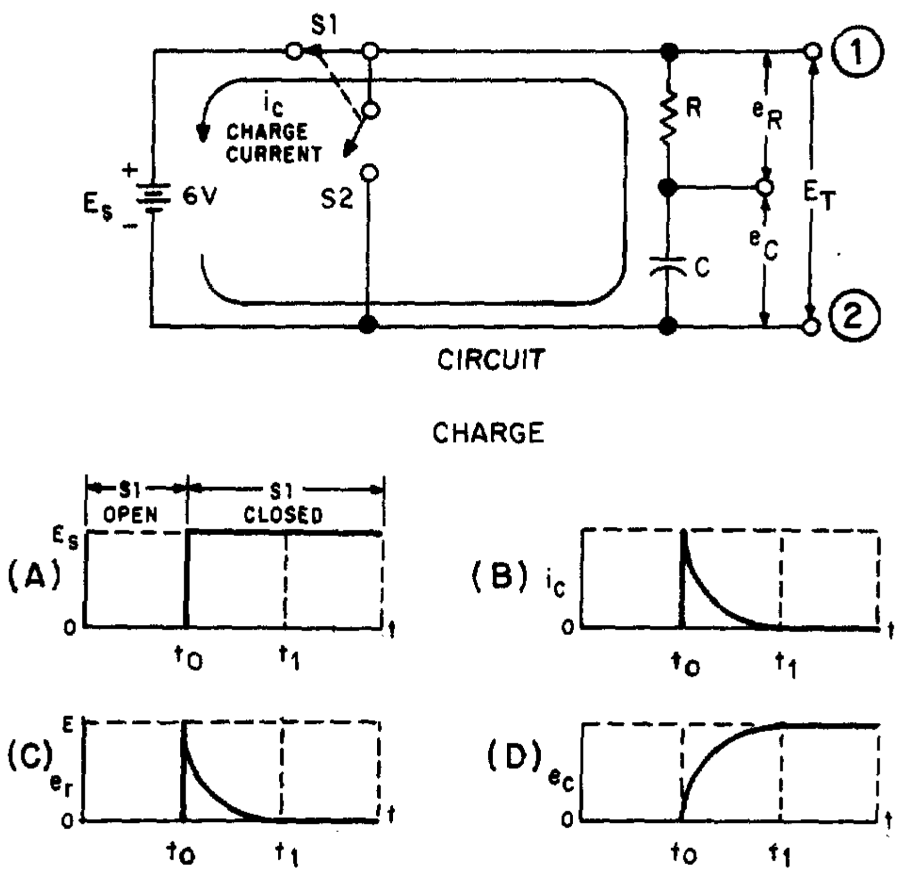


Figure 3-9.—Charge of an RC series circuit.

In explaining the charge and discharge cycles of an RC series circuit, the time interval from time t_0 (time zero, when the switch is first closed) to time t_1 (time one, when the capacitor reaches full charge or discharge potential) will be used. (Note that switches S1 and S2 move at the same time and can never both be closed at the same time.)

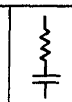
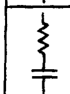
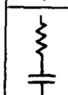
When switch S1 of the circuit in figure 3-9 is closed at t_0, the source voltage (E_S) is instantly felt across the entire circuit. Graph (A) of the figure shows an instantaneous rise at time t_0 from zero to source voltage (E_S = 6 volts). The total voltage can be measured across the circuit between points 1 and 2. Now look at graph (B) which represents the charging current in the capacitor (i_c). At time t_0, charging current is MAXIMUM. As time elapses toward time t_1, there is a continuous decrease in current flowing into the capacitor. The decreasing flow is caused by the voltage buildup across the capacitor. At time t_1, current flowing in the capacitor stops. At this time, the capacitor has reached full charge and has stored maximum energy in its electrostatic field. Graph (C) represents the voltage drop (e) across the resistor (R). The value of e_r is determined by the amount of current flowing through the resistor on its way to the capacitor. At time t_0 the current flowing to the capacitor is maximum. Thus, the voltage drop across the resistor is maximum (E = IR). As time progresses toward time t_1, the current flowing to the capacitor steadily decreases and causes the voltage developed across the resistor (R) to steadily decrease. When time t_1 is reached, current flowing to the capacitor is stopped and the voltage developed across the resistor has decreased to zero.

You should remember that capacitance opposes a change in voltage. This is shown by comparing graph (A) to graph (D). In graph (A) the voltage changed instantly from 0 volts to 6 volts across the circuit, while the voltage developed across the capacitor in graph (D) took the entire time interval from time t_o to time t_1 to reach 6 volts. The reason for this is that in the first instant at time t_0, maximum current flows through R and the entire circuit voltage is dropped across the resistor. The voltage impressed across the capacitor at t_0 is zero volts. As time progresses toward t_1, the decreasing current causes progressively less voltage to be dropped across the resistor (R), and more voltage builds up across the capacitor (C). At time t_1, the voltage felt across the capacitor is equal to the source voltage (6 volts), and the voltage dropped across the resistor (R) is equal to zero. This is the complete charge cycle of the capacitor.

As you may have noticed, the processes which take place in the time interval t_0 to t_1 in a series RC circuit are exactly opposite to those in a series LR circuit.

For your comparison, the important points of the charge cycle of RC and LR circuits are summarized in table 3-1.

Table 3-1.—Summary of Capacitive and Inductive Characteristics.

		TIME ZERO (t_0)	TIME BETWEEN t_0 AND t_1	TIME ONE (t_1)
CIRCUIT CURRENT		MAXIMUM	DECREASING	ZERO
		ZERO	INCREASING	MAXIMUM
VOLTAGE DEVELOPED ACROSS THE RESISTOR		MAXIMUM	DECREASING	ZERO
		ZERO	INCREASING	MAXIMUM
VOLTAGE DEVELOPED ACROSS CAPACITOR/ INDUCTOR		ZERO	INCREASING	MAXIMUM
		MAXIMUM	DECREASING	ZERO

DISCHARGE CYCLE

In figure 3-10 at time t_0, the capacitor is fully charged. When S1 is open and S2 closes, the capacitor discharge cycle starts. At the first instant, circuit voltage attempts to go from source potential (6 volts) to zero volts, as shown in graph (A). Remember, though, the capacitor during the charge cycle has stored energy in an electrostatic field.

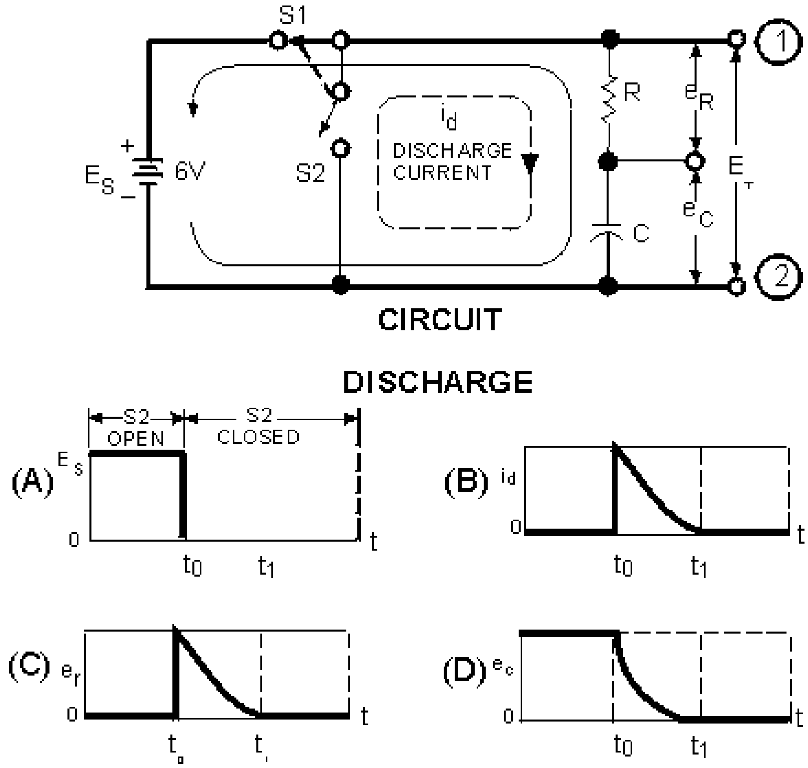


Figure 3-10.—Discharge of an RC Series circuit.

Because S2 is closed at the same time S1 is open, the stored energy of the capacitor now has a path for current to flow. At t_0, discharge current (i_d) from the bottom plate of the capacitor through the resistor (R) to the top plate of the capacitor (C) is maximum. As time progresses toward t_1, the discharge current steadily decreases until at time t_1 it reaches zero, as shown in graph (B).

The discharge causes a corresponding voltage drop across the resistor as shown in graph (C). At time t_0, the current through the resistor is maximum and the voltage drop (e_r) across the resistor is maximum. As the current through the resistor decreases, the voltage drop across the resistor decreases until at t_1 it has reached a value of zero. Graph (D) shows the voltage across the capacitor (e_c) during the discharge cycle. At time t_0 the voltage is maximum and as time progresses toward time t_1, the energy stored in the capacitor is depleted. At the same time the voltage across the resistor is decreasing, the voltage (e) across the capacitor is decreasing until at time t_1 the voltage (e_c) reaches zero.

By comparing graph (A) with graph (D) of figure 3-10, you can see the effect that capacitance has on a change in voltage. If the circuit had not contained a capacitor, the voltage would have ceased at the instant S1 was opened at time t_0. Because the capacitor is in the circuit, voltage is applied to the circuit until the capacitor has discharged completely at t_1. The effect of capacitance has been to oppose this change in voltage.

Q12. At what instant does the greatest voltage appear across the resistor in a series RC circuit when the capacitor is charging?

Q13. What is the voltage drop across the resistor in an RC charging circuit when the charge on the capacitor is equal to the battery voltage?

RC TIME CONSTANT

The time required to charge a capacitor to 63 percent (actually 63.2 percent) of full charge or to discharge it to 37 percent (actually 36.8 percent) of its initial voltage is known as the TIME CONSTANT (TC) of the circuit. The charge and discharge curves of a capacitor are shown in figure 3-11. Note that the charge curve is like the curve in figure 3-9, graph (D), and the discharge curve like the curve in figure 3-9, graph (B).

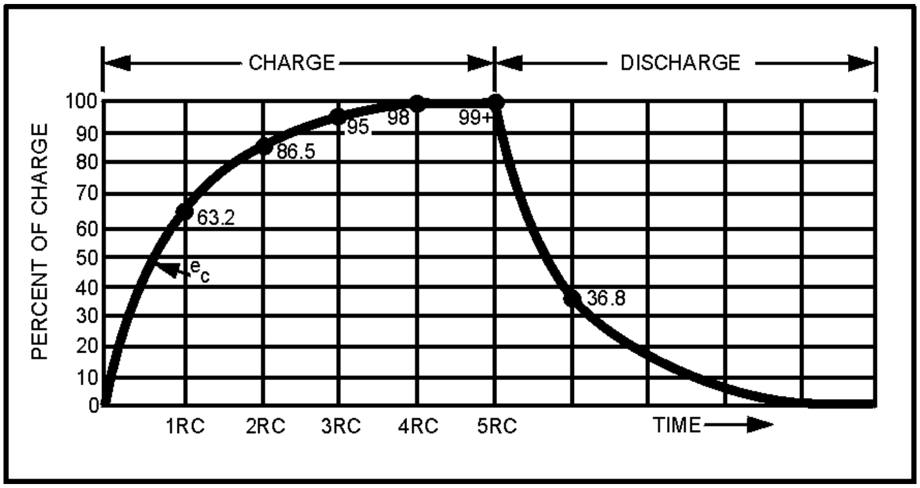


Figure 3-11.—RC time constant.

The value of the time constant in seconds is equal to the product of the circuit resistance in ohms and the circuit capacitance in farads. The value of one time constant is expressed mathematically as t = RC. Some forms of this formula used in calculating RC time constants are:

$$t \text{ (in seconds)} = R \text{ (in ohms)} \times C \text{ (in farads)}$$

$$t \text{ (in seconds)} = R \text{ (in megohms)} \times C \text{ (in microfarads)}$$

$$t \text{ (in microseconds)} = R \text{ (in ohms)} \times C \text{ (in microfarads)}$$

$$t \text{ (in microseconds)} = R \text{ (in megohms)} \times C \text{ (in picofarads)}$$

Q14. What is the RC time constant of a series RC circuit that contains a 12-megohm resistor and a 12-microfarad capacitor?

UNIVERSAL TIME CONSTANT CHART

Because the impressed voltage and the values of R and C or R and L in a circuit are usually known, a UNIVERSAL TIME CONSTANT CHART (fig. 3-12) can be used to find the time constant of the circuit. Curve A is a plot of both capacitor voltage during charge and inductor current during growth. Curve B is a plot of both capacitor voltage during discharge and inductor current during decay.

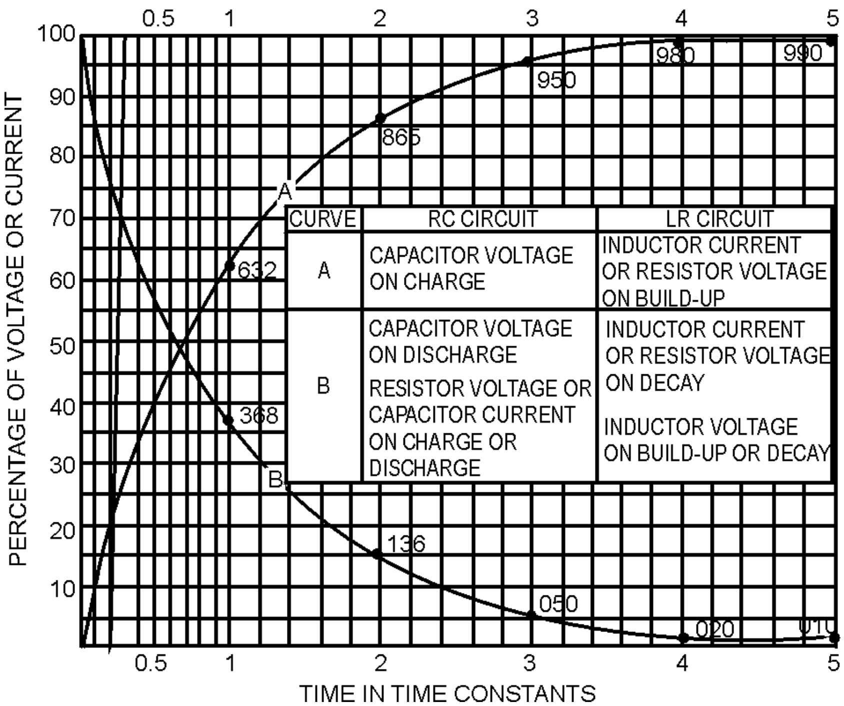


Figure 3-12.—Universal time constant chart for RC and RL circuit.

The time scale (horizontal scale) is graduated in terms of the RC or L/R time constants so that the curves may be used for any value of R and C or L and R. The voltage and current scales (vertical scales) are graduated in terms of percentage of the maximum voltage or current so that the curves may be used for any value of voltage or current. If the time constant and the initial or final voltage for the circuit in question are known, the voltages across the various parts of the circuit can be obtained from the curves for any time after the switch is closed, either on charge or discharge. The same reasoning is true of the current in the circuit.

The following problem illustrates how the universal time constant chart may be used.

An RC circuit is to be designed in which a capacitor (C) must charge to 20 percent (0.20) of the maximum charging voltage in 100 microseconds (0.0001 second). Because of other considerations, the resistor (R) must have a value of 20,000 ohms. What value of capacitance is needed?

$$\text{Given:} \quad \text{Percent of charge} = 20\% \ (.20)$$
$$t = 100\,\mu s$$
$$R = 20{,}000\,\Omega$$

Find: The capacitance of capacitor C.

Solution: Because the only values given are in units of time and resistance, a variation of the formula to find RC time is used:

$$RC = R \times C$$

where: 1 RC time constant $= R \times C$
and R is known.

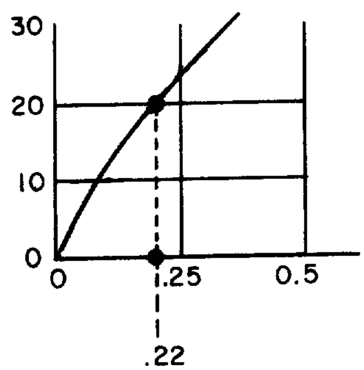
Transpose the formula to:

$$C = \frac{RC}{R}$$

Find the value of RC by referring to the universal time constant chart in figure 3-12 and proceed as follows:

- Locate the 20 point on the vertical scale at the left side of the chart (percentage).
- Follow the horizontal line from this point to intersect curve A.
- Follow an imaginary vertical line from the point of intersection on curve A downward to cross the RC scale at the bottom of the chart.

Note that the vertical line crosses the horizontal scale at about .22 RC as illustrated below:



The value selected from the graph means that a capacitor (including the one you are solving for) will reach twenty percent of full charge in twenty-two one hundredths (.22) of one RC time constant. Remember that it takes 100 µs for the capacitor to reach 20% of full charge. Since 100 µs is equal to .22 RC (twenty-two one-hundredths), then the time required to reach one RC time constant must be equal to:

$$.22\,RC = 100\,\mu s$$

$$RC = \frac{1}{.22} \times 100\,\mu s$$

$$RC = \frac{100\,\mu s}{.22}$$

$$RC = 454.54\,\mu s \text{ (rounded off to } 455\,\mu s)$$

$$RC = 455\,\mu s$$

Now use the following formula to find C:

$$C = \frac{RC}{R}$$

$$C = \frac{455\ \mu s}{20{,}000\ \text{ohms}}$$

$$C = 0.0227\ \mu F$$

$$C = .023\ \mu F$$

To summarize the above procedures, the problem and solution are shown below without the step by step explanation.

$$\begin{aligned} \text{Given:}\quad \text{Percent of charge} &= 20\%\ (.20) \\ t &= 100\ \mu s \\ R &= 20{,}000\ \text{ohms} \end{aligned}$$

Transpose the RC time constant formula as follows:

$$R \times C = RC$$

$$C = \frac{RC}{R}$$

Find: RC

$$.22\ RC = 100\ \mu s$$

$$RC = \frac{100\ \mu s}{.22}$$

$$RC = 455\ \mu s$$

Substitute the R and RC values into the formula:

$$C = \frac{RC}{R}$$

$$C = \frac{455\ \mu s}{20{,}000}\ \text{ohms}$$

$$C = .023\ \mu s$$

The graphs shown in figure 3-11 and 3-12 are not entirely complete. That is, the charge or discharge (or the growth or decay) is not quite complete in 5 RC or 5 L/R time constants. However, when the values reach 0.99 of the maximum (corresponding to 5 RC or 5 L/R), the graphs may be considered accurate enough for all practical purposes.

Q15. A circuit is to be designed in which a capacitor must charge to 40 percent of the maximum charging voltage in 200 microseconds. The resistor to be used has a resistance of 40,000 ohms. What size capacitor must be used? (Use the universal time constant chart in figure 3-12.)

CAPACITORS IN SERIES AND PARALLEL

Capacitors may be connected in series or in parallel to obtain a resultant value which may be either the sum of the individual values (in parallel) or a value less than that of the smallest capacitance (in series).

CAPACITORS IN SERIES

The overall effect of connecting capacitors in series is to move the plates of the capacitors further apart. This is shown in figure 3-13. Notice that the junction between C1 and C2 has both a negative and a positive charge. This causes the junction to be essentially neutral. The total capacitance of the circuit is developed between the left plate of C1 and the right plate of C2. Because these plates are farther apart, the total value of the capacitance in the circuit is decreased. Solving for the total capacitance (C_T) of capacitors connected in series is similar to solving for the total resistance (R_T) of resistors connected in parallel.

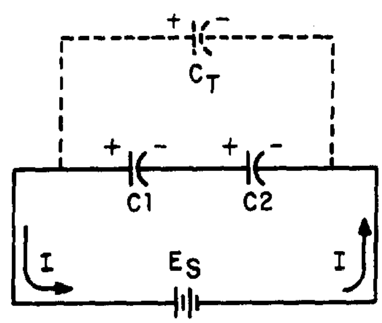


Figure 3-13.—Capacitors in series.

Note the similarity between the formulas for R_T and C_T:

$$R_T = \frac{1}{\frac{1}{R1} + \frac{1}{R2} + \ldots \frac{1}{R_n}}$$

$$C_T = \frac{1}{\frac{1}{C1} + \frac{1}{C2} + \ldots \frac{1}{C_n}}$$

If the circuit contains more than two capacitors, use the above formula. If the circuit contains only two capacitors, use the below formula:

$$C_T = \frac{C1 \times C2}{C1 + C2}$$

Note: All values for C_T, C1, C2, C3,... C_n should be in farads. It should be evident from the above formulas that the total capacitance of capacitors in series is less than the capacitance of any of the individual capacitors.

Example: Determine the total capacitance of a series circuit containing three capacitors whose values are 0.01 µF, 0.25 µF, and 50,000 pF, respectively.

Given: $C1 = 0.01\,\mu s$

$C2 = 0.25\,\mu s$

$C3 = 50{,}000\,pF$

Solution:

$$C_T = \frac{1}{\frac{1}{C1}+\frac{1}{C2}+\frac{1}{C3}}$$

$$C_T = \frac{1}{\frac{1}{.01\,\mu F}+\frac{1}{.25\,\mu F}+\frac{1}{50{,}000\,pF}}$$

$$C_T = \frac{1}{\frac{1}{1\times 10^{-8}}+\frac{1}{25\times 10^{-8}}+\frac{1}{5\times 10^{-8}}}F$$

$$C_T = \frac{1}{100\times 10^{6}+4\times 10^{6}+20\times 10^{6}}F$$

$$C_T = \frac{1}{124\times 10^{6}}F$$

$$C_T = 0.008\,\mu F$$

The total capacitance of 0.008µF is slightly smaller than the smallest capacitor (0.01µF).

CAPACITORS IN PARALLEL

When capacitors are connected in parallel, one plate of each capacitor is connected directly to one terminal of the source, while the other plate of each capacitor is connected to the other terminal of the source. Figure 3-14 shows all the negative plates of the capacitors connected together, and all the positive plates connected together. C_T, therefore, appears as a capacitor with a plate area equal to the sum of all the individual plate areas. As previously mentioned, capacitance is a direct function of plate area. Connecting capacitors in parallel effectively increases plate area and thereby increases total capacitance.

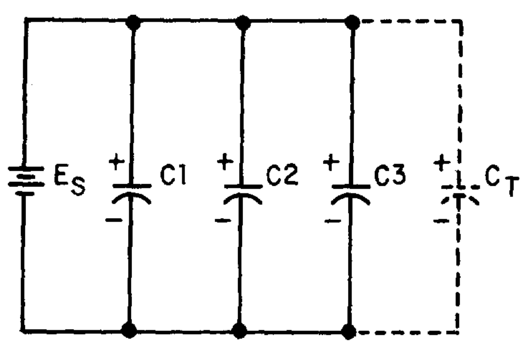


Figure 3-14.—Parallel capacitive circuit.

For capacitors connected in parallel the total capacitance is the sum of all the individual capacitances. The total capacitance of the circuit may by calculated using the formula:

$$C_T = C1 + C2 + C3 + ...C_n$$

where all capacitances are in the same units.

Example: Determine the total capacitance in a parallel capacitive circuit containing three capacitors whose values are 0.03 μF, 2.0 μF, and 0.25 μF, respectively.

$$\begin{aligned} \text{Given:} \quad & C1 = 0.03\,\mu F \\ & C2 = 2\,\mu F \\ & C3 = 0.25\,\mu F \\ \text{Solution:} \quad & C_T = C1 + C2 + C3 \\ & C_T = 0.03\,\mu F + 2.0\,\mu F + 0.25\,\mu F \\ & C_T = 2.28\,\mu F \end{aligned}$$

Q16. What is the total capacitance of a circuit that contains two capacitors (10 μF and 0.1 μF) wired together in series?

Q17. What is the total capacitance of a circuit in which four capacitors (10 μF, 21 μF, 0.1 μF and 2 μF) are connected in parallel?

FIXED CAPACITOR

A fixed capacitor is constructed in such manner that it possesses a fixed value of capacitance which cannot be adjusted. A fixed capacitor is classified according to the type of material used as its dielectric, such as paper, oil, mica, or electrolyte.

A PAPER CAPACITOR is made of flat thin strips of metal foil conductors that are separated by waxed paper (the dielectric material). Paper capacitors usually range in value from about 300 picofarads to about 4 microfarads. The working voltage of a paper capacitor rarely exceeds 600 volts. Paper capacitors are sealed with wax to prevent the harmful effects of moisture and to prevent corrosion and leakage.

Many different kinds of outer covering are used on paper capacitors, the simplest being a tubular cardboard covering. Some types of paper capacitors are encased in very hard plastic. These types are very rugged and can be used over a much wider temperature range than can the tubular cardboard type. Figure 3-15(A) shows the construction of a tubular paper capacitor; part 3-15(B) shows a completed cardboard-encased capacitor.

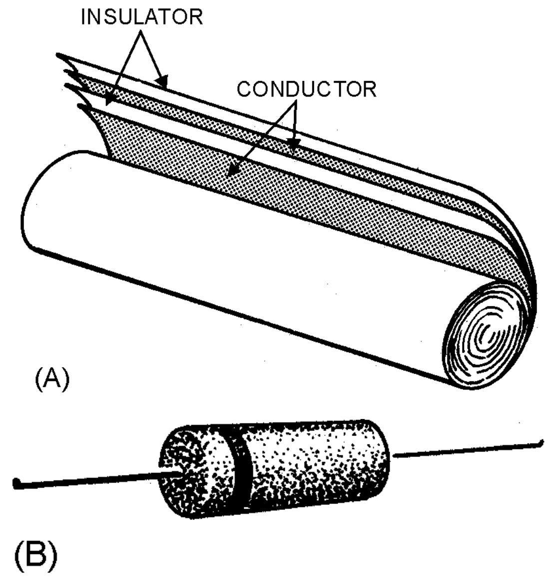


Figure 3-15.—Paper capacitor.

A MICA CAPACITOR is made of metal foil plates that are separated by sheets of mica (the dielectric). The whole assembly is encased in molded plastic. Figure 3-16(A) shows a cut-away view of a mica capacitor. Because the capacitor parts are molded into a plastic case, corrosion and damage to the plates and dielectric are prevented. In addition, the molded plastic case makes the capacitor mechanically stronger. Various types of terminals are used on mica capacitors to connect them into circuits. These terminals are also molded into the plastic case.

Mica is an excellent dielectric and can withstand a higher voltage than can a paper dielectric of the same thickness. Common values of mica capacitors range from approximately 50 picofarads to 0.02 microfarad. Some different shapes of mica capacitors are shown in figure 3-16(B).

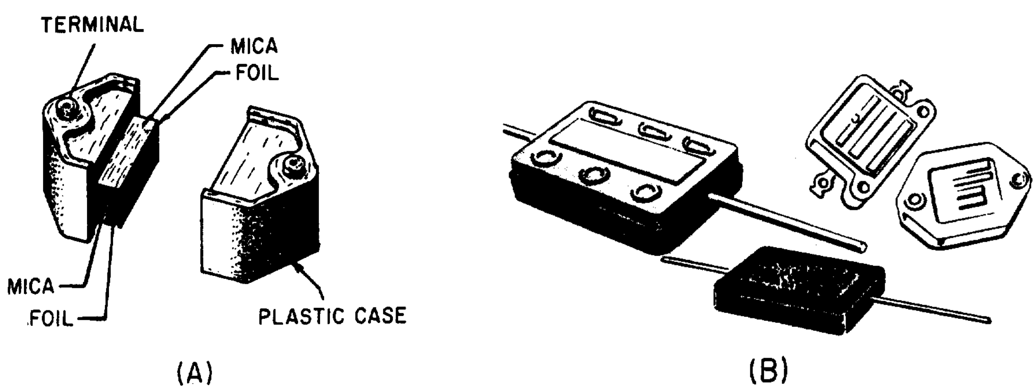


Figure 3-16.—Typical mica capacitors.

A CERAMIC CAPACITOR is so named because it contains a ceramic dielectric. One type of ceramic capacitor uses a hollow ceramic cylinder as both the form on which to construct the capacitor and as the dielectric material. The plates consist of thin films of metal deposited on the ceramic cylinder.

A second type of ceramic capacitor is manufactured in the shape of a disk. After leads are attached to each side of the capacitor, the capacitor is completely covered with an insulating moisture-proof coating. Ceramic capacitors usually range in value from 1 picofarad to 0.01 microfarad and may be used with voltages as high as 30,000 volts. Some different shapes of ceramic capacitors are shown in figure 3-17.

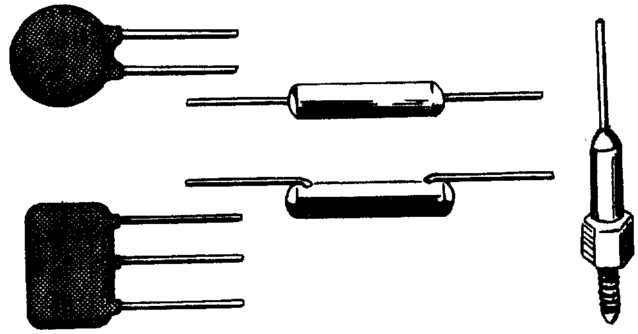
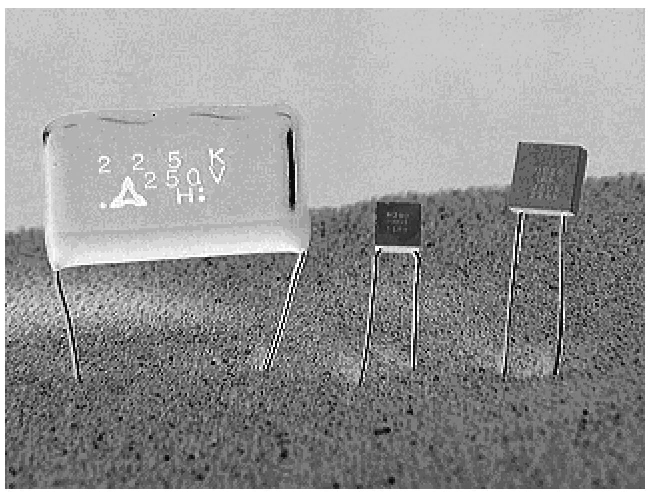


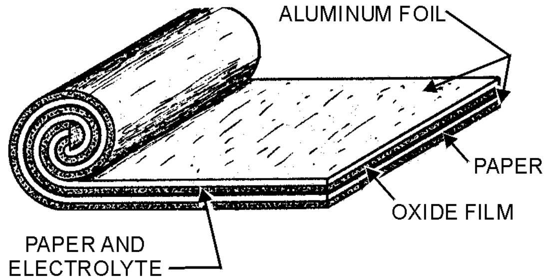
Figure 3-17.—Ceramic capacitors.



Examples of ceramic capacitors.

An ELECTROLYTIC CAPACITOR is used where a large amount of capacitance is required. As the name implies, an electrolytic capacitor contains an electrolyte. This electrolyte can be in the form of a liquid (wet electrolytic capacitor). The wet electrolytic capacitor is no longer in popular use due to the care needed to prevent spilling of the electrolyte.

A dry electrolytic capacitor consists essentially of two metal plates separated by the electrolyte. In most cases the capacitor is housed in a cylindrical aluminum container which acts as the negative terminal of the capacitor (see fig. 3-18). The positive terminal (or terminals if the capacitor is of the multisection type) is a lug (or lugs) on the bottom end of the container. The capacitance value(s) and the voltage rating of the capacitor are generally printed on the side of the aluminum case.



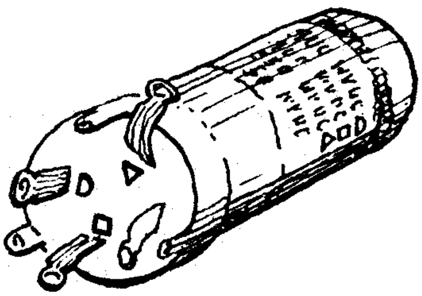


Figure 3-18.—Construction of an electrolytic capacitor.

An example of a multisection electrolytic capacitor is illustrated in figure 3-18(B). The four lugs at the end of the cylindrical aluminum container indicates that four electrolytic capacitors are enclosed in the can. Each section of the capacitor is electrically independent of the other sections. It is possible for one section to be defective while the other sections are still good. The can is the common negative connection to the four capacitors. Separate terminals are provided for the positive plates of the capacitors. Each capacitor is identified by an embossed mark adjacent to the lugs, as shown in figure 3-18(B). Note the identifying marks used on the electrolytic capacitor are the half moon, the triangle, the square, and no embossed mark. By looking at the bottom of the container and the identifying sheet pasted to the side of the container, you can easily identify the value of each section.

Internally, the electrolytic capacitor is constructed similarly to the paper capacitor. The positive plate consists of aluminum foil covered with an extremely thin film of oxide. This thin oxide film (which is formed by an electrochemical process) acts as the dielectric of the capacitor. Next to and in contact with the oxide is a strip of paper or gauze which has been impregnated with a paste-like electrolyte. The electrolyte acts as the negative plate of the capacitor. A second strip of aluminum foil is then placed against the electrolyte to provide electrical contact to the negative electrode (the electrolyte). When the three layers are in place they are rolled up into a cylinder as shown in figure 3-18(A).

An electrolytic capacitor has two primary disadvantages compared to a paper capacitor in that the electrolytic type is POLARIZED and has a LOW LEAKAGE RESISTANCE. This means that should the positive plate be accidentally connected to the negative terminal of the source, the thin oxide film dielectric will dissolve and the capacitor will become a conductor (i.e., it will short). The polarity of the terminals is normally marked on the case of the capacitor. Since an electrolytic capacitor is polarity sensitive, its use is ordinarily restricted to a dc circuit or to a circuit where a small ac voltage is superimposed on a dc voltage. Special electrolytic capacitors are available for certain ac applications, such as a motor starting capacitor. Dry electrolytic capacitors vary in size from about 4 microfarads to several thousand microfarads and have a working voltage of approximately 500 volts.

The type of dielectric used and its thickness govern the amount of voltage that can safely be applied to the electrolytic capacitor. If the voltage applied to the capacitor is high enough to cause the atoms of the dielectric material to become ionized, arcing between the plates will occur. In most other types of capacitors, arcing will destroy the capacitor. However, an electrolytic capacitor has the ability to be self-healing. If the arcing is small, the electrolytic will regenerate itself. If the arcing is too large, the capacitor will not self-heal and will become defective.

OIL CAPACITORS are often used in high-power electronic equipment. An oil-filled capacitor is nothing more than a paper capacitor that is immersed in oil. Since oil impregnated paper has a high dielectric constant, it can be used in the production of capacitors having a high capacitance value. Many capacitors will use oil with another dielectric material to prevent arcing between the plates. If arcing should occur between the plates of an oil-filled capacitor, the oil will tend to reseal the hole caused by the arcing. Such a capacitor is referred to as a SELF-HEALING capacitor.

VARIABLE CAPACITOR

A variable capacitor is constructed in such manner that its value of capacitance can be varied. A typical variable capacitor (adjustable capacitor) is the rotor-stator type. It consists of two sets of metal plates arranged so that the rotor plates move between the stator plates. Air is the dielectric. As the position of the rotor is changed, the capacitance value is likewise changed. This type of capacitor is used for tuning most radio receivers. Its physical appearance and its symbol are shown in figure 3-19.

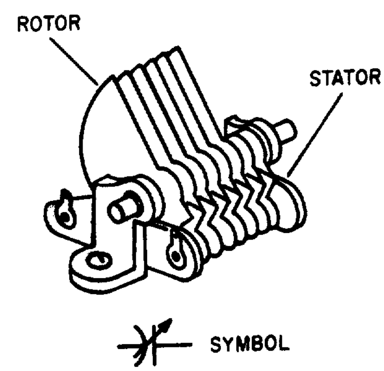


Figure 3-19.—Rotor-stator type variable capacitor.

Another type of variable capacitor (trimmer capacitor) and its symbol are shown in figure 3-20. This capacitor consists of two plates separated by a sheet of mica. A screw adjustment is used to vary the distance between the plates, thereby changing the capacitance.

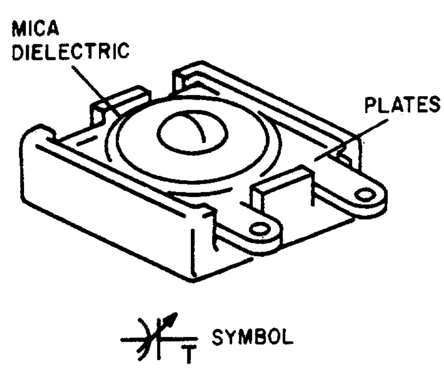


Figure 3-20.—Trimmer capacitor.

Q18.

a. An oxide-film dielectric is used in what type of capacitor?

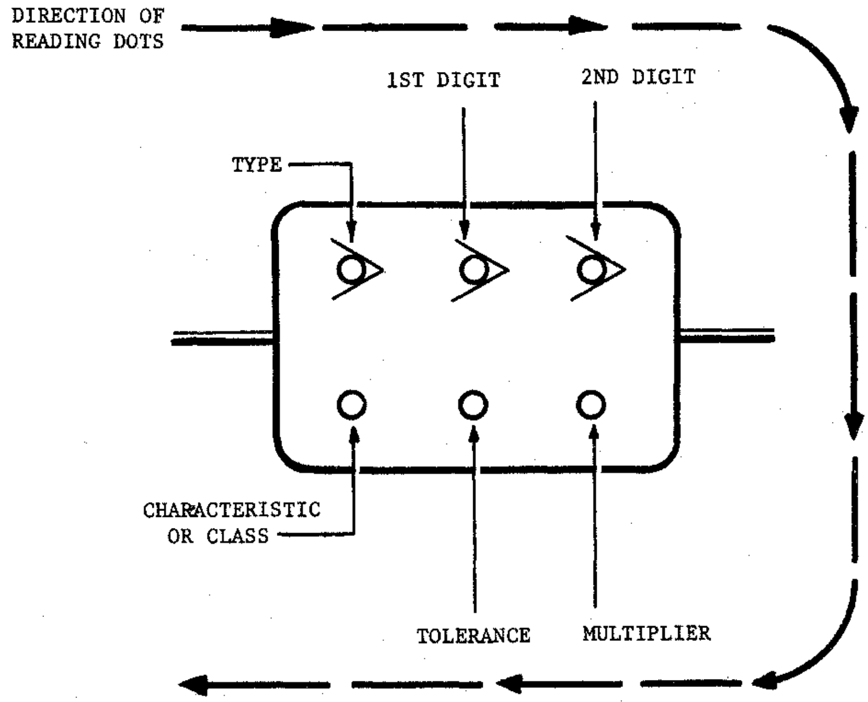
b. A screw adjustment is used to vary the distance between the plates of what type of capacitor?

COLOR CODES FOR CAPACITORS

Although the capacitance value may be printed on the body of a capacitor, it may also be indicated by a color code. The color code used to represent capacitance values is similar to that used to represent resistance values. The color codes currently in use are the Joint Army-Navy (JAN) code and the Radio Manufacturers' Association (RMA) code.

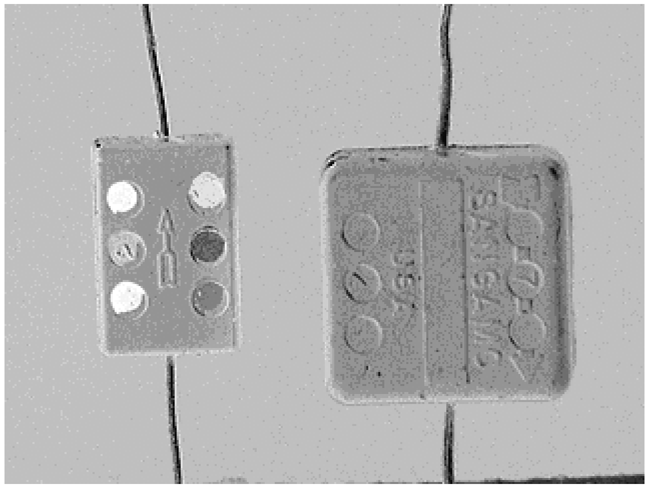
For each of these codes, colored dots or bands are used to indicate the value of the capacitor. A mica capacitor, it should be noted, may be marked with either three dots or six dots. Both the three- and the six-dot codes are similar, but the six-dot code contains more information about electrical ratings of the capacitor, such as working voltage and temperature coefficient.

The capacitor shown in figure 3-21 represents either a mica capacitor or a molded paper capacitor. To determine the type and value of the capacitor, hold the capacitor so that the three arrows point left to right (>). The first dot at the base of the arrow sequence (the left-most dot) represents the capacitor TYPE. This dot is either black, white, silver, or the same color as the capacitor body. Mica is represented by a black or white dot and paper by a silver dot or dot having the same color as the body of the capacitor. The two dots to the immediate right of the type dot indicate the first and second digits of the capacitance value. The dot at the bottom right represents the multiplier to be used. The multiplier represents picofarads. The dot in the bottom center indicates the tolerance value of the capacitor.

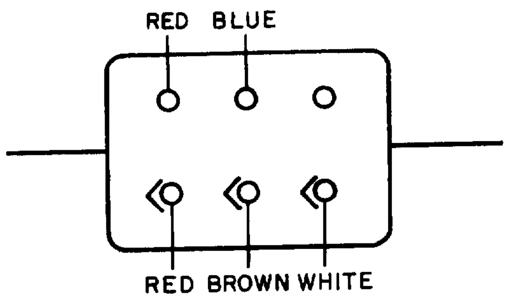


TYPE	COLOR	1ST DIGIT	2ND DIGIT	MULTIPLIER	TOLERANCE (PERCENT)	CHARACTERISTIC OR CLASS
JAN, MICA	BLACK	0	0	1.0	±20	APPLIES TO
	BROWN	1	1	10		TEMPERATURE
	RED	2	2	100	± 2	COEFFICIENT
	ORANGE	3	3	1,000	± 3	OR METHODS
	YELLOW	4	4	10,000	± 4	OF TESTING
	GREEN	5	5	100,000	± 5	
	BLUE	6	6	1,000,000	± 6	
	VIOLET	7	7	10,000,000	± 7	
	GRAY	8	8	100,000,000	± 8	
EIA, MICA	WHITE	9	9	1,000,000,000	+ 9	
	GOLD			.1		
MOLDED PAPER	SILVER			.01	±10	
	BODY				±20	

Figure 3-21.—6-dot color code for mica and molded paper capacitors.



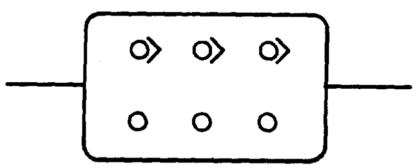
Example of mica capacitors.



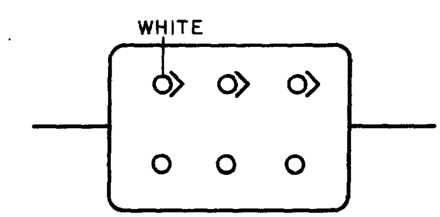
Example of mica capacitors.

To read the capacitor color code on the above capacitor:

1. Hold the capacitor so the arrows point left to right.

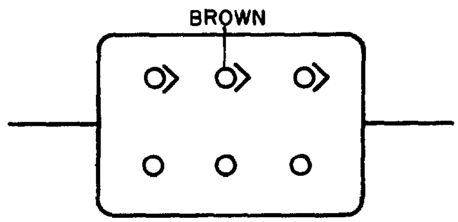


2. Read the first dot.



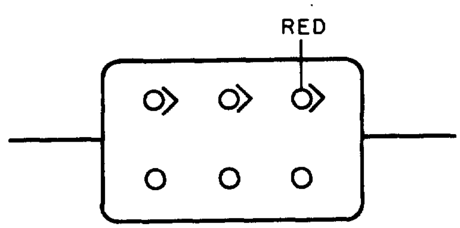
White = mica

3. Read the first digit dot.



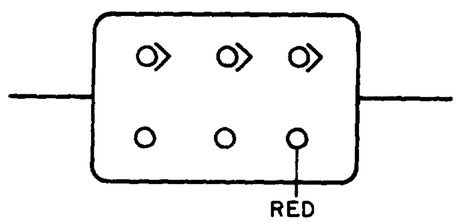
Brown = 1

4. Read the second digit dot and apply it to the first digit.



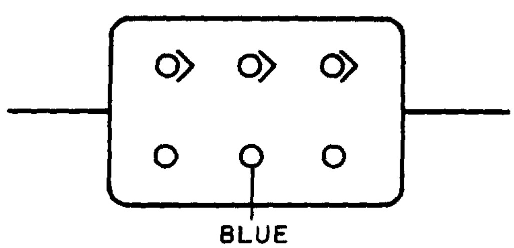
Red = 2 → 12

5. Read the multiplier dot and multiply the first two digits by multiplier. (Remember that the multiplier is in picofarads).



Red = 100 → 12 x 100 = 1200 pF

6. Lastly, read the tolerance dot.



Blue = ±6%

According to the above coding, the capacitor is a mica capacitor whose capacitance is 1200 pF with a tolerance of ±6%.

The capacitor shown in figure 3-22 is a tubular capacitor. Because this type of capacitor always has a paper dielectric, the type code is omitted. To read the code, hold the capacitor so the band closest to the end is on the left side; then read left to right. The last two bands (the fifth and sixth bands from the left) represent the voltage rating of the capacitor. This means that if a capacitor is coded red, red, red, yellow, yellow, yellow, it has the following digit values:

red = 2

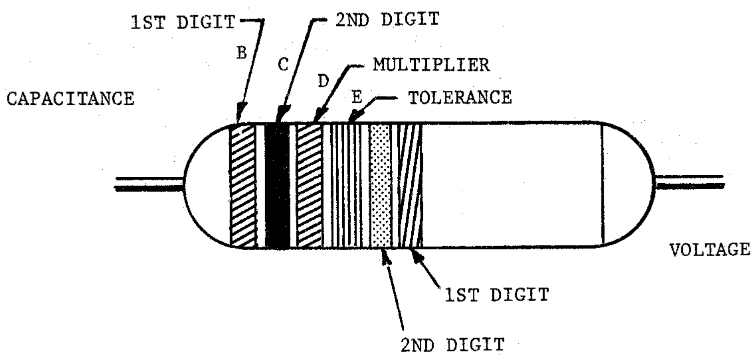
red = 2

red = × 100 pF

yellow = ±40%

yellow = 4

yellow = 4

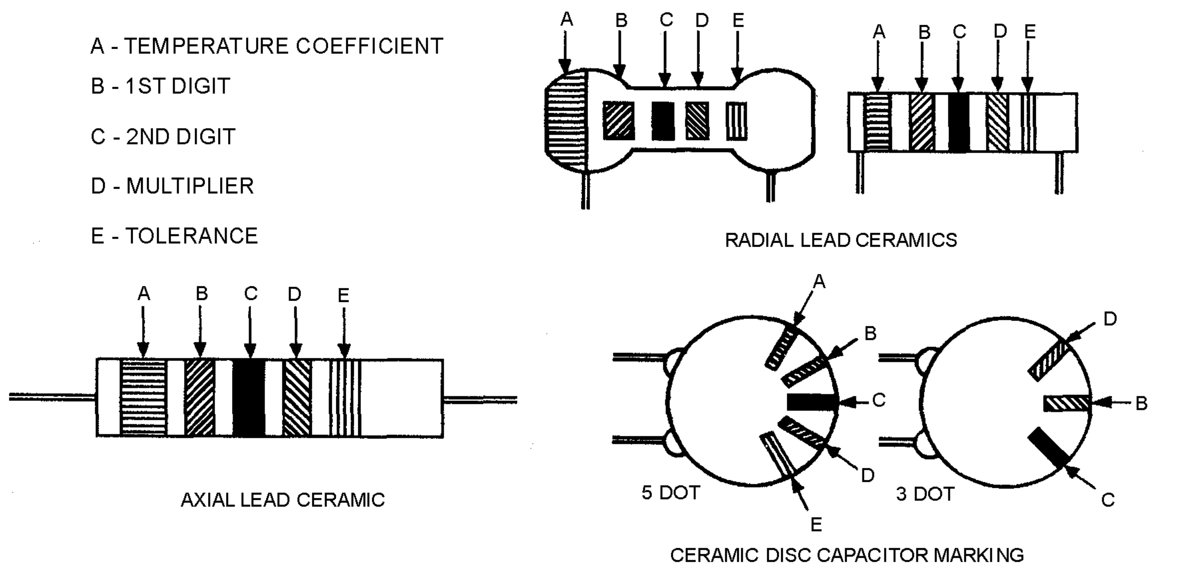


COLOR	CAPACITANCE 1ST DIGIT	CAPACITANCE 2ND DIGIT	CAPACITANCE MULTIPLIER	TOLERANCE (PERCENT)	VOLTAGE RATING 1ST DIGIT	VOLTAGE RATING 2ND DIGIT
BLACK	0	0	1	±20	0	0
BROWN	1	1	10		1	1
RED	2	2	100		2	2
ORANGE	3	3	1,000	±30	3	3
YELLOW	4	4	10,000	±40	4	4
GREEN	5	5	100,000	± 5	5	5
BLUE	6	6	1,000,000		6	6
VIOLET	7	7			7	7
GRAY	8	8			8	8
WHITE	9	9		±10	9	9

Figure 3-22.—6-band color code for tubular paper dielectric capacitors.

The six digits indicate a capacitance of 2200 pF with a ±40 percent tolerance and a working voltage of 44 volts.

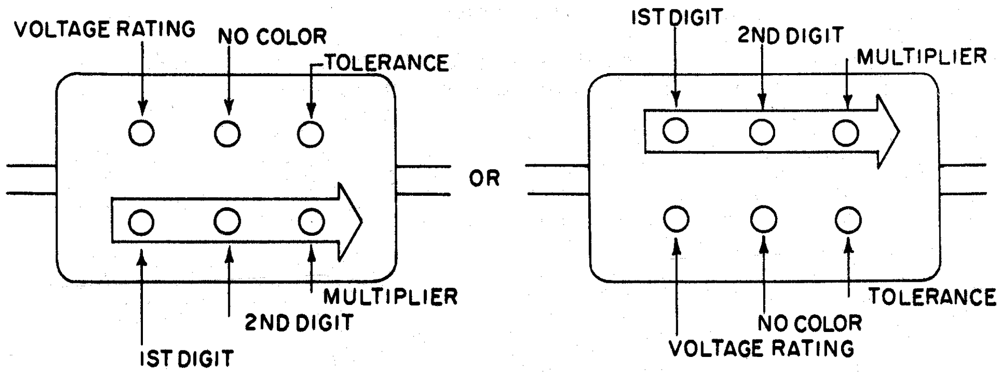
The ceramic capacitor is color coded as shown in figure 3-23 and the mica capacitor as shown in figure 3-24. Notice that this type of mica capacitor differs from the one shown in figure 3-21 in that the arrow is solid instead of broken. This type of mica capacitor is read in the same manner as the one shown in figure 3-21, with one exception: the first dot indicates the first digit. (Note: Because this type of capacitor is always mica, there is no need for a type dot.)



COLOR	1ST DIGIT	2ND DIGIT	MULTIPLIER	TOLERANCE MORE THAN 10pf (IN PERCENT)	TOLERANCE LESS THAN 10pf (IN pf)	TEMPERATURE COEFFICIENT*
BLACK	0	0	1.0	±20	±2.0	0
BROWN	1	1	10	±1		-30
RED	2	2	100	±2		-80
ORANGE	3	3	1,000			-150
YELLOW	4	4	10,000			-220
GREEN	5	5		±5	±0.5	-330
BLUE	6	6				-470
VIOLET	7	7				-750
GRAY	8	8	.01		±0.25	+30
WHITE	9	9	.1	±10	±1.0	+120 TO -750 (EIA)
						+500 TO -330 (JAN)
						+100 (JAN)
SILVER						
GOLD						BYPASS OR COUPLING (EIA)

* PARTS PER MILLION PER DEGREE CENTIGRADE

Figure 3-23.—Ceramic capacitor color code.

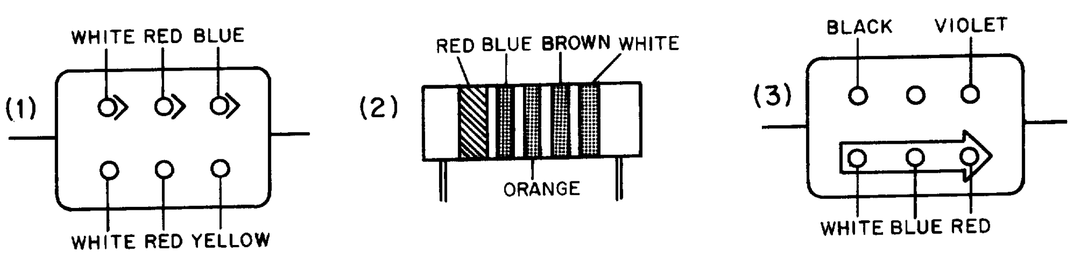


COLOR	1ST DIGIT	2ND DIGIT	MULTIPLIER	TOLERANCE (PERCENT)	VOLTAGE RATING
BLACK	0	0	1.0		
BROWN	1	1	10	± 1	100
RED	2	2	100	± 2	200
ORANGE	3	3	1,000	± 3	300
YELLOW	4	4	10,000	± 4	400
GREEN	5	5	100,000	± 5	500
BLUE	6	6	1,000,000	± 6	600
VIOLET	7	7	10,000,000	± 7	700
GRAY	8	8	100,000,000	± 8	800
WHITE	9	9	1,000,000,000	± 9	900
GOLD			.1		1000
SILVER			.01	±10	2000
BODY				±20	*

* WHERE NO COLOR IS INDICATED.THE VOLTAGE RATING MAY BE AS LOW AS 300 VOLTS.

Figure 3-24.—Mica capacitor color code.

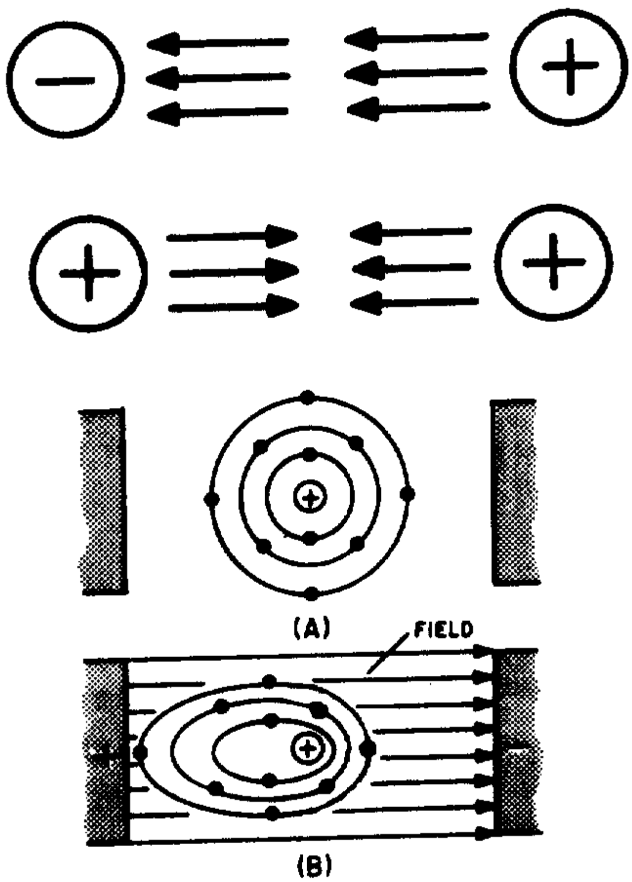
Q19. Examine the three capacitors shown below. What is the capacitance of each?



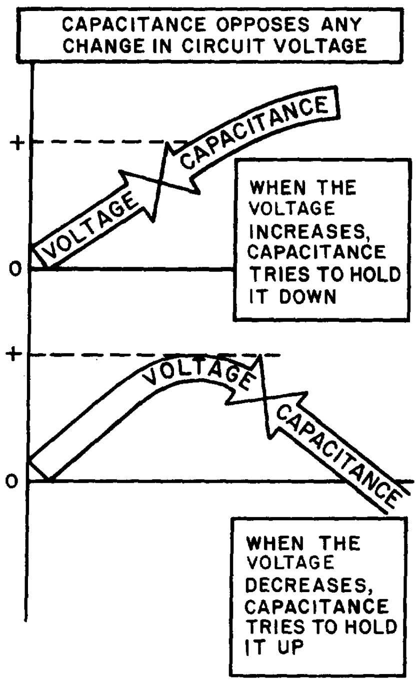
SUMMARY

Before going on to the next chapter, study the below summary to be sure that you understand the important points of this chapter.

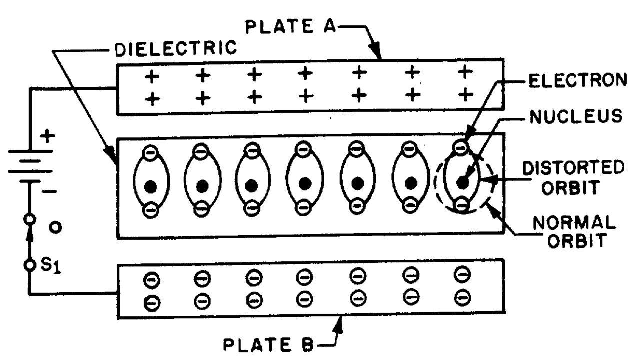
THE ELECTROSTATIC FIELD—When a charged body is brought close to another charged body, the bodies either attract or repel one another. (If the charges are alike they repel; if the charges are opposite they attract). The field that causes this effect is called the ELECTROSTATIC FIELD. The amount by which two charges attract or repel each other depends upon the size of the charges and the distance between the charges. The electrostatic field (force between two charged bodies) may be represented by lines of force drawn perpendicular to the charged surfaces. If an electron is placed in the field, it will move toward the positive charge.



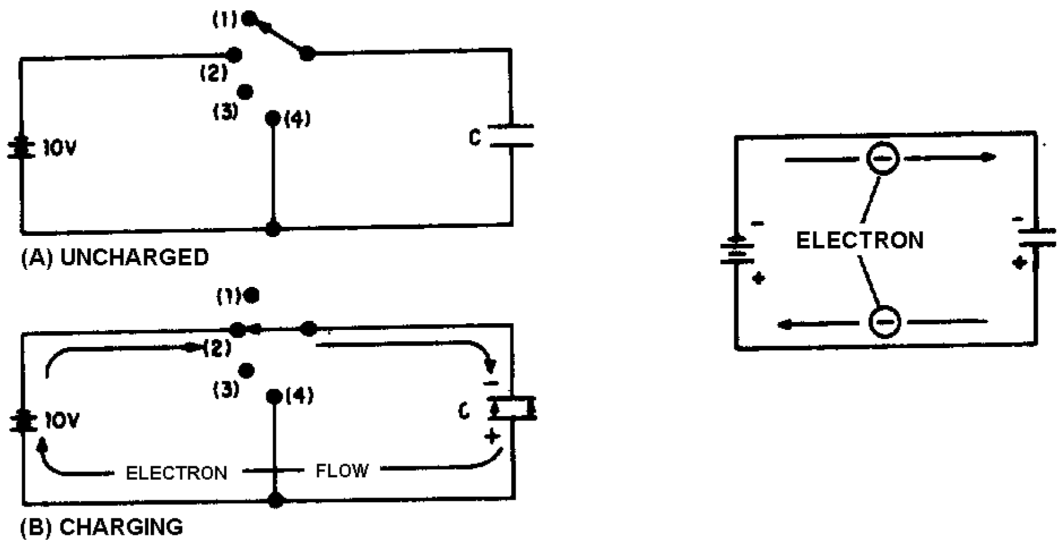
CAPACITANCE—Capacitance is the property of a circuit which OPPOSES any CHANGE in the circuit VOLTAGE. The effect of capacitance may be seen in any circuit where the voltage is changing. Capacitance is usually defined as the ability of a circuit to store electrical energy. This energy is stored in an electrostatic field. The device used in an electrical circuit to store this charge (energy) is called a CAPACITOR. The basic unit of measurement of capacitance is the FARAD (F). A one-farad capacitor will store one coulomb of charge (energy) when a potential of one volt is applied across the capacitor plates. The farad is an enormously large unit of capacitance. More practical units are the microfarad (μF) or the picofarad (pF).

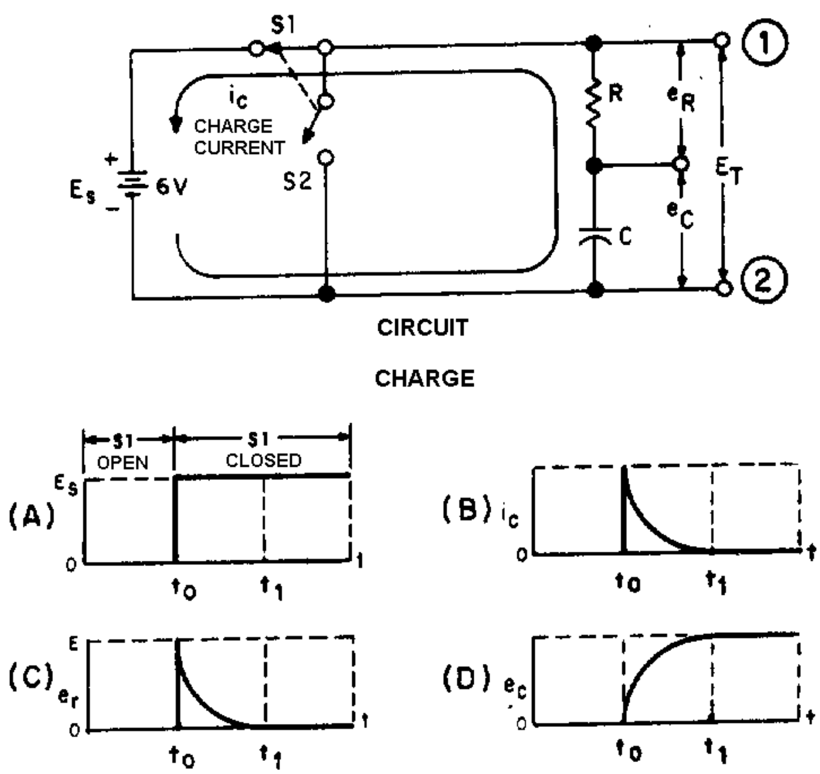


CAPACITOR—A capacitor is a physical device consisting of two pieces of conducting material separated by an insulating material. This insulating material is referred to as the DIELECTRIC. Because the dielectric is an insulator, NO current flows through the capacitor. If the dielectric breaks down and becomes a conductor, the capacitor can no longer hold a charge and is useless. The ability of a dielectric to hold a charge without breaking down is referred to as the dielectric strength. The measure of the ability of the dielectric material to store energy is called the dielectric constant. The dielectric constant is a relative value based on 1.0 for a vacuum.



CAPACITORS IN A DC CIRCUIT—When a capacitor is connected to the terminals of a battery, each plate of the capacitor becomes charged. The plate connected to the positive terminal loses electrons. Because this plate has a lack of electrons, it assumes a positive charge. The plate connected to the negative terminal gains electrons. Because the plate has an excess of electrons, it assumes a negative charge. This process continues until the charge across the plates equals the applied voltage. At this point current ceases to flow in the circuit. As long as nothing changes in the circuit, the capacitor will hold its charge and there will be no current in any part of the circuit. If the leads of the capacitor are now shorted together, current again flows in the circuit. Current will continue to flow until the charges on the two plates become equal. At this point, current ceases to flow. With a dc voltage source, current will flow in the circuit only long enough to charge (or discharge) the capacitor. Thus, a capacitor does NOT allow dc current to flow continuously in a circuit.





FACTORS AFFECTING CAPACITANCE—There are three factors affecting capacitance. One factor is the area of the plate surfaces. Increasing the area of the plate increases the capacitance. Another factor is the amount of space between the plates. The closer the plates, the greater will be the electrostatic field. A greater electrostatic field causes a greater capacitance. The plate spacing is determined by the thickness of the dielectric. The third factor affecting capacitance is the dielectric constant. The value of the dielectric constant is dependent upon the type of dielectric used.

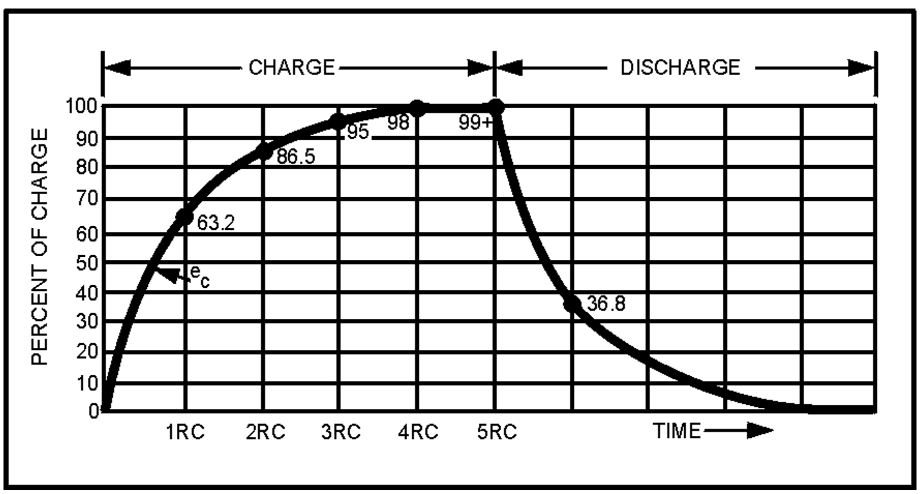
WORKING VOLTAGE—The working voltage of a capacitor is the maximum voltage that can be steadily applied to the capacitor without the capacitor breaking down (shorting). The working voltage depends upon the type of material used as the dielectric (the dielectric constant) and the thickness of the dielectric.

CAPACITOR LOSSES—Power losses in a capacitor are caused by dielectric leakage and dielectric hysteresis. Dielectric leakage loss is caused by the leakage current through the resistance in the dielectric. Although this resistance is extremely high, a small amount of current does flow. Dielectric hysteresis may be defined as an effect in a dielectric material similar to the hysteresis found in a magnetic material.

RC TIME CONSTANT—The time required to charge a capacitor to 63.2 percent of the applied voltage, or to discharge the capacitor to 36.8 percent of its charge. The time constant (t) is equal to the product of the resistance and the capacitance. Expressed as a formula:

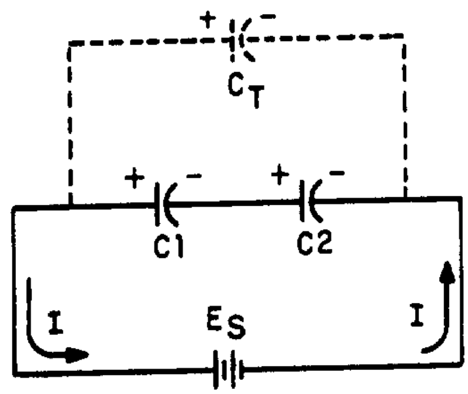
$$t = RC$$

where t is in seconds, R is in ohms, and C is in farads.



CAPACITORS IN SERIES—The effect of wiring capacitors in series is to increase the distance between plates. This reduces the total capacitance of the circuit. Total capacitance for series connected capacitors may be computed by the formula:

$$C_T = \frac{1}{\frac{1}{C1} + \frac{1}{C2} + \frac{1}{C3} + \ldots \frac{1}{C_n}}$$

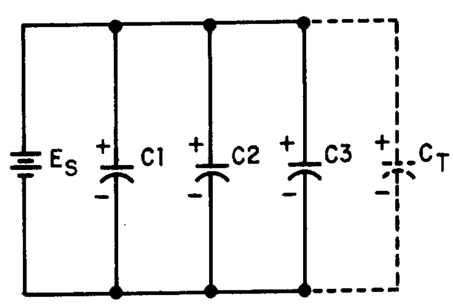


If an electrical circuit contains only two series connected capacitors, C_T may be computed using the following formula:

$$C_T = \frac{C1\,C2}{C1 + C2}$$

CAPACITORS IN PARALLEL—The effect of wiring capacitors in parallel is to increase the plate area of the capacitors. Total capacitance (C_T) may be found using the formula:

$$C_T = C1 + C2 \ldots + C_n$$



TYPES OF CAPACITORS—Capacitors are manufactured in various forms and may be divided into two main classes-fixed capacitors and variable capacitors. A fixed capacitor is constructed to have a constant or fixed value of capacitance. A variable capacitor allows the capacitance to be varied or adjusted.

ANSWERS TO QUESTIONS Q1. THROUGH Q19.

A1.

a. A capacitor is a device that stores electrical energy in an electrostatic field.

b. Capacitance is the property of a circuit which opposes changes in voltage.

A2.

a. They are polarized from positive to negative.

b. They radiate from a charged particle in straight lines and do not form closed loops.

c. They have the ability to pass through any known material.

d. They have the ability to distort the orbits of electrons circling the nucleus.

A3. Toward the positive charge.

A4. Two pieces of conducting material separated by an insulator.

A5. A farad is the unit of capacitance. A capacitor has a capacitance of 1 farad when a difference of 1 volt will charge it with 1 coulomb of electrons.

A6.

a. One microfarad equals 10^{-6} *farad.*

b. One picofarad equals 10^{-12} *farad.*

A7.

a. The area of the plates.

b. The distance between the plates.

c. The dielectric constant of the material between the plates.

A8.

4372 picofarads

$$C = .2249\left(\frac{KA}{d}\right)$$

$$C = .2249\left(\frac{81 \times 6}{.025}\right)$$

$$C = 4372 \text{ (Rounded off)}$$

A9.

a. Hysteresis

b. Dielectric leakage

A10.

a. It is the maximum voltage the capacitor can work without risk of damage.

b. 900 volts.

A11.

a. When the capacitor is charging, electrons accumulate on the negative plate and leave the positive plate until the charge on the capacitor is equal to the battery voltage.

b. When the capacitor is discharging, electrons flow from the negatively charged plate to the positively charged plate until the charge on each plate is neutral.

A12. At the instant of the initiation of the action.

A13. Zero.

A14.

144 seconds

$$t = R\,(\text{megohms}) \times C\,(\text{microfarads})$$
$$t = 12 \times 12$$
$$t = 144\ \text{seconds}$$

A15.

.01 microfarads 40% from the graph = .5

$$RC = \frac{200}{.5}$$
$$RC = 400\ \text{microseconds}$$
$$C = \frac{t}{R}$$
$$C = \frac{400\,\mu s}{40{,}000\,\Omega}$$
$$C = .01\,\mu F = 10{,}000\,pF$$

A16.

$. 1 \mu F$

$$C_T = \frac{C_1 C_2}{C_1 + C_2}$$

$$C_T = \frac{10 \times 0.1}{10 + 0.1} \mu F$$

$$C_T = \frac{1}{10.1} \mu F$$

$$C_T = .099 \mu F \text{ or } 0.1 \mu F$$

A17.

$33.1 \mu F$

$$C_T = C1 + C2 + C3 + C4$$

$$C_T = 10 \mu F + 21 \mu F + 0.1 \mu F + 2 \mu F$$

$$C_T = 33.1 \mu F$$

A18.

a. *Electrolytic capacitor*

b. *Trimmer capacitor*

A19.

a. *26 μF or 260,000 pF*

b. *630 pF*

c. *9600 pF*

CHAPTER 4

INDUCTIVE AND CAPACITIVE REACTANCE

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. State the effects an inductor has on a change in current and a capacitor has on a change in voltage.
2. State the phase relationships between current and voltage in an inductor and in a capacitor.
3. State the terms for the opposition an inductor and a capacitor offer to ac
4. Write the formulas for inductive and capacitive reactances.
5. State the effects of a change in frequency on X_L and X_C.
6. State the effects of a change in inductance on X_L and a change in capacitance on X_C.
7. Write the formula for determining total reactance (X); compute total reactance (X) in a series circuit; and indicate whether the total reactance is capacitive or inductive.
8. State the term given to the total opposition (Z) in an ac circuit.
9. Write the formula for impedance, and calculate the impedance in a series circuit when the values of X_C, X_L, and R are given.
10. Write the Ohm's law formulas used to determine voltage and current in an ac circuit.
11. Define true power, reactive power, and apparent power; state the unit of measurement for and the formula used to calculate each.
12. State the definition of and write the formula for power factor.
13. Given the power factor and values of X and R in an ac circuit, compute the value of reactance in the circuit, and state the type of reactance that must be connected in the circuit to correct the power factor to unity (1).
14. State the difference between calculating impedance in a series ac circuit and in a parallel ac circuit.

INDUCTIVE AND CAPACITIVE REACTANCE

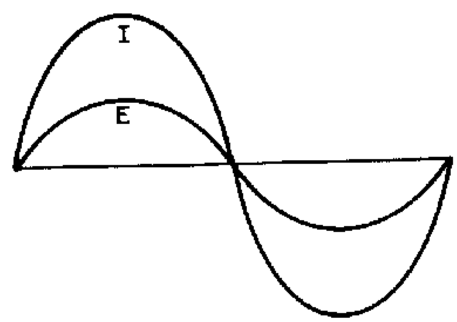
You have already learned how inductance and capacitance individually behave in a direct current circuit. In this chapter you will be shown how inductance, capacitance, and resistance affect alternating current.

INDUCTANCE AND ALTERNATING CURRENT

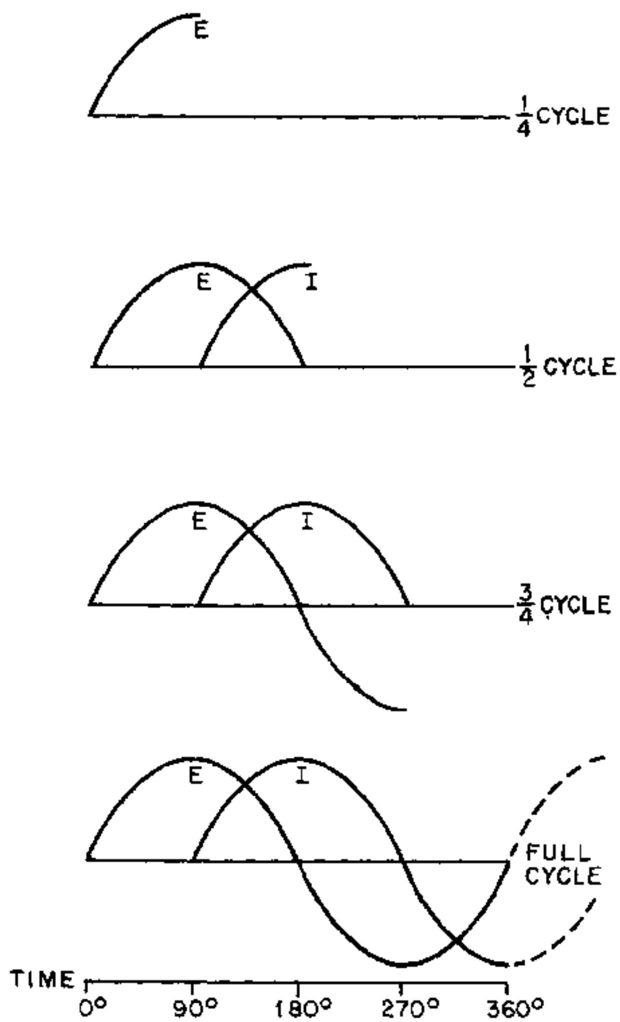
This might be a good place to recall what you learned about phase in chapter 1. When two things are in step, going through a cycle together, falling together and rising together, they are in phase. When they are out of phase, the angle of lead or lag-the number of electrical degrees by which one of the values leads or lags the other-is a measure of the amount they are out of step. The time it takes the current in an inductor to build up to maximum and to fall to zero is important for another reason. It helps illustrate a very useful characteristic of inductive circuits-the current through the inductor always lags the voltage across the inductor.

A circuit having pure resistance (if such a thing were possible) would have the alternating current through it and the voltage across it rising and failing together. This is illustrated in figure 4-1(A), which shows the sine waves for current and voltage in a purely resistive circuit having an ac source. The current and voltage do not have the same amplitude, but they are in phase.

In the case of a circuit having inductance, the opposing force of the counter emf would be enough to keep the current from remaining in phase with the applied voltage. You learned that in a dc circuit containing pure inductance the current took time to rise to maximum even though the full applied voltage was immediately at maximum. Figure 4-1(B) shows the wave forms for a purely inductive ac circuit in steps of quarter-cycles.



(A)



(B)

Figure 4-1.—Voltage and current waveforms in an inductive circuit.

With an ac voltage, in the first quarter-cycle (0º to 90º) the applied ac voltage is continually increasing. If there was no inductance in the circuit, the current would also increase during this first quarter-cycle. You know this circuit does have inductance. Since inductance opposes any change in current flow, no current flows during the first quarter-cycle. In the next quarter-cycle (90º to 180º) the voltage decreases back to zero; current begins to flow in the circuit and reaches a maximum value at the same instant the voltage reaches zero. The applied voltage now begins to build up to maximum in the other direction, to be followed by the resulting current. When the voltage again reaches its maximum at the end of the third quarter-cycle (270º) all values are exactly opposite to what they were during the first half-cycle. The applied voltage leads the resulting current by one quarter-cycle or 90 degrees. To complete the full 360º cycle of the voltage, the voltage again decreases to zero and the current builds to a maximum value.

You must not get the idea that any of these values stops cold at a particular instant. Until the applied voltage is removed, both current and voltage are always changing in amplitude and direction.

As you know the sine wave can be compared to a circle. Just as you mark off a circle into 360 degrees, you can mark off the time of one cycle of a sine wave into 360 electrical degrees. This relationship is shown in figure 4-2. By referring to this figure you can see why the current is said to lag the voltage, in a purely inductive circuit, by 90 degrees. Furthermore, by referring to figures 4-2 and 4-1(A) you can see why the current and voltage are said to be in phase in a purely resistive circuit. In a circuit having both resistance and inductance then, as you would expect, the current lags the voltage by an amount somewhere between 0 and 90 degrees.

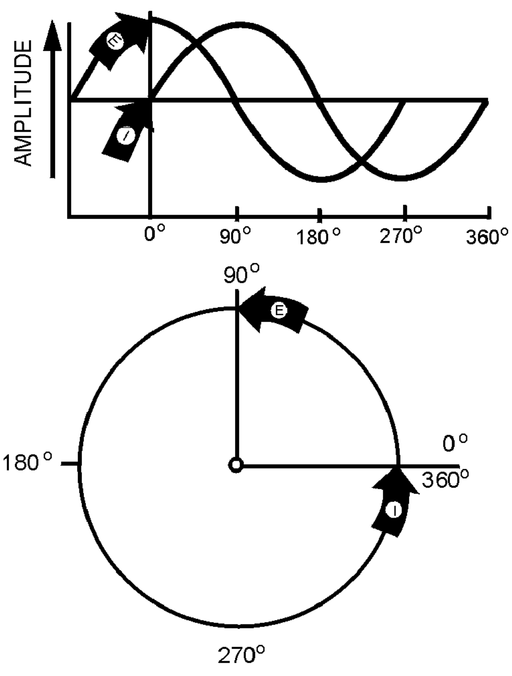


Figure 4-2.—Comparison of sine wave and circle in an inductive circuit.

A simple memory aid to help you remember the relationship of voltage and current in an inductive circuit is the word ELI. Since E is the symbol for voltage, L is the symbol for inductance, and I is the symbol for current; the word ELI demonstrates that current comes after (Lags) voltage in an inductor.

Q1. What effect does an inductor have on a change in current?

Q2. What is the phase relationship between current and voltage in an inductor?

INDUCTIVE REACTANCE

When the current flowing through an inductor continuously reverses itself, as in the case of an ac source, the inertia effect of the cemf is greater than with dc. The greater the amount of inductance (L), the greater the opposition from this inertia effect. Also, the faster the reversal of current, the greater this inertial opposition. This opposing force which an inductor presents to the FLOW of alternating current cannot be called resistance, since it is not the result of friction within a conductor. The name given to it is INDUCTIVE REACTANCE because it is the "reaction" of the inductor to the changing value of alternating current. Inductive reactance is measured in ohms and its symbol is X_L.

As you know, the induced voltage in a conductor is proportional to the rate at which magnetic lines of force cut the conductor. The greater the rate (the higher the frequency), the greater the cemf. Also, the induced voltage increases with an increase in inductance; the more ampere-turns, the greater the cemf. Reactance, then, increases with an increase of frequency and with an increase of inductance. The formula for inductive reactance is as follows:

$$X_L = 2\pi fL$$

Where:

X_L is inductive reactance in ohms.

2π is a constant in which the Greek letter π, called "pi" represents 3.1416 and $2 \times \pi =$ 6.28 approximately.

f is frequency of the alternating current in Hz.

L is inductance in henrys.

The following example problem illustrates the computation of X_L.

Given: $f = 60$ Hz

$L = 20$ H

Solution: $X_L = 2\pi fL$

$X_L = 6.28 \times 60 \text{ Hz} \times 20 \text{ H}$

$X_L = 7{,}536\ \Omega$

Q3. What is the term for the opposition an inductor presents to ac?

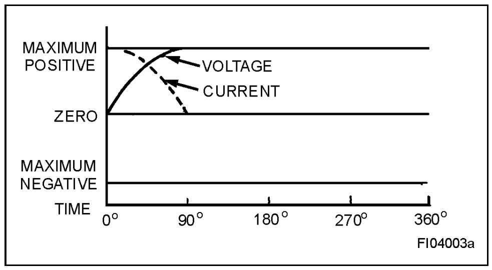
Q4. What is the formula used to compute the value of this opposition?

Q5. What happens to the value of X_L as frequency increases?

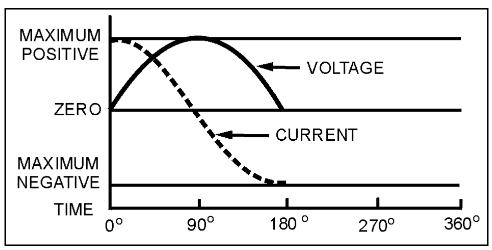
Q6. What happens to the value of X_L as inductance decreases?

CAPACITORS AND ALTERNATING CURRENT

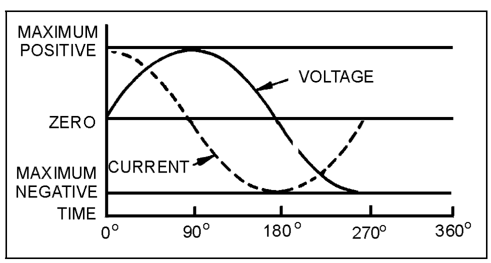
The four parts of figure 4-3 show the variation of the alternating voltage and current in a capacitive circuit, for each quarter of one cycle. The solid line represents the voltage across the capacitor, and the dotted line represents the current. The line running through the center is the zero, or reference point, for both the voltage and the current. The bottom line marks off the time of the cycle in terms of electrical degrees. Assume that the ac voltage has been acting on the capacitor for some time before the time represented by the starting point of the sine wave in the figure.



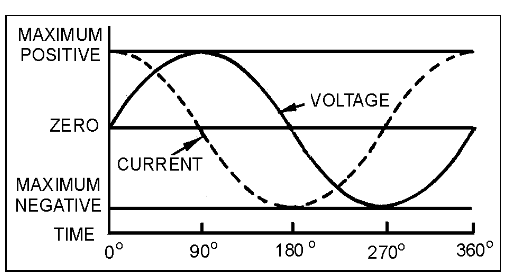
(A)



(B)



(C)



(D)

Figure 4-3.—Phase relationship of voltage and current in a capacitive circuit.

At the beginning of the first quarter-cycle (0º to 90º) the voltage has just passed through zero and is increasing in the positive direction. Since the zero point is the steepest part of the sine wave, the voltage is changing at its greatest rate. The charge on a capacitor varies directly with the voltage, and therefore the charge on the capacitor is also changing at its greatest rate at the beginning of the first quarter-cycle. In other words, the greatest number of electrons are moving off one plate and onto the other plate. Thus the capacitor current is at its maximum value, as part (A) of the figure shows.

As the voltage proceeds toward maximum at 90 degrees, its rate of change becomes less and less, hence the current must decrease toward zero. At 90 degrees the voltage across the capacitor is maximum, the capacitor is fully charged, and there is no further movement of electrons from plate to plate. That is why the current at 90 degrees is zero.

At the end of this first quarter-cycle the alternating voltage stops increasing in the positive direction and starts to decrease. It is still a positive voltage, but to the capacitor the decrease in voltage means that the plate which has just accumulated an excess of electrons must lose some electrons. The current flow, therefore, must reverse its direction. Part (B) of the figure shows the current curve to be below the zero line (negative current direction) during the second quarter-cycle (90º to 180º).

At 180 degrees the voltage has dropped to zero. This means that for a brief instant the electrons are equally distributed between the two plates; the current is maximum because the rate of change of voltage is maximum. Just after 180 degrees the voltage has reversed polarity and starts building up its maximum negative peak which is reached at the end of the third quarter-cycle (180º to 270º). During this third quarter-cycle the rate of voltage change gradually decreases as the charge builds to a maximum at 270 degrees. At this point the capacitor is fully charged and it carries the full impressed voltage. Because the capacitor is fully charged there is no further exchange of electrons; therefore, the current flow is zero at this point. The conditions are exactly the same as at the end of the first quarter-cycle (90º) but the polarity is reversed.

Just after 270 degrees the impressed voltage once again starts to decrease, and the capacitor must lose electrons from the negative plate. It must discharge, starting at a minimum rate of flow and rising to a maximum. This discharging action continues through the last quarter-cycle (270º to 360º) until the impressed-voltage has reached zero. At 360 degrees you are back at the beginning of the entire cycle, and everything starts over again.

If you examine the complete voltage and current curves in part D, you will see that the current always arrives at a certain point in the cycle 90 degrees ahead of the voltage, because of the charging and discharging action. You know that this time and place relationship between the current and voltage is called the phase relationship. The voltage-current phase relationship in a capacitive circuit is exactly opposite to that in an inductive circuit. <u>The current of a capacitor leads the voltage across the capacitor by 90 degrees</u>.

You realize that the current and voltage are both going through their individual cycles at the same time during the period the ac voltage is impressed. The current does not go through part of its cycle (charging or discharging), stop, and wait for the voltage to catch up. The amplitude and polarity of the voltage and the amplitude and direction of the current are continually changing. Their positions with respect to each other and to the zero line at any electrical instant-any degree between zero and 360 degrees-can be seen by reading upwards from the time-degree line. The current swing from the positive peak at zero degrees to the negative peak at 180 degrees is NOT a measure of the number of electrons, or the charge on the plates. It is a picture of the direction and strength of the current in relation to the polarity and strength of the voltage appearing across the plates.

At times it is convenient to use the word "ICE" to recall to mind the phase relationship of the current and voltage in capacitive circuits. I is the symbol for current, and in the word ICE it leads, or comes before, the symbol for voltage, E. C, of course, stands for capacitor. This memory aid is similar to the "ELI" used to remember the current and voltage relationship in an inductor. The phrase "ELI the ICE man" is helpful in remembering the phase relationship in both the inductor and capacitor.

Since the plates of the capacitor are changing polarity at the same rate as the ac voltage, the capacitor seems to pass an alternating current. Actually, the electrons do not pass through the dielectric, but their rushing back and forth from plate to plate causes a current flow in the circuit. It is convenient, however, to say that the alternating current flows "through" the capacitor. You know this is not true, but the expression avoids a lot of trouble when speaking of current flow in a circuit containing a capacitor. By the same short cut, you may say that the capacitor does not pass a direct current (if both plates are connected to a dc source, current will flow only long enough to charge the capacitor). With a capacitor type of hookup in a circuit containing both ac and dc, only the ac will be "passed" on to another circuit.

You have now learned two things to remember about a capacitor: <u>A capacitor will appear to conduct an alternating current and a capacitor will not conduct a direct current</u>.

Q7. What effect does the capacitor have on a changing voltage?

Q8. What is the phase relationship between current and voltage in a capacitor?

CAPACITIVE REACTANCE

So far you have been dealing with the capacitor as a device which passes ac and in which the only opposition to the alternating current has been the normal circuit resistance present in any conductor. However, capacitors themselves offer a very real opposition to current flow. This opposition arises from the fact that, at a given voltage and frequency, the number of electrons which go back and forth from plate to plate is limited by the storage ability-that is, the capacitance-of the capacitor. As the capacitance is increased, a greater number of electrons change plates every cycle, and (since current is a measure of the number of electrons passing a given point in a given time) the current is increased.

Increasing the frequency will also decrease the opposition offered by a capacitor. This occurs because the number of electrons which the capacitor is capable of handling at a given voltage will change plates more often. As a result, more electrons will pass a given point in a given time (greater current flow). The opposition which a capacitor offers to ac is therefore inversely proportional to frequency and to capacitance. This opposition is called CAPACITIVE REACTANCE. You may say that capacitive reactance decreases with increasing frequency or, for a given frequency, the capacitive reactance decreases with increasing capacitance. The symbol for capacitive reactance is X_C.

Now you can understand why it is said that the X_C varies inversely with the product of the frequency and capacitance. The formula is:

$$X_C = \frac{1}{2\pi fC}$$

Where:

X_C is capacitive reactance in ohms

f is frequency in Hertz

C is capacitance in farads

π is 6.28 (2×3.1416)

The following example problem illustrates the computation of X_C.

Given: $f = 100$ Hz

$C = 50\,\mu F$

Solution:

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{6.28 \times 100 \text{ Hz} \times 50\,\mu F}$$

$$X_C = \frac{1}{.0314}\Omega$$

$$X_C = 31.8\,\Omega \text{ or } 32\,\Omega$$

Q9. What is the term for the opposition that a capacitor presents to ac?

Q10. What is the formula used to compute this opposition?

Q11. What happens to the value of X_C as frequency decreases?

Q12. What happens to the value of X_C as capacitance increases?

REACTANCE, IMPEDANCE, AND POWER RELATIONSHIPS IN AC CIRCUITS

Up to this point inductance and capacitance have been explained individually in ac circuits. The rest of this chapter will concern the combination of inductance, capacitance, and resistance in ac circuits.

To explain the various properties that exist within ac circuits, the series RLC circuit will be used. Figure 4-4 is the schematic diagram of the series RLC circuit. The symbol shown in figure 4-4 that is marked E is the general symbol used to indicate an ac voltage source.

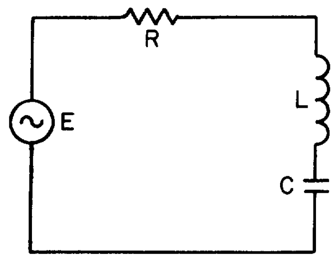


Figure 4-4.—Series RLC circuit.

REACTANCE

The effect of inductive reactance is to cause the current to lag the voltage, while that of capacitive reactance is to cause the current to lead the voltage. Therefore, since inductive reactance and capacitive reactance are exactly opposite in their effects, what will be the result when the two are combined? It is not hard to see that the net effect is a tendency to cancel each other, with the combined effect then equal to the difference between their values. This resultant is called REACTANCE; it is represented by the symbol X; and expressed by the equation $X = X_L - X_C$ or $X = X_C - X_L$. Thus, if a circuit contains 50 ohms of inductive reactance and 25 ohms of capacitive reactance in series, the net reactance, or X, is 50 ohms – 25 ohms, or 25 ohms of inductive reactance.

For a practical example, suppose you have a circuit containing an inductor of 100 µH in series with a capacitor of .001 µF, and operating at a frequency of 4 MHz. What is the value of net reactance, or X?

Given:
$$f = 4\,\text{MHz}$$
$$L = 100\,\mu\text{H}$$
$$C = .001\,\mu\text{F}$$

Solution:
$$X_L = 2\pi fL$$
$$X_L = 6.28 \times 4\,\text{MHz} \times 100\,\mu\text{H}$$
$$X_L = 2512\,\Omega$$
$$X_C = \frac{1}{2\pi fC}$$
$$X_C = \frac{1}{6.28 \times 4\text{ MHz} \times .001\,\mu\text{F}}$$
$$X_C = \frac{1}{.02512}\Omega$$
$$X_C = 39.8\,\Omega$$
$$X = X_L - X_C$$
$$X = 2512\,\Omega - 39.8\,\Omega$$
$$X = 2472.2\,\Omega \text{ (inductive)}$$

Now assume you have a circuit containing a 100 - µH inductor in series with a .0002-µF capacitor, and operating at a frequency of 1 MHz. What is the value of the resultant reactance in this case?

Given:

$$f = 1\,\text{MHz}$$
$$L = 100\,\mu H$$
$$C = .0002\,\mu F$$

Solution:

$$X_L = 2\pi fL$$
$$X_L = 6.28 \times 1\,\text{MHz} \times 100\,\mu H$$
$$X_L = 628\,\Omega$$
$$X_C = \frac{1}{2\pi fC}$$
$$X_C = \frac{1}{6.28 \times 1\,\text{MHz} \times .0002\,\mu F}$$
$$X_C = \frac{1}{.001256}\,\Omega$$
$$X_C = 796\,\Omega$$
$$X = X_C - X_L$$
$$X = 796\,\Omega - 628\,\Omega$$
$$X = 168\,\Omega\ (\text{capacitive})$$

You will notice that in this case the inductive reactance is smaller than the capacitive reactance and is therefore subtracted from the capacitive reactance.

These two examples serve to illustrate an important point: when capacitive and inductive reactance are combined in series, the smaller is always subtracted from the larger and the resultant reactance always takes the characteristics of the larger.

Q13. What is the formula for determining total reactance in a series circuit where the values of X_C and X_L are known?

Q14. What is the total amount of reactance (X) in a series circuit which contains an X_L of 20 ohms and an X_C of 50 ohms? (Indicate whether X is capacitive or inductive)

IMPEDANCE

From your study of inductance and capacitance you know how inductive reactance and capacitive reactance act to oppose the flow of current in an ac circuit. However, there is another factor, the resistance, which also opposes the flow of the current. Since in practice ac circuits containing reactance also contain resistance, the two combine to oppose the flow of current. This combined opposition by the resistance and the reactance is called the IMPEDANCE, and is represented by the symbol Z.

Since the values of resistance and reactance are both given in ohms, it might at first seem possible to determine the value of the impedance by simply adding them together. It cannot be done so easily, however. You know that in an ac circuit which contains only resistance, the current and the voltage will be in step (that is, in phase), and will reach their maximum values at the same instant. You also know that in an ac circuit containing only reactance the current will either lead or lag the voltage by one-quarter of a cycle or 90 degrees. Therefore, the voltage in a purely reactive circuit will differ in phase by 90 degrees from that in a purely resistive circuit and for this reason reactance and resistance are rot combined by simply adding them.

When reactance and resistance are combined, the value of the impedance will be greater than either. It is also true that the current will not be in step with the voltage nor will it differ in phase by exactly 90 degrees from the voltage, but it will be somewhere between the in-step and the 90-degree out-of-step conditions. The larger the reactance compared with the resistance, the more nearly the phase difference will approach 90º . The larger the resistance compared to the reactance, the more nearly the phase difference will approach zero degrees.

If the value of resistance and reactance cannot simply be added together to find the impedance, or Z, how is it determined? Because the current through a resistor is in step with the voltage across it and the current in a reactance differs by 90 degrees from the voltage across it, the two are at right angles to each other. They can therefore be combined by means of the same method used in the construction of a right-angle triangle.

Assume you want to find the impedance of a series combination of 8 ohms resistance and 5 ohms inductive reactance. Start by drawing a horizontal line, R, representing 8 ohms resistance, as the base of the triangle. Then, since the effect of the reactance is always at right angles, or 90 degrees, to that of the resistance, draw the line X_L, representing 5 ohms inductive reactance, as the altitude of the triangle. This is shown in figure 4-5. Now, complete the hypotenuse (longest side) of the triangle. Then, the hypotenuse represents the impedance of the circuit.

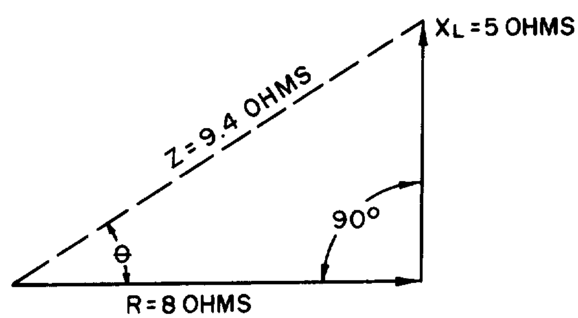


Figure 4-5.—Vector diagram showing relationship of resistance, inductive reactance, and impedance in a series circuit.

One of the properties of a right triangle is:

$$(\text{hypotenuse})^2 = (\text{base})^2 + (\text{altitude})^2$$

or,

$$\text{hypotenuse} = \sqrt{(\text{base})^2 + (\text{altitude})^2}$$

Applied to impedance, this becomes,

$$(\text{impedance})^2 = (\text{resistance})^2 + (\text{reactance})^2$$

or,

$$\text{impedance} = \sqrt{(\text{resistance})^2 + (\text{reactance})^2}$$

or,

$$Z = \sqrt{R^2 + X^2}$$

Now suppose you apply this equation to check your results in the example given above.

Given: $R = 8\,\Omega$

$X_L = 5\,\Omega$

Solution:

$$Z = \sqrt{R^2 + X_L^2}$$

$$Z = \sqrt{(8\,\Omega)^2 + (5\,\Omega)^2}$$

$$Z = \sqrt{64 + 25\,\Omega}$$

$$Z = \sqrt{89\,\Omega}$$

$$Z = 9.4\,\Omega$$

(See the Appendix III for a square Root Table.)

When you have a capacitive reactance to deal with instead of inductive reactance as in the previous example, it is customary to draw the line representing the capacitive reactance in a downward direction. This is shown in figure 4-6. The line is drawn downward for capacitive reactance to indicate that it acts in a direction opposite to inductive reactance which is drawn upward. In a series circuit containing capacitive reactance the equation for finding the impedance becomes:

$$Z = \sqrt{R^2 + X_C^2}$$

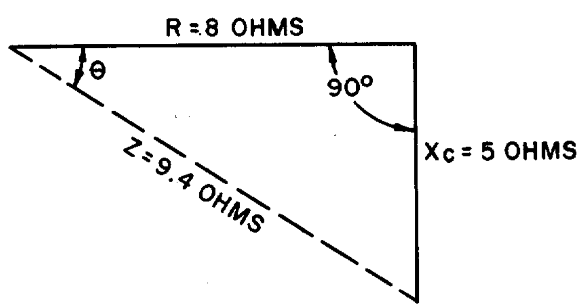


Figure 4-6.—Vector diagram showing relationship of resistance, capacitive reactance, and impedance in a series circuit.

In many series circuits you will find resistance combined with both inductive reactance and capacitive reactance. Since you know that the value of the reactance, X, is equal to the difference between the values of the inductive reactance, X_L, and the capacitive reactance, X_C, the equation for the impedance in a series circuit containing R, X_L, and X_C then becomes:

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

or,

$$Z = \sqrt{R^2 + X^2}$$

(Note: The formulas $Z = \sqrt{R^2 + X_L^2}$, $Z = \sqrt{R^2 + X_C^2}$, and $Z = \sqrt{R^2 + X^2}$ can be used to calculate Z only if the resistance and reactance are connected in series.)

In figure 4-7 you will see the method which may be used to determine the impedance in a series circuit consisting of resistance, inductance, and capacitance.

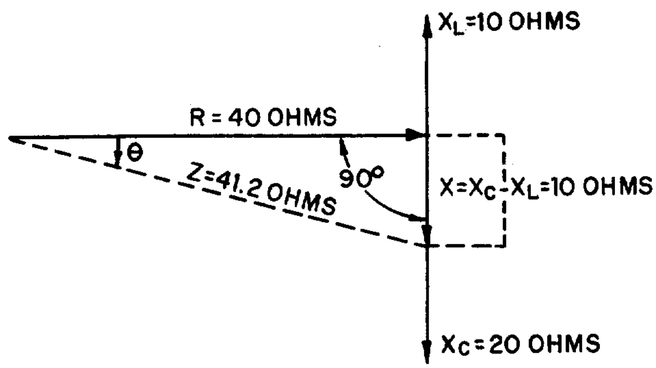


Figure 4-7.—Vector diagram showing relationship of resistance, reactance (capacitive and inductive), and impedance in a series circuit.

Assume that 10 ohms inductive reactance and 20 ohms capacitive reactance are connected in series with 40 ohms resistance. Let the horizontal line represent the resistance R. The line drawn upward from the end of R, represents the inductive reactance, X_L. Represent the capacitive reactance by a line drawn downward at right angles from the same end of R. The resultant of X_L and X_C is found by subtracting X_L from X_C. This resultant represents the value of X.

Thus:

$$X = X_C - X_L$$
$$X = 10 \text{ ohms}$$

The line, Z, will then represent the resultant of R and X. The value of Z can be calculated as follows:

Given: $X_L = 10\ \Omega$

$X_C = 20\ \Omega$

$R = 40\ \Omega$

$$
\begin{aligned}
\text{Solution:}\quad & X = X_c - X_L \\
& X = 20\,\Omega - 10\,\Omega \\
& X = 10\,\Omega \\
& Z = \sqrt{R^2 + X^2} \\
& Z = \sqrt{(40\,\Omega)^2 + (10\,\Omega^2)} \\
& Z = \sqrt{1600 + 100\,\Omega} \\
& Z = \sqrt{1700\,\Omega} \\
& Z = 41.2\,\Omega
\end{aligned}
$$

Q15. What term is given to total opposition to ac in a circuit?

Q16. What formula is used to calculate the amount of this opposition in a series circuit?

Q17. What is the value of Z in a series ac circuit where X_L = 6 ohms, X_C = 3 ohms, and R = 4 ohms?

OHMS LAW FOR AC

In general, Ohm's law cannot be applied to alternating-current circuits since it does not consider the reactance which is always present in such circuits. However, by a modification of Ohm's law which does take into consideration the effect of reactance we obtain a general law which is applicable to ac circuits. Because the impedance, Z, represents the combined opposition of all the reactances and resistances, this general law for ac is,

$$I = \frac{E}{Z}$$

this general modification applies to alternating current flowing in any circuit, and any one of the values may be found from the equation if the others are known.

For example, suppose a series circuit contains an inductor having 5 ohms resistance and 25 ohms inductive reactance in series with a capacitor having 15 ohms capacitive reactance. If the voltage is 50 volts, what is the current? This circuit can be drawn as shown in figure 4-8.

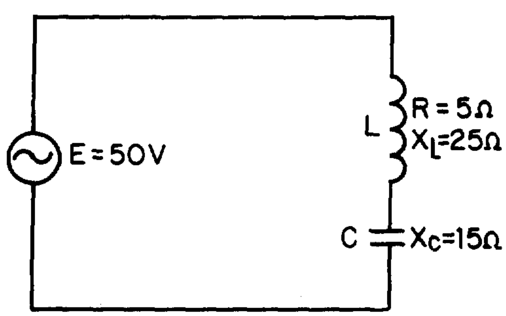


Figure 4-8.—Series LC circuit.

Given: $R = 5\,\Omega$

$$X_L = 25\,\Omega$$
$$X_C = 15\,\Omega$$
$$E = 50\,V$$

Solution:
$$X = X_L - X_C$$
$$X = 25\,\Omega - 15\,\Omega$$
$$X = 10\,\Omega$$
$$Z = \sqrt{R^2 + X^2}$$
$$Z = \sqrt{(5\,\Omega)^2 + (10\,\Omega)^2}$$
$$Z = \sqrt{25 + 100\,\Omega}$$
$$Z = \sqrt{125\,\Omega}$$
$$Z = 11.2\,\Omega$$
$$I = \frac{E}{Z}$$
$$I = \frac{50\,V}{11.2\,\Omega}$$
$$I = 4.46\,A$$

Now suppose the circuit contains an inductor having 5 ohms resistance and 15 ohms inductive reactance in series with a capacitor having 10 ohms capacitive reactance. If the current is 5 amperes, what is the voltage?

Given: $R = 5\,\Omega$

$X_L = 15\,\Omega$

$X_C = 10\,\Omega$

$I = 5A$

Solution: $X = X_L - X_C$

$X = 15\,\Omega - 10\,\Omega$

$X = 5\,\Omega$

$Z = \sqrt{R^2 + X^2}$

$Z = \sqrt{(5\,\Omega)^2 + (5\,\Omega)^2}$

$Z = \sqrt{25 + 25\,\Omega}$

$Z = \sqrt{50\,\Omega}$

$Z = 7.07\,\Omega$

$E = IZ$

$E = 5A \times 7.07\,\Omega$

$E = 35.35V$

Q18. What are the Ohm's law formulas used in an ac circuit to determine voltage and current?

POWER IN AC CIRCUITS

You know that in a direct current circuit the power is equal to the voltage times the current, or $P = E \times I$. If a voltage of 100 volts applied to a circuit produces a current of 10 amperes, the power is 1000 watts. This is also true in an ac circuit when the current and voltage are in phase; that is, when the circuit is effectively resistive. But, if the ac circuit contains reactance, the current will lead or lag the voltage by a certain amount (the phase angle). When the current is out of phase with the voltage, the power indicated by the product of the applied voltage and the total current gives only what is known as the APPARENT POWER. The TRUE POWER depends upon the phase angle between the current and voltage. The symbol for phase angle is θ (Theta).

When an alternating voltage is impressed across a capacitor, power is taken from the source and stored in the capacitor as the voltage increases from zero to its maximum value. Then, as the impressed voltage decreases from its maximum value to zero, the capacitor discharges and returns the power to the source. Likewise, as the current through an inductor increases from its zero value to its maximum value the field around the inductor builds up to a maximum, and when the current decreases from maximum to zero the field collapses and returns the power to the source. You can see therefore that no power is used up in either case, since the power alternately flows to and from the source. This power that is returned to the source by the reactive components in the circuit is called REACTIVE POWER.

In a purely resistive circuit <u>all of the power is consumed and none is returned to the source</u>; in a purely reactive circuit <u>no power is consumed and all of the power is returned to the source</u>. It follows that in a circuit which contains both resistance and reactance there must be some power dissipated in the resistance as well as some returned to the source by the reactance. In figure 4-9 you can see the relationship between the voltage, the current, and the power in such a circuit. The part of the power curve which is shown below the horizontal reference line is the result of multiplying a positive instantaneous value of current by a negative instantaneous value of the voltage, or vice versa. As you know, the product obtained by multiplying a positive value by a negative value will be negative. Therefore the power at that instant must be considered as negative power. In other words, during this time the reactance was returning power to the source.

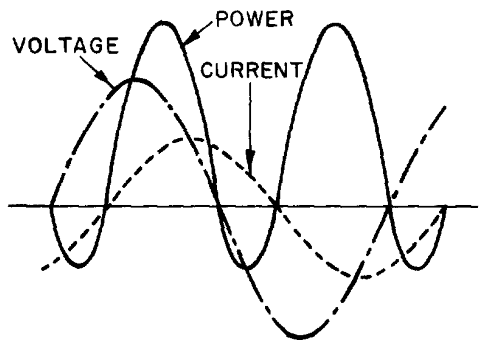


Figure 4-9.—Instantaneous power when current and voltage are out of phase.

The instantaneous power in the circuit is equal to the product of the applied voltage and current through the circuit. When the voltage and current are of the same polarity they are acting together and taking power from the source. When the polarities are unlike they are acting in opposition and power is being returned to the source. Briefly then, in an ac circuit which contains reactance as well as resistance, the apparent power is reduced by the power returned to the source, so that in such a circuit the net power, or true power, is always less than the apparent power.

Calculating True Power in AC Circuits

As mentioned before, the true power of a circuit is the power actually used in the circuit. This power, measured in watts, is the power associated with the total resistance in the circuit. To calculate true power, the voltage and current associated with the resistance must be used. Since the voltage drop across the resistance is equal to the resistance multiplied by the current through the resistance, true power can be calculated by the formula:

$$\text{True Power} = (I_R)^2 R$$

Where: True Power is measured in watts
I_R is resistive current in amperes
R is resistance in ohms

For example, find the true power of the circuit shown in figure 4-10.

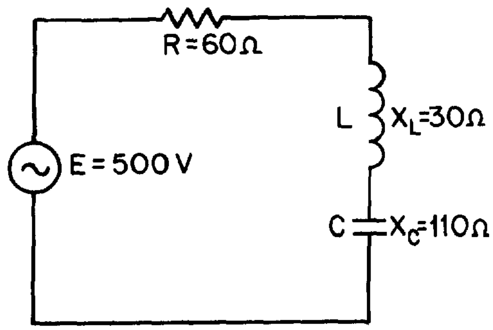


Figure 4-10.—Example circuit for determining power.

Given:

$$R = 60\ \Omega$$
$$X_L = 30\ \Omega$$
$$X_C = 110\ \Omega$$
$$E = 500\ V$$

Solution:

$$X = X_C - X_L$$
$$X = 110\Omega - 30\Omega$$
$$X = 80\ \Omega$$
$$Z = \sqrt{R^2 + X^2}$$
$$Z = \sqrt{(60\ \Omega)^2 + (80\ \Omega)^2}$$
$$Z = \sqrt{3600 + 6400\ \Omega}$$
$$Z = \sqrt{10{,}000\ \Omega}$$
$$Z = 100\ \Omega$$
$$I = \frac{E}{Z}$$
$$I = \frac{500\ V}{100\ \Omega}$$
$$I = 5\ A$$

Since the current in a series circuit is the same in all parts of the circuit:

$$\text{TruePower} = (I_R)^2 R$$
$$\text{TruePower} = (5\ A)^2 \times 60\ \Omega$$
$$\text{True Power} = 1500\ \text{watts}$$

Q19. What is the true power in an ac circuit?

Q20. What is the unit of measurement of true power?

Q21. What is the formula for calculating true power?

Calculating Reactive Power in AC Circuits

The reactive power is the power returned to the source by the reactive components of the circuit. This type of power is measured in Volt-Amperes-Reactive, abbreviated var.

Reactive power is calculated by using the voltage and current associated with the circuit reactance.

Since the voltage of the reactance is equal to the reactance multiplied by the reactive current, reactive power can be calculated by the formula:

$$\text{Reactive Power} = (I_X)^2 X$$

Where: Reactive power is measured in volt-amperes-reactive.

I_X is reactive current in amperes.

X is total reactance in ohms.

Another way to calculate reactive power is to calculate the inductive power and capacitive power and subtract the smaller from the larger.

$$\text{Reactive Power} = (I_L)^2 X_L - (I_C)^2 X_C$$

or

$$(I_C)^2 X_C - (I_L)^2 X_L$$

Where: Reactive power is measured in volt-amperes-reactive.

I_C is capacitive current in amperes.

X_C is capacitive reactance in ohms.

I_L is inductive current in amperes.

X_L is inductive reactance in ohms.

Either one of these formulas will work. The formula you use depends upon the values you are given in a circuit.

For example, find the reactive power of the circuit shown in figure 4-10.

$$\text{Given:} \quad X_L = 30\,\Omega$$
$$X_C = 110\,\Omega$$
$$X = 80\,\Omega$$
$$I = 5\,A$$

Since this is a series circuit, current (I) is the same in all parts of the circuit.

$$\text{Solution:} \quad \text{Reactive power} = (I_X)^2 X$$
$$\text{Reactive power} = (5\,A)^2 \times 80\,\Omega$$
$$\text{Reactive power} = 2{,}000\ \text{var}$$

To prove the second formula also works,

$$\text{Reactive power} = (I_C)^2 X_C - (I_L)^2 X_L$$
$$\text{Reactive power} = (5\,A)^2 \times 110\,\Omega - (5\,A)^2 \times 30\,\Omega$$
$$\text{Reactive power} = 2{,}750\ \text{var} - 750\ \text{var}$$
$$\text{Reactive power} = 2000\ \text{var}$$

Q22. What is the reactive power in an ac circuit?

Q23. What is the unit of measurement for reactive power?

Q24. What is the formula for computing reactive power?

Calculating Apparent Power in AC Circuits.

Apparent power is the power that appears to the source because of the circuit impedance. Since the impedance is the total opposition to ac, the apparent power is that power the voltage source "sees." Apparent power is the combination of true power and reactive power. Apparent power is not found by simply adding true power and reactive power just as impedance is not found by adding resistance and reactance.

To calculate apparent power, you may use either of the following formulas:

$$\text{Apparent power} = (I_Z)^2 Z$$

Where: Apparent power is measured in VA (volt-amperes)

I_Z is impedance current in amperes.

Z is impedance in ohms.

or

$$\text{Apparent power} = \sqrt{(\text{True power})^2 + (\text{reactive power})^2}$$

For example, find the apparent power for the circuit shown in figure 4-10.

Given: $Z = 100\ \Omega$

$I = 5\text{ A}$

Recall that current in a series circuit is the same in all parts of the circuit.

Solution:

$$\text{Apparent Power} = (I_Z)^2 Z$$

$$\text{Apparent power} = (5\text{ A})^2 \times 100\ \Omega$$

$$\text{Apparent power} = 2500\text{ VA}$$

or

Given:

$$\text{True power} = 1500\text{ W}$$

$$\text{Reactive power} = 2000\text{ var}$$

$$\text{Apparent power} = \sqrt{(\text{True power})^2 + (\text{reactive power})^2}$$

$$\text{Apparent power} = \sqrt{(1500\text{ W})^2 + (2000\text{ var})^2}$$

$$\text{Apparent power} = \sqrt{625 \times 10^4}\text{ VA}$$

$$\text{Apparent power} = 2500\text{ VA}$$

Q25. What is apparent power?

Q26. What is the unit of measurement for apparent power?

Q27. What is the formula for apparent power?

Power Factor

The POWER FACTOR is a number (represented as a decimal or a percentage) that represents the portion of the apparent power dissipated in a circuit.

If you are familiar with trigonometry, the easiest way to find the power factor is to find the cosine of the phase angle (θ). The cosine of the phase angle is equal to the power factor.

You do not need to use trigonometry to find the power factor. Since the power dissipated in a circuit is true power, then:

$$\text{Apparent Power} \times PF = \text{True Power},$$

$$\text{Therefore,} \qquad PF = \frac{\text{True Power}}{\text{Apparent Power}}$$

If true power and apparent power are known you can use the formula shown above.

Going one step further, another formula for power factor can be developed. By substituting the equations for true power and apparent power in the formula for power factor, you get:

$$PF = \frac{(I_R)^2 R}{(I_Z)^2 Z}$$

Since current in a series circuit is the same in all parts of the circuit, I_R equals I_Z. Therefore, in a series circuit,

$$PF = \frac{R}{Z}$$

For example, to compute the power factor for the series circuit shown in figure 4-10, any of the above methods may be used.

Given:

$$\text{True Power} = 1500V$$
$$\text{Apparent Power} = 2500\ VA$$

Solution:

$$PF = \frac{\text{True Power}}{\text{Apparent Power}}$$
$$PF = \frac{1500\ W}{2500\ VA}$$
$$PF = .6$$

Another method:

$$\text{Given:} \quad R = 60\ \Omega$$
$$Z = 100\ \Omega$$

$$\text{Solution:} \quad PF = \frac{R}{Z}$$
$$PF = \frac{60\ \Omega}{100\ \Omega}$$
$$PF = .6$$

If you are familiar with trigonometry you can use it to solve for angle θ and the power factor by referring to the tables in appendices V and VI.

$$\text{Given:} \quad R = 60\ \Omega$$
$$X = 80\ \Omega$$

$$\text{Solution:} \quad \tan\ \theta = \frac{X}{R}$$
$$\tan\ \theta = \frac{80\ \Omega}{60\ \Omega}$$
$$\tan\ \theta = 1.333$$
$$\theta = 53.1^\circ$$
$$PF = \cos\ \theta$$
$$PF = .6$$

NOTE: As stated earlier the power factor can be expressed as a decimal or percentage. In this example the decimal number .6 could also be expressed as 60%.

Q28. What is the power factor of a circuit?

Q29. What is a general formula used to calculate the power factor of a circuit?

Power Factor Correction

The apparent power in an ac circuit has been described as the power the source "sees". As far as the source is concerned the apparent power is the power that must be provided to the circuit. You also know that the true power is the power actually used in the circuit. The difference between apparent power and true power is wasted because, in reality, only true power is consumed. The ideal situation would be for apparent power and true power to be equal. If this were the case the power factor would be 1 (unity) or 100 percent. There are two ways in which this condition can exist. (1) If the circuit is purely resistive or (2) if the circuit "appears" purely resistive to the source. To make the circuit appear purely resistive there must be no reactance. To have no reactance in the circuit, the inductive reactance (X_L) and capacitive reactance (X_C) must be equal.

$$\text{Remember:} \quad X = X_L - X_c$$

$$\text{Therefore, when} \quad X_L = X_C, \; X = 0$$

The expression "correcting the power factor" refers to reducing the reactance in a circuit.

The ideal situation is to have no reactance in the circuit. This is accomplished by adding capacitive reactance to a circuit which is inductive and inductive reactance to a circuit which is capacitive. For example, the circuit shown in figure 4-10 has a total reactance of 80 ohms capacitive and the power factor was .6 or 60 percent. If 80 ohms of inductive reactance were added to this circuit (by adding another inductor) the circuit would have a total reactance of zero ohms and a power factor of 1 or 100 percent. The apparent and true power of this circuit would then be equal.

Q30. An ac circuit has a total reactance of 10 ohms inductive and a total resistance of 20 ohms. The power factor is .89. What would be necessary to correct the power factor to unity?

SERIES RLC CIRCUITS

The principles and formulas that have been presented in this chapter are used in all ac circuits. The examples given have been series circuits.

This section of the chapter will not present any new material, but will be an example of using all the principles presented so far. You should follow each example problem step by step to see how each formula used depends upon the information determined in earlier steps. When an example calls for solving for square root, you can practice using the square-root table by looking up the values given.

The example series RLC circuit shown in figure 4-11 will be used to solve for X_L, X_C, X, Z, I_T, true power, reactive power, apparent power, and power factor.

The values solved for will be rounded off to the nearest whole number.

First solve for X_L and X_C.

Given: $f = 60\,\text{Hz}$

$L = 27\,\text{mH}$

$C = 380\,\mu\text{F}$

Solution:

$$X_L = 2\pi fl$$

$$X_L = 6.28 \times 60\,\text{Hz} \times 27\,\text{mH}$$

$$X_L = 10\ \Omega$$

$$X_C = \frac{1}{2\pi fc}$$

$$X_C = \frac{1}{6.28 \times 60\,\text{Hz} \times 380\,\mu\text{F}}$$

$$X_C = \frac{1}{0.143}\Omega$$

$$X_C = 7\,\Omega$$

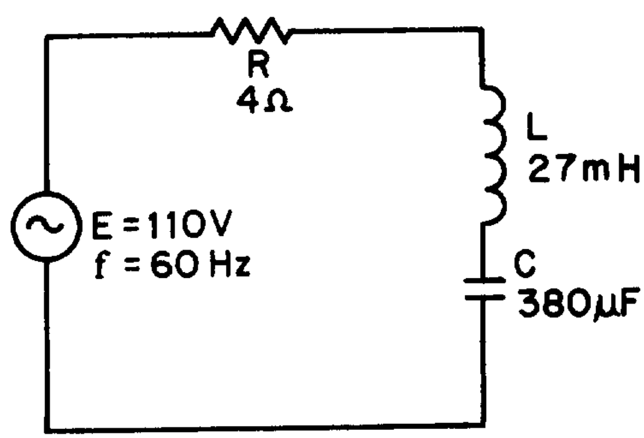


Figure 4-11.—Example series RLC circuit

Now solve for X

Given: $X_C = 7\ \Omega$

$X_L = 10\ \Omega$

Solution:

$$X = X_L - X_C$$

$$X = 10\,\Omega - 7\,\Omega$$

$$X = 3\,\Omega \text{ (Inductive)}$$

Use the value of X to solve for Z.

Given: $X = 3\ \Omega$

$R = 4\ \Omega$

Solution:

$$Z = \sqrt{X^2 + R^2}$$

$$Z = \sqrt{(3\ \Omega)^2 + (4\ \Omega)^2}$$

$$Z = \sqrt{9 + 16\ \Omega}$$

$$Z = \sqrt{25\ \Omega}$$

$$Z = 5\ \Omega$$

This value of Z can be used to solve for total current (I_T).

Given: $Z = 5\ \Omega$

$E = 110\ V$

Solution:

$$I_T = \frac{E}{Z}$$

$$I_T = \frac{110\ V}{5\ \Omega}$$

$$I_T = 22\ A$$

Since current is equal in all parts of a series circuit, the value of I_T can be used to solve for the various values of power.

Given:
$$I_T = 22\ A$$
$$R = 4\Omega$$
$$X = 3\Omega$$
$$Z = 5\Omega$$

Solution:
$$\text{True Power} = (I_R)^2 R$$
$$\text{True Power} = (22A)^2 \times 4\Omega$$
$$\text{True Power} = 1936\ W$$

$$\text{Reactive power} = (I_X)^2 X$$
$$\text{Reactive power} = (22A)^2 \times 3\Omega$$
$$\text{Reactive power} = 1452\ \text{var}$$

$$\text{Apparent power} = (I_Z)^2 Z$$
$$\text{Apparent Power} = (22A)^2 \times 5\Omega$$
$$\text{Apparent Power} = 2420\ VA$$

The power factor can now be found using either apparent power and true power or resistance and impedance. The mathematics in this example is easier if you use impedance and resistance.

Given:
$$R = 4\Omega$$
$$Z = 5\Omega$$

Solution:
$$PF = \frac{R}{Z}$$
$$PF = \frac{4\Omega}{5\Omega}$$
$$PF = .8 \text{ or } 80\%$$

PARALLEL RLC CIRCUITS

When dealing with a parallel ac circuit, you will find that the concepts presented in this chapter for series ac circuits still apply. There is one major difference between a series circuit and a parallel circuit that must be considered. The difference is that current is the same in all parts of a series circuit, whereas voltage is the same across all branches of a parallel circuit. Because of this difference, the total impedance of a parallel circuit must be computed on the basis of the current in the circuit.

You should remember that in the series RLC circuit the following three formulas were used to find reactance, impedance, and power factor:

$$X = X_L - X_C \text{ or } X = X_C - X_L$$

$$Z = \sqrt{(I_R)^2 + X^2}$$

$$PF = \frac{R}{Z}$$

When working with a parallel circuit you must use the following formulas instead:

$$I_X = I_L - I_C \text{ or } I_X = I_C - I_L$$

$$I_Z = \sqrt{(I_R)^2 + (I_X)^2}$$

$$PF = \frac{I_R}{I_Z}$$

(The impedance of a parallel circuit is found by the formula $Z = \frac{E}{I_Z}$)

NOTE: If no value for E is given in a circuit, any value of E can be assumed to find the values of I_L, I_C, I_X, I_R, and I_Z. The same value of voltage is then used to find impedance.

For example, find the value of Z in the circuit shown in figure 4-12.

Given:

$$E = 300\ V$$

$$R = 100\ \Omega$$

$$X_L = 50\ \Omega$$

$$X_C = 150\ \Omega$$

The first step in solving for Z is to calculate the individual branch currents.

Solution: $I_R = \frac{E}{R}$

$$I_R = \frac{300\ V}{100\ \Omega}$$

$$I_R = 3\ A$$

$$I_L = \frac{E}{X_L}$$

$$I_L = \frac{300\ V}{50\ \Omega}$$

$$I_L = 6\ A$$

$$I_C = \frac{E}{X_C}$$

$$I_C = \frac{300\ V}{150\ \Omega}$$

$$I_C = 2\ A$$

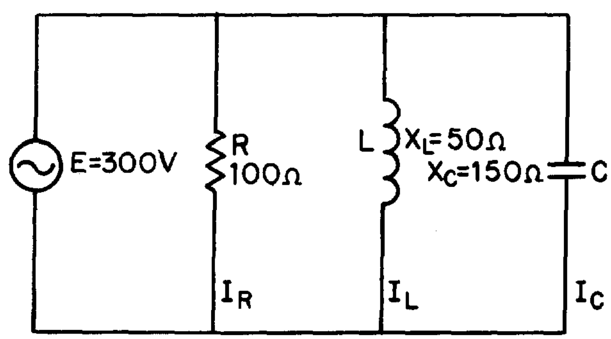


Figure 4-12.—Parallel RLC circuit.

Using the values for I_R, I_L, and I_C, solve for I_X and I_Z.

$$I_X = I_L - I_C$$

$$I_X = 6\ A - 2\ A$$

$$I_X = 4\ A \text{ (inductive)}$$

$$I_Z = \sqrt{(I_R)^2 + (I_X)^2}$$

$$I_Z = \sqrt{(3\ A)^2 + (4\ A)^2}$$

$$I_Z = \sqrt{25\ A}$$

$$I_Z = 5\ A$$

Using this value of I_Z, solve for Z.

$$Z = \frac{E}{I_Z}$$

$$Z = \frac{300\ V}{5\ A}$$

$$Z = 60\ \Omega$$

If the value for E were not given and you were asked to solve for Z, any value of E could be assumed. If, in the example problem above, you assume a value of 50 volts for E, the solution would be:

Given:

$$R = 100\ \Omega$$

$$X_L = 50\ \Omega$$

$$X_C = 150\ \Omega$$

$$E = 50\ V\ \text{(assumed)}$$

First solve for the values of current in the same manner as before.

Solution:

$$I_R = \frac{E}{R}$$

$$I_R = \frac{50\ V}{100\ \Omega}$$

$$I_R = .5\ A$$

$$I_L = \frac{E}{X_L}$$

$$I_L = \frac{50\ V}{50\ \Omega}$$

$$I_L = 1\ A$$

$$I_C = \frac{E}{X_C}$$

$$I_C = \frac{50\ V}{150\ \Omega}$$

$$I_C = .33\ A$$

Solve for I_X and I_Z.

$$I_X = I_L - I_C$$
$$I_X = 1A - .33A$$
$$I_X = .67A \text{ (Inductive)}$$
$$I_Z = \sqrt{(I_R)^2 + (I_X)^2}$$
$$I_Z = \sqrt{(0.5A)^2 + (0.67A)^2}$$
$$I_Z = \sqrt{0.6989A}$$
$$I_Z = 0.836A$$

Solve for Z.

$$Z = \frac{E}{I_Z}$$
$$Z = \frac{50\ V}{.836A}$$
$$Z = 60\ \Omega \text{ (rounded off)}$$

When the voltage is given, you can use the values of currents, I_R, I_X, and I_Z, to calculate for the true power, reactive power, apparent power, and power factor. For the circuit shown in figure 4-12, the calculations would be as follows.

To find true power,

Given: $R = 100\ \Omega$

$I_R = 3A$

Solution:

$$\text{True Power} = (I_R)^2X$$
$$\text{True Power} = (3A)^2 \times 75\ \Omega$$
$$\text{True Power} = 900\ W$$

To find reactive power, first find the value of reactance (X).

Given: $E = 300\text{ V}$

$I_X = 4\text{ A (Inductive)}$

Solution: $X = \dfrac{E}{I_X}$

$$X = \frac{300\text{ V}}{4\text{A}}$$

$$X = 75\ \Omega\text{ (Inductive)}$$

$$\text{Reactive power} = (I_X)^2 X$$

$$\text{Reactive power} = (4\text{A})^2 \times 75\ \Omega$$

$$\text{Reactive power} = 1200\text{ var}$$

To find apparent power,

Given: $Z = 60\ \Omega$

$I_Z = 5\text{A}$

Solution:

$$\text{Apparent Power} = (I_Z)^2 Z$$

$$\text{Apparent Power} = (5\text{A})^2 \times 60\ \Omega$$

$$\text{Apparent Power} = 1500\text{ VA}$$

The power factor in a parallel circuit is found by either of the following methods.

Given:

$$\text{True Power} = 900 \text{ W}$$
$$\text{Apparent Power} = 1500 \text{ VA}$$

Solution:

$$PF = \frac{\text{true power}}{\text{apparent power}}$$
$$PF = \frac{900 \text{ W}}{1500 \text{ VA}}$$
$$PF = .6$$

or

Given:

$$I_R = 3 \text{ A}$$
$$I_Z = 5 \text{ A}$$

Solution:

$$PF = \frac{I_R}{I_Z}$$
$$PF = \frac{3 \text{ A}}{5 \text{ A}}$$
$$PF = .6$$

Q31. What is the difference between calculating impedance in a series ac circuit and in a parallel ac circuit?

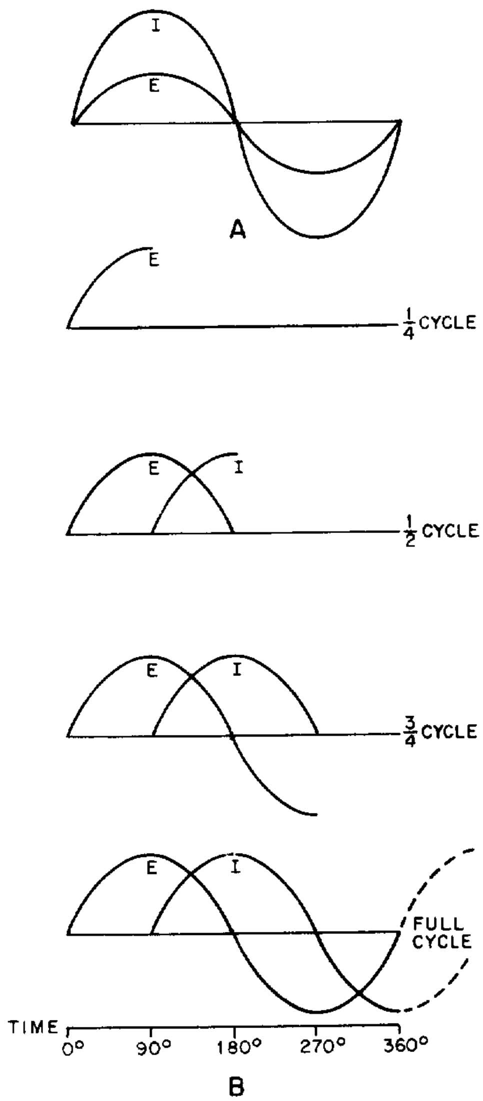
SUMMARY

With the completion of this chapter you now have all the building blocks for electrical circuits. The subjects covered from this point on will be based upon the concepts and relationships that you have learned. The following summary is a brief review of the subjects covered in this chapter.

INDUCTANCE IN AC CIRCUITS—An inductor in an ac circuit opposes any change in current flow just as it does in a dc circuit.

PHASE RELATIONSHIPS OF AN INDUCTOR—The current lags the voltage by 90º in an inductor (ELI).

INDUCTIVE REACTANCE—The opposition an inductor offers to ac is called inductive reactance. It will increase if there is an increase in frequency or an increase in inductance. The symbol is X_L, and the formula is $X_L = 2\pi fL$.

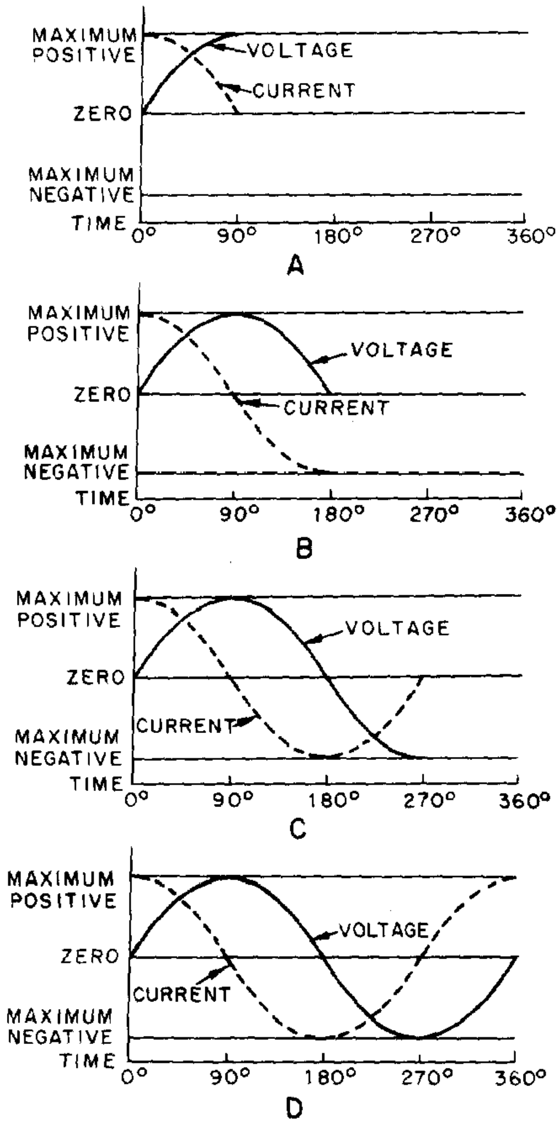


CAPACITANCE IN AC CIRCUITS—A capacitor in an ac circuit opposes any change in voltage just as it does in a dc circuit.

PHASE RELATIONSHIPS OF A CAPACITOR—The current leads the voltage by 90º in a capacitor (ICE).

CAPACITIVE REACTANCE—The opposition a capacitor offers to ac is called capacitive reactance. Capacitive reactance will decrease if there is an increase in frequency or an increase in capacitance. The symbol is X_C and the formula is

$$X_C = \frac{1}{2\pi f C}$$

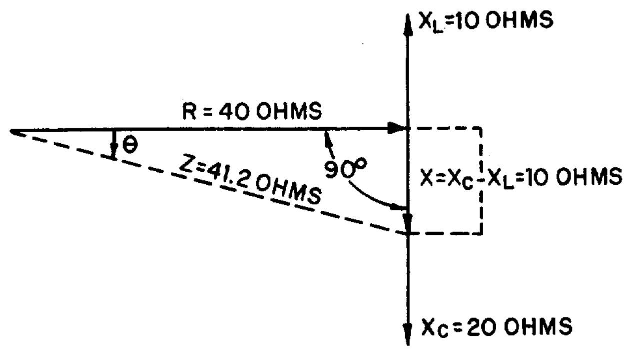


TOTAL REACTANCE—The total reactance of a series ac circuit is determined by the formula $X = X_L - X_C$ or $X = X_C - X_L$. The total reactance in a series circuit is either capacitive or inductive depending upon the largest value of X_C and X_L. In a parallel circuit the reactance is determined by

$$\frac{E}{I_X},$$

where $I_X = I_C - I_L$ or $I_X = I_L - I_C$. The reactance in a parallel circuit is either capacitive or inductive depending upon the largest value of I_L and I_C.

IMPEDANCE – The total opposition to a.c. is called impedance. The symbol is Z. In a series circuit $Z = \sqrt{R^2 + X^2}$. In a parallel circuit $I_Z = \sqrt{(I_R)^2 + (I_X)^2}$ and $Z = \frac{E}{I_Z}$.



PHASE ANGLE—The number of degrees that current leads or lags voltage in an ac circuit is called the phase angle. The symbol is θ.

OHM'S LAW FORMULAS FOR AC—The formulas derived for Ohm's law used in ac are: $E = IZ$ and $I = E/Z$.

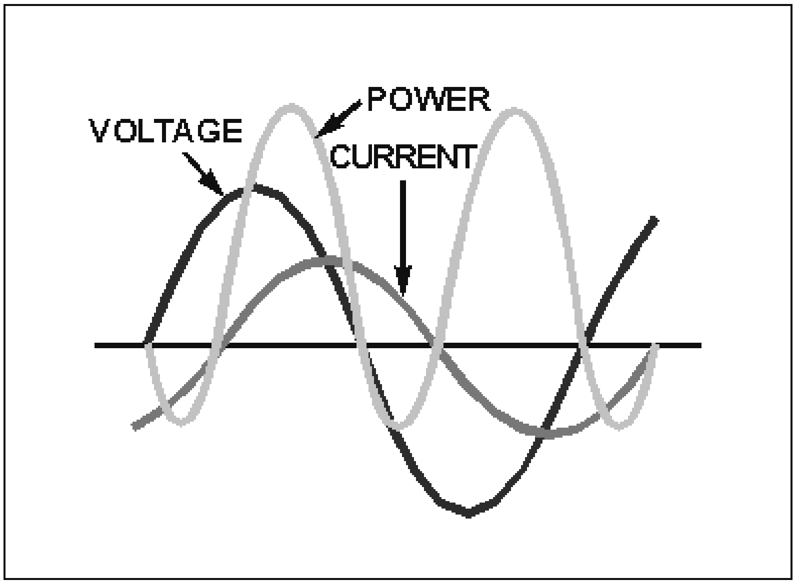
TRUE POWER—The power dissipated across the resistance in an ac circuit is called true power. It is measured in watts and the formula is: True Power = $(I_R)^2R$.

REACTIVE POWER—The power returned to the source by the reactive elements of the circuit is called reactive power. It is measured in volt-amperes reactive (var). The formula is: Reactive Power = $(I_X)^2X$.

APPARENT POWER—The power that appears to the source because of circuit impedance is called apparent power. It is the combination of true power and reactive power and is measured in volt-amperes (VA). The formulas are:

$$\text{Apparent Power} = (I_Z)^2 Z$$

$$\text{Apparent Power} = \sqrt{(\text{true power})^2 + (\text{reactive power})^2}$$



<u>POWER FACTOR</u>—The portion of the <u>apparent power</u> dissipated in a circuit is called the power factor of the circuit. It can be expressed as a decimal or a percentage. The formulas for power

factor are $PF = \frac{\text{true power}}{\text{apparent power}}$ or $PF = \cos\theta$. In a series circuit, $PF = \frac{R}{Z}$. In a parallel circuit, $Pf = \frac{I_R}{I_Z}$.

<u>POWER FACTOR CORRECTION</u>—To reduce losses in a circuit the power factor should be as close to unity or 100% as possible. This is done by adding capacitive reactance to a circuit when the total reactance is inductive. If the total reactance is capacitive, inductive reactance is added in the circuit.

ANSWERS TO QUESTIONS Q1. THROUGH Q31.

A1. An inductor opposes a change in current.

A2. Current lags voltage by 90º (ELI).

A3. Inductive reactance.

A4. $X_L = 2\pi fL$.

A5. X_L *increases.*

A6. X_L *decreases.*

A7. *The capacitor opposes any change in voltage.*

A8. *Current leads voltage by 90º (ICE).*

A9. *Capacitive reactance.*

A10.

$$X = \frac{1}{2\pi fC}$$

A11. *X_C increases.*

A12. *X_C decreases.*

A13. *$X = X_L - X_C$ or $X = X_C - X_L$*

A14. *30 Ω (capacitive).*

A15. *Impedance.*

A16.

$$Z = \sqrt{R^2 + X^2}.$$

A17. *$Z = 5\Omega$.*

A18.

$$E = IZ$$

$$I = \frac{E}{Z}.$$

A19. *True power is the power dissipated in the resistance of the circuit or the power actually used in the circuit.*

A20. *Watt.*

A21. *True Power = $(I_R)^2R$.*

A22. *Reactive power is the power returned to the source by the reactive components of the circuit.*

A23. *var.*

A24.

$$\text{Reactive Power} = (I_X)^2X \text{ or}$$

$$(I_C)^2X_C - (I_L)^2X_L \text{ or}$$

$$(I_L)^2X_L - (I_C)^2X_C.$$

A25. The power that appears to the source because of circuit impedance, or the combination of true power and reactive power.

A26. VA (volt-amperes).

A27.

$$\text{Apparent power} = (I_Z)^2 Z \text{ or}$$

$$\sqrt{(\text{true power})^2 + (\text{reactive power})^2}$$

A28. PF is a number representing the portion of apparent power actually dissipated in a circuit.

A29.

$$PF = \frac{\text{true power}}{\text{apparent power}} \text{ or } PF = \cos\theta.$$

A30. Add 10 ohms of capacitive reactance to the circuit.

A31. In a series circuit impedance is calculated from the values of resistance and reactance. In a parallel circuit, the values of resistive current and reactive current must be used to calculate total current (impedance current) and this value must be divided into the source voltage to calculate the impedance.

CHAPTER 5

TRANSFORMERS

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. State the meaning of "transformer action."
2. State the physical characteristics of a transformer, including the basic parts, common core materials, and main core types.
3. State the names given to the source and load windings of a transformer.
4. State the difference in construction between a high- and a low-voltage transformer.
5. Identify transformer symbols as to the type of transformer each symbol represents and the method used to denote transformer phasing.
6. State the meaning of a "no-load condition" and "exciting current" relative to a transformer.
7. State what causes voltage to be developed across the secondary of a transformer and the effect of cemf in a transformer.
8. State the meaning of leakage flux and its effect on the coefficient of coupling.
9. Identify a transformer as step up or step down and state the current ratio of a transformer when given the turns ratio.
10. Solve for primary voltage, secondary voltage, primary current and number of turns in the secondary given various transformer values.
11. State the mathematical relationship between the power in the primary and the power in the secondary of a transformer and compute efficiency of a transformer.
12. State the three power losses in a transformer.
13. State the reason a transformer should not be operated at a lower frequency than that specified for the transformer.
14. List five different types of transformers according to their applications.
15. State the standard color coding for a power transformer.
16. State the general safety precautions you should observe when working with transformers and other electrical components.

TRANSFORMERS

The information in this chapter is on the construction, theory, operation, and the various uses of transformers. Safety precautions to be observed by a person working with transformers are also discussed.

A TRANSFORMER is a device that transfers electrical energy from one circuit to another by electromagnetic induction (transformer action). The electrical energy is always transferred without a change in frequency, but may involve changes in magnitudes of voltage and current. Because a transformer works on the principle of electromagnetic induction, it must be used with an input source voltage that varies in amplitude. There are many types of power that fit this description; for ease of explanation and understanding, transformer action will be explained using an ac voltage as the input source.

In a preceding chapter you learned that alternating current has certain advantages over direct current. One important advantage is that when ac is used, the voltage and current levels can be increased or decreased by means of a transformer.

As you know, the amount of power used by the load of an electrical circuit is equal to the current in the load times the voltage across the load, or P = EI. If, for example, the load in an electrical circuit requires an input of 2 amperes at 10 volts (20 watts) and the source is capable of delivering only 1 ampere at 20 volts, the circuit could not normally be used with this particular source. However, if a transformer is connected between the source and the load, the voltage can be decreased (stepped down) to 10 volts and the current increased (stepped up) to 2 amperes. Notice in the above case that the power remains the same. That is, 20 volts times 1 ampere equals the same power as 10 volts times 2 amperes.

Q1. What is meant by "transformer action?"

BASIC OPERATION OF A TRANSFORMER

In its most basic form a transformer consists of:

- A primary coil or winding.
- A secondary coil or winding.
- A core that supports the coils or windings.

Refer to the transformer circuit in figure 5-1 as you read the following explanation: The primary winding is connected to a 60 hertz ac voltage source. The magnetic field (flux) builds up (expands) and collapses (contracts) about the primary winding. The expanding and contracting magnetic field around the primary winding cuts the secondary winding and induces an alternating voltage into the winding. This voltage causes alternating current to flow through the load. The voltage may be stepped up or down depending on the design of the primary and secondary windings.

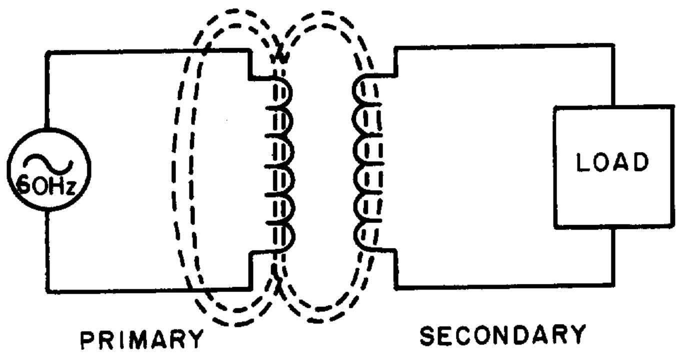


Figure 5-1.—Basic transformer action.

Q2. What are, the three basic parts of a transformer?

THE COMPONENTS OF A TRANSFORMER

Two coils of wire (called windings) are wound on some type of core material. In some cases the coils of wire are wound on a cylindrical or rectangular cardboard form. In effect, the core material is air and the transformer is called an AIR-CORE TRANSFORMER. Transformers used at low frequencies, such as 60 hertz and 400 hertz, require a core of low-reluctance magnetic material, usually iron. This type of transformer is called an IRON-CORE TRANSFORMER. Most power transformers are of the iron-core type. The principle parts of a transformer and their functions are:

- The CORE, which provides a path for the magnetic lines of flux.
- The PRIMARY WINDING, which receives energy from the ac source.
- The SECONDARY WINDING, which receives energy from the primary winding and delivers it to the load.
- The ENCLOSURE, which protects the above components from dirt, moisture, and mechanical damage.

CORE CHARACTERISTICS

The composition of a transformer core depends on such factors as voltage, current, and frequency. Size limitations and construction costs are also factors to be considered. Commonly used core materials are air, soft iron, and steel. Each of these materials is suitable for particular applications and unsuitable for others. Generally, air-core transformers are used when the voltage source has a high frequency (above 20 kHz). Iron-core transformers are usually used when the source frequency is low (below 20 kHz). A soft-iron-core transformer is very useful where the transformer must be physically small, yet efficient. The iron-core transformer provides better power transfer than does the air-core transformer. A transformer whose core is constructed of laminated sheets of steel dissipates heat readily; thus it provides for the efficient transfer of power. The majority of transformers you will encounter in Navy equipment contain laminated-steel cores. These steel laminations (see figure 5-2) are insulated with a nonconducting material, such as varnish, and then formed into a core. It takes about 50 such laminations to make a core an inch thick. The purpose of the laminations is to reduce certain losses which will be discussed later in this chapter. An important point to remember is that the most efficient transformer core is one that offers the best path for the most lines of flux with the least loss in magnetic and electrical energy.

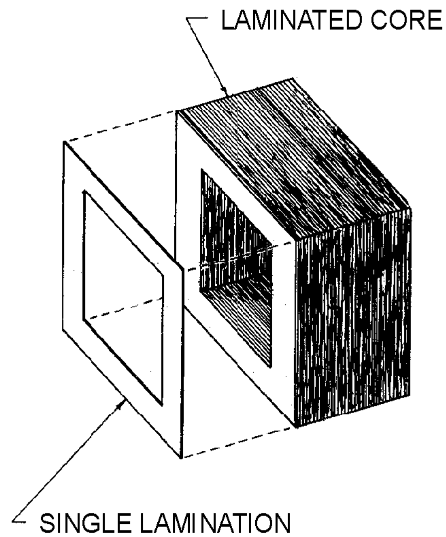


Figure 5-2.—Hollow-core construction.

Q3. What are three materials commonly used as the core of a transformer?

Hollow-Core Transformers

There are two main shapes of cores used in laminated-steel-core transformers. One is the HOLLOW-CORE, so named because the core is shaped with a hollow square through the center. Figure 5-2 illustrates this shape of core. Notice that the core is made up of many laminations of steel. Figure 5-3 illustrates how the transformer windings are wrapped around both sides of the core.

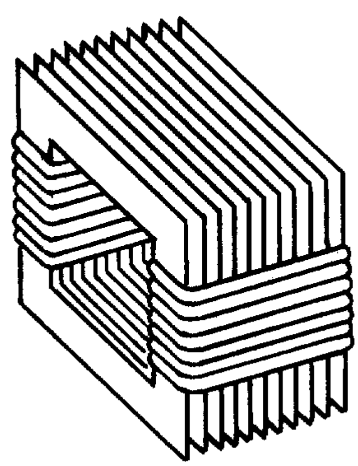


Figure 5-3.—Windings wrapped around laminations.

Shell-Core Transformers

The most popular and efficient transformer core is the SHELL CORE, as illustrated in figure 5-4. As shown, each layer of the core consists of E- and I-shaped sections of metal. These sections are butted together to form the laminations. The laminations are insulated from each other and then pressed together to form the core.

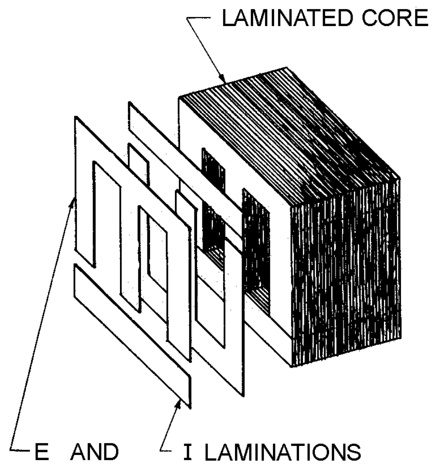


Figure 5-4.—Shell-type core construction.

Q4. What are the two main types of cores used in transformers?

TRANSFORMER WINDINGS

As stated above, the transformer consists of two coils called WINDINGS which are wrapped around a core. The transformer operates when a source of ac voltage is connected to one of the windings and a load device is connected to the other. The winding that is connected to the source is called the PRIMARY WINDING. The winding that is connected to the load is called the SECONDARY WINDING. (Note: In this chapter the terms "primary winding" and "primary" are used interchangeably; the term: "secondary winding" and "secondary" are also used interchangeably.)

Figure 5-5 shows an exploded view of a shell-type transformer. The primary is wound in layers directly on a rectangular cardboard form.

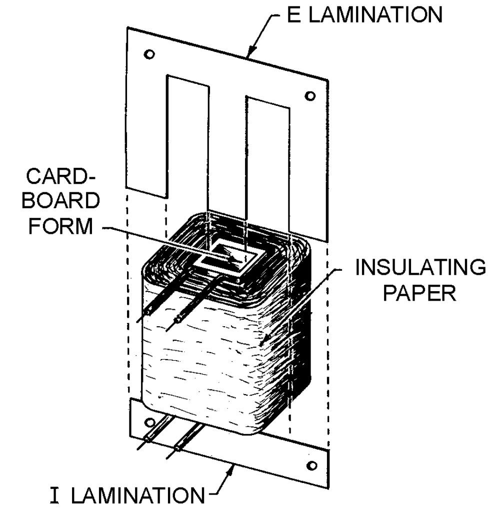


Figure 5-5.—Exploded view of shell-type transformer construction.

In the transformer shown in the cutaway view in figure 5-6, the primary consists of many turns of relatively small wire. The wire is coated with varnish so that each turn of the winding is insulated from every other turn. In a transformer designed for high-voltage applications, sheets of insulating material, such as paper, are placed between the layers of windings to provide additional insulation.

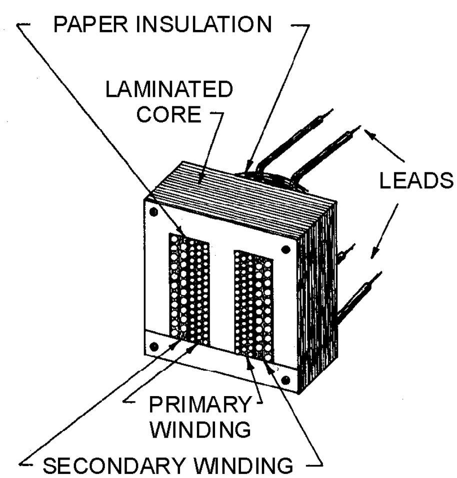


Figure 5-6.—Cutaway view of shell-type core with windings.

When the primary winding is completely wound, it is wrapped in insulating paper or cloth. The secondary winding is then wound on top of the primary winding. After the secondary winding is complete, it too is covered with insulating paper. Next, the E and I sections of the iron core are inserted into and around the windings as shown.

The leads from the windings are normally brought out through a hole in the enclosure of the transformer. Sometimes, terminals may be provided on the enclosure for connections to the windings. The figure shows four leads, two from the primary and two from the secondary. These leads are to be connected to the source and load, respectively.

Q5. Which transformer windings are connected to an ac source voltage and to a load, respectively?

Q6. A transformer designed for high-voltage applications differs in construction in what way from a transformer designed for low-voltage applications?

SCHEMATIC SYMBOLS FOR TRANSFORMERS

Figure 5-7 shows typical schematic symbols for transformers. The symbol for an air-core transformer is shown in figure 5-7(A). Parts (B) and (C) show iron-core transformers. The bars between the coils are used to indicate an iron core. Frequently, additional connections are made to the transformer windings at points other than the ends of the windings. These additional connections are called TAPS. When a tap is connected to the center of the winding, it is called a CENTER TAP. Figure 5-7(C) shows the schematic representation of a center-tapped iron-core transformer.

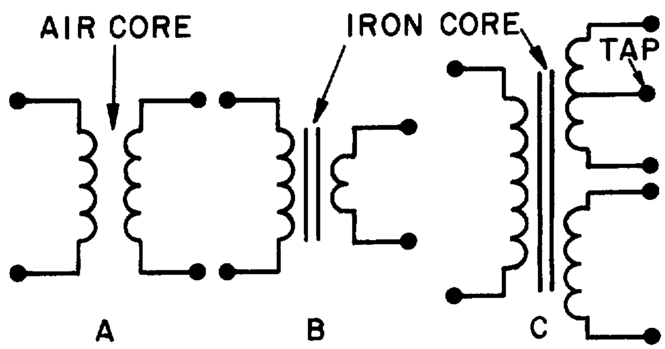
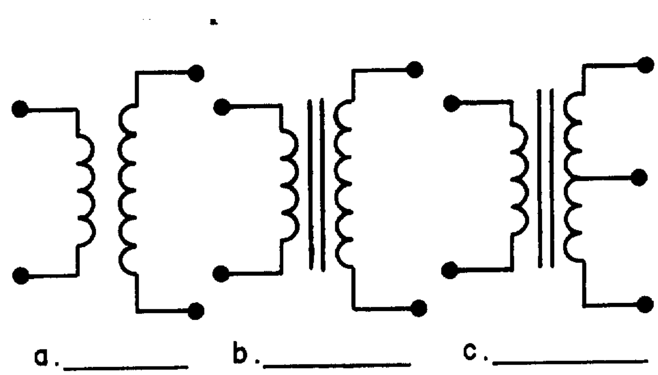


Figure 5-7.—Schematic symbols for various types of transformers.

Q7. Identify the below schematic symbols of transformers by labeling them in the blanks provided.



HOW A TRANSFORMER WORKS

Up to this point the chapter has presented the basics of the transformer including transformer action, the transformer's physical characteristics, and how the transformer is constructed. Now you have the necessary knowledge to proceed into the theory of operation of a transformer.

NO-LOAD CONDITION

You have learned that a transformer is capable of supplying voltages which are usually higher or lower than the source voltage. This is accomplished through mutual induction, which takes place when the changing magnetic field produced by the primary voltage cuts the secondary winding.

A no-load condition is said to exist when a voltage is applied to the primary, but no load is connected to the secondary, as illustrated by figure 5-8. Because of the open switch, there is no current flowing in the secondary winding. With the switch open and an ac voltage applied to the primary, there is, however, a very small amount of current called EXCITING CURRENT flowing in the primary. Essentially, what the exciting current does is "excite" the coil of the primary to create a magnetic field. The amount of exciting current is determined by three factors: (1) the amount of voltage applied (E_a), (2) the resistance (R) of the primary coil's wire and core losses, and (3) the X_L which is dependent on the frequency of the exciting current. These last two factors are controlled by transformer design.

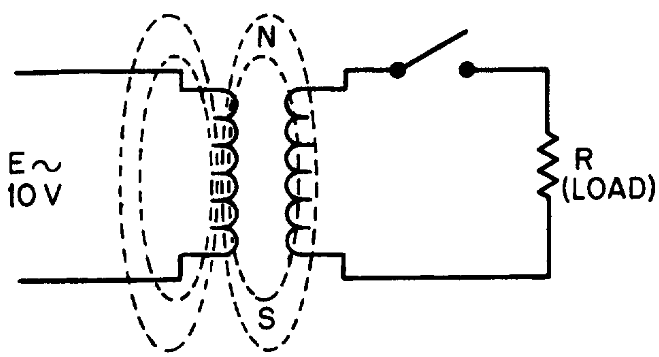


Figure 5-8.—Transformer under no-load conditions.

This very small amount of exciting current serves two functions:

1. Most of the exciting energy is used to maintain the magnetic field of the primary.

2. A small amount of energy is used to overcome the resistance of the wire and core losses which are dissipated in the form of heat (power loss).

Exciting current will flow in the primary winding at all times to maintain this magnetic field, but no transfer of energy will take place as long as the secondary circuit is open.

Q8. What is meant by a "no-load condition" in a transformer circuit?

PRODUCING A COUNTER EMF

When an alternating current flows through a primary winding, a magnetic field is established around the winding. As the lines of flux expand outward, relative motion is present, and a counter emf is induced in the winding. This is the same counter emf that you learned about in the chapter on inductors. Flux leaves the primary at the north pole and enters the primary at the south pole. The counter emf induced in the primary has a polarity that opposes the applied voltage, thus opposing the flow of current in the primary. It is the counter emf that limits exciting current to a very low value.

Q9. What is meant by "exciting current" in a transformer?

INDUCING A VOLTAGE IN THE SECONDARY

To visualize how a voltage is induced into the secondary winding of a transformer, again refer to figure 5-8. As the exciting current flows through the primary, magnetic lines of force are generated. During the time current is increasing in the primary, magnetic lines of force expand outward from the primary and cut the secondary. As you remember, a voltage is induced into a coil when magnetic lines cut across it. Therefore, the voltage across the primary causes a voltage to be induced across the secondary.

Q10. What is the name of the emf generated in the primary that opposes the flow of current in the primary?

Q11. What causes a voltage to be developed across the secondary winding of a transformer?

PRIMARY AND SECONDARY PHASE RELATIONSHIP

The secondary voltage of a simple transformer may be either in phase or out of phase with the primary voltage. This depends on the direction in which the windings are wound and the arrangement of the connections to the external circuit (load). Simply, this means that the two voltages may rise and fall together or one may rise while the other is falling.

Transformers in which the secondary voltage is in phase with the primary are referred to as LIKE-WOUND transformers, while those in which the voltages are 180 degrees out of phase are called UNLIKE-WOUND transformers.

Dots are used to indicate points on a transformer schematic symbol that have the same instantaneous polarity (points that are in phase).

The use of phase-indicating dots is illustrated in figure 5-9. In part (A) of the figure, both the primary and secondary windings are wound from top to bottom in a clockwise direction, as viewed from above the windings. When constructed in this manner, the top lead of the primary and the top lead of the secondary have the SAME polarity. This is indicated by the dots on the transformer symbol. A lack of phasing dots indicates a reversal of polarity.

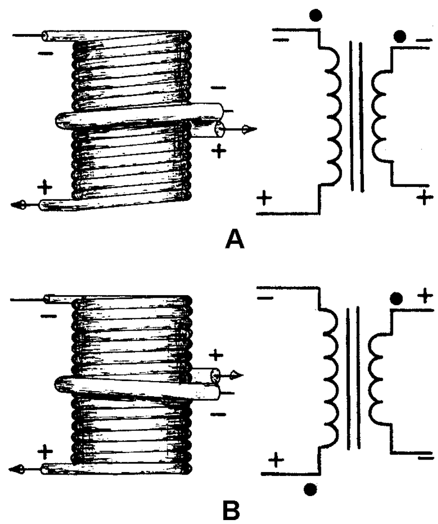
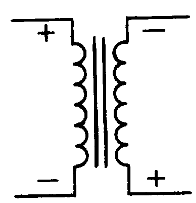


Figure 5-9.—Instantaneous polarity depends on direction of winding.

Part (B) of the figure illustrates a transformer in which the primary and secondary are wound in opposite directions. As viewed from above the windings, the primary is wound in a clockwise direction from top to bottom, while the secondary is wound in a counterclockwise direction. Notice that the top leads of the primary and secondary have OPPOSITE polarities. This is indicated by the dots being placed on opposite ends of the transformer symbol. Thus, the polarity of the voltage at the terminals of the secondary of a transformer depends on the direction in which the secondary is wound with respect to the primary.

Q12. What is the phase relationship between the voltage induced in the secondary of an unlike-wound transformer and the counter emf of the primary winding?

Q13. Draw dots on the below symbol to indicate the phasing of the transformer.



COEFFICIENT OF COUPLING

The COEFFICIENT OF COUPLING of a transformer is dependent on the portion of the total flux lines that cuts both primary and secondary windings. Ideally, all the flux lines generated by the primary should cut the secondary, and all the lines of the flux generated by the secondary should cut the primary. The coefficient of coupling would then be one (unity), and maximum energy would be transferred from the primary to the secondary. Practical power transformers use high-permeability silicon steel cores and close spacing between the windings to provide a high coefficient of coupling.

Lines of flux generated by one winding which do not link with the other winding are called LEAKAGE FLUX. Since leakage flux generated by the primary does not cut the secondary, it cannot induce a voltage into the secondary. The voltage induced into the secondary is therefore less than it would be if the leakage flux did not exist. Since the effect of leakage flux is to lower the voltage induced into the secondary, the effect can be duplicated by assuming an inductor to be connected in series with the primary. This series LEAKAGE INDUCTANCE is assumed to drop part of the applied voltage, leaving less voltage across the primary.

Q14. What is "leakage flux?"

Q15. What effect does flux leakage in a transformer have on the coefficient of coupling (K) in the transformer?

TURNS AND VOLTAGE RATIOS

The total voltage induced into the secondary winding of a transformer is determined mainly by the RATIO of the number of turns in the primary to the number of turns in the secondary, and by the amount of voltage applied to the primary. Refer to figure 5-10. Part (A) of the figure shows a transformer whose primary consists of ten turns of wire and whose secondary consists of a single turn of wire. You know that as lines of flux generated by the primary expand and collapse, they cut BOTH the ten turns of the primary and the single turn of the secondary. Since the length of the wire in the secondary is approximately the same as the length of the wire in each turn in the primary, EMF INDUCED INTO THE SECONDARY WILL BE THE SAME AS THE EMF INDUCED INTO EACH TURN IN THE PRIMARY. This means that if the voltage applied to the primary winding is 10 volts, the counter emf in the primary is almost 10 volts. Thus, each turn in the primary will have an induced counter emf of approximately one-tenth of the total applied voltage, or one volt. Since the same flux lines cut the turns in both the secondary and the primary, each turn will have an emf of one volt induced into it. The transformer in part (A) of figure 5-10 has only one turn in the secondary, thus, the emf across the secondary is one volt.

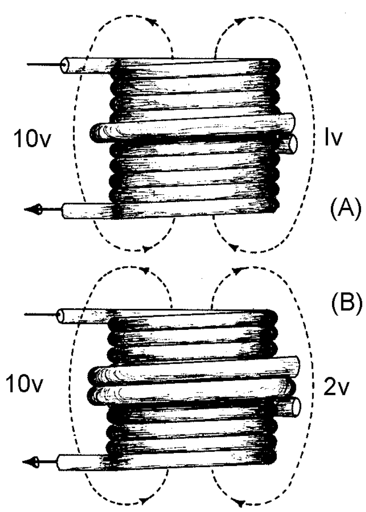


Figure 5-10.—Transformer turns and voltage ratios.

The transformer represented in part (B) of figure 5-10 has a ten-turn primary and a two-turn secondary. Since the flux induces one volt per turn, the total voltage across the secondary is two volts. Notice that the volts per turn are the same for both primary and secondary windings. Since the counter emf in the primary is equal (or almost) to the applied voltage, a proportion may be set up to express the value of the voltage induced in terms of the voltage applied to the primary and the number of turns in each winding. This proportion also shows the relationship between the number of turns in each winding and the voltage across each winding. This proportion is expressed by the equation:

$$\frac{E_S}{E_P} = \frac{N_S}{N_P}$$

Where:

N_P = number of turns in the primary

E_P = voltage applied to the primary

E_S = voltage induced in the secondary

N_S = number of turns in the secondary

Notice the equation shows that the ratio of secondary voltage to primary voltage is equal to the ratio of secondary turns to primary turns. The equation can be written as:

$$E_P N_S = E_S N_P$$

The following formulas are derived from the above equation:

$$\text{Transposing for } E_S: \quad E_S = \frac{E_P N_S}{N_P}$$

$$\text{Transposing for } E_P: \quad E_P = \frac{E_S N_P}{N_S}$$

If any three of the quantities in the above formulas are known, the fourth quantity can be calculated. Example. A transformer has 200 turns in the primary, 50 turns in the secondary, and 120 volts applied to the primary (E_p). What is the voltage across the secondary (E_S)?

Given:

$N_P = 200$ turns

$N_S = 50$ turns

$E_P = 120$ volts

$E_S = ?$ volts

Solution: $$E_S = \frac{E_P N_S}{N_P}$$

Substitution: $$E_S = \frac{120 \text{ volts} \times 50 \text{ turns}}{200 \text{ turns}}$$

$$E_S = 30 \text{ volts}$$

Example. There are 400 turns of wire in an iron-core coil. If this coil is to be used as the primary of a transformer, how many turns must be wound on the coil to form the secondary winding of the transformer to have a secondary voltage of one volt if the primary voltage is five volts?

$$\text{Given:} \quad N_P = 400 \text{ turns}$$
$$E_P = 5 \text{ volts}$$
$$E_S = 1 \text{ volt}$$
$$N_S = ? \text{ turns}$$

$$\text{Solution:} \quad E_P N_S = E_S N_P$$

Transposing for N_S:

$$N_S = \frac{E_S N_P}{E_P}$$

$$\text{Substitution:} \quad N_S = \frac{1 \text{ volt} \times 400 \text{ turns}}{5 \text{ volts}}$$
$$N_S = 80 \text{ turns}$$

Note: The ratio of the voltage (5:1) is equal to the turns ratio (400:80). Sometimes, instead of specific values, you are given a turns or voltage ratio. In this case, you may assume any value for one of the voltages (or turns) and compute the other value from the ratio. For example, if a turn ratio is given as 6:1, you can assume a number of turns for the primary and compute the secondary number of turns (60:10, 36:6, 30:5, etc.).

The transformer in each of the above problems has fewer turns in the secondary than in the primary. As a result, there is less voltage across the secondary than across the primary. A transformer in which the voltage across the secondary is less than the voltage across the primary is called a STEP-DOWN transformer. The ratio of a four-to-one step-down transformer is written as 4:1. A transformer that has fewer turns in the primary than in the secondary will produce a greater voltage across the secondary than the voltage applied to the primary. A transformer in which the voltage across the secondary is greater than the voltage applied to the primary is called a STEP-UP transformer. The ratio of a one-to-four step-up transformer should be written as 1:4. Notice in the two ratios that the value of the primary winding is always stated first.

Q16. Does 1:5 indicate a step-up or step-down transformer?

Q17. A transformer has 500 turns on the primary and 1500 turns on the secondary. If 45 volts are applied to the primary, what is the voltage developed across the secondary? (Assume no losses)

Q18. A transformer has a turns ratio of 7:1. If 5 volts is developed across the secondary, what is the voltage applied to the primary? (Note: E_S is given, what is E_P ?)

Q19. A transformer has 60 volts applied to its primary and 420 volts appearing across its secondary. If there are 800 turns on the primary, what is the number of turns in the secondary?

EFFECT OF A LOAD

When a load device is connected across the secondary winding of a transformer, current flows through the secondary and the load. The magnetic field produced by the current in the secondary interacts with the magnetic field produced by the current in the primary. This interaction results from the mutual inductance between the primary and secondary windings.

MUTUAL FLUX

The total flux in the core of the transformer is common to both the primary and secondary windings. It is also the means by which energy is transferred from the primary winding to the secondary winding. Since this flux links both windings, it is called MUTUAL FLUX. The inductance which produces this flux is also common to both windings and is called mutual inductance.

Figure 5-11 shows the flux produced by the currents in the primary and secondary windings of a transformer when source current is flowing in the primary winding.

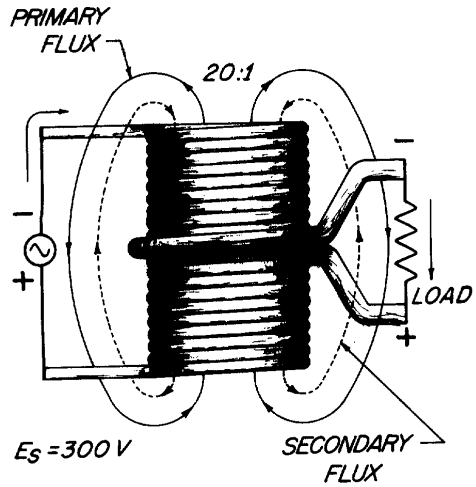


Figure 5-11.—Simple transformer indicating primary- and secondary-winding flux relationship.

When a load resistance is connected to the secondary winding, the voltage induced into the secondary winding causes current to flow in the secondary winding. This current produces a flux field about the secondary (shown as broken lines) which is in opposition to the flux field about the primary (Lenz's law). Thus, the flux about the secondary cancels some of the flux about the primary. With less flux surrounding the primary, the counter emf is reduced and more current is drawn from the source. The additional current in the primary generates more lines of flux, nearly reestablishing the original number of total flux lines.

TURNS AND CURRENT RATIOS

The number of flux lines developed in a core is proportional to the magnetizing force (IN AMPERE-TURNS) of the primary and secondary windings. The ampere-turn ($I \times N$) is a measure of magnetomotive force; it is defined as the magnetomotive force developed by one ampere of current flowing in a coil of one turn. The flux which exists in the core of a transformer surrounds both the primary and secondary windings. Since the flux is the same for both windings, the ampere-turns in both the primary and secondary windings must be the same.

Therefore:

$$I_P N_P = I_S N_S$$

Where:

$I_P N_P$ = ampere – turns in the primary winding

$I_S N_S$ = ampere – turns in the secondary winding

By dividing both sides of the equation by $I_P N_S$, you obtain:

$$\frac{N_P}{N_S} = \frac{I_S}{I_P}$$

Since: $$\frac{E_S}{E_P} = \frac{N_S}{N_P}$$

Then: $$\frac{E_P}{E_S} = \frac{N_P}{N_S}$$

And: $$\frac{E_P}{E_S} = \frac{I_S}{I_P}$$

Where:

E_P = voltage applied to the primary in volts

E_S = voltage across the secondary in volts

I_P = current in the primary in amperes

I_S = current in the secondary in amperes

Notice the equations show the current ratio to be the inverse of the turns ratio and the voltage ratio. This means, a transformer having less turns in the secondary than in the primary would step down the voltage, but would step up the current. Example: A transformer has a 6:1 voltage ratio. Find the current in the secondary if the current in the primary is 200 milliamperes.

$$
\begin{aligned}
&\text{Given:} \quad && E_P = 6\,V \text{ (assumed)} \\
& && E_S = 1\,V \\
& && I_P = 200\,mA \text{ or } 0.2A \\
& && I_S = ?
\end{aligned}
$$

$$\text{Solution:} \quad \frac{E_P}{E_S} = \frac{I_S}{I_P}$$

Transposing for I_S:

$$I_S = \frac{E_P I_P}{E_S}$$

Substitution:

$$I_S = \frac{6\,V \times 0.2A}{1V}$$

$$I_S = 1.2A$$

The above example points out that although the voltage across the secondary is one-sixth the voltage across the primary, the current in the secondary is six times the current in the primary.

The above equations can be looked at from another point of view. The expression

$$\frac{N_p}{N_s}$$

is called the transformer TURNS RATIO and may be expressed as a single factor. Remember, the turns ratio indicates the amount by which the transformer increases or decreases the voltage applied to the primary. For example, if the <u>secondary</u> of a transformer has two times as many turns as the <u>primary</u>, the voltage induced into the <u>secondary</u> will be <u>two times</u> the voltage across the <u>primary</u>. If the secondary has one-half as many turns as the primary, the voltage across the secondary will be one-half the voltage across the primary. However, the turns ratio and the current ratio of a transformer have an inverse relationship. Thus, a 1:2 step-up transformer will have one-half the current in the secondary as in the primary. A 2:1 step-down transformer will have twice the current in the secondary as in the primary.

Example: A transformer with a turns ratio of 1:12 has 3 amperes of current in the secondary. What is the value of current in the primary?

$$
\begin{aligned}
&\text{Given:} \quad && N_P = 1 \text{ turn (assumed)} \\
& && N_S = 12 \text{ turns} \\
& && I_S = 3A \\
& && L_P = ?
\end{aligned}
$$

$$\text{Solution:} \quad \frac{N_P}{N_S} = \frac{I_S}{I_P}$$

Transposing for I_P:

$$I_P = \frac{N_S I_S}{N_P}$$

Substitution: $$I_P = \frac{12 \text{ turns} \times 3\text{ A}}{1 \text{ turn}}$$

$$I_P = 36\text{ A}$$

Q20. A transformer with a turns ratio of 1:3 has what current ratio?

Q21. A transformer has a turns ratio of 5:1 and a current of 5 amperes flowing in the secondary. What is the current flowing in the primary? (Assume no losses)

POWER RELATIONSHIP BETWEEN PRIMARY AND SECONDARY WINDINGS

As just explained, the turns ratio of a transformer affects current as well as voltage. If voltage is doubled in the secondary, current is halved in the secondary. Conversely, if voltage is halved in the secondary, current is doubled in the secondary. In this manner, all the power delivered to the primary by the source is also delivered to the load by the secondary (minus whatever power is consumed by the transformer in the form of losses). Refer again to the transformer illustrated in figure 5-11. The turns ratio is 20:1. If the input to the primary is 0.1 ampere at 300 volts, the power in the primary is $P = E \times I = 30$ watts. If the transformer has no losses, 30 watts is delivered to the secondary. The secondary steps down the voltage to 15 volts and steps up the current to 2 amperes. Thus, the power delivered to the load by the secondary is $P = E \times I = 15 \text{ volts} \times 2 \text{ amps} = 30$ watts.

The reason for this is that when the number of turns in the secondary is decreased, the opposition to the flow of the current is also decreased. Hence, more current will flow in the secondary. If the turns ratio of the transformer is increased to 1:2, the number of turns on the secondary is twice the number of turns on the primary. This means the opposition to current is doubled. Thus, voltage is doubled, but current is halved due to the increased opposition to current in the secondary. The important thing to remember is that <u>with the exception of the power consumed within the transformer, all power delivered to the primary by the source will be delivered to the load.</u> The form of the power may change, but the power in the secondary almost equals the power in the primary.

As a formula:

$$P_S = P_P - P_L$$

Where:

P_S = power delivered to the load by the secondary

P_P = power delivered to the primary by the source

P_L = power losses in the transformer

TRANSFORMER LOSSES

Practical power transformers, although highly efficient, are not perfect devices. Small power transformers used in electrical equipment have an 80 to 90 percent efficiency range, while large, commercial powerline transformers may have efficiencies exceeding 98 percent.

The total power loss in a transformer is a combination of three types of losses. One loss is due to the dc resistance in the primary and secondary windings. This loss is called COPPER loss or I^2R loss. The two other losses are due to EDDY CURRENTS and to HYSTERESIS in the core of the transformer. Copper loss, eddy-current loss, and hysteresis loss result in undesirable conversion of electrical energy into heat energy.

Q22. What is the mathematical relationship between the power in the primary (P_P) and power in the secondary (P_S) of a transformer?

Copper Loss

Whenever current flows in a conductor, power is dissipated in the resistance of the conductor in the form of heat. The amount of power dissipated by the conductor is directly proportional to the resistance of the wire, and to the square of the current through it. The greater the value of either resistance or current, the greater is the power dissipated. The primary and secondary windings of a transformer are usually made of low-resistance copper wire. The resistance of a given winding is a function of the diameter of the wire and its length. Copper loss can be minimized by using the proper diameter wire. Large diameter wire is required for high-current windings, whereas small diameter wire can be used for low-current windings.

Eddy-Current Loss

The core of a transformer is usually constructed of some type of ferromagnetic material because it is a good conductor of magnetic lines of flux.

Whenever the primary of an iron-core transformer is energized by an alternating-current source, a fluctuating magnetic field is produced. This magnetic field cuts the conducting core material and induces a voltage into it. The induced voltage causes random currents to flow through the core which dissipates power in the form of heat. These undesirable currents are called EDDY CURRENTS.

To minimize the loss resulting from eddy currents, transformer cores are LAMINATED. Since the thin, insulated laminations do not provide an easy path for current, eddy-current losses are greatly reduced.

Hysteresis Loss

When a magnetic field is passed through a core, the core material becomes magnetized. To become magnetized, the domains within the core must align themselves with the external field. If the direction of the field is reversed, the domains must turn so that their poles are aligned with the new direction of the external field.

Power transformers normally operate from either 60 Hz, or 400 Hz alternating current. Each tiny domain must realign itself twice during each cycle, or a total of 120 times a second when 60 Hz alternating current is used. The energy used to turn each domain is dissipated as heat within the iron core. This loss, called HYSTERESIS LOSS, can be thought of as resulting from molecular friction. Hysteresis loss can be held to a small value by proper choice of core materials.

TRANSFORMER EFFICIENCY

To compute the efficiency of a transformer, the input power to and the output power from the transformer must be known. The input power is equal to the product of the voltage applied to the primary and the current in the primary. The output power is equal to the product of the voltage across the secondary and the current in the secondary. The difference between the input power and the output power represents a power loss. You can calculate the percentage of efficiency of a transformer by using the standard efficiency formula shown below:

$$\text{Efficiency (in \%)} = \frac{P_{out}}{P_{in}} \times 100$$

Where:

P_{out} = total output power delivered to the load

P_{in} = total input power

Example. If the input power to a transformer is 650 watts and the output power is 610 watts, what is the efficiency?

Solution:

$$\text{Efficiency} = \frac{P_{out}}{P_{in}} \times 100$$

$$\text{Efficiency} = \frac{610\,W}{650\,W} \times 100$$

$$\text{Efficiency} = 93.8\%$$

Hence, the efficiency is approximately 93.8 percent, with approximately 40 watts being wasted due to heat losses.

Q23. Name the three power losses in a transformer.

Q24. The input power to a transformer is 1,000 watts and the output power is 500 watts. What is the efficiency of the transformer, expressed as a percentage?

TRANSFORMER RATINGS

When a transformer is to be used in a circuit, more than just the turns ratio must be considered. The voltage, current, and power-handling capabilities of the primary and secondary windings must also be considered.

The maximum voltage that can safely be applied to any winding is determined by the type and thickness of the insulation used. When a better (and thicker) insulation is used between the windings, a higher maximum voltage can be applied to the windings.

The maximum current that can be carried by a transformer winding is determined by the diameter of the wire used for the winding. If current is excessive in a winding, a higher than ordinary amount of power will be dissipated by the winding in the form of heat. This heat may be sufficiently high to cause the insulation around the wire to break down. If this happens, the transformer may be permanently damaged.

The power-handling capacity of a transformer is dependent upon its ability to dissipate heat. If the heat can safely be removed, the power-handling capacity of the transformer can be increased. This is sometimes accomplished by immersing the transformer in oil, or by the use of cooling fins. The power-handling capacity of a transformer is measured in either the volt-ampere unit or the watt unit.

Two common power generator frequencies (60 hertz and 400 hertz) have been mentioned, but the effect of varying frequency has not been discussed. If the frequency applied to a transformer is increased, the inductive reactance of the windings is increased, causing a greater ac voltage drop across the windings and a lesser voltage drop across the load. However, an increase in the frequency applied to a transformer should not damage it. But, if the frequency applied to the transformer is decreased, the reactance of the windings is decreased and the current through the transformer winding is increased. If the decrease in frequency is enough, the resulting increase in current will damage the transformer. For this reason a transformer may be used at frequencies above its normal operating frequency, but not below that frequency.

Q25. Why should a transformer designed for 400 hertz operation not be used for 60 hertz operation?

TYPES AND APPLICATIONS OF TRANSFORMERS

The transformer has many useful applications in an electrical circuit. A brief discussion of some of these applications will help you recognize the importance of the transformer in electricity and electronics.

POWER TRANSFORMERS

Power transformers are used to supply voltages to the various circuits in electrical equipment. These transformers have two or more windings wound on a laminated iron core. The number of windings and the turns per winding depend upon the voltages that the transformer is to supply. Their coefficient of coupling is 0.95 or more.

You can usually distinguish between the high-voltage and low-voltage windings in a power transformer by measuring the resistance. The low-voltage winding usually carries the higher current and therefore has the larger diameter wire. This means that its resistance is less than the resistance of the high-voltage winding, which normally carries less current and therefore may be constructed of smaller diameter wire.

So far you have learned about transformers that have but one secondary winding. The typical power transformer has several secondary windings, each providing a different voltage. The schematic symbol for a typical power-supply transformer is shown in figure 5-12. For any given voltage across the primary, the voltage across each of the secondary windings is determined by the number of turns in each secondary. A winding may be center-tapped like the secondary 350 volt winding shown in the figure. To center tap a winding means to connect a wire to the center of the coil, so that between this center tap and either terminal of the winding there appears one-half of the voltage developed across the entire winding. Most power transformers have colored leads so that it is easy to distinguish between the various windings to which they are connected. Carefully examine the figure which also illustrates the color code for a typical power transformer. Usually, red is used to indicate the high-voltage leads, but it is possible for a manufacturer to use some other color(s).

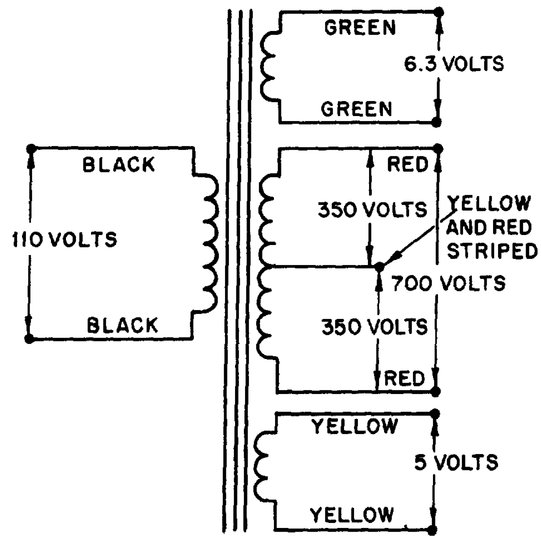


Figure 5-12.—Schematic diagram of a typical power transformer.

There are many types of power transformers. They range in size from the huge transformers weighing several tons-used in power substations of commercial power companies-to very small ones weighing as little as a few ounces-used in electronic equipment.

AUTOTRANSFORMERS

It is not necessary in a transformer for the primary and secondary to be separate and distinct windings. Figure 5-13 is a schematic diagram of what is known as an AUTOTRANSFORMER. Note that a single coil of wire is "tapped" to produce what is electrically a primary and secondary winding. The voltage across the secondary winding has the same relationship to the voltage across the primary that it would have if they were two distinct windings. The movable tap in the secondary is used to select a value of output voltage, either higher or lower than E_P, within the range of the transformer. That is, when the tap is at point A, E_S is less than E_P; when the tap is at point B, E_S is greater than E_P.

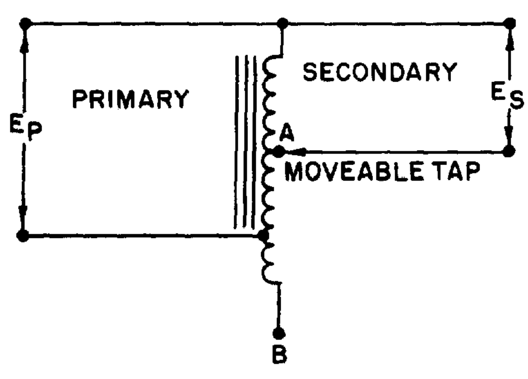


Figure 5-13.—Schematic diagram of an autotransformer.

AUDIO-FREQUENCY TRANSFORMERS

Audio-frequency (af) transformers are used in af circuits as coupling devices. Audio-frequency transformers are designed to operate at frequencies in the audio frequency spectrum (generally considered to be 15 Hz to 20kHz). They consist of a primary and a secondary winding wound on a laminated iron or steel core. Because these transformers are subjected to higher frequencies than are power transformers, special grades of steel such as silicon steel or special alloys of iron that have a very low hysteresis loss must be used for core material. These transformers usually have a greater number of turns in the secondary than in the primary; common step-up ratios being 1 to 2 or 1 to 4. With audio transformers the impedance of the primary and secondary windings is as important as the ratio of turns, since the transformer selected should have its impedance match the circuits to which it is connected.

RADIO-FREQUENCY TRANSFORMERS

Radio-frequency (rf) transformers are used to couple circuits to which frequencies above 20,000 Hz are applied. The windings are wound on a tube of nonmagnetic material, have a special powdered-iron core, or contain only air as the core material. In standard broadcast radio receivers, they operate in a frequency range of from 530 kHz to 1550 kHz. In a short-wave receiver, rf transformers are subjected to frequencies up to about 20 MHz - in radar, up to and even above 200 MHz.

IMPEDANCE-MATCHING TRANSFORMERS

For maximum or optimum transfer of power between two circuits, it is necessary for the impedance of one circuit to be matched to that of the other circuit. One common impedance-matching device is the transformer. To obtain proper matching, you must use a transformer having the correct turns ratio. The number of turns on the primary and secondary windings and the impedance of the transformer have the following mathematical relationship:

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}}$$

Because of this ability to match impedances, the impedance-matching transformer is widely used in electronic equipment.

Q26. List five different types of transformers according to their applications.

Q27. The leads to the primary and to the high-voltage secondary windings of a power transformer usually are of what color?

SAFETY

EFFECTS OF CURRENT ON THE BODY

Before learning safety precautions, you should look at some of the possible effects of electrical current on the human body. The following table lists some of the probable effects of electrical current on the human body.

AC 60 Hz (mA)	DC (mA)	Effects
0-1	0-4	Perception
1-4	4-15	Surprise
4-21	15-80	Reflex action
21-40	80-160	Muscular inhibition
40-100	160-300	Respiratory failure
Over 100	Over 300	Usually fatal

Note in the above chart that a current as low as 4 mA can be expected to cause a reflex action in the victim, usually causing the victim to jump away from the wire or other component supplying the current. While the current should produce nothing more than a tingle of the skin, the quick action of trying to get away from the source of this irritation could produce other effects (such as broken limbs or even death if a severe enough blow was received at a vital spot by the shock victim).

It is important for you to recognize that the resistance of the human body cannot be relied upon to prevent a fatal shock from a voltage as low as 115 volts or even less. Fatalities caused by human contact with 30 volts have been recorded. Tests have shown that body resistance under unfavorable conditions may be as low as 300 ohms, and possibly as low as 100 ohms (from temple to temple) if the skin is broken. Generally direct current is not considered as dangerous as an equal value of alternating current. This is evidenced by the fact that reasonably safe "let-go currents" for 60 hertz, alternating current, are 9.0 milliamperes for men and 6.0 milliamperes for women, while the corresponding values for direct current are 62.0 milliamperes for men and 41.0 milliamperes for women. Remember, the above table is a fist of probable effects. The actual severity of effects will depend on such things as the physical condition of the work area, the physiological condition and resistance of the body, and the area of the body through which the current flows. Thus, based on the above information, you MUST consider every voltage as being dangerous.

ELECTRIC SHOCK

Electric shock is a jarring, shaking sensation you receive from contact with electricity. You usually feel like you have received a sudden blow. If the voltage and resulting current are sufficiently high, you may become unconscious. Severe burns may appear on your skin at the place of contact; muscular spasms may occur, perhaps causing you to clasp the apparatus or wire which caused the shock and be unable to turn it loose.

RESCUE AND CARE OF SHOCK VICTIMS

The following procedures are recommended for rescue and care of electric shock victims:

1. Remove the victim from electrical contact at once, but DO NOT endanger yourself. You can do this by:

   - Throwing the switch if it is nearby
   - Cutting the cable or wires to the apparatus, using an ax with a wooden handle while taking care to protect your eyes from the flash when the wires are severed

- Using a dry stick, rope, belt, coat, blanket, shirt or any other nonconductor of electricity, to drag or push the victim to safety

2. Determine whether the victim is breathing. If the victim is not breathing, you must apply artificial ventilation (respiration) without delay, even though the victim may appear to be lifeless. DO NOT STOP ARTIFICIAL RESPIRATION UNTIL MEDICAL AUTHORITY PRONOUNCES THE VICTIM DEAD.

3. Lay the victim face up. The feet should be about 12 inches higher than the head. Chest or head injuries require the head to be slightly elevated. If there is vomiting or if facial injuries have occurred which cause bleeding into the throat, the victim should be placed on the stomach with the head turned to one side and 6 to 12 inches lower than the feet.

4. Keep the victim warm. The injured person's body heat must be conserved. Keep the victim covered with one or more blankets, depending on the weather and the person's exposure to the elements. Artificial means of warming, such as hot water bottles should not be used.

5. Drugs, food, and liquids should not be administered if medical attention will be available within a short time. If necessary, liquids may be administered. Small amounts of warm salt water, tea or coffee should be used. Alcohol, opiates, and other depressant substances must never be administered.

6. Send for medical personnel (a doctor if available) at once, but do NOT under any circumstances leave the victim until medical help arrives.

For complete coverage of administering artificial respiration, and on treatment of burn and shock victims, refer to *Standard First Aid Training Course*, NAVEDTRA 10081 (Series).

SAFETY PRECAUTIONS FOR PREVENTING ELECTRIC SHOCK

You must observe the following safety precautions when working on electrical equipment:

1. Never work alone. Another person may save your life if you receive an electric shock.

2. Work on energized circuits ONLY WHEN ABSOLUTELY NECESSARY. Power should be tagged out, using approved tagout procedures, at the nearest source of electricity.

3. Stand on an approved insulating material, such as a rubber mat.

4. Discharge power capacitors before working on deenergized equipment. Remember, a capacitor is an electrical power storage device.

5. When you must work on an energized circuit, wear rubber gloves and cover as much of your body as practical with an insulating material (such as shirt sleeves). This is especially important when you are working in a warm space where sweating may occur.

6. Deenergize equipment prior to hooking up or removing test equipment.

7. Work with only one hand inside the equipment. Keep the other hand clear of all obstacles that may provide a path, such as a ground, for current to flow.

8. Wear safety goggles. Sparks could damage your eyes, as could the cooling liquids in some components such as transformers should they overheat and explode.

9. Keep a cool head and think about the possible consequences before performing any action. Carelessness is the cause of most accidents. Remember the best technician is NOT necessarily the fastest one, but the one who will be on the job tomorrow.

Q28. What is the cause of most accidents?

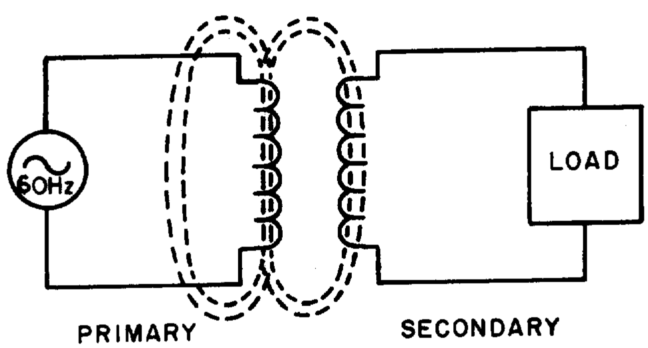
Q29. Before working on electrical equipment containing capacitors, what should you do to the capacitors?

Q30. When working on electrical equipment, why should you use only one hand?

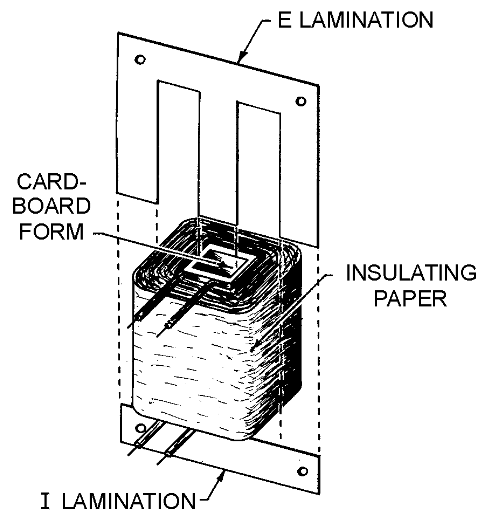
SUMMARY

As a study aid and for future reference, the important points of this chapter have been summarized below.

BASIC TRANSFORMER—The basic transformer is an electrical device that transfers alternating-current energy from one circuit to another circuit by magnetic coupling of the primary and secondary windings of the transformer. This is accomplished through mutual inductance (M). The coefficient of coupling (K) of a transformer is dependent upon the size and shape of the coils, their relative positions, and the characteristic of the core between the two coils. An ideal transformer is one where all the magnetic lines of flux produced by the primary cut the entire secondary. The higher the K of the transformer, the higher is the transfer of the energy. The voltage applied to the primary winding causes current to flow in the primary. This current generates a magnetic field, generating a counter emf (cemf) which has the opposite phase to that of the applied voltage. The magnetic field generated by the current in the primary also cuts the secondary winding and induces a voltage in this winding.

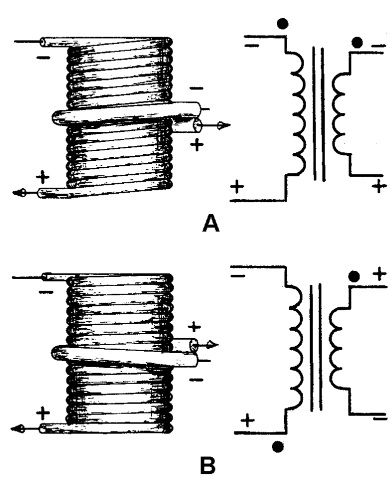


TRANSFORMER CONSTRUCTION—A TRANSFORMER consists of two coils of insulated wire wound on a core. The primary winding is usually wound onto a form, then wrapped with an insulating material such as paper or cloth. The secondary winding is then wound on top of the primary and both windings are wrapped with insulating material. The windings are then fitted onto the core of the transformer. Cores come in various shapes and materials. The most common materials are air, soft iron, and laminated steel. The most common types of transformers are the shell-core and the hollow-core types. The type and shape of the core is dependent on the intended use of the transformer and the voltage applied to the current in the primary winding.

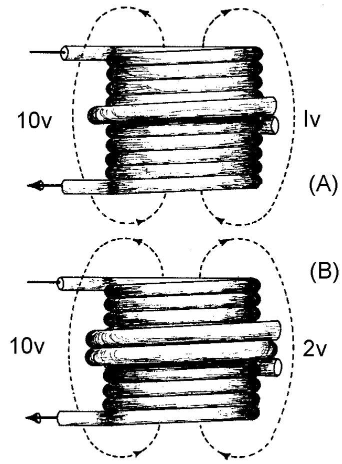


EXCITING CURRENT—When voltage is applied to the primary of a transformer, exciting current flows in the primary. The current causes a magnetic field to be set up around both the primary and the secondary windings. The moving flux causes a voltage to be induced into the secondary winding, countering the effects of the counter emf in the primary.

PHASE—When the secondary winding is connected to a load, causing current to flow in the secondary, the magnetic field decreases momentarily. The primary then draws more current, restoring the magnetic field to almost its original magnitude. The phase of the current flowing in the secondary circuit is dependent upon the phase of the voltage impressed across the primary and the direction of the winding of the secondary. If the secondary were wound in the same direction as the primary, the phase would be the same. If wound opposite to the primary, the phase would be reversed. This is shown on a schematic drawing by the use of phasing dots. The dots mean that the leads of the primary and secondary have the same phase. The lack of phasing dots on a schematic means a phase reversal.



TURNS RATIO—The TURNS RATIO of a transformer is the ratio of the number of turns of wire in the primary winding to the number of turns in the secondary winding. When the turns ratio is stated, the number representing turns on the primary is always stated first. For example, a 1:2 turns ratio means the secondary has twice the number of turns as the primary. In this example, the voltage across the secondary is two times the voltage applied to the primary.

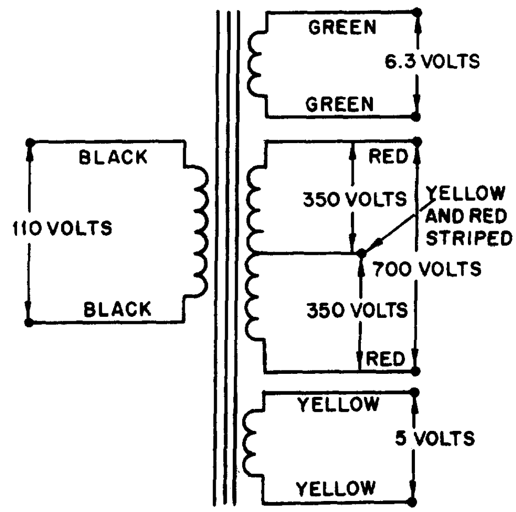


POWER AND CURRENT RATIOS—The power and current ratios of a transformer are dependent on the fact that power delivered to the secondary is always equal to the power delivered to the primary minus the losses of the transformer. This will always be true, regardless of the number of secondary windings. Using the law of power and current, it can be stated that current through the transformer is the inverse of the voltage or turns ratio, with power remaining the same or less, regardless of the number of secondaries.

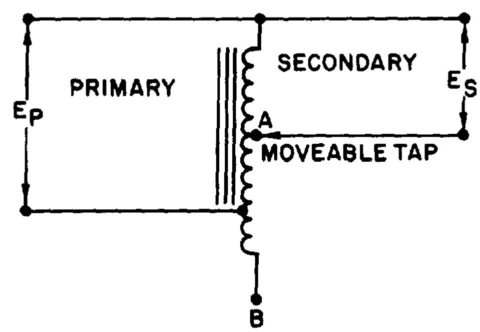
TRANSFORMER LOSSES—Transformer losses have two sources-copper loss and magnetic loss. Copper losses are caused by the resistance of the wire (I^2R). Magnetic losses are caused by eddy currents and hysteresis in the core. Copper loss is a constant after the coil has been wound and therefore a measureable loss. Hysteresis loss is constant for a particular voltage and current. Eddy-current loss, however, is different for each frequency passed through the transformer.

TRANSFORMER EFFICIENCY—The amplitude of the voltage induced in the secondary is dependent upon the efficiency of the transformer and the turns ratio. The efficiency of a transformer is related to the power losses in the windings and core of the transformer. Efficiency (in percent) equals $P_{out}/P_{in} \times 100$. A perfect transformer would have an efficiency of 1.0 or 100%.

POWER TRANSFORMER—A transformer with two or more windings wound on a laminated iron core. The transformer is used to supply stepped up and stepped down values of voltage to the various circuits in electrical equipment.



AUTOTRANSFORMER—A transformer with a single winding in which the entire winding can be used as the primary and part of the winding as the secondary, or part of the winding can be used as the primary and the entire winding can be used as the secondary.



AUDIO-FREQUENCY TRANSFORMER—A transformer used in audio-frequency circuits to transfer af signals from one circuit to another.

RADIO-FREQUENCY TRANSFORMER—A transformer used in a radio-frequency circuit to transfer rf signals from one circuit to another.

IMPEDANCE-MATCHING TRANSFORMER—A transformer used to match the impedance of the source and the impedance of the load. The mathematical relationship of the turns and impedance of the transformer is expressed by the equation:

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}}$$

ANSWERS TO QUESTIONS Q1. THROUGH Q30.

A1. *The transfer of energy from one circuit to another circuit by electromagnetic induction.*

A2. *Primary winding; secondary winding; core.*

A3. *Air; soft iron; steel.*

A4. *Hollow-core type; shell-core type.*

A5. *Primary to source; secondary to load.*

A6. *Additional insulation is provided between the layers of windings in the high-voltage transformer.*

A7.

a. air-core transformer

b. iron-core transformer

c. iron-core center tapped transformer

A8. *A voltage is applied to the primary, but no load is connected to the secondary.*

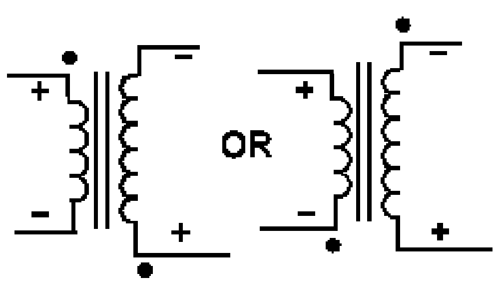
A9. *Exciting current is the current that flows in the primary of a transformer with the secondary open (no load attached).*

A10. *Self-induced or counter emf.*

A11. *The magnetic lines generated by the current in the primary cut the secondary windings and induce a voltage into them.*

A12. *In phase. Remember, the cemf of the primary is 180 degrees out of phase with the applied voltage. The induced voltage of the secondary of an unlike-wound transformer is also 180 degrees out of phase with the primary voltage.*

A13.



Note: Remember the dots indicate areas of like polarity, NOT a particular polarity.

A14. *Lines of flux generated by one winding which do not link the other winding.*

A15. *It causes K to be less than unity (1).*

A16. *Step up.*

A17.

$$\frac{E_S}{E_P} = \frac{N_S}{N_P} \text{ or}$$

$$E_S = \frac{E_P N_S}{N_P} = \frac{45\text{ V} \times 1500\text{ turns}}{500\text{ turns}} = 135\text{ V}$$

A18.

$$\frac{E_P}{E_S} = \frac{N_P}{N_S} \text{ or}$$

$$E_P = \frac{N_P E_S}{N_S} = \frac{7\text{ turns} \times 5\text{ V}}{1\text{ turn}} = 35\text{ V}$$

A19.

$$\frac{E_P}{E_S} = \frac{N_P}{N_S} \text{ or}$$

$$N_S = \frac{E_S N_P}{E_P} = \frac{420\text{ V} \times 800\text{ turns}}{60\text{ V}} = 5600\text{ turns}$$

A20.

$$\frac{N_P}{N_S} = \frac{I_S}{I_P} = \frac{1}{3} = 3:1 \text{ current ratio}$$

(Turns ratio and current ratio have an inverse relationship.)

A21.

$$\frac{N_P}{N_S} = \frac{I_S}{I_P} \text{ or}$$

$$I_P = \frac{N_S I_S}{N_P} = \frac{1\text{ turn} \times 5\text{A}}{5\text{ turns}} = 1\text{A}$$

A22.

$$P_S = P_P - P_L$$

A23. *Copper loss, eddy-current loss, and hysteresis loss.*

A24.

$$\text{Eff (in \%)} = \frac{P_{out}}{P_{in}} \times 100 = \frac{500\text{ W}}{1000\text{ W}}$$
$$\times 100 = 0.5 \times 100 = 50\%$$

A25. *The inductive reactance at 60 hertz would be too low. The resulting excessive current would probably damage the transformer.*

A26.

- *a.* *Power transformer*
- *b.* *Autotransformer*
- *c.* *Impedance-matching transformer*
- *d.* *Audio-frequency transformer*
- *e.* *Radio-frequency transformer*

A27. *Primary leads-black; secondary leads-red.*

A28. *Carelessness.*

A29. *Discharge them by shorting them to ground.*

A30. *To minimize the possibility of providing a path for current through your body.*

APPENDIX I

GLOSSARY

AIR-CORE TRANSFORMER—A transformer composed of two or more coils, which are wound around a non-metallic core.

ALTERNATING CURRENT—An electrical current which constantly changes amplitude and changes polarity at regular intervals.

APPARENT POWER—That power apparently available for use in an ac circuit containing a reactive element. It is the product of effective voltage times effective current expressed in voltamperes. It must be multiplied by the power factor to obtain true power available.

AVERAGE VALUE OF AC—The average of all the instantaneous values of one-half cycle of alternating current.

CAPACITANCE—The property of an electrical circuit that opposes changes in voltage.

CAPACITOR—An electrical device capable of storing electrical energy in an electrostatic field.

CAPACITIVE REACTANCE—The opposition offered to the flow of an alternating current by capacitance, expressed in ohms. The symbol for capacitive reactance is X_c.

CHARGE CYCLE—The period of time that a capacitor in an electrical circuit is storing a charge.

COIL—An inductive device created by looping turns of wire around a core.

COPPER LOSS (I^2R LOSS)—The power lost due to the resistance of the conductors. In transformers the power lost due to heating because of current flow (I) through the resistance (R) of the windings.

CORE—Any material that affords a path for magnetic flux lines in a coil.

COUNTER EMF—Counter electromotive force; an electromotive force (voltage) induced in a coil that opposes the applied voltage.

COUPLING, COEFFICIENT OF—An expression of the extent to which two inductors are coupled by magnetic lines of force. This is expressed as a decimal or percentage of maximum possible coupling and represented by the letter K.

CYCLE—One complete positive and one complete negative alternation of a current or voltage.

DIELECTRIC—An insulator; a term applied to the insulating material between the plates of a capacitor.

DIELECTRIC CONSTANT—The ratio of the capacitance of a capacitor with a dielectric between the electrodes to the capacitance of a capacitor with air between the electrodes.

DIELECTRIC HYSTERESIS LOSS—Power loss of a capacitor due to the changes in orientation of electron orbits in the dielectric caused by rapid reversal in polarity of line voltage. The higher the frequency, the greater the loss.

DIELECTRIC LEAKAGE—Power loss of a capacitor due to the leakage of current through the dielectric. Also relates to leakage resistance, the higher the leakage resistance, the lower the dielectric leakage.

DISPLACEMENT CURRENT—The current which appears to flow through a capacitor.

EDDY CURRENT—Induced circulating currents in a conducting material that are caused by a varying magnetic field.

EDDY CURRENT LOSS—Losses caused by random current flowing in the core of a transformer. Power is lost in the form of heat.

EFFECTIVE VALUE—Same as root-mean-square.

ELECTROMAGNETIC INDUCTION—The production of a voltage in a coil due to the change in the number of magnetic lines of force (flux linkage) passing through the coil.

ELECTROMAGNETISM—The generation of a magnetic field around a current carrying conductor.

ELECTROMOTIVE FORCE (emf)—The force (voltage) that produces an electric current in a circuit

ELECTROSTATIC FIELD—The field of influence between two charged bodies.

EXCITING CURRENT—The current that flows in the primary winding of a transformer, which produces a magnetic flux field. Also called magnetizing current.

FARAD—The basic unit of capacitance. A capacitor has a capacitance of 1 farad when a voltage change of 1 volt per second across it produces a current of 1 ampere.

FLUX—Electrostatic or magnetic lines of force.

FREQUENCY (f)—The number of complete cycles per second existing in any form of wave motion; such as the number of cycles per second of an alternating current.

HENRY (H)—The electromagnetic unit of inductance or mutual inductance. The inductance of a circuit is 1 henry when a current variation of 1 ampere per second induces 1 volt. It is the basic unit of inductance. In radio, smaller units are used such as the millihenry (mH), which is one-thousandth of a henry (H), and the microhenry (μH) which is one-millionth of a henry.

HERTZ (Hz)—The basic unit of frequency equal to one cycle per second.

HYSTERESIS—The time lag of the magnetic flux in a magnetic material behind the magnetizing force producing it, caused by the molecular friction of the molecules trying to align themselves with the magnetic force applied to the material.

HYSTERESIS LOSS—The power loss in an iron-core transformer or other alternating-current device as a result of magnetic hysteresis.

IMPEDANCE—The total opposition offered to the flow of an alternating current. It may consist of any combination of resistance, inductive reactance, and capacitive reactance. The symbol for impedance is Z.

IN PHASE—Applied to the condition that exists when two waves of the same frequency pass through their maximum and minimum values of like polarity at the same instant.

INDUCTANCE—The property of a circuit which tends to oppose a change in the existing current flow. The symbol for inductance is L.

INDUCTIVE COUPLING—Coupling of two coils by means of magnetic lines of force. In transformers, coupling applied through magnetic lines of force between the primary and secondary windings.

INDUCTIVE REACTANCE—The opposition to the flow of an alternating current caused by the inductance of a circuit, expressed in ohms. Identified by the letter X_L.

INSTANTANEOUS VALUE—The magnitude at any particular instant when a value is continually varying with respect to time.

LAG—The amount one wave is behind another in time; expressed in electrical degrees.

LAMINATED CORE—A core built up from thin sheets of metal insulated from each other and used in transformers.

LEAD—The opposite of lag. Also a wire or connection.

LEAKAGE FLUX—Magnetic flux lines produced by the primary winding which do not link the turns of the secondary winding.

LEAKAGE RESISTANCE—The electrical resistance which opposes the flow of current through the dielectric of a capacitor. The higher the leakage resistance the slower the capacitor will discharge or leak across the dielectric.

LEFT-HAND-RULE FOR GENERATORS—A rule or procedure used to determine the direction of electron current flow in the windings of a generator.

LENZ'S LAW—The current induced in a circuit due to its motion in a magnetic field or to a change in its magnetic flux in such a direction as to exert a mechanical force opposing the motion or to oppose the change in flux.

MAGNETIC FIELD—Region in which the magnetic forces created by a permanent magnet or by a current-carrying conductor or coil can be detected.

MAGNETIC LINES OF FORCE—Imaginary lines used for convenience to designate the direction in which magnetic forces are acting as a result of magneto-motive force.

MUTUAL FLUX—The total flux in the core of a transformer that is common to both the primary and secondary windings. The flux links both windings.

MUTUAL INDUCTANCE—A circuit property existing when the relative position of two inductors causes the magnetic lines of force from one to link with the turns of the other. The symbol for mutual inductance is M.

NEGATIVE ALTERNATION—The negative half of an ac waveform

NO-LOAD CONDITION—The condition that exists when an electrical source or secondary of a transformer is operated without an electrical load.

PEAK-TO-PEAK—The measure of absolute magnitude of an ac waveform, measured from the greatest positive alternation to greatest negative alternation.

PEAK VALUE—The maximum instantaneous value of a varying current, voltage, or power. It is equal to 1.414 times the effective value of a sine wave.

PERIOD TIME—The time required to complete one cycle of a waveform.

PHASE—The angular relationship between two alternating currents or voltages when the voltage or current is plotted as a function of time. When the two are in phase, the angle is zero, and both reach their peak simultaneously. When out of phase, one will lead or lag the other; at the instant when one is at its peak, the other will not be at peak value and (depending on the phase angle) may differ in polarity as well as magnitude.

PHASE ANGLE—The number of electrical degrees of lead or lag between the voltage and current waveforms in an ac circuit.

PHASE DIFFERENCE—The time in electrical degrees by which one wave leads or lags another.

POLARIZATION—The magnetic orientation of molecules in a magnetizable material in a magnetic field, whereby tiny internal magnets tend to line up in the field.

POSITIVE ALTERNATION—The positive half of an ac waveform.

POWER FACTOR—The ratio of the actual power of an alternating or pulsating current, as measured by a wattmeter, to the apparent power, as indicated by ammeter and voltmeter readings. The power factor of an inductor, capacitor, or insulator is an expression of their losses

POWER LOSS—The electrical power supplied to a circuit that does no work, usually dissipated as heat.

PRIMARY WINDING—The winding of a transformer connected to the electrical source.

Q—Figure of merit of efficiency of a circuit or coil. Ratio of inductive reactance to resistance in servos. Relationship between stored energy (capacitance) and rate of dissipation in certain types of electric elements, structures, or materials.

RADIO FREQUENCY (RF)—Any frequency of electrical energy capable of propagation into space.

RATIO—The value obtained by dividing one number by another, indicating their relative proportions.

RC CONSTANT—Time constant of a resistor-capacitor circuit; equal in seconds to the resistance value in ohms multiplied by the capacitance value in farads.

REACTANCE—The opposition offered to the flow of an alternating current by the inductance, capacitance, or both, in any circuit.

RESONANCE—The condition existing in a circuit when the values of inductance, capacitance, and the applied frequency are such that the inductive reactance and capacitive reactance cancel each other.

RLC CIRCUIT—An electrical circuit which has the properties of resistance, inductance, and capacitance.

RMS—Abbreviation of Root Mean Square.

ROOT MEAN SQUARE (RMS)—The equivalent heating value of an alternating current or voltage, as compared to a direct current or voltage. It is 0.707 times the peak value of a sine wave.

SECONDARY—The output coil of a transformer.

SELF-INDUCTION—The production of a counter-electromotive force in a conductor when its own magnetic field collapses or expands with a change in current in the conductor

SINE WAVE—The curve traced by the projection on a uniform time scale of the end of a rotating arm, or vector. Also known as a sinusoidal wave.

THETA—The greek letter (θ) used to represent phase angle.

TIME CONSTANT—The time required to charge a capacitor to 63.2 percent of maximum voltage or discharge to 36.8 percent of its final voltage. The time required for the current in an inductor to increase to 63.2 percent of maximum current or decay to 36.8 percent of its final current.

TRANSFORMER—A device composed of two or more coils, linked by magnetic lines of force, used to transfer electrical energy from one circuit to another.

TRANSFORMER EFFICIENCY—The ratio of output power to input power, generally expressed as a percentage.

$$\text{Efficiency} = \frac{P_{out}}{P_{in}} \times 100$$

TRANSFORMER, STEP-DOWN—A transformer so constructed that the number of turns in the secondary winding is less than the number of turns in the primary winding. This construction will provide less voltage in the secondary circuit than in the primary circuit.

TRANSFORMER, STEP-UP—A transformer so constructed that the number of turns in the secondary winding is more than the number of turns in the primary winding. This construction will provide more voltage in the secondary circuit than in the primary circuit.

TRUE POWER—The power dissipated in the resistance of the circuit, or the power actually used in the circuit.

TURN—One complete loop of a conductor about a core.

TURNS RATIO—The ratio of number of turns in the primary winding to the number of turns in the secondary winding of a transformer.

UNIVERSAL TIME CONSTANT CHART—A chart used to find the time constant of a circuit if the impressed voltage and the values of R and C or R and L are known.

WAVEFORM—The shape of the wave obtained when instantaneous values of an ac quantity are plotted against time in rectangular coordinates.

WAVELENGTH (λ)—The distance, usually expressed in meters, traveled by a wave during the time interval of one complete cycle. It is equal to the velocity of light divided by the frequency.

WORKING VOLTAGE—The maximum voltage that a capacitor may operate at without the risk of damage.

VAR—Abbreviation for Volt-Amperes Reactive.

VECTOR—A line used to represent both direction and magnitude.

APPENDIX II

GREEK ALPHABET

Name	Upper Case	Lower Case	Designates
Alpha…….	Α	α	Angles.
Beta……...	Β	β	Angles, flux density.
Gamma…..	Γ	γ	Conductivity.
Delta……..	Δ	ó	Variation of a quantity, increment.
Epsilon…...	Ε	ε	Base of natural logarithms (2.71828).
Zeta………	Ζ	ζ	Impedance, coefficients, coordinates.
Eta……….	Η	η	Hysteresis coefficient, efficiency, magnetizing force.
Theta……..	Θ	θ	Phase angle.
Iota……….	Ι	ι	
Kappa……	Κ	κ	Dielectric constant, coupling coefficient, susceptibility.
Lambda…..	Λ	λ	Wavelength.
Mu……….	Μ	μ	Permeability, micro, amplification factor.
Nu………..	Ν	ν	Reluctivity.
Xi………...	Ξ	ξ	
Omicron….	Ο	ο	
Pi………...	Π	π	3.1416
Rho………	Ρ	ρ	Resistivity.
Sigma…….	Σ	σ	
Tau……….	Τ	τ	Time constant, time-phase displacement.
Upsilon…..	Υ	υ	
Phi……….	Φ	ϕ	Angles, magnetic flux.
Chi……….	Χ	χ	
Psi………..	Ψ	ψ	Dielectric flux, phase difference.
Omega…...	Ω	ω	Ohms (upper case), angular velocity (2 π f) (lower case).

APPENDIX III

SQUARE AND SQUARE ROOTS

N	N^2	$\sqrt{N}$	N	N^2	$\sqrt{N}$	N	N^2	$\sqrt{N}$
1	1	1.000	41	1681	6.4031	81	6561	9.0000
2	4	1.414	42	1764	6.4807	82	6724	9.0554
3	9	1.732	43	1849	6.5574	83	6889	9.1104
4	16	2.000	44	1936	6.6332	84	7056	9.1652
5	25	2.236	45	2025	6.7082	85	7225	9.2195
6	36	2.449	46	2116	6.7823	86	7396	9.2736
7	49	2.646	47	2209	6.8557	87	7569	9.3274
8	64	2.828	48	2304	6.9282	88	7744	9.3808
9	81	3.000	49	2401	7.0000	89	7921	9.4340
10	100	3.162	50	2500	7.0711	90	8100	9.4868
11	121	3.3166	51	2601	7.1414	91	8281	9.5394
12	144	3.4641	52	2704	7.2111	92	8464	9.5917
13	169	3.6056	53	2809	7.2801	93	8649	9.6437
14	196	3.7417	54	2916	7.3485	94	8836	9.6954
15	225	3.8730	55	3025	7.4162	95	9025	9.7468
16	256	4.0000	56	3136	7.4833	96	9216	9.7980
17	289	4.1231	57	3249	7.5498	97	9409	9.8489
18	324	4.2426	58	3364	7.6158	98	9604	9.8995
19	361	4.3589	59	3481	7.6811	99	9801	9.9499
20	400	4.4721	60	3600	7.7460	100	10000	10.0000
21	441	4.5826	61	3721	7.8102	101	10201	10.0499
22	484	4.6904	62	3844	7.8740	102	10404	10.0995
23	529	4.7958	63	3969	7.9373	103	10609	10.1489
24	576	4.8990	64	4096	8.0000	104	10816	10.1980
25	625	5.0000	65	4225	8.0623	105	11025	10.2470
26	676	5.0990	66	4356	8.1240	106	11236	10.2956
27	729	5.1962	67	4489	8.1854	107	11449	10.3441
28	784	5.2915	68	4624	8.2462	108	11664	10.3923
29	841	5.3852	69	4761	8.3066	109	11881	10.4403
30	900	5.4772	70	4900	8.3666	110	12100	10.4881
31	961	5.5678	71	5041	8.4261	111	12321	10.5357
32	1024	5.6569	72	5184	8.4853	112	12544	10.5830
33	1089	5.7447	73	5329	8.5440	113	12769	10.6301
34	1156	5.8310	74	5476	8.6023	114	12996	10.6771
35	1225	5.9161	75	5625	8.6603	115	13225	10.7238
36	1296	6.0000	76	5776	8.7178	116	13456	10.7703
37	1369	6.0828	77	5929	8.7750	117	13689	10.8167
38	1444	6.1644	78	6084	8.8318	118	13924	10.8628
39	1521	6.2450	79	6241	8.8882	119	14161	10.9087
40	1600	6.3246	80	6400	8.9443	120	14400	10.9545

For numbers up to 120. For larger numbers divide into factors smaller than 120.

Examples: $\sqrt{225}$ and $\sqrt{16200}$

$$225 = 5 \times 45$$
$$\sqrt{225} = \sqrt{5} \times \sqrt{45}$$
$$\sqrt{225} = 2.236 \times 6.7082$$
$$\sqrt{225} = 15$$

$$16200 = 100 \times 81 \times 2$$
$$\sqrt{16200} = \sqrt{100} \times \sqrt{81} \times \sqrt{2}$$
$$\sqrt{16200} = 10 \times 9 \times 1.414$$
$$\sqrt{16200} = 127.26$$

APPENDIX IV

USEFUL AC FORMULAS

PERIOD TIME (t)

$$t = \frac{1}{f}$$

FREQUENCY (f)

$$f = \frac{1}{t}$$

AVERAGE VOLTAGE OR CURRENT

$$E_{avg} = 0.636 \times E_{max}$$
$$I_{avg} = 0.636 \times I_{max}$$

EFFECTIVE VALUE OF VOLTAGE OR CURRENT

$$E_{eff} = 0.707 \times E_{max}$$
$$I_{eff} = 0.707 \times I_{max}$$

MAXIMUM VOLTAGE OR CURRENT

$$I_{max} = 1.414 \times I_{eff}$$
$$E_{max} = 1.414 \times E_{eff}$$

OHM'S LAW OF AC CIRCUIT CONTAINING ONLY RESISTANCE

$$I_{eff} = \frac{E_{eff}}{R}$$
$$I_{avg} = \frac{E_{avg}}{R}$$
$$I_{max} = \frac{E_{max}}{R}$$

L/R TIME CONSTANT (TC)

$$TC(\text{in seconds}) = \frac{L(\text{henrys})}{R(\text{ohms})}$$

MUTUAL INDUCTANCE (M)

$$M = K\sqrt{L_1 L_2}$$

TOTAL INDUCTANCE (L_T) Series without magnetic coupling

$$L_T = L_1 + L_2 + L_3 \ldots L_n$$
with magnetic coupling
$$L_T = L_1 + L_2 \pm 2M$$

TOTAL INDUCTANCE (L_T) PARALLEL
(No magnetic coupling)

$$\frac{1}{L_T} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} \ldots \frac{1}{L_n}$$

CAPACITANCE (C)

$$C(\text{farads}) = \frac{Q(\text{coulombs})}{E(\text{volts})}$$
$$C = 0.2249\frac{KA}{d}$$

RC TIME CONSTANT (t)

$$t(\text{in seconds}) = R(\text{ohms}) \times C(\text{farads})$$
$$t(\text{in seconds}) = R(\text{Mohms}) \times C(\mu F)$$
$$t(\mu s) = R(\text{ohms}) \times C(\mu F)$$
$$t(\mu s) = R(\text{Mohms}) \times C(pF)$$

TOTAL CAPACITANCE (C_T) SERIES

$$C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} \cdots \frac{1}{C_n}}$$

$$C_T = \frac{C_1 \times C_2}{C_1 + C_2}$$

TOTAL CAPACITANCE (C_T) PARALLEL

$$C_T = C_1 + C_2 + C_3 \ldots C_n$$

INDUCTIVE REACTANCE (X_L)

$$X_L = 2\pi f L$$

CAPACITIVE REACTANCE (X_C)

$$X_C = \frac{1}{2\pi f C}$$

IMPEDANCE (Z)

$$Z = \sqrt{R^2 + X^2}$$

OHM'S LAW FOR REACTIVE CIRCUITS

$$I = \frac{E}{X_L} \text{ or } I = \frac{E}{X_C}$$

OHM'S LAW FOR CIRCUITS CONTAINING RESISTANCE AND REACTANCE

$$I = \frac{E}{Z}$$

REACTIVE POWER

$$= I_L^2 X_L - I_C^2 X_C$$
$$= I_C^2 X_C - I_L^2 X_L$$

APPARENT POWER

$$= I_Z^2 Z$$
$$= \sqrt{(\text{true power})^2 + (\text{reactive power})^2}$$

POWER FACTOR (PF)

$$PF = \frac{(I_R)^2 R}{(I_Z)^2 Z}$$

VOLTAGE ACROSS THE SECONDARY (E_s)

$$E_S = \frac{E_P N_S}{N_S}$$

VOLTAGE ACROSS THE PRIMARY (E_p)

$$E_P = \frac{E_S N_P}{N_S}$$

CURRENT ACROSS THE SECONDARY (I_s)

$$I_P = \frac{E_S I_S}{E_P}$$

CURRENT ACROSS THE PRIMARY (I_p)

$$I_P = \frac{E_S I_S}{E_P}$$

TRANSFORMER EFFICIENCY

$$= \frac{P_{out}}{P_{in}} \times 100$$

APPENDIX V

TRIGONOMETRIC FUNCTIONS

In a right triangle, there are several relationships which always hold true. These relationships pertain to the length of the sides of a right triangle, and the way the lengths are affected by the angles between them. An understanding of these relationships, called trigonometric functions, is essential for solving problems in a-c circuits such as power factor, impedance, voltage drops, and so forth.

To be a RIGHT triangle, a triangle must have a "square" corner; one in which there is exactly 90º between two of the sides. Trigonometric functions do not apply to any other type of triangle. This type of triangle is shown in figure V-1.

By use of the trigonometric functions, it is possible to determine the UNKNOWN length of one or more sides of a triangle, or the number of degrees in UNKNOWN angles, depending on what is presently known about the triangle. For instance, if the lengths of any two sides are known, the third side and both angles θ (theta) and Φ (phi) may be determined. The triangle may also be solved if the length of any one side and one of the angles(θ or Φ in fig. V-1) are known.

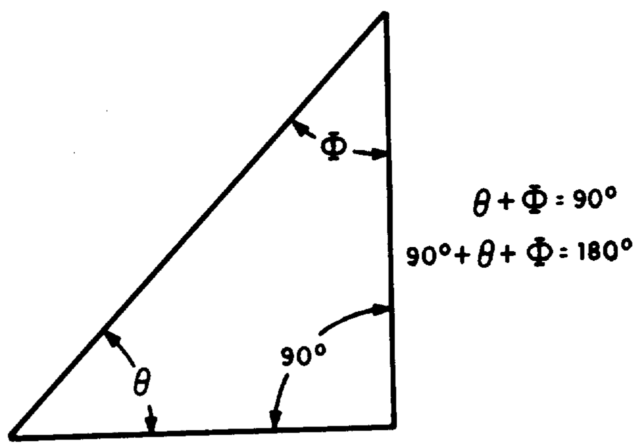


Figure V-1.—A right triangle.

The first basic fact of triangles is that IN ANY RIGHT TRIANGLE, THE SUM OF THE THREE ANGLES FORMED INSIDE THE TRIANGLE MUST ALWAYS EQUAL 180º . If one angle is always 90º (a right angle) then the sum of the other two angles must always be 90º .

$$\theta + \Phi = 90^\circ$$

and $$90^\circ + \theta + \Phi = 180^\circ$$

or $$90^\circ + 90^\circ = 180^\circ$$

thus, if angle θ is known, Φ may be quickly determined.

For instance, if θ ; is 30º m what is Φ ?

$$90^\circ + 30^\circ + \Phi = 180^\circ$$

$$\text{Transposing} \quad \Phi = 180^\circ - 90^\circ - 30^\circ$$

$$\Phi = 60^\circ$$

Also, if θ is known, Φ may be determined in the same manner.

The second basic fact you must understand is that FOR EVERY DIFFERENT COMBINATION OF ANGLES IN A TRIANGLE, THERE IS A DEFINITE RATIO BETWEEN THE LENGTHS OF THE THREE SIDES. Consider the triangle in figure V-2, consisting of the base, side B; the altitude, side A; and the hypotenuse, side C. (The hypotenuse is always the longest side, and is always opposite the 90º angle.)

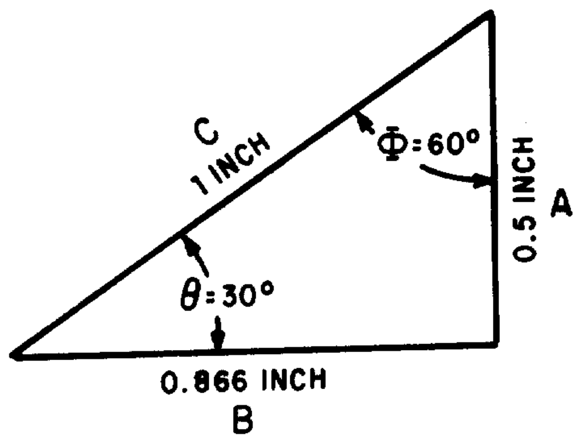


Figure V-2.—A 30º - 60º - 90º triangle.

If angle θ is 30º , Φ must be 60º . With θ equal to 30º , the ratio of the length of side B to side C is 0.866 to 1. That is, if the hypotenuse is 1 inch long, the side adjacent to θ, side B, is 0.866 inch long. Also, with θ equal to 30º , the ratio of side A to side C is 0.5 to 1. That is, with the hypotenuse 1 inch long, the side opposite to θ (side A) is 0.5 inch long. With θ still at 30º , side A is 0.5774 of the length of B. With the combination of angles given (30º -60º -90º) these are the ONLY ratios of lengths that will "fit" to form a right triangle.

Note that three ratios are shown to exist for the given value of θ: the ratio B\C which is always referred to as the COSINE ratio of θ, the ratio A\C, which is always the SINE ratio of θ, and the ratio A\B, which is always the TANGENT ratio of θ. If θ changes, all three ratios change, because the lengths of the sides (base and altitude) change. There is a set of ratios for every increment between 0º and 90º . These angular ratios, or sine, cosine, and tangent functions, are listed for each degree and tenth of degree in a table at the end of this appendix. In this table, the length of the hypotenuse of a triangle is considered fixed. Thus, the ratios of length given refer to the manner in which sides A and B vary with relation to each other and in relation to side C, as angle θ is varied from 0º to 90º .

The solution of problems in trigonometry (solution of triangles is much simpler when the table of trigonometric functions is used properly. The most common ways in which it is used will be shown by solving a series of exemplary problems.

Problem 1: If the hypotenuse of the triangle (side C) in figure V-3 is 10 inches long, and angle θ is 33°, how long are sides B and A?

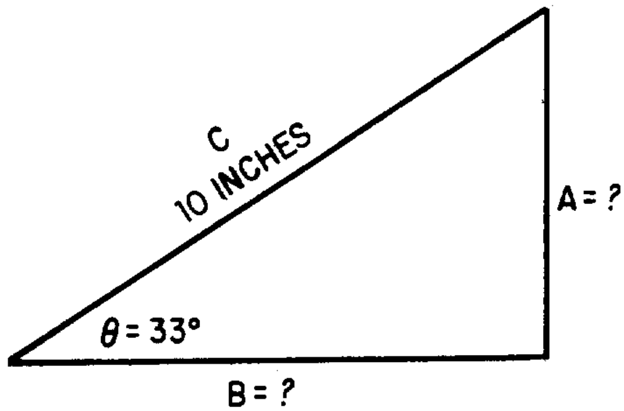


Figure V-3.—Problem 1.

Solution: The ratio B/C is the cosine function. By checking the table of functions, you will find that the cosine of 33° is 0.8387. This means that the length of B is 0.8387 the length of side C. If side C is 10 inches long, then side B must be 10×0.8387, or 8.387 inches in length. To determine the length of side A, use the sine function, the ratio A\C. Again consulting the table of functions, you will find that the sine of 33° is 0.5446. Thus, side A must be 10×0.5446, or 5.446 inches in length.

Problem 2: The triangle in figure V-4 has a base 74.2 feet long, and hypotenuse 100 feet long. What is θ, and how long is side A?

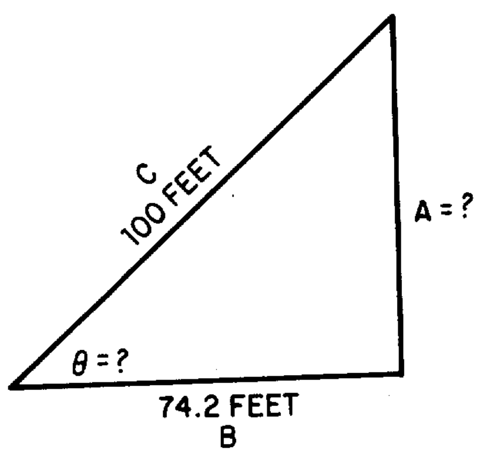


Figure V-4.—Problem 2.

Solution: When no angles are given, you Must always solve for a known angle first. The ratio B\C is the cosine of the unknown angle θ; therefore 74.2/100 or 0.742, is the cosine of the unknown angle. Locating 0.742 as a cosine value in the table, you find that it is the cosine of 42.1°. That is, $\theta = 42.1^\circ$. With θ known, side A is solved for by use of the sine ratio A/C. The sine of 42.1°, according to the table, is 0.6704. Therefore, side A is 100×0.6704, or 67.04 feet long.

Problem 3: In the triangle in figure V-5, the base is 3 units long, and the altitude is 4 units. What is θ, and how long is the hypotenuse?

Solution: With the information given, the tangent of θ may be determined. Tan θ = A/B = 4/3 = 1.33.

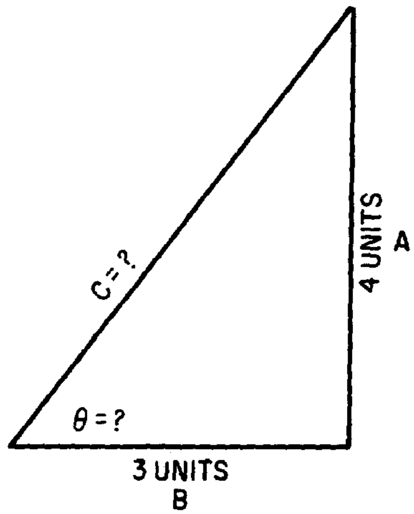


Figure V-5.—Problem 3.

Locating the value 1.33 as a tangent value in the table of functions, you find it to be the tangent of 53.1° . Therefore, θ = 53.1° . Once θ is known, either the sine or cosine ratio may be used to determine the length of the hypotenuse. The cosine of 53.1° is 0.6004. This indicates that the base of 3 units is 0.6004, the length of the hypotenuse. Therefore, the hypotenuse is 3/0.6004, or 5 units in length. Using the sine ratio, the hypotenuse is 4/0.7997, or 5 units in length.

In the foregoing explanations and problems, the sides of triangles were given in inches, feet, and units. In applying trigonometry to a-c circuit problems, these units of measure will be replaced by such values given in ohms, amperes, volts, and watts. Angle θ will be the phase angle between (source) voltage and circuit current. However, the solution of these a-c problems is accomplished in exactly the same manner as the foregoing problems. Only the units and some terminology are changed.

APPENDIX VI

TRIGONOMETRIC TABLES

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
0.0	0.0000	1.0000	0.00000		90.0	4.0	0.6976	0.9976	0.6993	140.301	86.0
0.1	0.0175	1.0000	0.00175	573.0	0.9	0.1	0.7150	0.9974	0.7168	130.951	0.9
0.2	0.0349	1.0000	0.00349	286.5	0.8	0.2	0.7324	0.9973	0.7344	130.617	0.8
0.3	0.0524	1.0000	0.00524	191.0	0.7	0.3	0.7498	0.9972	0.17519	130.30	0.7
0.4	0.0698	1.0000	0.00698	143.24	0.6	0.4	0.7672	0.9971	0.7695	120.996	0.6
0.5	0.0873	1.0000	0.00873	114.59	0.5	0.5	0.7846	0.9969	0.7870	120.706	0.5
0.6	0.1047	0.9999	0.10147	95.49	0.4	0.6	0.8020	0.9968	0.8046	120.429	0.4
0.7	0.1222	0.9999	0.01222	81.85	0.3	0.7	0.8194	0.9966	0.8221	120.163	0.3
0.8	0.1396	0.9999	0.01396	71.62	0.2	0.8	0.8368	0.9965	0.8397	110.909	0.2
0.9	0.1571	0.9999	0.01571	63.66	0.1	0.9	0.8542	0.9963	0.8573	110.664	0.1
1.0	0.1745	0.9998	0.1746	570.20	89.0	5.0	0.8716	0.9962	0.8749	110.430	85.0
0.1	0.1920	0.9998	0.1920	520.8	0.9	0.1	0.8889	0.9960	0.89215	110.205	0.9
0.2	0.2094	0.9998	0.2095	470.74	0.8	0.2	0.9063	0.9959	0.9101	10.988	0.8
0.3	0.2269	0.9997	0.2269	440.7	0.7	0.3	0.9237	0.9957	0.9277	10.780	0.7
0.4	0.2443	0.9997	0.2444	40.92	0.6	0.4	0.9411	0.9956	0.9453	10.579	0.6
0.5	0.2618	0.9997	0.2619	380.19	0.5	0.5	0.9585	0.9954	0.9629	10.385	0.5
0.6	0.2792	0.9996	0.2793	350.80	0.4	0.6	0.9758	0.9952	0.9805	10.199	0.4
0.7	0.2967	0.9996	0.2968	330.69	0.4	0.7	0.9932	0.9951	0.9981	10.19	0.3
0.8	0.3141	0.9995	0.3143	310.82	0.2	0.8	0.10106	0.9949	0.10158	90.845	0.2
0.9	0.3316	0.9995	0.3317	30.14	0.1	0.9	0.10279	0.9947	0.10334	90.677	0.1
2.0	0.3490	0.9994	0.3492	280.64	88.0	6.0	0.10453	0.9945	0.10510	90.514	84.0
0.1	0.3664	0.9993	0.3667	270.27	0.9	0.1	0.10626	0.9943	0.10687	90.357	0.9
0.2	0.3839	0.9993	0.3842	260.3	0.8	0.2	0.1080	0.9942	0.10863	90.205	0.8
0.3	0.4013	0.9992	0.4016	240.90	0.7	0.3	0.10973	0.9940	0.11040	90.58	0.7
0.4	0.4188	0.9991	0.4191	230.86	0.6	0.4	0.11147	0.9938	0.11217	80.915	0.6
0.5	0.4362	0.9990	0.4366	220.90	0.5	0.5	0.11320	0.9936	0.11394	80.777	0.5
0.6	0.4536	0.9990	0.4541	220.2	0.4	0.6	0.11494	0.9934	0.11570	80.643	0.4
0.7	0.4711	0.9989	0.4716	210.20	0.3	0.7	0.11667	0.9932	0.11747	80.513	0.3
0.8	0.4885	0.9988	0.4891	20.45	0.2	0.8	0.11840	0.9930	0.11924	80.386	0.2
0.9	0.5059	0.9987	0.5066	190.74	0.1	0.9	0.12014	0.9928	0.12101	80.264	0.1
3.0	0.5234	0.9986	0.5241	190.81	87.0	7.0	0.12187	0.9925	0.12278	80.144	83.0
0.1	0.5408	0.9985	0.5416	180.464	0.9	0.1	0.12360	0.9923	0.12456	80.28	0.9
0.2	0.5582	0.9984	0.5591	170.886	0.8	0.2	0.12533	0.9921	0.12633	70.916	0.8
0.3	0.5756	0.9983	0.5766	170.343	0.7	0.3	0.12706	0.9919	0.12810	70.806	0.7
0.4	0.5931	0.9982	0.5941	160.832	0.6	0.4	0.12880	0.9917	0.12988	70.70	0.6
0.5	0.6105	0.9981	0.6116	160.350	0.5	0.5	0.13053	0.9914	0.13165	70.596	0.5
0.6	0.6279	0.9980	0.6291	150.895	0.4	0.6	0.13226	0.9912	0.13343	70.495	0.4
0.7	0.6453	0.9979	0.6467	150.464	0.3	0.7	0.13399	0.9910	0.13521	70.396	0.3
0.8	0.6627	0.9978	0.6642	150.56	0.2	0.8	0.13572	0.9907	0.13698	70.30	0.2
0.9	0.6802	0.9977	0.6817	140.669	0.1	0.9	0.13744	0.9905	0.13876	70.207	0.1
	cos	sin	cot	tan	deg		cos	sin	cot	tan	

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
8.0	0.13917	0.9903	0.14054	70.115	82.0	12.0	0.2079	0.9781	0.2126	40.705	78.0
0.1	0.14090	0.990	0.14232	70.26	0.9	0.1	0.2096	0.9778	0.2144	40.665	0.9
0.2	0.14263	0.9898	0.14410	60.940	0.8	0.2	0.2133	0.9774	0.2162	40.625	0.8
0.3	0.14436	0.9895	0.14588	60.855	0.7	0.3	0.2130	0.9770	0.2180	40.586	0.7
0.4	0.14608	0.9893	0.14767	60.772	0.6	0.4	0.2147	0.9767	0.2199	40.548	0.6
0.5	0.14781	0.9890	0.14945	60.691	0.5	0.5	0.2164	0.9763	0.2217	40.511	0.5
0.6	0.14954	0.9888	0.15124	60.612	0.4	0.6	0.2181	0.9759	0.2235	40.474	0.4
0.7	0.15126	0.9885	0.15302	60.535	0.3	0.7	0.2198	0.9755	0.2254	40.437	0.3
0.8	0.15299	0.9882	0.15481	60.460	0.2	0.8	0.2215	0.9751	0.2272	40.402	0.2
0.9	0.15471	0.9880	0.15660	60.386	0.1	0.9	0.2233	0.9748	0.2290	40.366	0.1
9.0	0.15643	0.9877	0.15836	60.314	81.0	13.0	0.2250	0.9744	0.2309	40.331	77.0
0.1	0.15816	0.9874	0.16017	60.243	0.9	0.1	0.2267	0.9740	0.2327	40.297	0.9
0.2	0.15988	0.9871	0.16196	60.174	0.8	0.2	0.2284	0.9736	0.2345	40.264	0.8
0.3	0.16160	0.9869	0.16376	60.107	0.7	0.3	0.230	0.9732	0.2364	40.230	0.7
0.4	0.16333	0.9866	0.16555	60.41	0.6	0.4	0.2317	0.9728	0.2382	40.198	0.6
0.5	0.16505	0.9863	0.16734	50.976	0.5	0.5	0.2334	0.9724	0.2401	40.165	0.5
0.6	0.16677	0.9860	0.16914	50.912	0.4	0.6	0.2351	0.9720	0.2419	40.134	0.4
0.7	0.16849	0.9857	0.17093	50.850	0.3	0.7	0.2368	0.9715	0.2438	40.102	0.3
0.8	0.17021	0.9854	0.17273	50.789	0.2	0.8	0.2385	0.9711	0.2456	40.71	0.2
0.9	0.17193	0.9851	0.17453	50.730	0.1	0.9	0.2402	0.9707	0.2475	40.41	0.1
10.0	0.1736	0.9848	0.1763	50.671	80.0	14.0	0.2419	0.9703	0.2493	40.11	76.0
0.1	0.1754	0.9845	0.1781	50.614	0.9	0.1	0.2436	0.9699	0.2512	30.981	0.9
0.2	0.1771	0.9842	0.1799	50.558	0.8	0.2	0.2453	0.9694	0.2530	30.952	0.8
0.3	0.1788	0.9839	0.1817	50.503	0.7	0.3	0.2470	0.9680	0.2549	30.923	0.7
0.4	0.1805	0.9836	0.1835	50.449	0.6	0.4	0.2487	0.9686	0.2568	30.895	0.6
0.5	0.1822	0.9833	0.1853	50.396	0.5	0.5	0.2504	0.9681	0.2586	30.867	0.5
0.6	0.1840	0.9829	0.1871	50.343	0.4	0.6	0.2521	0.9677	0.2605	30.839	0.4
0.7	0.1857	0.9826	0.1890	50.292	0.3	0.7	0.2538	0.9673	0.2623	30.812	0.3
0.8	0.1874	0.9823	0.1908	50.242	0.2	0.8	0.2554	0.9668	0.2642	30.785	0.2
0.9	0.1891	0.9820	0.1926	50.193	0.1	0.9	0.2571	0.9664	0.2661	30.758	0.1
11.0	0.1908	0.9816	0.1944	50.145	79.0	15.0	0.2588	0.9659	0.2679	30.732	75.0
0.1	0.1925	0.9813	0.1962	50.97	0.9	0.1	0.2605	0.9655	0.2698	30.706	0.9
0.2	0.1942	0.9810	0.1980	50.50	0.8	0.2	0.2622	0.9650	0.2717	30.681	0.8
0.3	0.1959	0.9806	0.1998	50.5	0.7	0.3	0.2639	0.9646	0.2736	30.655	0.7
0.4	0.1977	0.9803	0.2016	40.959	0.6	0.4	0.2656	0.9641	0.2754	30.630	0.6
0.5	0.1994	0.9799	0.2035	40.915	0.5	0.5	0.2672	0.9636	0.2773	30.606	0.5
0.6	0.2011	0.9796	0.2053	40.872	0.4	0.6	0.2689	0.9632	0.2792	30.582	0.4
0.7	0.2028	0.9792	0.2071	40.829	0.3	0.7	0.2706	0.9627	0.2811	30.558	0.3
0.8	0.2045	0.9789	0.2089	40.787	0.2	0.8	0.2723	0.9622	0.2830	30.534	0.2
0.9	0.2062	0.9785	0.2107	40.745	0.1	0.9	0.2740	0.9617	0.2849	30.511	0.2
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
16.0	0.2756	0.9613	0.2867	30.487	74.0	20.0	0.3420	0.9397	0.3640	20.747	70.0
0.1	0.2773	0.9608	0.2886	30.465	0.9	0.1	0.3437	0.9391	0.3659	20.733	0.9
0.2	0.2790	0.9603	0.2905	30.442	0.8	0.2	0.3453	0.9385	0.3679	20.718	0.8
0.3	0.2807	0.9598	0.2924	30.420	0.7	0.3	0.3469	0.9379	0.3699	20.703	0.7
0.4	0.2823	0.9593	0.2943	30.398	0.6	0.4	0.3486	0.9373	0.3719	20.689	0.6
0.5	0.2840	0.9588	0.2962	30.376	0.5	0.5	0.3502	0.9367	0.3739	20.675	0.5
0.6	0.2857	0.9583	0.2981	30.354	0.4	0.6	0.3518	0.9361	0.3759	20.660	0.4
0.7	0.2874	0.9578	0.300	30.333	0.3	0.7	0.3535	0.9354	0.3779	20.646	0.3
0.8	0.2890	0.9573	0.3019	30.312	0.2	0.8	0.3551	0.9348	0.3799	20.633	0.2
0.9	0.2907	0.9568	0.3038	30.291	0.1	0.9	0.3567	0.9342	0.3819	20.619	0.1
17.0	0.2924	0.9563	0.3067	30.271	73.0	21.0	0.3584	0.9336	0.3839	20.605	69.0
0.1	0.2940	0.9558	0.3076	30.271	0.9	0.1	0.360	0.9330	0.3859	20.592	0.9
0.2	0.2957	0.9553	0.3096	30.230	0.8	0.2	0.3616	0.9323	0.3879	20.578	0.8
0.3	0.2974	0.9548	0.3115	30.211	0.7	0.3	0.3633	0.9317	0.3899	20.565	0.7
0.4	0.2990	0.9542	0.3134	30.191	0.6	0.4	0.3649	0.9311	0.3919	20.552	0.6
0.5	0.307	0.9537	0.3153	30.172	0.5	0.5	0.3665	0.9304	0.3939	20.539	0.5
0.6	0.3024	0.9532	0.3172	30.152	0.4	0.6	0.3681	0.9298	0.3959	20.526	0.4
0.7	0.3040	0.9527	0.3191	30.133	0.3	0.7	0.3697	0.9291	0.3979	20.513	0.3
0.8	0.3057	0.9521	0.3211	30.115	0.2	0.8	0.3714	0.9285	0.400	20.50	0.2
0.9	0.3074	0.9516	0.3230	30.96	0.1	0.9	0.3730	0.9278	0.4020	20.488	0.1
18.0	0.3090	0.9511	0.3249	30.78	72.0	22.0	0.3746	0.9272	0.4040	20.475	68.0
0.1	0.3107	0.9505	0.3269	30.60	0.9	0.1	0.3762	0.9265	0.4061	20.463	0.9
0.2	0.3123	0.950	0.3288	30.42	0.8	0.2	0.3778	0.9259	0.4081	20.450	0.8
0.3	0.3140	0.9494	0.3307	30.24	0.7	0.3	0.3795	0.9252	0.4101	20.438	0.7
0.4	0.3156	0.9489	0.3327	30.6	0.6	0.4	0.3811	0.9245	0.4122	20.426	0.6
0.5	0.3173	0.9483	0.3346	20.989	0.5	0.5	0.3727	0.9239	0.4142	20.414	0.5
0.6	0.3190	0.9478	0.3365	20.971	0.4	0.6	0.3843	0.9232	0.4163	20.402	0.4
0.7	0.3206	0.9472	0.3385	20.954	0.3	0.7	0.3859	0.9225	0.4183	20.391	0.3
0.8	0.3223	0.9466	0.3404	20.937	0.2	0.8	0.3875	0.9219	0.4204	20.379	0.2
0.9	0.3239	0.9461	0.3424	20.921	0.1	0.9	0.3891	0.9212	0.4224	20.367	0.1
19.0	0.3256	0.9455	0.3443	20.904	71.0	23.0	0.3907	0.9205	0.4245	20.356	67.0
0.1	0.3272	0.9449	0.3463	20.888	0.9	0.1	0.3923	0.9198	0.4265	20.344	0.9
0.2	0.3289	0.9444	0.3482	20.872	0.8	0.2	0.3939	0.9191	0.4286	20.333	0.8
0.3	0.3305	0.9438	0.3502	20.856	0.7	0.3	0.3955	0.9184	0.4307	20.322	0.7
0.4	0.3322	0.9432	0.3522	20.840	0.6	0.4	0.3971	0.9178	0.4327	20.311	0.6
0.5	0.3338	0.9426	0.3541	20.824	0.5	0.5	0.3987	0.9171	0.4348	20.30	0.5
0.6	0.3355	0.9421	0.3561	20.808	0.4	0.6	0.403	0.9164	0.4369	20.289	0.4
0.7	0.3371	0.9415	0.3581	20.793	0.3	0.7	0.4019	0.9157	0.4390	20.278	0.3
0.8	0.3387	0.9409	0.360	20.778	0.2	0.8	0.4035	0.9150	0.4411	20.267	0.2
0.9	0.3403	0.9403	0.3620	20.762	0.1	0.9	0.4051	0.9143	0.4431	20.257	0.1
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
24.0	0.4067	0.9135	0.4452	20.246	66.0	28.0	0.4695	0.8829	0.5317	10.881	62.0
0.1	0.4083	0.9128	0.4473	20.236	0.9	0.1	0.4710	0.8821	0.5340	10.873	0.9
0.2	0.4099	0.9121	0.4494	20.225	0.8	0.2	0.4726	0.8813	0.5362	10.865	0.8
0.3	0.4115	0.9114	0.4515	20.215	0.7	0.3	0.4741	0.8805	0.5384	10.857	0.7
0.4	0.4131	0.9107	0.4536	20.204	0.6	0.4	0.4756	0.8796	0.5407	10.849	0.6
0.5	0.4147	0.910	0.4557	20.194	0.5	0.5	0.4772	0.8788	0.5430	10.842	0.5
0.6	0.4163	0.9092	0.4578	20.184	0.4	0.6	0.4787	0.8780	0.5452	10.834	0.4
0.7	0.4179	0.9085	0.4599	20.174	0.3	0.7	0.4802	0.8771	0.5475	10.827	0.3
0.8	0.4195	0.9078	0.4621	20.164	0.2	0.8	0.4818	0.8763	0.5498	10.819	0.2
0.9	0.4210	0.9070	0.4642	20.154	0.1	0.9	0.4833	0.8755	0.5520	10.811	0.1
25.0	0.4226	0.9063	0.4663	20.145	65.0	29.0	0.4848	0.8746	0.5543	10.804	61.0
0.1	0.4242	0.9056	0.4684	20.135	0.9	0.1	0.4863	0.8738	0.5566	10.797	0.9
0.2	0.4258	0.9048	0.4706	20.125	0.8	0.2	0.4879	0.8729	0.5589	10.789	0.8
0.3	0.4274	0.9041	0.4727	20.116	0.7	0.3	0.4894	0.8721	0.5612	10.782	0.7
0.4	0.4289	0.9033	0.4748	20.106	0.6	0.4	0.4909	0.8712	0.5635	10.775	0.6
0.5	0.4305	0.9028	0.4770	20.97	0.5	0.5	0.4924	0.8704	0.5658	10.767	0.5
0.6	0.4321	0.9018	0.4791	20.87	0.4	0.6	0.4939	0.8695	0.5681	10.760	0.4
0.7	0.4337	0.9011	0.4813	20.78	0.3	0.7	0.4955	0.8686	0.5704	10.753	0.3
0.8	0.4352	0.903	0.4834	20.69	0.2	0.8	0.4970	0.8678	0.5726	10.746	0.2
0.9	0.4368	0.8996	0.4856	20.59	0.1	0.9	0.4985	0.8669	0.5750	10.739	0.1
26.0	0.4384	0.8988	0.4877	20.50	64.0	30.0	0.500	0.8660	0.5774	10.7321	60.0
0.1	0.4399	0.8980	0.4899	20.41	0.9	0.1	0.5015	0.8652	0.5797	10.7251	0.9
0.2	0.4415	0.8973	0.4921	20.32	0.8	0.2	0.5030	0.8643	0.5820	10.7162	0.8
0.3	0.4431	0.8965	0.4942	20.23	0.7	0.3	0.5045	0.8634	0.5844	10.7113	0.7
0.4	0.4446	0.8957	0.4964	20.14	0.6	0.4	0.5040	0.8625	0.5867	10.7045	0.6
0.5	0.4462	0.8949	0.4986	20.6	0.5	0.5	0.5075	0.8616	0.5890	10.6977	0.5
0.6	0.4478	0.8942	0.508	10.997	0.4	0.6	0.5090	0.8607	0.5914	10.6909	0.4
0.7	0.4493	0.8934	0.5029	10.988	0.3	0.7	0.5105	0.8599	0.5938	10.6842	0.3
0.8	0.4509	0.8926	0.5051	10.980	0.2	0.8	0.5120	0.8590	0.5961	10.6715	0.2
0.9	0.4524	0.8918	0.5073	10.971	0.1	0.9	0.5135	0.8581	0.5985	10.6709	0.1
27.0	0.4540	0.8910	0.5095	10.963	63.0	31.0	0.5150	0.8572	0.609	10.6643	59.0
0.1	0.4555	0.8902	0.5117	10.954	0.9	0.1	0.5165	0.8643	0.6032	10.6577	0.9
0.2	0.4571	0.8894	0.5139	10.946	0.8	0.2	0.5180	0.8554	0.6056	10.6512	0.8
0.3	0.4586	0.8886	0.5161	10.937	0.7	0.3	0.5195	0.8545	0.6080	10.6447	0.7
0.4	0.4602	0.8878	0.5184	10.929	0.6	0.4	0.5210	0.8536	0.6104	10.6383	0.6
0.5	0.4617	0.8870	0.5206	10.921	0.5	0.5	0.5225	0.8526	0.6128	10.6319	0.5
0.6	0.4633	0.8862	0.5228	10.913	0.4	0.6	0.5240	0.8517	0.6152	10.6255	0.4
0.7	0.4648	0.8854	0.5250	10.905	0.3	0.7	0.5255	0.8508	0.6176	10.6191	0.3
0.8	0.4664	0.8846	0.5272	10.897	0.2	0.8	0.5270	0.8499	0.620	10.6128	0.2
0.9	0.4679	0.8838	0.5295	10.889	0.1	0.9	0.5284	0.8490	0.6224	10.6066	0.1
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
32.0	0.5299	0.8480	0.6249	10.603	58.0	36.0	0.5878	0.8090	0.7265	10.3764	54.0
0.1	0.5314	0.8471	0.6273	10.5941	0.9	0.1	0.5892	0.8080	0.7292	10.3713	0.9
0.2	0.5329	0.8462	0.6297	10.5880	0.8	0.2	0.5906	0.8070	0.7319	10.3663	0.8
0.3	0.5344	0.8453	0.6322	10.5818	0.7	0.3	0.5920	0.8059	0.7346	10.3613	0.7
0.4	0.5358	0.8443	0.6346	10.5757	0.6	0.4	0.5934	0.8049	0.7373	10.3564	0.6
0.5	0.5373	0.8434	0.6371	10.5697	0.5	0.5	0.5948	0.8039	0.740	10.3514	0.5
0.6	0.5388	0.8425	0.6395	10.5637	0.4	0.6	0.5962	0.8028	0.7427	10.3465	0.4
0.7	0.5402	0.8415	0.6420	10.5577	0.3	0.7	0.5976	0.8018	0.7454	10.3416	0.3
0.8	0.5417	0.8406	0.6445	10.5517	0.2	0.8	0.5990	0.807	0.7481	10.3367	0.2
0.9	0.5432	0.8396	0.6469	10.5458	0.1	0.9	0.604	0.7997	0.7508	10.3319	0.1
33.0	0.5446	0.8387	0.6494	10.5399	57.0	37.0	0.6018	0.7986	0.7536	10.3270	53.0
0.1	0.5461	0.8377	0.6519	10.5340	0.9	0.1	0.6032	0.7976	0.7563	10.3222	0.9
0.2	0.5476	0.8368	0.6544	10.5282	0.8	0.2	0.6046	0.7965	0.7590	10.3175	0.8
0.3	0.5490	0.8358	0.6569	10.5224	0.7	0.3	0.6060	0.7955	0.7518	10.3127	0.7
0.4	0.5505	0.8348	0.6594	10.5166	0.6	0.4	0.6074	0.7944	0.7646	10.3079	0.6
0.5	0.5519	0.8339	0.6619	10.5108	0.5	0.5	0.6088	0.7934	0.7673	10.3032	0.5
0.6	0.5534	0.8329	0.6644	10.5051	0.4	0.6	0.6101	0.7923	0.7701	10.2985	0.4
0.7	0.5548	0.8320	0.6669	10.4994	0.3	0.7	0.6115	0.7912	0.7729	10.2938	0.3
0.8	0.5563	0.8310	0.6694	10.4938	0.2	0.8	0.6129	0.7902	0.7757	10.2892	0.2
0.9	0.5577	0.830	0.6720	10.4882	0.1	0.9	0.6143	0.7891	0.7785	10.2846	0.1
34.0	0.5592	0.8290	0.6745	10.4826	56.0	38.0	0.6157	0.7880	0.7813	10.2799	52.0
0.1	0.5606	0.8281	0.6771	10.4770	0.9	0.1	0.6170	0.7869	0.7841	10.2753	0.9
0.2	0.5621	0.8271	0.6796	10.4715	0.8	0.2	0.6184	0.7859	0.7869	10.2708	0.8
0.3	0.5635	0.8261	0.6822	10.4659	0.7	0.3	0.6198	0.7848	0.7898	10.2662	0.7
0.4	0.5650	0.8251	0.6847	10.4605	0.6	0.4	0.6211	0.7837	0.7926	10.2617	0.6
0.5	0.5664	0.8241	0.6873	10.4550	0.5	0.5	0.6225	0.7826	0.7954	10.2572	0.5
0.6	0.5678	0.8231	0.6899	10.4496	0.4	0.6	0.6239	0.7815	0.7983	10.2527	0.4
0.7	0.5693	0.8221	0.6924	10.4442	0.3	0.7	0.6252	0.7804	0.8012	10.2482	0.3
0.8	0.5707	0.8211	0.6950	10.4388	0.2	0.8	0.6266	0.7793	0.8040	10.2437	0.2
0.9	0.5721	0.8202	0.6970	10.4335	0.1	0.9	0.6280	0.7782	0.8069	10.2393	0.1
35.0	0.5736	0.8192	0.702	10.4281	55.0	39.0	0.6293	0.7771	0.8098	10.2349	51.0
0.1	0.5750	0.8181	0.7028	10.4229	0.9	0.1	0.6307	0.7760	0.8127	10.2305	0.9
0.2	0.5764	0.8171	0.7054	10.4176	0.8	0.2	0.6320	0.7749	0.8156	10.2261	0.8
0.3	0.5779	0.8161	0.7080	10.4124	0.7	0.3	0.6334	0.7738	0.8185	10.2218	0.7
0.4	0.5793	0.8151	0.7107	10.4071	0.6	0.4	0.6347	0.7727	0.8214	10.2174	0.6
0.5	0.5807	0.8141	0.7133	10.4019	0.5	0.5	0.6361	0.7716	0.8243	10.2131	0.5
0.6	0.5821	0.8131	0.7159	10.3968	0.4	0.6	0.6474	0.7705	0.8273	10.2088	0.4
0.7	0.5835	0.8121	0.7186	10.3916	0.3	0.7	0.6388	0.7694	0.8302	10.2045	0.3
0.8	0.5850	0.8111	0.7212	10.3865	0.2	0.8	0.6401	0.7683	0.8332	10.202	0.2
0.9	0.5864	0.810	0.7239	10.3814	0.1	0.9	0.6414	0.7672	0.8361	10.1960	0.1
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
40.0	0.6428	0.7660	0.8291	10.1918	50.0	43.0	0.6820	0.7314	0.9325	10.724	47.0
0.1	0.6441	0.7649	0.8421	10.1875	0.9	0.1	0.6833	0.7302	0.9358	10.686	0.9
0.2	0.6455	0.7638	0.8451	10.1833	0.8	0.2	0.6845	0.7290	0.9391	10.649	0.8
0.3	0.6468	0.7627	0.8481	10.1792	0.7	0.3	0.6858	0.7278	0.9424	10.612	0.7
0.4	0.6481	0.7615	0.8511	10.1750	0.6	0.4	0.6871	0.7266	0.9457	10.575	0.6
0.5	0.6494	0.7604	0.8541	10.1708	0.5	0.5	0.6884	0.7254	0.9490	10.538	0.5
0.6	0.6508	0.7593	0.8571	10.1667	0.4	0.6	0.6896	0.7242	0.9523	10.501	0.4
0.7	0.6521	0.7581	0.8601	10.1626	0.3	0.7	0.6909	0.7230	0.9556	10.464	0.3
0.8	0.6534	0.7570	0.8632	10.1585	0.2	0.8	0.6921	0.7218	0.9590	10.428	0.2
0.9	0.6547	0.7559	0.8662	10.1544	0.1	0.9	0.6934	0.7206	0.9623	10.392	0.1
41.0	0.6561	0.7547	0.8693	10.1504	49.0	44.0	0.6947	0.7193	0.9657	10.355	46.0
0.1	0.6574	0.7536	0.8724	10.1463	0.9	0.1	0.6959	0.7181	0.9691	10.319	0.9
0.2	0.6587	0.7524	0.8754	10.1423	0.8	0.2	0.6972	0.7169	0.9725	10.283	0.8
0.3	0.660	0.7513	0.8785	10.1383	0.7	0.3	0.6984	0.7157	0.9759	10.247	0.7
0.4	0.6613	0.7501	0.8816	10.1343	0.6	0.4	0.6997	0.7145	0.9793	10.212	0.6
0.5	0.6626	0.7490	0.8847	10.1303	0.5	0.5	0.709	0.7133	0.9827	10.176	0.5
0.6	0.6639	0.7478	0.8878	10.1263	0.4	0.6	0.7022	0.7120	0.9861	10.141	0.4
0.7	0.6652	0.7466	0.8910	10.1224	0.3	0.7	0.7034	0.7108	0.9896	10.105	0.3
0.8	0.6665	0.7455	0.8941	10.1184	0.2	0.8	0.6794	0.7337	0.9260	10.799	0.2
0.9	0.6678	0.7443	0.8972	10.1145	0.1	0.9	0.6807	0.7325	0.9293	10.761	0.1
42.0	0.6691	0.7431	0.904	10.1106	48.0	45.0	0.7071	0.7071	1.0000	1.0000	45.0
0.1	0.6704	0.7420	0.9036	10.1067	0.9						
0.2	0.6717	0.7408	0.9067	10.1028	0.8						
0.3	0.6730	0.7396	0.9099	10.990	0.7						
0.4	0.6743	0.7385	0.9131	10.951	0.6						
0.5	0.6756	0.7373	0.9163	10.913	0.5						
0.6	0.6769	0.7361	0.9195	10.875	0.4						
0.7	0.6782	0.7349	0.9228	10.837	0.3						
0.8	0.6794	0.7337	0.9260	10.799	0.2						
0.9	0.6807	0.7325	0.9293	10.761	0.1						
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

Assignment Questions

Information: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Concepts of Alternating Current," pages 1-1 through 1-33.

1-1. Alternating current can be defined as current that varies in

1. amplitude and direction
2. magnitude and phase
3. amplitude and time
4. time and phase

1-2. Before a 120-volt dc source can be used to power a 12-volt load, the voltage must be reduced. Which of the following methods can be used?

1. A resistor placed in parallel with the load
2. A resistor placed in series with the load
3. A step-down transformer placed in series with load
4. A step-down transformer placed in parallel-with the load

1-3. Alternating current has replaced direct current in modern transmission systems because it has which of the following advantages?

1. Ac can be transmitted with no line loss
2. Ac can be transmitted at higher current levels
3. Ac can be transmitted at lower voltage levels
4. Ac can be readily stepped up or down

1-4. A waveform is a graphic plot of what quantities?

1. Current versus time
2. Amplitude versus time
3. Voltage versus amplitude
4. Magnitude versus amplitude

1-5. Which of the following properties surrounds a current-carrying conductor?

1. A magnetic field
2. A repulsive force
3. An attractive force
4. An electrostatic field

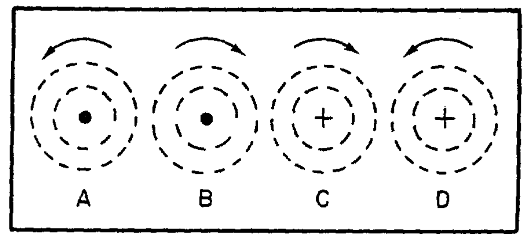


Figure 1A.—Conductors, cross-sectional view.

IN ANSWERING QUESTIONS 1-6 AND 1-7, REFER TO FIGURE 1A.

1-6. The direction of the magnetic field is correctly depicted by which of the followings

1. A and B
2. B and D
3. A and C
4. B and C

IN ANSWERING QUESTION 1-7, REFER TO FIGURE 1A AND IGNORE THE MAGNETIC FIELD ARROWS SHOWN IN THE FIGURE.

1-7. In which conductors will the magnetic fields (a) aid, and (b) oppose each other?

1. (a) A and C, (b) C and D
2. (a) A and D, (b) B and C
3. (a) A and C, (b) B and D
4. (a) A and B, (b) A and D

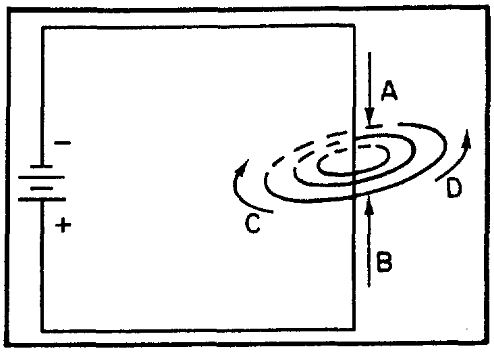


Figure 1B.—A magnetic field surrounding a current-carrying conductor.

IN ANSWERING QUESTIONS 1-8, REFER TO FIGURE 1B.

1-8. In figure 1B, the direction of the magnetic field surrounding the conductor is correctly indicated by what arrow?

1. A
2. B
3. C
4. D

1-9. Which of the following statements accurately describes the magnetic field surrounding a current-carrying conductor?

1. It is parallel to and equal along all parts of the conductor
2. It is parallel to and maximum at the most negative part of the conductor
3. It is perpendicular to and equal along all parts of the conductor
4. It is perpendicular to the conductor and maximum at the most negative point of the conductor

1-10. Which of the following factors determine(s) the intensity of a magnetic field surrounding a coil?

1. The amount of current flow through the coil
2. The type of core material
3. The number of turns in the conductor
4. All of the above

1-11. When you grasp a coil in your left hand with your thumb pointing in the direction of the north pole, your fingers will be wrapped around the coil in the direction of the

1. voltage potential
2. magnetic field
3. current flow
4. south pole

1-12. The power consumed in a conductor in realigning the atoms which set up the magnetic field is known as what type of loss?

1. Hysteresis loss
2. Magnetic loss
3. Field loss
4. Heat loss

1-13. The magnetic field surrounding a straight conductor is (a) what shape, and (b) is in what position relative to the conductor?

1. (a) Linked oblong
   (b) Parallel
2. (a) Concentric circles
   (b) Parallel
3. (a) Linked oblong
   (b) Perpendicular
4. (a) Concentric circles
   (b) Perpendicular

1-14. Why is a two-pole magnetic field set up around a coil?

1. Because separate lines of magnetic force link and combine their effects
2. Because concentric lines of force cross at right angles and combine.
3. Because lines of force are separated and bent at the coil ends
4. Because separate lines of force are attracted to the two poles of the coi

1-15. When a conductor is moving parallel to magnetic lines of force, (a) what relative number of magnetic lines are cut, and (b) what relative value of emf is induced?

1. (a) Minimum, (b) maximum
2. (a) Minimum, (b) minimum
3. (a) Maximum, (b) maximum
4. (a) Maximum, (b) minimum

1-16. When the induced voltage in a conductor rotating in a magnetic field is plotted against the degrees of rotation, the plot will take what shape?

1. A circle
2. A sine curve
3. A square wave
4. A straight line

1-17. When a loop of wire is rotated through 360º in a magnetic field, the induced voltage will be zero at which of the following points?

1. 45º
2. 90º
3. 180º
4. 270º

1-18. When a loop of wire is rotated 360º in a magnetic field, at what points will the induced voltage reach its maximum (a) positive, and (b) negative values?

1. (a) 0º , (b) 180º
2. (a) 0º , (b) 270º
3. (a) 90º , (b) 180º
4. (a) 90º , (b) 270º

1-19. When a coil of wire makes eight complete revolutions through a single magnetic field, (a) what total number of alternations of voltage will be generated and, (b) what total number of cycles of ac will be generated?

1. (a) 32, (b) 16
2. (a) 16, (b) 8
3. (a) 8, (b) 4
4. (a) 4, (b) 2

1-20. According to the left-hand rule for generators, when your thumb points in the direction of rotation, your (a) forefinger and (b) your middle finger will indicate the relative directions of what quantities?

1. (a) Current,
   (b) Magnetic flux, south to north
2. (a) Current,
   (b) Magnetic flux, north to south
3. (a) Magnetic flux, south to north,
   (b) Current
4. (a) Magnetic flux, north to south,
   (b) Current

1-21. Continuous rotation of a conductor through magnetic lines of force will produce what type of (a) voltage and (b) waveform?

1. (a) Ac, (b) sine wave
2. (a) Dc, (b) continuous level
3. (a) Ac, (b) sawtooth
4. (a) Dc, (b) pulsating wave

1-22. What is the term for the number of complete cycles of ac produced in one second?

1. Period
2. Waveform
3. Frequency
4. Wavelength

1-23. What is the unit of measurement for frequency?

1. Cycle
2. Hertz
3. Period
4. Maxwell

1-24. A loop of wire rotating at 60 rpm in a magnetic field will produce an ac voltage of what frequency?

1. 1 Hz
2. 60 Hz
3. 120 Hz
4. 360 Hz

1-25. An ac voltage of 250 hertz has a period of

1. 0.004 second
2. 0.025 second
3. 0.4 second
4. 2.5 seconds

1-26. What is the approximate frequency of an ac voltage that has a period of .0006 second?

1. 6 Hz
2. 16.67 Hz
3. 600 Hz
4. 1667 Hz

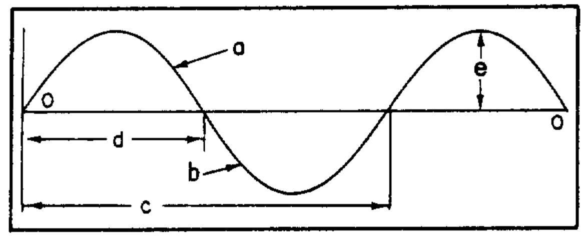


Figure 1C.—Components of a sinewave.

IN ANSWERING QUESTIONS 1-27 THROUGH 1-36, REFER TO FIGURE 1C. IN ANSWERING QUESTIONS 1-27 THROUGH 1-31, SELECT FROM COLUMN B THE COMPONENT THAT IS DESCRIBED BY THE TERM IN COLUMN A.

	A. TERM	B. COMPONENT
1-27.	Period	1. a
1-28.	Negative alternation	2. b
		3. c
1-29.	Wavelength	4. d
1-30.	One-half cycle	
1-31.	Positive alternation	

1-32. Component a is a measure of what quantity?

1. Frequency
2. Polarity
3. Amplitude
4. Time

1-33. Component a differs from component b in which of the following characteristics

1. Frequency
2. Polarity
3. Amplitude
4. Period

1-34. Component c would represent what quantities if it were expressed as (a) physical distance, and (b) time?

1. (a) Frequency (b) period
2. (a) Period (b) wavelength
3. (a) Frequency (b) wavelength
4. (a) Wavelength (b) period

1-35. The combined values of components a and b represent what ac value?

1. Peak-to-peak value
2. Average value
3. Effective value
4. Instantaneous value

1-36. A peak voltage is represented by which of the following components?

1. a
2. c
3. d
4. e

1-37. An ac voltage has a frequency of 350 Hz. In two seconds, what total number of times will the peak value of voltage be generated?

1. 350 times
2. 700 times
3. 1400 times
4. 2800 times

1-38. The value of current of an ac waveform taken at any particular moment of time is what type of value?

1. Average value
2. Effective value
3. Instantaneous value
4. Peak-to-peak value

1-39. While the value of an ac voltage may be expressed as one of several values, the accepted practice is to express it as what type value?

1. Average value
2. Instantaneous value
3. Peak-to-peak value
4. Effective value

1-40. The total of ten instantaneous values of an alternation divided by ten is equal to what value?

1. The peak value
2. The average value
3. The instantaneous value
4. The effective value

1-41. Which of the following mathematical formulas is used to find the average value of voltage for an ac voltage?

1. $E_{avg} = 0.707 \times E_{max}$
2. $E_{avg} = 1.414 \times E_{eff}$
3. $E_{avg} = 0.636 \times E_{max}$
4. $E_{avg} = 0.226 \times E_{eff}$

1-42. What is the average value of all of the instantaneous voltages occurring during one cycle of an ac waveform with a peak value of 60 volts?

1. 0 volts
2. 38 volts
3. 76 volts
4. 128 volts

1-43. If an ac voltage has an E_{max} of 220 volts, what is E_{avg}?

1. 50 volts
2. 140 volts
3. 156 volts
4. 311 volts

1-44. If an ac waveform has a peak-to-peak value of 28 volts, what is E_{avg}?

1. 40 volts
2. 20 volts
3. 18 volts
4. 9 volts

1-45. If an ac waveform has a peak value of 4.5 amperes, what is its average value?

1. 2.9 amperes
2. 3.2 amperes
3. 5.7 amperes
4. 6.4 amperes

1-46. If the average value of current of an ac waveform is 1.2 amperes, what is its maximum value of current?

1. 0.8 amperes
2. 0.9 amperes
3. 1.7 amperes
4. 1.9 amperes

1-47. The value of alternating current that will heat a resistor to the same temperature as an equal value of direct current is known as

1. I_{avg}
2. I_{eff}
3. I_{in}
4. I_{max}

1-48. The rms value for an ac voltage is equal to what other ac value?

1. E_{avg}
2. E_{max}
3. E_{eff}
4. E_{in}

1-49. What value will result by squaring all values for E_{inst}, averaging these values, and then taking the square root of that average?

1. E_{avg}
2. E_{max}
3. E_{eff}
4. E_{in}

1-50. The accepted, nominal value for household power in the United States is 120-volts, 60 Hz. What is the value of maximum voltage?

1. 170 volts
2. 120 volts
3. 85 volts
4. 76 volts

1-51. An ac voltmeter is usually calibrated to read which of the following ac values?

1. Average
2. Effective
3. Peak
4. Peak-to-peak

1-52. If the maximum value for an ac voltage is known, the E_{eff} can be found by using which of the following formulas?

1. $E_{eff} = E_{max}/.636$
2. $E_{eff} = E_{max}/.707$
3. $E_{eff} = E_{max} \times .707$
4. $E_{eff} = E_{max} \times 1.414$

1-53. If the I_{eff} of an ac waveform is 3.25 amperes, what is I_{max}?

1. 4.6 amperes
2. 2.3 amperes
3. 2.1 amperes
4. 1.6 amperes

1-54. If the rms value of the voltage of an ac waveform is 12.4 volt what is its average value? (Hint: compute E_{max} first.)

1. 8 volts
2. 11 volts
3. 15 volts
4. 18 volts

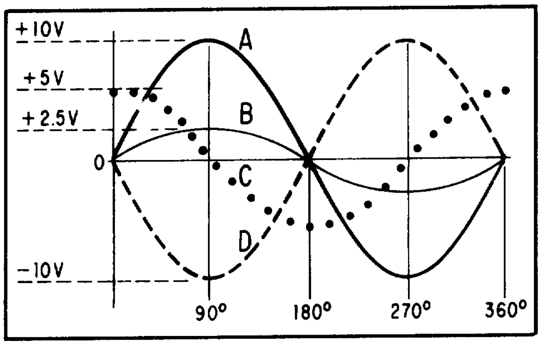


Figure 1D.—Phase relationship of sinewaves.

IN ANSWERING QUESTIONS 1-55 THROUGH 1-60, REFER TO FIGURE 1D.

1-55. What two waveforms are in phase?

1. A and B
2. A and C
3. C and D
4. B and C

1-56. What is the phase difference, if any, between waveform B and C?

1. B is 225° out of phase with C
2. B is 180° out of phase with C
3. B is 90° out of phase with C
4. None; they are in phase

1-57. What is the phase difference, if any, between waveform A and D?

1. A is 270° out of phase with D
2. A is 180° out of phase with D
3. A is 90° out of phase with D
4. None; they are in phase

1-58. If the voltage represented by waveform A is summed to the voltage represented by waveform D, what is the resultant voltage?

1. 20 volts
2. 15 volts
3. 10 volts
4. 0 volts

1-59. What is E_{in} at 90° that results from adding waveform B to waveform D?

1. +7.5 volts
2. +2.5 volts
3. -7.5 volts
4. -10 volts

1-60. What is the phase difference between waveform A and waveform C?

1. A lags C by 90°
2. A leads C by 90°
3. A leads C by 180°
4. A lags C by 180°

1-61. Which of the following is an important rule to remember when using Ohm's Law to solve ac circuit problems?

1. Always solve for resistance first
2. Give the answer as effective value
3. Never mix values
4. Convert all given values to effective before attempting to solve

1-62. An ac circuit is composed of three 20-ohm resistors connected in parallel. The average voltage supplied to this circuit is 62-volts ac. What is the maximum current?

1. 9.3 amperes
2. 14.6 amperes
3. 17.5 amperes
4. 22.5 amperes

1-63. If the ac source in question 1-62 is raised to an average value of 120 volts, what is the I_{eff}?

1. 11.48 amperes
2. 12.70 amperes
3. 20.01 amperes
4. 25.52 amperes

1-64. If E_{eff} is 150 volts and I_{max} is 4.5 amperes, what is the total resistance (R_T) of a circuit?

1. 21.2Ω
2. 23.6Ω
3. 33.3Ω
4. 47.1Ω

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Inductance," pages 2-1 through 2-27.

2-1. The property of inductance offers opposition to which of the following quantities?

1. Constant current
2. Constant voltage
3. Changes in current
4. Changes in voltage

2-2. What is the symbol for inductance?

1. L
2. H
3. X_L
4. IND

2-3. What is the unit of measurement for inductance?

1. Ohm
2. Rel
3. Farad
4. Henry

2-4. If 9 volts are induced in a conductor when the current changes by 4.5 amperes in one second, what is the total inductance of the circuit?

1. 1.5 henries
2. 2.0 henries
3. 13.5 henries
4. 40.0 henries

2-5. What physical property is similar to inductance?

1. Mass
2. Motion
3. Velocity
4. Inertia

2-6. The difference in potential across a resistor, created by current through the resistor is an example of which of the following forces?

1. Resistive
2. Inertia
3. Inductive
4. Electromotive

2-7. When a magnetic field moves through a stationary conductor, the electrons in orbit are affected in what manner?

1. They are dislodged from orbit
2. They move closer to their nucleus
3. They move closer to other orbiting electrons
4. They bunch up on one side of the nucleus

2-8. When electrons are moved in a conductor by a magnetic field, a force known by which of the following terms is created?

1. Voltage
2. Electromotive
3. Potential difference
4. All of the above

2-9. Self-induced emf is also known as what force?

1. magnetic force
2. Inertial force
3. Electromotive force
4. Counter electromotive force

2-10. According to Lenz's Law, the induced emf produced by a change in current in an inductive circuit tends to have what effect on the current?

1. It aids a rise in current and opposes fall in current
2. It aids a fall in current and opposes a rise in current
3. It opposes either a rise or a fall in current
4. It aids either a rise or fall in current

2-11. The direction of the induced voltage in an inductor may be found by application of which of the following rules?

1. The left-hand rule for inductors
2. The left-hand rule for generators
3. The right-hand rule for conductors
4. The right-hand rule for motors

2-12. The left-hand rule for generators states that the thumb of the left hand points in the direction of motion of the

1. conductor
2. magnetic field
3. generator poles
4. induced current

2-13. When source voltage is removed from a current-carrying conductor, a voltage will be induced in the conductor by which of the following actions?

1. The decreasing voltage
2. The collapsing magnetic field
3. The reversal of current
4. The reversing electrical field

2-14. The property of inductance is present in which of the following electrical circuits?

1. An ac circuit
2. A dc circuit
3. A resistive circuit
4. Each of the above

2-15. How are inductors classified?

1. By core type
2. By conductor type
3. By the number of turns
4. By the direction of the windings on the core

2-16. Normally, most coils have cores composed of either air or

1. copper
2. carbon
3. soft iron
4. carbon steel

2-17. The hollow form of nonmagnetic material found in the center of an air-core coil has what purpose?

1. To focus the magnetic flux
2. To support the windings
3. To act as a low resistance path for flux
4. To serve as a container for the core

2-18. Which of the following factors will NOT affect the value of inductance of a coil?

1. Number of coil turns
2. Diameter of the coil
3. Conductor tensility
4. Core materials used

2-19. When the number of turns is increased in a coil from 2 to 4, the total inductance will increase by a factor of

1. eight
2. two
3. six
4. four

2-20. Why do large diameter coils have greater inductance than smaller diameter coils, all other factors being the same?

1. Large diameter coils have more wire and thus more flux
2. Large diameter coils have less resistance
3. Small diameter coils have less resistance
4. Small diameter coils have large cemfs which oppose current flow

2-21. If the radius of a coil is doubled, its inductance is increased by what factor?

1. One
2. Two
3. Eight
4. Four

2-22. If the length of a coil is doubled while the number of turns is kept the same, this will have (a) what effect on inductance and (b) by what factor?

1. (a) Decrease, (b) by 1/4
2. (a) Decrease, (b) by 1/2
3. (a) Increase, (b) by 2 times
4. (a) Increase, (b) by 4 times

2-23. A soft iron core will increase inductance because it has which of the following characteristics?

1. Low permeability and low reluctance
2. Low permeability and high reluctance
3. High permeability and high reluctance
4. High permeability and low reluctance

2-24. An increase in the permeability of the core of a coil will increase which of the following coil characteristics?

1. Magnetic flux
2. Reluctance
3. Resistance
4. Conductance

2-25. If a coil is wound in layers, its inductance will be greater than that of a similar single-layer coil because of a higher

1. permeability
2. flux linkage
3. reluctance
4. conductance

2-26. Regardless of the method used, inductance of a coil can ONLY be increased by increasing what coil characteristic?

1. Transconductance
2. Reluctance
3. Flux linkage
4. Conductance

2-27. What is the symbol used to denote the basic unit of measurement for inductance?

1. L
2. H
3. I
4. F

2-28. What does the Greek letter Delta signify as in "ΔI" or "Δt"?

1. The values are constant
2. The values are average
3. The values are changing
4. The values are effective

2-29. An electrical circuit contains a coil. When the current varies 2.5 amperes in one second, 7.5 volts are induced in the coil. What is the value of inductance of the coil?

1. 1 henry
2. 2.2 henries
3. 3.0 henries
4. 4 henries

2-30. An ac electrical current varies 1.5 amperes in one second and is applied to a 10-henry coil. What is the value of the emf induced across the coil?

1. 1.0 volt
2. 1.5 volts
3. 11.5 volts
4. 15.0 volts

2-31. If a coil is rated at 10 henries, what is its value in (a) millihenries and (b) microhenries?

1. (a) 10,000 mH, (b) 10,000,000 μH
2. (a) 10,000 mH, (b) 1,000,000 μH
3. (a) 1,000 mH, (b) 1,000,000 μH
4. (a) 1,000 mH, (b) 100,000 μH

THIS SPACE LEFT BLANK INTENTIONALLY.

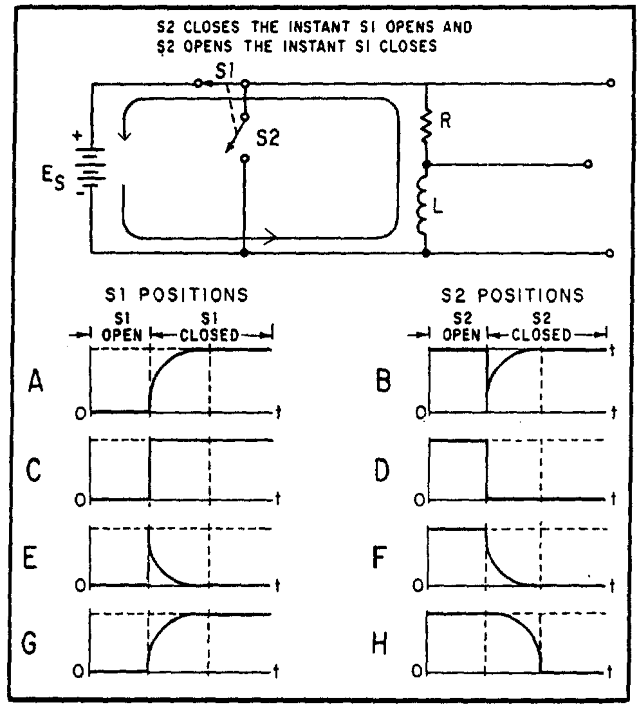


Figure 2A.—LR circuit characteristics.

IN ANSWERING QUESTIONS 2-32 THROUGH 2-40, REFER TO FIGURE 2A.

2-32. What waveform is an illustration of the voltage (E_S) present across the voltage divider when switch S1 is closed?

1. B
2. C
3. E
4. G

2-33. The voltage dropped across R when switch S1 is closed is depicted in which of the following waveforms?

1. G
2. H
3. F
4. C

2-34. The voltage developed across L when switch S1 is closed is depicted in what waveform?

1. A
2. B
3. E
4. H

2-35. Which of the following waveforms depicts growth current (I_g) through the coil (L)?

1. A
2. D
3. E
4. H

2-36. What waveform depicts the voltage developed across R when switch S2 is closed?

1. E
2. F
3. H
4. D

2-37. During the first instant when switch S1 is closed, maximum voltage is dropped across

1. the battery
2. the resistor
3. the coil
4. both the coil and resistor

2-38. During the first instant when switch S1 is closed, current is maximum in which, if any, of the following parts of the h circuit?

1. The battery
2. The coil
3. The resistor
4. None of the above

2-39. In the first instant when switch S1 is closed, the entire battery-voltage is used to overcome the

1. resistance of R
2. resistance of L
3. emf developed in R
4. emf developed in L

2-40. When switch S2 is closed, energy is supplied to the circuit by the

1. battery through S_2
2. battery through S_1
3. collapsing magnetic field of L_1
4. expanding magnetic field of L_1

2-41. One L/R time constant is equal to the time required for the current in an inductor to reach what portion of its maximum value?

1. 63.2%
2. 37.8%
3. 25.2%
4. 12.8%

2-42. Maximum current will flow in an LR circuit after a minimum of how many time constants have elapsed?

1. One
2. Five
3. Three
4. Four

2-43. The maximum current in an LR circuit is 20 amperes. What total current will be flowing in the circuit at the end of the second time constant of the charge cycle?

1. 20.0 amperes
2. 17.3 amperes
3. 12.6 amperes
4. amperes

2-44. Refer to the circuit described in question 2-43. Circuit current will increase by what amount during the second time constant?

1. 17.3 amperes
2. 12.6 amperes
3. 7.6 amperes
4. 4.7 amperes

2-45. An LR circuit has a maximum current of 30 mA. At the end of the first time constant of the discharge cycle, what total current will be flowing in the circuit?

1. 11 mA
2. 19 mA
3. 26 mA
4. 28 mA

2-46. An LR circuit contains a 150-ohm resistor and a 2-henry coil. What is the time value of one L/R time constant?

1. 7.5 seconds
2. .75 seconds
3. 1.33 seconds
4. .0133 seconds

2-47. An LR circuit has a time constant of .05 second and an inductor with a value of .60 henry. What value of resistor is required?

1. 5 ohms
2. 12 ohms
3. 24 ohms
4. 64 ohms

2-48. An LR circuit is composed of a coil of .5 henry and a 10-ohm resistor. The maximum current in the circuit is 5 amperes. After the circuit is energized, how long will it take for the current to reach maximum value?

1. 1.0 second
2. 0.05 second
3. 0.25 second
4. 5.0 seconds

2-49. Inductors experience copper loss for what reason?

1. Because of flux leakage in the copper core
2. Because the reactance of an inductor is greater than the resistance of an inductor
3. Because all inductors have resistance which dissipates power
4. Because the inertia of the magnetic field must be overcome every time the direction of current changes

2-50. Copper loss of an inductor can be calculated by the use of which of the following formulas?

1. $P = I^2R$
2. $P = I^2E$
3. $P = \frac{E}{R^2}$
4. $P = \frac{I^2}{R}$

2-51. What term applies to the power loss in an iron core inductor due to the current induced in the core?

1. Iron loss
2. Heat loss
3. Hysteresis loss
4. Eddy-current loss

2-52. Power consumed by an iron core inductor in reversing the magnetic field of the core is termed as what type of loss?

1. Iron loss
2. Heat loss
3. Hysteresis loss
4. Eddy-current loss

2-53. When does mutual inductance occur between inductors?

1. Whenever eddy-currents do not exist
2. Whenever the flux of one inductor causes an emf to be induced in another inductor
3. Whenever the effect of one inductor is aided by another inductor
4. Whenever the effect of one inductor is opposed by another inductor

2-54. Mutual inductance between two coils is affected by which of the following factors?

1. Material of the windings
2. Physical dimensions of the coils
3. Direction of the coil windings
4. Hysteresis characteristics of the coils

2-55. The coefficient of coupling between two coils is a measure of what factor?

1. The turns ratio of the coils
2. The distance between the coils
3. The relative positive of the coils
4. The magnetic flux ratio linking the coils

2-56. Two coils have a coefficient of coupling of .7 and are rated at 12 µH and 3 µH respectively. What is their total mutual inductance?

1. 4.2 µH
2. 5.2 µH
3. 7.0 µH
4. 10.5 µH

2-57. An electrical circuit contains four non-coupled inductors in a series configuration. The inductors have the following values: 2 µH, 3.5 µH, 6 µH, and 1 µH. What is the total inductance (L_T) of the circuit?

1. 45.0 µH
2. 42.5 µH
3. 12.5 µH
4. 11.5 µH

2-58. Two inductors of 3.6 µH and 7.3 µH are wired together in series and they aid each other. The mutual inductance for the circuit is 3.6µH. What is the total inductance (L_T) Of the circuit?

1. 17.5 µH
2. 18.1 µH
3. 24.8 µH
4. 34.4 µH

2-59. An electrical circuit contains three non-coupled inductors of 3.3 µH, 4.5 µH, and 2.0 µH wired in parallel. What is the total inductance of the circuit?

1. 9.8 µH
2. 3.6 µH
3. 0.98 µH
4. 0.28 µH

ASSIGNMENT 3

Textbook assignment: Chapter 3, "Capacitance," pages 3-1 through 3-41.

3-1. Capacitance and inductance in a circuit are similar in which of the following ways?

1. Both oppose current
2. Both aid voltage
3. Both cause the storage of energy
4. Both prevent the storage of energy

3-2. Capacitance is defined as the property of a circuit that

1. opposes a change in voltage
2. aids a change in voltage
3. opposes a change in current
4. aids a change in current

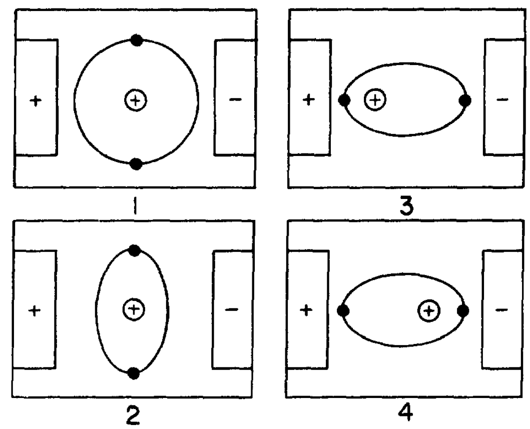
3-3. A capacitor is a device that stores energy in a/an

1. electrostatic field
2. electromagnetic field
3. induced field
4. molecular field

3-4. Electrostatic fields have what effect on (a) free electrons, and (b) bound electrons?

1. (a) Attracts them to the negative charges
   (b) Frees them from their orbits
2. (a) Attracts them to the positive charges
   (b) Frees them from their orbits
3. (a) Attracts them to the negative charges
   (b) Distorts their orbits
4. (a) Attracts them to the positive charges
   (b) Distorts their orbits

3-5. The influence of a charge on an electron orbit is correctly depicted by which of the following illustrations?



3-6. Electrostatic lines of force radiate from a charged particle along what type of lines?

1. Straight lines
2. Curved lines
3. Elliptical lines
4. Orbital lines

THIS SPACE LEFT BLANK INTENTIONALLY.

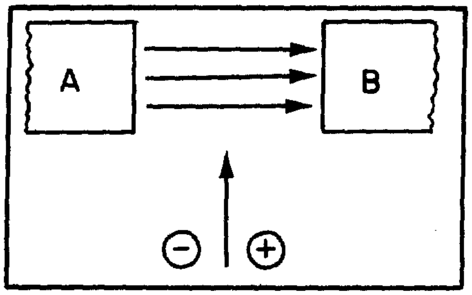


Figure 3A.—Electron and proton entering an electrostatic field.

IN ANSWERING QUESTION 3-7, REFER TO FIGURE 3A.

3-7. When the illustrated electron and proton enter the electrostatic field, toward what plate(s), will the (a) electron and, (b) proton be deflected?

1. (a) A (b) B
2. (a) B (b) A
3. (a) A (b) A
4. (a) B (b) B

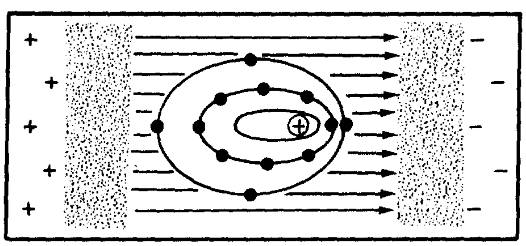


Figure 3B.—Effect of electrostatic lines of force.

IN ANSWERING QUESTION 3-8, REFER TO FIGURE 3B.

3-8. If the charges on the two plates are reversed, what will happen to the electrons?

1. They will dislodge from the atom
2. They will stay where they are
3. They will go back to circular orbits
4. They will distort in the opposite direction

3-9. Which of the following combinations describe(s) a simple capacitor?

1. Two copper plates separated by an iron plate
2. Two copper plates separated by a sheet of mica
3. Two iron plates separated by an air gap
4. Both 2 and 3 above

3-10. A capacitor that stores 6 coulombs of electrons when a potential of 2 volts is applied across its terminals has what total value of capacitance?

1. 12 farads
2. 8 farads
3. 3 farads
4. 6 farads

IN ANSWERING QUESTIONS 3-11 THROUGH 3-14, MATCH THE TERMS IN COLUMN B WITH THEIR MATHEMATICAL VALUES IN COLUMN A.

	A. VALUES		B. TERMS
3-11.	.000001 F	1.	Farad
3-12.	1×10^{-12} F	2.	Microfarad
3-13.	1×10^{-6} F	3.	Picofarad
3-14.	1×10^{0} F		

3-15. A capacitor of .0069 microfarad has which of the following capacitance values when measured in picofarads?

1. .000069 pF
2. 6900 pF
3. 6.9×10^{-9} pF
4. Both 2 and 3 above, individually, are correct

3-16. Which of the following characteristics of a capacitor can be varied WITHOUT altering its capacitance?

1. Area of the plates
2. Thickness of the dielectric
3. Material of the dielectric
4. Thickness of the plates

3-17. Which of the following actions will increase the capacitance of a capacitor?

1. The plates are moved closer together
2. The plates are moved farther apart
3. The dielectric constant is decreased
4. Both 2 and 3 above

3-18. Two capacitors are identical with the exception of the material used for the dielectric. Which of the following combinations of dielectric material will cause capacitor (b) to have a larger capacitance than capacitor (a)?

1.	(a) Glass	(b) Paraffin paper
2.	(a) Glycerine	(b) Pure water
3.	(a) Petroleum	(b) Air
4.	(a) Paraffin paper	(b) Petroleum

3-19. Two capacitors are identical with the exception of the material used for the dielectric. Which of the following combinations of dielectric materials will cause the capacitors to have almost the same capacitance?

1. Glass, paraffin paper
2. Mica, petroleum
3. Vacuum, air
4. Petroleum, rubber

3-20. A capacitor is composed of two plates. Each plate has an area of 7 square inches. The plates are separated by a 2-inch thick paraffin paper dielectric. What is its capacitance?

1. 2.76 μF
2. 2.76 pF
3. 5.51 μF
4. 5.51 pF

3-21. The maximum voltage that can be applied to a capacitor without causing current flow through the dielectric is called

1. breaking voltage
2. limiting voltage
3. conduction voltage
4. working voltage

3-22. A capacitor with a working voltage of 300 volts would normally have what maximum effective voltage applied to it?

1. 200 volts
2. 250 volts
3. 300 volts
4. 350 volts

3-23. An ac voltage of 350 volts effective can be safely applied to a capacitor with which of the following working voltages?

1. 550 volts
2. 400 volts
3. 350 volts
4. 250 volts

3-24. Which, if any, of the following conditions may cause a capacitor to suffer power losses?

1. Dielectric hysteresis
2. Plate loading
3. Plate heating
4. None of the above

3-25. Rapid reversals in the polarity of the line voltage applied to a capacitor will cause what type of capacitor power loss?

1. Dielectric-leakage
2. Dielectric-hysteresis
3. Plate-loading
4. Plate-leakage

3-26. What type of dielectric is LEAST sensitive to power dielectric-hysteresis losses?

1. Pure water
2. Air
3. Vacuum
4. Mica

3-27. As the current through a capacitor increases, which of the following types of capacitor losses will increase?

1. Dielectric-hysteresis
2. Dielectric-leakage
3. Plate-leakage
4. Plate-breakdown

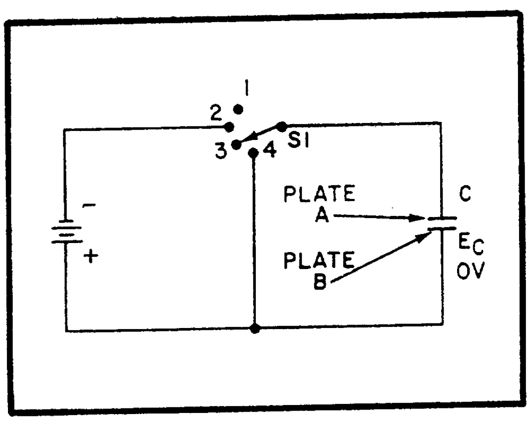


Figure 3C.—Simple capacitor circuit.

IN ANSWERING QUESTIONS 3-28 THROUGH 3-33, REFER TO FIGURE 3C.

3-28. Assume that the switch has been moved from position 4 to the position shown. Which of the following conditions now exists?

1. An electrostatic field exists between the capacitor's plates
2. No potential difference exists across the capacitor
3. Current flow is at its maximum
4. Energy is being stored in the capacitor's electrostatic field

3-29. To charge the capacitor, the switch must be in what position?

1. 1
2. 2
3. 3
4. 4

3-30. Which of the following are paths for current flow when the capacitor is charging?

1. Plate A, Plate B, Batt (+) and -Batt (–), Plate A
2. Batt (+) and -Batt (–), Plate A
Batt (–) and Batt (+), Plate B
3. Batt (–), Plate A and Plate B,
Batt (+), Batt (–)
4. Batt (+), Plate B and Plate A,
Batt (–), Batt (+)

3-31. When the switch is placed in position 4, after being in position 2, which of the following conditions exits within the circuit?

1. E_C is increasing
2. L_C is increasing
3. Electrical energy is stored in the capacitor
4. Stored energy is returned to the circuit

3-32. With S1 in position 4, which of the following is the path for current flow?

1. Plate B, Plate A, S1, Plate B
2. Plate A, Plate B, S1, Plate A
3. Plate A, S1, Plate B
4. Plate B, S1, Plate A

3-33. If the illustrated capacitor has a value of 50 μf and a potential difference of 300 volts exists across its plates, what is the total number of coulombs it contains?

1. 0.015
2. 0.15
3. 1.50
4. 15.0

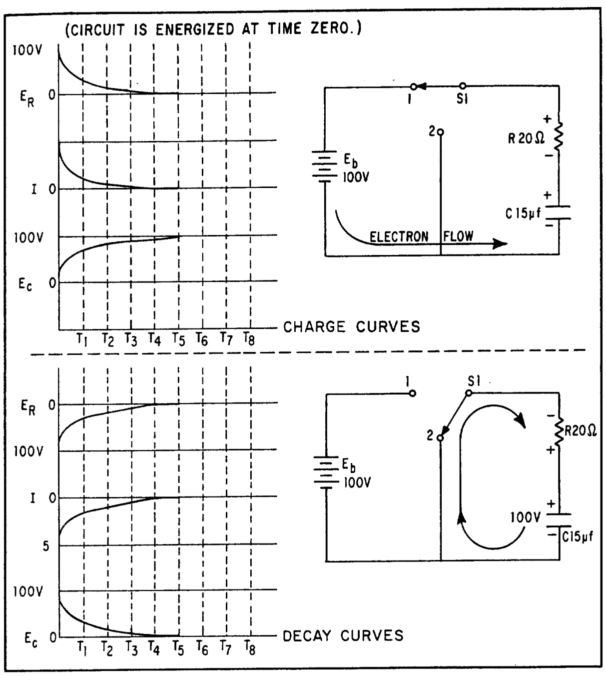


Figure 3D.—RC charge and discharge circuits.

IN ANSWERING QUESTIONS 3-34 THROUGH 3-41, REFER TO FIGURE 3D.

3-34. The greatest rate of change in current occurs between what two times?

1. $T_1 - T_2$
2. $T_2 - T_3$
3. $T_4 - T_5$
4. $T_0 - T_1$

3-35. At what instant does the maximum voltage appear across the resistor?

1. T_1
2. T_2
3. T_5
4. T_0

3-36. When the charge on the capacitor is equal to 100 volts, what is the voltage drop across the resistor?

1. 100 volts
2. 63 volts
3. 27 volts
4. 0 volts

3-37. After the capacitor has reached full charge, S1 is placed in position 2. The greatest rate of change in current is between what two times?

1. $T_1 - T_2$
2. $T_2 - T_3$
3. $T_4 - T_5$
4. $T_0 - T_1$

3-38. The capacitor will be completely discharged at what minimum time interval?

1. T_1
2. T_5
3. T_3
4. T_4

3-39. What is the RC time constant for the circuit?

1. 300 sec
2. 35 sec
3. 300 µsec
4. 35 µsec

3-40. What total time will it take the capacitor to charge to 98 volts? (You may use figure 3-11 of your text, or figure 3 located on this page.)

1. 140 µsec
2. 1200 µsec
3. 140 sec
4. 1200 sec

3-41. After the capacitor has reached full charge, S1 is moved to position 2. What total number of RC time constants will it take for the capacitor to discharge to 5 volts?

1. One
2. Two
3. Three
4. Four

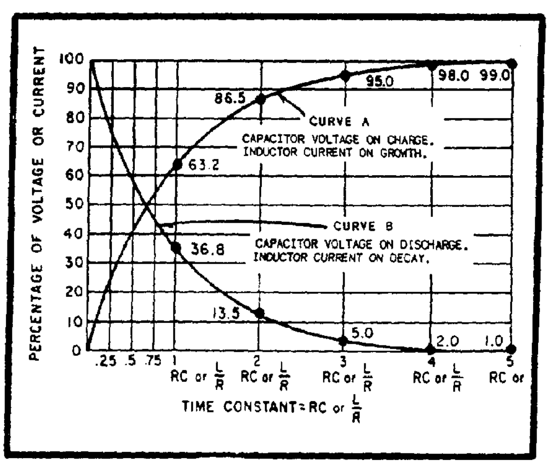


Figure 3E.—Universal time constant chart for RC and RL circuit.

IN ANSWERING QUESTION 3-42, REFER TO FIGURE 3E ABOVE.

3-42. An RC circuit is designed in which a capacitor must charge to 55 percent (.55) of the maximum charging voltage in 200 microseconds. The resistor has a value of 30,000 ohms. What value of capacitance is needed?

1. 0.0089 pF
2. 89.0 pF
3. 0.0089 µF
4. 89.0 µF

MATCH THE CAPACITOR CONFIGURATION IN COLUMN B WITH THE CHARACTERISTICS IN COLUMN A.

	A. CHARACTERISTICS	B. CONFIGURATION
3-43.	Increases total capacitance	1. Capacitors in parallel
3-44.	Effectively moves plates further apart	2. Capacitors in series
3-45.	Increases plate area	
3-46.	Total capacitance is found by adding all capacitances	
3-47.	Decreases total capacitance	
3-48.	Similar to resistors in parallel	

3-49. A circuit contains four parallel-connected capacitors of 33 µF each. What is the total capacitance of the circuit?

1. 8.3 µF
2. 33.0 µF
3. 183.0 µF
4. 132.0 µF

3-50. A circuit contains two series-connected capacitors of 15 µF, and 1500 pF. What is the total capacitance of the circuit?

1. 0.0015 pF
2. 150.0 pF
3. 0.0015 µF
4. 0.1500 µF

3-51. A circuit contains two 10 μF capacitors wired together in a parallel configuration. The two parallel-wired capacitors are wired in series with a 20 μF capacitor and a 20 K ohm resistor. Which of the following expresses the RC time constant for this circuit?

1. .2 sec
2. 2 sec
3. 20,000 sec
4. Both 2 and 3 above

3-52. How are fixed capacitors classified?

1. By their plate size
2. By their dielectric materials
3. By the thickness of their dielectric materials
4. By the thickness of their conductors

3-53. Which of the following types of capacitors are referred to as self-healing?

1. Ceramic
2. Paper
3. Oil
4. Mica

MATCH THE CAPACITOR TYPE IN COLUMN B WITH THE CHARACTERISTIC IN COLUMN A.

	A. CHARACTERISTIC		B. TYPE
3-54.	Has an oxide film dielectric	1.	Electrolytic
3-55.	Can be adjusted by a screw setting	2.	Trimmer
3-56.	A polarized capacitor	3.	Mica
3-57.	Has a waxed paper dielectric	4.	Paper
3-58.	An adjustable capacitor with a mica dielectric		

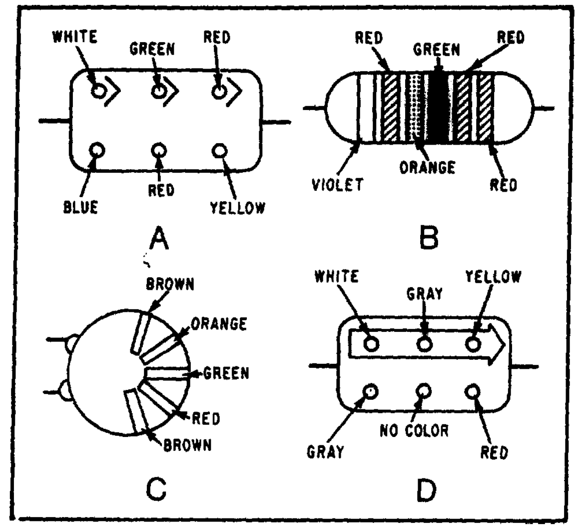


Figure 3F.—Types of capacitors.

IN ANSWERING QUESTIONS 3-59 THROUGH 3-63, REFER TO FIGURE 3F AND TO THE ASSOCIATED PAGES IN YOUR TEXTBOOK.

3-59. Capacitor A is what type of capacitor?

1. Electrolytic
2. Ceramic
3. Paper
4. Mica

3-60. Capacitor B is what type of capacitor?

1. Mica
2. Paper
3. Ceramic
4. Electrolytic

3-61. What is the capacitance of capacitor B?

1. 2,200,000 pF
2. 2,200,000 μF
3. 72,000 pF
4. 72,000 μF

3-62. What is the (a) temperature coefficient and (b) multiplier of capacitor C?

1. (a) -30 (b) 100
2. (a) -30 (b) 1000
3. (a) -330 (b) 100
4. (a) -330 (b) 1000

3-63. What is the (a) capacitance, and (b) voltage rating of capacitor D?

1. (a) 4800 μF (b) 200 volts
2. (a) 4800 pF (b) 200 volts
3. (a) 98,000 μF (b) 800 volts
4. (a) 980,000 pF (b) 800 volts

ASSIGNMENT 4

Textbook assignment: Chapter 4, "Inductive and Capacitive Reactance," pages 4-1 through 4-40. Chapter 5, "Transformers," pages 5-1 through 5-31.

4-1. Inductance has what effect, if any, on a change in (a) current, and (b) voltage?

1. (a) No effect (b) aids it
2. (a) Aids it (b) no effect
3. (a) Opposes it (b) no effect
4. (a) No effect (b) opposes it

4-2. Voltage leads current in which of the following types of circuits?

1. Resistive
2. Capacitive
3. Both 1 and 2 above
4. Inductive

4-3. Opposition to the flow of current by a coil in an ac circuit is represented by what symbol?

1. R
2. X_L
3. L
4. H

4-4. What is the opposition offered by a coil to (a) the flow of alternating current and (b) a change in current?

1. (a) Resistance (b) Inductance
2. (a) Reactance (b) Counterreactance
3. (a) Reactance (b) Inductance
4. (a) Resistance (b) Reactance

4-5. The formula $2\pi fL$ is used to determine what electrical quantity?

1. Resistance
2. Inductance
3. Counterreaction
4. Inductive reactance

4-6. An inductive circuit contains a 200-µH coil and the ac voltage applied is at a frequency of 120 Hz. What is the value of reactance for the circuit?

1. 0.15 Ω
2. 1.50 Ω
3. 7.50 Ω
4. 75.0 Ω

4-7. If the frequency applied to a circuit with a 200-µH coil is increased from 120 Hz to 50 kHz, what will be the value of reactance for the circuit?

1. 1.0.75 Ω
2. 2.7.5 Ω
3. 3.62.8 Ω
4. 628.0 Ω

4-8. A capacitor will (a) conduct what type of current, and (b) block what type of current?

1. (a) Dc (b) All ac
2. (a) All ac (b) Dc
3. (a) Dc (b) Ac above 60 Hz
4. (a) Ac above 60 Hz (b) Dc

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 4-9 THROUGH 4-13, SELECT FROM COLUMN B THE PROPERTY THAT CAUSES THE ELECTRICAL EFFECT IN COLUMN A.

A	B
4-9. Opposition to ac but not dc	1. Inductive reactance
4-10. Causes a phase shift between voltage and current	2. Capacitive reactance
4-11. Increases with an increase in frequency	3. Both 1 and 2 above
4-12. Causes current to lead voltage by 90º	4. Resistance
4-13. Decreases with an increase in frequency.	

4-14. An electrical circuit contains a 25-µF capacitor and operates from a 60-Hz ac source. What is the value of capacitive reactance of the circuit?

1. 0.00106 Ω
2. 0.0106 Ω
3. 10.6 Ω
4. 106.2 Ω

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 4-15 AND 4-16 USE THE FOLLOWING INFORMATION: A SERIES CIRCUIT HAS AN INDUCTIVE REACTANCE OF 56Ω, A CAPACITIVE REACTANCE OF 25Ω, AND OPERATES AT A FREQUENCY OF 400 HZ.

4-15. What formula should you use to determine the total reactance for the circuit?

1. $X = 2\pi fL$
2. $X = \frac{1}{2\pi fC}$
3. $X = X_L - X_C$
4. $X = X_C - X_L$

4-16. What is the total value of reactance for the circuit?

1. 31 Ω
2. 81 Ω
3. 1,400 Ω
4. 14,067 Ω

4-17. What term is used to express the total opposition to ac in an electrical circuit?

1. Reactance
2. Impedance
3. Resistance
4. Conductance

4-18. A series ac circuit has the following values: $X_L = 5\Omega$, $X_C = 6\Omega$, and $R = 7\Omega$. What is the value of Z?

1. 1.00 Ω
2. 3.03 Ω
3. 7.07 Ω
4. 14.14 Ω

4-19. A series circuit contains an inductor having 12 ohms of resistance and 30 ohms of inductive reactance in series with a capacitor having 21 ohms of capacitive reactance. The applied voltage is 100 volts. What is the value of current for the circuit?

1. 6.6 amps
2. 8.4 amps
3. 15.0 amps
4. 25.6 amps

4-20. A series circuit contains an inductor having 12 ohms of resistance and 64 ohms of inductive reactance in series with a capacitor having 69 ohms of capacitive reactance. If the current through the circuit is 6.5 amperes, what is the value of the voltage applied to the circuit?

1. 26.5 volts
2. 55.5 volts
3. 75.5 volts
4. 84.5 volts

4-21. True power in a circuit is dissipated in what circuit element?

1. Resistance
2. Reactance
3. Capacitance
4. Inductance

4-22. In a purely reactive circuit, what happens to power?

1. It is dissipated across the reactive loads
2. It is cancelled by the reactive elements
3. It is stored in the reactive elements
4. It is returned to the source

4-23. True power is measured in what unit?

1. Watt
2. Volt-ampere
3. Var
4. P_t-watt

4-24. An ac series circuit has the following characteristics: R = 8 ohms, X_C = 100 ohms, X_L = 70 ohms, and E = 220 V. What is the value of true power for the circuit?

1. 46 W
2. 57 W
3. 268 W
4. 402 W

4-25. What is the unit of measurement for reactive power?

1. Watt
2. Var
3. Volt-ampere
4. Volt-ohm

4-26. An ac series circuit has the following values: I = 7.5 amps, $X_L = 80\Omega$, and $X_C = 35\Omega$. What is the value of reactive power for the circuit?

1. 2531 var
2. 1567 var
3. 1283 var
4. 861 var

4-27. Apparent power in an ac circuit is a combination of which of the following factors?

1. Applied power and true power
2. Reactive power and true power
3. Applied power and the power returned to the source
4. Reactive power and the power returned to the source

4-28. What is the unit of measurement for apparent power?

1. Watt
2. Var
3. Volt-ampere
4. Volt-ohm

4-29. An ac circuit dissipates 800 watts across its resistance and returns 600 var to the source. What is the value of the apparent power of the circuit?

1. 200 VA
2. 500 VA
3. 1000 VA
4. 1400 VA

4-30. The portion of apparent power dissipated in an ac circuit can be calculated by which of the following formulas?

1. $PF = (I_R)^2 R$
2. $PF = (I_Z)^2$
3. $PF = \frac{(I_Z)^2 Z}{(I_Z)^2 R}$
4. $PF = \frac{(I_R)^2 R}{(I_Z)^2 Z}$

4-31. A series ac circuit has a X_C of 110 ohms, an X_L of 30 ohms, and a circuit resistance of 22 ohms. What is the power factor of this circuit?

1. .91
2. .27
3. .20
4. .13

An RLC series a.c. circuit has the following values:

$E = 65$ volts

$f = 120$ Hz

$R = 12$ ohms

$L = 30$ mH

$C = 450\ \mu F$

Figure 4A.—Circuit characteristics.

IN ANSWERING QUESTIONS 4-32 THROUGH 4-36, REFER TO FIGURE 4A.

4-32. What is the value of X?

1. 19.7 Ω
2. 27.8 Ω
3. 31.6 Ω
4. 42.3 Ω

4-33. What is the value of Z?

1. 23 Ω
2. 28 Ω
3. 33 Ω
4. 38 Ω

4-34. What is the value of I_T for the circuit?

1. 1.8 A
2. 2.8 A
3. 3.4 A
4. 4.4 A

4-35. What is the value of true power?

1. 67 W
2. 83 W
3. 94 W
4. 125 W

4-36. What is the power factor?

1. .46
2. .52
3. .73
4. .88

4-37. When impedance is calculated for a parallel ac circuit, an intermediate value must first be calculated. The intermediate value must then be divided into the source voltage to derive impedance. What is this intermediate value?

1. Reactance
2. Resistance
3. Power factor
4. Total current

4-38. Which of the following defines transformer action?

1. The transfer of energy from one circuit to another through electromagnetic action
2. The transfer of energy from one circuit to another through electrostatic action
3. The development of counter electromotive force where a magnetic field cuts a coil
4. The development of a voltage across a coil as it cuts through a magnetic field

4-39. Which of the following is NOT a necessary element in a basic transformer?

1. A core
2. A primary winding
3. A secondary winding
4. A magnetic shield

4-40. What three materials are most commonly used for transformer cores?

1. Copper, soft iron, and air
2. Copper, soft iron, and steel
3. Air, copper, and steel
4. Air, soft iron, and steel

4-41. The two types of transformer cores most commonly used are the shell-core and the

1. I-core
2. E-core
3. hollow-core
4. laminated-core

4-42. What is the major difference between the primary and secondary windings of a transformer?

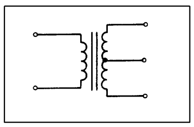
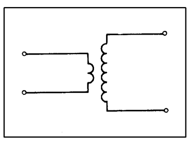
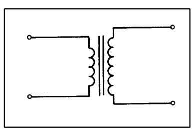
1. The primary has more turns than the secondary
2. The secondary has more insulation than the primary
3. The primary is connected to the source; the secondary is connected to the load
4. The primary is connected to the load; the secondary is connected to the source

4-43. What is the principal difference between a high-voltage transformer and a low-voltage transformer?

1. A high-voltage transformer has more turns of wire than a low-voltage transformer
2. A high-voltage transformer uses a hollow-core, while a low-voltage transformer uses a shell-type core
3. A high-voltage transformer uses a shell-type core, while a low-voltage transformer uses a hollow-core
4. A high-voltage transformer has more insulation between the layers of windings than does a low-voltage transformer

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 4-44 THROUGH 4-46, SELECT FROM COLUMN B THE TRANSFORMER CHARACTERISTICS THAT ARE IDENTIFIED IN THE TRANSFORMER SCHEMATICS IN COLUMN A.

	A. SCHEMATIC	B. TRANSFORMER TYPE
4-44.		1. Air-core 2. Iron-core 3. Center-tapped 4. Iron-core with center tap
4-45.		
4-46.		

4-47. When the secondary of a transformer is NOT connected to a circuit, the transformer is said to be operating under which of the following conditions?

1. Uncoupled
2. No-load
3. Loaded
4. Open

4-48. What term applies to the current in the primary of a transformer that creates the magnetic field?

1. Exciting current
2. Primary current
3. Magnetizing current
4. Counter current

4-49. In the primary of a transformer, what opposes the current from the source?

1. The impedance
2. The forward emf
3. The self-induced emf
4. The exciting current

4-50. What is the source of the magnetic flux that develops secondary voltage in a transformer?

1. Primary emf
2. Secondary counter emf
3. Primary exciting current
4. Secondary exciting current

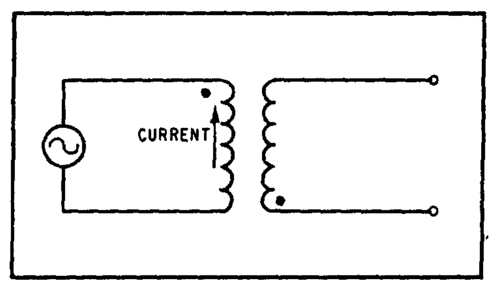


Figure 4B.—Transformer polarity.

IN ANSWERING QUESTION 4-51, REFER TO FIGURE 4B.

4-51. The illustrated transformer is (a) what type, and (b) in what direction is the current flowing in the secondary?

1. (a) Like-wound (b) ↑
2. (a) Unlike-wound (b) ↑
3. (a) Like-wound (b) ↓
4. (a) Unlike-wound (b) ↓

4-52. Which of the following terms applies to the flux from the primary that does NOT cut the secondary

1. Lost flux
2. Leakage flux
3. Uncoupled flux
4. Coefficient flux

4-53. What is the main cause for the coefficient of coupling of a transformer being less than unity?

1. Counter emf
2. Induced emf
3. Uncoupled flux
4. Leakage flux

4-54. A transformer has a source voltage of 50 volts ac, with a turns ratio of 1:6. The coefficient of coupling is 1.0. What is the voltage of the secondary winding?

1. 150
2. 300
3. 500
4. 600

4-55. A transformer has a unity coefficient of coupling with a 5:1 turns ratio; 20 volts are induced in the secondary. What is the primary voltage?

1. 100 volts
2. 50 volts
3. 10 volts
4. 4 volts

4-56. A transformer has a unity coefficient of coupling. Thirty-five volts applied to its primary induces 105 volts in its secondary. The secondary is composed of 99 turns. What is the number of turns in the primary?

1. 11 turns
2. 22 turns
3. 33 turns
4. 44 turns

4-57. A transformer secondary has 20 amperes of current flowing at 60 volts potential. The applied voltage is 10 volts. What is (a) the turns ratio of the transformer and (b) what total current is flowing in the primary?

1. (a) 6:1, (b) 3.3 amperes
2. (a) 1:6, (b) 120 amperes
3. (a) 1:2, (b) 10 amperes
4. (a) 2:1, (b) 120 amperes

4-58. A 2:1 transformer delivers 30 watts to the load and 3 watts of power are lost to internal losses. What total power is drawn from the source?

1. 63 watts
2. 57 watts
3. 33 watts
4. 27 watts

4-59. What is the efficiency of the transformer described in question 4-58?

1. 33 %
2. 46 %
3. 53 %
4. 91 %

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 4-60 THROUGH 4-62, SELECT FROM COLUMN B THE TERM THAT DESCRIBES THE TYPE OF POWER LOSS IN COLUMN A.

	A. LOSS TYPE		B. TERMS
4-60.	Power lost in realigning domains	1.	Copper loss
4-61.	Power dissipated by the resistance of the windings	2.	Eddy-current loss
4-62.	Power loss caused by random core currents	3.	Hysteresis loss
		4.	Leakage Loss

4-63. A transformer designed for a low frequency will NOT be damaged when used at higher frequencies. What change within the transformer, limits transformer current to a safe value at higher frequencies?

1. Increased hysteresis loss
2. Increased inductive reactance
3. Increased leakage flux
4. Increased eddy-current loss

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 4-64 THROUGH 4-67, SELECT THE TRANSFORMER TYPE FROM COLUMN B THAT PERFORMS THE TASK OR HAS THE CHARACTERISTICS DESCRIBED IN COLUMN A.

	A. TASK		B. TRANSFORMER TYPE
4-64.	Used above 20 kHz	1.	Power
4-65.	The secondary is a tapped primary	2.	Autotransformer
4-66.	Used to deliver voltage from a source to a load	3.	Audio-Frequency
4-67.	Can be used to match impedance in a sound system	4.	Radio-Frequency

4-68. What wire colors conventionally identify the secondary center tap of a power transformer?

1. Black and yellow
2. Red and white
3. Black and red
4. Red and yellow

4-69. Before starting to work on any electrical equipment, you should first determine that the equipment is in which of the following conditions?

1. Connected
2. Deenergized
3. Energized
4. Operational

4-70. A person is working on electrical equipment. The power is secured and tagged. The technician receives a shock on the hand. What safety precaution was overlooked?

1. The technician was not standing on approved rubber matting
2. The technician had not discharged the equipment's capacitors
3. The technician was working on energized equipment
4. The technician had two hands in the equipment

4-71. When working on electrical equipment, why should you use only one hand?

1. The free hand can be used to turn off the power in case of shock
2. The free hand can be used to pull the other hand free in case of muscle contraction from shock
3. The free hand will ensure that you are properly grounded
4. The free hand will minimize the possibility of creating a low resistance path to ground through your body

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 3—Introduction to Circuit Protection, Control, and Measurement

NAVEDTRA 14175

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Sailor’s Creed

“I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country’s Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all.”

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: Subjects of circuit measurement, circuit protection devices, and circuit control devices are presented to give the student background information on topics that may be encountered in daily work.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and the occupational standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards*, NAVPERS 18068, found on line at https://buperscd.technology.navy.mil /bup_updt/upd_CD/BUPERS/enlistedManOpen.htm.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
FCC(SW) James L. Hicks

Corrections and minor modifications prepared in
January 2003 by
ETC Scott Collie

NAVSUP Logistics Tracking Number
0504-LP-026-8280

TABLE OF CONTENTS

Assignments follow Appendix IV.

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 5 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

(THIS PAGE IS INTENTIONALLY LEFT BLANK.)

CHAPTER 1

CIRCUIT MEASUREMENT

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the NRTC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completion of this chapter you will be able to:

1. State two ways circuit measurement is used, why in-circuit meters are used, and one advantage of out-of-circuit meters.
2. State the way in which a compass reacts to a conducting wire including the compass reaction to increasing and decreasing dc and ac high and low frequencies.
3. State how a d'Arsonval meter movement reacts to dc.
4. State the purpose of a rectifier as used in ac meters.
5. State the meaning of the term "damping" as it applies to meter movements and describe two methods by which damping is accomplished.
6. Identify average value as the value of ac measured and effective value (rms) as the ac value indicated on ac meter scales.
7. Identify three meter movements that measure dc or ac without the use of a rectifier.
8. State the electrical quantity measured by an ammeter, the way in which an ammeter is connected in a circuit, and the effect of an ammeter upon a circuit.
9. Define ammeter sensitivity.
10. State the method used to allow an ammeter to measure different ranges and the reason for using the highest range when connecting an ammeter to a circuit.
11. List the safety precautions for ammeter use.
12. State the electrical quantity measured by a voltmeter, the way in which a voltmeter is connected in a circuit, the way in which a voltmeter affects the circuit being measured, and the way in which a voltmeter is made from a current reacting meter movement.
13. Define voltmeter sensitivity.
14. State the method used to allow a voltmeter to measure different ranges and the reason for using the highest range when connecting a voltmeter to a circuit.

15. Identify the type of meter movement that reacts to voltage and the most common use of this movement.

16. List the safety precautions for voltmeter use.

17. State the electrical quantity measured by an ohmmeter, the second use of an ohmmeter, and the way in which an ohmmeter is connected to a resistance being measured.

18. State the method used to allow an ohmmeter to measure different ranges and the area of an ohmmeter scale that should be used when measuring resistance.

19. State the two types of ohmmeters and the way in which each can be identified.

20. List the safety precautions for ohmmeter use.

21. State the primary reason for using a megger and the method of using it.

22. Identify normal and abnormal indications on a megger.

23. List the safety precautions for megger use.

24. State how a multimeter differs from other meters, the reason a multimeter is preferred over separate meters, and the way in which a multimeter is changed from a voltage measuring device to a current measuring device.

25. State the reason the ac and dc scales of a multimeter differ, the reason for having a mirror on the scale of a multimeter, and the proper way of reading a multimeter using the mirror.

26. List the safety precautions for multimeter use.

27. State the purpose of a hook-on type voltameter.

28. State the electrical quantity measured by a wattmeter and a watt-hour meter.

29. Identify the two types of frequency meters.

30. Identify the type of meter and interpret the meter reading from scale presentations of an ammeter; a voltmeter; an ohmmeter; a megger; a multimeter (current, voltage, and resistance examples); a wattmeter; a watt-hour meter; and a frequency meter (vibrating reed and moving-disk types).

CIRCUIT MEASUREMENT

This chapter will acquaint you with the basics of circuit measurement and some of the devices used to measure voltage, current, resistance, power, and frequency. There are other quantities involved in electrical circuits, such as capacitance, inductance, impedance, true power, and effective power. It is possible to measure any circuit quantity once you are able to select and use the proper circuit measuring device. You will NOT know all there is to know about circuit measuring devices (test equipment) when you finish this chapter. That is beyond the scope of this chapter and even beyond the scope of this training series. However, more information on test equipment is provided in another portion of this training series.

A question which you might ask before starting this chapter is "Why do I need to know about circuit measurement?"

If you intend to accomplish anything in the field of electricity and electronics, you must be aware of the forces acting inside the circuits with which you work. Modules 1 and 2 of this training series introduced you to the physics involved in the study of electricity and to the fundamental concepts of direct and alternating current. The terms voltage (volts), current (amperes), and resistance (ohms) were explained, as well as the various circuit elements; e.g., resistors, capacitors, inductors, transformers, and batteries.

In explaining these terms and elements to you, schematic symbols and schematic diagrams were used. In many of these schematic diagrams, a meter was represented in the circuit, as shown in figure 1-1.

As you recall, the current in a dc circuit with 6 volts across a 6-ohm resistor is 1 ampere. The @(UPPERCASE A) in figure 1-1 is the symbol for an ammeter. An ammeter is a device that measures current. The name "ammeter" comes from the fact that it is a meter used to measure current (in amperes), and thus is called an AMpere METER, or AMMETER. The ammeter in figure 1-1 is measuring a current of 1 ampere with the voltage and resistance values given.

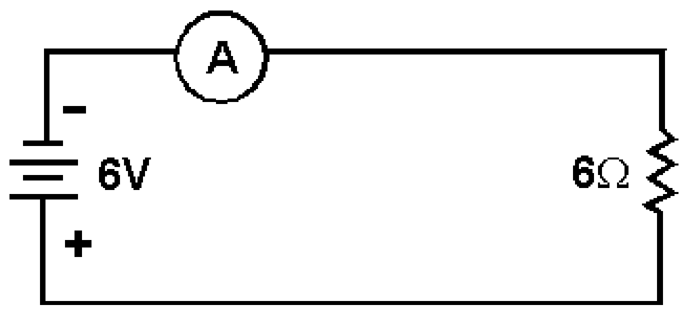


Figure 1-1.—A simple representative circuit.

In the discussion and explanation of electrical and electronic circuits, the quantities in the circuit (voltage, current, and resistance) are important. If you can measure the electrical quantities in a circuit, it is easier to understand what is happening in that circuit. This is especially true when you are troubleshooting defective circuits. By measuring the voltage, current, capacitance, inductance, impedance, and resistance in a circuit, you can determine why the circuit is not doing what it is supposed to do. For instance, you can determine why a radio is not receiving or transmitting, why your automobile will not start, or why an electric oven is not working. Measurement will also assist you in determining why an electrical component (resistor, capacitor, inductor) is not doing its job.

The measurement of the electrical parameters quantities in a circuit is an essential part of working on electrical and electronic equipment.

INTRODUCTION TO CIRCUIT MEASUREMENT

Circuit measurement is used to monitor the operation of an electrical or electronic device, or to determine the reason a device is not operating properly. Since electricity is invisible, you must use some sort of device to determine what is happening in an electrical circuit. Various devices called test equipment are used to measure electrical quantities. The most common types of test equipment use some kind of metering device.

IN-CIRCUIT METERS

Some electrical and electronic devices have meters built into them. These meters are known as in-circuit meters. An in-circuit meter is used to monitor the operation of the device in which it is installed. Some examples of in-circuit meters are the generator or alternator meter on some automobiles; the voltage, current, and frequency meters on control panels at electrical power plants; and the electrical power meter that records the amount of electricity used in a building.

It is not practical to install an in-circuit meter in every circuit. However, it is possible to install an in-circuit meter in each critical or representative circuit to monitor the operation of a piece of electrical equipment. A mere glance at or scan of the in-circuit meters on a control board is often sufficient to tell if the equipment is working properly.

While an in-circuit meter will indicate that an electrical device is not functioning properly, the cause of the malfunction is determined by troubleshooting. Troubleshooting is the process of locating and repairing faults in equipment after they have occurred. Since troubleshooting is covered elsewhere in this training series, it will be mentioned here only as it applies to circuit measurement.

OUT-OF-CIRCUIT METERS

In troubleshooting, it is usually necessary to use a meter that can be connected to the electrical or electronic equipment at various testing points and may be moved from one piece of equipment to another. These meters are generally portable and self-contained, and are known as out-of-circuit meters.

Out-of-circuit meters are more versatile than in-circuit meters in that the out-of-circuit meter can be used wherever you wish to connect it. Therefore, the out-of-circuit meter is more valuable in locating the cause of a malfunction in a device.

Q1. What are two ways that circuit measurement is used?

Q2. Why are in-circuit meters used?

Q3. What is one advantage of an out-of-circuit meter when it is compared with an in-circuit meter?

BASIC METER MOVEMENTS

The meter movement is, as the name implies, the part of a meter that moves. A meter movement converts electrical energy into mechanical energy. There are many different types of meter movements. The first one you will learn about is based upon a principle with which you are already familiar. That principle is the interaction of magnetic fields.

COMPASS AND CONDUCTING WIRE

You know that an electrical conductor in which current flows has a magnetic field generated around it. If a compass is placed close to the conductor, the compass will react to that magnetic field (fig. 1-2).

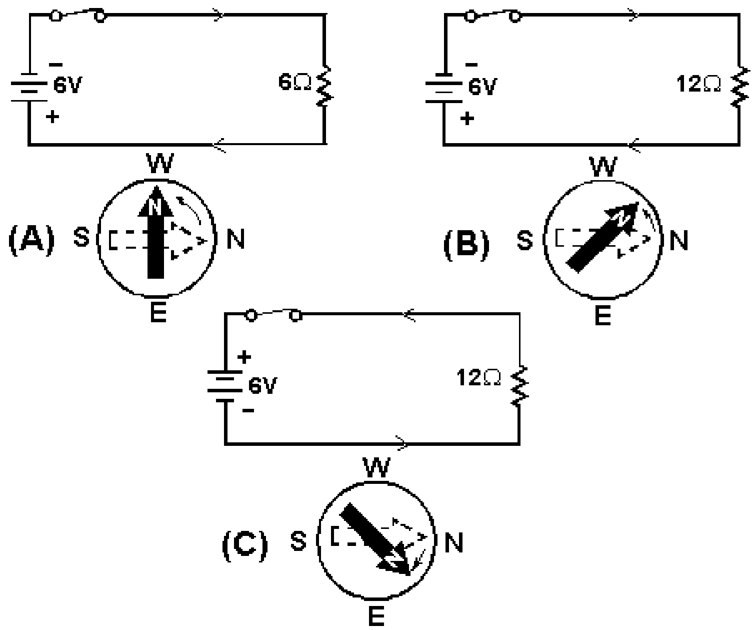


Figure 1-2.—Compass and conductor with direct current.

If the battery is disconnected, the north end of the compass needle will point to magnetic north, as illustrated in figure 1-2(A) by the broken-line compass needle pointing to the right. When the battery is connected, current flows through the circuit and the compass needle aligns itself with the magnetic field of the conductor, as indicated by the solid compass needle. The strength of the magnetic field created around the conductor is dependent upon the amount of current.

In figure 1-2(A), the resistance in the circuit is 6 ohms. With the 6-volt battery shown, current in the circuit is 1 ampere. In figure 1-2(B), the resistance has been changed to 12 ohms. With the 6-volt battery shown, current in the circuit is 1/2 or .5 ampere. The magnetic field around the conductor in figure 1-2(B) is weaker than the magnetic field around the conductor in figure 1-2(A). The compass needle in figure 1-2(B) does not move as far from magnetic north.

If the direction of the current is reversed, the compass needle will move in the opposite direction because the polarity of the magnetic field has reversed.

In figure 1-2(C), the battery connections are reversed and the compass needle now moves in the opposite direction.

You can construct a crude meter to measure current by using a compass and a piece of paper. By using resistors of known values, and marking the paper to indicate a numerical value, as in figure 1-3, you have a device that measures current.

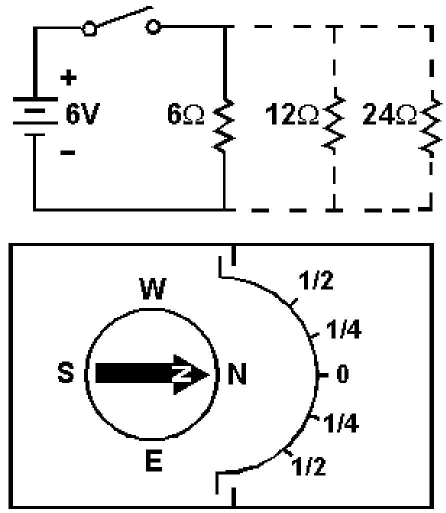


Figure 1-3.—A simple meter from a compass.

This is, in fact, the way the first GALVANOMETERS were developed. A galvanometer is an instrument that measures small amounts of current and is based on the electromagnetic principle. A galvanometer can also use the principles of electrodynamics, which will be covered later in this topic.

The meter in figure 1-3 is not very practical for electrical measurement. The amount the compass needle swings depends upon the closeness of the compass to the conductor carrying the current, the direction of the conductor in relation to magnetic north, and the influence of other magnetic fields. In addition, very small amounts of current will not overcome the magnetic field of the Earth and the needle will not move.

Q4. How does a compass react when placed close to a current carrying conductor?

Q5. If the amount of current in the conductor changes, what happens to the magnetic field around the conductor?

Q6. How does the compass needle react to a decreased magnetic field?

PERMANENT-MAGNET MOVING-COIL MOVEMENT

The compass and conducting wire meter can be considered a fixed-conductor moving-magnet device since the compass is, in reality, a magnet that is allowed to move. The basic principle of this device is the interaction of magnetic fields-the field of the compass (a permanent magnet) and the field around the conductor (a simple electromagnet).

A permanent-magnet moving-coil movement is based upon a fixed permanent magnet and a coil of wire which is able to move, as in figure 1-4. When the switch is closed, causing current through the coil, the coil will have a magnetic field which will react to the magnetic field of the permanent magnet. The bottom portion of the coil in figure 1-4 will be the north pole of this electromagnet. Since opposite poles attract, the coil will move to the position shown in figure 1-5.

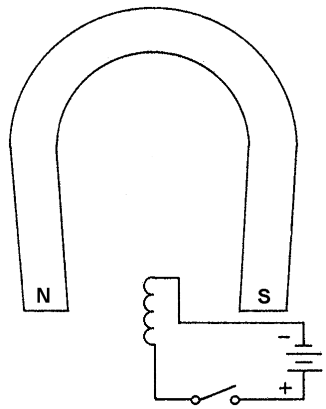


Figure 1-4.—A movable coil in a magnetic field (no current).

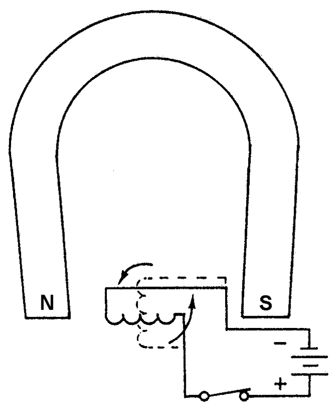


Figure 1-5.—A movable coil in a magnetic field (current).

The coil of wire is wound on an aluminum frame, or bobbin, and the bobbin is supported by jeweled bearings which allow it to move freely. This is shown in figure 1-6.

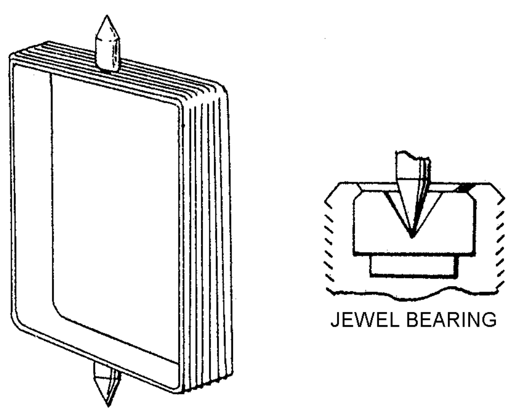


Figure 1-6.—A basic coil arrangement.

To use this permanent-magnet moving-coil device as a meter, two problems must be solved. First, a way must be found to return the coil to its original position when there is no current through the coil. Second, a method is needed to indicate the amount of coil movement.

The first problem is solved by the use of hairsprings attached to each end of the coil as shown in figure 1-7. These hairsprings can also be used to make the electrical connections to the coil. With the use of hairsprings, the coil will return to its initial position when there is no current. The springs will also tend to resist the movement of the coil when there is current through the coil. When the attraction between the magnetic fields (from the permanent magnet and the coil) is exactly equal to the force of the hairsprings, the coil will stop moving toward the magnet.

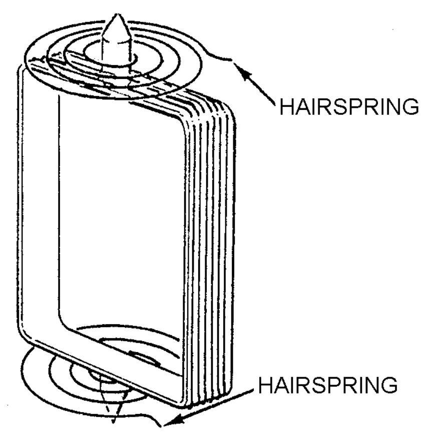


Figure 1-7.—Coil and hairsprings.

As the current through the coil increases, the magnetic field generated around the coil increases. The stronger the magnetic field around the coil, the farther the coil will move. This is a good basis for a meter.

But, how will you know how far the coil moves? If a pointer is attached to the coil and extended out to a scale, the pointer will move as the coil moves, and the scale can be marked to indicate the amount of current through the coil. This is shown in figure 1-8.

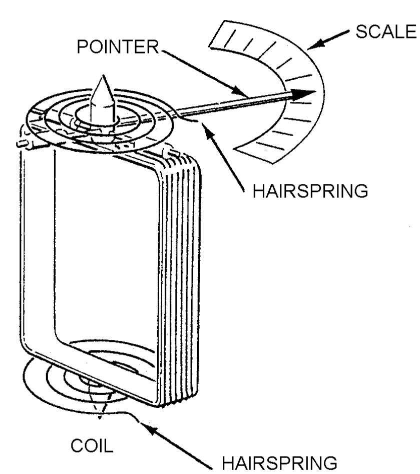


Figure 1-8.—A complete coil.

Two other features are used to increase the accuracy and efficiency of this meter movement. First, an <u>iron core</u> is placed inside the coil to <u>concentrate the magnetic fields</u>. Second, curved <u>pole pieces</u> are attached to the magnet to ensure that the turning force on the coil increases steadily as the current increases.

The meter movement as it appears when fully assembled is shown in figure 1-9.

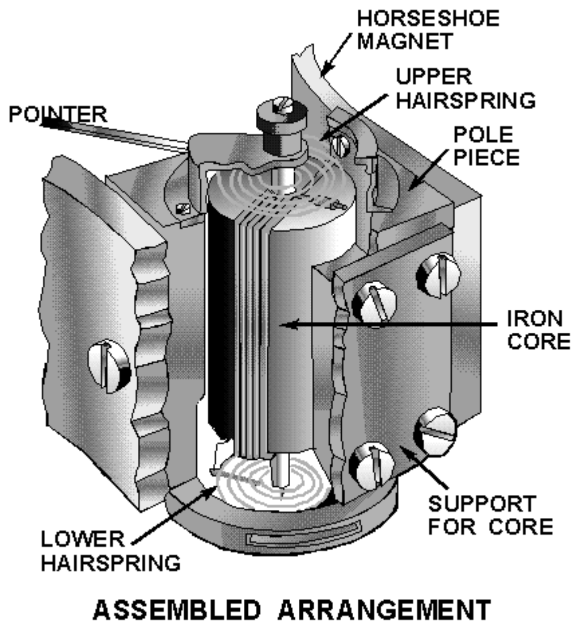


Figure 1-9.—Assembled meter movement.

This permanent-magnet moving-coil meter movement is the basic movement in most measuring instruments. It is commonly called the d'Arsonval movement because it was first employed by the Frenchman d'Arsonval in making electrical measurements. Figure 1-10 is a view of the d'Arsonval meter movement used in a meter.

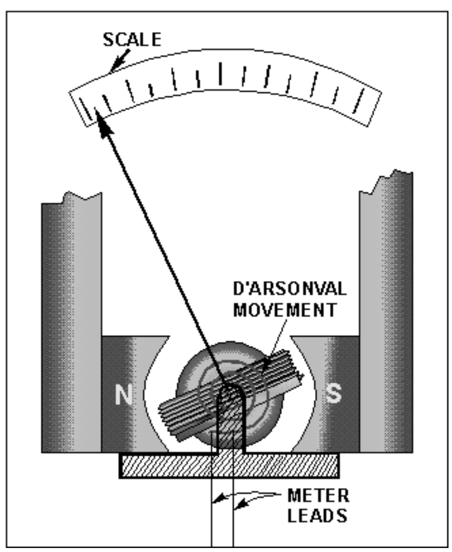


Figure 1-10.—A meter using d'Arsonval movement.

Q7. What type of meter movement is the d'Arsonval meter movement?

Q8. What is the effect of current flow through the coil in a d'Arsonval meter movement?

Q9. What are three functions of the hairsprings in a d'Arsonval meter movement?

COMPASS AND ALTERNATING CURRENT

Up to this point, only direct current examples have been used. What happens with the use of alternating current? Figure 1-11 shows a magnet close to a conductor carrying alternating current at a frequency of 1 hertz.

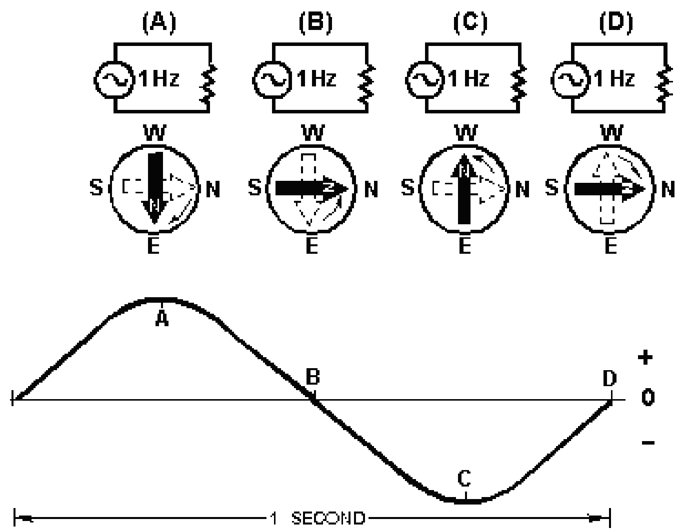


Figure 1-11.—Compass and conductor with ac.

The compass needle will swing toward the east part of the compass (down) as the current goes positive, as represented in figure 1-11(A). (The sine wave of the current is shown in the lower portion of the figure to help you visualize the current in the conductor.)

In figure 1-11(B), the current returns to zero, and the compass needle returns to magnetic north (right). As the current goes negative, as in figure 1-11(C), the compass needle swings toward the west portion of the compass (up). The compass needle returns to magnetic north as the current returns to zero as shown in figure 1-11(D).

This cycle of the current going positive and negative and the compass swinging back and forth will continue as long as there is alternating current in the conductor.

If the frequency of the alternating current is increased, the compass needle will swing back and forth at a higher rate (faster). At a high enough frequency, the compass needle will not swing back and forth, but simply vibrate around the magnetic north position. This happens because the needle cannot react fast enough to the very rapid current alternations. The compass (a simple meter) will indicate the average value of the alternating current (remember the average value of a sine wave is zero) by vibrating around the zero point on the meter (magnetic north). This is not of much use if you wish to know the value of the alternating current. Some device, such as a rectifier, is needed to allow the compass to react to the alternating current in a way that can be useful in measuring the current.

RECTIFIER FOR AC MEASUREMENT

A rectifier is a device that changes alternating current to a form of direct current. The way in which this is done will be covered later in this training series. For now, it is necessary to know only the information presented in figure 1-12.

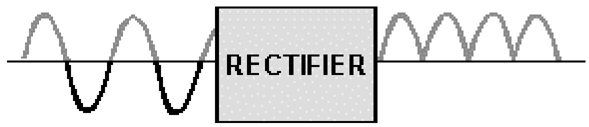


Figure 1-12.—Rectifier action.

Figure 1-12 shows that an alternating current passed through a rectifier will come out as a "pulsating direct current."

What happens to the compass now? Figure 1-13 answers that question.

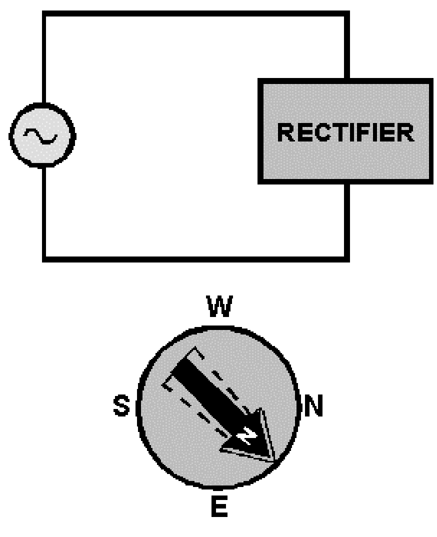


Figure 1-13.—Compass and conductor; rectified ac.

When the compass is placed close to the wire and the frequency of the alternating current is high enough, the compass will vibrate around a point that represents the average value of the pulsating direct current, as shown in figure 1-13.

Q10. How would a compass react when placed close to a conductor carrying alternating current at a low frequency?

Q11. How would the compass react if the alternating current through the conductor was a high frequency?

Q12. What is the purpose of a rectifier in a meter?

By connecting a rectifier to a d'Arsonval meter movement, an alternating current measuring device is created.

When ac is converted to pulsating dc, the d'Arsonval movement will react to the average value of the pulsating dc (which is the average value of one-half of the sine wave). Another characteristic of using a rectifier concerns the fact that the d'Arsonval meter movement is capable of indicating current in only one direction. If the d'Arsonval meter movement were used to indicate alternating current without a rectifier, or direct current of the wrong polarity, the movement would be severely damaged. The pulsating dc is current in a single direction, and so the d'Arsonval meter movement can be used as long as proper polarity is observed.

DAMPING

A problem that is created by the use of a rectifier and d'Arsonval meter movement is that the pointer will vibrate (oscillate) around the average value indication. This oscillation will make the meter difficult to read.

The process of "smoothing out" the oscillation of the pointer is known as DAMPING. There are two basic techniques used to damp the pointer of a d'Arsonval meter movement.

The first method of damping comes from the d'Arsonval meter movement itself. In the d'Arsonval meter movement, current through the coil causes the coil to move in the magnetic field of the permanent magnet. This movement of the coil (conductor) through a magnetic field causes a current to be induced in the coil opposite to the current that caused the movement of the coil. This induced current will act to damp oscillations. In addition to this method of damping, which comes from the movement itself, most meters use a second method of damping.

The second method of damping used in most meter movements is an airtight chamber containing a vane (like a windmill vane) attached to the coil (fig.1-14).

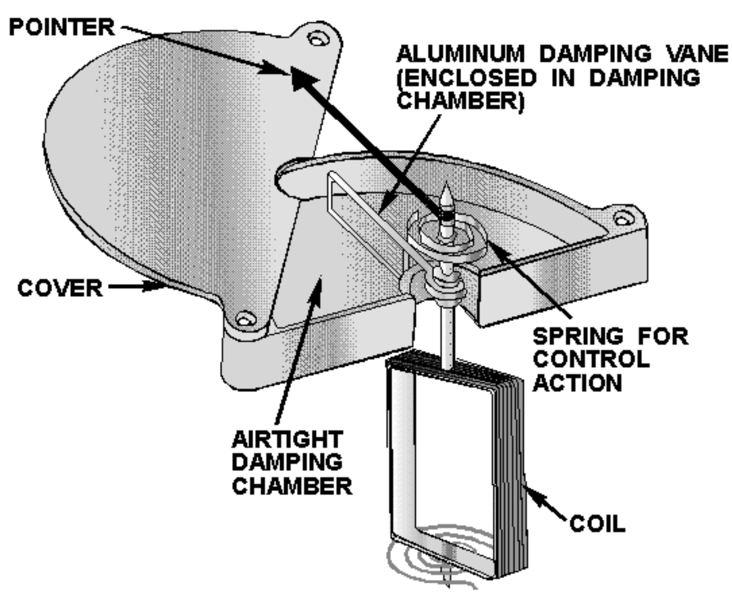


Figure 1-14.—A typical meter damping system.

As the coil moves, the vane moves within the airtight chamber. The action of the vane against the air in the chamber opposes the coil movement and damps the oscillations.

Q13. How can a d'Arsonval meter movement be adapted for use as an ac meter?

Q14. What is damping?

Q15. What are two methods used to damp a meter movement?

Q16. What value does a meter movement react to (actually measure) when measuring ac?

Q17. What value is indicated on the scale of an ac meter?

An additional advantage of damping a meter movement is that the damping systems will act to slow down the coil and help keep the pointer from overshooting its rest position when the current through the meter is removed.

INDICATING ALTERNATING CURRENT

Another problem encountered in measuring ac is that the meter movement reacts to the average value of the ac. The value used when working with ac is the effective value (rms value). Therefore, a different scale is used on an ac meter. The scale is marked with the effective value, even though it is the average value to which the meter is reacting. That is why an ac meter will give an incorrect reading if used to measure dc.

OTHER METER MOVEMENTS

The d'Arsonval meter movement (permanent-magnet moving-coil) is only one type of meter movement. Other types of meter movements can be used for either ac or dc measurement without the use of a rectifier.

When galvanometers were mentioned earlier in this topic, it was stated that they could be either electromagnetic or electrodynamic. Electrodynamic meter movements will be discussed at this point.

ELECTRODYNAMIC METER MOVEMENT

An electrodynamic movement uses the same basic operating principle as the basic moving-coil meter movement, except that the permanent magnet is replaced by fixed coils (fig. 1-15). A moving coil, to which the meter pointer is attached, is suspended between two field coils and connected in series with these coils. The three coils (two field coils and the moving coil) are connected in series across the meter terminals so that the same current flows through each.

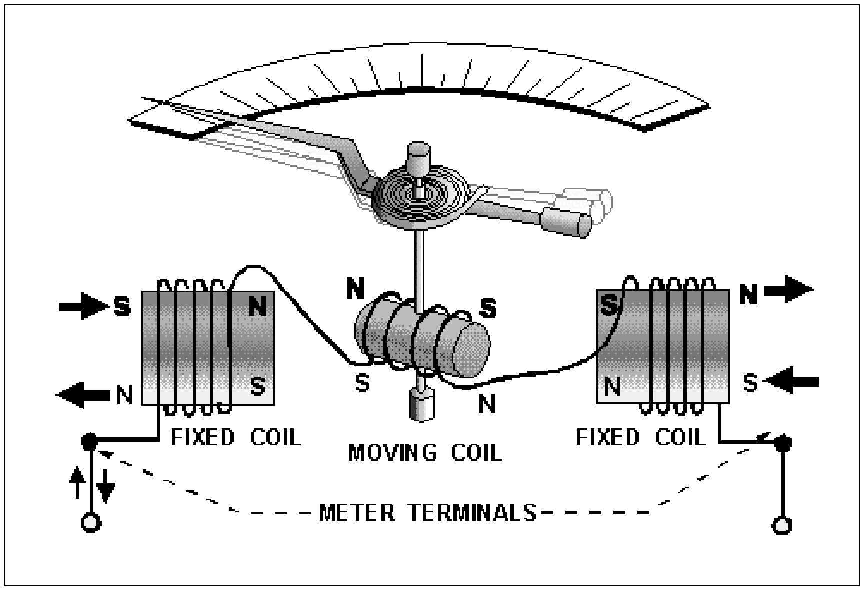


Figure 1-15.—Electrodynamic meter movement.

Current flow in either direction through the three coils causes a magnetic field to exist between the field coils. The current in the moving coil causes it to act as a magnet and exert a turning force against a spring. If the current is reversed, the field polarity and the polarity of the moving coil reverse at the same time, and the turning force continues in the original direction. Since reversing the current direction does not reverse the turning force, this type of meter can be used to measure both ac and dc if the scale is changed. While some voltmeters and ammeters use the electrodynamic principle of operation, the most important application is in the wattmeter. The wattmeter, along with the voltmeter and the ammeter, will be discussed later in this topic.

MOVING-VANE METER MOVEMENTS

The moving-vane meter movement (sometimes called the moving-iron movement) is the most commonly used movement for ac meters. The moving-vane meter operates on the principle of magnetic repulsion between like poles (fig.1-16). The current to be measured flows through a coil, producing a magnetic field which is proportional to the strength of the current. Suspended in this field are two iron vanes. One is in a fixed position, the other, attached to the meter pointer, is movable. The magnetic field magnetizes these iron vanes with the same polarity regardless of the direction of current flow in the coil. Since like poles repel, the movable vane pulls away from the fixed vane, moving the meter pointer. This motion exerts a turning force against the spring. The distance the vane will move against the force of the spring depends on the strength of the magnetic field, which in turn depends on the coil current.

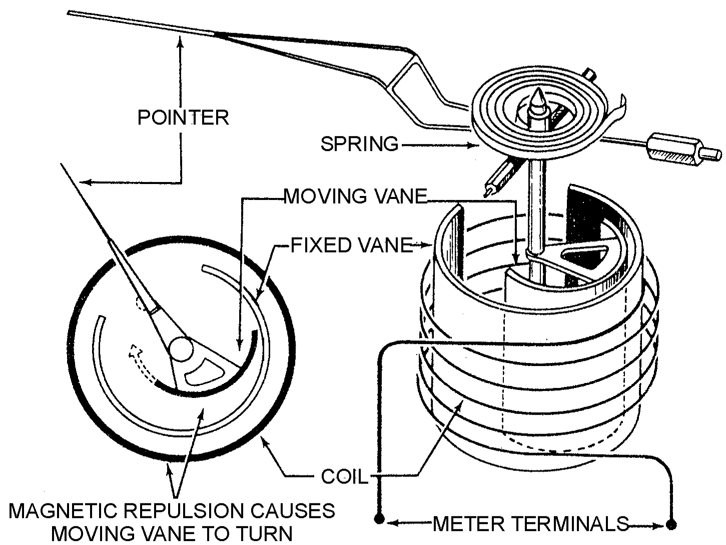


Figure 1-16.—Moving-vane meter movement.

These meters are generally used at 60-hertz ac, but may be used at other ac frequencies. By changing the meter scale to indicate dc values rather than ac rms values, moving-vane meters will measure dc current and dc voltage. This is not recommended due to the residual magnetism left in the vanes, which will result in an error in the instrument.

One of the major disadvantages of this type of meter movement occurs due to the high reluctance of the magnetic circuit. This causes the meter to require much more power than the D'Arsonval meter to produce a full scale deflection, thereby reducing the meters sensitivity.

HOT-WIRE AND THERMOCOUPLE METER MOVEMENTS

Hot-wire and thermocouple meter movements both use the heating effect of current flowing through a resistance to cause meter deflection. Each uses this effect in a different manner. Since their operation depends only on the heating effect of current flow, they may be used to measure both direct current and alternating current of any frequency on a single scale.

The hot-wire meter movement deflection depends on the expansion of a high-resistance wire caused by the heating effect of the wire itself as current flows through it. (See fig. 1-17.) A resistance wire is stretched taut between the two meter terminals, with a thread attached at a right angle to the center of the wire. A spring connected to the opposite end of the thread exerts a constant tension on the resistance wire. Current flow heats the wire, causing it to expand. This motion is transferred to the meter pointer through the thread and a pivot.

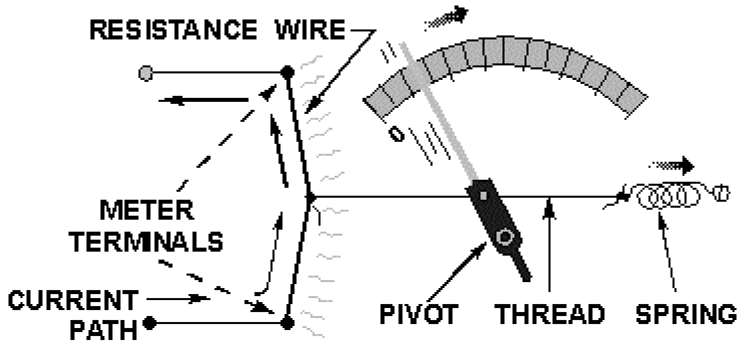


Figure 1-17.—Hot-wire meter movement.

The thermocouple meter consists of a resistance wire across the meter terminals, which heats in proportion to the amount of current. (See fig. 1-18.) Attached to this wire is a small thermocouple junction of two unlike metal wires, which connect across a very sensitive dc meter movement (usually a d'Arsonval meter movement). As the current being measured heats the heating resistor, a small current (through the thermocouple wires and the meter movement) is generated by the thermocouple junction. The current being measured flows through only the resistance wire, not through the meter movement itself. The pointer turns in proportion to the amount of heat generated by the resistance wire.

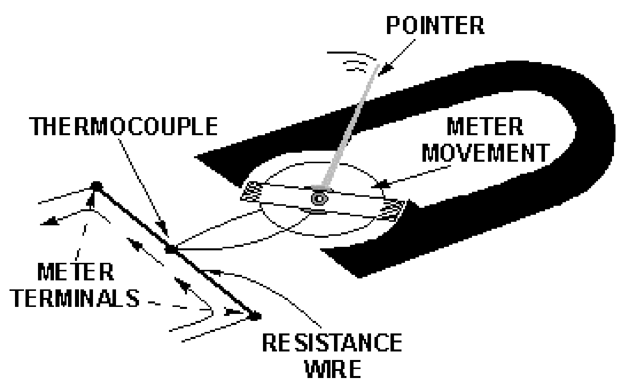


Figure 1-18.—A thermocouple meter.

Q18. List three meter movements that can measure either ac or dc without the use of a rectifier.

Q19. What electrical property is used by all the meter movements discussed so far?

AMMETERS

An ammeter is a device that measures current. Since all meter movements have resistance, a resistor will be used to represent a meter in the following explanations. Direct current circuits will be used for simplicity of explanation.

AMMETER CONNECTED IN SERIES

In figure 1-19(A), R_1 and R_2 are in series. The total circuit resistance is $R_2 + R_2$ and total circuit current flows through both resistors. In figure 1-19(B), R_1 and R_2 are in parallel. The total circuit resistance is

$$\frac{1}{\frac{1}{R_1} + \frac{1}{R_2}}$$

and total circuit current does not flow through either resistor.

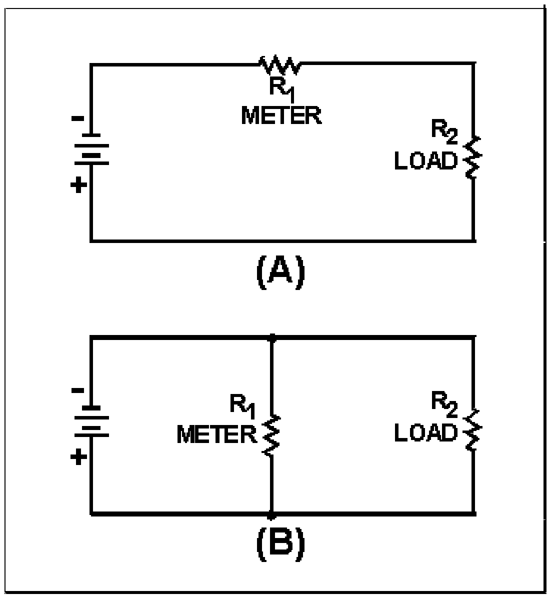


Figure 1-19.—A series and a parallel circuit.

If R_1 represents an ammeter, the only way in which total circuit current will flow through the meter (and thus be measured) is to have the meter (R_1) in series with the circuit load (R_2), as shown in figure 1-19(A).

In complex electrical circuits, you are not always concerned with total circuit current. You may be interested in the current through a particular component or group of components. In any case, an ammeter is always connected in series with the circuit you wish to test. Figure 1-20 shows various circuit arrangements with the ammeter(s) properly connected for measuring current in various portions of the circuit.

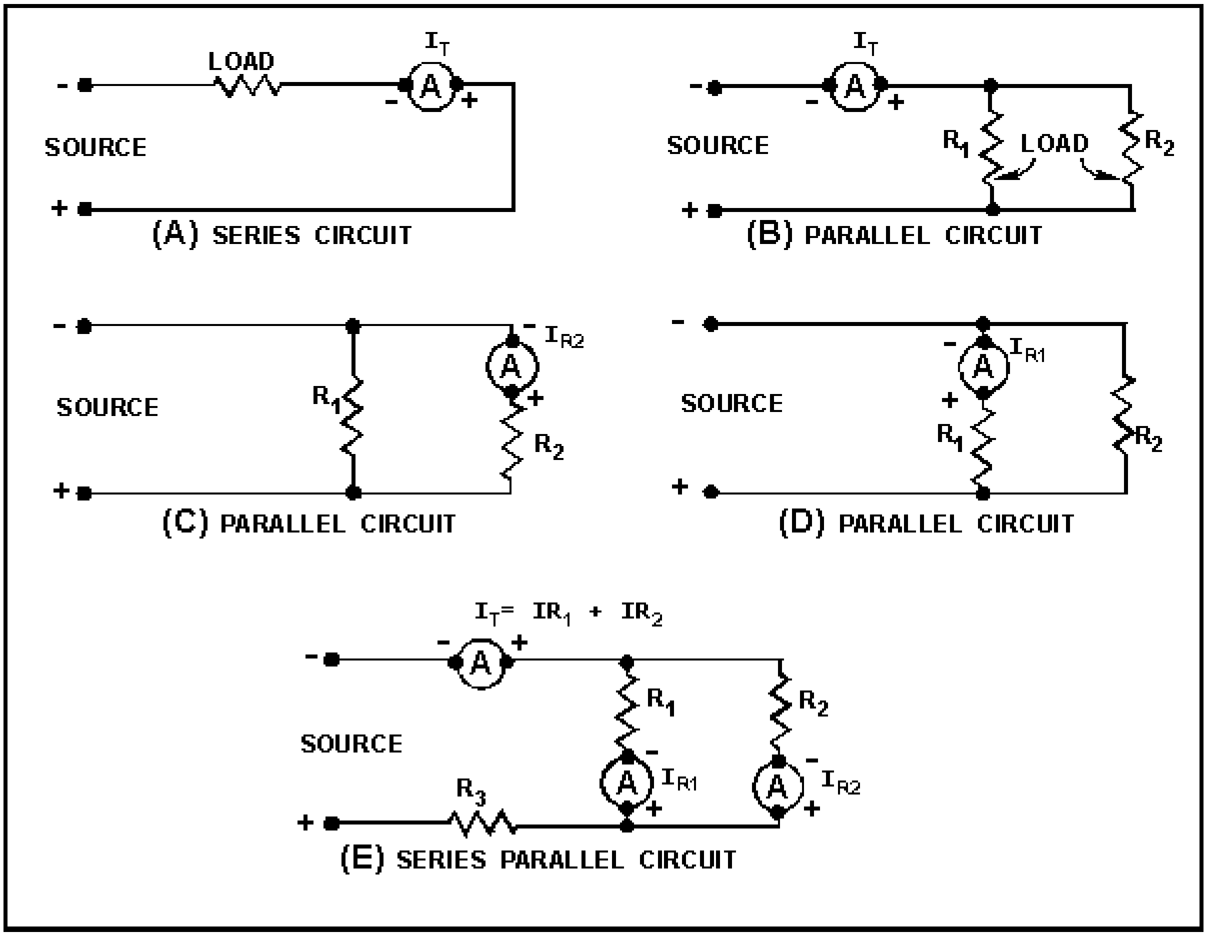


Figure 1-20.—Proper ammeter connections.

Connecting an ammeter in parallel would give you not only an incorrect measurement, it would also damage the ammeter, because too much current would pass through the meter.

EFFECT ON CIRCUIT BEING MEASURED

The meter affects the circuit resistance and the circuit current. If R_1 is removed from the circuit in figure 1-19(A), the total circuit resistance is R_2. Circuit current

$$(I) \text{ equals } \frac{E}{R_2},$$

with the meter (R_1) in the circuit, circuit resistance is $R_1 + R_2$ and circuit current

$$(I) \frac{E}{R_2 + R_2}.$$

The smaller the resistance of the meter (R_1), the less it will affect the circuit being measured. (R_1 represents the total resistance of the meter; not just the resistance of the meter movement.)

AMMETER SENSITIVITY

Ammeter sensitivity is the amount of current necessary to cause full scale deflection (maximum reading) of the ammeter. The smaller the amount of current, the more "sensitive" the ammeter. For example, an ammeter with a maximum current reading of 1 milliampere would have a sensitivity of 1 milliampere, and be more sensitive than an ammeter with a maximum reading of 1 ampere and a sensitivity of 1 ampere. Sensitivity can be given for a meter movement, but the term "ammeter sensitivity" usually refers to the entire ammeter and not just the meter movement. An ammeter consists of more than just the meter movement.

AMMETER RANGES

If you have a meter movement with a sensitivity of 1 milliampere, you can connect it in series with a circuit and measure currents up to 1 milliampere. But what do you do to measure currents over 1 milliampere?

To answer this question, look at figure 1-21. In figure 1-21(A), 10 volts are applied to two resistors in parallel. R_1 is a 10-ohm resistor and R_2 is a 1.11-ohm resistor. Since voltage in parallel branches is equal-

$$I_{R1} = \frac{10V}{10\Omega} = 1A$$

$$I_{R2} = \frac{10V}{1.11\Omega} = 9A$$

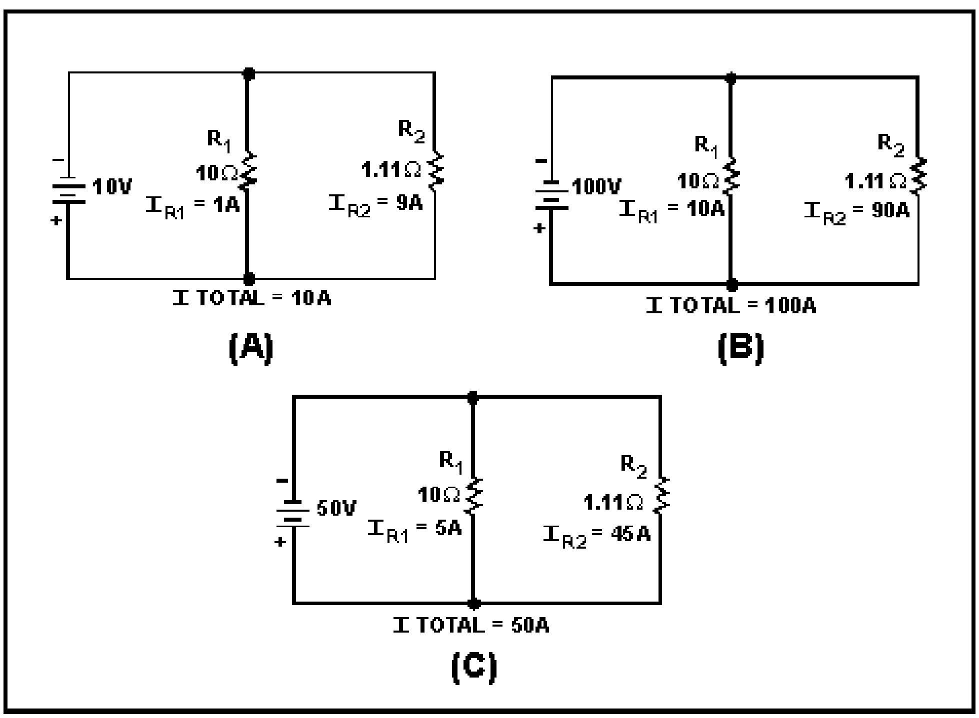


Figure 1-21.—Current in a parallel circuit.

In figure 1-21(B), the voltage is increased to 100 volts. Now,

$$I_{R1} = \frac{100V}{10\Omega} = 10A$$

$$I_{R2} = \frac{100V}{1.11\Omega} = 90A$$

In figure 1-21(C), the voltage is reduced from 100 volts to 50 volts. In this case,

$$I_{R1} = \frac{50V}{10\Omega} = 5A$$

$$I_{R2} = \frac{50V}{1.11\Omega} = 45A$$

Notice that the relationship (ratio) of I_{R1} and I_{R2} remains the same. I_{R2} is nine times greater than I_{R1} and I_{R1} has one-tenth of the total current.

If R_1 is replaced by a meter movement that has 10 ohms of resistance and a sensitivity of 10 amperes, the reading of the meter will represent one-tenth of the current in the circuit and R_2 will carry nine-tenths of the current. R_2 is a SHUNT resistor because it diverts, or shunts, a portion of the current from the meter movement (R_1). By this method, a 10-ampere meter movement will measure current up to 100 amperes. By adding a second scale to the face of the meter, the current can be read directly.

By adding several shunt resistors in the meter case, with a switch to select the desired resistor, the ammeter will be capable of measuring several different maximum current readings or ranges.

Most meter movements in use today have sensitivities of from 5 microamperes to 1 milliampere. Figure 1-22 shows the circuit of meter switched to higher ranges, the shunt an ammeter that uses a meter movement with a sensitivity of 100 microamperes and shunt resistors. This ammeter has five ranges (100 microamperes; 1, 10, and 100 milliamperes; 1 ampere) selected by a switch.

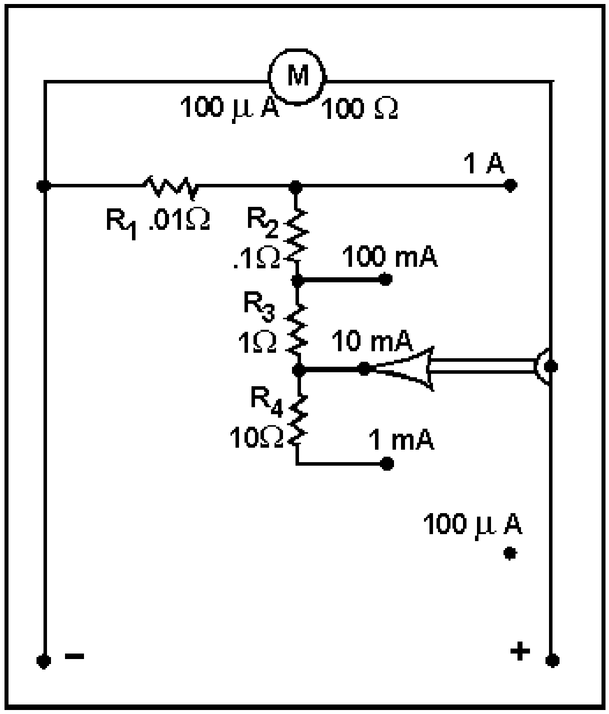


Figure 1-22.—An ammeter with internal shunt resistors.

By adding several shunt resistors in the meter case, with a switch to select the desired resistor, the ammeter will be capable of measuring several different maximum current readings or ranges.

Most meter movements in use today have sensitivities of from 5 microamperes to 1 milliampere. Figure 1-22 shows the circuit of meter switched to higher ranges, the shunt an ammeter that uses a meter movement with a sensitivity of 100 microamperes and shunt resistors. This ammeter has five ranges (100 microamperes; 1, 10, and 100 milliamperes; 1 ampere) selected by a switch.

With the switch in the 100 microampere position, all the current being measured will go through the meter movement. None of the current will go through any of the shunt resistors. If the ammeter is switched to the 1 milliampere position, the current being measured will have parallel paths of the meter movement and all the shunt resistors (R_1 , R_2, R_3, and R_4). Now, only a portion of the current will go through the meter movement and the rest of the current will go through the shunt resistors. When the meter is switched to the 10-milliampere position (as shown in fig. 1-22), only resistors R_1, R_2, and R_3 shunt the meter. Since the resistance of the shunting resistance is less than with R_4 in the circuit (as was the case in the 1-milliampere position), more current will go through the shunt resistors and less current will go through the meter movement. As the resistance decreases and more current goes through the shunt resistors. As long as the current to be measured does not exceed the range selected, the meter movement will never have more than 100 microamperes of current through it.

Shunt resistors are made with close tolerances. That means if a shunt resistor is selected with a resistance of .01 ohms (as R_1 in fig. 1-22), the actual resistance of that shunt resistor will not vary from that value by more than 1 percent. Since a shunt resistor is used to protect a meter movement and to allow accurate measurement, it is important that the resistance of the shunt resistor is known very accurately.

Figure 1-22 represents an ammeter with internal shunts. The shunt resistors are inside the meter case and selected by a switch. For limited current ranges (below 50 amperes), internal shunts are most often employed.

For higher current ranges (above 50 amperes) ammeters that use external shunts are used. The external shunt resistor serves the same purpose as the internal shunt resistor. The external shunt is connected in series with the circuit to be measured and in parallel with the ammeter. This shunts (bypasses) the ammeter so only a portion of the current goes through the meter. Each external shunt will be marked with the maximum current value that the ammeter will measure when that shunt is used. Figure 1-23 shows an ammeter that is designed to use external shunts and a d'Arsonval meter movement. Figure 1-23(A) shows the internal construction of the meter and the way in which the external shunt is connected to the meter and to the circuit being measured. Figure 1-23(C) shows some typical external shunts.

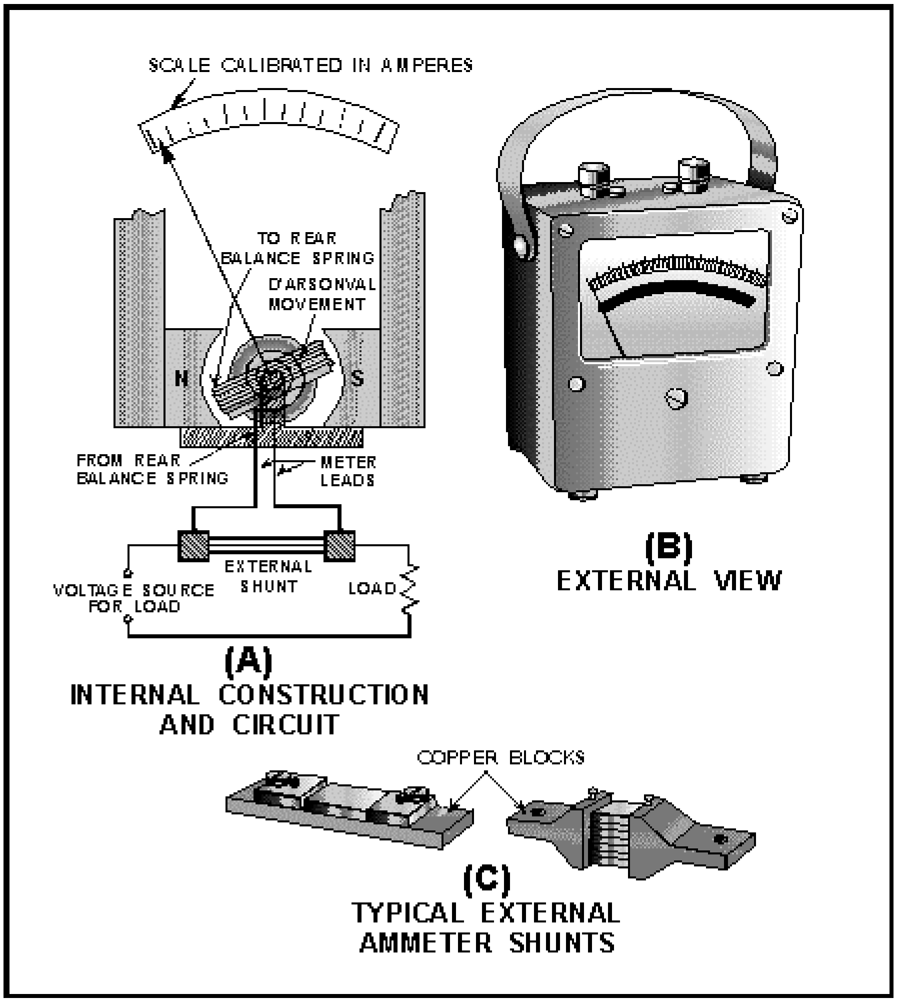


Figure 1-23.—An ammeter employing the d'Arsonval principle and external shunts.

A shunt resistor is nothing more than a resistor in parallel with the meter movement. To measure high currents, very small resistance shunts are used so the majority of the current will go through the shunt. Since the total resistance of a parallel circuit (the meter movement and shunt resistor) is always less than the resistance of the smallest resistor, as an ammeter's range is increased, its resistance decreases. This is important because the load resistance of high-current circuits is smaller than the load resistance of low-current circuits. To obtain accurate measurements, it is necessary that the ammeter resistance be much less than the load resistance, since the ammeter is connected in series with the load.

Q20. What electrical property does an ammeter measure?

Q21. How is an ammeter connected to the circuit under test?

Q22. How does an ammeter affect the circuit being measured?

Q23. How is the ammeter's effect on the circuit being measured kept to a minimum?

Q24. What is ammeter sensitivity?

Q25. What is used to allow an ammeter to measure different ranges?

Range Selection

Part of the correct use of an ammeter is the proper use of the range selection switch. If the current to be measured is larger than the scale of the meter selected, the meter movement will have excessive current and will be damaged. Therefore, it is important to always start with the highest range when you use an ammeter. If the current can be measured on several ranges, use the range that results in a reading near the middle of the scale. Figure 1-24 illustrates these points.

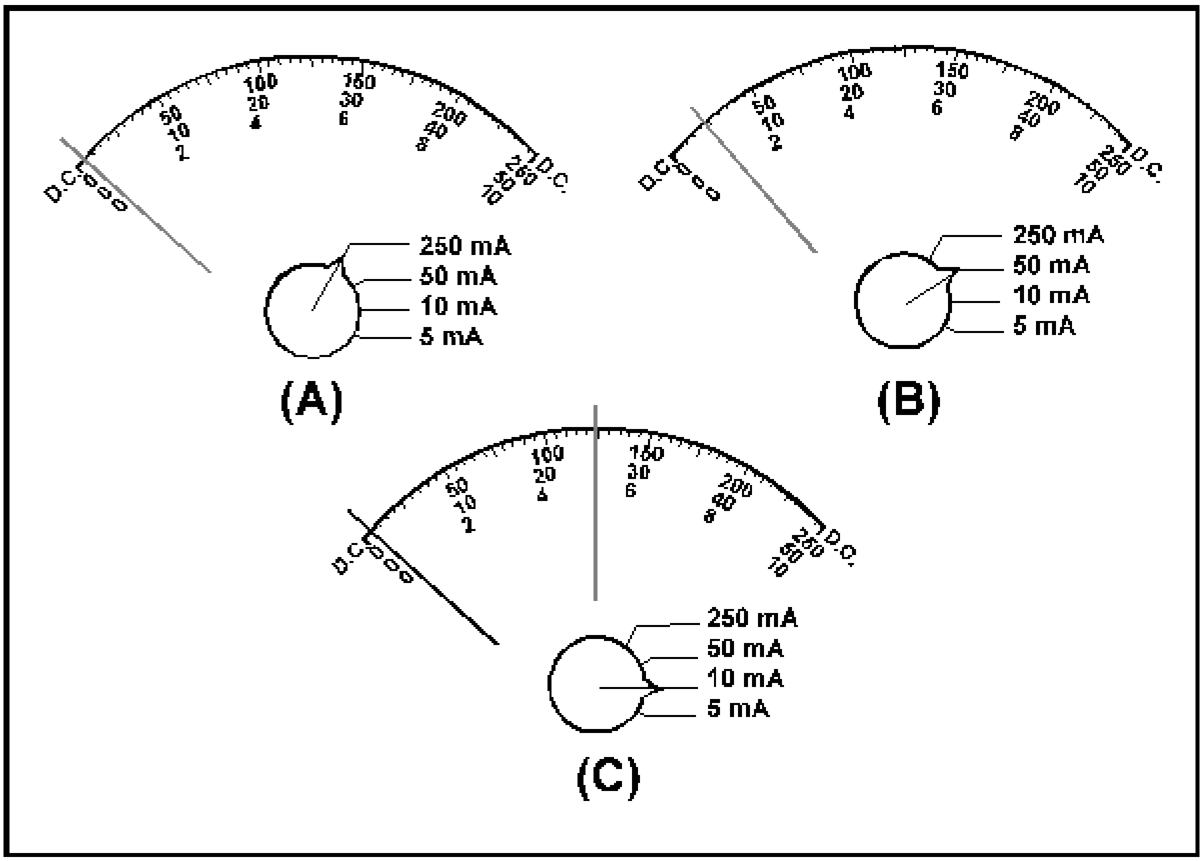


Figure 1-24.—Reading an ammeter at various ranges.

Figure 1-24(A) shows the initial reading of a circuit. The highest range (250 milliamperes) has been selected and the meter indication is very small. It would be difficult to properly interpret this reading with any degree of accuracy. Figure 1-24(B) shows the second reading, with the next largest range (50 milliamperes). The meter deflection is a little greater. It is possible to interpret this reading as 5 milliamperes. Since this approximation of the current is less than the next range, the meter is switched as shown in figure 1-24(C). The range of the meter is now 10 milliamperes and it is possible to read the meter indication of 5 milliamperes with the greatest degree of accuracy. Since the current indicated is equal to (or greater than) the next range of the ammeter (5 milliamperes), the meter should NOT be switched to the next range.

AMMETER SAFETY PRECAUTIONS

When you use an ammeter, certain precautions must be observed to prevent injury to yourself or others and to prevent damage to the ammeter or the equipment on which you are working. The following list contains the MINIMUM precautions to observe when using an ammeter.

- Ammeters must always be connected in series with the circuit under test.
- Always start with the highest range of an ammeter.
- Deenergize and discharge the circuit completely before you connect or disconnect the ammeter.
- In dc ammeters, observe the proper circuit polarity to prevent the meter from being damaged.
- Never use a dc ammeter to measure ac.
- Observe the general safety precautions of electrical and electronic devices.

Q26. Why should you use the highest range of an ammeter for the initial measurement?

Q27. What range of an ammeter is selected for the final measurement?

Q28. List the six safety precautions for the use of ammeters.

Q29. Why will an ammeter be damaged if connected in parallel with the circuit to be measured?

VOLTMETERS

All the meter movements discussed so far react to current, and you have been shown how ammeters are constructed from those meter movements. It is often necessary to measure circuit properties other than current. Voltage measurement, for example, is accomplished with a **VOLTMETER**.

VOLTMETERS CONNECTED IN PARALLEL

While ammeters are always connected in series, voltmeters <u>are always connected in parallel</u>. Figure 1-25 (and the following figures) use resistors to represent the voltmeter movement. Since a meter movement can be considered as a resistor, the concepts illustrated are true for voltmeters as well as resistors. For simplicity, dc circuits are shown, but the principles apply to both ac and dc voltmeters.

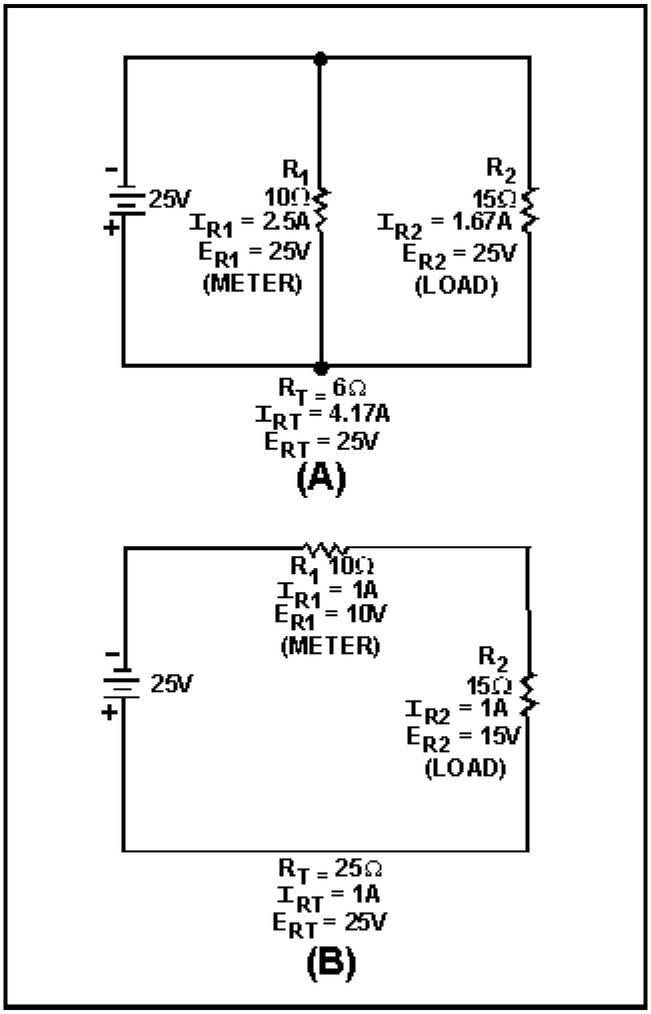


Figure 1-25.—Current and voltage in series and parallel circuits.

Figure 1-25(A) shows two resistors connected in parallel. Notice that the voltage across both resistors is equal. In figure 1-25(B) the same resistors are connected in series. In this case, the voltage across the resistors is not equal. If R_1 represents a voltmeter, the only way in which it can be connected to measure the voltage of R_2 is in parallel with R_2, as in figure 1-25(A).

LOADING EFFECT

A voltmeter has an effect on the circuit being measured. This is called **LOADING** the circuit. Figure 1-26 illustrates the loading effect and the way in which the loading effect is kept to a minimum.

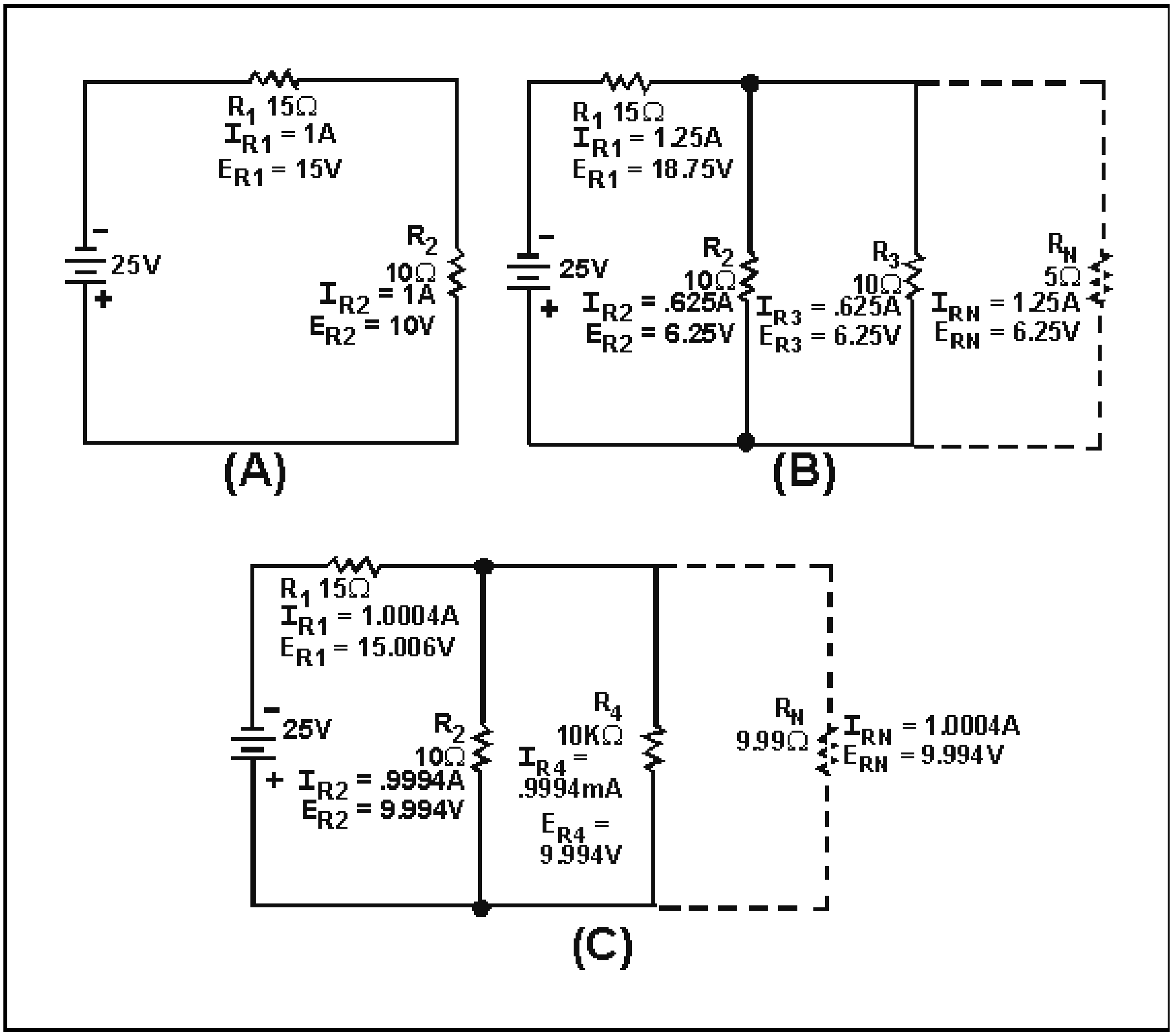


Figure 1-26.—The loading effect.

In figure 1-26(A), a series circuit is shown with R_1 equaling 15 ohms and R_2 equaling 10 ohms. The voltage across R_2 (E_{R2}) equals 10 volts. If a meter (represented by R_3) with a resistance of 10 ohms is connected in parallel with R_2, as in figure 1-26(B), the combined resistance of R_2 and R_3 (R_n) is equal to 5 ohms. The voltage across R_2 and R_3 is now 6.25 volts, and that is what the meter will indicate. Notice that the voltage across R_1 and the circuit current have both increased. The addition of the meter (R_3) has loaded the circuit.

In figure 1-26(C), the low-resistance meter (R_3) is replaced by a higher resistance meter (R_4) with a resistance of 10 kilohms. The combined resistance of R_2 and R_4 (R_n) is equal to 9.99 ohms. The voltage across R_2 and R_4 is now 9.99 volts, the value that will be indicated on the meter. This is much closer to the voltage across R_2, with no meter (R_3 or R_4) in the circuit. Notice that the voltage across R, and the circuit current in figure 1-26(C) are much closer to the values in 1-26(A). The current (I_{R4}) through the meter (R_4) in figure 1-26(C) is also very small compared to the current (I_{R2}) through R_2. In figure 1-26(C) the meter (R_4) has much less effect on the circuit and does not load the circuit as much. Therefore, <u>a voltmeter should have a high resistance compared to the circuit being measured, to minimize the loading effect</u>.

Q30. What electrical quantity is measured by a voltmeter?

Q31. How is a voltmeter connected to the circuit to be measured?

Q32. What is the loading effect of a voltmeter?

Q33. How is the loading effect of a voltmeter kept to a minimum?

MAKING A VOLTMETER FROM A CURRENT SENSITIVE METER MOVEMENT

The meter movements discussed earlier in this chapter have all reacted to current. Various ways have been shown in which these movements can be used in ammeters. If the current and resistance are known, the voltage can be calculated by the formula $E = IR$. A meter movement has a known resistance, so as the movement reacts to the current, the voltage can be indicated on the scale of the meter.

In figure 1-27(A), a voltmeter (represented by R_2) connected across a 10-ohm resistor with 10 volts applied. The current through the voltmeter (R_2) is .1 milliamperes. In figure 1-27(B), the voltage is increased to 100 volts. Now, the current through the voltmeter (R_2) is 1 milliampere. The voltage has increased by a factor of 10 and so has the current. This illustrates that the current through the meter is proportional to the voltage being measured.

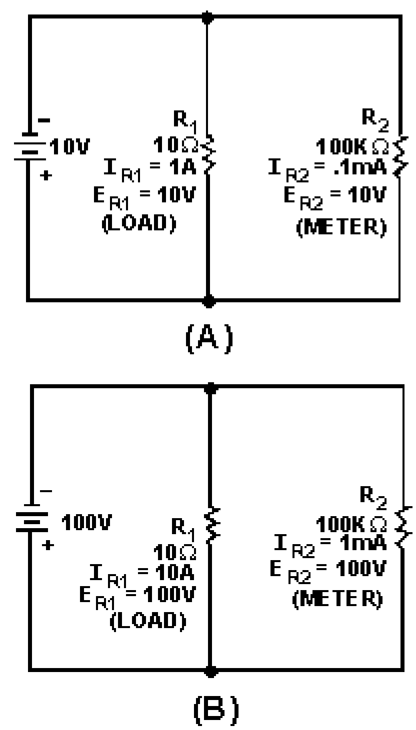


Figure 1-27.—Current and voltage in parallel circuit.

SENSITIVITY OF VOLTMETERS

Voltmeter sensitivity is expressed in ohms per volt (Ω/V). It is the resistance of the voltmeter at the full-scale reading in volts. Since the voltmeter's resistance does not change with the position of the pointer, the total resistance of the meter is the sensitivity multiplied by the full-scale voltage reading. The higher the sensitivity of a voltmeter, the higher the voltmeter's resistance. Since high resistance voltmeters have less loading effect on circuits, a high-sensitivity meter will provide a more accurate voltage measurement.

To determine the sensitivity of a meter movement, you need only to divide 1 by the amount of current needed to cause full-scale deflection of the meter movement. The manufacturer usually marks meter movements with the amount of current needed for full-scale deflection and the resistance of the meter. With these figures, you can calculate the sensitivity

$$\left(\frac{1}{\text{full - scale current}}\right)$$

and the full-scale voltage reading full-scale current (full-scale current × resistance).

For example, if a meter has a full-scale current of 50µA and a resistance of 960Ω, the sensitivity could be calculated as:

$$\text{Sensitivity} = \frac{1}{\text{full - scale current}}$$

$$\text{Sensitivity} = \frac{1}{50\mu A}$$

$$\text{Sensitivity} = 20k\Omega\ /\text{volt}$$

The full-scale voltage reading would be calculated as:

Full-scale voltage reading = full-scale current × resistance

Full-scale voltage reading = 50µA × 960Ω

Full-scale voltage reading = 48mV

RANGES

Table 1-1 shows the figures for most meter movements in use today.

Table 1-1.—Meter Movement Characteristics

CURRENT TO DEFLECT FULL SCALE	RESISTANCE	SENSITIVITY	VOLTAGE FULL SCALE
1mA	100 Ω	1 k Ω/VOLT	.1 V
50 µA	960 Ω	20 k Ω/VOLT	.048 V
5 µA	5750 Ω	200 k Ω/VOLT	.029 V

Notice that the meter movements shown in table 1-1 will indicate .029 volts to .I volt at full scale, and the sensitivity ranges from 1000 ohms per volt to 200,000 ohms per volt. The higher sensitivity meters indicate smaller amounts of voltage. Since most voltage measurements involve voltage larger than .1 volt, a method must be used to extend the voltage reading.

Figure 1-28 illustrates the method of increasing the voltage range of a voltmeter.

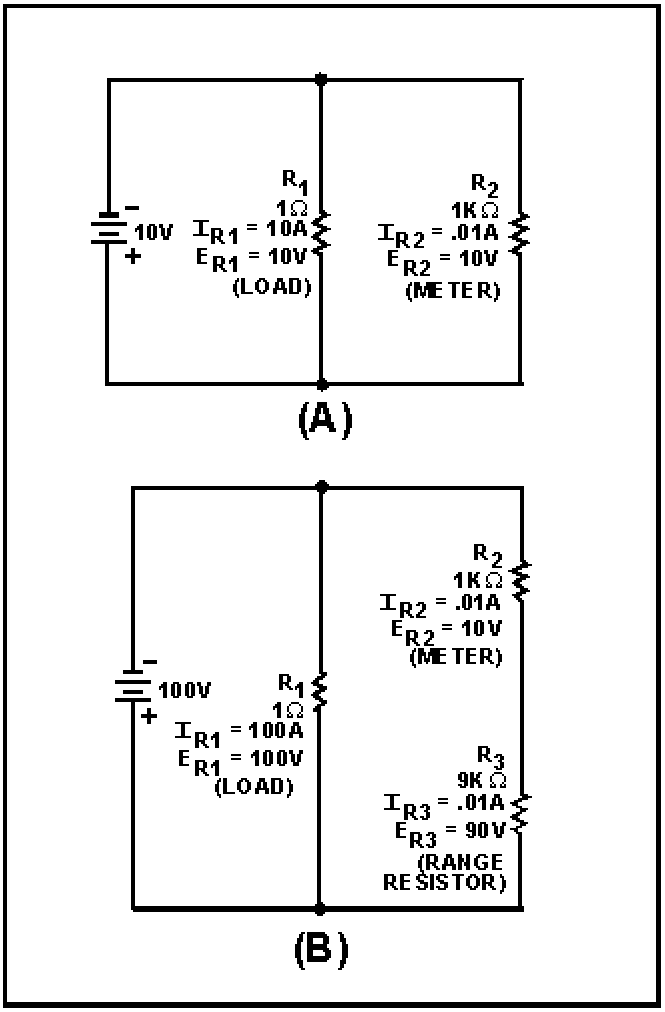


Figure 1-28.—A voltmeter and a range resistor.

In figure 1-28(A), a voltmeter with a range of 10 volts and a resistance of 1 kilohm (R_2) is connected in parallel to resistor R_1. The meter has .01 ampere of current (full-scale deflection) and indicates 10 volts. In figure 1-28(B), the voltage has been increased to 100 volts. This is more than the meter can measure. A 9 kilohm resistor (R_3) is connected in series with the meter (R_2). The meter (R_2) now has .01 ampere of current (full-scale deflection). But since R_3 has increased the voltage capability of the meter, the meter indicates 100 volts. R_3 has changed the range of the meter.

Voltmeters can be constructed with several ranges by the use of a switch and internal resistors. Figure 1-29 shows a voltmeter with a meter movement of 100 ohms and 1 milliampere full-scale deflection with 5 ranges of voltage through the use of a switch. In this way a voltmeter can be used to measure several different ranges of voltage.

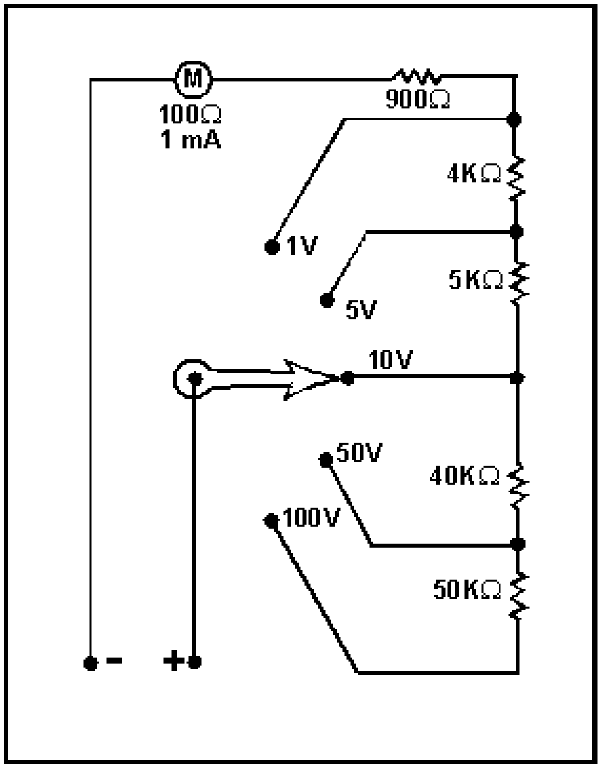


Figure 1-29.—A voltmeter with internal range resistors.

The current through the meter movement is determined by the voltage being measured. If the voltage measured is higher than the range of the voltmeter, excess current will flow through the meter movement and the meter will be damaged. Therefore, you should always start with the highest range of a voltmeter and switch the ranges until a reading is obtained near the center of the scale. Figure 1-30 illustrates these points.

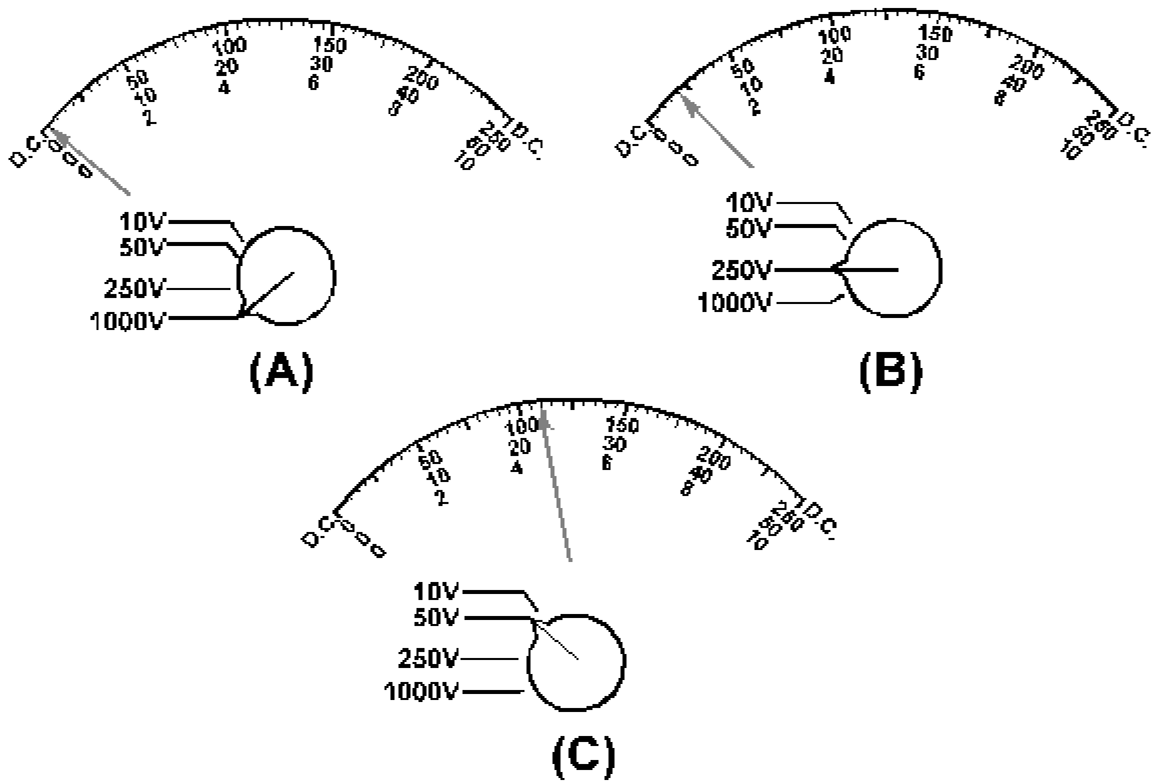


Figure 1-30.—Reading a voltmeter at various ranges.

In figure 1-30(A) the meter is in the 1000-volt range. The pointer is barely above the 0 position. It is not possible to accurately read this voltage. In figure 1-30(B) the meter is switched to the 250 volt range. From the pointer position it is possible to approximate the voltage as 20 volts. Since this is well below the next range, the meter is switched, as in figure 1-30(C). With the meter in the 50-volt range, it is possible to read the voltage as 22 volts. Since this is more than the next range of the meter (10 volts), the meter would not be switched to the next (lower) scale.

Q34. How is it possible to use a current sensitive meter movement to measure voltage?

Q35. What is voltmeter sensitivity?

Q36. What method is used to allow a voltmeter to have several ranges?

Q37. Why should you always use the highest range when connecting a voltmeter to a circuit?

ELECTROSTATIC METER MOVEMENT

The final meter movement covered in this chapter is the ELECTROSTATIC METER MOVEMENT. The other meter movements you have studied all react to current, the electrostatic meter movement reacts to voltage.

The mechanism is based on the repulsion of like charges on the plates of a capacitor. The electrostatic meter movement is actually a large variable capacitor in which one set of plates is allowed to move. The movement of the plates is opposed by a spring attached to the plates. A pointer that indicates the value of the voltage is attached to these movable plates. As the voltage increases, the plates develop more torque. To develop sufficient torque, the plates must be large and closely spaced. A very high voltage is necessary to provide movement, therefore, electrostatic voltmeters are used only for **HIGH VOLTAGE** measurement.

VOLTMETER SAFETY PRECAUTIONS

Just as with ammeters, voltmeters require safety precautions to prevent injury to personnel and damage to the voltmeter or equipment. The following is a list of the **MINIMUM** safety precautions for using a voltmeter.

- Always connect voltmeters in parallel.
- Always start with the highest range of a voltmeter.
- Deenergize and discharge the circuit completely before connecting or disconnecting the voltmeter.
- In dc voltmeters, observe the proper circuit polarity to prevent damage to the meter.
- Never use a dc voltmeter to measure ac voltage.
- Observe the general safety precautions of electrical and electronic devices.

Q38. What type of meter movement reacts to voltage rather than current?

Q39. What is the only use for the voltage sensitive meter movement?

Q40. List the six safety precautions for the use of voltmeters.

OHMMETERS

The two instruments most commonly used to check the continuity (a complete circuit), or to measure the resistance of a circuit or circuit element, are the **OHMMETER** and the **MEGGER** (megohm meter). The ohmmeter is widely used to measure resistance and check the continuity of electrical circuits and devices. Its range usually extends to only a few megohms. The megger is widely used for measuring insulation resistance, such as between a wire and the outer surface of the insulation, and insulation resistance of cables and insulators. The range of a megger may extend to more than 1,000 megohms.

The ohmmeter consists of a dc ammeter, with a few added features. The added features are:

1. A dc source of potential (usually a 3-volt battery)
2. One or more resistors (one of which is variable)
3. A simple ohmmeter circuit is shown in figure 1-31.

The ohmmeter's pointer deflection is controlled by the amount of battery current passing through the moving coil. Before measuring the resistance of an unknown resistor or electrical circuit, the test leads of the ohmmeter are first shorted together, as shown in figure 1-31. With the leads shorted, the meter is calibrated for proper operation on the selected range. While the leads are shorted, meter current is maximum and the pointer deflects a maximum amount, somewhere near the zero position on the ohms scale. Because of this current through the meter with the leads shorted, it is necessary to remove the test leads when you are finished using the ohmmeter. If the leads were left connected, they could come in contact with each other and discharge the ohmmeter battery. When the variable resistor (rheostat) is adjusted properly, with the leads shorted, the pointer of the meter will come to rest exactly on the zero position. This indicates **ZERO RESISTANCE** between the test leads, which, in fact, are shorted together. The zero reading of a series-type ohmmeter is on the right-hand side of the scale, where as the zero reading for an ammeter or a voltmeter is generally to the left-hand side of the scale. (There is another type of ohmmeter which is discussed a little later on in this chapter.) When the test leads of an ohmmeter are separated, the pointer of the meter will return to the left side of the scale. The interruption of current and the spring tension act on the movable coil assembly, moving the pointer to the left side (∞) of the scale.

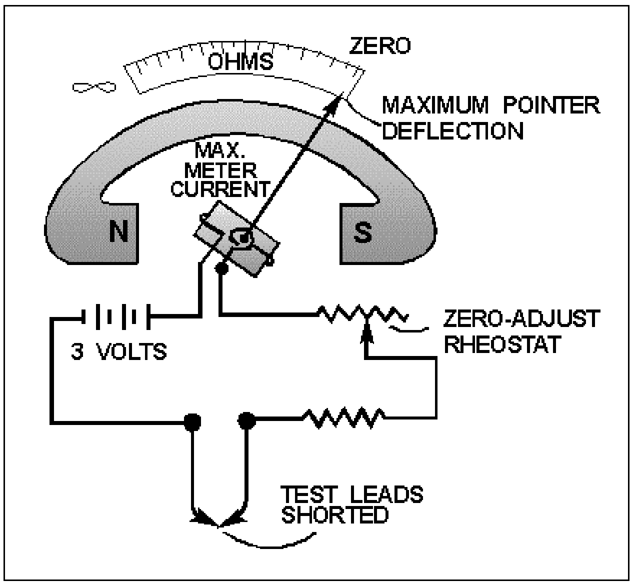


Figure 1-31.—A simple ohmmeter circuit.

USING THE OHMMETER

After the ohmmeter is adjusted for zero reading, it is ready to be connected in a circuit to measure resistance. A typical circuit and ohmmeter arrangement is shown in figure 1-32.

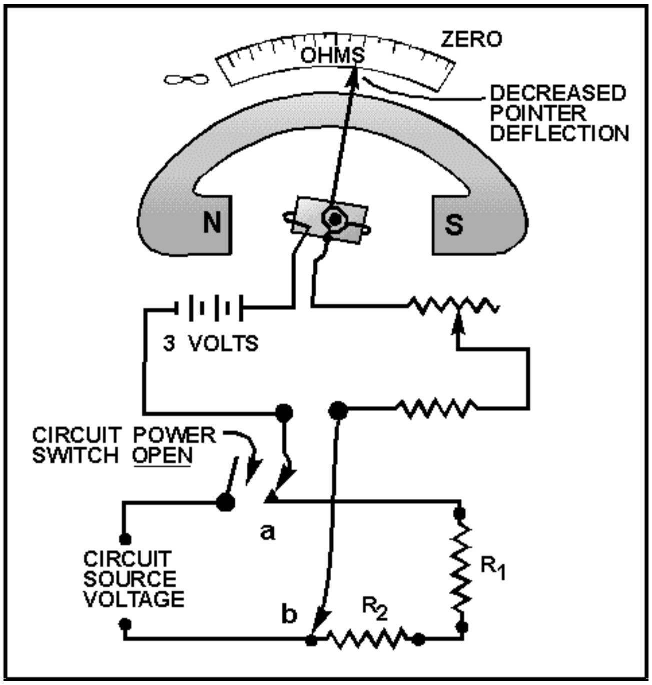


Figure 1-32.—Measuring circuit resistance with an ohmmeter.

The power switch of the circuit to be measured should always be in the **OFF** position. This prevents the source voltage of the circuit from being applied across the meter, which could cause damage to the meter movement.

The test leads of the ohmmeter are connected in series with the circuit to be measured (fig. 1-32). This causes the current produced by the 3-volt battery of the meter to flow through the circuit being tested. Assume that the meter test leads are connected at points a and b of figure 1-32. The amount of current that flows through the meter coil will depend on the total resistance of resistors R_1 and R_2, and the resistance of the meter. Since the meter has been preadjusted (zeroed), the amount of coil movement now depends solely on the resistance of $R_1$and R_2. The inclusion of R_1 and R_2 raises the total series resistance, decreasing the current, and thus decreasing the pointer deflection. The pointer will now come to rest at a scale figure indicating the combined resistance of R_1 and R_2. If R_1 or R_2, or both, were replaced with a resistor(s) having a larger value, the current flow in the moving coil of the meter would be decreased further. The deflection would also be further decreased, and the scale indication would read a still higher circuit resistance. Movement of the moving coil is proportional to the amount of current flow.

OHMMETER RANGES

The amount of circuit resistance to be measured may vary over a wide range. In some cases it may be only a few ohms, and in others it may be as great as 1,000,000 ohms (1 megohm). To enable the meter to indicate any value being measured, with the least error, scale multiplication features are used in most ohmmeters. For example, a typical meter will have four test lead jacks-**COMMON**, $R \times 1$, $R \times 10$, and $R \times 100$. The jack marked **COMMON** is connected internally through the battery to one side of the moving coil of the ohmmeter. The jacks marked $R \times 1$, $R \times 10$, and $R \times 100$ are connected to three different size resistors located within the ohmmeter. This is shown in figure 1-33.

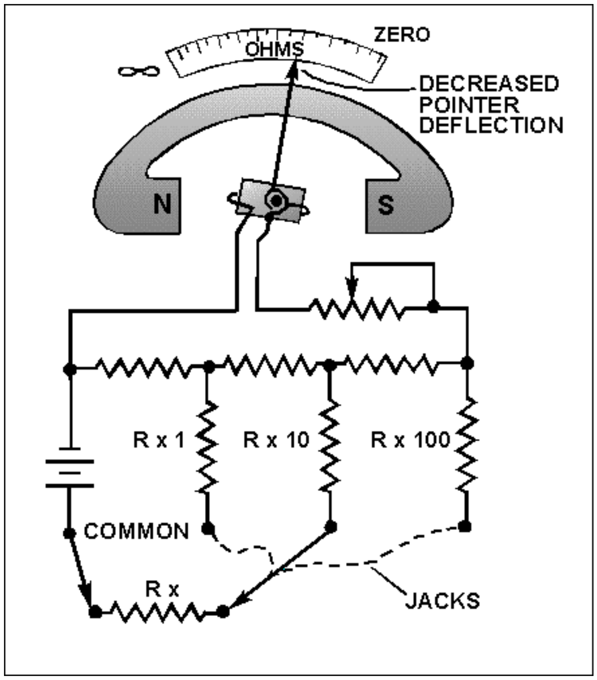


Figure 1-33.—An ohmmeter with multiplication jacks.

Some ohmmeters are equipped with a selector switch for selecting the multiplication scale desired, so only two test lead jacks are necessary. Other meters have a separate jack for each range, as shown in figure 1-33. The range to be used in measuring any particular unknown resistance (R_x in figure 1-33) depends on the approximate value of the unknown resistance. For instance, assume the ohmmeter in figure 1-33 is calibrated in divisions from 0 to 1,000. If R_x is greater than 1,000 ohms, and the $R \times 1$ range is being used, the ohmmeter cannot measure it. This occurs because the combined series resistance of resistor $R \times 1$ and R_x is too great to allow sufficient battery current to flow to deflect the pointer away from infinity (∞). (Infinity is a quantity larger than the largest quantity you can measure.) The test lead would have to be plugged into the next range, $R \times 10$. With this done, assume the pointer deflects to indicate 375 ohms. This would indicate that R_x has 375 ohms $\times$ 10, or 3,750 ohms resistance. The change of range caused the deflection because resistor $R \times 10$ has about 1/10 the resistance of resistor $R \times 1$. Thus, selecting the smaller series resistance permitted a battery current of sufficient amount to cause a useful pointer deflection. If the $R \times 100$ range were used to measure the same 3,750-ohm resistor, the pointer would deflect still further, to the 37.5-ohm position. This increased deflection would occur because resistor $R \times 100$ has about 1/10 the resistance of resistor $R \times 10$.

The foregoing circuit arrangement allows the same amount of current to flow through the meter's moving coil whether the meter measures 10,000 ohms on the $R \times 10$ scale, or 100,000 ohms on the $R \times 100$ scale.

It always takes the same amount of current to deflect the pointer to a certain position on the scale (midscale position for example), regardless of the multiplication factor being used. Since the multiplier resistors are of different values, it is necessary to **ALWAYS** "zero" adjust the meter for each multiplication fact or selected.

You should select the multiplication factor (range) that will result in the pointer coming to rest as near as possible to the midpoint of the scale. This enables you to read the resistance more accurately, because the scale readings are more easily interpreted at or near midpoint.

Q41. What electrical quantity is measured by an ohmmeter?

Q42. What other measurement can an ohmmeter make?

Q43. How is a series-type ohmmeter connected to the circuit being measured?

Q44. What is used to provide the ohmmeter with several ranges?

Q45. What area of an ohmmeter scale should be used when measuring circuits?

SHUNT OHMMETER

The ohmmeter described to this point is known as a series ohmmeter, because the resistance to be measured is in series with the internal resistors and the meter movement of the ohmmeter. Another type of ohmmeter is the **SHUNT OHMMETER**. In the shunt ohmmeter, the resistance to be measured shunts (is in parallel with) the meter movement of the ohmmeter. The most obvious way to tell the difference between the series and shunt ohmmeters is by the scale of the meter. Figure 1-34 shows the scale of a series ohmmeter and the scale of a shunt ohmmeter.

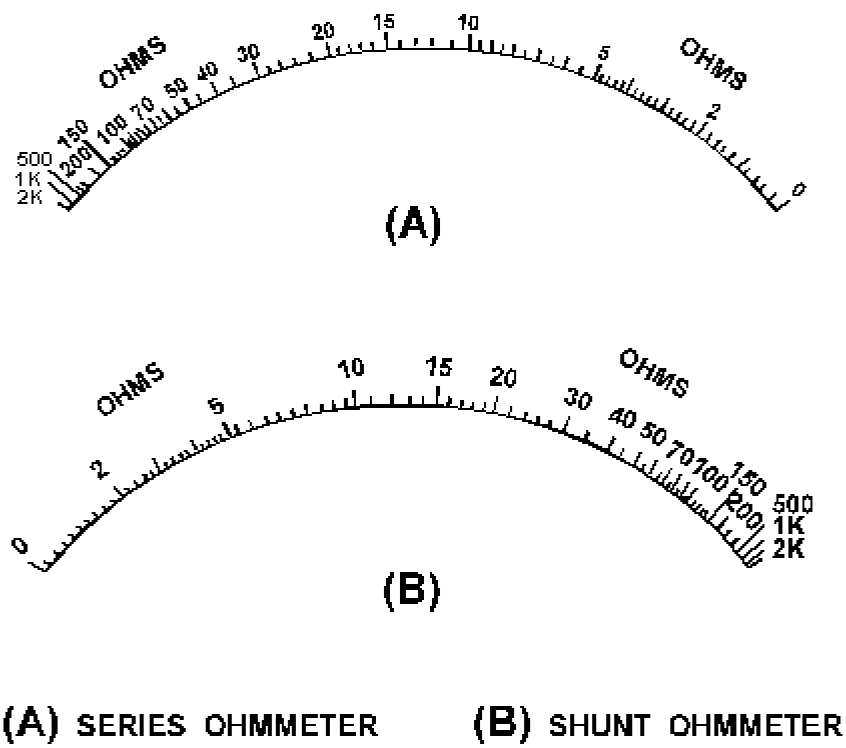


Figure 1-34.—Series and shunt ohmmeter scales.

Figure 1-34(A) is the scale of a series ohmmeter. Notice "0" is on the right and "∞" is on the left. Figure 1-34(B) is the scale of a shunt ohmmeter. In the shunt ohmmeter "∞" is on the right and "0" is on the left. A shunt ohmmeter circuit is shown in figure 1-35.

In figure 1-35, R_1 is a rheostat used to adjust the ∞ reading of the meter (full-scale deflection). R_2, R_3, and R_4 are used to provide the R × 1, R × 10, and R × 100 ranges. Points A and B represent the meter leads. With no resistance connected between points A and B the meter has full-scale current and indicates

∞. If a resistance is connected between points A and B, it shunts some of the current from the meter movement and the meter movement reacts to this lower current. Since the scale of the meter is marked in ohms, the resistance of the shunting resistor (between points A and B) is indicated. Notice that the switch has an **OFF** position, as well as positions for R × 1, R × 10, and R × 100. This is provided to stop current flow and prevents the battery from being discharged while the meter is not being used.

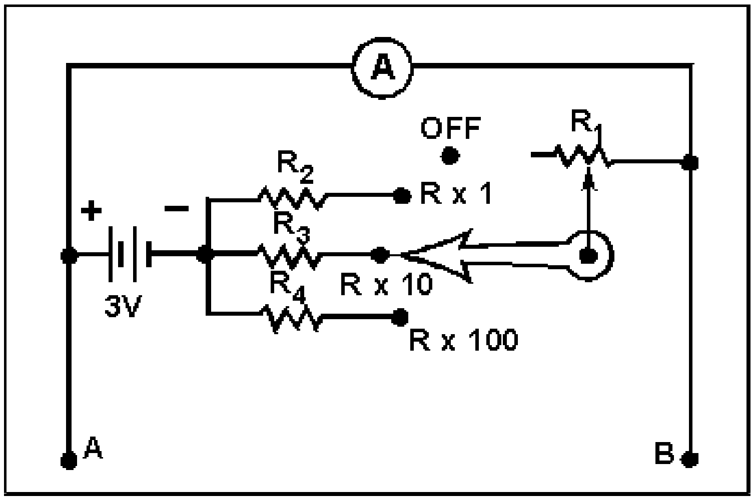


Figure 1-35.—A shunt ohmmeter with internal range resistors.

The shunt ohmmeter is connected to the circuit to be measured in the same way the series ohmmeter is connected. The only difference is that on the shunt ohmmeter the ∞ reading is adjusted, while on the series ohmmeter the 0 reading is adjusted. Shunt ohmmeters are not commonly used because they are limited generally to measuring resistances from 5 ohms to 400 ohms. If you use a shunt ohmmeter, be certain to switch it to the **OFF** position when you are finished using it.

Q46. What are the two types of ohmmeters?

Q47. What is the most obvious difference between the two types of ohmmeters?

Q48. List the four safety precautions observed when using ohmmeters.

OHMMETER SAFETY PRECAUTIONS

The following safety precautions and operating procedures for ohmmeters are the **MINIMUM** necessary to prevent injury and damage.

- Be certain the circuit is deenergized and discharged before connecting an ohmmeter.
- Do not apply power to a circuit while measuring resistance.
- When you are finished using an ohmmeter, switch it to the OFF position if one is provided and remove the leads from the meter.
- Always adjust the ohmmeter for 0 (or ∞ in shunt ohmmeter) after you change ranges before making the resistance measurement.

MEGOHMMETER

An ordinary ohmmeter cannot be used for measuring resistance of multimillions of ohms, such as in conductor insulation. To adequately test for insulation break down, it is necessary to use a much higher potential than is furnished by the battery of an ohmmeter. This potential is placed between the conductor and the outside surface of the insulation.

An instrument called a **MEGOHMMETER (MEGGER)** is used for these tests. The megger (fig. 1-36) is a portable instrument consisting of two primary elements: (1) a hand-driven dc generator, G, which supplies the high voltage for making the measurement, and (2) the instrument portion, which indicates the value of the resistance being measured. The instrument portion is of the opposed-coil type, as shown in figure 1-36(A). Coils a and b are mounted on the movable member c with a fixed relationship to each other, and are free to turn as a unit in a magnetic field. Coil b tends to move the pointer counterclockwise, and coil a tends to move the pointer clockwise.

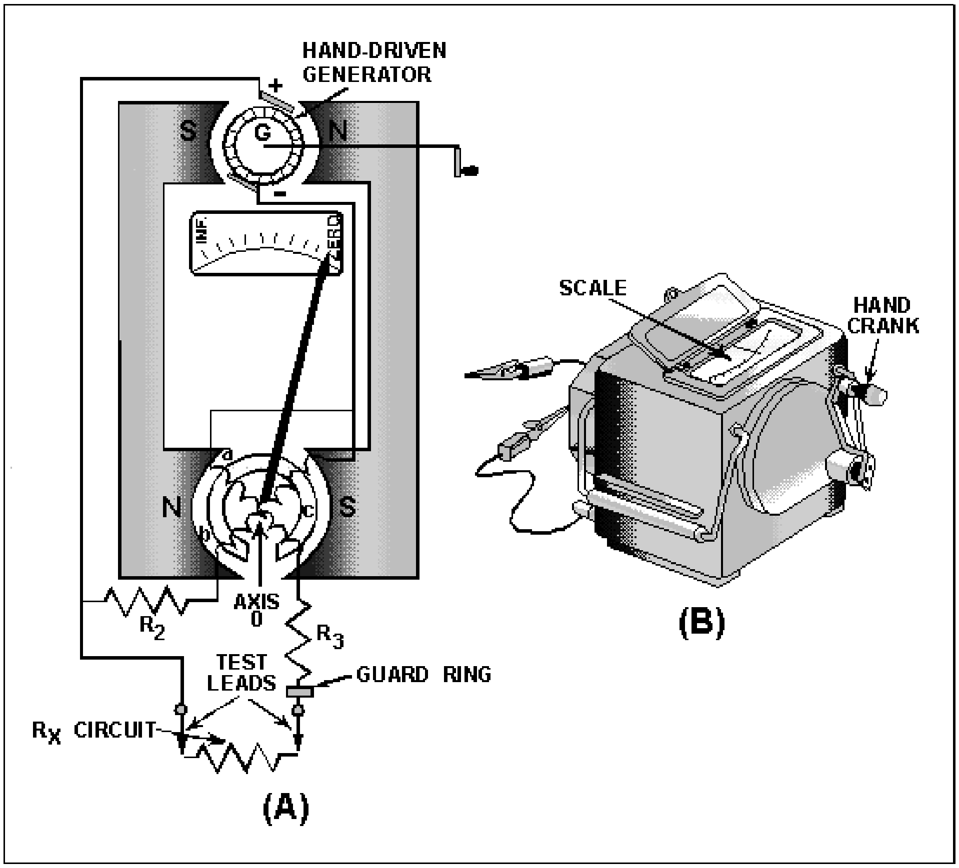


Figure 1-36.—A megger internal circuit.

Coil a is connected in series with R_3 and the unknown resistance, R_x, to be measured. The combination of coil, R_3, and R_x forms a direct series path between the positive (+) and negative (−) brushes of the dc generator. Coil b is connected in series with R_2 and this combination is also connected across the generator. There are no restraining springs on the movable member of the instrument portion of the megger. Therefore, when the generator is not operated, the pointer floats freely and may come to rest at any position on the scale.

The guard ring intercepts leakage current. Any leakage currents intercepted are shunted to the negative side of the generator. They do not flow through coil a; therefore, they do not affect the meter reading.

If the test leads are open-circuited, no current flows in coil a. However, current flows internally through coil b, and deflects the pointer to infinity, which indicates a resistance too large to measure. When a resistance such as R_x is connected between the test leads, current also flows in coil a, tending to move the pointer clockwise. At the same time, coil b still tends to move the pointer counterclockwise. Therefore, the moving element, composed of both coils and the pointer, comes to rest in a position at which the two forces are balanced. This position depends upon the value of the external resistance, which controls the relative amount of current in coil a. Because changes in voltage affect both coil a and coil b in the same proportion, the position of the moving system is independent of the voltage. If the test leads are short-circuited, the pointer rests at zero because the current in coil a is relatively large. The instrument is not damaged under these circumstances because the current is limited by R_3.

The external view of one type of megger is shown in figure 1-36(B).

Navy meggers are usually rated at 500 volts. To avoid excessive test voltages, most meggers are equipped with friction clutches. When the generator is cranked faster than its rated speed, the clutch slips and the generator speed and output voltage are not permitted to exceed their rated values. When extremely high resistances-for example, 10,000 megohms or more-are to be measured, a high voltage is needed to cause sufficient current flow to actuate the meter movement. For extended ranges, a 1,000-volt generator is available.

When a megger is used, the generator voltage is present on the test leads. This voltage could be hazardous to you or to the equipment you are checking. Therefore, **NEVER TOUCH THE TEST LEADS WHILE THE MEGGER IS BEING USED** and isolate the item you are checking from the equipment before using the megger.

Using the Megger

To use a megger to check wiring insulation, connect one test lead to the insulation and the other test lead to the conductor, after isolating the wiring from the equipment. Turn the hand crank until the slip clutch just begins to slip and note the meter reading. Normal insulations should read infinity. Any small resistance reading indicates the insulation is breaking down.

Megger Safety Precautions

When you use a megger, you could be injured or damage equipment you are working on if the following **MINIMUM** safety precautions are not observed.

- Use meggers on high-resistance measurements only (such as insulation measurements or to check two separate conductors on a cable).

- Never touch the test leads while the handle is being cranked.

- Deenergize and discharge the circuit completely before connecting a megger.

- Disconnect the item being checked from other circuitry, if possible, before using a megger.

Q49. What is the primary use of a megger?

Q50. What is the procedure for using a megger to check the insulation of a conductor?

Q51. What is a normal indication on a megger when checking insulation?

Q52. List the four safety precautions observed when using a megger.

MULTIMETER

A **MULTIMETER** is the most common measuring device used in the Navy. The name multimeter comes from **MULTI**ple **METER**, and that is exactly what a multimeter is. It is a dc ammeter, a dc voltmeter, an ac voltmeter, and an ohmmeter, all in one package. Figure 1-37 is a picture of a typical multimeter.

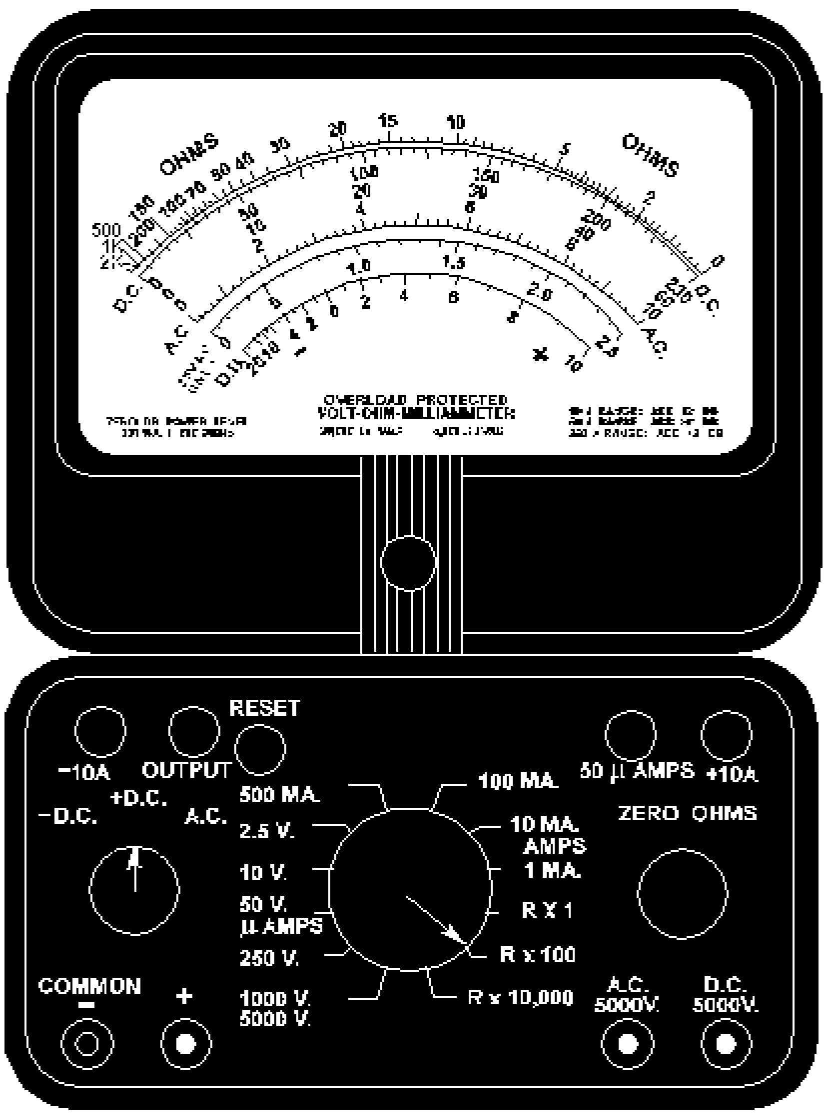


Figure 1-37.—A typical multimeter.

The multimeter shown in figure 1-37 may look complicated, but it is very easy to use. You have already learned about ammeters, voltmeters, and ohmmeters; the multimeter is simply a combination of these meters.

Most multimeters use a d'Arsonval meter movement and have a built-in rectifier for ac measurement. The lower portion of the meter shown in figure 1-37 contains the function switches and jacks (for the meter leads).

The use of the jacks will be discussed first. The **COMMON** or -jack is used in all functions is plugged into the **COMMON** jack. The +jack is used for the second meter lead for any of the functions printed in large letters beside the **FUNCTION SWITCH** (the large switch in the center). The other jacks have specific functions printed above or below them and are self-explanatory (the output jack is used with the dB scale, which will not be explained in this chapter). To use one of the special function jacks, except +10 amps, one lead is plugged into the **COMMON** jack, and the **FUNCTION SWITCH** is positioned to point to the special function (small letters). For example, to measure a very small current (20 microamperes), one meter lead would be plugged into the **COMMON** jack, the other meter lead would be plugged into the 50A AMPS jack, and the **FUNCTION SWITCH** would be placed in the 50V/IA AMPS position. To measure currents above 500 milliamperes, the +10A and −10A jacks would be used on the meter with one exception. One meter lead and the FUNCTION SWITCH would be placed in the 10MA/AMPS position.

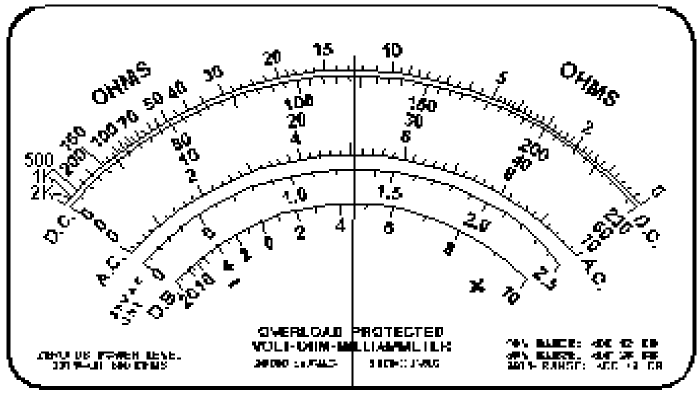
MULTIMETER CONTROLS

As described above, the **FUNCTION SWITCH** is used to select the function desired; the −DC, +DC, AC switch selects dc or ac (the rectifier), and changes the polarity of the dc functions. To measure resistance, this switch should be in the +DC position.

The **ZERO OHMS** control is a potentiometer for adjusting the 0 reading on ohmmeter functions. Notice that this is a series ohmmeter. The **RESET** is a circuit breaker used to protect the meter movement (circuit breakers will be discussed in chapter 2 of this module). Not all multimeters have this protection but most have some sort of protection, such as a fuse. When the multimeter is not in use, it should have the leads disconnected and be switched to the highest voltage scale and AC. These switch positions are the ones most likely to prevent damage if the next person using the meter plugs in the meter leads and connects the meter leads to a circuit without checking the function switch and the dc/ac selector.

MULTIMETER SCALES

The numbers above the uppermost scale in figure 1-38 are used for resistance measurement. If the multimeter was set to the R × 1 function, the meter reading would be approximately 12.7 ohms.



FUNCTION SWITCH	-D.C./+D.C. A.C.	INDICATION
5000 V	+ d.c.	+ 2420.00Vd.c.
1000 V	- d.c.	- 482.00Vd.c.
250 V	+ d.c.	+ 121.00Vd.c.
50 V	a.c.	24.90Va.c.
10 V	a.c.	4.99Va.c.
2.5 V	a.c.	1.28Va.c.
10 A	+ d.c.	4.82Ad.c.
500 mA	a.c.	249.00mAa.c.
100 mA	a.c.	49.90mAa.c.
10mA	+ d.c.	4.82mAd.c.
50μA	+ d.c.	24.20μAd.c.
R x 100	+ d.c.	1.27kΩ

Figure 1-38.—A multimeter scale and reading.

The numbers below the uppermost scale are used with the uppermost scale for dc voltage and direct current, and the same numbers are used with the scale just below the numbers for ac voltage and alternating current. Notice the difference in the dc and ac scales. This is because the ac scale must indicate effective ac voltage and current. The third scale from the top and the numbers just below the scale are used for the 2.5-volt ac function only. The lowest scale (labeled DB) will not be discussed. The manufacturer's technical manual will explain the use of this scale.

The table in figure 1-38 shows how the given needle position should be interpreted with various functions selected.

As you can see, a multimeter is a very versatile measuring device and is much easier to use than several separate meters.

Q53. What is a multimeter?

Q54. Why is a multimeter preferred over separate meters?

Q55. How is a multimeter changed from a voltage measuring device to a resistance measuring device?

Q56. Why is the dc scale on a multimeter different than the ac scale?

Table 1-2 illustrates an interesting point about multimeters. It was mentioned earlier in this chapter that both voltmeters and ammeters have an effect upon the circuits they measure.

Table 1-2.—Multimeter Movements

CURRENT TO DEFLECT FULL SCALE	METER MOVEMENT RESISTANCE	VOLTMETER SENSITIVITY	VOLTAGE FULL SCALE	SHUNT RESISTOR	OVERALL RESISTANCE
1mA	100 Ω	1 k Ω /VOLT	.1 V	NA	100 Ω
50 μA	960 Ω	20 k Ω /VOLT	.048 V	50.5 Ω	48 Ω
5 μA	5750 Ω	200 k Ω/VOLT	.029 V	29.146Ω	28.999 Ω

To keep this effect to a minimum, it is necessary that the voltmeter have a high resistance (sensitivity expressed in ohms per volt) and the ammeter have a low resistance compared to the circuit being measured.

Table 1-2 shows the figures associated with three meter movements available for use in multimeters. The last two columns indicate the value of shunt resistance and the overall resistance of the shunt and meter movement necessary to compensate all three movements to an ammeter sensitivity (full-scale current) of 1 milliampere. Notice that as the voltmeter sensitivity increases, the resistance of the ammeter decreases. This shows how a meter movement used in a voltmeter will have a high effective resistance and the same meter movement used in an ammeter will have a low effective resistance because of the shunt resistors.

PARALLAX ERROR

Most multimeters (and some other meters) have a mirror built into the scale. Figure 1-39 shows the arrangement of the scale and mirror.

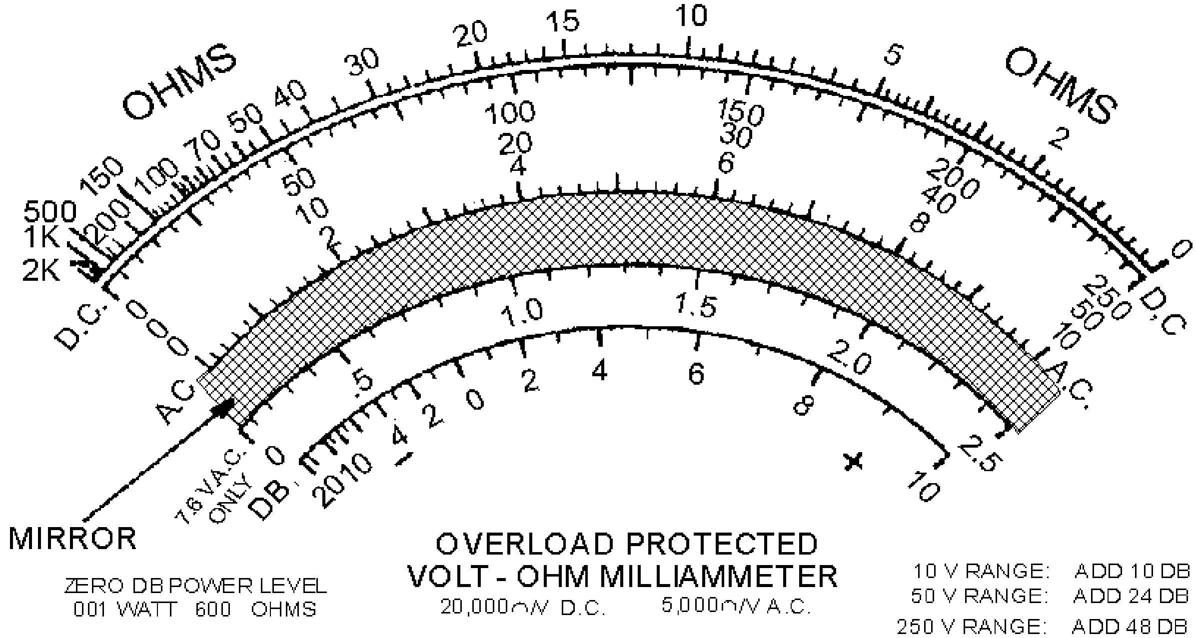


Figure 1-39.—A multimeter scale with mirror.

The purpose of the mirror on the scale of a meter is to aid in reducing **PARALLAX ERROR**. Figure 1-40 will help you understand the idea of parallax.

Figure 1-40(A) shows a section of barbed wire fence as you would see it from one side of the fence. Figure 1-40(B) shows the fence as it would appear if you were to look down the fine of fence posts and were directly in line with the posts. You see only one post because the other posts, being in line, are hidden behind the post you can see. Figure 1-40(C) shows the way the fence would appear if you moved to the right of the line of posts. Now the fence posts appear to the right of the post closest to you. Figure 1-40(D) shows the line of fence posts as you would see them if you moved to the left of the front post. This apparent change in position of the fence posts is called **PARALLAX**.

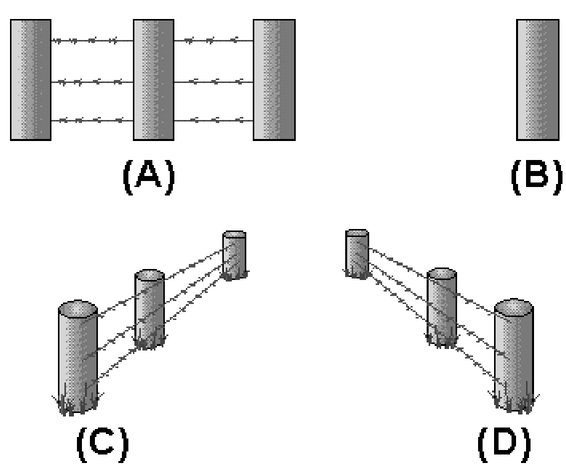


Figure 1-40.—Parallax illustration (barbed-wire fence).

Parallax can be a problem when you are reading a meter. Since the pointer is slightly above the scale (to allow the pointer to move freely), you must look straight at the pointer to have a correct meter reading. In other words, you must be in line with the pointer and the scale. Figure 1-41 shows the effect of parallax error.

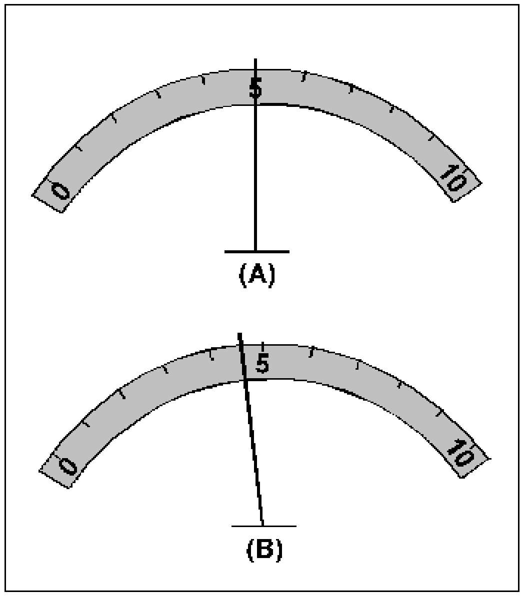


Figure 1-41.—A parallax error in a meter reading.

Figure 1-41 (A) shows a meter viewed correctly. The meter reading is 5 units. Figure 1-41(B) shows the same meter as it would appear if you were to look at it from the right. The correct reading (5) appears to the right of the pointer because of parallax.

The mirror on the scale of a meter, shown in figure 1-39, helps get rid of parallax error. If there is any parallax, you will be able to see the image of the pointer in the mirror. If you are looking at the meter correctly (no parallax error) you will not be able to see the image of the pointer in the mirror because the image will be directly behind the pointer. Figure 1-42 shows how a mirror added to the meter in figure 1-41 shows parallax error. Figure 1-42(A) is a meter with an indication of 5 units. There is no parallax error in this reading and no image of the pointer is seen in the mirror. Figure 1-42(B) shows the same meter as viewed from the right. The parallax error is shown and the image of the pointer is shown in the mirror.

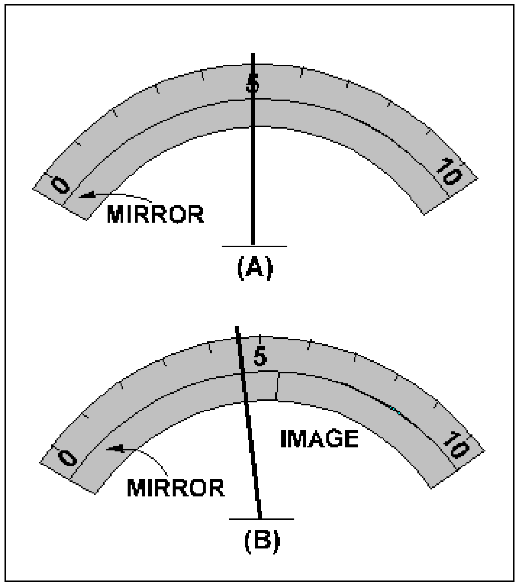


Figure 1-42.—A parallax error on a meter with a mirrored scale.

MULTIMETER SAFETY PRECAUTIONS

As with other meters, the incorrect use of a multimeter could cause injury or damage. The following safety precautions are the **MINIMUM** for using a multimeter.

- Deenergize and discharge the circuit completely before connecting or disconnecting a multimeter.
- Never apply power to the circuit while measuring resistance with a multimeter.
- Connect the multimeter in series with the circuit for current measurements, and in parallel for voltage measurements.
- Be certain the multimeter is switched to ac before attempting to measure ac circuits.
- Observe proper dc polarity when measuring dc.
- When you are finished with a multimeter, switch it to the OFF position, if available. If there is no OFF position, switch the multimeter to the highest ac voltage position.
- Always start with the highest voltage or current range.
- Select a final range that allows a reading near the middle of the scale.
- Adjust the "0 ohms" reading after changing resistance ranges and before making a resistance measurement.
- Be certain to read ac measurements on the ac scale of a multimeter.
- Observe the general safety precautions for electrical and electronic devices.

Q57. What is the reason for having a mirror on the scale of a multimeter?

Q58. How is the mirror on a multimeter used?

Q59. List the 11 safety precautions for multimeters.

OTHER METERS

In addition to the ammeter, voltmeter, ohmmeter, and multimeter, you will probably use many other types of measuring devices. Many of those measuring devices (test equipment) are discussed later in this training series. The following brief discussion of a few additional meters will introduce you to some of common measuring devices you will use in working on electrical and electronic circuits.

HOOK-ON TYPE VOLTAMMETER

The hook-on ac ammeter consists essentially of a current transformer with a split core and a rectifier-type instrument connected to the secondary. The primary of the current transformer is the conductor through which the current to be measured flows. The split core permits the instrument to be "hooked on" the conductor without disconnecting it. Therefore, the current flowing through the conductor may be measured safely and easily, as shown in figure 1-43.

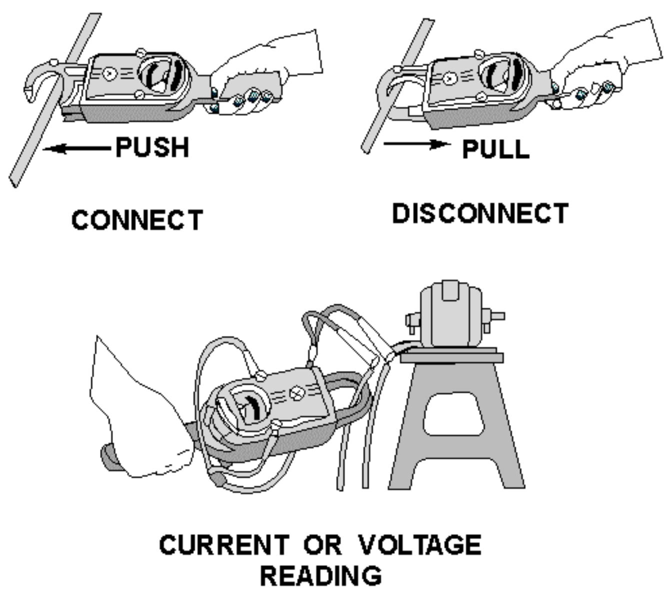


Figure 1-43.—A hook-on type voltammeter.

The instrument is usually constructed so that voltages also may be measured. However, in order to read voltage, the meter switch must be set to **VOLTS**, and leads must be connected from the voltage terminals on the meter to terminals across which the voltage is to be measured.

WATTMETER

Electric power is measured by means of a wattmeter. This instrument is of the electrodynamic type. It consists of a pair of fixed coils, known as current coils, and a movable coil known as the potential coil. (See fig. 1-44.) The fixed coils are made up of a few turns of a comparatively large conductor. The potential coil consists of many turns of fine wire. It is mounted on a shaft, carried in jeweled bearings, so that it may turn inside the stationary coils. The movable coil carries a needle which moves over a suitably marked scale. Spiral coil springs hold the needle to a zero position.

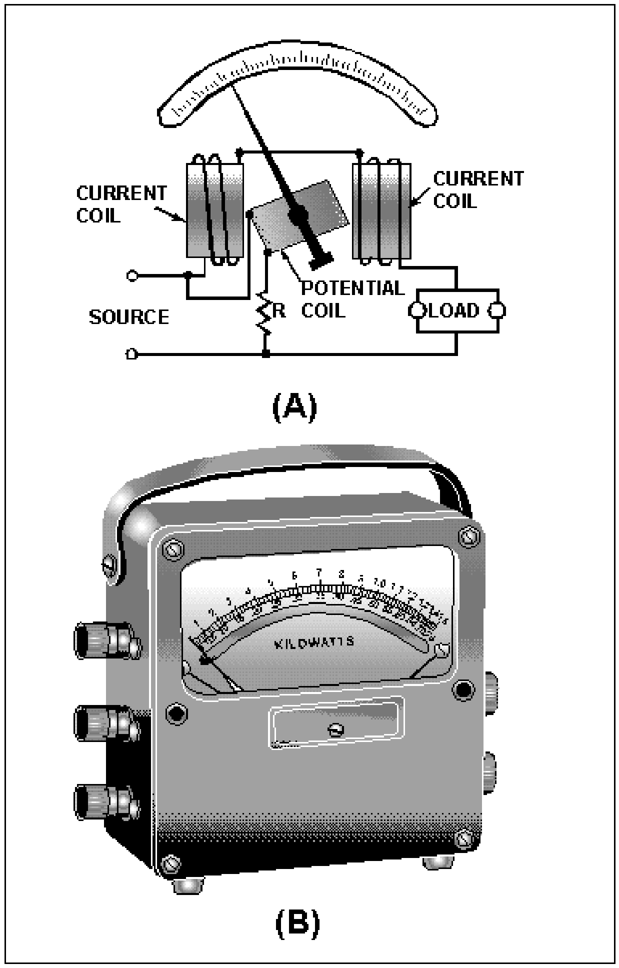


Figure 1-44.—A simplified electrodynamic wattmeter circuit.

The current coil (stationary coil) of the wattmeter is connected in series with the circuit (load), and the potential coil (movable coil) is connected across the line. When line current flows through the current coil of a wattmeter, a field is set up around the coil. The strength of this field is proportional to the line current and in phase with it. The potential coil of the wattmeter generally has a high-resistance resistor connected in series with it. This is for the purpose of making the potential-coil circuit of the meter as purely resistive as possible. As a result, current in the potential circuit is practically in phase with line voltage. Therefore, when voltage is applied to the potential circuit, current is proportional to and in phase with the line voltage.

The actuating force of a wattmeter comes from the field of its current coil and the field of its potential coil. The force acting on the movable coil at any instant (tending to turn it) is proportional to the instantaneous values of line current and voltage.

The wattmeter consists of two circuits, either of which will be damaged if too much current is passed through them. This fact is to be especially emphasized in the case of wattmeters, because the reading of the instrument does not serve to tell the user that the coils are being overheated. If an ammeter or voltmeter is overloaded, the pointer will be indicating beyond the upper limit of its scale. In the wattmeter, both the current and potential circuits may be carrying such an overload that their insulation is burning, and yet the pointer may be only part way up the scale. This is because the position of the pointer depends upon the power factor of the circuit as well as upon the voltage and current. Thus, a low power-factor circuit will give a very low reading on the wattmeter even when the current and potential circuits are loaded to the maximum safe limit. This safe rating is generally given on the face of the instrument. A wattmeter is always distinctly rated, not in watts but in volts and amperes. Figure 1-45 shows the proper way to connect a wattmeter in various circuits.

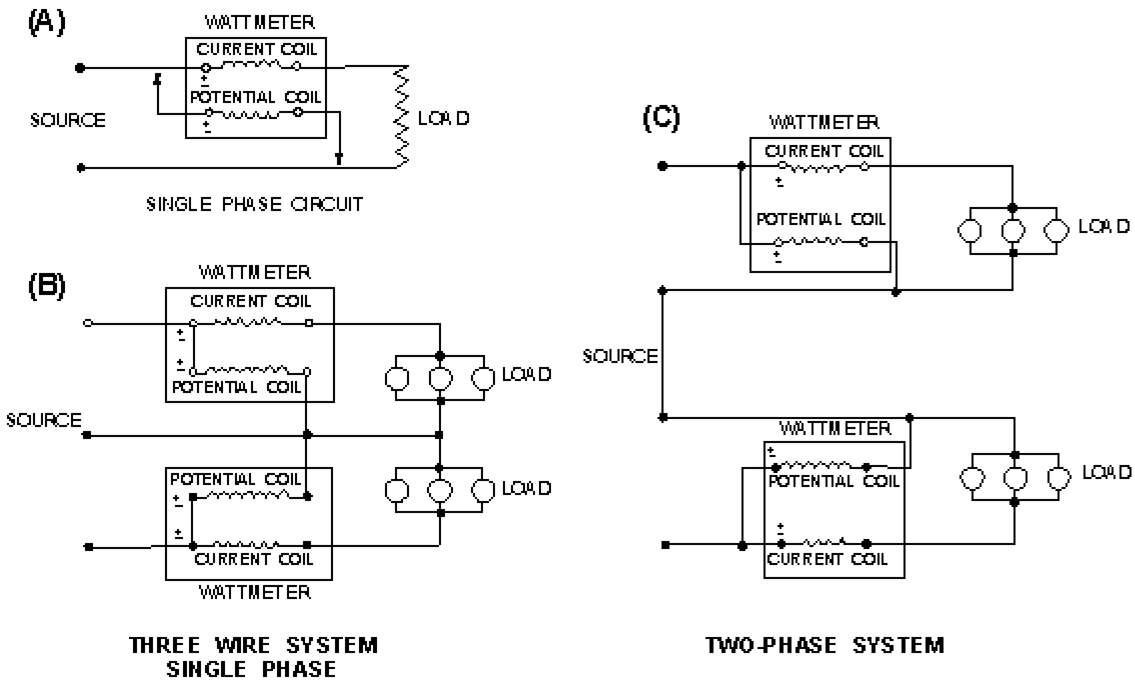


Figure 1-45.—A wattmeter connected in various circuits. TWO-PHASE SYSTEM

WATT-HOUR METER

The watt-hour meter is an instrument for measuring energy. Since energy is the product of power and time, the watt-hour meter must take into consideration both of these factors.

In principle, the watt-hour meter is a small motor whose instantaneous speed is proportional to the **POWER** passing through it. The total revolutions in a given time are proportional to the total **ENERGY**, or watt-hours, consumed during that time.

The following directions should be followed when reading the dials of a watt-hour meter. The meter, in this case, is a four-dial type.

The pointer on the right-hand dial (fig. 1-46) registers 1 kilowatt-hour, or 1,000 watt-hours, for each division of the dial. A complete revolution of the hand on this dial will move the hand of the second dial one division and register 10 kilowatt-hours, or 10,000 watt-hours. A complete revolution of the hand of the second dial will move the third hand one division and register 100 kilowatt-hours or 100,000 watt-hours, and so on.

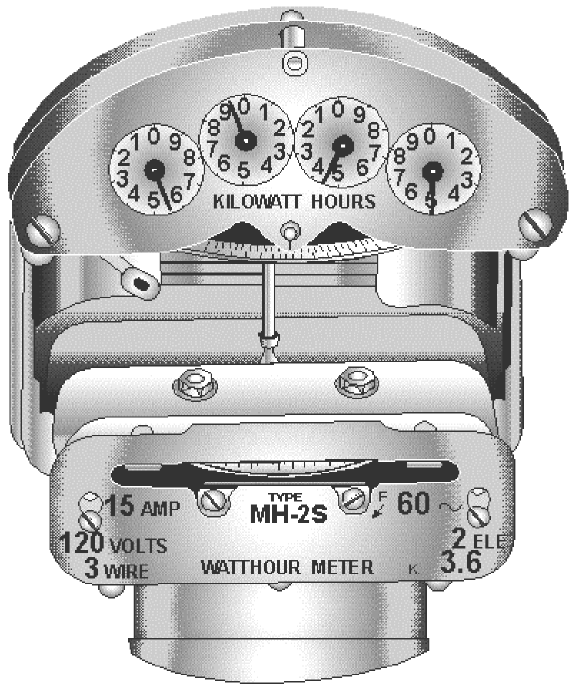


Figure 1-46.—Watt-hour meter.

Accordingly, you must read the hands from left to right, and add three zeros to the reading of the lowest dial to obtain the reading of the meter in watt-hours. The dial hands should always be read as indicating the figure which they have **LAST PASSED**, and not the one they are approaching.

Q60. Why would you use a hook-on voltameter instead of a multimeter?

Q61. What electrical quantity is measured by a wattmeter?

Q62. What electrical quantity is measured by a watt-hour meter?

Q63. What is the quantity shown on the watt-hour meter in figure 1-46?

FREQUENCY METERS

All alternating voltage sources are generated at a set frequency or range of frequencies. A frequency meter provides a means of measuring this frequency. Two common types of frequency meters are the vibrating-reed frequency meter and the moving-disk frequency meter.

Vibrating-Reed Frequency Meter

The vibrating-reed frequency meter is one of the simplest devices for indicating the frequency of an ac source. Vibrating-reed frequency meters are usually in-circuit meters. They are used on power panels to monitor the frequency of ac. A simplified diagram of one type of vibrating-reed frequency meter is shown in figure 1-47.

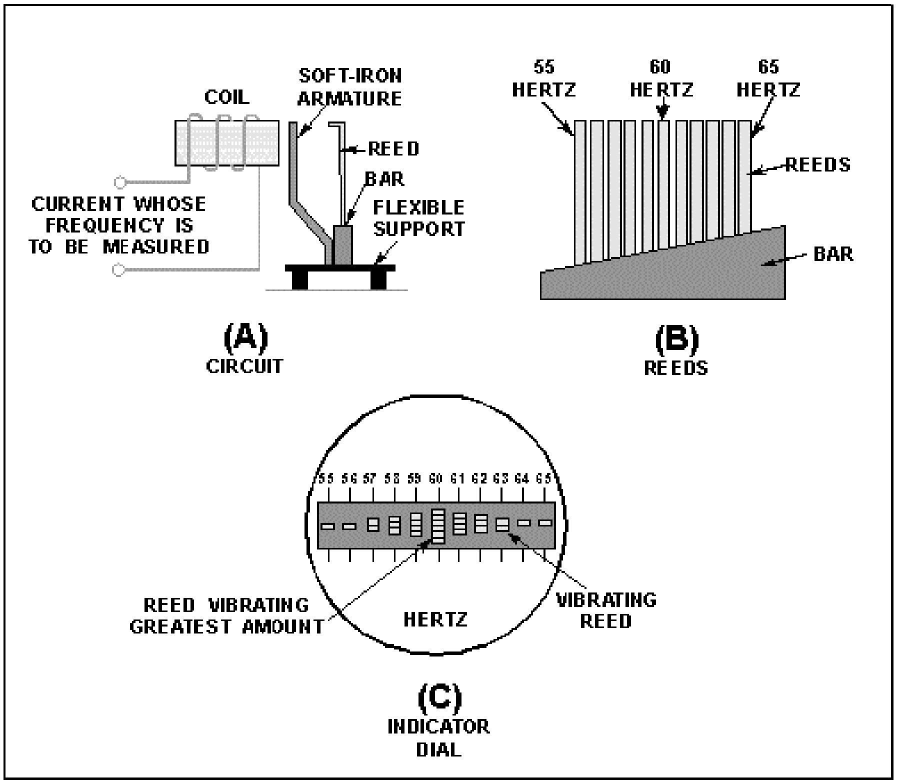


Figure 1-47.—Simplified disc of a vibrating-reed frequency meter. INDICATOR DIAL

The current whose frequency is to be measured flows through the coil and exerts maximum attraction on the soft-iron armature **TWICE** during each cycle (fig. 1-47(A)). The armature is attached to the bar, which is mounted on a flexible support. Reeds having natural vibration frequencies of 110, 112, 114, and so forth, up to 130 hertz are mounted on the bar (fig. 1-47(B)). The reed having a frequency of 110 hertz is marked 55 hertz; the one having a frequency of 112 hertz is marked 56 hertz; the one having a frequency of 120 hertz is marked 60 hertz; and so forth.

When the coil is energized with a current having a frequency between 55 and 65 hertz, all the reeds are vibrated slightly; but, the reed having a natural frequency closest to that of the energizing current (whose frequency is to be measured) vibrates more.

The frequency is read from the scale value opposite the reed having the greatest vibration.

In some instruments the reeds are the same lengths, but are weighted by different amounts at the top so that they will have different natural rates of vibration.

An end view of the reeds is shown in the indicator dial of figure 1-47(C). If the current has a frequency of 60 hertz per second, the reed marked "60" hertz will vibrate the amount, as shown.

Moving-Disk Frequency Meter

Moving-disk frequency meters are most commonly out-of-circuit meters. They can be used to spot check the frequency of power sources or equipment signals.

A moving-disk frequency meter is shown in figure 1-48. One coil tends to turn the disk clockwise, and the other, counterclockwise. Magnetizing coil A is connected in series with a large value of resistance. Coil B is connected in series with a large inductance and the two circuits are supplied in parallel by the source.

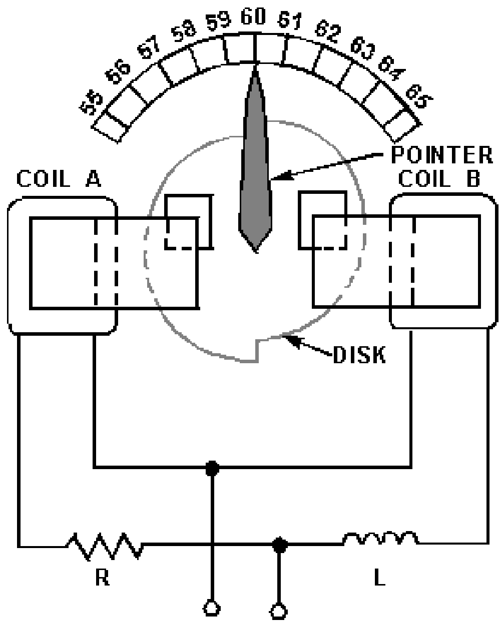


Figure 1-48.—Simplified diagram of a moving-disk frequency meter.

For a given voltage, the current through coil A is practically constant. However, the current through coil B varies with the frequency. At a higher frequency the inductive reactance is greater and the current through coil B is less; the reverse is true at a lower frequency. The disk turns in the direction determined by the stronger coil.

A perfectly circular disk would tend to turn continuously. This is not desirable, and so the disk is constructed so that it will turn only a certain amount clockwise or counterclockwise about the center position, which is commonly marked 60 hertz on commercial equipment. To prevent the disk from turning more than the desired amount, the left half of the disk is mounted so that when motion occurs, the same amount of disk area will always be between the poles of coil A. Therefore, the force produced by coil A to rotate the disk is constant for a constant applied voltage. The right half of the disk is offset, as shown in the figure. When the disk rotates clockwise, an increasing area will come between the poles of coil B; when it rotates counterclockwise, a decreasing area will come between the poles of coil B. The greater the area between the poles, the greater will be the disk current and the force tending to turn the disk.

If the frequency applied to the frequency meter should decrease, the reactance offered by L would decrease and the field produced by coil B would increase. The field produced by coil A would remain the same. Thus, the force produced by coil B would tend to move the disk and the pointer counterclockwise until the area between the poles was reduced enough to make the two forces equal. The scale is calibrated to indicate the correct frequency.

If the frequency is constant and the voltage is changed, the currents in the two coils-and therefore the opposing forces-change by the same amount. Thus, the indication of the instrument is not affected by a change in voltage.

Q64. What are two types of frequency meters?

Q65. What type of meter is shown and what is the value of the quantity being measured for each meter in figure1-49?

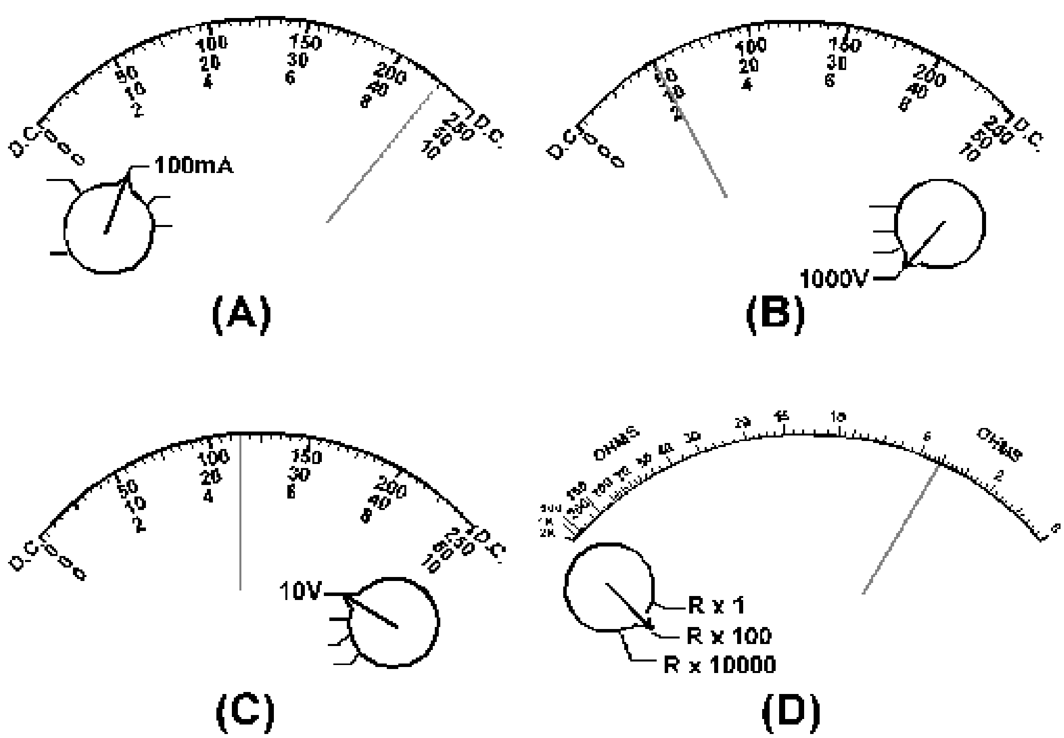


Figure 1-49.—Meter recognition.

Q66. What meter reading is shown on each multimeter in each part of figure 1-50?

Q67. Which part of figure1-50 shows the switch positions the multimeter should be left in when the meter is secured?

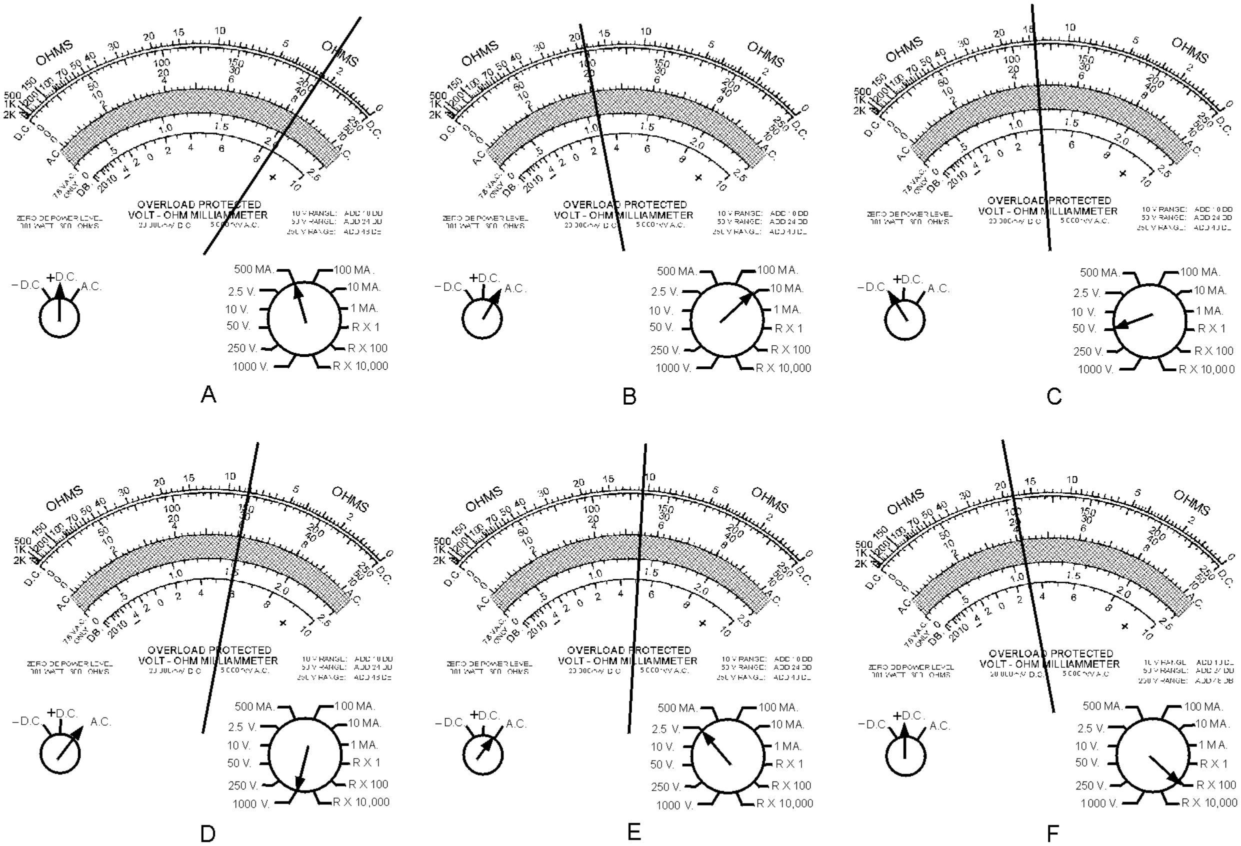


Figure 1-50.—Multimeter reading practice.

Q68. What type of meter is shown and what is the value of the quantity being measured for each meter in figure 1-51?

Q69. If the insulation of a conductor was being measured in figure1-51 (A), would the reading indicate a good insulation?

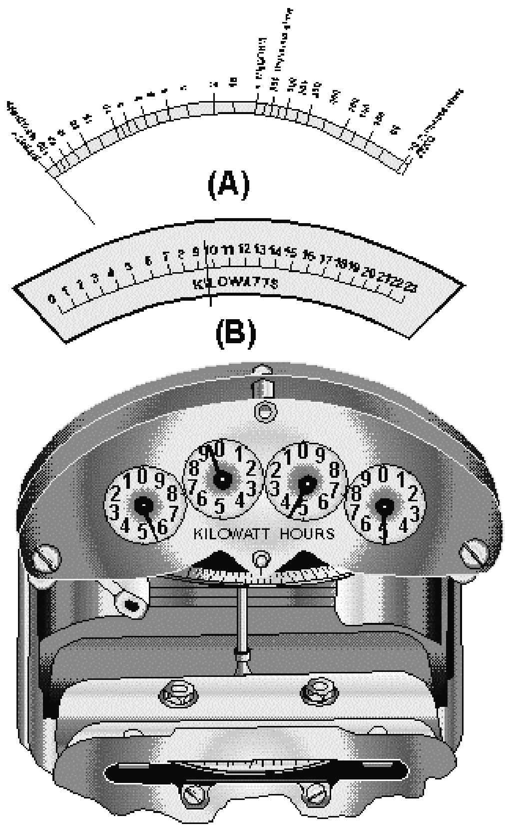


Figure 1-51.—Meter reading practice.

Q70. What type of frequency meter is shown and what is the value indicated for each meter in figure 1-52?

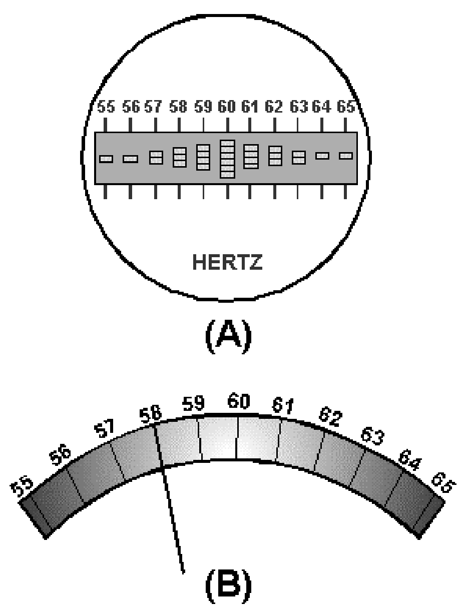


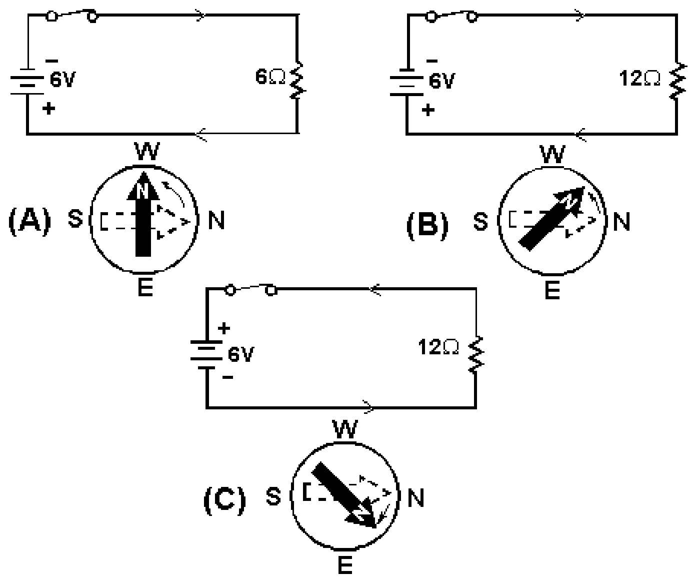
Figure 1-52.—Frequency meter reading.

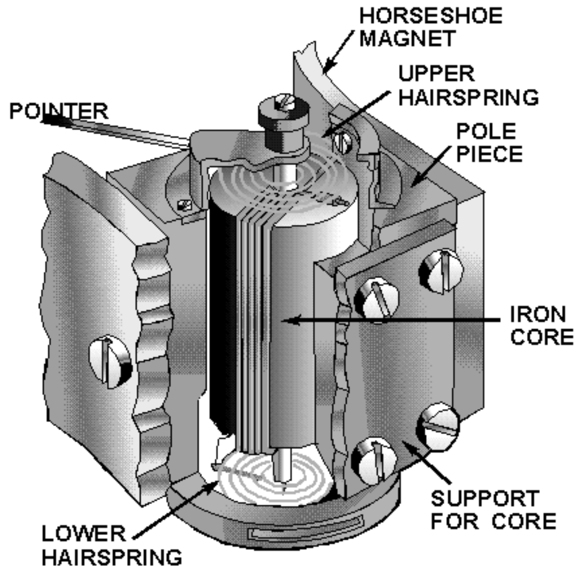
SUMMARY

The important points of this chapter are summarized in the following summary. You should be familiar with these points before continuing with the study of electricity.

CIRCUIT MEASUREMENT is used to monitor the operation of a piece of electrical or electronic equipment and determine the reason the equipment is not functioning properly. In-circuit meters monitor the operation of equipment and out-of-circuit meters can be used on more than one device.

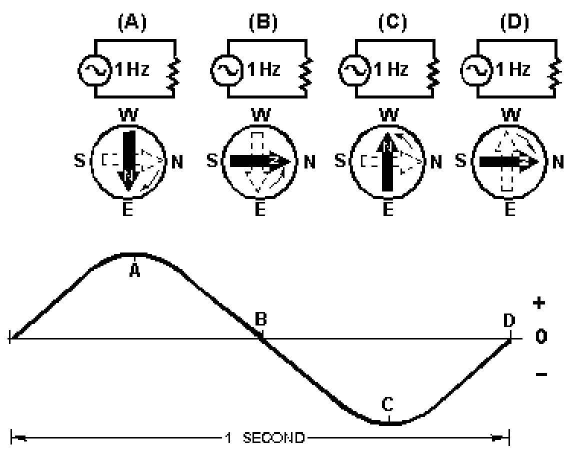
A compass will react to the magnetic field around a conducting wire. As the current increases, the compass movement increases. If the current decreases, the compass movement is less. If the current direction changes, the compass movement changes direction. **PERMANENT-MAGNET MOVING-COIL** meter movement (d’Arsonval movement) uses the interaction of magnetic fields to produce movement.

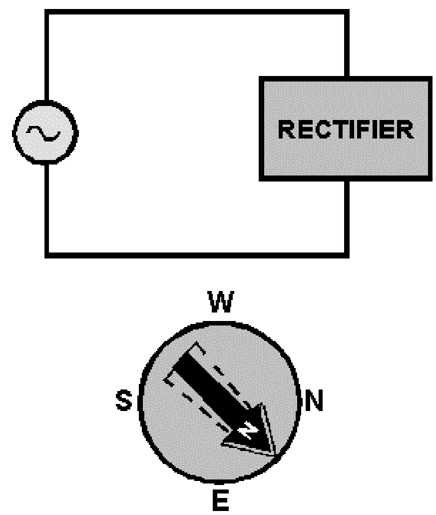




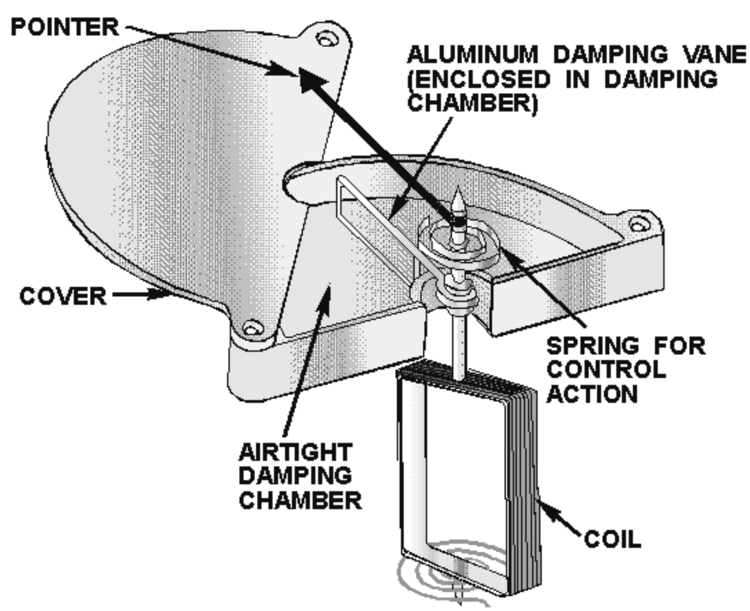
ASSEMBLED ARRANGEMENT

If a compass is placed close to a conductor with ac, the compass will follow the current alternations if the ac is of low frequency. A rectifier will allow the compass to react to the average value of the ac.



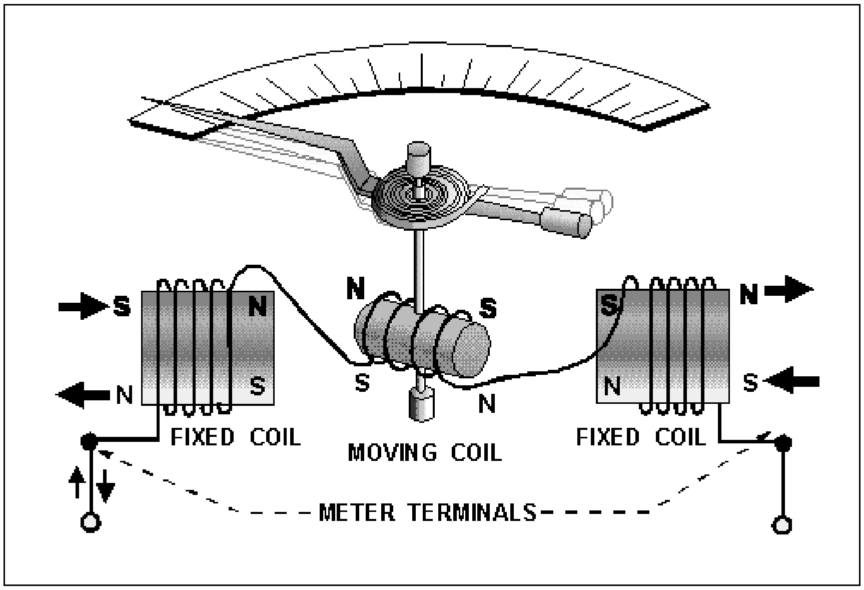


DAMPING is used to smooth out the vibration and to help prevent overshooting of the meter pointer. Damping in a d'Arsonval meter movement is accomplished by the emf caused by the coil movement. A second damping system uses a vane attached to the coil in an airtight chamber. A meter movement reacts to the average value of ac, but the scale is calibrated to read the effective (rms) value.

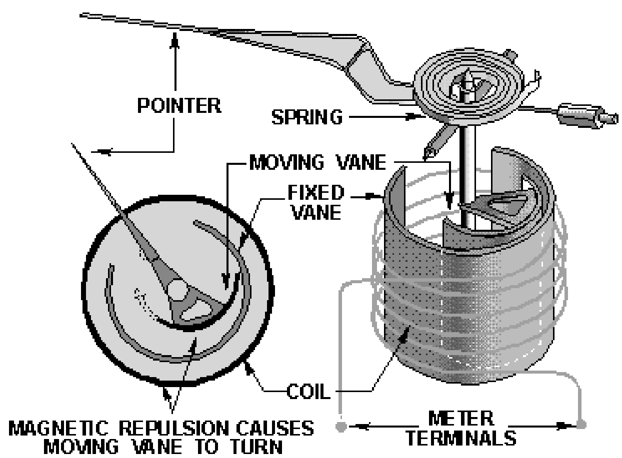


There are meter movements that will measure either ac or dc without the use of a rectifier. They are the **ELECTRODYNAMIC**, **MOVING-VANE**, and **HOT-WIRE** or **THERMOCOUPLE** movements.

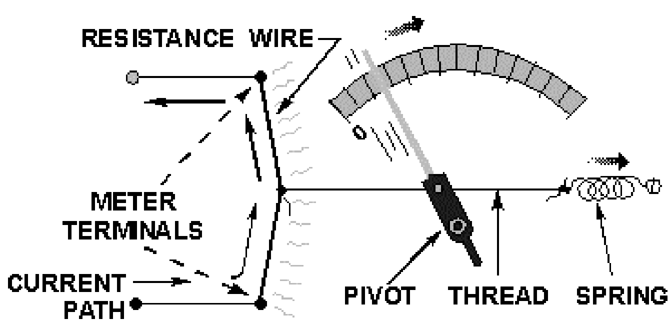
Electrodynamic meter movements are usually used in wattmeters. They operate much like a d'Arsonval meter movement, except field coils are used instead of a permanent magnet.



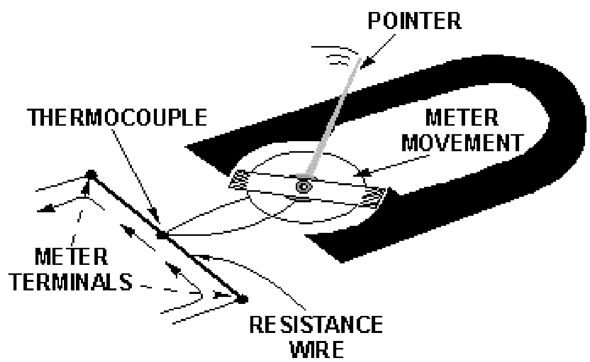
The moving-vane meter movement operates on the principle of magnetic repulsion of like poles. This movement will measure either current or voltage.



The hot-wire movement is only used to measure current. It is based on the expansion of a wire heated by current through the wire.



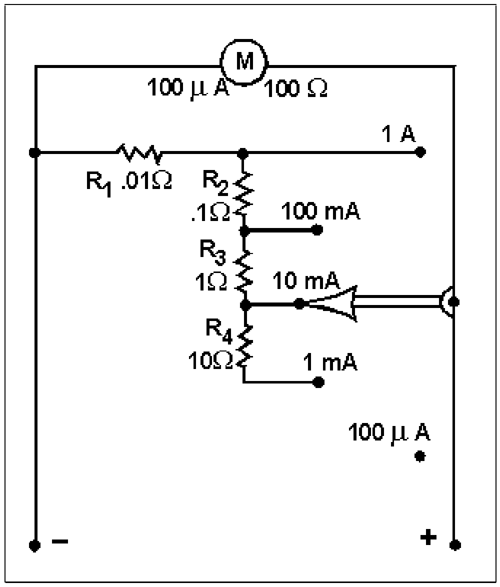
The thermocouple movement uses the current being developed in a thermocouple when the heat of a resistive wire is transferred to the thermocouple. The developed current is measured by a very sensitive dc ammeter. This movement will measure only current.



An **AMMETER** measures current. It is always connected in series with the circuit being measured. An ammeter should have a small resistance so the effect of the ammeter on the circuit will be kept to a minimum. Ammeter sensitivity is the amount of current that causes 0 full scale deflection of the ammeter. Shunt F resistors are used to provide an ammeter’s ranges.

The following **SAFETY PRECAUTIONS** should be observed when using an ammeter.

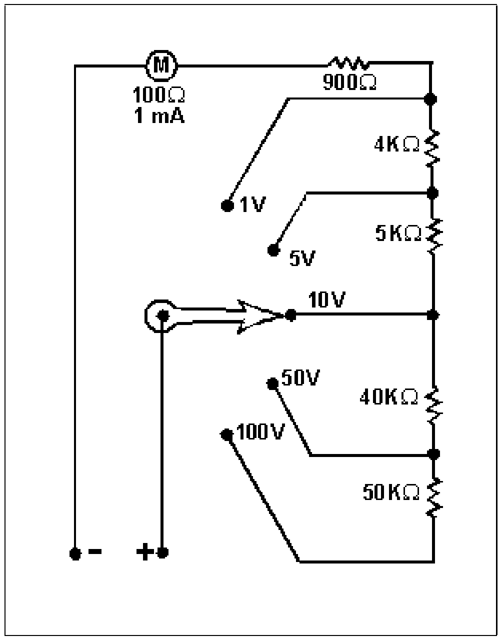
- Always connect an ammeter in series.
- Always start with the highest range.
- Deenergize and discharge the circuit before connecting or disconnecting an ammeter.
- Never use a dc ammeter to measure ac.
- In dc ammeters, observe the proper polarity.



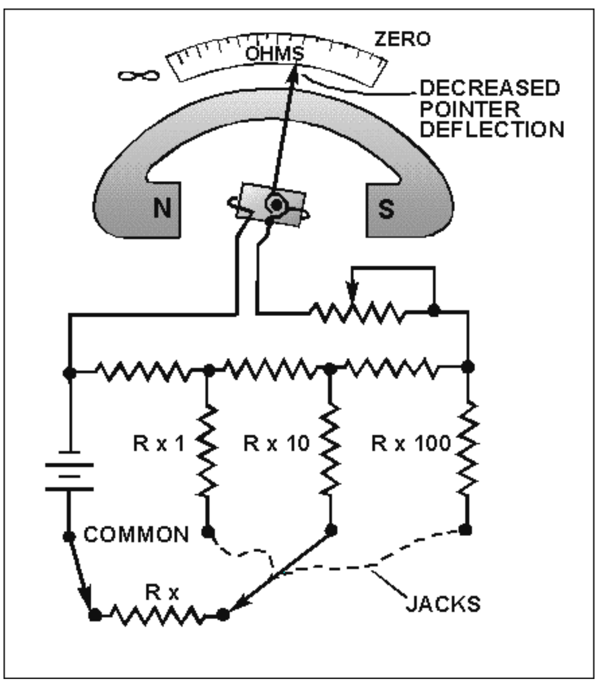
VOLTMETERS are used to measure voltage. They are always connected in parallel with the circuit being measured. A voltmeter should have a high resistance compared to the circuit being measured to minimize the loading effect. Since the resistance of a meter movement is constant, a voltmeter can be made from a current-sensitive meter movement by the use of range resistors and an appropriate scale. Voltmeter sensitivity is expressed in ohms per volt.

An electrostatic meter movement reacts to voltage rather than current and is used only for high-voltage measurements. The following **SAFETY PRECAUTIONS** should be observed when using a voltmeter.

- Always connect a voltmeter in parallel.
- Always start with the highest range.
- Deenergize and discharge the circuit before connecting or disconnecting the voltmeter.
- Never use a dc voltmeter to measure an ac voltage.
- On a dc voltmeter, observe the proper polarity.



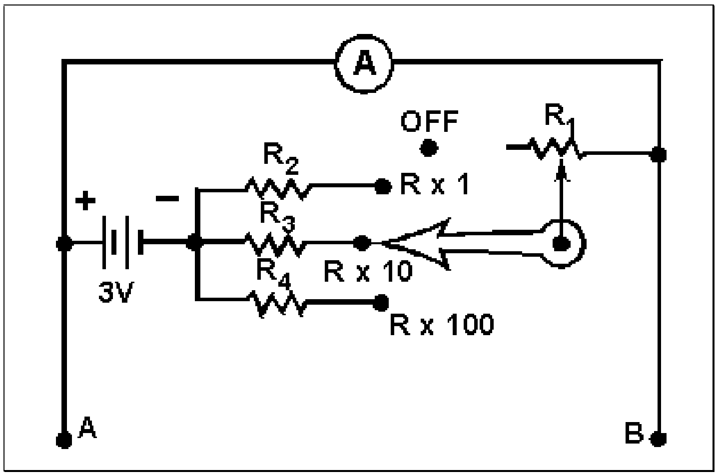
OHMMETERS are used to measure resistance and to check continuity. An ohmmeter is connected in series with the resistance being measured. The ohmmeter range which allows a midscale indication should be selected. Resistors are used to allow an ohmmeter to have several ranges. In a **SERIES OHMMETER** the resistors are used in series with the resistance being measured. Series ohmmeters have the 0 indication on the right side of the scale.



A **SHUNT OHMMETER's** internal range resistors are in parallel with the resistance being measured. A shunt ohmmeter will have the 0 indication on the left side of the scale.

The following **SAFETY PRECAUTIONS** should be observed when using an ohmmeter.

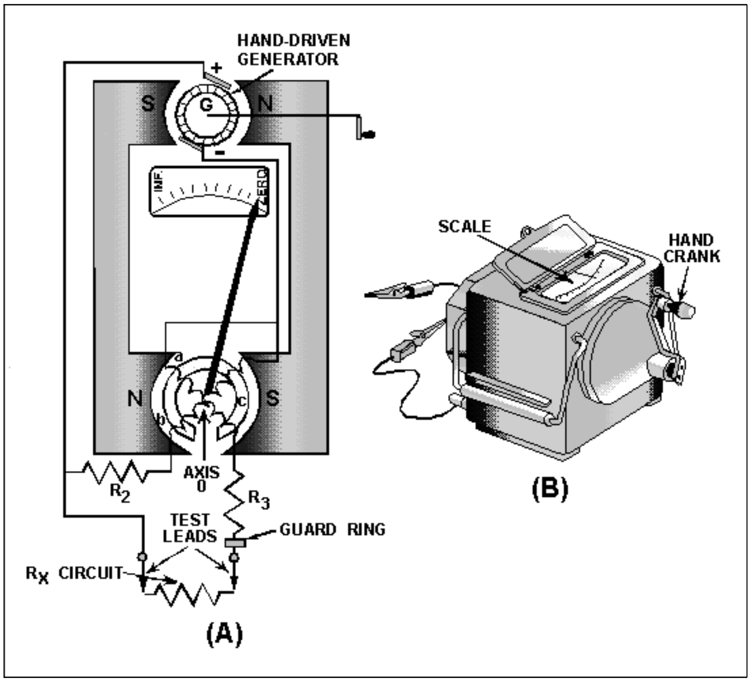
- Deenergize and discharge the circuit before connecting an ohmmeter.
- Do not apply power while measuring resistance.
- Switch ohmmeters OFF, if a setting is provided, or to the highest range and remove the meter leads from the meter when finished measuring resistance.
- Adjust the ohmmeter after changing ranges and before measuring resistance.



A **MEGOHMMETER (MEGGER)** is used to measure very large resistances, such as the insulation of wiring. To use a megger, isolate the resistance being measured from other circuits, connect the meter leads, turn the hand crank, and note the meter indication. Normal insulation will indicate infinity.

The following **SAFETY PRECAUTIONS** should be observed when using a megger. Use meggers for high-resistance measurements only.

- Never touch the test leads while the handle is being cranked.
- Deenergize and discharge the circuit completely before connecting a megger.
- Disconnect the item being checked from other circuitry, if possible, before using megger.

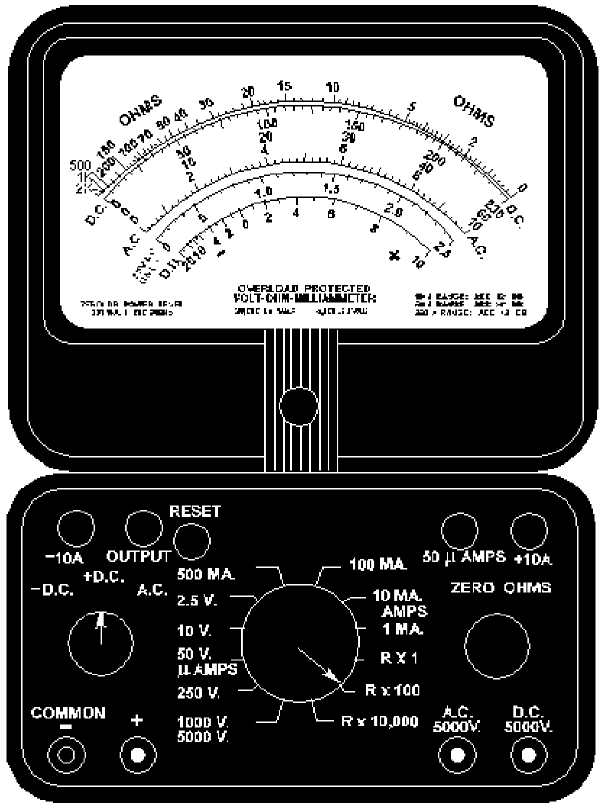


A **MULTIMETER** is a single meter that combines the functions of a dc ammeter, a dc voltmeter, an ac ammeter, an ac voltmeter, and an ohmmeter. It is more convenient to have one meter with several functions than several meters each with a single function. The various functions of a multimeter are selected by use of the appropriate function switch positions, jacks, and meter scales. A mirror may be used on the scale of a multimeter to eliminate parallax error.

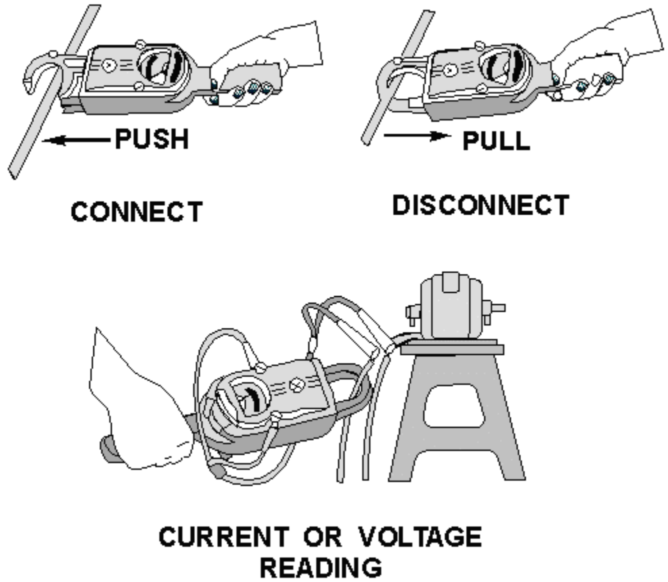
The following **SAFETY PRECAUTIONS** should be observed when using a multimeter.

- Deenergize and discharge the circuit completely before connecting a multimeter.
- Never apply power to the circuit while measuring resistance with a multimeter.
- Connect the multimeter in series with the circuit for current measurements and in parallel for voltage measurements.
- Be certain the multimeter is switched to ac before attempting to measure ac circuits.
- Observe proper dc polarity when measuring dc circuits.
- When you are finished with a multimeter, switch it to the **OFF** position, if available. If there is no **OFF** position, switch the multimeter to the highest ac voltage position.
- Always start with the highest voltage or current range.
- Select a final range that allows a reading near the middle of the scale.
- Adjust the "0 ohms" reading after changing resistance ranges and before making a resistance measurement.

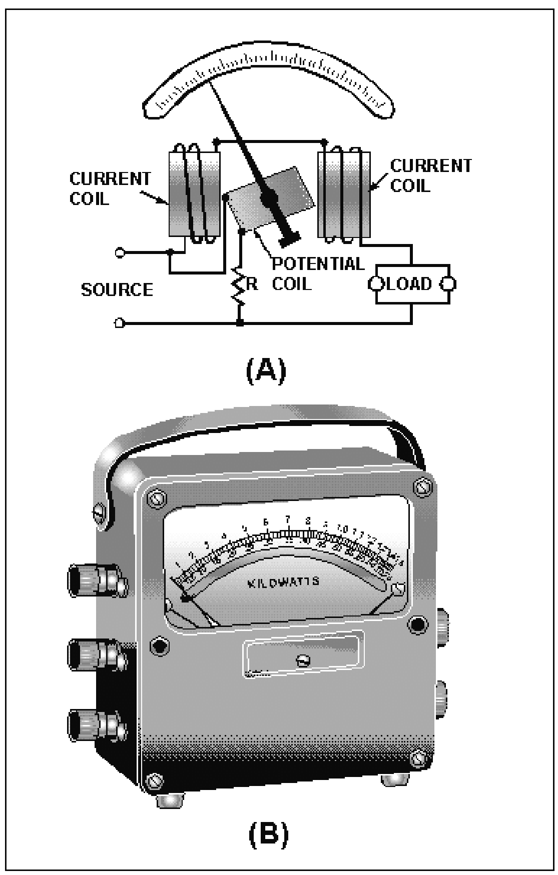
- Be certain to read ac measurements on the ac scale of a multimeter.



A **HOOK-ON TYPE VOLTAMETER** allows you to measure current safely and easily (with no need to disconnect the wiring of the circuit). A hook-on type voltameter uses a split-core transformer to measure current.

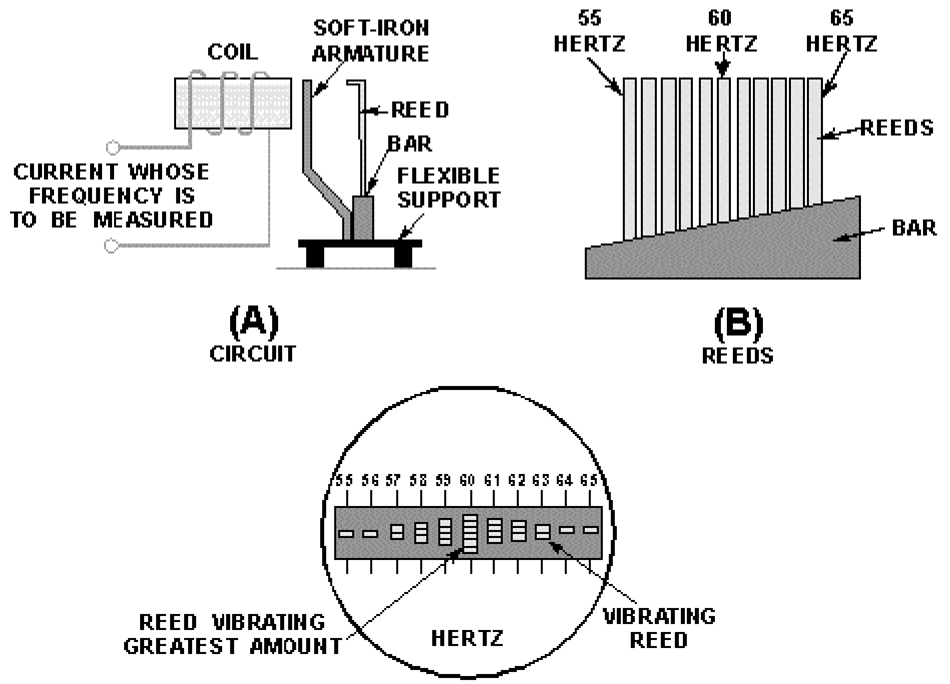


A **WATTMETER** is usually an electrodynamic meter and is used to measure power.



A **WATT-HOUR METER** is basically a small motor whose instantaneous speed is proportional to the power through the motor. The total revolutions in a given time are proportional to the total energy, or watt-hours, used during that time. Watt-hour meters measure energy.

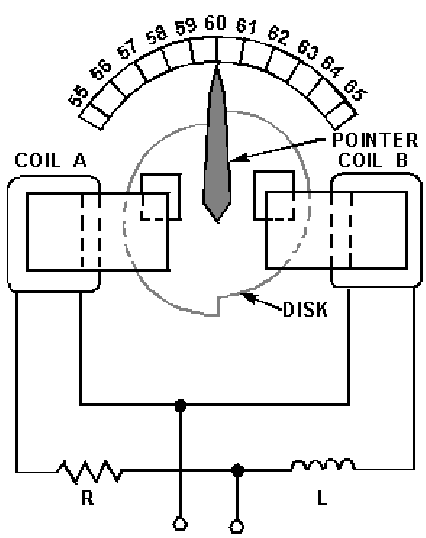
FREQUENCY METERS are used to measure the frequency of an ac signal. The two basic types are the vibrating-reed frequency meter which is usually used as an in-circuit meter, and the moving-disk frequency meter which is usually used as an out-of-circuit meter.



(A) CIRCUIT

(B) REEDS

(C) INDICATOR DIAL



ANSWERS TO QUESTIONS Q1. THROUGH Q70.

A1. Circuit measurement is used to (1) monitor the operation of a piece of electrical or electronic equipment and (2) determine the reason a piece of electrical or electronic equipment is not functioning properly.

A2. In-circuit meters are used to monitor the operation of electrical or electronic devices.

A3. Out-of-circuit meters can be used on more than one electrical or electronic device.

A4. The compass needle swings away from magnetic north and aligns itself with the magnetic field around the conductor.

A5. If the current increases the magnetic field increases; if the current decreases the magnetic field decreases.

A6. The compass needle will not be deflected as far from magnetic north.

A7. A permanent-magnet moving-coil meter movement used in most electrical and electronic meters.

A8. A magnetic field is generated around the coil and the attraction of this field with the permanent magnet causes the coil to move.

A9. To return the pointer to its rest position when there is no current flow; to oppose the coil movement when there is current flow; to provide electrical connections for the coil.

A10. The compass needle would swing back and forth as the current changed from positive to negative.

A11. The compass needle would vibrate rapidly around the zero-current point ac meter (magnetic north).

A12. A rectifier changes alternating current to pulsating direct current and allows a dc meter to measure ac.

A13. By the use of a rectifier

A14. The process of "smoothing out" the oscillation in a meter movement.

A15. As the coil moves through the field of the permanent magnet, a current is induced in the coil opposing the movement of the coil; and a vane can be attached to the coil and placed in the airtight chamber so that the movement of the vane opposes the movement of the coil.

A16. Average value.

A17. Effective value (rms).

A18. Electrodynamic, moving vane, and hot-wire or thermocouple.

A19. Current.

A20. Current.

A21. In series.

A22. Since the ammeter is a resistor in series with the load, it increases the resistance of the circuit and lowers circuit current.

A23. The resistance of the ammeter must be much smaller than the circuit load.

A24. The amount of current that will cause full-scale deflection.

A25. Shunt resistors (internal or external).

A26. To prevent damage to the meter movement from excessive current.

A27. A range that allows a meter reading near the center of the scale.

A28.

- *a. Always connect an ammeter in series.*
- *b. Always start with the highest range.*
- *c. In dc ammeters, observe the proper polarity.*
- *d. Deenergize and discharge the circuit before connecting or disconnecting the ammeter.*
- *e. Never use a dc ammeter to measure ac current.*
- *f. Observe the general safety precautions of electric and electronic devices.*

A29. Since the ammeter has a small resistance compared to the load, it will have very high current if it is connected in parallel. This high current will damage the meter.

A30. Voltage.

A31. In parallel.

A32. The connection of a voltmeter adds a resistance in parallel with the circuit changing the total circuit resistance, and loads the circuit.

A33. A voltmeter must have a high resistance compared to the circuit being measured.

A34. Since the resistance of a meter movement remains the same as the pointer is deflected, the amount of current through the movement is proportional to the voltage applied. Therefore, only the scale of the movement must be changed.

A35. It is an indication of the resistance of the meter expressed in ohms per volt. The total resistance of the meter is the sensitivity multiplied by the full-scale voltage.

A36. The use of resistors in series with the meter movement.

A37. To prevent excess current through the meter movement.

A38. Electrostatic.

A39. High-voltage measurement.

A40.

a. Always connect a voltmeter in parallel.

b. Always start with the highest range.

c. Deenergize and discharge the circuit before connecting or disconnecting the voltmeter.

d. In a dc voltmeter, observe the proper polarity.

e. Never use a dc voltmeter to measure ac voltage.

f. Observe the general safety precautions of electric and electronic devices.

A41. Resistance.

A42. Circuit continuity.

A43. The ohmmeter is connected in series with the resistance to be measured.

A44. An ohmmeter has several internal range resistors and a switch or a series of jacks to select the proper range.

A45. The middle of the scale.

A46. Series and shunt.

A47. Series ohmmeters have 0 on the right end of the scale and ∞ on the left end of the scale. Shunt ohmmeters are the opposite.

A48.

a. Deenergize and discharge the circuit before connecting an ohmmeter.

b. Do not apply power to a circuit while measuring resistance.

c. Switch ohmmeters to the OFF position, if provided, or to highest range and remove meter leads from the meter when finished measuring resistance.

d. Adjust the ohmmeter after changing resistance range and before measuring reading indicates the resistance.

A49. To measure high resistance.

A50. Connect one lead to the insulation and one lead to the conductor. Turn the handcrank until it starts to slip. Note the reading.

A51. Infinity.

A52.

a. Use meggers for high-resistance measurement only.

b. Never touch the test leads when the handle is being cranked.

c. Deenergize and discharge the circuit completely before connecting a megger.

d. Disconnect the item being checked from other circuitry, if possible, before using a megger.

A53. A single measuring device capable of performing the functions of a dc voltmeter and ammeter, an ac voltmeter and ammeter, and an ohmmeter.

A54. It is much more convenient to have one meter with several functions than several meters each with a single function.

A55. By changing the position of the function switch.

A56. The meter movement reacts to average ac voltage and current and the effective value is desired.

A57. To stop parallax error

A58. Make sure no image of the pointer is visible in the mirror when reading the meter.

A59.

a. Deenergize and discharge the circuit completely before connecting or disconnecting a multimeter.

b. Never apply power to the circuit while measuring resistance with a multimeter.

c. Connect the multimeter in series with the circuit for current measurements, and in parallel for voltage measurements.

d. Be certain the multimeter is switched to ac before attempting to measure ac circuits.

e. Observe proper dc polarity when measuring dc.

f. When you are finished with a multimeter, switch it to the OFF position, if available. If there is no OFF position, switch the multimeter to the highest ac voltage position.

g. Always start with the highest voltage or current range.

h. Select a final range that allows a reading near the middle of the scale.

i. Adjust the "0 ohms" reading after changing resistance ranges and before making a resistance measurement.

j. Be certain to read ac measurements on the ac scale of a multimeter.

k. Observe the general safety precautions for electrical and electronic devices.

A60. To measure current safely and easily (with no need to disconnect the wiring of the circuit).

A61. Power.

A62. Energy.

A63. 5.945 megawatt-hours, or 5,945 kilowatt-hours, or 5,945, 000 watt-hours.

A64. Vibrating reed and moving disk.

A65.

- *a. A dc ammeter, 90 mA dc*
- *b. A dc voltmeter, 200 V dc*
- *c. An ac voltmeter, 4.6 V ac*
- *d. An ohmmeter, 400 ohms*

A66. (A) 410 mA dc; (B) 3.9 mA ac; (C) -22 V dc; (D) 600 V ac; (E) 1.4 V ac; (F) 1.9 kohms (1900 Ω).

A67. Figure1-50(D).

A68. (A) Megger (megohmmeter), infinity; (B) Wattmeter, 9.5 kilowatts (9,500 watts). (C) Watt-hour meter, 2.693 megawatt-hours 2,693 kilowatt-hours) (2,693,000 watt-hours).

A69. Yes.

A70. (A) Vibrating-reed, 60Hz. (B) Moving-disk, 58 Hz.

CHAPTER 2

CIRCUIT PROTECTION DEVICES

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. State the reasons circuit protection is needed and three conditions requiring circuit protection.
2. Define a direct short, an excessive current condition, and an excessive heat condition.
3. State the way in which circuit protection devices are connected in a circuit.
4. Identify two types of circuit protection devices and label the schematic symbols for each type.
5. Identify a plug-type and a cartridge-type fuse (open and not open) from illustrations.
6. List the three characteristics by which fuses are rated and state the meaning of each rating. Identify a plug-type and a cartridge-type fuse (open and not open) from illustrations.
7. List the three categories of time delay rating for fuses and state a use for each type of time-delay rated fuse.
8. List the three categories of time delay rating for fuses and state a use for each type of time-delay rated fuse. Identify fuses as to voltage, current, and time delay ratings using fuses marked with the old military, new military, old commercial, and new commercial systems. List the three categories of time delay rating for fuses and state a use for each type of time-delay rated fuse.
9. Identify a clip-type and a post-type fuse holder from illustrations and identify the connections used on a post-type fuse holder for power source and load connections.
10. List the methods of checking for an open fuse, the items to check when replacing a fuse, the safety precautions to be observed when checking and replacing fuses, and the conditions to be checked for when conducting preventive maintenance on fuses.
11. Select a proper replacement and substitute fuse from a listing of fuses.
12. List the five main components of a circuit breaker and the three types of circuit breaker trip elements.
13. Describe the way in which each type of trip element reacts to excessive current.
14. Define the circuit breaker terms trip-free and nontrip-free and state one example for the use of each of these types of circuit breakers.
15. List the three time delay ratings of circuit breakers.
16. Define selective tripping, state why it is used, and state the way in which the time delay ratings of circuit breakers are used to design a selective tripping system.
17. Identify the factors used in selecting circuit breakers.

18. List the steps to follow before starting work on a circuit breaker and the items to be checked when maintaining circuit breakers.

CIRCUIT PROTECTION DEVICES

Electricity, like fire, can be either helpful or harmful to those who use it. A fire can keep people warm and comfortable when it is confined in a campfire or a furnace. It can be dangerous and destructive if it is on the loose and uncontrolled in the woods or in a building. Electricity can provide people with the light to read by or, in a blinding flash, destroy their eyesight. It can help save people's lives, or it can kill them. While we take advantage of the tremendous benefits electricity can provide, we must be careful to protect the people and systems that use it.

It is necessary then, that the mighty force of electricity be kept under control at all times. If for some reason it should get out of control, there must be a method of protecting people and equipment. Devices have been developed to protect people and electrical circuits from currents and voltages outside their normal operating ranges. Some examples of these devices are discussed in this chapter.

While you study this chapter, it should be kept in mind that a circuit protection device is used to keep an undesirably large current, voltage, or power surge out of a given part of an electrical circuit.

INTRODUCTION

An electrical unit is built with great care to ensure that each separate electrical circuit is fully insulated from all the others. This is done so that the current in a circuit will follow its intended path. Once the unit is placed into service, however, many things can happen to alter the original circuitry. Some of the changes can cause serious problems if they are not detected and corrected. While circuit protection devices cannot correct an abnormal current condition, they can indicate that an abnormal condition exists and protect personnel and circuits from that condition. In this chapter, you will learn what circuit conditions require protection devices and the types of protection devices used.

CIRCUIT CONDITIONS REQUIRING PROTECTION DEVICES

As has been mentioned, many things can happen to electrical and electronic circuits after they are in use. Chapter 1 of this module contains information showing you how to measure circuit characteristics to help determine the changes that can occur in them. Some of the changes in circuits can cause conditions that are dangerous to the circuit itself or to people living or working near the circuits. These potentially dangerous conditions require circuit protection. The conditions that require circuit protection are direct shorts, excessive current, and excessive heat.

Direct Short

One of the most serious troubles that can occur in a circuit is a DIRECT SHORT. Another term used to describe this condition is a SHORT CIRCUIT. The two terms mean the same thing and, in this chapter, the term direct short will be used. This term is used to describe a situation in which some point in the circuit, where full system voltage is present, comes in direct contact with the ground or return side of the circuit. This establishes a path for current flow that contains only the very small resistance present in the wires carrying the current.

According to Ohm's law, if the resistance in a circuit is extremely small, the current will be extremely large. Therefore, when a direct short occurs, there will be a very large current through the wires. Suppose, for instance, that the two leads from a battery to a motor came in contact with each other. If the leads were bare at the point of contact, there would be a direct short. The motor would stop running because all the current would be flowing through the short and none through the motor. The battery would become discharged quickly (perhaps ruined) and there could be the danger of fire or explosion.

The battery cables in our example would be large wires capable of carrying heavy currents. Most wires used in electrical circuits are smaller and their current carrying capacity is limited. The size of wire used in any given circuit is determined by space considerations, cost factors, and the amount of current the wire is expected to carry under normal operating conditions. Any current flow greatly in excess of normal, such as there would be in the case of a direct short, would cause a rapid generation of heat in the wire.

If the excessive current flow caused by the direct short is left unchecked, the heat in the wire will continue to increase until some portion of the circuit burns. Perhaps a portion of the wire will melt and open the circuit so that nothing is damaged other than the wire involved. The probability exists, however, that much greater damage will result. The heat in the wire can char and burn the insulation of the wire and that of other wires bundled with it, which can cause more shorts. If a fuel or oil leak is near any of the hot wires, a disastrous fire might be started.

Excessive Current

It is possible for the circuit current to increase without a direct short. If a resistor, capacitor, or inductor changes value, the total circuit impedance will also change in value. If a resistor <u>decreases</u> in ohmic value, the total circuit resistance decreases. If a capacitor has a dielectric leakage, the capacitive reactance <u>decreases</u>. If an inductor has a partial short of its winding, inductive reactance decreases. Any of these conditions will cause an <u>increase</u> in circuit current. Since the circuit wiring and components are designed to withstand normal circuit current, an increase in current would cause overheating (just as in the case of a direct short). Therefore, excessive current without a direct short will cause the same problems as a direct short.

Excessive Heat

As you have read, most of the problems associated with a direct short or excessive current concern the heat generated by the higher current. The damage to circuit components, the possibility of fire, and the possibility of hazardous fumes being given off from electrical components are consequences of excessive heat. It is possible for excessive heat to occur without a direct short or excessive current. If the bearings on a motor or generator were to fail, the motor or generator would overheat. If the temperature around an electrical or electronic circuit were to rise (through failure of a cooling system for example), excessive heat would be a problem. No matter what the cause, if excessive heat is present in a circuit, the possibility of damage, fire, and hazardous fumes exists.

Q1. Why are circuit protection devices necessary?

Q2. What are the three conditions that require circuit protection?

Q3. What is a direct short?

Q4. What is an excessive current condition?

Q5. What is an excessive heat condition?

CIRCUIT PROTECTION DEVICES

All of the conditions mentioned are potentially dangerous and require the use of circuit protection devices. Circuit protection devices are used to stop current flow or open the circuit. To do this, a circuit protection device must ALWAYS be connected in series with the circuit it is protecting. If the protection device is connected in parallel, current will simply flow around the protection device and continue in the circuit.

A circuit protection device operates by opening and interrupting current to the circuit. The opening of a protection device shows that something is wrong in the circuit and should be corrected before the current is restored. When a problem exists and the protection device opens, the device should isolate the faulty circuit from the other unaffected circuits, and should respond in time to protect unaffected components in the faulty circuit. The protection device should NOT open during normal circuit operation.

The two types of circuit protection devices discussed in this chapter are fuses and circuit breakers.

Fuses

A fuse is the simplest circuit protection device. It derives its name from the Latin word "fusus," meaning "to melt." Fuses have been used almost from the beginning of the use of electricity. The earliest type of fuse was simply a bare wire between two connections. The wire was smaller than the conductor it was protecting and, therefore, would melt before the conductor it was protecting was harmed. Some "copper fuse link" types are still in use, but most fuses no longer use copper as the fuse element (the part of the fuse that melts). After changing from copper to other metals, tubes or enclosures were developed to hold the melting metal. The enclosed fuse made possible the addition of filler material, which helps to contain the arc that occurs when the element melts.

For many low power uses, the finer material is not required. A simple glass tube is used. The use of a glass tube gives the added advantage of being able to see when a fuse is open. Fuses of this type are commonly found in automobile lighting circuits.

Figure 2-1 shows several fuses and the symbols used on schematics.

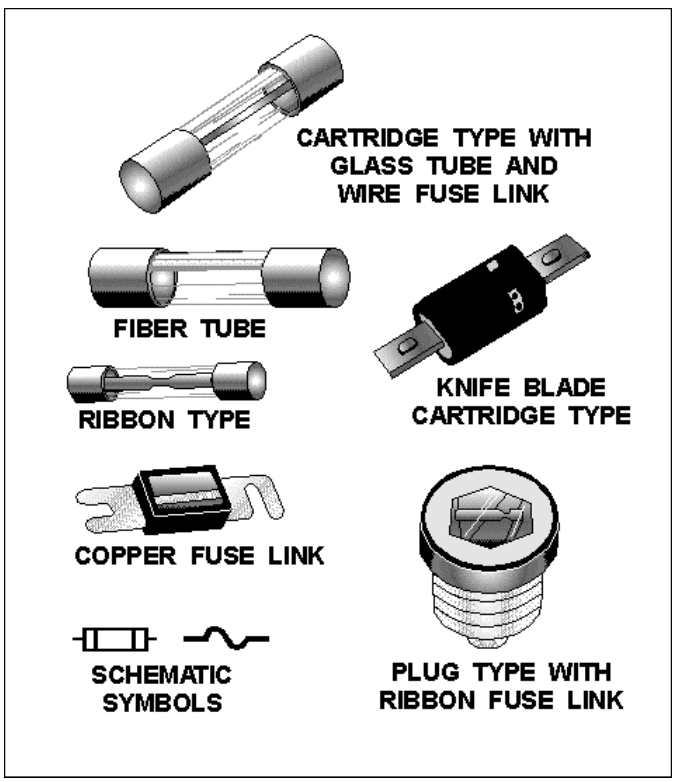


Figure 2-1.—Typical fuses and schematic symbols.

Circuit Breakers

While a fuse protects a circuit, it is destroyed in the process of opening the circuit. Once the problem that caused the increased current or heat is corrected, a new fuse must be placed in the circuit. A circuit protection device that can be used more than once solves the problems of replacement fuses. Such a device is safe, reliable, and tamper proof. It is also resettable, so it can be reused without replacing any parts. This device is called a CIRCUIT BREAKER because it breaks (opens) the circuit.

The first compact, workable circuit breaker was developed in 1923. It took 4 years to design a device that would interrupt circuits of 5000 amperes at 120 volts ac or dc. In 1928 the first circuit breaker was placed on the market. A typical circuit breaker and the appropriate schematic symbols are shown in figure 2-2.

Figure 2-2.—Typical circuit breaker and schematic symbols.

Q6. How are circuit protection devices connected to the circuit they are intended to protect and why are they connected in this way?

Q7. What are the two types of circuit protection devices?

Q8. Label the schematic symbols shown in figure 2-3 below.

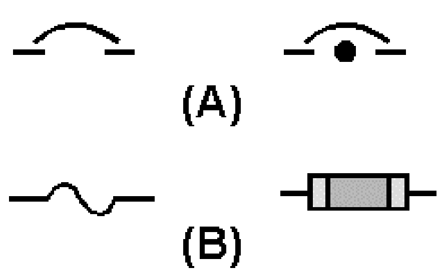


Figure 2-3.—Schematic symbols.

FUSE TYPES

Fuses are manufactured in many shapes and sizes. In addition to the copper fuse link already described, figure 2-1 shows other fuse types. While the variety of fuses may seem confusing, there are basically only two types of fuses: plug-type fuses and cartridge fuses. Both types of fuses use either a single wire or a ribbon as the fuse element (the part of the fuse that melts). The condition (good or bad) of some fuses can be determined by visual inspection. The condition of other fuses can only be determined with a meter. In the following discussion, visual inspection will be described. The use of meters to check fuses will be discussed later in this chapter.

PLUG-TYPE FUSE

The plug-type fuse is constructed so that it can be screwed into a socket mounted on a control panel or electrical distribution center. The fuse link is enclosed in an insulated housing of porcelain or glass. The construction is arranged so the fuse link is visible through a window of mica or glass. Figure 2-4 shows a typical plug-type fuse.

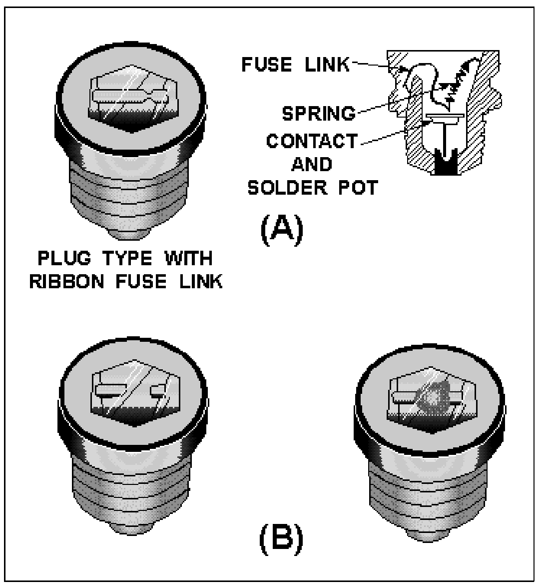


Figure 2-4.—Plug-type fuses:

Figure 2-4, view A, sows a good plug-type fuse. Notice the construction and the fuse link. In figure 2-4, view B, the same type of fuse is shown after the fuse link has melted. Notice the window showing the indication of this open fuse. The indication could be either of the ones shown in figure 2-4, view B.

The plug-type fuse is used primarily in low-voltage, low-current circuits. The operating range is usually up to 150 volts and from 0.5 ampere to 30 amperes. This type of fuse is found in older circuit protection devices and is rapidly being replaced by the circuit breaker.

CARTRIDGE FUSE

The cartridge fuse operates exactly like the plug-type fuse. In the cartridge fuse, the fuse link is enclosed in a tube of insulating material with metal ferrules at each end (for contact with the fuse holder). Some common insulating materials are glass, bakelite, or a fiber tube filled with insulating powder.

Figure 2-5 shows a glass-tube fuse. In figure 2-5, view A, notice the fuse link and the metal ferrules. Figure 2-5, view B, shows a glass-tube fuse that is open. The open fuse link could appear either of the ways shown in figure 2-5, view B.

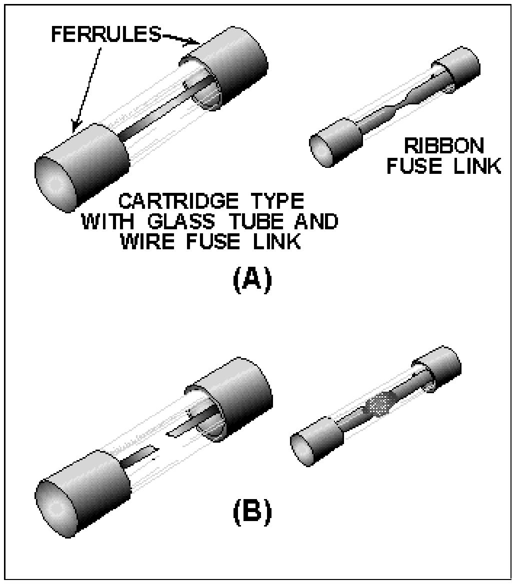


Figure 2-5.—Cartridge-tube fuse.

Cartridge fuses are available in a variety of physical sizes and are used in many different circuit applications. They can be rated at voltages up to 10,000 volts and have current ratings of from 1/500 (.002) ampere to 800 amperes. Cartridge fuses may also be used to protect against excessive heat and open at temperatures of from 165° F to 410°F (74°C to 210°C).

Q9. Label the fuses shown in figure 2-6 according to type.

Q10. Identify the open fuses shown in figure 2-6.

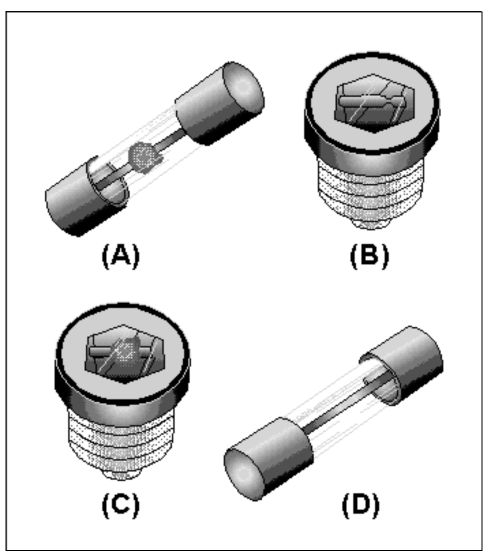


Figure 2-6.—Fuse recognition.

FUSE RATINGS

You can determine the physical size and type of a fuse by looking at it, but you must know other things about a fuse to use it properly. Fuses are rated by current, voltage, and time-delay characteristics to aid in the proper use of the fuse. To select the proper fuse, you must understand the meaning of each of the fuse ratings.

CURRENT RATING

The current rating of a fuse is a value expressed in amperes that represents the current the fuse will allow without opening. The current rating of a fuse is always indicated on the fuse.

To select the proper fuse, you must know the normal operating current of the circuit. If you wish to protect the circuit from overloads (excessive current), select a fuse rated at 125 percent of the normal circuit current. In other words, if a circuit has a normal current of 10 amperes, a 12.5-ampere fuse will provide overload protection. If you wish to protect against direct shorts only, select a fuse rated at 150 percent of the normal circuit current. In the case of a circuit with 10 amperes of current, a 15 ampere fuse will protect against direct shorts, but will not be adequate protection against excessive current.

VOLTAGE RATING

The voltage rating of a fuse is NOT an indication of the voltage the fuse is designed to withstand while carrying current. The voltage rating indicates the ability of the fuse to quickly extinguish the arc after the fuse element melts and the maximum voltage the open fuse will block. In other words, once the fuse has opened, any voltage less than the voltage rating of the fuse will not be able to "jump" the gap of the fuse. Because of the way the voltage rating is used, it is a maximum rms voltage value. You must always select a fuse with a voltage rating equal to or higher than the voltage in the circuit you wish to protect.

TIME DELAY RATING

There are many kinds of electrical and electronic circuits that require protection. In some of these circuits, it is important to protect against temporary or transient current increases. Sometimes the device being protected is very sensitive to current and cannot withstand an increase in current. In these cases, a fuse must open very quickly if the current increases.

Some other circuits and devices have a large current for short periods and a normal (smaller) current most of the time. An electric motor, for instance, will draw a large current when the motor starts, but normal operating current for the motor will be much smaller. A fuse used to protect a motor would have to allow for this large temporary current, but would open if the large current were to continue.

Fuses are time delay rated to indicate the relationship between the current through the fuse and the time it takes for the fuse to open. The three time delay ratings are delay, standard, and fast.

Delay

A delay, or slow-blowing, fuse has a built-in delay that is activated when the current through the fuse is greater than the current rating of the fuse. This fuse will allow temporary increases in current (surge) without opening. Some delay fuses have two elements; this allows a very long time delay. If the over-current condition continues, a delay fuse will open, but it will take longer to open than a standard or a fast fuse.

Delay fuses are used for circuits with high surge or starting currents, such as motors, solenoids, and transformers.

Standard

Standard fuses have no built-in time delay. Also, they are not designed to be very fast acting. Standard fuses are sometimes used to protect against direct shorts only. They may be wired in series with a delay fuse to provide faster direct short protection. For example, in a circuit with a 1-ampere delay fuse, a 5-ampere standard fuse may be used in addition to the delay fuse to provide faster protection against a direct short.

A standard fuse can be used in any circuit where surge currents are not expected and a very fast opening of the fuse is not needed. A standard fuse opens faster than a delay fuse, but slower than a fast rated fuse.

Standard fuses can be used for automobiles, lighting circuits, or electrical power circuits.

Fast

Fast fuses are designed to open very quickly when the current through the fuse exceeds the current rating of the fuse. Fast fuses are used to protect devices that are very sensitive to increased current. A fast fuse will open faster than a delay or standard fuse.

Fast fuses can be used to protect delicate instruments or semiconductor devices.

Figure 2-7 will help you understand the differences between delay, standard, and fast fuses. Figure 2-7 shows that, if a 1-ampere rated fuse had 2 amperes of current through it, (200% of the rated value), a fast fuse would open in about .7 second, a standard rated fuse would open in about 1.5 seconds, and a delay rated fuse would open in about 10 seconds. Notice that in each of the fuses, the time required to open the fuse decreases as the rated current increases.

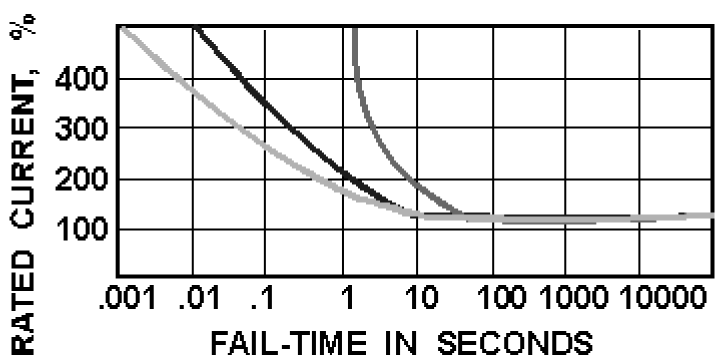


Figure 2-7.—Time required for fuse to open.

Q11. In what three ways are fuses rated?

Q12. What does the current rating of a fuse indicate?

Q13. What does the voltage rating of a fuse indicate?

Q14. What are the three time delay ratings of fuses?

Q15. Give an example of a device you could protect with each type of time delay fuse.

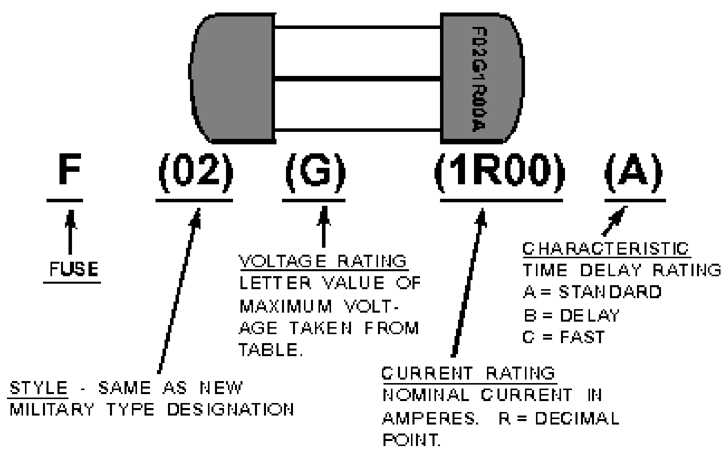
IDENTIFICATION OF FUSES

Fuses have identifications printed on them. The printing on the fuse will identify the physical size, the type of fuse, and the fuse ratings. There are four different systems used to identify fuses. The systems are the old military designation, the new military designation, the old commercial designation, and the new commercial designation. All four systems are presented here, so you will be able to identify a fuse no matter which designation is printed on the fuse.

You may have to replace an open fuse that is identified by one system with a good fuse that is identified by another system. The designation systems are fairly simple to understand and cross-reference once you are familiar with them.

OLD MILITARY DESIGNATION

Figure 2-8 shows a fuse with the old military designation. The tables in the lower part of the figure show the voltage and current codes used in this system. The upper portion of the figure is the explanation of the old military designation. The numbers and letters in parentheses are the coding for the fuse shown in figure 2-8.



VOLTAGE CODE LETTER	VOLTAGE (VOLTS OR LESS)
A	32
B	52
C	90
D	125
G	250
H	500
J	1,000
K	2,500
N	5,000
P	10,000

VOLTAGE CODE (OLD)

CURRENT CODE	CURRENT (AMPERES)
R002	.002 = 1/500
R005	.005 = 1/200
R010	.010 = 1/100
R031	.031 = 1/32
R750	.750 = 3/4
1R50	1.500 = 1 1/2

CURRENT CODE (OLD)

Figure 2-8.—Old type military fuse designation.

The old military designation always starts with "F," which stands for fuse. Next, the set of numbers (02) indicates the style. Style means the construction and dimensions (size) of the fuse. Following the style is a letter that represents the voltage rating of the fuse (G). The voltage code table in figure 2-8 shows each voltage rating letter and its meaning in volts. In the example shown, the voltage ratings is G, which means the fuse should be used in a circuit where the voltage is 250 volts or less. After this is a set of three numbers and the letter "R," which represent the current rating of the fuse. The "R" indicates the decimal point. In the example shown, the current rating is 1R00 or 1.00 ampere. Some other examples of the current rating are shown in the current code table of figure 2-8. The final letter in the old military designation (A) indicates the time delay rating of the fuse.

While the old military designation is still found on some fuses, the voltage and current ratings must be "translated," since they use letters to represent numerical values. The military developed the new military designations to make fuse identification easier.

NEW MILITARY DESIGNATION

Figure 2-9 is an example of a fuse coded in the new military designation. The fuse identified in the example in figure 2-9 is the same type as the fuse used as an example in figure 2-8.

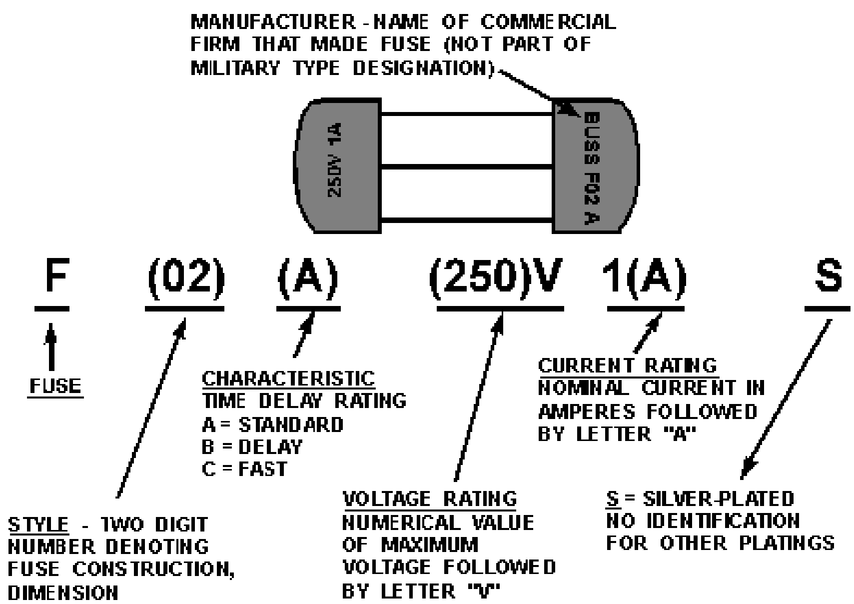


Figure 2-9.—New type military fuse designation.

The new military designation always start with the letter "F," which stands for fuse. The set of numbers (02) next to this indicates the style. The style numbers are identical to the ones used in the old military designation and indicate the construction and dimensions of the fuse. Following the style designation is a single letter (A) that indicates the time delay rating of the fuse. This is the same time delay rating code as indicated in the old military designation, but the position of this letter in the coding is changed to avoid confusing the "A" for standard time delay with the "A" for ampere. Following the time delay rating is the voltage rating of the fuse (250) V. In the old military designation, a letter was used to indicate the voltage rating. In the new military designation, the voltage is indicated by numbers followed by a "V," which stands for volts or less. After the voltage rating, the current rating is given by numbers followed by the letter "A." The current rating may be a whole number (1A), a fraction (1/500 A), a whole number and a fraction (1 1/2A), a decimal (0.250A), or a whole number and a decimal (1.50A). If the ferrules of the fuse are silver-plated, the current rating will be followed by the letter "S." If any other plating is used, the current rating will be the last part of the fuse identification.

As you can see, the new military designation is much easier to understand than the old military designation.

You may find a fuse coded in one of the commercial designations. The commercial designations are fairly easy to understand and figure 2-10 shows the old and new commercial designations for the same type of fuse that was used in figures 2-8 and 2-9.

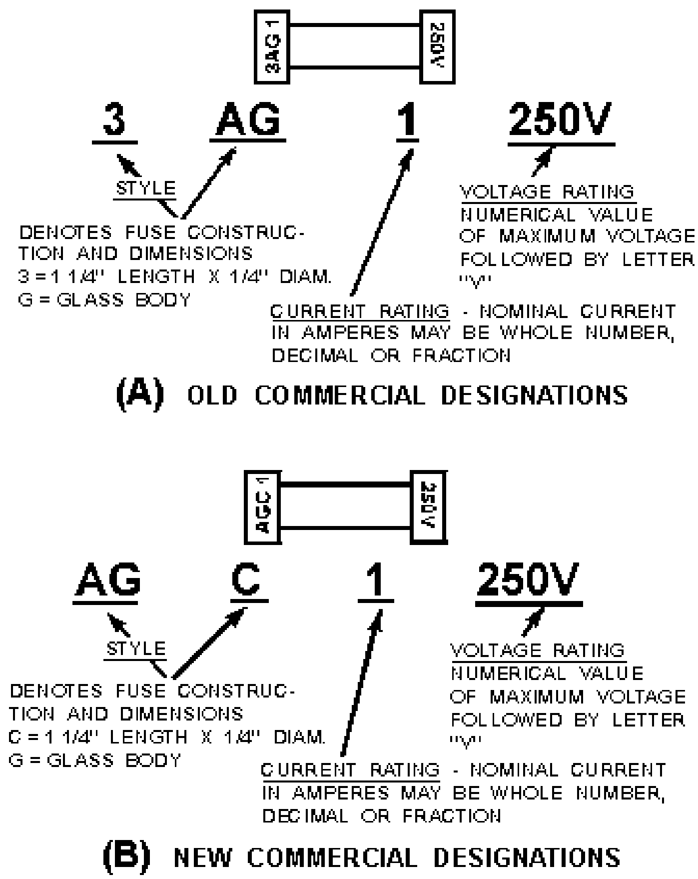


Figure 2-10.—Commercial designations for fuses:

OLD COMMERCIAL DESIGNATION

Figure 2-10, view A, shows the old commercial designation for a fuse. The first part of the designation is a combination of letters and numbers (three in all) that indicates the style and time delay characteristics. This part of the designation (3AG) is the information contained in the style and time delay rating portions of military designations.

In the example shown, the code 3AG represents the same information as the underlined portions of F02 G 1R00 A from figure 2-8 (Old Military Designation) and F02A 250VIAS from figure 2-9 (New Military Designation). The only way to know the time delay rating of this fuse is to look it up in the manufacturer's catalog or in a cross-reference listing to find the military designation. The catalog will tell you the physical size, the material from which the fuse is constructed, and the time delay rating of the fuse. A 3AG fuse is a glass-bodied fuse, 1/4 inch × 1 1/4 inches (6.35 millimeters × 31.8 millimeters) and has a standard time delay rating.

Following the style designation is a number that is the current rating of the fuse (1). This could be a whole number, a fraction, a whole number and a fraction, a decimal, or a whole number and a decimal. Following the current rating is the voltage rating; which, in turn, is followed by the letter "V," which stands for volts or less (250V).

NEW COMMERCIAL DESIGNATION

Figure 2-10, view B, shows the new commercial designation for fuses. It is the same as the old commercial designation except for the style portion of the coding. In the old commercial system, the style was a combination of letters and numbers. In the new commercial system, only letters are used. In the example shown, 3AG in the old system becomes AGC in the new system. Since "C" is the third letter of the alphabet, it is used instead of the "3" used in the old system. Once again, the only way to find out the time delay rating is to look up this coding in the manufacturer's catalog or to use a cross-reference listing. The remainder of the new commercial designation is exactly the same as the old commercial designation.

Q16. What are the voltage, current, and time delay ratings for a fuse with the designation

(a) F02DIR50B?

(b) F02A250V$\frac{1}{8}$A ?

Q17. What are the voltage and current ratings for a fuse designated

(a) 3AG$\frac{1}{16}$125V?

(b) FNA$\frac{15}{100}$250V?

Q18. What is the new military designation for a fuse with the old military designation F05A20ROB?

FUSEHOLDERS

For a fuse to be useful, it must be connected to the circuit it will protect. Some fuses are "wired in" or soldered to the wiring of circuits, but most circuits make use of FUSEHOLDERS. A fuseholder is a device that is wired into the circuit and allows easy replacement of the fuse.

Fuseholders are made in many shapes and sizes, but most fuseholders are basically either clip-type or post-type. Figure 2-11 shows a typical clip-type and post-type fuseholder.

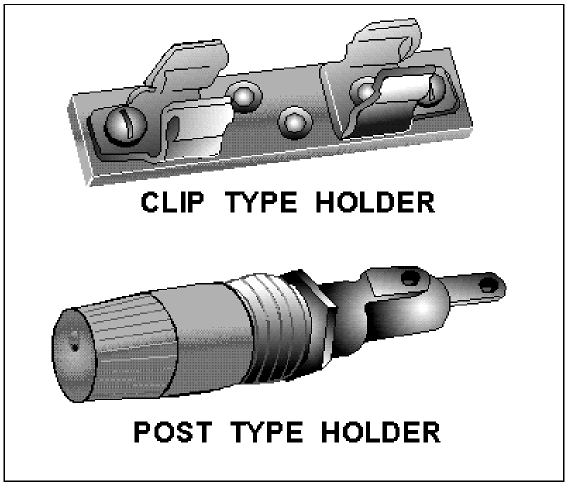


Figure 2-11.—Typical fuseholders.

CLIP-TYPE FUSEHOLDER

The clip-type fuseholder is used for cartridge fuses. The ferrules or knife blade of the fuse are held by the spring tension of the clips. These clips provide the electrical connection between the fuse and the circuit. If a glass-bodied fuse is used, the fuse can be inspected visually for an open without removing the fuse from the fuse holder. Clip-type fuseholders are made in several sizes to hold the many styles of fuses. The clips maybe made for ferrules or knife blade cartridge fuses. While the base of a clip-type fuseholder is made from insulating material, the clips themselves are conductors. The current through the fuse goes through the clips and care must be taken to not touch the clips when there is power applied. If the clips are touched, with power applied, a severe shock or a short circuit will occur.

POST-TYPE FUSEHOLDERS

Post-type fuseholders are made for cartridge fuses. The post-type fuseholder is much safer because the fuse and fuse connections are covered with insulating material. The disadvantage of the post-type fuseholder is that the fuse must be removed to visually check for an open. The post-type fuseholder has a cap that screws onto the body of the fuseholder. The fuse is held in this cap by a spring-type connector and, as the cap is screwed on, the fuse makes contact with the body of the fuseholder. When the cap and fuse are removed from the body of the fuseholder, the fuse is removed from the circuit and there is no danger of shock or short circuit from touching the fuse.

Post-type fuseholders are usually mounted on the chassis of the equipment in which they are used. After wires are connected to the fuseholder, insulating sleeves are placed over the connections to reduce the possibility of a short circuit. Notice the two connections on the post-type fuseholder of figure 2-11. The connection on the right is called the center connector. The other connector is the outside connector. The outside connector will be closer to the equipment chassis. (The threads and nut shown are used to fasten the fuseholder to the chassis.) The possibility of the outside connector coming in contact with the chassis (causing a short circuit) is much higher than the possibility of the center conductor contacting the chassis. The power source should always be connected to the center connector so the fuse will open if the outside connector contacts the chassis. If the power source were connected to the outside connector, and the outside connector contacted the chassis, there would be a direct short, but the fuse would not open.

Q19. Label the fuseholders in figure 2-12.

Q20. Which connector should you use to connect the (a) power source and (b) load to the fuseholder shown in figure 2-12(A)?

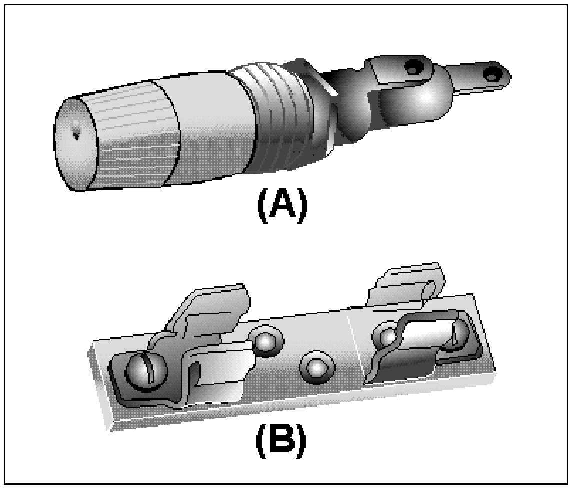


Figure 2-12.—Fuseholder identification.

CHECKING AND REPLACEMENT OF FUSES

A fuse, if properly used, should not open unless something is wrong in the circuit the fuse is protecting. When a fuse is found to be open, you must determine the reason the fuse is open. Replacing the fuse is not enough.

Before you look for the cause of an open fuse, you must be able to determine if the fuse is open.

CHECKING FOR AN OPEN FUSE

There are several ways of checking for an open fuse. Some fuses and fuseholders have indicators built in to help you find an open fuse; also, a multimeter can be used to check fuses. The simplest way to check glass-bodied fuses, and the method you should use first, is visual inspection.

Visual Inspection

An open glass-bodied fuse can usually be found by visual inspection. Earlier in this chapter, figures 2-4 and 2-5 showed you how an open plug-type and an open glass-bodied cartridge-type fuse would look. If the fuse element is not complete, or if the element has been melted onto the glass tube, the fuse is open.

It is not always possible to tell if a fuse is open by visual inspection. Fuses with low current ratings have elements that are so small, it is sometimes not possible to know if the fuse link is complete simply by looking at it. If the fuse is not glass-bodied, it will not be possible to check the fuse visually. Also, sometimes a fuse will look good, but will, in fact, be open. Therefore, while it is sometimes possible to know if a fuse is open by visual inspection, it is not possible to be sure a fuse is good just by looking at it.

Fuse Indicators

Some fuses and fuseholders have built-in indicators to show when a fuse is open. Examples of these open-fuse indicators are shown in figure 2-13. Figure 2-13, view A, shows a cartridge-type fuse with an open-fuse indicator. The indicator is spring loaded and held by the fuse link. If the fuse link opens, the spring forces the indicator out. Some manufacturers color the indicator so it is easier to see in the open-fuse position.

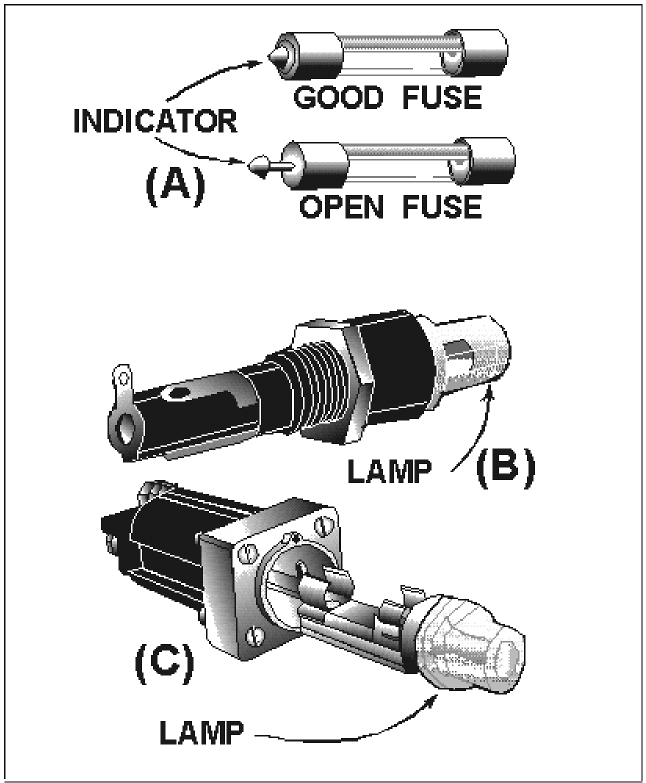


Figure 2-13.—Open fuse indicators: Clip-type fuseholder with an indicating lamp.

Figure 2-13, view B, shows a plug-type fuseholder with an indicating lamp in the fuse cap. If the fuse opens, the lamp in the fuse cap will light. Figure 2-13, view C, shows a clip-type fuseholder with an indicating lamp.

Just as in visual checking, the indicator can show an open fuse. Since the indicator may not always work, you cannot be sure a fuse is good just because there is no open-fuse indication.

Checking Fuses with a Meter

The only sure method of determining if a fuse is open is to use a meter. An ohmmeter can be used to check for an open fuse by removing the fuse from the circuit and checking for continuity through the fuse (0 ohms). If the fuse is not removed from the circuit, and the fuse is open, the ohmmeter may measure the circuit resistance. This resistance reading might lead you to think the fuse is good. You must be careful when you use an ohmmeter to check fuses with small current ratings (such as 1/32 ampere or less), because the current from the ohmmeter may be larger than the current rating of the fuse. For most practical uses, a small current capacity fuse can be checked out of the circuit through the use of a resistor. The ohmic value of the resistor is first measured and then placed in series with the fuse. The continuity reading on the ohmmeter should be of the same value, or close to it, as the original value of the resistor. This method provides protection for the fuse by dropping the voltage across the resistor. This in turn decreases the power in the form of heat at the fuse. Remember, it is heat which melts the fuse element.

A voltmeter can also be used to check for an open fuse. The measurement is taken between each end of the fuse and the common or ground side of the line. If voltage is present on <u>both</u> sides of the fuse (from the voltage source and to the load), the fuse is not open. Another method commonly used, is to measure <u>across</u> the fuse with the voltmeter. If NO voltage is indicated on the meter, the fuse is good, (not open).

Remember there is no voltage drop across a straight piece of wire. Some plug-type fuseholders have test points built in to allow you to check the voltage. To check for voltage on a clip-type fuseholder, check each of the clips. The advantage of using a voltmeter to check for an open fuse is that the circuit does not have to be deenergized and the fuse does not have to be removed.

WARNING

PERSONNEL MAY BE EXPOSED TO HAZARDOUS VOLTAGE

Safety Precautions When Checking a Fuse

Since a fuse has current through it, you must be very careful when checking for an open fuse to avoid being shocked or damaging the circuit. The following safety precautions will protect you and the equipment you are using.

- Turn the power off and discharge the circuit before removing a fuse.
- Use a fusepuller (an insulated tool) when you remove a fuse from a clip-type fuseholder.
- When you check a fuse with a voltmeter, be careful to avoid shocks and short circuits.
- When you use an ohmmeter to check fuses with low current ratings, be careful to avoid opening the fuse by excessive current from the ohmmeter.

Q21. What are three methods for determining if a fuse is open?

Q22. You have just checked a fuse with an ohmmeter and find that the fuse is shorted. What should you do?

Q23. You have just checked a 1/500-ampere fuse with an ohmmeter and find it is open. Checking the replacement fuse shows the replacement fuse is open also. Why would the replacement fuse indicate open?

Q24. How could you check a 1/500-ampere fuse with an ohmmeter?

Q25. List the safety precautions to be observed when checking fuses.

REPLACEMENT OF FUSES

After an open fuse is found and the trouble that caused the fuse to open has been corrected, the fuse must be replaced. Before you replace the fuse, you must be certain the replacement fuse is the proper type and fits correctly.

Proper Type of Replacement Fuse

To be certain a fuse is the proper type, <u>check the technical manual</u> for the equipment. The parts list will show you the proper fuse identification for a replacement fuse. Obtain the exact fuse specified, if possible, and check the identification number of the replacement fuse against the parts list.

If you cannot obtain a direct replacement, use the following guidelines:

- Never use a fuse with a higher current rating, a lower voltage rating, or a slower time delay rating than the specified fuse.

- The best substitution fuse is a fuse with the same current and time delay ratings and a higher voltage rating.

- If a lower current rating or a faster time delay rating is used, the fuse may open under normal circuit conditions.

- Substitute fuses must have the same style (physical dimensions) as the specified fuse.

Proper Fit of Replacement Fuses

When you have obtained a proper replacement fuse, you must make certain it will fit correctly in the fuseholder. If the fuseholder is corroded, the fuse will not fit properly. In addition, the corrosion can cause increased resistance or heating. Clean corroded terminals with fine sandpaper so that all corrosion is removed. Do **NOT** lubricate the terminals. If the terminals are badly pitted, replace the fuseholder. Be certain the replacement fuseholder is the correct size and type by checking the parts list in the technical manual for the equipment.

After you check for and correct any corrosion problems, be certain the fuse fits tightly in the fuseholder. When you insert the fuse in the cap of a plug-type fuseholder, the fuse should fit tightly. A small amount of pressure should be needed to insert the fuse and cap into the fuseholder body.

In clip-type fuseholders, the clips can be easily bent out of shape. This causes an incorrect fit, which in time could cause an equipment malfunction. Figure 2-14 shows examples of correct and incorrect fuse contacts for clip-type fuseholders used with knifeblade and ferrule cartridge fuses. The clips shown in the left picture of each row have the correct contact. The three pictures on the right of each row show incorrect contact. Notice how the clips are not contacting completely with the knifeblade or ferrules. This incomplete contact can. cause corrosion at the contacts, which in turn can create a high resistance and drop some of the circuit voltage at this point.

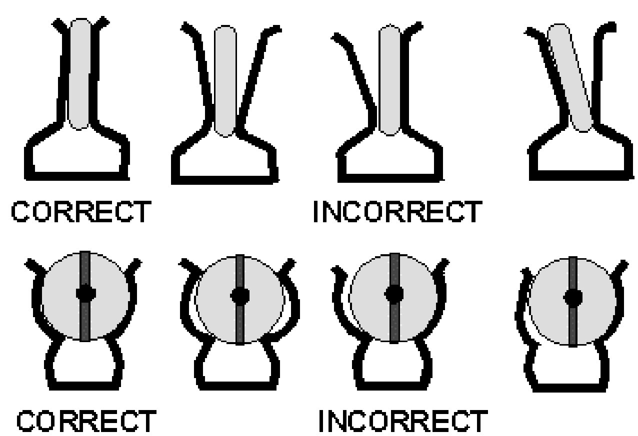


Figure 2-14.—Contact between clips and fuses.

If the fuse clips do not make complete contact with the fuse, try to bend the clips back into shape. If the clips cannot be repaired by bending, replace the fuseholder or use clip clamps. Clip clamps are shown in figure 2-15.

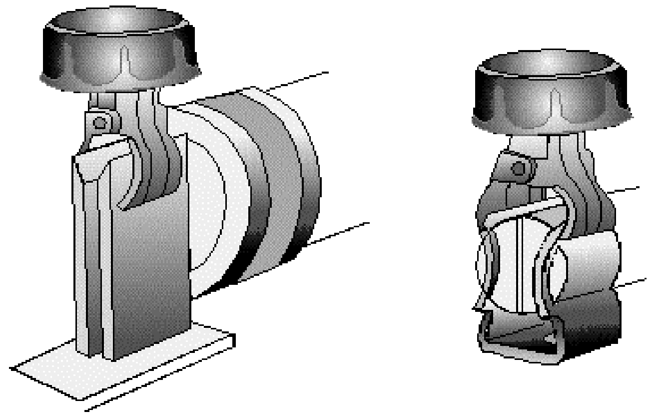


Figure 2-15.—Clip clamps.

Safety Precautions When Replacing Fuses

The following safety precautions will prevent injury to personnel and damage to equipment. These are the MINIMUM safety precautions for replacing fuses.

- Be sure the power is off in the circuit and the circuit is discharged before replacing a fuse.
- Use an identical replacement fuse if possible.
- Remove any corrosion from the fuseholder before replacing the fuse.
- Be certain the fuse properly fits the fuseholder.

PREVENTIVE MAINTENANCE OF FUSES

Preventive maintenance of fuses consists of checking for the following conditions and correcting any discrepancies.

1. IMPROPER FUSE. Check the fuse installed against that recommended in the technical manual for the equipment. If an incorrect fuse is installed, replace it with the correct fuse.
2. CORROSION. Check for corrosion on the fuseholder terminals or the fuse itself. If corrosion is present, remove it with fine sandpaper.
3. IMPROPER FIT. Check for contact between the fuse and fuseholder. If a piece of paper will fit between the fuse and the clips on a clip-type fuseholder, there is improper contact. If the fuse is not held in the cap of a plug-type fuseholder, the contacts are too loose.
4. OPEN FUSES. Check fuses for opens. If any fuse is open, repair the trouble that caused the open fuse and replace the fuse.

Q26. You have removed an open fuse from a fuseholder and repaired the cause of the fuse opening. The parts list specifies a fuse coded F02BI25V½A. There are no fuses available with that identification. In the following list, indicate if the fuse is a direct replacement, a good substitute, or not acceptable. For the fuses that are good substitutes, number them in order of preference and explain why they are numbered that way. If the fuse is not acceptable, explain why.

(a) F03BI25V½A

(b) F02BI25V⅜A

(c) F02GR500B

(d) F02B32V½A

(e) F02DR500B

(f) F02A250V⅝A

(g) F02AI25V½A

Q27. What two things should you check before replacing a fuse?

Q28. List the safety precautions to be observed when replacing a fuse.

Q29. What conditions should you check for when conducting preventive maintenance on fuses?

CIRCUIT BREAKERS

A circuit breaker is a circuit protection device that, like a fuse, will stop current in the circuit if there is a direct short, excessive current, or excessive heat. Unlike a fuse, a circuit breaker is reusable. The circuit breaker does not have to be replaced after it has opened or broken the circuit. Instead of replacing the circuit breaker, you reset it.

Circuit breakers can also be used as circuit control devices. By manually opening and closing the contacts of a circuit breaker, you can switch the power on and off. Circuit control devices will be covered in more detail in the next chapter.

Circuit breakers are available in a great variety of sizes and types. It would not be possible to describe every type of circuit breaker in use today, but this chapter will describe the basic types of circuit breakers and their operational principles.

Circuit breakers have five main components, as shown in figure 2-16. The components are the frame, the operating mechanism, the arc extinguishers and contacts, the terminal connectors, and the trip elements.

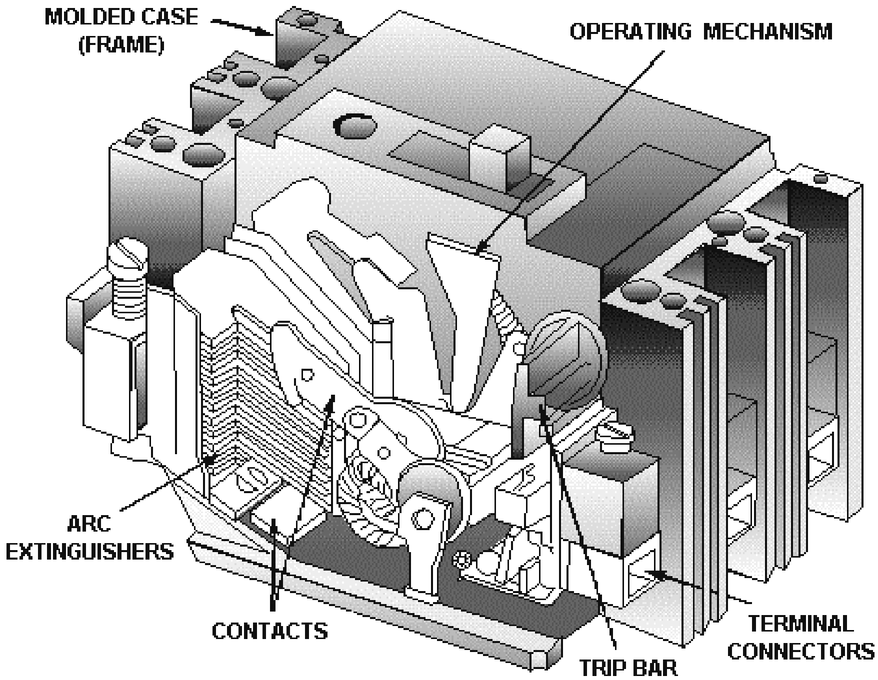


Figure 2-16.—Circuit breaker components.

The FRAME provides an insulated housing and is used to mount the circuit breaker components (fig. 2-17). The frame determines the physical size of the circuit breaker and the maximum allowable voltage and current.

The OPERATING MECHANISM provides a means of opening and closing the breaker contacts (turning, the circuit ON and OFF). The toggle mechanism shown in figure 2-17 is the quick-make, quick-break type, which means the contacts snap open or closed quickly, regardless of how fast the handle is moved. In addition to indicating whether the breaker is ON or OFF, the operating mechanism handle indicates when the breaker has opened automatically (tripped) by moving to a position between ON and OFF. To reset the circuit breaker, the handle must first be moved to the OFF position, and then to the ON position.

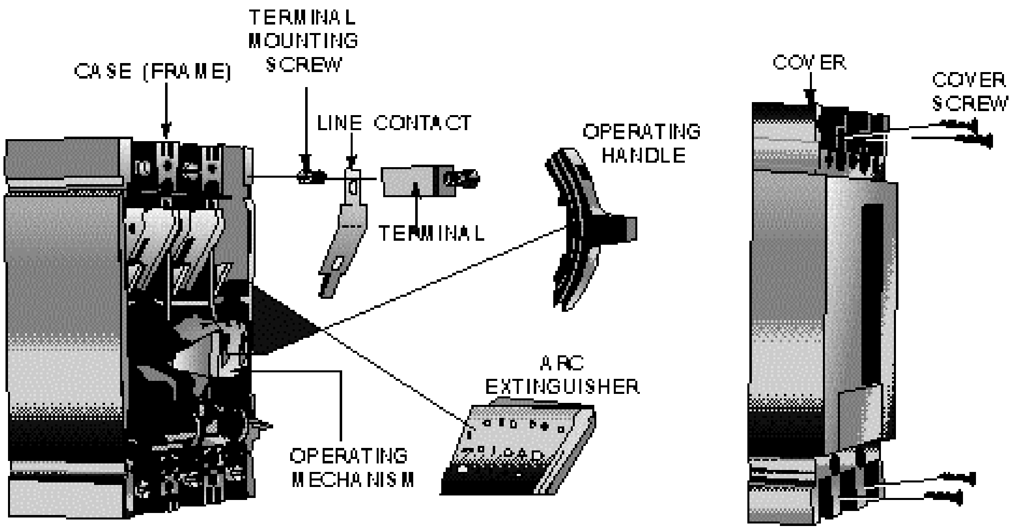


Figure 2-17.—Circuit breaker construction.

The ARC EXTINGUISHER confines, divides, and extinguishes the arc drawn between contacts each time the circuit breaker interrupts current. The arc extinguisher is actually a series of contacts that open gradually, dividing the arc and making it easier to confine and extinguish. This is shown in figure 2-18. Arc extinguishers are generally used in circuit breakers that control a large amount of power, such as those found in power distribution panels. Small power circuit breakers (such as those found in lighting panels) may not have arc extinguishers.

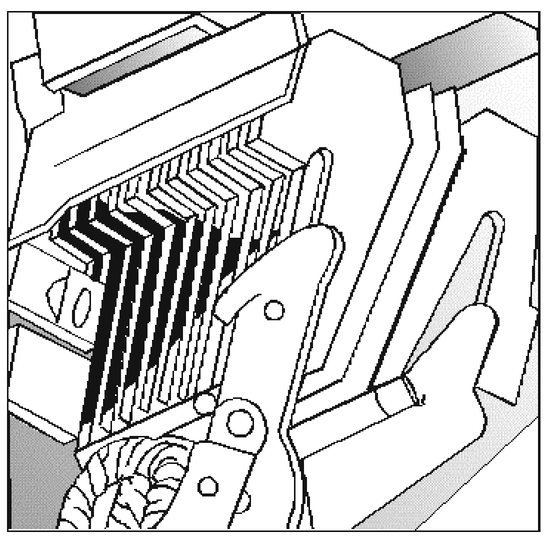


Figure 2-18.—Arc extinguisher action.

TERMINAL CONNECTORS are used to connect the circuit breaker to the power source and the load. They are electrically connected to the contacts of the circuit breaker and provide the means of connecting the circuit breaker to the circuit.

The TRIP ELEMENT is the part of the circuit breaker that senses the overload condition and causes the circuit breaker to trip or break the circuit. This chapter will cover the thermal, magnetic, and thermal-magnetic trip units used by most circuit breakers. (Some circuit breakers make use of solid-state trip units using current transformers and solid-state circuitry.)

THERMAL TRIP ELEMENT

A thermal trip element circuit breaker uses a bimetallic element that is heated by the load current. The bimetallic element is made from strips of two different metals bonded together. The metals expand at different rates as they are heated. This causes the bimetallic element to bend as it is heated by the current going to the load. Figure 2-19 shows how this can be used to trip the circuit breaker.

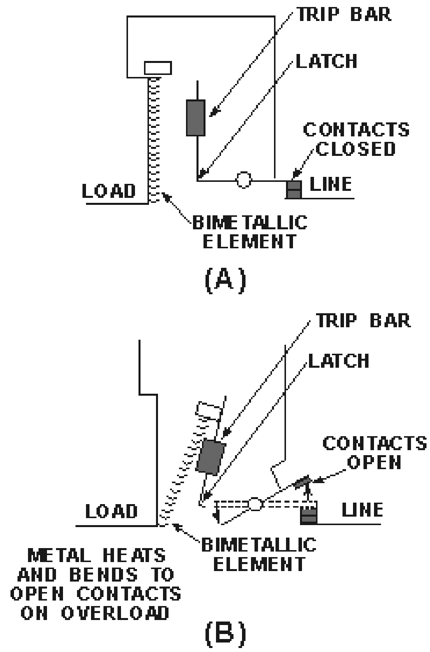


Figure 2-19.—Thermal trip element action: A. Trip element with normal current; B. Contacts open.

Figure 2-19, view A, shows the trip element with normal current. The bimetallic element is not heated excessively and does not bend. If the current increases (or the temperature around the circuit breaker increases), the bimetallic element bends, pushes against the trip bar, and releases the latch. Then, the contacts open, as shown in figure 2-19, view B.

The amount of time it takes for the bimetallic element to bend and trip the circuit breaker depends on the amount the element is heated. A large overload will heat the element quickly. A small overload will require a longer time to trip the circuit breaker.

MAGNETIC TRIP ELEMENT

A magnetic trip element circuit breaker uses an electromagnet in series with the circuit load as in figure 2-20. With normal current, the electromagnet will not have enough attraction to the trip bar to move it, and the contacts will remain closed as shown in figure 2-20, view A. The strength of the magnetic field of the electromagnet increases as current through the coil increases. As soon as the current in the circuit becomes large enough, the trip bar is pulled toward the magnetic element (electromagnet), the contacts are opened, and the current stops, as shown in figure 2-20, view B.

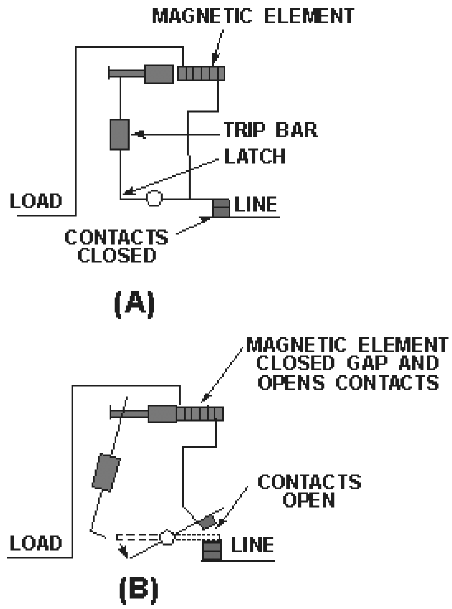


Figure 2-20.—Magnetic trip element action; Closed contacts;

The amount of current needed to trip the circuit breaker depends on the size of the gap between the trip bar and the magnetic element. On some circuit breakers, this gap (and therefore the trip current) is adjustable.

THERMAL-MAGNETIC TRIP ELEMENT

The thermal trip element circuit breaker, like a delay fuse, will protect a circuit against a small overload that continues for a long time. The larger the overload, the faster the circuit breaker will trip. The thermal element will also protect the circuit against temperature increases. A magnetic circuit breaker will trip instantly when the preset current is present. In some applications, both types of protection are desired. Rather than use two separate circuit breakers, a single trip element combining thermal and magnetic trip elements is used. A thermal-magnetic trip element is shown in figure 2-21.

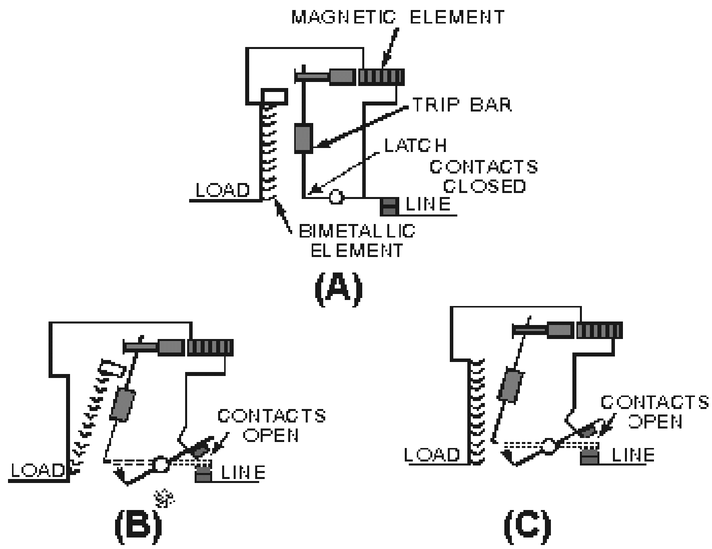


Figure 2-21.—Thermal-magnetic element action:

In the thermal-magnetic trip element circuit breaker, a magnetic element (electromagnet) is connected in series with the circuit load, and a bimetallic element is heated by the load current. With normal circuit current, the bimetallic element does not bend, and the magnetic element does not attract the trip bar, as shown in figure 2-21, view A.

If the temperature or current increases over a sustained period of time, the bimetallic element will bend, push the trip bar and release the latch. The circuit breaker will trip as shown in figure 2-21, view B.

If the current suddenly or rapidly increases enough, the magnetic element will attract the trip bar, release the latch, and the circuit breaker will trip, as shown in figure 2-21, view C. (This circuit breaker has tripped even though the thermal element has not had time to react to the increased current.)

Q30. What are the five main components of a circuit breaker?

Q31. What are the three types of circuit breaker trip elements?

Q32. How does each type of trip element react to an overload?

TRIP-FREE/NONTRIP-FREE CIRCUIT BREAKERS

Circuit breakers are classified as being trip free or nontrip free. A trip-free circuit breaker is a circuit breaker that will trip (open) even if the operating mechanism (ON-OFF switch) is held in the ON position. A nontrip-free circuit breaker can be reset and/or held ON even if an overload or excessive heat condition is present. In other words, a nontrip-free circuit breaker can be bypassed by holding the operating mechanism ON.

Trip-free circuit breakers are used on circuits that cannot tolerate overloads and on nonemergency circuits. Examples of these are precision or current sensitive circuits, nonemergency lighting circuits, and nonessential equipment circuits.Nontrip-free circuit breakers are used for circuits that are essential for operations. Examples of these circuits are emergency lighting, required control circuits, and essential equipment circuits.

TIME DELAY RATINGS

Circuit breakers, like fuses, are rated by the amount of time delay. In circuit breakers the ratings are instantaneous, short time delay, and longtime delay. The delay times of circuit breakers can be used to provide for SELECTIVE TRIPPING.

Selective tripping is used to cause the circuit breaker closest to the faulty circuit to trip. This will remove power from the faulty circuit without affecting other, nonfaulty circuits. Figure 2-22 should help you understand selective tripping.

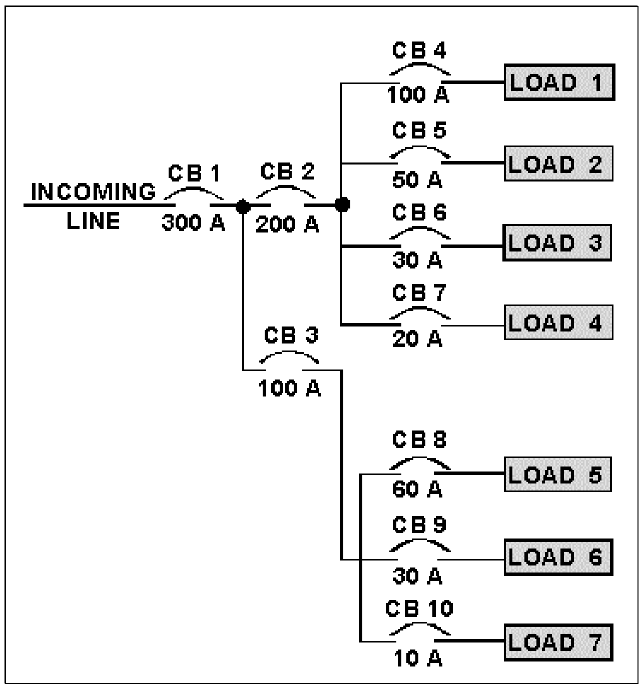


Figure 2-22.—Use of circuit breakers in a power distribution system.

Figure 2-22 shows a power distribution system using circuit breakers for protection. Circuit breaker 1 (CB1) has the entire current for all seven loads through it. CB2 feeds loads 1, 2, 3, and 4 (through CB4, CB5, CB6, and CB7), and CB3 feeds loads 5, 6, and 7 (through CB8, CB9, and CB10). If all the circuit breakers were rated with the same time delay, an overload on load 5 could cause CB1, CB3, and CB8 to trip. This would remove power from all seven loads, even though load 5 was the only circuit with an overload.

Selective tripping would have CB1 rated as long time delay, CB2 and CB3 rated as short time delay, and CB4 through CB10 rated as instantaneous. With this arrangement, if load 5 had an overload, only CB8 would trip. CB8 would remove the power from load 5 before CB1 or CB3 could react to the overload. In this way, only load 5 would be affected and the other circuits would continue to operate.

PHYSICAL TYPES OF CIRCUIT BREAKERS

All the circuit breakers presented so far in this chapter have been physically large, designed to control large amounts of power, and used a type of toggle operating mechanism. Not all circuit breakers are of this type. The circuit breaker in figure 2-23 is physically large and controls large amounts of power; but the operating mechanism is not a toggle. Except for the difference in the operating mechanism, this circuit breaker is identical to the circuit breakers already presented.

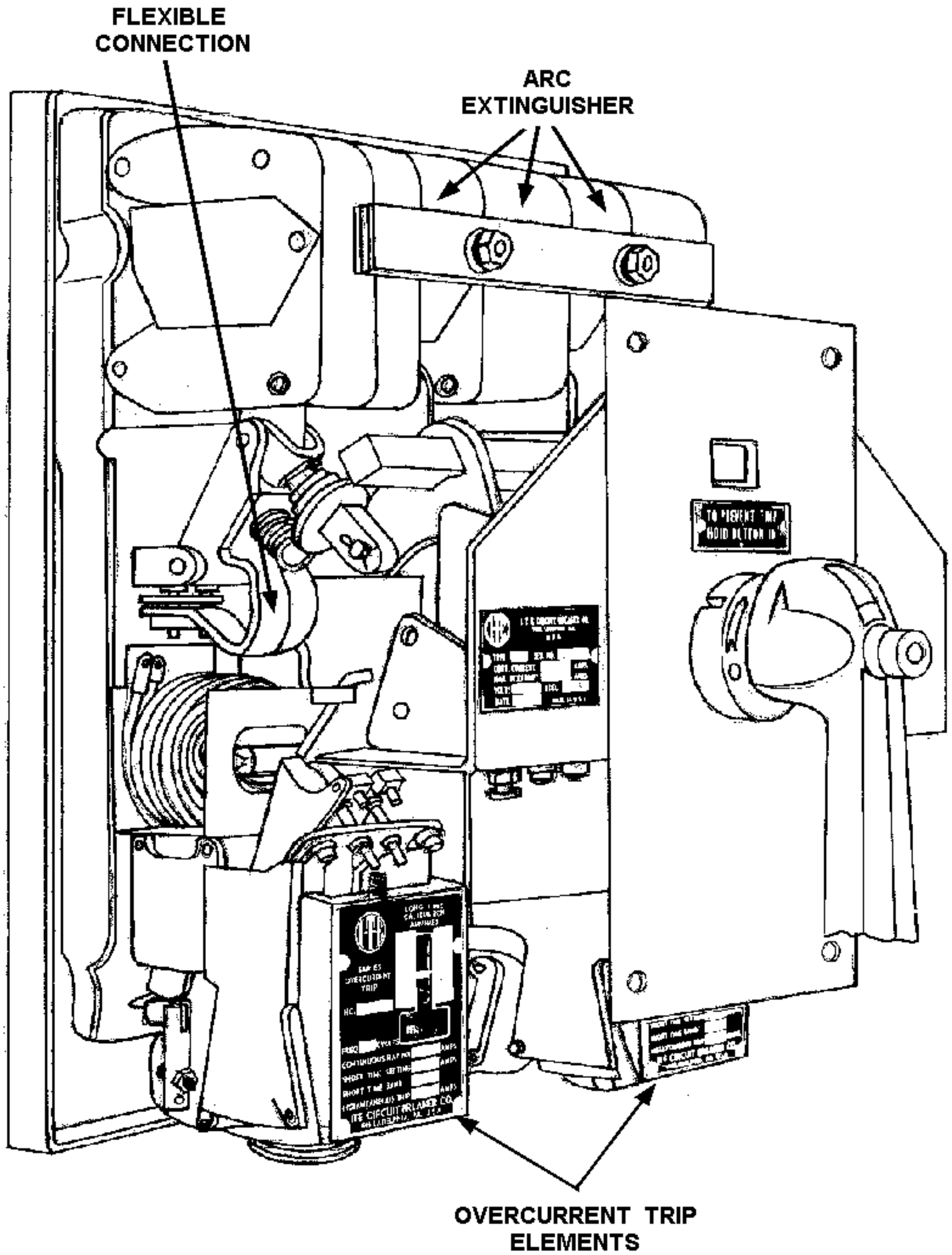


Figure 2-23.—Circuit breaker with an operating handle.

Circuit breakers used for low power protection, such as 28-volt dc, 30 amperes, can be physically small. With low power use, arc extinguishers are not required, and so are not used in the construction of these circuit breakers. Figure 2-24 shows a low power circuit breaker of the push-button or push-pull type. This circuit breaker has a thermal trip element (the bimetallic disk) and is nontrip-free. The push button is the operating mechanism of this circuit breaker.

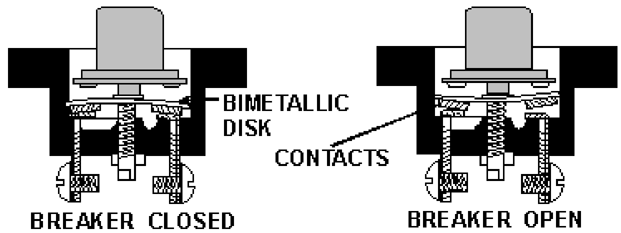


Figure 2-24.—Push-button circuit breaker.

You will find other physical types of circuit breakers as you work with electrical circuits. They are found in power distribution systems, lighting panels, and even on individual pieces of equipment. Regardless of the physical size and the amount of power through the circuit breaker, the basic operating principles of circuit breakers apply.

Q33. What is a trip-free circuit breaker?

Q34. What is a nontrip-free circuit breaker?

Q35. Where should you use a trip-free circuit breaker?

Q36. Where should you use a nontrip-free circuit breaker?

The magnetic trip element makes use of a magnetic element (electromagnet). If current reaches a preset quantity, the magnetic element attracts the trip bar and releases the latch.

The thermal-magnetic trip element combines the actions of the bimetallic and magnetic elements in a single trip element. If either the bimetal element or the magnetic element reacts, the circuit breaker will trip.

Q37. What are the three time delay ratings for circuit breakers?

Q38. What is selective tripping and why is it used?

Q39. If the power distribution system shown in figure 2-22 uses selective tripping, what is the time delay rating for each of the circuit breakers shown?

Q40. What factors are used to select a circuit breaker?

Q41. What type of circuit breaker is used on a multimeter?

CIRCUIT BREAKER MAINTENANCE

Circuit breakers require careful inspection and periodic cleaning. Before you attempt to work on circuit breakers, check the applicable technical manual carefully. When you work on shipboard circuit breakers, the approval of the electrical or engineering officer must be obtained before starting work. Be certain to remove all power to the circuit breaker before you work on it. Tag the switch that removes the power to the circuit breaker to ensure that power is not applied while you are working.

Once approval has been obtained, the incoming power has been removed, the switch tagged, and you have checked the technical manual, you may begin to check the circuit breaker. Manually operate the circuit breaker several times to be sure the operating mechanism works smoothly. Inspect the contacts for pitting caused by arcing or corrosion. If pitting is present, smooth the contacts with a fine file or number 00 sandpaper. Be certain the contacts make proper contact when the operating mechanism is ON.

Check the connections at the terminals to be certain the terminals and wiring are tight and free from corrosion. Check all mounting hardware for tightness and wear. Check all components for wear. Clean the circuit breaker completely.

When you have finished working on the circuit breaker, restore power and remove the tag from the switch that applies power to the circuit.

Q42. What steps are to be taken before beginning work on a circuit breaker?

Q43. What items are you to check when working on a circuit breaker?

SUMMARY

This chapter has provided the information to enable you to have a basic understanding of circuit protection devices. The following is a summary of the main points in this chapter.

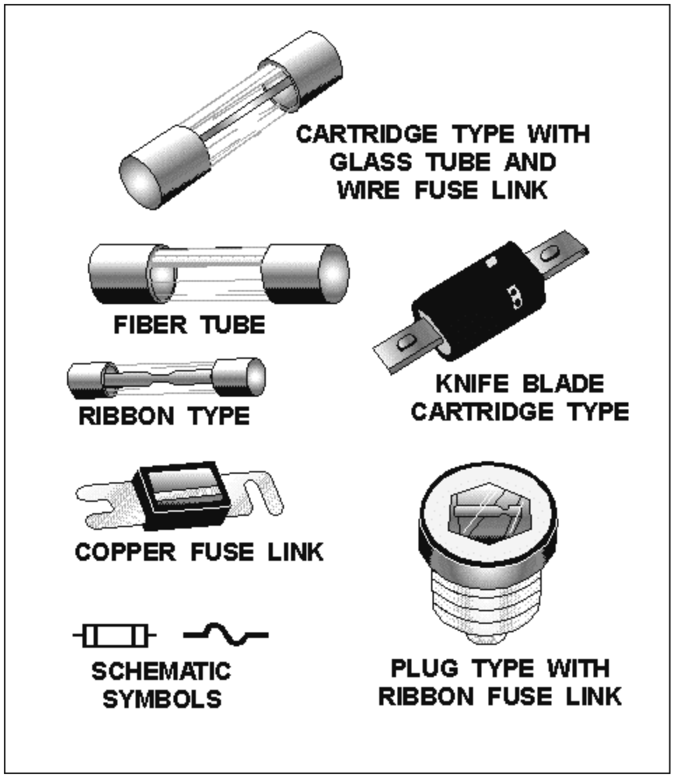
CIRCUIT PROTECTION DEVICES are needed to protect personnel and circuits from hazardous conditions. The hazardous conditions can be caused by a direct short, excessive current, or excessive heat. Circuit protection devices are always connected in series with the circuit being protected.

A DIRECT SHORT is a condition in which some point in the circuit, where full system voltage is present, comes in direct contact with the ground or return side of the circuit.

EXCESSIVE CURRENT describes a condition that is not a direct short but in which circuit current increases beyond the designed current carrying ability of the circuit.

EXCESSIVE HEAT describes a condition in which the heat in or around a circuit increases to a higher than normal level.

FUSES and **CIRCUIT BREAKERS** are the two types of circuit protection devices discussed in this chapter.



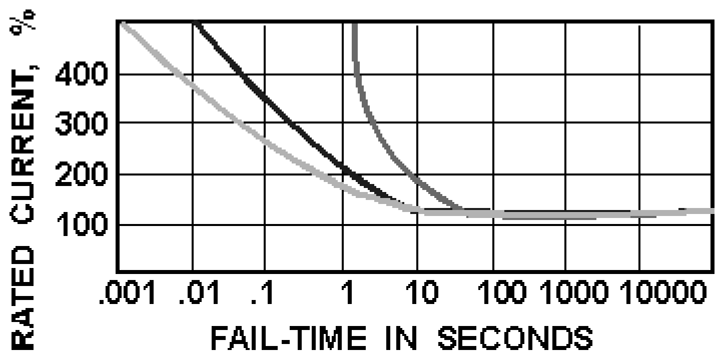
PLUG-TYPE FUSES are used in low-voltage, low-current circuits. This type fuse is rapidly being replaced by the circuit breaker.

CARTRIDGE FUSES are available in a wide range of physical sizes and voltage and current ratings. This type fuse is the most commonly used fuse.

The **CURRENT RATING** of a fuse is a value expressed in amperes that represents the amount of current the fuse will allow to flow without opening.

The **VOLTAGE RATING** of a fuse indicates the ability of the fuse to quickly extinguish the arc after the fuse element melts and the maximum voltage the open fuse will block.

The **TIME DELAY RATING** of a fuse indicates the relationship between the current through the fuse and the time it takes for the fuse to open. The three time delay ratings for fuses are DELAY, STANDARD, and FAST.

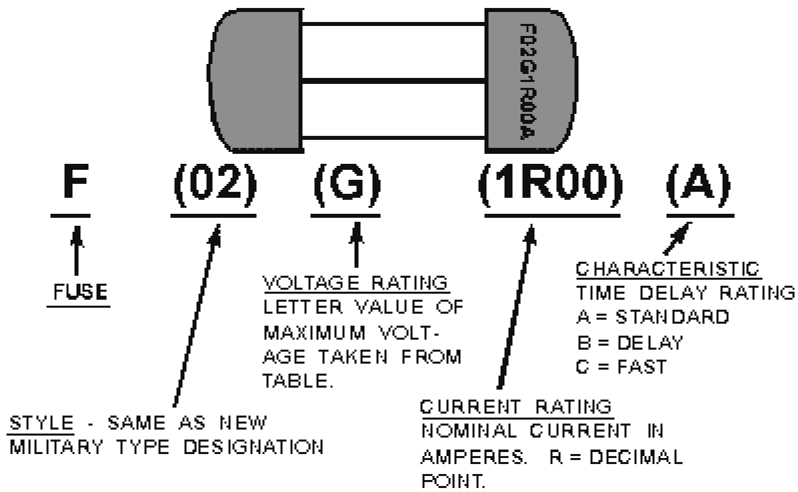


DELAY FUSES allow surge currents without opening. They are used to protect motors, solenoids, and transformers.

STANDARD FUSES have neither a time delay nor a fast acting characteristic. They are used in automobiles, lighting circuits and electrical power circuits.

FAST FUSES open very quickly with any current above the current rating of the fuse. They are used to protect delicate instruments or semiconductor devices.

The **OLD MILITARY FUSE DESIGNATION** is a system of fuse identification that uses coding to represent the current, voltage, and time-delay rating of the fuse. New fuses purchased by the Navy will no longer use this designation.



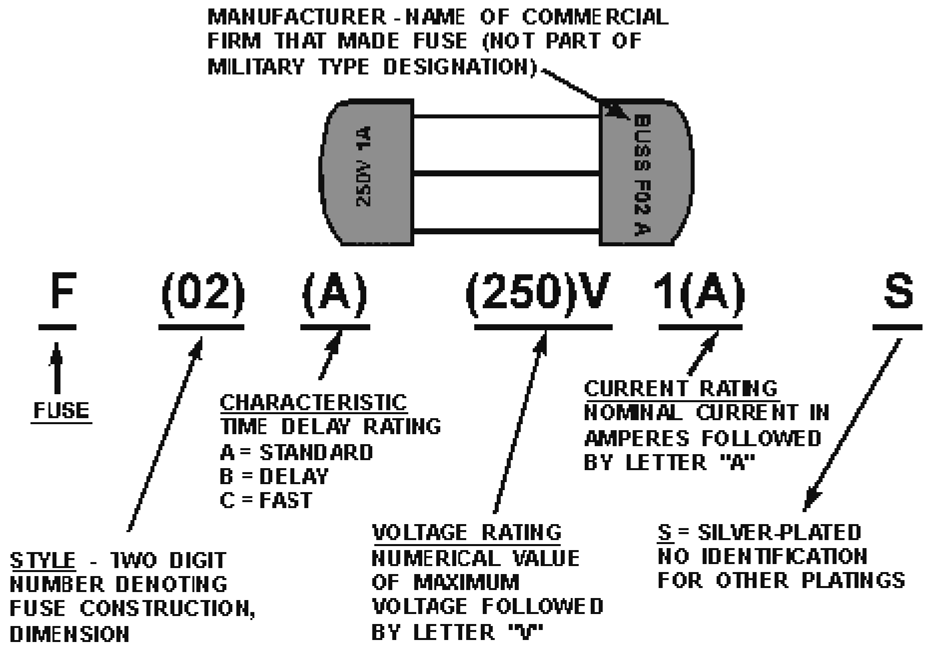
VOLTAGE CODE LETTER	VOLTAGE (VOLTS OR LESS)
A	32
B	52
C	90
D	125
G	250
H	500
J	1,000
K	2,500
N	5,000
P	10,000

VOLTAGE CODE (OLD)

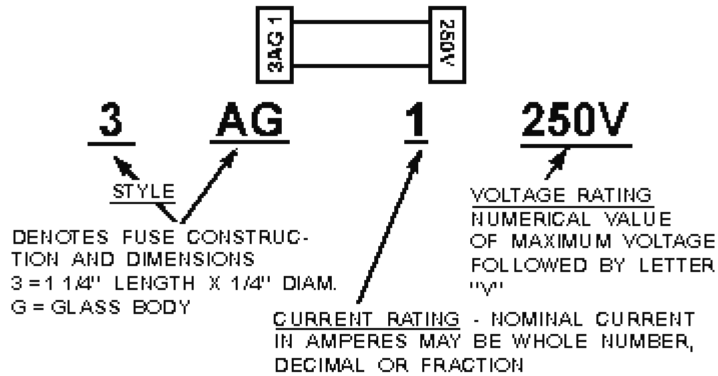
CURRENT CODE	CURRENT (AMPERES)
R002	.002 = 1/500
R005	.005 = 1/200
R010	.010 = 1/100
R031	.031 = 1/32
R750	.750 = 3/4
1R50	1.500 = 1 1/2

CURRENT CODE (OLD)

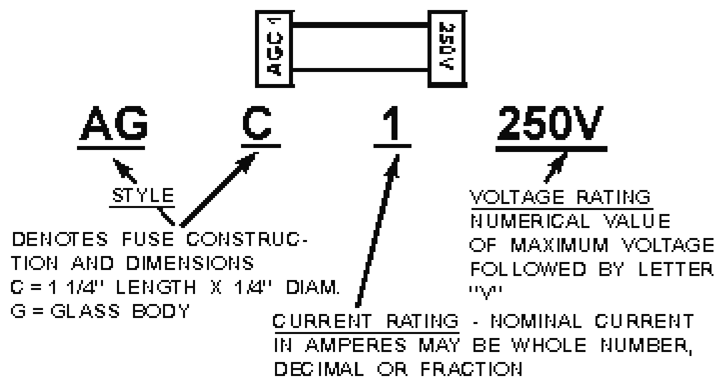
The **NEW MILITARY FUSE DESIGNATION** is the system used to identify fuses purchased by the Navy at the present time. The coding of current and voltage ratings has been replaced with direct printing of these ratings.



The **OLD COMMERCIAL FUSE DESIGNATION** was used by the fuse manufacturers to identify fuses. The current and voltage ratings are printed on the fuse, but the time delay rating is contained in the style coding of the fuse.



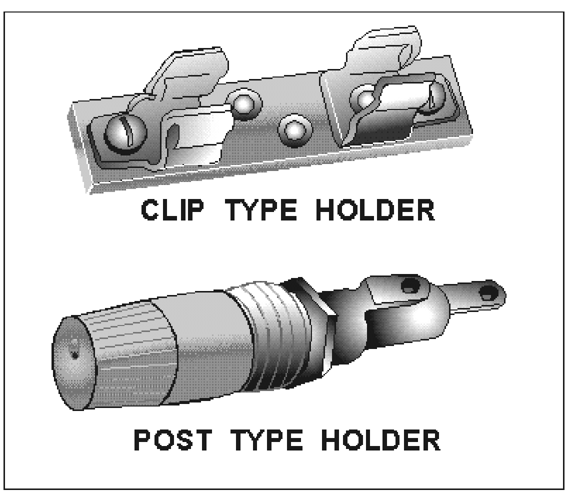
(A) OLD COMMERCIAL DESIGNATIONS



(B) NEW COMMERCIAL DESIGNATIONS

The **NEW COMMERCIAL FUSE DESIGNATION** is currently used by fuse manufacturers to identify fuses. It is similar to the old commercial fuse designation with the difference being in the style coding portion of the designation.

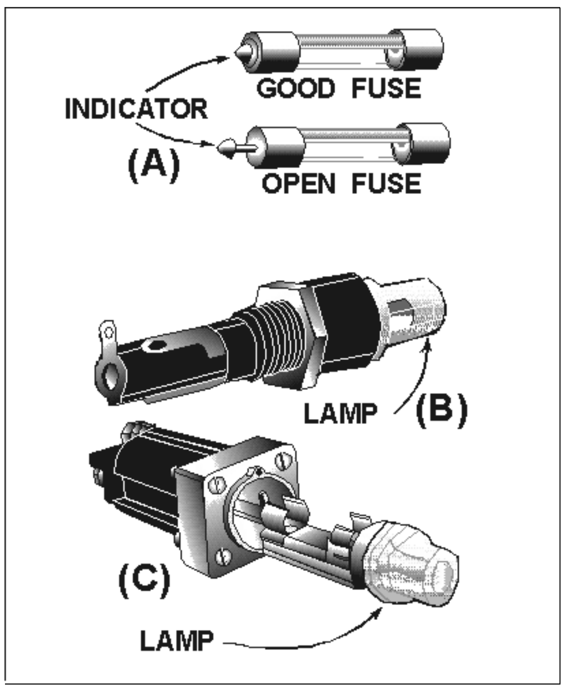
FUSEHOLDERS are used to allow easy replacement of fuses in a circuit.



The **CLIP-TYPE** has clips to connect the ferrules or knifeblades of the fuse to the circuit. The POST-TYPE is an enclosed fuseholder. The center connection of the post type should be connected to the power source and the outside connector should be connected to the load.

OPEN FUSES can be found by VISUAL INSPECTION, FUSE INDICATORS, or by the use of a METER. The following SAFETY PRECAUTIONS should be observed when checking a fuse:

- Turn the power off and discharge the circuit before removing a FUSE.
- Use a fusepuller when you remove a fuse from clip-type fuseholders.
- When you check a fuse with a voltmeter, be careful to avoid shocks and short circuits.
- When you use an ohmmeter to check fuses with low current ratings, be careful to avoid opening the fuse by excessive current.

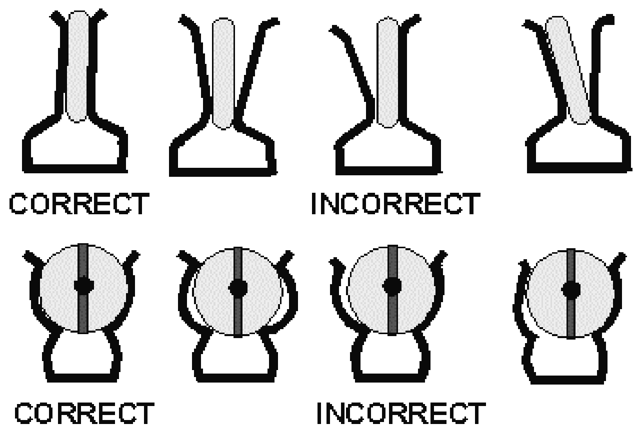


REPLACEMENT FUSES must be of the proper type. Check the technical manual parts list to find the identification of the proper fuse. If a substitute fuse must be used, the following guidelines should be followed:

- Never use a fuse with a higher current rating, a lower voltage rating, or a slower time delay rating than the specified fuse.
- The best substitution fuse is a fuse with the same current and time delay ratings and a higher voltage rating.

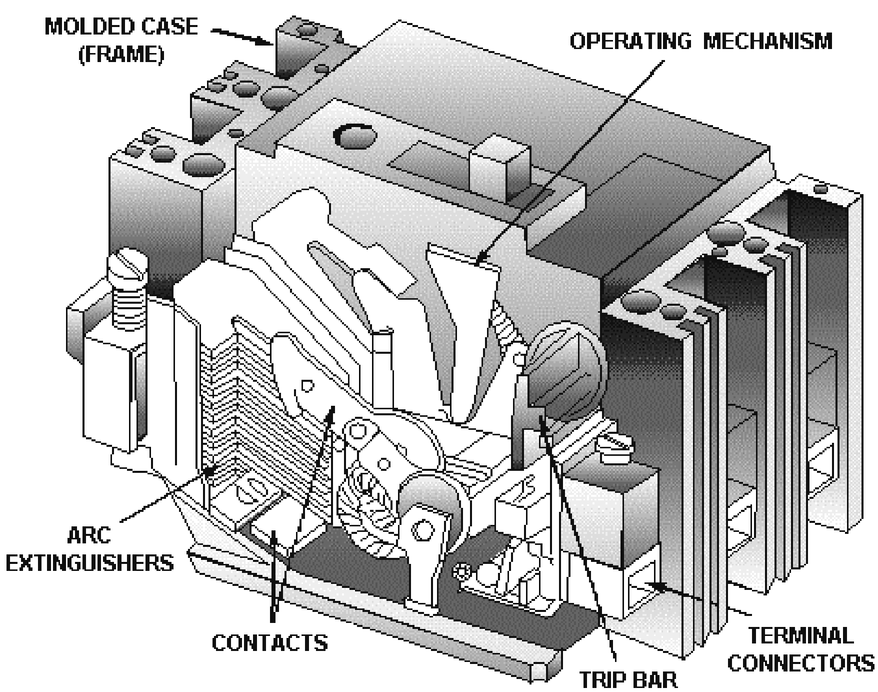
- If a lower current rating, or a lower time delay rating is used, the fuse may open under normal circuit conditions. Substitute fuses must have the same style (physical dimensions) as the specified fuse.

PROPER FIT between the fuse and fuseholder is essential. If the clips on clip-type fuseholders are sprung, the clips should be reformed, or clip clamps should be used. Any corrosion on fuses or fuseholders must be removed with fine sandpaper.

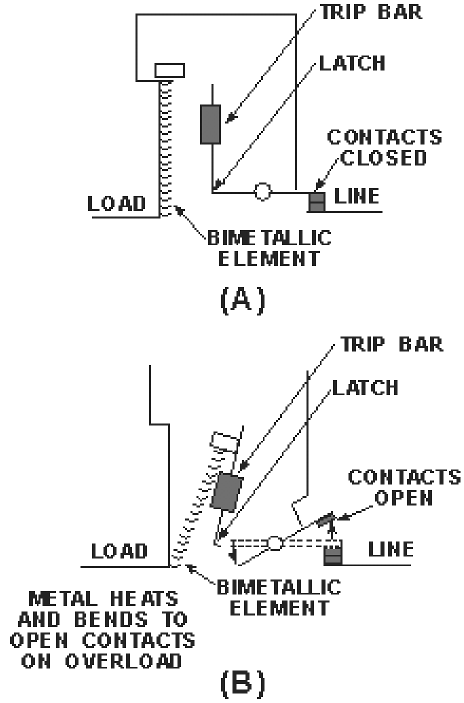


PREVENTIVE MAINTENANCE of fuses involves checking for the proper fuse, corrosion, proper fit, and open fuses; and correcting any discrepancies.

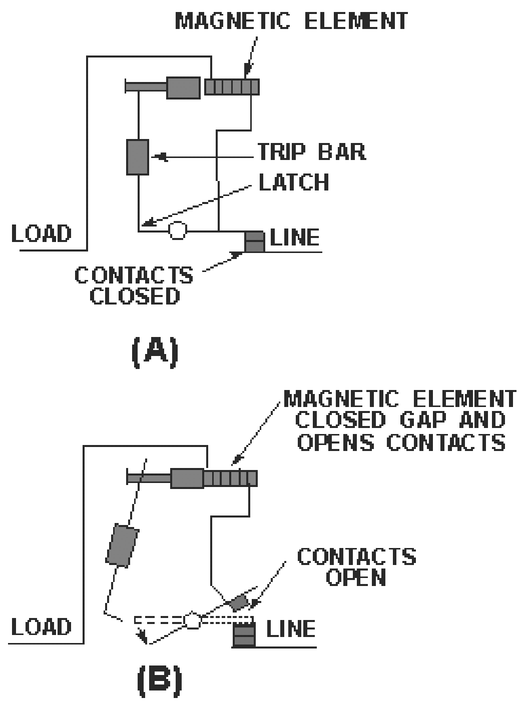
CIRCUIT BREAKERS have five main components: The frame, the operating mechanism, the arc extinguisher, the terminal connectors, and the trip element.



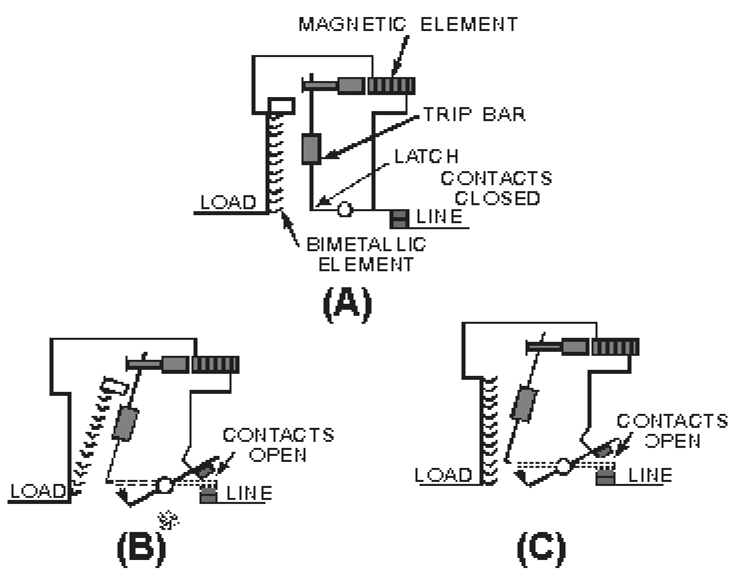
A **THERMAL TRIP ELEMENT** uses a bimetallic element that is heated by load current and bends due to this heating. If current (or temperature) increases above normal, the bimetallic element bends to push against a trip bar and opens the circuit.



A **MAGNETIC TRIP ELEMENT** uses an electromagnet in series with the load current to attract the trip bar and open the circuit if excessive current is present.



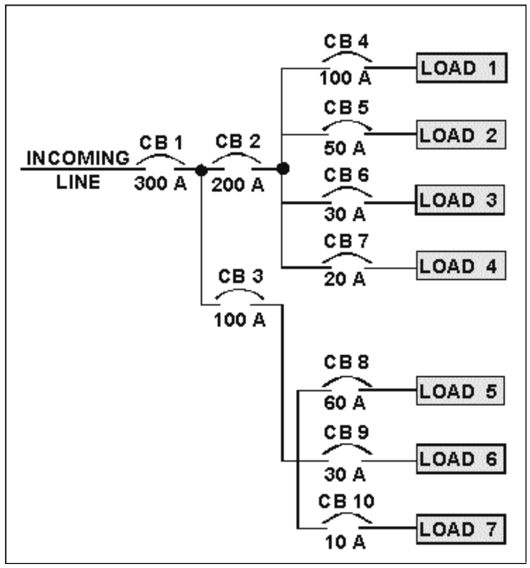
A **THERMAL-MAGNETIC TRIP ELEMENT** combines the thermal and magnetic trip elements into a single unit. A TRIP-FREE circuit breaker will trip (open) even if the operating mechanism is held in the ON position. A TRIP-FREE circuit breaker would be used on non-essential circuits.



A **NONTRIP-FREE** circuit breaker can be bypassed by holding the operating mechanism ON. A NONTRIP-FREE circuit breaker would be used for emergency or essential equipment circuits.

The **TIME DELAY RATINGS** of circuit breakers are INSTANTANEOUS, SHORT TIME DELAY, and LONG TIME DELAY.

SELECTIVE TRIPPING is used to cause the circuit breaker closest to the faulty circuit to trip, isolating the faulty circuit without affecting other nonfaulty circuits. This is accomplished by using an instantaneous circuit breaker close to the load, a short time delay circuit breaker at the next junction, and a long time delay circuit breaker at the main junction box.



The **FACTORS** used to select a circuit breaker are the power requirements of the circuit and the physical space available.

When **WORKING ON CIRCUIT BREAKERS**, the following items should be done BEFORE working on the circuit breaker: Check the applicable technical manual, obtain the approval of the electrical or engineering officer (for shipboard circuit breakers), remove power from the circuit breaker, and tag the switch that removes power from the circuit breaker. The following items should be checked and discrepancies corrected when working on circuit breakers: Check the operating mechanism for smooth operation, check the contacts for pitting, check the terminals for tightness and corrosion, check the mounting hardware for tightness and wear, check all components for wear, and check the entire circuit breaker for cleanliness.

ANSWERS TO QUESTIONS Q1. THROUGH Q43.

A1. To protect people and circuits from possible hazardous conditions.

A2. A direct short, excessive current, and excessive heat.

A3. A condition in which some point in the circuit where full system voltage is present comes in contact with the ground or return side of the circuit.

A4. A condition that is not a direct short but in which circuit current increases beyond the designed current carrying ability of the circuit.

A5. A condition in which the heat in or around the circuit increases to a higher than normal level

A6. In series, so total current will be stopped when the device opens.

A7. Fuses and circuit breakers.

A8.

- *a. circuit breaker*
- *b. fuse.*

A9.

- *a. cartridge*
- *b. plug*
- *c. plug*
- *d. cartridge.*

A10. A, C.

A11. Current, voltage, and time delay.

A12. The amount of current the fuse will allow without opening.

A13. The ability of the fuse to quickly extinguish the arc after the fuse element melts and the maximum voltage that cannot jump across the gap of the fuse after the fuse opens.

A14. Delay, standard, and fast.

A15. Delay-Motors, solenoids, or transformers. Standard-Automobiles, lighting or electrical power circuits. Fast-Delicate instruments or semiconductor devices.

A16.

- *a. 125 volts or less, 1.5 amperes, delay*
- *b. 250 volts or less, 1/8 ampere standard*

A17.

- *a. 125 volts or less, 1/16 ampere*
- *b. 250 volts or less, .15 ampere*

A18. F05B32V20A.

A19.

- *a. Post-type fuseholder*
- *b. Clip-type fuseholder*

A20.

a. Center connector

b. Outside connector

A21. Visual inspection, indicators, and using a meter.

A22. Put it back in the circuit. A good fuse will have zero ohms of resistance.

A23. The ohmmeter causes more than 1/500 ampere through the fuse when you check the fuse, thus it opens the fuse.

A24. Use a resistor in series with the fuse when you check it with the ohmmeter.

A25. Turn the power off and discharge the circuit before you remove fuses. Use a fuse puller (an insulated tool) when you remove fuses front clip-type fuse holders. When you check fuses with a voltmeter, be careful to avoid shocks and short circuits.

A26.

a. Not acceptable-wrong style

b. Substitute #3-smaller current rating

c. Substitute #1-identical, except higher voltage rating

d. Not acceptable-lower voltage rating

e. Direct replacement

f. Not acceptable-higher current rating

g. Substitute #2-Faster time delay rating

A27. Check for the proper type of replacement fuse and proper fit.

A28. Be sure the power is off in the circuit and the circuit is discharged before replacing a fuse. Use an identical replacement fuse if possible. Remove any corrosion from the fuseholders before replacing the fuses.

A29. Improper fuse, corrosion, improper fit, and open fuse.

A30. Frame, operating mechanism, arc extinguishers, terminal connectors, and trip element.

A31. Thermal, magnetic, and thermal-magnetic.

A32. The thermal trip element makes use of a bimetallic element that bends with an increase in temperature or current. The bending causes the trip bar to be moved releasing the latch.

A33. A circuit breaker that will trip even if the operating mechanism is held ON.

A34. A circuit breaker that can be overridden if the operating mechanism is held ON.

A35. In current sensitive or nonemergency systems.

A36. In emergency or essential circuits.

A37. Instantaneous, short time delay, and long time delay.

A38. It is the use of time delay ratings to cause the circuit breaker closest to the faulty circuit to trip. This isolates the faulty circuit without affecting other circuits.

A39. CB1-long time delay; CB2, CB3-short time delay; CB4 through CB10-instantaneous.

A40. The power requirements of the circuit and the physical space available.

A41. A push button or push-pull circuit breaker (small size, low power).

A42. Check the applicable technical manual, obtain the approval of the electrical or engineering officer (for shipboard circuit breakers), remove power from the circuit breaker, and tag the switch that supplies power to the circuit breaker.

A43. Check the operating mechanism for smooth operation, check the contacts for pitting, check the terminals for tightness and corrosion, check the mounting hardware for tightness and wear, check all components for wear, and check the entire circuit breaker for cleanliness.

CHAPTER 3

CIRCUIT CONTROL DEVICES

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. State three reasons circuit control devices are used and list three general types of circuit control devices.
2. Identify the schematic symbols for a switch, a solenoid, and a relay.
3. State the difference between a manual and an automatic switch and give an example of each.
4. State the reason multicontact switches are used.
5. Identify the schematic symbols for the following switches:
   - Single-pole, double-throw
   - Double-pole, single-throw
   - Double-pole, double-throw
   - Single-break
   - Double-break
   - Rotary
   - Wafer
6. State the characteristics of a switch described as a rocker switch.
7. State the possible number of positions for a single-pole, double-throw switch.
8. Identify a type of momentary switch.
9. State the type of switch used to prevent the accidental energizing or deenergizing of a circuit.
10. State the common name for an accurate snap-acting switch.
11. State the meaning of the current and voltage rating of a switch.
12. State the two types of meters you can use to check a switch.
13. Select the proper substitute switch from a list.
14. State the conditions checked for in preventive maintenance of switches.

15. State the operating principle and one example of a solenoid.

16. State the ways in which a solenoid can be checked for proper operation.

17. State the operating principle of a relay and how it differs from a solenoid.

18. State the two types of relays according to use.

19. State the ways in which a relay can be checked for proper operation and the procedure for servicing it.

CIRCUIT CONTROL DEVICES

Circuit control devices are used everywhere that electrical or electronic circuits are used. They are found in submarines, computers, aircraft, televisions, ships, space vehicles, medical instruments, and many other places. In this chapter you will learn what circuit control devices are, how they are used, and some of their characteristics.

INTRODUCTION

Electricity existed well before the beginning of recorded history. Lightning was a known and feared force to early man, but the practical uses of electricity were not recognized until the late 18th century. The early experimenters in electricity controlled power to their experiments by disconnecting a wire from a battery or by the use of a clutch between a generator and a steam engine. As practical uses were found for electricity, a convenient means for turning power on and off was needed.

Telegraph systems, tried as early as the late 1700s and perfected by Morse in the 1830s, used a mechanically operated contact lever for opening and closing the signal circuit. This was later replaced by the hand-operated contact lever or "key."

Early power switches were simple hinged beams, arranged to close or open a circuit. The blade-and-jaw knife switch with a wooden, slate, or porcelain base and an insulated handle, was developed a short time later. This was the beginning of circuit control devices.

Modern circuit control devices can change their resistance from a few milliohms (when closed) to well over 100,000 megaohms (when open) in a couple of milliseconds. In some circuit control devices, the movement necessary to cause the device to open or close is only .001 inch (.025 millimeters).

NEED FOR CIRCUIT CONTROL

Circuit control, in its simplest form, is the application and removal of power. This can also be expressed as turning a circuit on and off or opening and closing a circuit. Before you learn about the types of circuit control devices, you should know why circuit control is needed.

If a circuit develops problems that could damage the equipment or endanger personnel, it should be possible to remove the power from that circuit. The circuit protection devices discussed in the last chapter will remove power automatically if current or temperature increase enough to cause the circuit protection device to act. Even with this protection, a manual means of control is needed to allow you to remove power from the circuit before the protection device acts.

When you work on a circuit, you often need to remove power from it to connect test equipment or to remove and replace components. When you remove power from a circuit so that you can work on it, be sure to "tag out" the switch to ensure that power is not applied to the circuit while you are working. When work has been completed, power must be restored to the circuit. This will allow you to check the proper operation of the circuit and place it back in service. After the circuit has been checked for proper operation, remove the tag from the power switch.

Many electrical devices are used some of the time and not needed at other times. Circuit control devices allow you to turn the device on when it is needed and off when it is not needed.

Some devices, like multimeters or televisions, require the selection of a specific function or circuit. A circuit control device makes possible the selection of the particular circuit you wish to use.

TYPES OF CIRCUIT CONTROL DEVICES

Circuit control devices have many different shapes and sizes, but most circuit control devices are either SWITCHES, SOLENOIDS, or RELAYS.

Figure 3-1 shows an example of each of these types of circuit control devices and their schematic symbols.

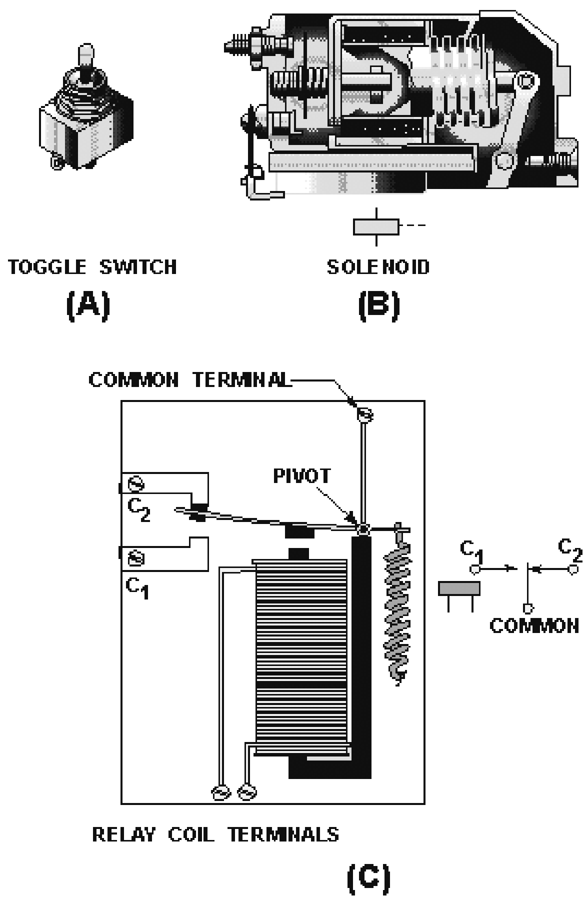


Figure 3-1.—Typical circuit control devices: RELAY COIL TERMINALS

Figure 3-1, view A, is a simple toggle switch and the schematic symbol for this switch is shown below it. Figure 3-1, view B, is a cutaway view of a solenoid. The schematic symbol below the solenoid is one of the schematic symbols used for this solenoid. Figure 3-1, view C, shows a simple relay. One of the schematic symbols for this relay is shown next to the relay.

Q1. What are three reasons circuit control is needed?

Q2. What are the three types of circuit control devices?

Q3. Label the schematic symbols shown in figure 3-2.

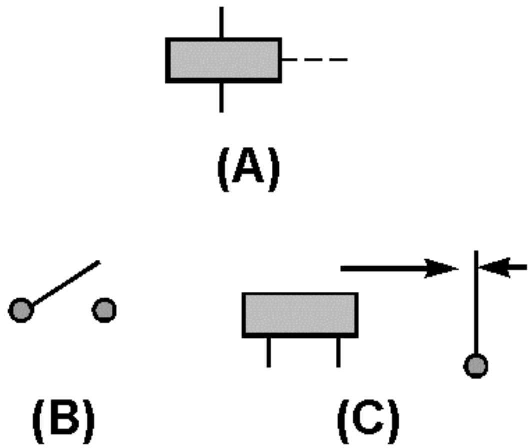


Figure 3-2.—Schematic symbol recognition.

SWITCH TYPES

There are thousands and thousands of switch applications found in home, industry, and the Navy. Hundreds of electrical switches work for you everyday to perform functions you take for granted. Some switches operate by the touch of a finger and many others are operated automatically.

Switches are used in the home to turn off the alarm clock, to control the stove, to turn on the refrigerator light, to turn on and control radios and televisions, hair dryers, dishwashers, garbage disposals, washers and dryers, as well as to control heating and air conditioning. A typical luxury automobile with power seats and windows might have as many as 45 switches.

Industry uses switches in a wide variety of ways. They are found in the business office on computers, copy machines, electric typewriters, and other equipment. A factory or shop may use thousands of switches and they are found on almost every piece of machinery. Switches are used on woodworking machinery, metal working machinery, conveyors, automation devices, elevators, hoists, and lift trucks.

The Navy uses switches in a number of ways. A typical aircraft could have over 250 switches to control lights, electronic systems, and to indicate whether the landing gear is up or down. Ships, fire control systems, and missile launchers are also controlled by electrical switches. In fact, almost all electrical or electronic devices will have at least one switch.

Switches are designed to work in many different environments from extreme high pressure, as in a submarine, to extreme low pressure, as in a spacecraft. Other environmental conditions to consider are high or low temperature, rapid temperature changes, humidity, liquid splashing or immersion, ice, corrosion, sand or dust, fungus, shock or vibration, and an explosive atmosphere.

It would not be possible to describe all the different switches used. This chapter will describe the most common types of switches.

MANUAL SWITCHES

A manual switch is a switch that is controlled by a person. In other words, a manual switch is a switch that you turn on or off. Examples of common manual switches are a light switch, the ignition switch on a motor vehicle, or the channel selector on a television. You may not think of the channel selector as a switch that you use to turn something on or off, but that is what it does. The channel selector is used to turn on the proper circuit and allows the television to receive the channel you have selected.

AUTOMATIC SWITCH

An automatic switch is a switch that is controlled by a mechanical or electrical device. You do not have to turn an automatic switch on or off. Two examples of automatic switches are a thermostat and the distributor in a motor vehicle. The thermostat will turn a furnace or air conditioner on or off by responding to the temperature in a room. The distributor electrically turns on the spark plug circuit at the proper time by responding to the mechanical rotation of a shaft. Even the switch that turns on the light in a refrigerator when the door is opened is an automatic switch.

Automatic switches are not always as simple as the examples given above. Limit switches, which sense some limit such as fluid level, mechanical movement, pressure (altitude or depth under water), or an electrical quantity, are automatic switches. Computers use and control automatic switches that are sometimes quite complicated.

Basically, any switch that will turn a circuit on or off without human action is an automatic switch.

MULTICONTACT SWITCHES

Switches are sometimes used to control more than one circuit or to select one of several possible circuits. An example of a switch controlling more than one circuit is the AM/FM selector on a radio. This switch enables you to control either the AM or FM portion of the radio with a single switch. An example of a switch that selects one of several circuits is the channel selector of a television set. These switches are called MULTICONTACT switches because they have more than one contact or MULTI(ple) CONTACTS.

Number of Poles and Number of Throws

Multicontact switches (other than rotary switches, which will be covered later) are usually classified by the number of POLES and number of THROWS. Poles are shown in schematics as those contacts through which current enters the switch; they are connected to the movable contacts. Each pole may be connected to another part of the circuit through the switch by "throwing" the switch (movable contacts) to another position. This action provides an individual conduction path through the switch for each pole connection. The number of THROWS indicates the number of different circuits that can be controlled by each pole. By counting the number of points where current enters the switch (from the schematic symbol or the switch itself), you can determine the number of poles. By counting the number of different points each pole can connect with, you can determine the number of throws.

Figure 3-3 will help you understand this concept by showing illustrations of various multicontact switches and their schematic symbols.

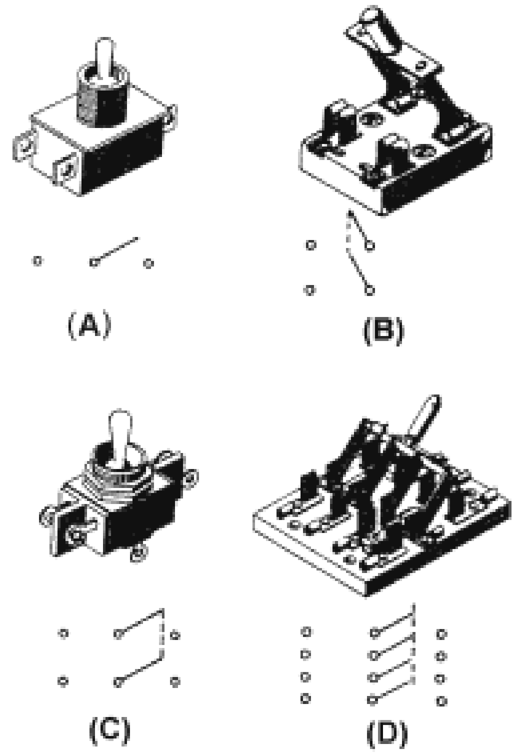


Figure 3-3.—Multicontact switches.

Figure 3-3(A) shows a single-pole, double-throw switch. The illustration shows three terminals (connections) on this switch. The schematic symbol for the switch is also shown.

The center connection of the schematic symbol represents the point at which current enters the switch. The left and right connections represent the two different points to which this current can be switched. From the schematic symbol, it is easy to determine that this is a single-pole, double-throw switch.

Now look at figure 3-3(B). The switch is shown with its schematic symbol. The schematic symbol has two points at which current can enter the switch, so this is a double-pole switch. Each of the poles is mechanically connected (still electrically separate) to one point, so this is a single-throw switch. Only one throw is required to route two separate circuit paths through the switch.

Figure 3-3(C) shows a double-pole, double-throw switch and its schematic symbol. Figure 3-3(D) shows a four-pole, double-throw switch and its schematic symbol.

It might help you to think of switches with more than one pole as several switches connected together mechanically. For example, the knife switch shown in figure 3-3(D) could be thought of as four single-pole, double-throw switches mechanically connected together.

Q4. What is the difference between a manual and an automatic switch?

Q5. What is one example of a manual switch?

Q6. What is one example of an automatic switch?

Q7. Why are multicontact switches used?

Q8. Label the schematic symbols shown in figure 3-4 as to number of poles and number of throws.

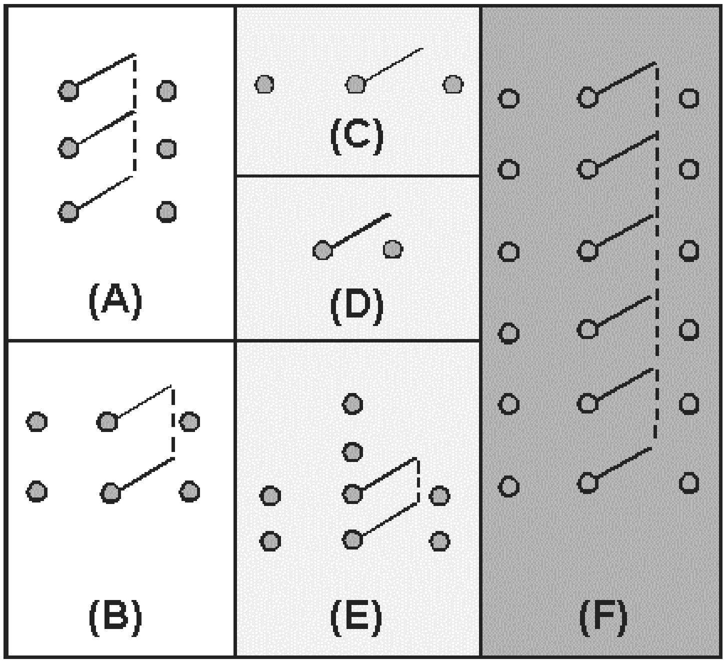


Figure 3-4.—Schematic symbols of switches.

Single-Break and Double-Break Switches

Switches can also be classified as SINGLE-BREAK or DOUBLE-BREAK switches. This refers to the number of places in which the switch opens or breaks the circuit. All of the switches shown so far have been single-break switches. A double-break switch is shown in figure 3-5. The schematic symbol shown in figure 3-5(A) shows that this switch breaks the circuit in two places (at both terminals). The upper part of the schematic symbol indicates that these contacts are in the open position and the circuit will close when the switch is acted upon (manually or automatically). The lower symbol shows closed contacts. These contacts will open the circuit when the switch is acted upon.

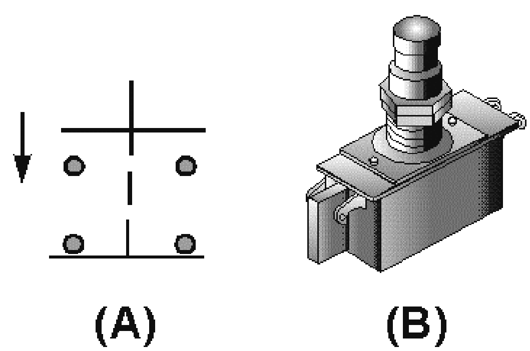


Figure 3-5.—Double-break pushbutton switch.

Figure 3-5(B) is a picture of the switch. This switch is called a pushbutton switch because it has a button that must be pushed to change the switch contact connections. Notice that the switch has four terminals. The schematic symbol in figure 3-5(A) shows that when one set of contacts is open, the other set of contacts is closed. This switch is a double-pole, single-throw, double-break switch.

The number of poles in a switch is independent of the number of throws and whether it is a single or double break switch. The number of throws in a switch is independent of the number of poles and whether it is a single or double break switch. In other words, each characteristic of a switch (poles, throws, break) is not determined by either of the other characteristics. Figure 3-6 shows the schematic symbols for several different switch configurations.

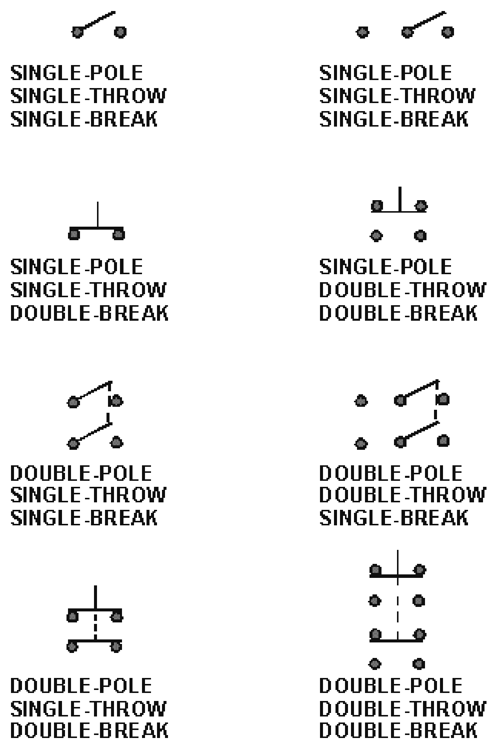


Figure 3-6.—Schematic symbols of switch configurations.

Rotary Switches

A rotary switch is a midcontact switch part of the schematic with the contacts arranged in a full or partial circle. Instead of a pushbutton or toggle, the mechanism used to select the contact moves in a circular motion and must be turned. Rotary switches can be manual or automatic switches. An automobile distributor, the ignition switch on a motor vehicle, and the channel selector on some television sets are rotary switches.

The automobile distributor cap and rotor are an example of the simplest form of an automatic rotary switch. Figure 3-7 shows a portion of an automobile ignition system with the distributor cap and rotor shown. The rotor is the portion of this switch that moves (rotates) and selects the circuit (spark plug). The rotor does not actually touch the contacts going to the spark plugs, but the signal (spark) jumps the gap between the rotor and the contacts. This switch has one input (the rotor) and six positions (one for each spark plug). The schematic diagram for this rotary switch is shown below the illustration of the distributor cap.

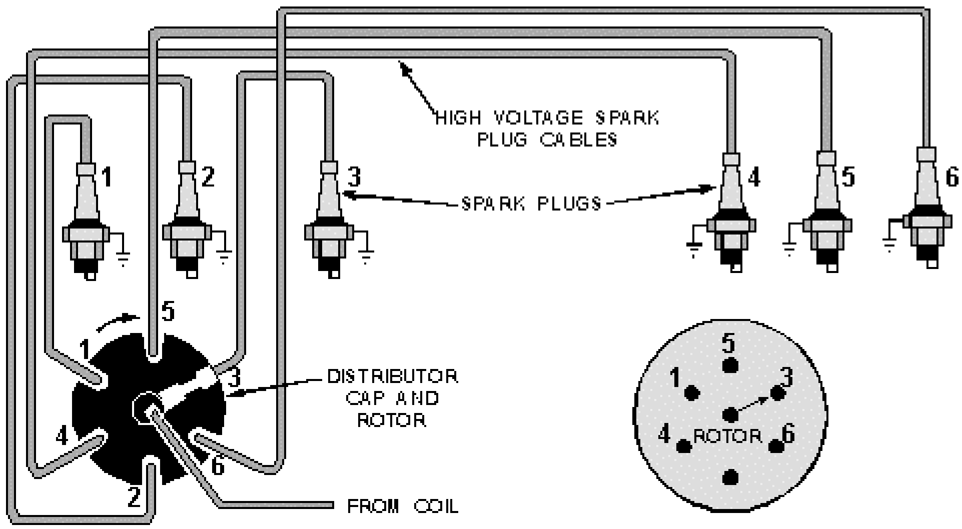


Figure 3-7.—Rotary switch in automobile ignition system.

The rotor in the distributor rotates continually (when in use) in one direction and makes a complete circle. This is not true for all rotary switches. The ignition switch in an automobile is also a rotary switch. It usually has four positions (accessory, off, on, start). Unlike the rotor, it does not rotate continually when in use, can be turned in either direction, and does not move through a complete circle.

Some rotary switches are made with several layers or levels. The arrangement makes possible the control of several circuits with a single switch. Figure 3-8 is an illustration of a rotary switch with two layers. Each layer has a selector and 20 contacts. As this switch is rotated, both layers select a single circuit (contact) of the 20.

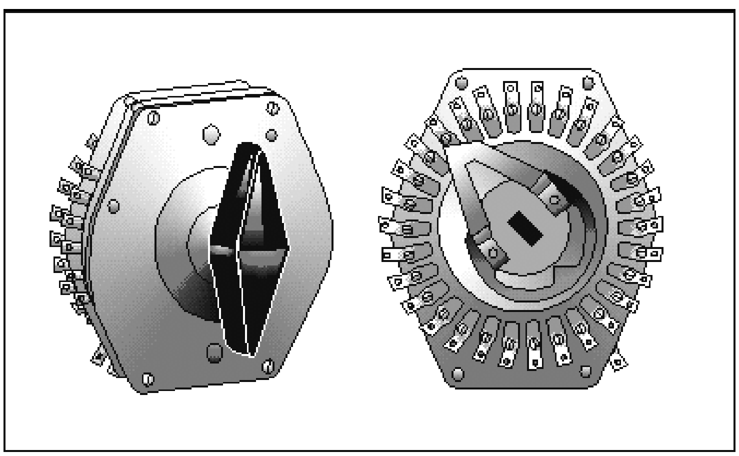


Figure 3-8.—Two-layer rotary switch.

The channel selector on some television sets is a multilayer rotary switch. It is also called a WAFER SWITCH. In a wafer switch, each layer is known as a wafer.

The schematic of the wafer is always drawn to represent the wafer as it would look if viewed from opposite the operating handle or mechanism. If the wafer has contacts on both sides, two drawings are used to show the two sides of the wafer. The two drawings are labeled "front" and "rear." The drawing labeled "front" represents the side of the wafer closest to the operating mechanism.

Figure 3-9(A) shows one wafer of a wafer switch and its schematic symbol. Contact 1 is the point at which current enters the wafer. It is always connected to the movable portion of the wafer. With the wafer in the position shown, contact 1 is connected to both contact 5 and 6 through the movable portion. If the movable portion was rotated slightly clockwise, contact 1 would only be connected to contact 5. This arrangement is known as MAKE BEFORE BREAK because the switch makes a contact before breaking the old contact.

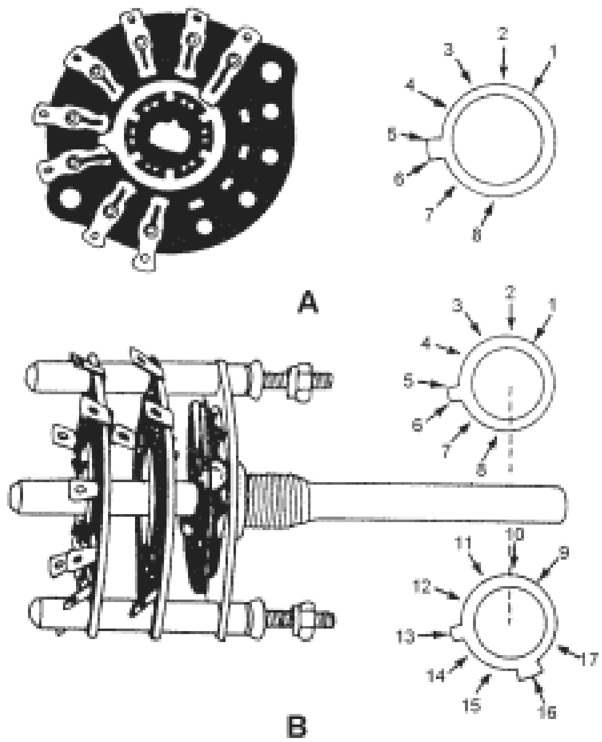


Figure 3-9.—Wafer switch.

Figure 3-9(B) is an illustration of the entire switch and its schematic symbol. Since the switch has two wafers mechanically connected by the shaft of the switch, the shaft rotates the movable portion of both wafers at the same time. This is represented on the schematic symbol by the dotted line connecting the two wafers.

The upper wafer of the schematic symbol is the wafer closest to the control mechanism, and is identical to the wafer shown in figure 3-9(A). When switches have more than one wafer, the first wafer shown is always the wafer closest to the operating mechanism. The lower wafer on the schematic diagram is the wafer farthest away from the operating mechanism. Contact 9 of this wafer is connected to the movable portion and is the point at which current enters the wafer. In the position shown, contact 9 is connected to both contact 13 and 16. If the switch is rotated slightly clockwise, contact 9 would no longer be connected to contact 13. A further clockwise movement would connect contact 9 to contact 12. This arrangement is called BREAK BEFORE MAKE. Contact 9 will also be connected to contact 15 at the same time as it is connected to contact 12.

Q9. Label the switch schematics shown in figure 3-10A through 3-10G.

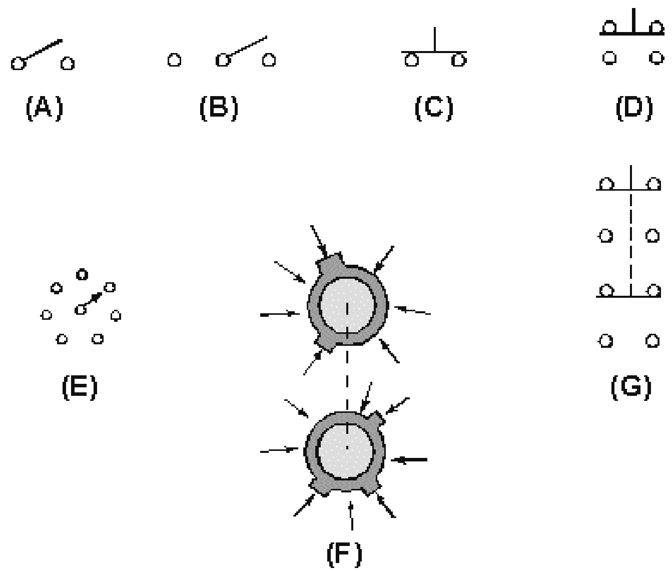


Figure 3-10.—Switch schematic symbols.

OTHER TYPES OF SWITCHES

You have learned that switches are classified by the number of poles, throws, and breaks. There are other factors used to describe a switch such as the type of actuator and the number of positions. In addition, switches are classified by whether the switch has momentary contacts or is locked into or out of position and whether or not the switch is snap-acting.

Type of Actuator

In addition to the pushbutton, toggle, and knife actuated switches already described, switches can have other actuators. There are rocker switches, paddle switches, keyboard switches and mercury switches (in which a small amount of mercury makes the electrical contact between two conductors).

Number of Positions

Switches are also classified by the number of positions of the actuating device. Figure 3-11 shows three toggle switches, the toggle positions, and schematic diagrams of the switch. Figure 3-11(A) is a single-pole, single-throw, two-position switch. The switch is marked to indicate the ON position (when the switch is closed) and the OFF position (when the switch is open). Figure 3-11(B) is a single-pole, double-throw, three-position switch. The switch markings show two ON positions and an OFF position. When this switch is OFF, no connection is made between any of the terminals. In either of the ON positions, the center terminal is connected to one of the outside terminals. (The outside terminals are not connected together in any position of the switch.) Figure 3-11(C) is a single-pole, double-throw, two-position switch. There is no OFF position. In either position of this switch, the center terminal is connected to one of the outside terminals.

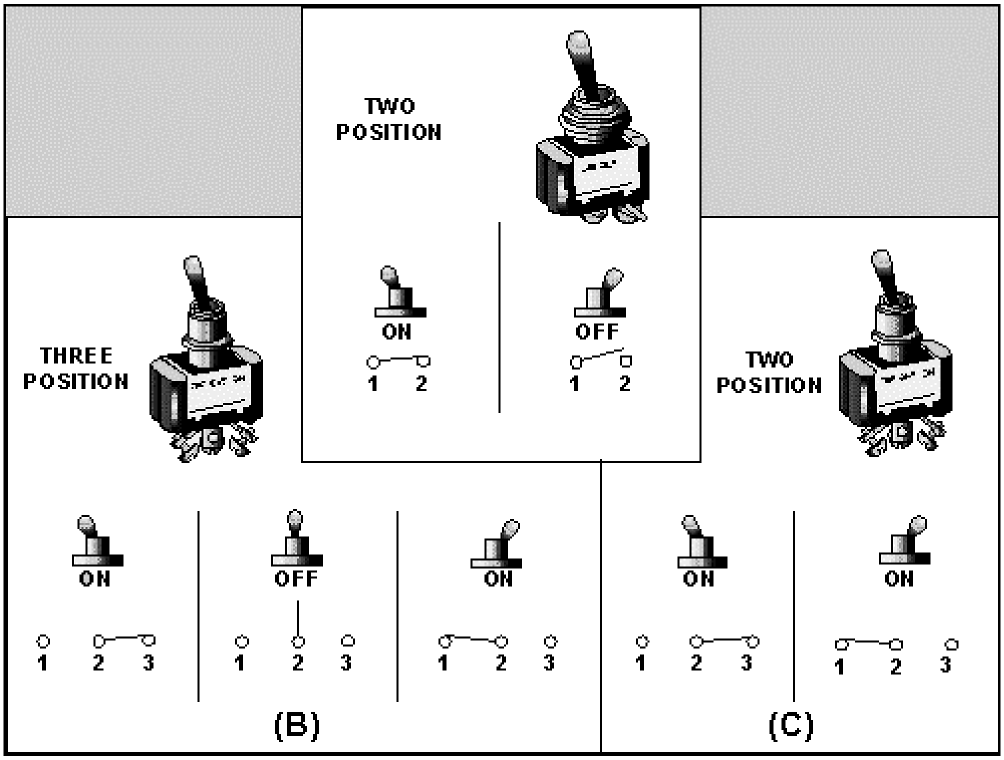


Figure 3-11.—Two- and three-position switches.

Momentary and Locked Position Switches

In some switches, one or more of the switch positions are MOMENTARY. This means that the switch will only remain in the momentary position as long as the actuator is held in that position. As soon as you let go of the actuator, the switch will return to a non-momentary position. The starter switch on an automobile is an example of a momentary switch. As soon as you release the switch, it no longer applies power to the starter.

Another type of switch can be LOCKED IN or OUT of some of the switch positions. This locking prevents the accidental movement of the switch. If a switch has locked-in positions, the switch cannot be moved from those positions accidentally (by the switch being bumped or mistaken for an unlocked switch). If the switch has locked-out positions, the switch cannot be moved into those positions accidentally. Figure 3-12 shows a three-position, locking switch.

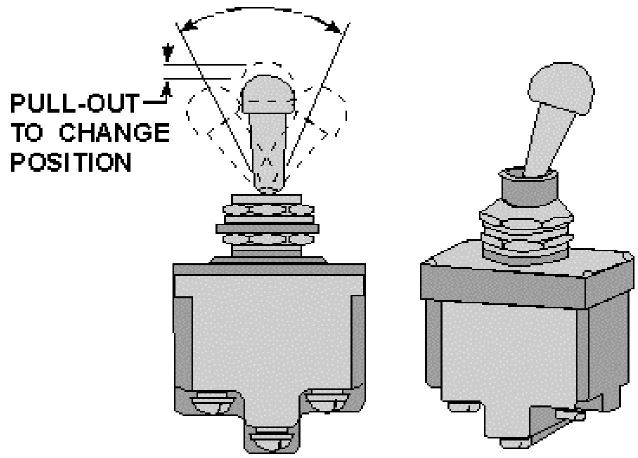


Figure 3-12.—Three-position locking switch.

Snap-Acting Switches

A SNAP-ACTING switch is a switch in which the movement of the switch mechanism (contacts) is relatively independent of the activating mechanism movement. In other words, in a toggle switch, no matter how fast or slow you move the toggle, the actual switching of the circuit takes place at a fixed speed. The snap-acting switch is constructed by making the switch mechanism a leaf spring so that it "snaps" between positions. A snap-acting switch will always be in one of the positions designed for that switch. The switch cannot be "between" positions. A two-position, single-pole, double-throw, snap-acting switch could not be left in an OFF position.

Accurate Snap-Acting Switches

An ACCURATE SNAP-ACTING SWITCH is a snap-acting switch in which the operating point is pre-set and very accurately known. The operating point is the point at which the plunger causes the switch to "switch." The accurate snap-acting switch is commonly called a MICROSWITCH. A microswitch is shown in figure 3-13.

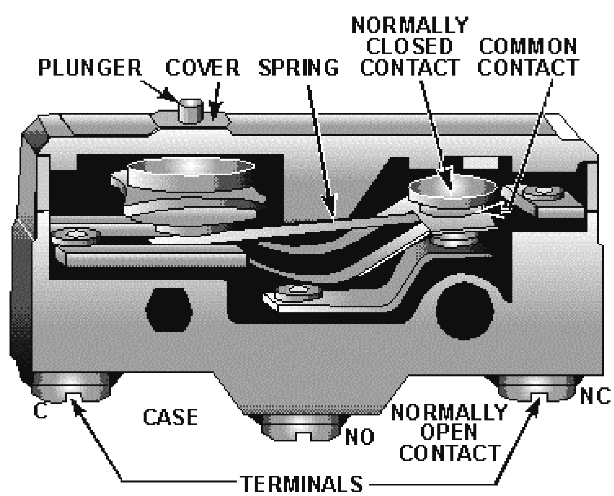


Figure 3-13.—Accurate snap-acting switch (microswitch).

The full description of the microswitch shown in figure 3-13 is a two-position, single-pole, double-throw, single-break, momentary-contact, accurate, snap-acting switch. Notice the terminals marked C, NO, and NC. These letters stand for common, normally open, and normally closed. The common terminal is connected to the normally closed terminal until the plunger is depressed. When the plunger is depressed, the spring will "snap" into the momentary position and the common terminal will be connected to the normally open terminal. As soon as the plunger is released, the spring will "snap" back to the original condition.

This basic accurate snap-acting switch is used in many applications as an automatic switch. Several different methods are used to actuate this type of switch. Some of the more common actuators and their uses are shown in figure 3-14.

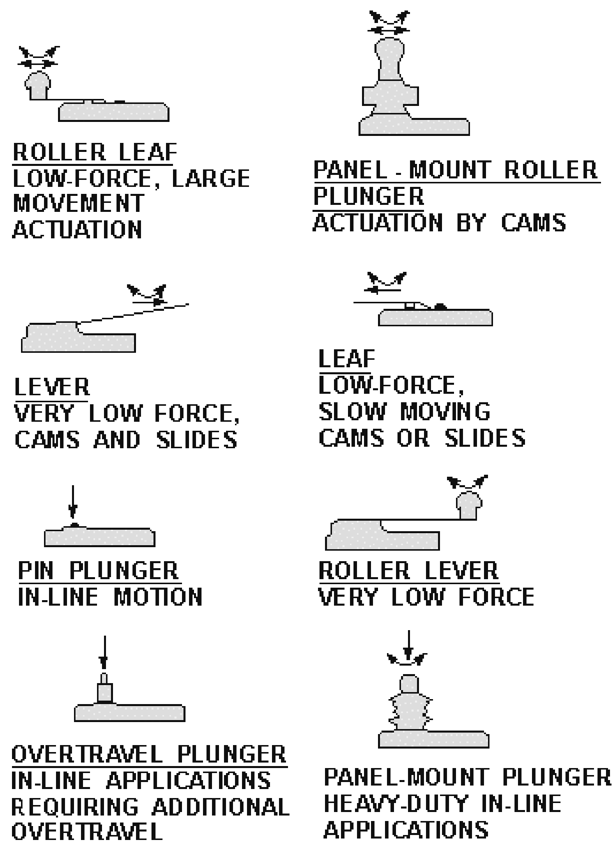


Figure 3-14.—Common actuators and their uses for accurate snap-acting switches.

Q10. What classification of a switch is used when you describe it as a rocker switch?

Q11. In describing a switch by the number of positions of the actuator, what are the two possible configurations for a single-pole, double-throw switch?

Q12. What type of switch should be used to control a circuit that requires a temporary actuation signal?

Q13. What type of switch is used if it is necessary to guard against a circuit being accidentally turned on or off?

Q14. What is the common name used for an accurate snap-acting switch?

SWITCH RATING

Switches are rated according to their electrical characteristics. The rating of a switch is determined by such factors as contact size, contact material, and contact spacing. There are two basic parts to a switch rating-the current and voltage rating. For example, a switch may be rated at 250 volts dc, 10 amperes. Some switches have more than one rating. For example, a single switch may be rated at 250 volts dc, 10 amperes; 500 volts ac, 10 amperes; and 28 volts dc, 20 amperes. This rating indicates a current rating that depends upon the voltage applied.

CURRENT RATING OF A SWITCH

The current rating of a switch refers to the maximum current the switch is designed to carry. This rating is dependent on the voltage of the circuit in which the switch is used. This is shown in the example given above. The current rating of a switch should never be exceeded. If the current rating of a switch is exceeded, the contacts may "weld" together making it impossible to open the circuit.

VOLTAGE RATING OF A SWITCH

The voltage rating of a switch refers to the maximum voltage allowable in the circuit in which the switch is used. The voltage rating may be given as an ac voltage, a dc voltage, or both. The voltage rating of a switch should never be exceeded. If a voltage higher than the voltage rating of the switch is applied to the switch, the voltage may be able to "jump" the open contacts of the switch. This would make it impossible to control the circuit in which the switch was used.

Q15. What is the current rating of a switch?

Q16. What is the voltage rating of a switch?

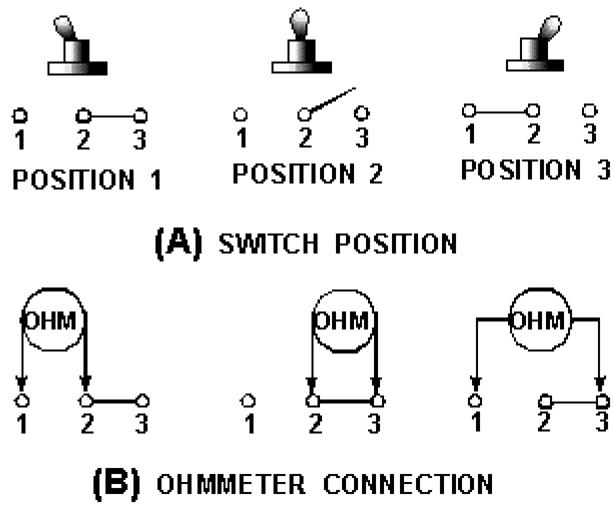
MAINTENANCE AND REPLACEMENT OF SWITCHES

Switches are usually a very reliable electrical component. This means, they don't fail very often. Most switches are designed to operate 100,000 times or more without failure if the voltage and current ratings are not exceeded. Even so, switches do fail. The following information will help you in maintaining and changing switches.

CHECKING SWITCHES

There are two basic methods used to check a switch. You can use an ohmmeter or a voltmeter. Each of these methods will be explained using a single-pole, double-throw, single-break, three-position, snap-acting, toggle switch.

Figure 3-15 is used to explain the method of using an ohmmeter to check a switch. Figure 3-15(A) shows the toggle positions and schematic diagrams for the three switch positions. Figure 3-15(B) shows the ohmmeter connections used to check the switch while the toggle is in position 1. Figure 3-15(C) is a table showing the switch position, ohmmeter connection, and correct ohmmeter reading for those conditions.



SWITCH POSITION	OHMMETER CONNECTION	CORRECT READING
1	1 - 2	∞
1	2 - 3	0
1	1 - 3	∞
2	1 - 2	∞
2	2 - 3	∞
2	1 - 3	∞
3	1 - 2	0
3	2 - 3	∞
3	1 - 3	∞

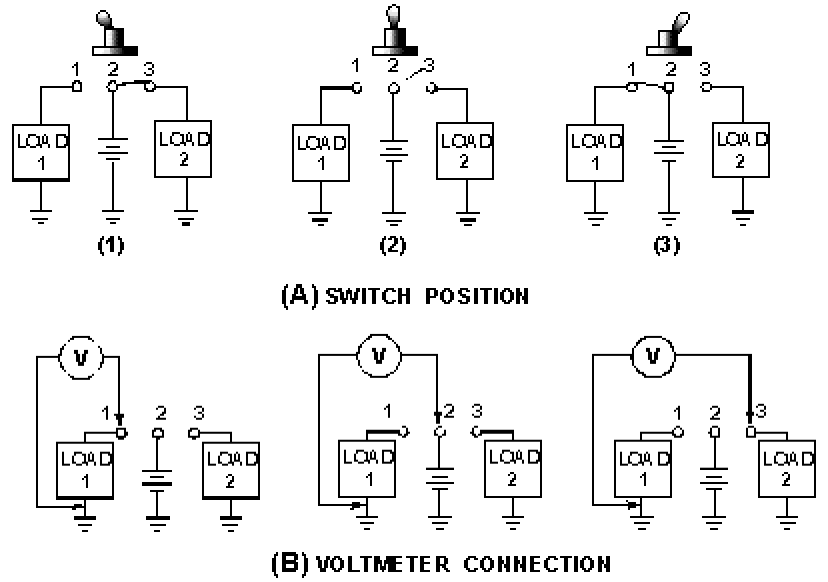
(C) TABLE OF CORRECT READINGS

Figure 3-15.—Table of correct readings.

With the switch in position 1 and the ohmmeter connected to terminals 1 and 2 of the switch, the ohmmeter should indicate (∞). When the ohmmeter is moved to terminals 2 and 3, the ohmmeter should indicate zero ohms. With the ohmmeter connected to terminals 1 and 3, the indication should be (∞).

As you remember from chapter 1, before the ohmmeter is used, power must be removed from the circuit and the component being checked should be isolated from the circuit. The best way to isolate the switch is to remove it from the circuit completely. This is not always practical, and it is sometimes necessary to check a switch while there is power applied to it. In these cases, you would not be able to use an ohmmeter to check the switch, but you can check the switch by the use of a voltmeter.

Figure 3-16(A) shows a switch connected between a power source (battery) and two loads. In figure 3-16(B), a voltmeter is shown connected between ground and each of the three switch terminals while the switch is in position 1. Figure 3-16(C) is a table showing the switch position, voltmeter connection, and the correct voltmeter reading.



SWITCH POSITION	VOLTMETER CONNECTION	CORRECT READING
1	1	0V
1	2	VOLTAGE
1	3	VOLTAGE
2	1	0V
2	2	VOLTAGE
2	3	0V
3	1	VOLTAGE
3	2	VOLTAGE
3	3	0V

(C) TABLE OF CORRECT READINGS

Figure 3-16.—Table of correct readings.

With the switch in position 1 and the voltmeter connected between ground and terminal 1, the voltmeter should indicate no voltage (OV). When the voltmeter is connected to terminal 2, the voltmeter should indicate the source voltage. With the voltmeter connected to terminal 3, the source voltage should also be indicated. The table in figure 3-16(C) will show you the correct readings with the switch in position 2 or 3.

REPLACEMENT OF SWITCHES

When a switch is faulty, it must be replaced. The technical manual for the equipment will specify the exact replacement switch. If it is necessary to use a substitute switch, the following guidelines should be used. The substitute switch must have all of the following characteristics.

- At least the same number of poles.
- At least the same number of throws.
- The same number of breaks.
- At least the same number of positions.

- The same configuration in regard to momentary or locked positions.
- A voltage rating equal to or higher than the original switch.
- A current rating equal to or higher than the original switch.
- A physical size compatible with the mounting.

In addition, the type of actuator (toggle, pushbutton, rocker, etc.) should be the same as the original switch. (This is desirable but not necessary. For example, a toggle switch could be used to replace a rocker switch if it were acceptable in all other ways.)

The number of poles and throws of a switch can be determined from markings on the switch itself. The switch case will be marked with a schematic diagram of the switch or letters such as SPST for single-pole, single-throw. The voltage and current ratings will also be marked on the switch. The number of breaks can be determined from the schematic marked on the switch or by counting the terminals after you have determined the number of poles and throws. The type of actuator, number of positions, the momentary and locked positions of the switch can all be determined by looking at the switch and switching it to all the positions.

PREVENTIVE MAINTENANCE OF SWITCHES

As already mentioned, switches do not fail very often. However, there is a need for preventive maintenance of switches. Periodically switches should be checked for corrosion at the terminals, smooth and correct operation, and physical damage. Any problems found should be corrected immediately. Most switches can be inspected visually for corrosion or damage. The operation of the switch may be checked by moving the actuator. When the actuator is moved, you can feel whether the switch operation is smooth or seems to have a great deal of friction. To check the actual switching, you can observe the operation of the equipment or check the switch with a meter.

Q17. What two types of meters can be used to check a switch?

Q18. If a switch must be checked with power applied, what type of meter is used?

Q19. A double-pole, double-throw, single-break, three-position, toggle switch is faulty. This switch has a momentary position 1 and is locked out opposition 3. The voltage and current ratings for the switch are 115 volt dc, 5 amperes. No direct replacement is available. From switches A through I, in table 3-1, indicate if the switch is acceptable or not acceptable as a substitute. Of the acceptable switches, rank them in order of choice. If the switch is unacceptable, give the reason.

Q20. What should you check when performing preventive maintenance on a switch?

Table 3-1.—Replacement Switches and Their Characteristics

	POLES	THROWS	BREAKS	NUMBER OF POSITIONS	MOMENTARY POSITIONS	LOCKED POSITIONS	ACTUATOR	RATING
A	2	1	1	2	—	—	PUSH BUTTON	115Vdc 5A
B	2	2	2	3	1	OUT-3	TOGGLE	150Vdc 5A
C	2	2	1	3	1	OUT-3	ROCKER	115Vdc 10A
D	1	2	1	3	1	OUT-3	TOGGLE	115Vdc 5A
E	2	2	1	3	—	OUT-3	ROCKER	150Vdc 10A
F	2	2	1	3	1	OUT-3	TOGGLE	150Vdc 10A
G	2	2	1	3	1	IN-3	TOGGLE	115Vdc 10A
H	2	2	1	3	1	OUT-3	ROCKER	115Vdc 3A
I	2	2	1	3	1	OUT-3	ROCKER	28Vdc 5A

SOLENOIDS

A SOLENOID is a control device that uses electromagnetism to convert electrical energy into mechanical motion. The movement of the solenoid may be used to close a set of electrical contacts, cause the movement of a mechanical device, or both at the same time.

Figure 3-17 is a cutaway view of a solenoid showing the solenoid action. A solenoid is an electromagnet formed by a conductor wound in a series of loops in the shape of a spiral. Inserted within this coil is a soft-iron core and a movable plunger. The soft-iron core is pinned or held in an immovable position. The movable plunger (also soft iron) is held away from the core by a spring when the solenoid is deenergized.

When current flows through the conductor, it produces a magnetic field. The magnetic flux produced by the coil results in establishing north and south poles in both the core and the plunger. The plunger is attracted along the lines of force to a position at the center of the coil. As shown in figure 3-17, the deenergized position of the plunger is partially out of the coil due to the action of the spring. When voltage is applied, the current through the coil produces a magnetic field. This magnetic field draws the plunger within the coil, resulting in mechanical motion. When the coil is deenergized, the plunger returns to its normal position because of spring action. The effective strength of the magnetic field on the plunger varies according to the distance between the plunger and the core. For short distances, the strength of the field is strong; and as distances increase, the strength of the field drops off quite rapidly.

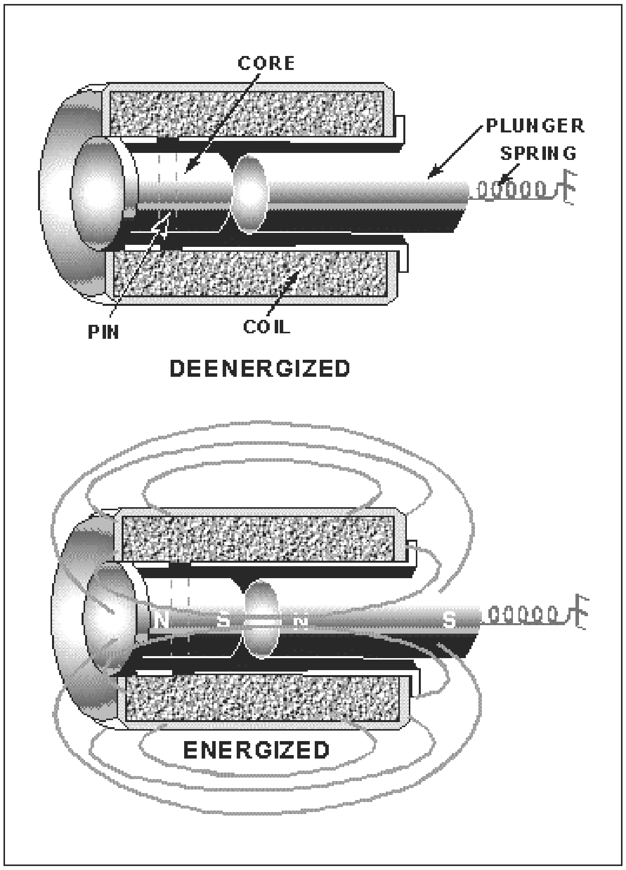


Figure 3-17.—Solenoid action.

While a solenoid is a control device, the solenoid itself is energized by some other control device such as a switch or a relay. One of the distinct advantages in the use of solenoids is that a mechanical movement can be accomplished at a considerable distance from the control device. The only link necessary between the control device and the solenoid is the electrical wiring for the coil current. The solenoid can have large contacts for the control of high current. Therefore, the solenoid also provides a means of controlling high current with a low current switch. For example, the ignition switch on an automobile controls the large current of a starter motor by the use of a solenoid. Figure 3-18 shows a cutaway view of a starter motor-solenoid combination and a section of the wiring for the solenoid. Notice that the solenoid provides all electrical contact for current to the starter motor as well as a mechanical movement of the shift lever.

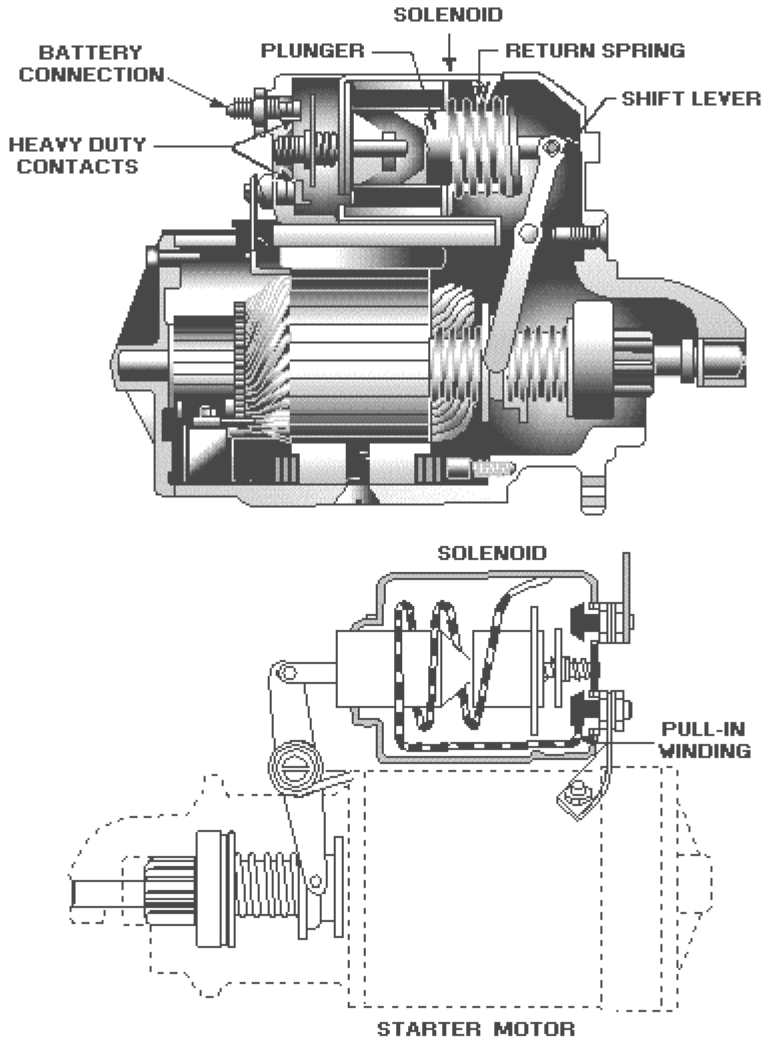


Figure 3-18.—Starter motor and solenoid.

MAINTENANCE OF SOLENOIDS

If you suspect that a solenoid is not working properly, the first step in troubleshooting it is a good visual inspection. Check the connections for poor soldering, loose connections, or broken wires. The plunger should be checked for cleanliness, binding, mechanical failure, and improper alignment. The mechanism that the solenoid is connected to (actuates) should also be checked for proper operation.

The second step is to check the energizing voltage with a voltmeter. If the voltage is too low, the result is less current flowing through the coil and a weak magnetic field. A weak magnetic field can result in slow or poor operation. Low voltage could also result in chatter or no operation at all. If the energizing voltage is too high, it could damage the solenoid by causing overheating or arcing. In either case, the voltage should be reset to the proper value so that further damage or failure of the solenoid will not result.

The solenoid coil should then be checked for opens, shorts, and proper resistance with an ohmmeter. If the solenoid coil is open, current cannot flow through it and the magnetic field is lost. A short results in fewer turns and higher current in the coil. The net result of a short is a weak magnetic field. A high-resistance coil will reduce coil current and also result in a weak magnetic field. A weak magnetic field will cause less attraction between the plunger and the core of the coil. This will result in improper operation similar to that caused by low voltage. If the coil is open, shorted, or has changed in resistance, the solenoid should be replaced.

Finally, you should check the solenoid to determine if the coil is shorted to ground. If a short to ground is found, the short should be removed to restore the solenoid to proper operation.

Q21. What is the operating principle of a solenoid?

Q22. What is one example of the use of a solenoid?

Q23. If a solenoid is not operating properly, what items should be checked?

RELAYS

The RELAY is a device that acts upon the same fundamental principle as the solenoid. The difference between a relay and a solenoid is that a relay does not have a movable core (plunger) while the solenoid does. Where multipole relays are used, several circuits may be controlled at once.

Relays are electrically operated control switches, and are classified according to their use as POWER RELAYS or CONTROL RELAYS. Power relays are called CONTACTORS; control relays are usually known simply as relays.

The function of a contactor is to use a relatively small amount of electrical power to control the switching of a large amount of power. The contactor permits you to control power at other locations in the equipment, and the heavy power cables need be run only through the power relay contacts.

Only lightweight control wires are connected from the control switches to the relay coil. Safety is also an important reason for using power relays, since high power circuits can be switched remotely without danger to the operator.

Control relays, as their name implies, are frequently used in the control of low power circuits or other relays, although they also have many other uses. In automatic relay circuits, a small electric signal may set off a chain reaction of successively acting relays, which then perform various functions.

In general, a relay consists of a magnetic core and its associated coil, contacts, springs, armature, and the mounting. Figure 3-19 illustrates the construction of a relay. When the coil is energized, the flow of current through the coil creates a strong magnetic field which pulls the armature downward to contact C1, completing the circuit from the common terminal to C1. At the same time, the circuit to contact C2, is opened.

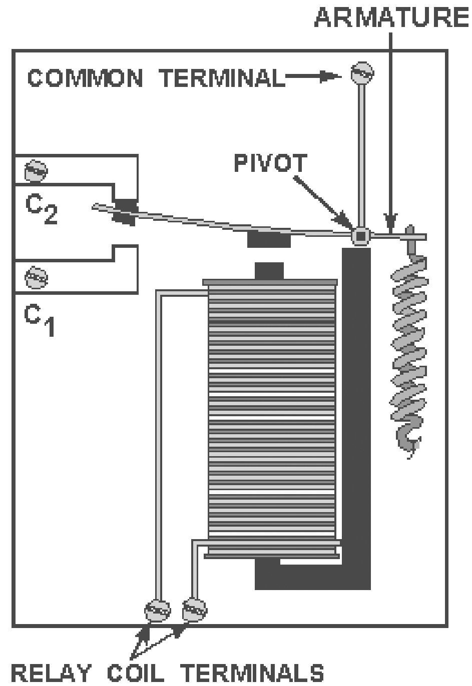


Figure 3-19.—Relay construction.

A relay can have many different types of contacts. The relay shown in figure 3-19 has contacts known as "break-make" contacts because they break one circuit and make another when the relay is energized. Figure 3-20 shows five different combinations of relay contacts and the names given to each.

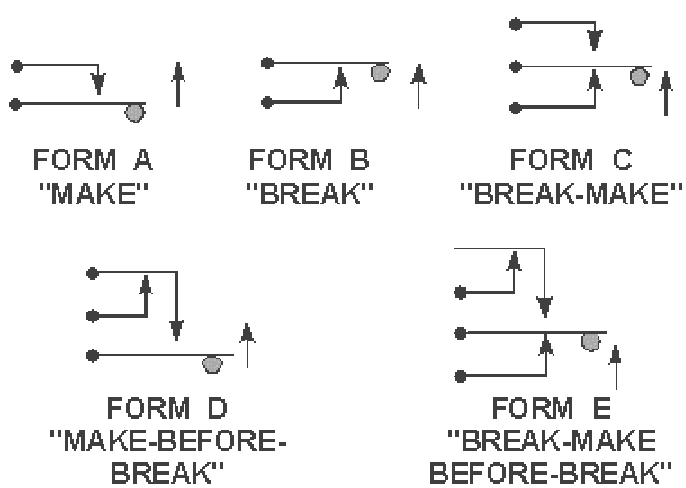


Figure 3-20.—Contact combinations.

A single relay can have several different types of contact combinations. Figure 3-21 is the contact arrangement on a single relay that has four different contact combinations. (The letters next to the contacts are the "forms" shown in figure 3-20.)

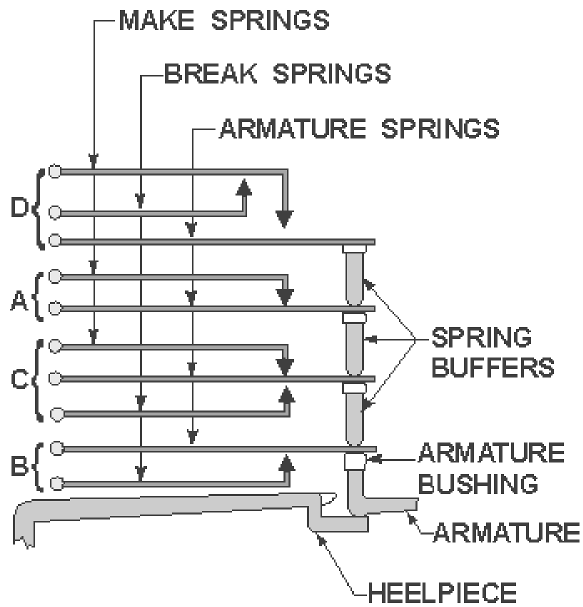


Figure 3-21.—Relay contact arrangement.

One type of relay with multiple sets of contacts is the clapper relay shown in figure 3-22. As the circuit is energized, the clapper is pulled to the magnetic coil. This physical movement of the armature of the clapper forces the pushrod and movable contacts upward. Any number of sets of contacts may be built onto the relay; thus, it is possible to control many different circuits at the same time. This type of relay can be a source of trouble because the motion of the clapper armature does not necessarily assure movement of all the movable contacts. Referring to figure 3-22, if the pushrod were broken, the clapper armature might push the lower movable contact upward but not move the upper movable contact.

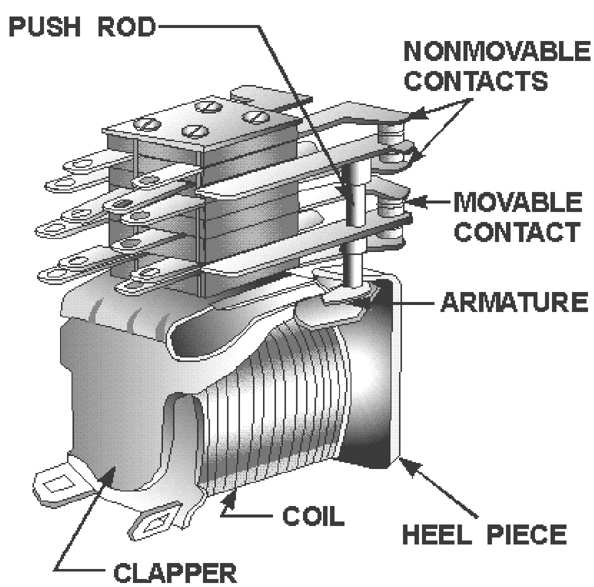


Figure 3-22.—Clapper-type relay.

Some equipment requires a "warm-up" period between the application of power and some other action. For example, vacuum tubes (covered later in this training series) require a delay between the application of filament power and high voltage. A time-delay relay will provide this required delay.

A thermal time-delay relay (fig. 3-23) is constructed to produce a delayed action when energized. Its operation depends on the thermal action of a bimetallic element similar to that used in a thermal circuit breaker. A heater is mounted around or near the element. The movable contact is mounted on the element itself. As the heat causes the element to bend (because of the different thermal expansion rates), the contacts close.

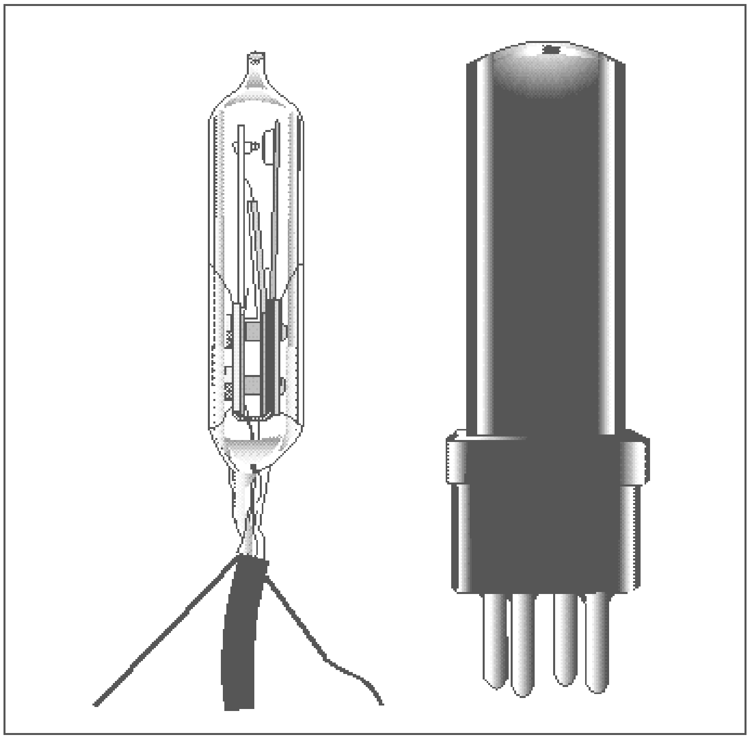


Figure 3-23.—A thermal time-delay relay.

Relays can be described by the method of packaging; open, semisealed, and sealed. Figure 3-24 shows several different relays and illustrates these three types of packaging.

Figure 3-24 (E), (G) and (H) are open relays. The mechanical motion of the contacts can be observed and the relays are easily available for maintenance. Figure 3-24 (A), (B) and (C) are semisealed relays. The covers provide protection from dust, moisture, and other foreign material but can be removed for maintenance.

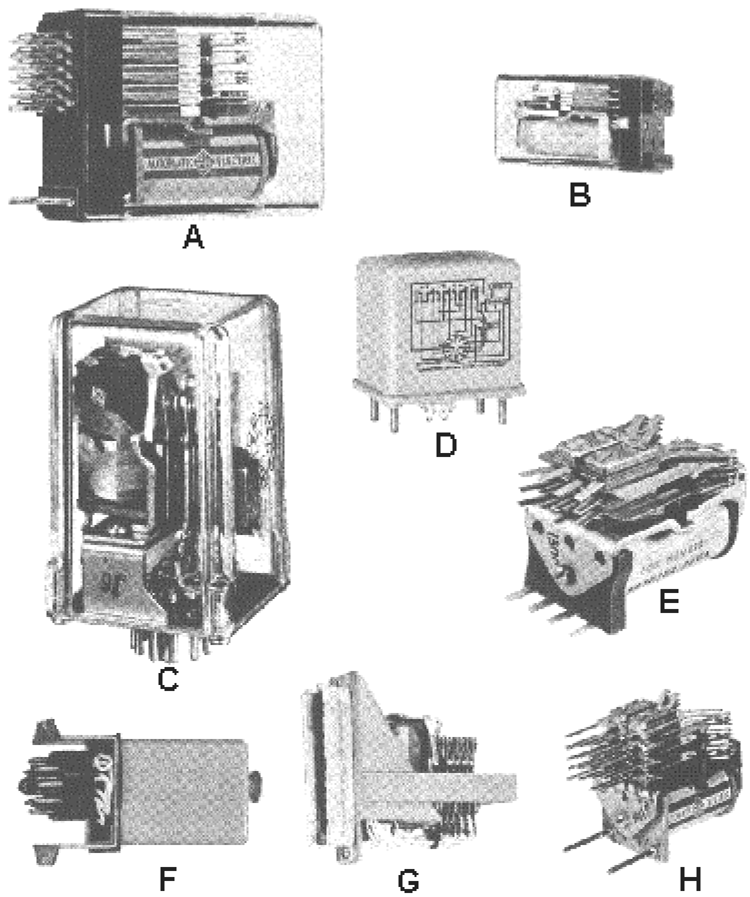


Figure 3-24.—Relay enclosures.

The clear plastic or glass covers provide a means of observing the operation of the relay without removal of the cover. Figure 3-24 (D) and (F) are examples of a hermetically sealed relay. These relays are protected from temperature or humidity changes as well as dust and other foreign material. Since the covers cannot be removed, the relays are also considered to be tamper-proof. With metal or other opaque covers, the operation of the relay can be "felt" by placing your finger on the cover and activating the relay.

Q24. What is the operating principle of a relay?

Q25. How does a relay differ from a solenoid?

Q26. What are the two classifications of relays?

MAINTENANCE OF RELAYS

The relay is one of the most dependable electromechanical devices in use, but like any other mechanical or electrical device, relays occasionally wear out or become inoperative. Should an inspection determine that a relay is defective, the relay should be removed immediately and replaced with another of the same type. You should be sure to obtain the same type relay as a replacement. Relays are rated in voltage, amperage, type of service, number of contacts, and similar characteristics.

Relay coils usually consist of a single coil. If a relay fails to operate, the coil should be tested for open circuit, short circuit, or short to ground. An open coil is a common cause of relay failure.

During preventive maintenance you should check for charred or burned insulation on the relay and for darkened or charred terminal leads. Both of these indicate overheating, and the likelihood of relay breakdown. One possible cause for overheating is that the power terminal connectors are not tight. This would allow arcing at the connection.

The build-up of film on the contact surfaces of a relay is another cause of relay trouble. Although film will form on the contacts by the action of atmospheric and other gases, grease film is responsible for a lot of contact trouble. Carbon build-up which is caused by the burning of a grease film or other substance (during arcing), also can be troublesome. Carbon forms rings on the contact surfaces and as the carbon rings build-up, the relay contacts are held open.

When current flows in one direction through a relay, a problem called "cone and crater" may be created at the contacts. The crater is formed by transfer of metal from one contact to the other contact, the deposit being in the shape of a cone. This condition is shown in figure 3-25(A).

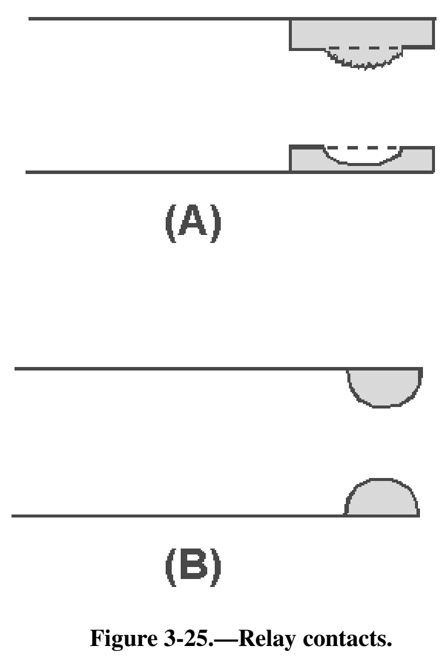


Figure 3-25.—Relay contacts.

Some relays are equipped with ball-shaped contacts and, in many applications, this type of contact is considered superior to a flat surface. Figure 3-25(B) shows a set of ball-shaped contacts. Dust or other substances are not as readily deposited on a ball-shaped surface. In addition, a ball-shaped contact penetrates film more easily than a flat contact. When you clean or service ball-shaped relay contacts, be careful to avoid flattening or otherwise altering the rounded surfaces of the contacts, YOU could damage a relay if you used sandpaper or emery cloth to clean the contacts. Only a burnishing tool, shown in figure 3-26 should be used for this purpose.

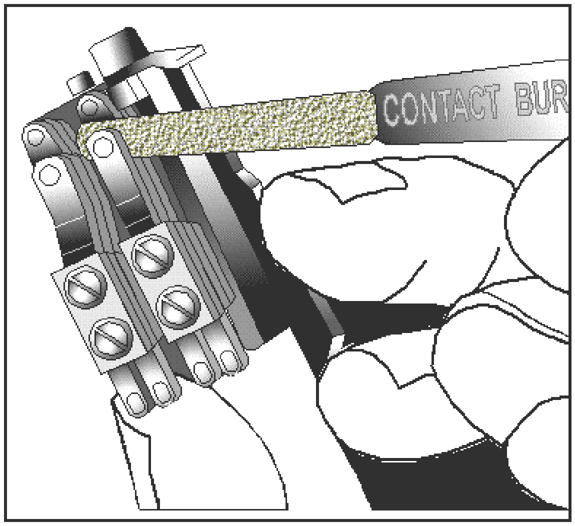


Figure 3-26.—Burnishing tool.

You should not touch the surfaces of the burnishing tool that are used to clean the relay contacts. After the burnishing, tool is used, it should be cleaned with alcohol.

Contact clearances or gap settings must be maintained in accordance with the operational specifications of the relay. When a relay has bent contacts, you should use a point bender (shown in figure 3-27) to straighten the contacts. The use of any other tool could cause further damage and the entire relay would have to be replaced.

Figure 3-27.—Point bender.

Cleanliness must be emphasized in the removal and replacement of covers on semi sealed relays. The entry of dust or other foreign material can cause poor contact connection. When the relay is installed in a position where there is a possibility of contact with explosive fumes, extra care should be taken with the cover gasket. Any damage to, or incorrect seating of the gasket increases the possibility of igniting the vapors.

Q27. How can you determine if a relay is operating (changing from one position to the other)?

Q28. What items should be checked on a relay that is not operating properly?

Q29. What is used to clean the contacts of a relay?

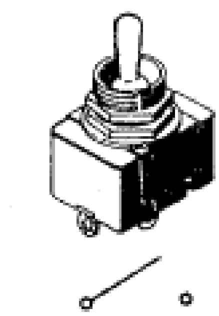
Q30. What tool is used to set contact clearances on a relay?

SUMMARY

This chapter has provided you with basic information on circuit control devices. The following is a summary of the main points in this chapter.

CIRCUIT CONTROL DEVICES are used to apply or remove power and to select a function or circuit within a device.

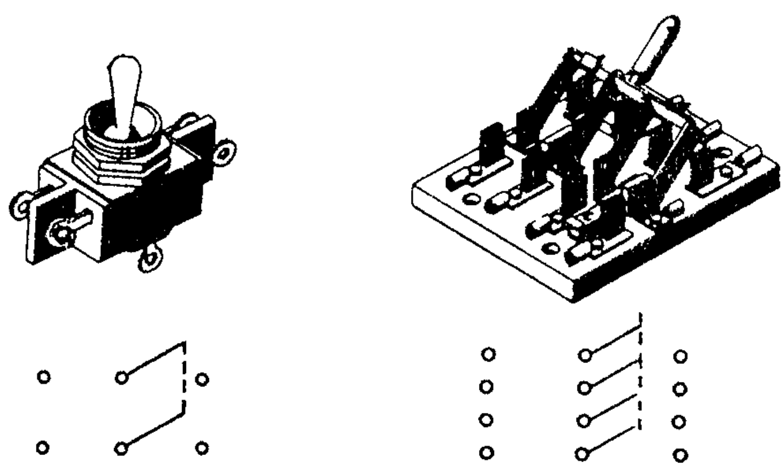
A **SWITCH** is one type of circuit control device. Switches are classified in many different ways.



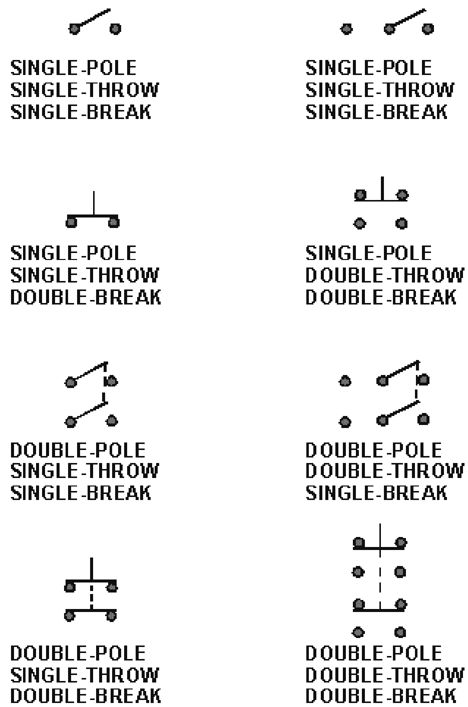
TOGGLE SWITCH

A **MANUAL SWITCH** must be tuned ON or OFF by a person. An AUTOMATIC SWITCH will turn a circuit ON or OFF without the action of a person by using mechanical or electrical devices.

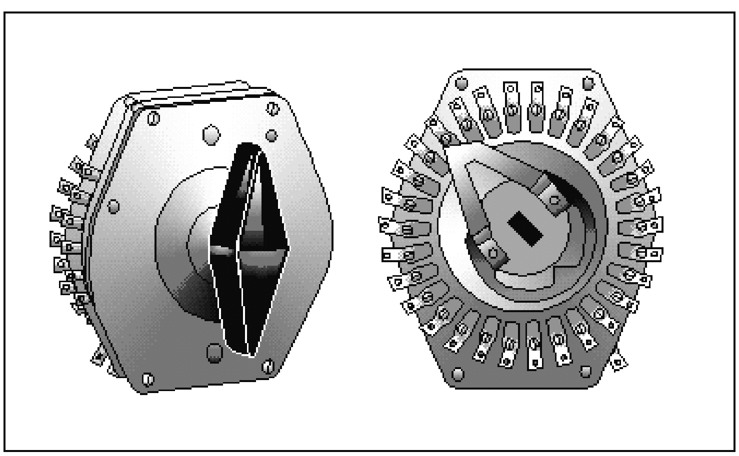
MULTICONTACT SWITCHES make possible the control of more than one circuit or the selection of one of several possible circuits with a single switch.



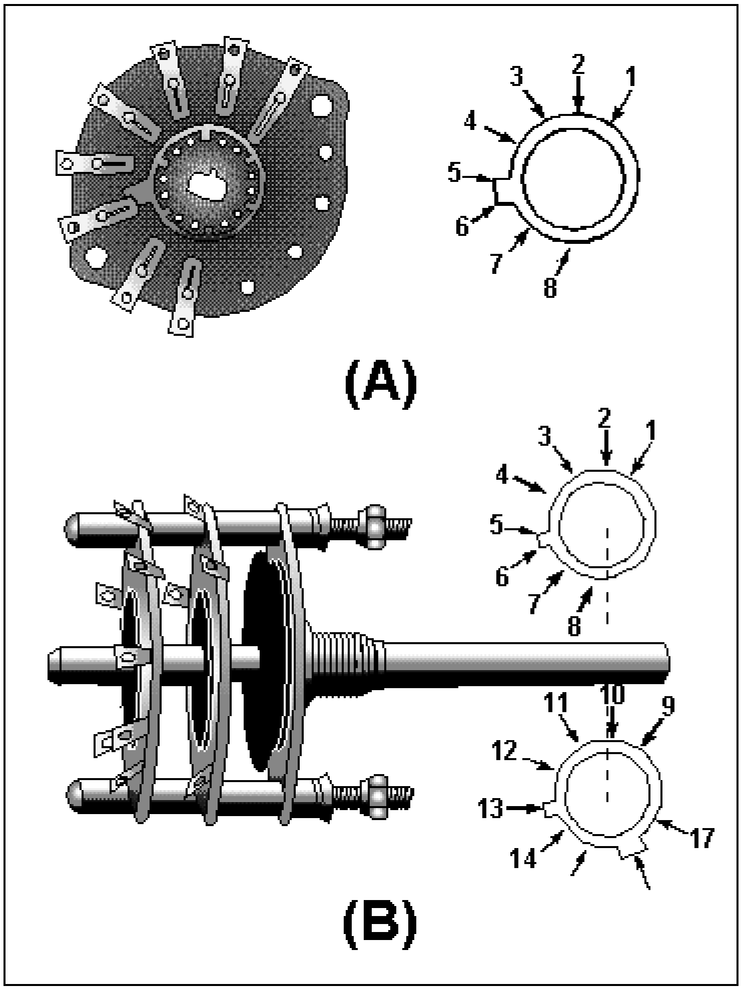
The **POLES** of a switch are the points at which current can enter the switch. The number of THROWS is the number of possible circuits that can be connected to each pole. The number of BREAKS is the number of points at which the switch breaks the circuit.



A **ROTARY SWITCH** is a multicontact switch with contacts arranged in a circular or semicircular manner.

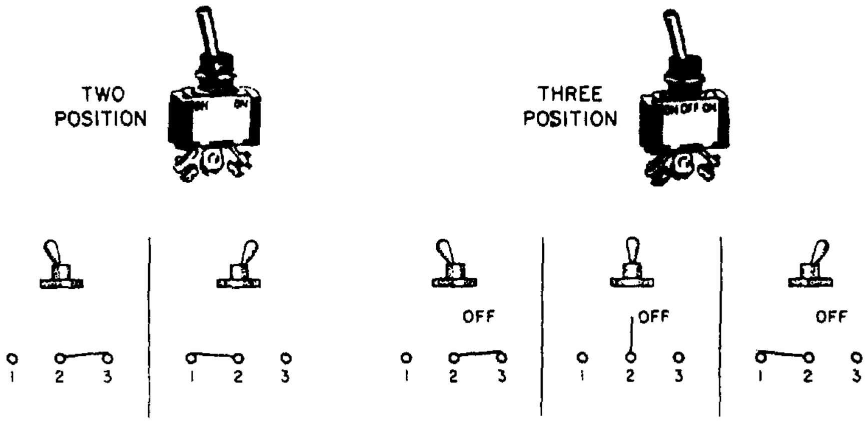


A **WAFER SWITCH** is a rotary switch in which the contacts are on wafers. The wafers are mechanically connected by the shaft of the switch.



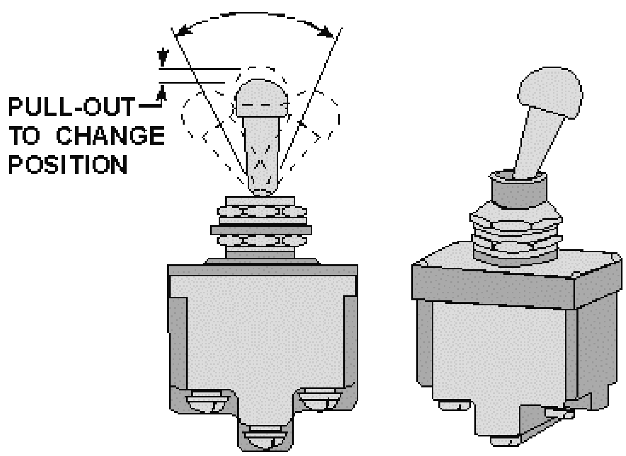
The **ACTUATOR** of a switch is the portion of the switch which is moved to cause the switch to change contact positions. The actuator could be a toggle, a pushbutton, a rocker, or, in the case of a rotary switch, a shaft and handle.

The **NUMBER OF POSITIONS** of a switch refers to the number of points at which the actuator can select a contact configuration.



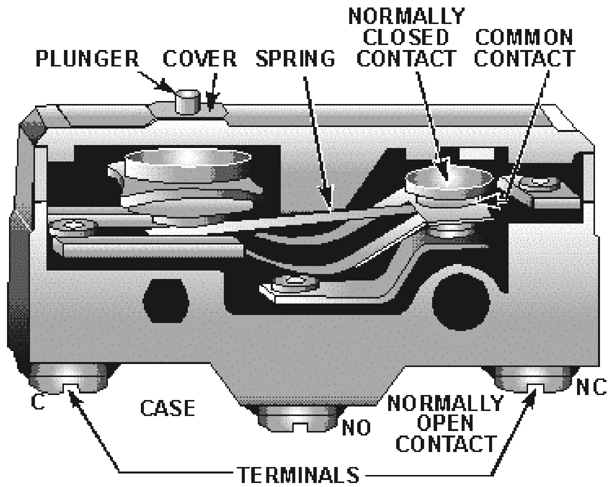
A **MOMENTARY POSITION** of a switch is one in which the actuator will only stay as long as force is applied to the actuator. When the force is removed, the actuator (and switch) will return to a non-momentary position.

A **LOCKED POSITION** of a switch is used to prevent the accidental movement of the actuator to or from a specific position.



A **SNAP-ACTING SWITCH** is one in which the movement of the switch contacts is relatively independent of the actuator movement. This is accomplished by using a leaf spring for the common contact of the switch.

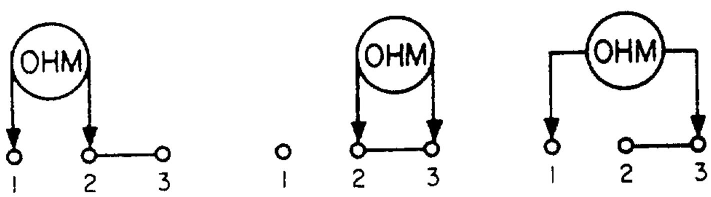
A **MICROSWITCH** is an accurate snap-acting switch and the operating point is preset and very accurately known.

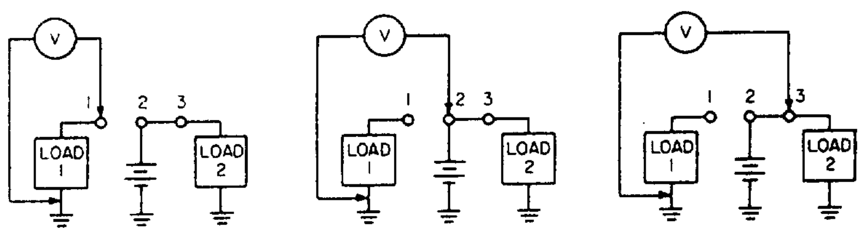


The **VOLTAGE RATING** of a switch is the maximum voltage the switch is designed to control. A voltage higher than the voltage rating may be able to "jump" the open contacts of the switch.

The **CURRENT RATING** of a switch is the maximum current the switch is designed to carry; it is dependent on the voltage rating. Any current higher than the current rating may cause the contacts of the switch to melt and "weld" together.

The contacts of a switch can be checked with an ohmmeter if power is removed or with a voltmeter if power is applied to the switch. To check a switch, the actuator should be checked for smooth and correct operation, the terminals should be checked for evidence of corrosion, and the physical condition of the switch should be determined. If a substitute switch must be used to replace a faulty switch, the substitute must have all of the following:

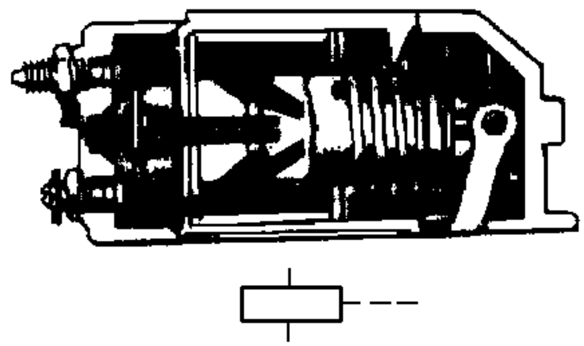




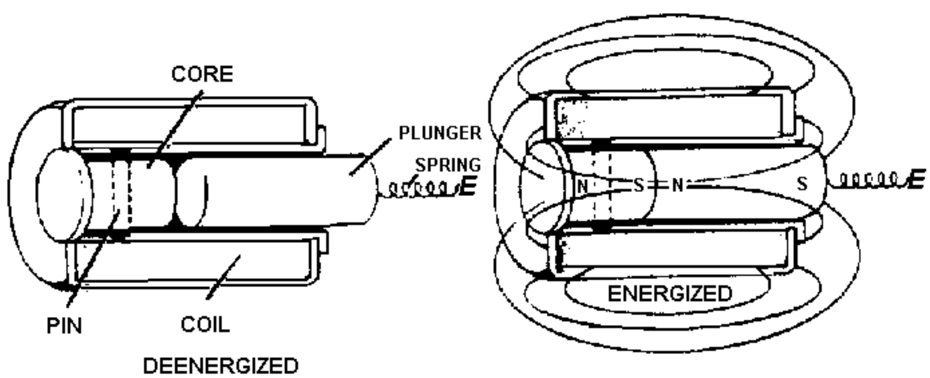
At least the same number of poles, throws, and positions; the same number of breaks and an identical configuration in regard to momentary and locked positions; and a voltage and current rating equal to or higher than the original switch. In addition, the substitute must be of a physical size compatible with the mounting, and must have the same type actuator as the original switch.

A **SOLENOID** is a control device that uses electromagnetism to convert electrical energy into a mechanical motion. The magnetic field of the coil and core will attract the plunger of a solenoid when current flows through the coil. When current is removed, the spring attached to the plunger will cause the plunger to return to its original position.

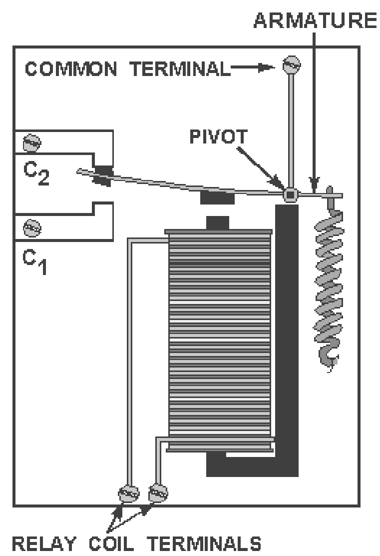
If a solenoid fails to operate, check the terminal connections, the plunger and attached mechanism for smooth operation, the energizing voltage, and the coil of the solenoid.



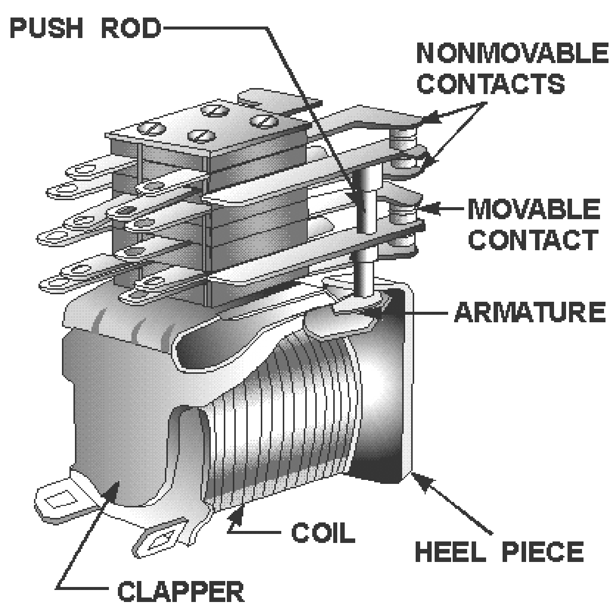
SOLENOID



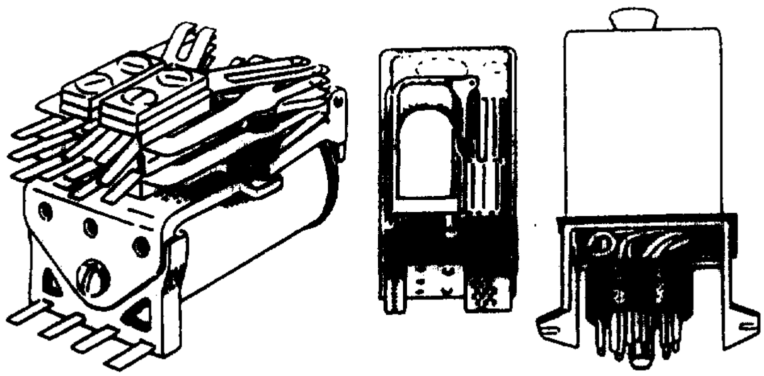
A **RELAY** is an electromagnetic control device that differs from the solenoid in that the solenoid uses a movable core (plunger) while the relay has fixed core. Relays are classified as CONTROL RELAYS, which control low power COMMON circuits and POWER RELAYS or CONTACTORS which control high power circuits.



CLAPPER RELAYS use a clapper (armature) to move contact positions and accomplish the switching of circuits.

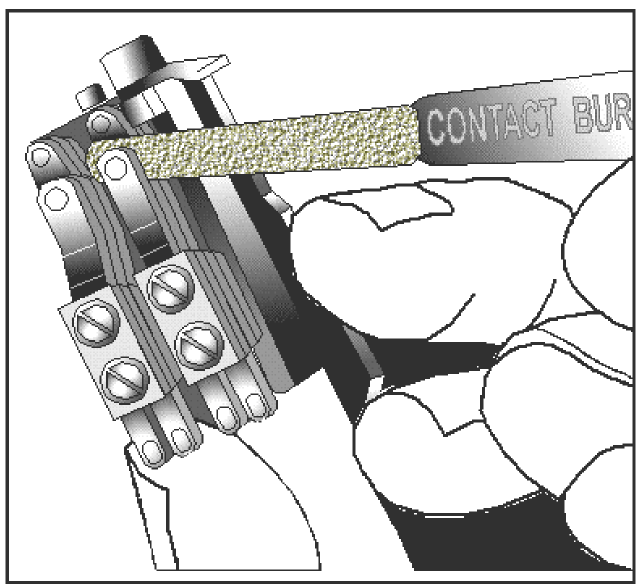


Relays are described by the type of enclosure. A relay may be OPEN, SEMISEALED, or SEALED.



If a relay fails to function, the movement of the contacts should be observed; the coil should be checked for opens or shorts; the terminal leads should be checked for burned or charred insulation; and the contact surfaces should be checked for carbon, arcing, and contact spacing.

A **BURNISHING TOOL** is used to clean the contacts of a relay. Files, sandpaper, and emery cloth should NOT be used.



A **POINT BENDER** is used to adjust contact spacing of a relay. No other tool should be used.

ANSWERS TO QUESTIONS Q1. THROUGH Q30.

A1. To remove power from a malfunctioning device; to remove power from a device you wish to work on and restore power when the work is completed; to turn devices on and off as the device is needed; to select the function or circuit desired within a device.

A2. Switches, solenoids, and relays.

A3.

- *a. Solenoid.*
- *b. Switch.*
- *c. Relay.*

A4. A manual switch must be turned on or off by a person. An automatic switch turns a circuit on or off without the action of a person (by using mechanical or electrical devices).

A5. A light switch, an ignition switch, television channel selector, etc.

A6. A thermostat, an automobile distributor, a limit switch, etc.

A7. Multicontact switches make possible the control of more than one circuit or the selection of one of several possible circuits with a single switch.

A8.

a. Three-pole, single-throw (triple-pole, single-throw)

b. Double-pole, double-throw

c. Single-pole, double-throw

d. Single-pole, single-throw

e. Double-pole, triple-throw

f. Six-pole, double-throw

A9.

a. Single-pole, single-throw, single-break

b. Single-pole, double-throw, single-break

c. Single-pole, single-throw, double-break

d. Single-pole, double-throw, double-break

e. Rotary

f. Wafer

g. Double-pole, double-throw, double-break

A10. The type of actuator.

A11. Two-position and three-position.

A12. A momentary switch.

A13. A locked-position switch.

A14. A microswitch.

A15. The maximum current a switch is designed to carry.

A16. The maximum voltage allowable in the circuit in which the switch is installed.

A17. An ohmmeter and a voltmeter.

A18. A voltmeter.

A19.

a. Not acceptable-single throw.

b. Not acceptable-double break.

c. Acceptable-choice #2 (different actuator).

d. Not acceptable-single pole.

e. Not acceptable-no momentary position.

f. Acceptable-choice #1 (higher rating).

g. Not acceptable-locked position incorrect.

h. Not acceptable-current rating too low.

i. Not acceptable-voltage rating too low.

A20. The switch operation for smooth and correct operation, the terminals for corrosion, and the physical condition of the switch.

A21. The magnetic field created in a coil of wire and core will attract a soft iron plunger when current flows through the coil.

A22. A starter motor and solenoid.

A23. The connections, the plunger, the mechanism that the solenoid actuates, the energizing voltage, and the coil of the solenoid.

A24. The magnetic field created in a coil of wire will attract aft armature causing a movement in sets of contacts.

A25. The solenoid provides a mechanical movement of a plunger (a moveable core) while the core of a relay is fixed.

A26. Control relays and power relays (contactors).

A27. By observing the movement of the contacts if the relay is open or sealed with a transparent cover. If the relay has an opaque cover, you can "feel" the operation of the relay by placing your finger on the cover.

A28. The coil should be checked for opens, shorts, or a short to ground; terminal leads should be checked for charred or burned insulation; the contact surfaces should be checked for film, carbon, arcing, and contact spacing.

A29. A burnishing tool.

A30. A point bender

APPENDIX I

GLOSSARY

ACTUATOR—The part of a switch that is acted upon to cause the switch to change contact connections; e.g., toggle, pushbutton, and rocker.

AMMETER—A meter used to measure current.

ARC EXTINGUISHER—The part of a circuit breaker that confines and divides the arc that occurs when the contacts of the circuit breaker open.

ARMATURE—In a relay, the movable portion of the relay.

BREAK—In a switch, the number of breaks refers to the number of points at which the switch opens the circuit; e.g., single break and double break.

BURNISHING TOOL—A tool used to clean and polish contacts on a relay.

CONTINUITY—An uninterrupted, complete path for current.

DAMPING—The process of smoothing out oscillations. In a meter, damping is used to keep the pointer of the meter from overshooting the correct reading.

D'ARSONVAL METER MOVEMENT—A name used for the permanent-magnet moving-coil movement used in most meters.

DIRECT SHORT—A connection between two points in a circuit, such as between a component and ground.

ELECTRODYNAMIC METER MOVEMENT—A meter movement using fixed field coils and a moving coil; usually used in wattmeters.

ELECTROMAGNETISM—The relationship between magnetism and electricity.

ELECTROSTATIC METER MOVEMENT—A meter movement that uses the electrostatic repulsion of two sets of charged plates (one fixed and the other movable). This meter movement reacts to voltage rather than to current and is used to measure high voltage.

FERRULES—The cylindrical metallic ends of a cartridge fuse.

FREQUENCY METER—A meter used to measure the frequency of an ac signal.

GALVANOMETER—A meter used to measure small values of current by electromagnetic or electrodynamic means.

HOT WIRE METER MOVEMENT—A meter movement that uses the expansion of a heated wire to move the pointer of a meter; measures dc or ac.

IN-CIRCUIT METER—A meter permanently installed in a circuit; used to monitor circuit operation.

LOADING EFFECT—The effect of a voltmeter upon the circuit being measured which results in an inaccurate measurement. Loading effect is minimized by using a voltmeter with an internal resistance many times higher than the resistance of the circuit being measured.

MAGNETIC TRIP ELEMENT—A circuit breaker trip element that uses the increasing magnetic attraction of a coil with increased current to open the circuit.

MEGGER—Common name for a megohmmeter.

MEGOHMMETER—A meter that measures very large values of resistance; usually used to check for insulation breakdown in wires.

METER—A device used to measure an electrical quantity; e.g., current, voltage, and frequency.

METER MOVEMENT—The part of a meter that moves.

MOVING-IRON METER MOVEMENT—Same as moving-vane meter movement.

MOVING-VANE METER MOVEMENT—A meter movement that uses the magnetic repulsion of the like poles created in iron vanes by current through a coil of wire; most commonly used movement for ac meters.

MULTIMETER—A single meter combining the functions of an ammeter, a voltmeter, and an ohmmeter.

NONTRIP-FREE CIRCUIT BREAKER—A circuit breaker that can be held ON during an overcurrent condition.

OHMMETER—A meter used to measure resistance.

OUT-OF-CIRCUIT METER—A meter which is not permanently installed in a circuit. Usually portable and self-contained, these meters are used to check the operation of a circuit or to isolate troubles within a circuit.

PARALLAX ERROR—The error in meter readings that results when you look at a meter from some position other than directly in line with the pointer and meter face. A mirror mounted on the meter face aids in eliminating parallax error.

POINT BENDER—A tool used to adjust the contact spacing on a relay.

POLE—(1) One end of a magnet. (2) The number of points at which current can enter a switch; e.g., single pole, double pole, and three pole.

POLE PIECE—A piece of ferromagnetic material used to control the distribution of magnetic lines of force; i.e., concentrate the lines of force in a particular place or evenly distribute the lines of force over a wide area.

RANGES—The several upper limits a meter will measure as selectable by a switch or by jacks; e.g., a voltmeter may have ranges of 1 volt, 2.5 volts, 10 volts, 25 volts, and 100 volts.

RECTIFIER—A device used to convert ac to pulsating dc.

RELAY—An electromagnetic device with one or more sets of contacts which changes position by the magnetic attraction of a coil to an armature.

RELUCTANCE—The resistance of a magnetic path to the flow of magnetic lines of force through it.

ROTARY SWITCH—A multicontact switch with contacts arranged in a circular or semi-circular manner.

SENSITIVITY—(1) For an ammeter: the amount of current that will cause full-scale deflection of the meter. (2) For a voltmeter: the ratio of the voltmeter resistance divided by the full-scale reading of the meter, expressed in ohms-per-volt.

SHORT CIRCUIT—An unintentional current path between two components in a circuit or between a component and ground which is usually caused by a malfunction in the circuit.

SHUNT RESISTOR—A resistor in parallel. In an ammeter, shunt resistors are used to provide range capability.

SNAP-ACTING—Changing position quickly with the aid of a spring.

SOLENOID—An electromagnetic device that changes electrical energy into mechanical motion; based upon the attraction of a movable iron plunger to the core of an electromagnet.

SWITCH—A device used to open or close a circuit.

TEST EQUIPMENT—A general term applied to devices used to test electrical and electronic circuits.

THERMAL TRIP ELEMENT—A circuit breaker trip element that uses the increased bending of a bimetallic strip caused by increased current to open a circuit.

THERMAL-MAGNETIC TRIP ELEMENT—A single circuit breaker trip element that combines the action of a thermal and a magnetic trip element.

THERMOCOUPLE METER MOVEMENT—A meter movement that uses the current induced in a thermocouple by the heating of a resistive element to measure the current in a circuit; used to measure ac or dc.

THROW—In a switch, the number of different circuits each pole can control; e.g., single throw and double throw.

TRIP-ELEMENT—The part of a circuit breaker that senses any overload condition and causes the circuit breaker to open the circuit.

TRIP-FREE CIRCUIT BREAKER—A circuit breaker that will open a circuit even if the operating 5 mechanism is held in the ON position.

TROUBLESHOOTING—The process of locating and repairing faults in electrical or electronic equipment.

VOLTMETER—A meter used to measure voltage.

WAFER SWITCH—A rotary switch in which the contacts are arranged on levels. Each level is electrically independent but mechanically connected by the shaft of the switch.

WATT-HOUR METER—A meter used to measure electrical energy.

WATTMETER—A meter used to measure electrical power.

APPENDIX II

LAWS OF EXPONENTS

The International Symbols Committee has adopted prefixes for denoting decimal multiples of units. The National Bureau of Standards has followed the recommendations of this committee, and has adopted the following list of prefixes:

Numbers	Powers of ten	Prefixes	Symbols
1,000,000,000,000	1012	tera	T
1,000,000,000	109	giga	G
1,000,000	106	mega	M
1,000	103	kilo	k
100	102	hecto	h
10	10	deka	da
.1	10-1	deci	d
.01	10-2	centi	c
.001	10-3	milli	m
.000001	10 -6	micro	u
.000000001	10-9	nano	n
.000000000001	10-12	Pico	p
.000000000000001	10-15	femto	f
.000000000000000001	10-18	Atto	a

To multiply like (with same base) exponential quantities, add the exponents. In the language of algebra the rule is $a^m \times a^n = a^{m+n}$

$$10^4 \times 10^2 = 10^{4+2} = 10^6$$

$$0.003 \times 825.2 = 3 \times 10^3 \times 8.252 \times 10^2$$

$$= 24.756 \times 10^1 = 2.4756$$

To divide exponential quantities, subtract the exponents. In the language of algebra the rule is

$$\frac{a^m}{a^n} = a^{m-n} \quad \text{or} \quad \frac{10^8}{10^2} = 10^6$$

$$\frac{3{,}000}{0.015} = \frac{(3 \times 10^3)}{(1.5 \times 10^{-2})} = 2 \times 10^5 = 200{,}000$$

To raise an exponential quantity to a power, multiply the exponents. In the language of algebra $(x^m)^n = x^{mn}$.

$$(10^3)^4 = 10^{3\times4} = 10^{12}$$

$$2{,}500^2 = (2.5\times10^3)^2 = 6.25\times10^6 = 6{,}250{,}00$$

Any number (except zero) raised to the zero power is one. In the language of algebra $x^0 = 1$

$$\frac{x^3}{x^3} = 1$$

$$\frac{10^4}{10^4} = 1$$

Any base with a negative exponent is equal to 1 divided by the base with an equal positive exponent. In the language of algebra $x^{-a} = 1/x^a$

$$10^{-2} = \frac{1}{10^2} = \frac{1}{100}$$

$$5a^{-3} = \frac{5}{a^3}$$

$$(6a)^{-1} = \frac{1}{6a}$$

To raise a product to a power, raise each factor of the product to that power.

$$(2\times10)^2 = 2^2\times10^2$$

$$3{,}000 = (3\times10^3)^3 = 27\times10^9$$

To find the nth root of an exponential quantity, divide the exponent by the index of the root. Thus, the nth root of $a^m = a^m$.

$$\sqrt{x^6 = x^{\frac{6}{2}} = x^3}$$

$$\sqrt[3]{64\times10^3 = 4\times10 = 40}$$

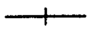
APPENDIX III

SCHEMATIC SYMBOLS

Contacts, Switches, Contactors, and Relays

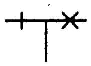
Switching Function

Conducting, closed contact (break)

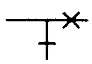


Nonconducting, open contact (make)

Application: transfer



OR



Electrical Contact

Fixed contact

Fixed contact for jack, key, relay, switch, etc

IEC

OR

Fixed contact with momentary contact (automatic return)

NOTE When this symbol (representing a contact with automatic return) is used on a diagram for international use, the convention should be so noted on the diagram or associated documentation. IEC

IEC 

‡ Sleeve

 OR 

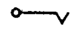
Moving Contact

‡ The broken line -—- indicates where line connection to a symbol is made and is not part of the symbol.

Adjustable or sliding contact for resistor, inductor, etc

OR

Locking



Nonlocking

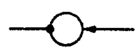
Segment; bridging contact

OR

Vibrator reed

Vibrator split reed

Rotating contact (slip ring) and brush



Basic Contact Assemblies

The standard method of showing a contact is by a symbol indicating the circuit condition it produces when the actuating device is in the deenergized or nonoperated position. The actuating device may be of a mechanical, electrical, or other nature, and a clarifying note may be necessary with the symbol to explain the proper point at which the contact functions; for example, the point where a contact closes or opens as a function of changing pressure, level, flow, voltage, current, etc. In cases where it is desirable to show contacts in the energized or operated condition and where confusion may result, a clarifying note shall be added to the drawing.

Auxiliary switches or contacts for circuit breakers, etc, may be designated as follows:

(a) Closed when device is energized or operated position.
(b) Closed when device is in deenergized or nonoperated position.
(aa) Closed when operating mechanism of main device is in energized or operated position.
(bb) Closed when operated mechanism of main device is in deenergized or nonoperated position.

See American National Standard Manual and Automatic Station Control, Supervisory, and Associated Telemetering Equipment, C37.2-1970, for further details.

In the parallel-line contact symbols shown below, the length of the parallel lines shall be approximately 1¼ times the width of the gap

Closed contact (make)

IEC 

OR

IEC 

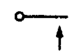
OR

IEC

Open contact (break)

IEC 

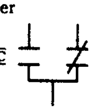
OR

IEC 

OR.

IEC 

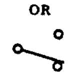
Transfer

IEC 

OR

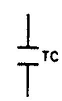
IEC 

OR

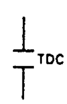
IEC 

Make-before-break

Application: open contact with time closing (TC) or time-delay closing (TDC) feature

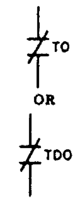


OR

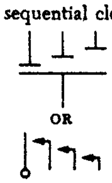


Contacts, Switches, Contactors, and Relays

Application: closed contact with time opening (TO) or time-delay opening (TDO) feature

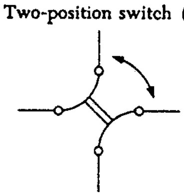


Time sequential closing

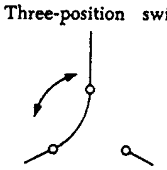


Multiway transfer switch

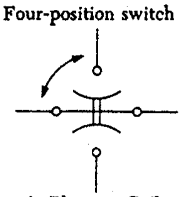
Two-position switch (90° step)

IEC 

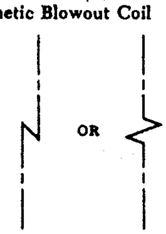
Three-position switch (120° step)

IEC 

Four-position switch (45° step)

IEC 

‡ **Magnetic Blowout Coil**

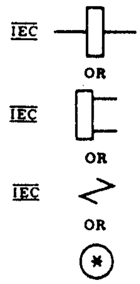


Operating Coil
Relay Coil

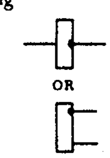
---

‡ The broken line -—- indicates where line connection to a symbol is made and is not part of the symbol.

NOTE The asterisk is not part of the symbol. Always replace the asterisk by a device designation.

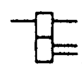


Semicircular dot indicates inner end of winding



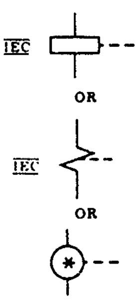
Application: multiwinding coil (2 windings shown)

NOTE The ends of a given winding shall be shown directly opposite each other on opposite sides of the core, or adjacent to each other on the same side of the core.

IEC  See Note

Electromagnetic actuator (solenoid), with mechanical linkage shown

NOTE The mechanical linkage may be omitted if the intent is clear.



* See Notes

Switch

See also FUSE

Fundamental symbols for contacts, mechanical connections, etc, may be used for switch symbols.

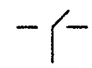
The standard method of showing switches is in a position with no operating force applied. For switches that may be in any of two or more positions with no operating force applied, and for switches actuated by some mechanical device (as in air-pressure, liquid-level, rate-of-flow, etc, switches), a clarifying note may be necessary to explain the point at which the switch functions.

When the basic switch symbols are shown in the closed position on a diagram, terminals must be added for clarity.

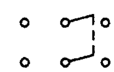
Single-throw, general

IEC 

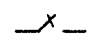
Double-throw, general



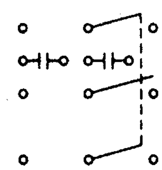
Application: 2-pole double-throw switch with terminals shown



Knife switch , general

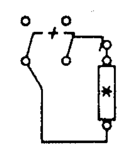


Application: 3-pole double-throw knife switch with auxiliary contacts and terminals



Application: 2-pole field-discharge knife switch with terminals and discharge resistor

NOTE The asterisk is not part of the symbol. Always add identification within or adjacent to the rectangle.



*See Note

Switch with horn gap

Sector switch

IEC 

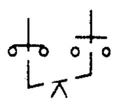
Pushbutton Momentary or Spring-Return

Circuit closing (make)

Circuit opening (break)

Two-circuit

Two-Circuit, Maintained or Not Spring-Return

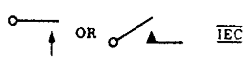


Nonlocking Switch, Momentary or Spring-Return

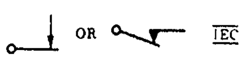
The symbols to the left are commonly used for spring buildups in key switches, relays, and jacks.

The symbols to the right are commonly used for toggle switches.

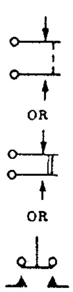
Circuit closing (make)



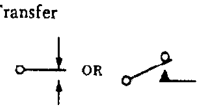
Circuit opening (break)



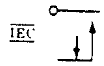
Two-circuit



Transfer



Make-before-break

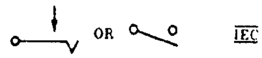


Locking Switch

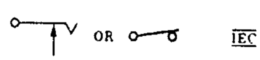
The symbols to the left are commonly used for spring buildups in key switches and jacks.

The symbols to the right are commonly used for toggle switches.

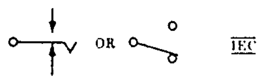
Circuit closing (make)



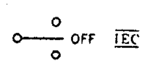
Circuit opening (break)



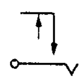
Transfer, 2-position



Transfer, 3-position



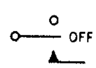
Make-before-break



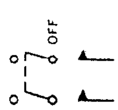
Combination Locking and Nonlocking Switch

Commonly used for toggle switches

3-position, 1-pole: circuit closing (make), off, momentary circuit closing (make)

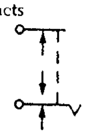


3-position, 2-pole: circuit closing (make), off, momentary circuit closing (make)

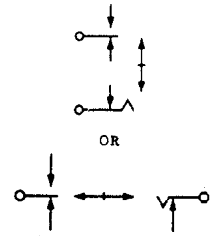


Key-Type Switch
Lever Switch

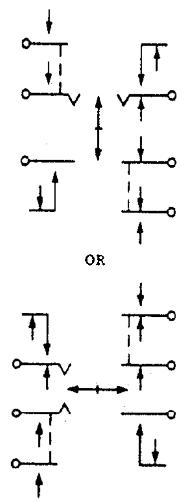
2-position with locking transfer and break contacts



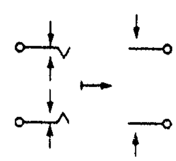
3-position with nonlocking transfer and locking break contacts



3-position, multicontact combination



2-position, half of key switch normally operated, multicontact combination

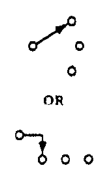


Selector or Multiposition Switch

The position in which the switch is shown may be indicated by a note or designation of switch position.

General (for power and control diagrams)

Any number of transmission paths may be shown.



If specific type identification is required: circuit closing

Switches with Time-Delay Feature

NOTE The point of the arrow indicates the direction of switch operation in which contact action is delayed.

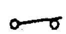
Open switch with time-delay closing (TDC) feature

TDC

OR

See Note

Closed switch with time-delay opening (TDO) feature



TDO

OR

See Note

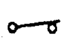
Open switch with time-delay opening (TDO) feature

TDO

OR

See Note

Closed switch with time-delay closing (TDC) feature



TDC

OR

See Note

Flow-Actuated Switch

Closes on increase in flow

Opens on increase in flow

Liquid-Level-Actuated Switch

Closes on rising level

Opens on rising level

Pressure- or Vacuum-Actuated Switch

Closes on rising pressure

Opens on rising pressure

Temperature-Actuated Switch

Closes on rising temperature

IEC

OR

Opens on rising temperature

IEC

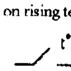
OR

Thermostat

NOTE The t° symbol shall be shown or be replaced by data giving the nominal or specific operating temperature of the device.

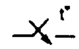
NOTE If clarification of direction of contact operation is needed, a directional arrow may be added. The arrowhead shall point in the direction of rising temperature operation. A directional arrow shall always be shown for central-off (neutral) position devices.

Closes on rising temperature



See Note

With contact-motion direction clarified

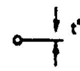


See Note

Opens on rising temperature

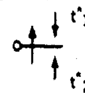
See Note

Transfers on rising temperature



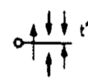
See Note

Transfer, with intended central-off (neutral) position



See Notes

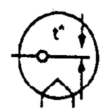
Application: multifunction, typical



See Notes

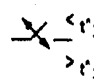
With integral heater and transfer contacts

Use only if essential to indicate integral heater details.



See Notes

Application: with operating temperatures indicated



See Notes

Flasher
Self-Interrupting Switch

OR

Foot-Operated Switch
Foot Switch

Opens by foot pressure

Closes by foot pressure

Switch Operated by Shaft Rotation and Responsive to Speed or Direction

Speed

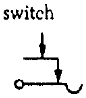
Plugging: to stop drive after it has come practically to rest

Anti-plugging: to prevent plugging of drive

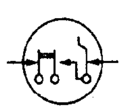
Centrifugal switch (opening on increasing speed)

Switches with Specific Features

Hook switch

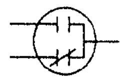


Telephone dial (switch)

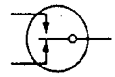


TYPICAL

Switch in evacuated envelope, 1-pole double-throw



OR



Mushroom-head safety feature

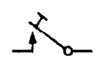
Application to 2-circuit pushbutton switch.

Key-operated lock switch

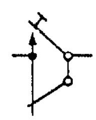
Use appropriate standard symbol and add key designation or other information in note.

Telegraph Key

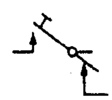
Simple



Simple with shorting switch

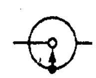


Open-circuit or pole-changing



Governor (Contact-making) Speed Regulator

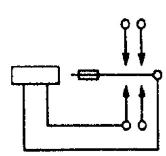
Contacts open or closed as required (shown here as closed).



Vibrator, Interrupter

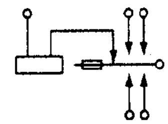
Typical shunt drive (with terminals shown)

Show contacts as required.



Typical separate drive (with terminals shown)

Show contacts as required.



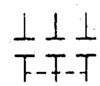
Contactor

See also CIRCUIT BREAKER

Fundamental symbols for contacts, coils, mechanical connections, etc, are the basis of contactor symbols and should be used to represent contactors on complete diagrams. Complete diagrams of contactors consist of combinations of fundamental symbols for control coils, mechanical connections, etc, in such configurations as to represent the actual device. Mechanical interlocking should be indicated by notes.

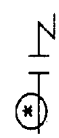
Manually operated 3-pole contactor

IEC

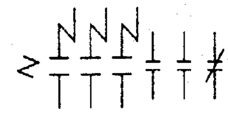


Electrically operated 1-pole contactor with series blowout coil

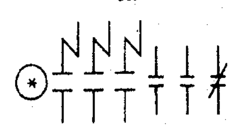
OR



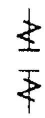
Electrically operated 3-pole contactor with series blowout coils; 2 open and 1 closed auxiliary contacts (shown smaller than the main contacts)



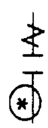
OR



Electrically operated 1-pole contactor with shunt blowout coil



OR



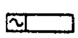
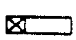
Relay

See OPERATING COIL; RELAY COIL

Fundamental symbols for contacts, mechanical connections, coils, etc, are the basis of relay symbols and should be used to represent relays on complete diagrams.

The following letter combinations or symbol elements may be used with relay symbols. The requisite number of these letters or symbol elements may be used to show what special features a relay possesses.

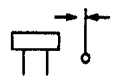
The terms "slow" and "fast" are relative, and the degree is not to be noted by a multiplicity of the same relay symbol on a diagram. Relays that are direct-current operated are not marked to indicate dc operation.

IEC  AC Alternating-current or ringing relay
D Differential
DB Double-biased (biased in both directions)
DP Dashpot
EP Electrically polarized
FO Fast-operate
FR Fast-release
L Latching
MG Marginal
ML Magnetic-latching (remanent)
NB No bias
NR Nonreactive
 P Magnetically polarized using biasing spring, or having magnet bias
SA Slow-operate and slow-release
IEC  SO Slow-operate
 SR Slow-release
SW Sandwich-wound to improve balance to longitudinal currents

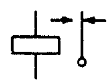
The proper poling for a polarized relay shall be shown by the use of + and − designations applied to the winding leads. The interpretation of this shall be that a voltage applied with the polarity as indicated shall cause the armature to move toward the contact shown nearer the coil on the diagram. If the relay is equipped with numbered terminals, the proper terminal numbers shall also be shown.

Basic

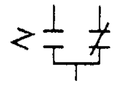
Application: relay with transfer contacts

IEC 

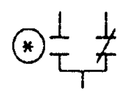
OR

IEC 

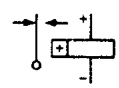
OR

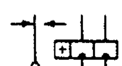
IEC 

OR

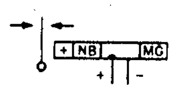


Application: polarized relay with transfer contacts (two typical types shown)





Application: polarized (no bias) marginal relay with transfer contacts



Relay, thermally operated

Activating device for thermally operated relay

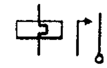
Time of delay may be shown.

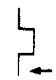
Contacts may be shown separately from the operating device.

IEC 

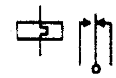
OR

With normally open contacts shown (two typical types)

IEC 

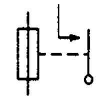


With transfer contacts shown

IEC 

Thermal relay, one-time type, not reusable

Normally open contact type shown.

IEC 

Inertia Switch (operated by sudden deceleration)

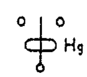
 IEC

NOTE This symbol is commonly used on diagrams for aerospace applications.

Mercury Switch

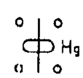
Leveling

Three terminal



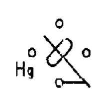
IEC

Four terminal



IEC

With acceleration cutoff (four terminal)



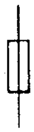
IEC

Circuit Protectors

Fuse (one-time thermal current-overload device)

General

IEC



OR

OR

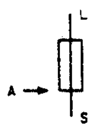
Fuse, supply side indicated by a thick line

IEC

Fuse with alarm contact

NOTE When fuse blows, alarm bus A is connected to power supply bus S. The letters S (supply), L (load), and A (alarm circuit) are for explanation only, and are not part of the symbol.

IEC



See Note

Isolating fuse-switch; high-voltage primary fuse cutout, dry

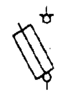
IEC

High-voltage primary fuse cutout, oil

OR

Isolating fuse-switch for on-load switching

IEC



Current Limiter (for power cable)

The arrowheads in this case are filled.

NOTE Use appropriate number of single-line diagram symbols.

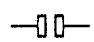
See Note

Lightning Arrester
Arrester (electric surge, etc)
Gap

General

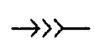
Carbon block; telephone protector block

The sides of the rectangle shall be approximately in the ratio of 1 to 2 and the space between rectangles shall be approximately equal to the width of a rectangle.



Electrolytic or aluminum cell

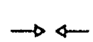
This symbol is not composed of arrowheads.



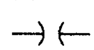
Horn gap

Protective gap

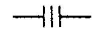
These triangles shall not be filled.



Sphere gap

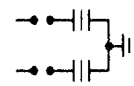


Valve or film element



Multigap, general

Application: gap plus valve plus ground, 2-pole

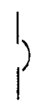


Circuit Breaker

If it is desired to show the condition causing the breaker to trip, the relay protective-function symbols in item 9.5.1 may be used alongside the breaker symbol.

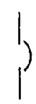
General

IEC



Air circuit breaker, if distinction is needed; for alternating-current circuit breakers rated at 1,500 volts or less and for all direct-current circuit breakers

IEC

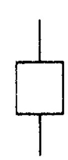


Network protector

Circuit breaker

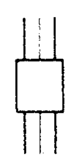
The symbol in the right column is for a 3-pole breaker.

NOTE On a power diagram, the symbol may be used without other identification. On a composite drawing where confusion with the general circuit element symbol (item 16.1) may result, add the identifying letters CB inside or adjacent to the square.



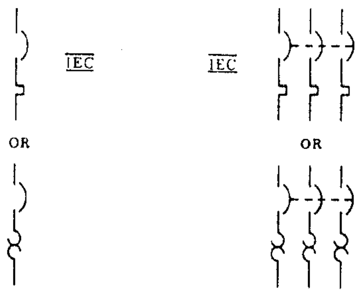
IEC

IEC



See Note

Application: 3-pole circuit breaker with thermal-overload device in all 3 poles



Application: 3-pole circuit breaker with magnetic-overload device in all 3 poles

Application: 3-pole circuit breaker, drawout type

Protective Relay

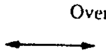
Fundamental symbols for contacts, coils, mechanical connections, etc, are the basis of relay symbols and should be used to represent relays on complete diagrams.

See RELAY COIL; OPERATING COIL and RELAY

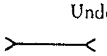
Relay protective functions

The following symbols may be used to indicate protective functions, or device-function numbers may be placed in the circle or adjacent to the basic symbol (see American National Standard for Manual and Automatic Station Control, Supervisory, and Associated Telemetering Equipments, C37.2-1970).

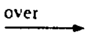
Over, general



Under, general

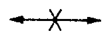


Direction, general; directional over

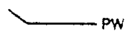


Balance, general

Differential, general



Pilot wire, general



Carrier current, general

Operating quantity

The operating quantity is indicated by the following letters or symbols placed either on or immediately above the relay protective-function symbols shown above.

C	*Current
Z	Distance
F	Frequency
GP	Gas pressure
ϕ	Phase
W	Power
S	Synchronism
T	Temperature
V	Voltage

* The use of the letter may be omitted in the case of current, and the absence of such letter presupposes that the relay operates on current.

Ground relays

Relays operative on residual current only are so designated by attaching the ground symbol to the relay protective-function symbol. Note that the zero phase-sequence designation given below may be used instead when desirable.

Phase-sequence quantities

Operations on phase-sequence quantities may be indicated by the use of the conventional subscripts 0, 1, and 2 after the letter indicating the operating quantity.

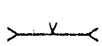
Applications

Overcurrent

Directional overcurrent

Directional residual overcurrent

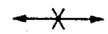
Undervoltage



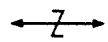
Power directional

Balanced current

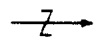
Differential current



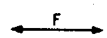
Distance



Directional distance



Overfrequency



Overtemperature

Phase balance

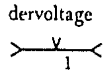
Phase sequence

Pilot wire, differential-current

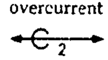
Pilot wire, directional-comparison

Carrier pilot

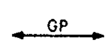
Positive phase-sequence undervoltage



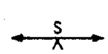
Negative phase-sequence overcurrent



Gas-pressure (Buchholz)



Out-of-step



Meter
Instrument

NOTE The asterisk is not part of the symbol. Always replace the asterisk by one of the following letter combinations, depending on the function of the meter or instrument, unless some other identification is provided in the circle and explained on the diagram.

* See Note

A	Ammeter IEC
AH	Ampere-hour meter
C	Coulombmeter
CMA	Contact-making (or breaking) ammeter
CMC	Contact-making (or breaking) clock
CMV	Contact-making (or breaking) voltmeter
CRO	Oscilloscope Cathode-ray oscillograph
DB	DB (decibel) meter Audio level/meter
DBM	DBM (decibels referred to 1 milliwatt) meter
DM	Demand meter
DTF	Demand-totalizing relay
F	Frequency meter
GD	Ground detector
I	Indicating meter
INT	Integrating meter
μA or UA	Microammeter
MA	Milliammeter
NM	Noise meter
OHM	Ohmmeter
OP	Oil pressure meter
OSCG	Oscillograph, string
PF	Power factor meter
PH	Phasemeter
PI	Position indicator
RD	Recording demand meter
REC	Recording meter
RF	Reactive factor meter
SY	Synchroscope
t°	Temperature meter
THC	Thermal converter
TLM	Telemeter
TT	Total time meter Elapsed time meter
V	Voltmeter IEC
VA	Volt-ammeter
VAR	Varmeter
VARH	Varhour meter
VI	Volume indicator Audio-level meter
VU	Standard volume indicator Audio-level meter
W	Wattmeter IEC
WH	Watthour meter

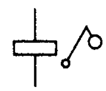
Galvanometer

IEC 

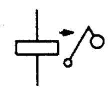
OR

Electromagnetically Operated Counter
Message Register

General



With make contact



APPENDIX IV

CROSS REFERENCE OF MILITARY AND COMMERCIAL FUSE DESIGNATIONS

OLD MILITARY	NEW MILITARY	OLD COMMERCIAL	NEW COMMERCIAL
F02GR010A	F02A 250V 1/100A	3AG 1/100 250V	AGC 1/100 250V
F02GR031A	F02A 250V 1/32A	3AG 1/32 250V	AGC 1/32 250V
F02GR062A	F02A 250V 1/16A	3AG 1/16 250V	AGC 1/16 250V
F02G1R50A	F02A 250V 1 1/2A	3AG 1 1/2 250V	AGC 1 1/250V
F02G2R00A	F02A 250V 2A	3AG 2 250V	AGC 2 250V
F02GR010B	F02B 250V 1/100A	3AG 1/100 250V	MDL 1/100 250V
F02GR031B	F02B 250V 1/32A	3AG 1/32 250V	MDL 1/32 250V
F02GR375B	F02B 250V 3/8A	3AG 3/8 250V	MDL 3/8 250V
F02D1R50B	F02B 125V 1 1/2A	3AG 1 1/2 125V	MDL 1 1/2 125V
F02D2R00B	F02B 125V 2A	3AG 2 125V	MDL 2 125V
F03G1R00A	F03A 250V 1A	3AB 1 250V	ABC 1 250V
F03G3R00A	F03A 250V 3A	3AB 3 250V	ABC 3 250V
F03G10R0A	F03A 250V 10A	3AB 10 250V	ABC 10 250V
F03G12R0A	F03A 250V 12A	3AB 12 250V	ABC 12 250V
F03D20R0A	F03A 125V 20A	3AB 20 125V	ABC 20 125V
F03D30R0A	F03A 125V 30A	3AB 30 125V	ABC 30 125V
F04A10R0A	F02A 32V 10A	3AG 10 32V	AGC 10 32V
F04A15R0A	F02A 32V 15A	3AG 15 32V	AGC 15 32V
F04A5R00B	F02B 32V 5A	3AG 5 32V	MDL 5 32V
F04A10R0B	F02B 32V 10A	3AG 10 32V	MDL 10 32V
F05A10R0A	F05A 32V 10A	4AG 10 32V	AGS 10 32V
F05A15R0A	F05A 32V 15A	4AG 15 32V	AGS 15 32V
F05A10R0B	F05B 32V 10A	4AG 10 32V	MDM 10 32V
F05A15R0B	F05B 32V 15A	4AG 15 32V	MDM 15 32V
F06G1R00A	F06A 250V 1A	4AB 1 250V	ABS 1 250V
F06G2R00A	F06A 250V 2A	4AB 2 250V	ABS 2 250V
F07A5R00A	F07A 32V 5A	5AG 5 32V	AGU 5 32V
F07A10R0A	F07A 32V 10A	5AG 10 32V	AGU 10 32V
F07A5R00B	F07B 32V 5A	5AG 5 32V	MDR 5 32V
F07A10R0B	F07B 32V 10A	5AG 10 32V	MDR 10 32V
F08G1R00A	F07A 250V 1A	5AG 1 250V	AGU 1 250V
F08G2R00A	F07A 250V 2A	5AG 2 250V	AGU 2 250V
F08G1R00B	F09B 250V 1A	5AG 1 250V	FNM 1 250V
F08G2R00B	F09B 250V 2A	5AG 2 250V	FNM 2 250V

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Circuit Protection Devices," pages 1-1 through 1-73.

1-1. Circuit measurement is used for which of the following purposes?

1. To find the weight of a circuit
2. To increase the power used in a circuit
3. To discover the length and width of a circuit
4. To determine the reason a circuit is not functioning properly

1-2. An in-circuit meter is used for which of the following purposes?

1. To reduce circuit losses
2. To monitor circuit operation
3. To control power to a circuit
4. To prevent circuit overload conditions

1-3. Out-of-circuit meters have which of the following advantages over in circuit meters?

1. They can be used on more than one device
2. They are lighter weight
3. They are more rugged
4. All of the above

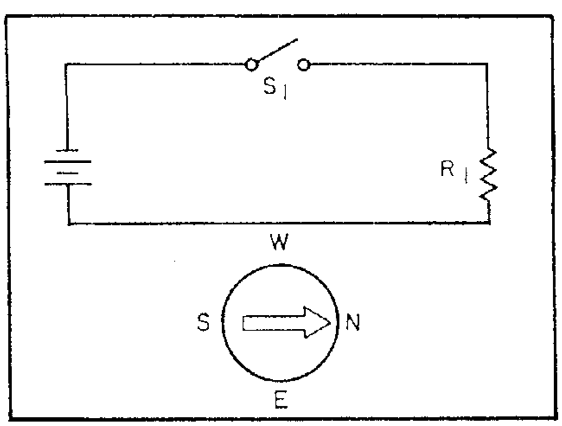


Figure 1-A.—Compass and wire (dc).

IN ANSWERING QUESTIONS 1-4 THROUGH 1-6, REFER TO FIGURE 1-A.

1-4. When S1 is closed, the compass needle will align itself in which of the following manners?

1. With magnetic north
2. With geographic north
3. Parallel to the conductor
4. With the magnetic field around the wire

1-5. If the current through the conductor is decreased, what happens to the magnetic field around the conductor?

1. It reverses
2. It decreases
3. It increases
4. It oscillates

1-6. When the current through the conductor decreases, the compass needle will react in which of the following manners?

1. Point more to magnetic north
2. Move farther away from magnetic north
3. Swing back and forth from east to west
4. Vibrate rapidly back and forth around magnetic north

1-7. The d'Arsonval meter movement is based on which of the following principles?

1. Moving vane
2. Electrostatic
3. Electrodynamic
4. Permanent-magnet moving-coil

1-8. Current through a meter results in the pointer. In d'Arsonval meter movement, what force produces this deflection?

1. Thermocouple action
2. Electrostatic repulsion
3. Mechanical spring tension
4. The interaction of magnetic fields

1-9. The hairsprings in a d'Arsonval meter movement perform which of the following functions?

1. They keep the pointer in the position of the last indication when current is removed
2. They aid the movement of the pointer when there is current through the meter
3. They make electrical connections to the meter movement
4. All of the above

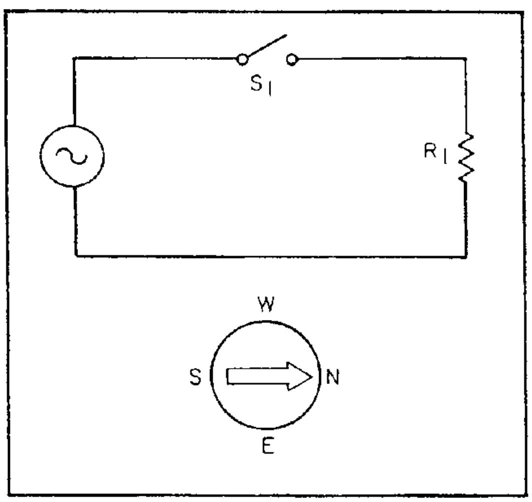


Figure 1-B.—Compass and wire (ac).

IN ANSWERING QUESTIONS 1-10 AND 1-11, REFER TO FIGURE 1-B.

1-10. If the frequency of the ac source is 5 Hz, how will the compass needle react when S1 is closed?

1. It will swing back and forth
2. It will point directly at the wire
3. It will point directly away from the wire
4. It will vibrate rapidly around magnetic north

1-11. If the frequency of the ac source is 200 Hz, how will the compass needle react when S1 is closed?

1. It will swing back and forth
2. It will point directly at the wire
3. It will point directly away from the wire
4. It will vibrate rapidly around magnetic north

1-12. What device allows a d'Arsonval meter movement to measure ac by converting ac to pulsating dc?

1. A pulsator
2. A modulator
3. A rectifier
4. A converter

1-13. What is meant by the term "meter damping"?

1. Moistening the felt pads
2. Smoothing the oscillations of the pointer
3. Preventing excessive current through the coil
4. Compensating for electromagnetic induced interference

1-14. Which of the following methods is used to dampen a meter?

1. Mount the meter in a mu-metal case
2. Install a fuse in one of the input leads
3. Incorporate an airtight chamber containing a van
4. Provide a fluid reservoir and sponge arrangement next to the pads

1-15. A d'Arsonval meter movement reacts to which of the following values of voltage?

1. Peak
2. Average
3. Effective
4. Peak-to-peak

1-16. What value of ac is indicated by a meter scale?

1. Peak
2. Average
3. Effective
4. Peak-to-peak

1-17. Which of the following meter movements will measure either ac or dc without the use of a rectifier?

1. GMS
2. d'Arsonval
3. Electrostatic
4. Electrodynamic

1-18. What electrical property is reacted to by the electrodynamic, d'Arsonval, moving-vane, and thermocouple meter movements?

1. Power
2. Current
3. Voltage
4. Resistance

1-19. What electrical property is measured by an ammeter?

1. Power
2. Current
3. Voltage
4. Resistance

1-20. How are ammeters connected in an electrical circuit?

1. In series with the load
2. In parallel with the load
3. In accordance with Lenz's Law
4. In series-parallel with the load

1-21. How does an ammeter affect the circuit being measured?

1. It acts as a resistances in series and lowers the circuit current
2. It acts as a resistance in series and raises the circuit current
3. It acts as a resistance in parallel and lowers the circuit current
4. It acts as a resistance in parallel and raises the circuit current

1-22. How is the effect that an ammeter produces in a circuit kept to a minimum?

1. By using a large resistor in series with the ammeter
2. By using a large capacitor in parallel with the ammeter
3. By ensuring that the meter resistance is low compared to circuit resistance
4. By ensuring that the meter resistance is high compared to circuit resistance

1-23. The ammeter with the greatest sensitivity has which of the following characteristics?

1. The lowest amount of current for full-scale deflection indication
2. The highest amount of current for full-scale deflection indication
3. A low ratio of internal resistance to full-scale deflection indication
4. A high ratio of internal resistance to full-scale deflection indication

1-24. Ammeters measure various ranges through the addition of which of the following components?

1. Shunt resistors in series with the meter movement
2. Shunt resistors in parallel with the meter movement
3. Capacitors in series with the meter movement
4. Capacitors in parallel with the meter movement

1-25. What range of an ammeter should you use for an initial measurement?

1. The lowest range
2. The highest range
3. The mid-scale range

1-26. What portion of the ammeter scale should be used to take a final reading?

1. The upper half
2. The lower half
3. The mid-scale portion
4. Anywhere on the meter face

1-27. When, if ever, can you use a dc ammeter to measure ac values?

1. When the ac is high frequency
2. For low values
3. Always
4. Never

1-28. Which of the following safety precautions should be observed prior to connecting an ammeter into a circuit?

1. Switch to the highest range
2. Observe proper dc polarity
3. Deenergize the circuit
4. All of the above

1-29. What electrical property is measured by a voltmeter?

1. Power
2. Current
3. Voltage
4. Resistance

1-30. A voltmeter should be connected in an electrical circuit in what manner?

1. In series with the load
2. In parallel with the load
3. In accordance with Lenz's Law
4. In series-parallel with the load

1-31. A voltmeter has an effect on the circuit being measured; what is this effect called?

1. Loading
2. Damping
3. Rectification
4. Eddy-current drag

1-32. To keep the effect of a voltmeter on a circuit to a minimum, the internal resistance of the voltmeter must have which of the following relationships to the circuit load?

1. Equal to
2. Lower than
3. Higher than
4. In proportion to

1-33. Which of the following types of meters can be made from a current sensitive meter movement?

1. Ammeter
2. Ohmmeter
3. Voltmeter
4. Each of the above

1-34. A voltmeter has a high sensitivity when it has which of the following characteristics?

1. Low deflection indication
2. High deflection indication
3. Low ratio of internal resistance to full-scale deflection indication
4. High ratio of internal resistance to full-scale deflection indication

1-35. Which of the following configurations extends the range of a voltmeter?

1. A resistor in series with the meter movement
2. A resistor in parallel with the meter movement
3. A capacitor in series with the meter movement
4. A capacitor in parallel with the meter movement

1-36. What voltmeter range should be used for initial measurements?

1. The lowest
2. The highest
3. The mid-scale

1-37. The electrostatic meter movement reacts to which of the following electrical properties?

1. Power
2. Current
3. Voltage
4. Resistance

1-38. Electrostatic meter movements are used to measure which of the following current/voltage values?

1. Low voltage
2. Low current
3. High voltage
4. High current

1-39. Which of the following safety precautions should be observed when a voltmeter is used?

1. Deenergize the circuit before connecting the meter
2. Start with the lowest range of the meter
3. Connect the meter in series with the circuit
4. All of the above

1-40. What electrical property is measured with an ohmmeter?

1. Power
2. Current
3. Voltage
4. Resistance

1-41. An ohmmeter is used to check for which of the following conditions?

1. Continuity
2. Overheating
3. Overcurrent
4. Undercurrent

1-42. How should an ohmmeter be connected in an electrical circuit?

1. In series with the load
2. In parallel with the load
3. In parallel with the source
4. In series-parallel with the load

1-43. An ohmmeter can measure different ranges because of the use of which of the following components?

1. Range coils
2. Range resistors
3. Range capacitors
4. Range potentiometers

1-44. What area of an ohmmeter scale should be used when a measurement is taken?

1. Upper half
2. Lower half
3. Mid-scale portion
4. Anywhere on the meter face

1-45. Ohmmeter are classified by type. What are the two types of ohmmeters?

1. Series and shunt
2. Normal and reverse
3. Full- and half-scale

1-46. What is the most obvious differences in the two types of ohmmeters?

1. The ranges of the meters
2. The scales of the meters
3. The power sources of the meters
4. The size of the test leads of the meters

1-47. Which of the following safety precautions should be observed when an ohmmeter is used?

1. Always start with the highest scale of the meter
2. Deenergize the circuit before connecting the meter
3. Observe proper polarity
4. All of the above

1-48. Meggers (megohmmeters) are used to measure which of the following quantities?

1. Low voltage
2. High voltage
3. Low resistance
4. High resistance

1-49. When a megger is used to check the insulation of a wire, which of the following indications should be considered normal?

1. ∞
2. 0
3. 500 V
4. 1000 V

1-50. Which of the following safety precautions should be observed when a megger is used?

1. Do not use a dc megger to measure circuits that are powered by ac
2. Always start with the highest scale selection of the meter
3. Do not touch the meter leads when a measurement is being taken
4. All of the above

1-51. A multimeter can be used to measure which of the following electrical properties?

1. Voltage
2. Current
3. Resistance
4. Each of the above

1-52. The function switch on a multimeter does NOT perform which of the following functions?

1. Selection of the meter range
2. Determination of the proper scale
3. Selection of ac or dc capability
4. Changing of the multimeter from an ammeter to a voltmeter

1-53. One of the problems encountered in building a multimeter is that the meter movement gives different readings for the same values of ac and dc. Which of the following features of a multimeter will solve this problem?

1. A rectifier
2. An ac/dc switch
3. Separate scales for ac and dc
4. A mirror on the face of the meter

1-54. Why is there a mirror on the face of a multimeter?

1. To illuminate the meter face
2. To aid in reducing parallax error
3. To reduce the friction between the pointer and the meter face
4. To compensate for the difference in ac and dc measurements

1-55. If the mirror on the face of a multimeter is used properly, where will the image of the pointer appear?

1. Hidden behind the pointer
2. Barely visible on either side of the pointer
3. Clearly visible to the left of the pointer
4. Clearly visible to the right of the pointer

1-56. Which of the following safety precautions does NOT apply to a multimeter?

1. Observe proper dc polarity when measuring dc
2. Deenergize the circuit before connecting the meter
3. Be sure the meter is switched to ac for ac measurements
4. Never apply power to the circuit when measuring voltage with the meter

1-57. If a multimeter has no OFF position, and it is returned to storage, on which of the following positions should the meter be set?

1. +dc; highest voltage range
2. –dc; higher resistance range
3. Ac; highest voltage range
4. Ac; highest current range

1-58. When the current in a conductor is measured without the conductor being disconnected, which of the following meters could be used?

1. Multimeter
2. Hook-on voltameter
3. Induction wattmeter
4. Transformer voltmeter

1-59. Which of the following electrical quantities is measured by a wattmeter?

1. Power
2. Energy
3. Voltage
4. Current

1-60. Which of the following electrical quantities is measured by a watt hour meter?

1. Power
2. Energy
3. Voltage
4. Current

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Circuit Protection Devices," pages 2-1 through 2-42.

2-1. Circuit protection devices are used for which of the following purposes?

1. To protect people
2. To protect circuits
3. To guard against hazardous conditions
4. All of the above

2-2. Which of the following conditions does NOT require the use of a circuit protection device?

1. Direct short
2. High resistance
3. Excessive current
4. Abnormal heating

2-3. When a point in a circuit, where full system voltage is present, comes in direct contact with the ground or return side of the circuit, which of the following terms applies?

1. Direct short
2. High resistance
3. Excessive current
4. Abnormal heating

2-4. When circuit current increases beyond the designed current carrying capability of the circuit, which of the following terms applies?

1. Direct short
2. High resistance
3. Excessive current
4. Abnormal heating

2-5. If the bearings of a generator were to fail, which of the following circuit conditions would probably occur?

1. Direct short
2. High resistance
3. Excessive current
4. Abnormal heating

2-6. How are circuit protection devices connected to the circuit they are protecting?

1. Alongside
2. In series
3. In parallel
4. In series-parallel

2-7. Which of the following two are circuit protection devices?

1. Electrical plugs and CO_2 cartridges
2. CO_2 cartridges and circuit breakers
3. Fuses and circuit breakers
4. Fuses and electrical plugs

THIS SPACE LEFT BLANK INTENTIONALLY.

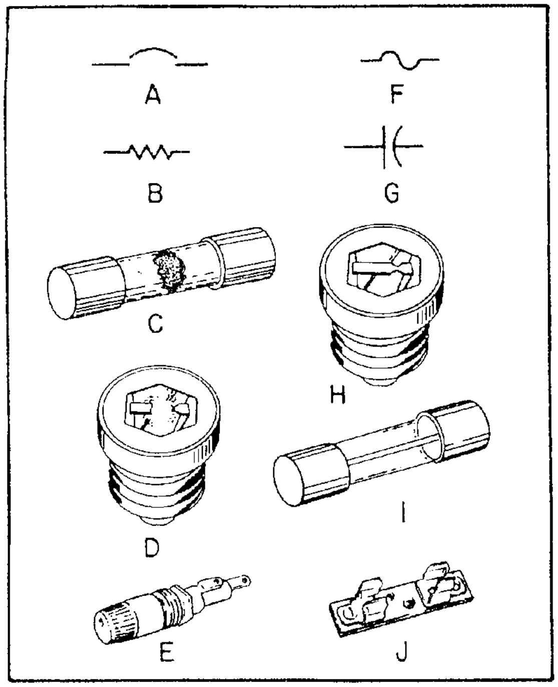


Figure 2-A.—Recognition practice.

IN ANSWERING QUESTIONS 2-8 THROUGH 2-13, REFER TO FIGURE 2-A.

2-8. Which of the following is the schematic symbol for a fuse?

1. A
2. B
3. F
4. G

2-9. Which of the following is the schematic symbol for a circuit breaker?

1. A
2. B
3. F
4. G

2-10. Which of the following is an illustration of an open cartridge fuse?

1. C
2. D
3. E
4. I

2-11. Which of the following is an illustration of a good cartridge fuse?

1. C
2. D
3. E
4. I

2-12. Which of the following is an illustration of a good plug-type fuse?

1. C
2. D
3. E
4. H

2-13. Which of the following is an illustration of an open plug-type fuse?

1. C
2. D
3. E
4. H

2-14. Which of the following factors is NOT used to rate fuses?

1. Size
2. Current
3. Voltage
4. Time delay

2-15. A fuse current rating has which of the following definitions?

1. The maximum current that can flow through a circuit without causing the circuit to overheat
2. The maximum current that will flow through a circuit if there is a direct short
3. The maximum current that will flow through a fuse without opening the fuse
4. The maximum current that will not "jump" an open fuse

2-16. A fuse voltage rating has which of the following definitions?

1. The maximum voltage that can exist in a circuit without causing the circuit to overheat
2. The maximum voltage that can exist in a circuit if there is a direct short
3. The maximum voltage across a fuse that will not cause the fuse to open
4. The maximum voltage across a fuse that will not jump the open fuse

---

IN ANSWERING QUESTIONS 2-17 THROUGH 2-19, MATCH THE TIME-DELAY RATING LISTED IN COLUMN B TO THE ELECTRICAL DEVICE LISTED IN COLUMN A. NOT ALL ITEMS IN COLUMN B WILL BE USED.

	A. ELECTRICAL DEVICES		B. TIME-DELAY RATINGS
2-17.	Electric motor	1.	Fast
2-18.	Lighting circuit	2.	Delay
2-19.	Meter Movement	3.	Standard
		4.	Intermediate

2-20. What is the voltage rating for a fuse with the designation F03D1R00B?

1. 32 V or less
2. 125 V or less
3. 250 V or less
4. 500 V or less

2-21. What is the current rating for a fuse with the designation F03B0R50B?

1. 1/2 amp
2. 1.5 amp
3. 3 amp
4. 50 amp

2-22. What is the time-delay rating for a fuse with the designation F03A20R0C?

1. Fast
2. Delay
3. Standard
4. Intermediate

2-23. What is the voltage rating for a fuse with the designation F02B250V10AS?

1. 10 V or less
2. 32 V or less
3. 52 V or less
4. 250 V or less

2-24. What is the current rating for a fuse with the designation F03A125V5A?

1. 125 amp
2. 5 amp
3. 3 amp
4. 1/8 amp

2-25. What is the time-delay rating for a fuse with the designation F04C125V2AS?

1. Fast
2. Delay
3. Standard
4. Intermediate

2-26. What is the voltage rating for a fuse with the designation 3AG20125V?

1. 20 V or less
2. 90 V or less
3. 125 V or less
4. 250 V or less

2-27. What is the current rating for a fuse with the designation 3AG1032V?

1. 1 amp
2. 2 amp
3. 3 amp
4. 10 amp

2-28. What is the voltage rating for a fuse with the designation AGC5125V?

1. 12 V or less
2. 25 V or less
3. 51 V or less
4. 125 V or less

2-29. What is the current rating for a fuse with the designation AGC2125V?

1. 1 amp
2. 2 amp
3. 3 amp
4. 25 amp

2-30. What is the new military designation for a fuse with the old, military designation F03D1R50B?

1. F03A125V1.5A
2. F02B125V1.5A
3. F03A250V11/2A
4. F03B125V1.5A

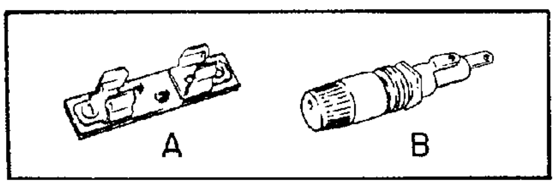


Figure 2-B.—Fuseholder identification.

IN ANSWERING QUESTIONS 2-31 THROUGH 2-34, REFER TO FIGURE 2-B.

2-31. What type of fuseholder is shown in figure 2-B(A)?

1. Clip
2. Post
3. Bayonet
4. Spring-loaded

2-32. What type of fuseholder is shown in figure 2-B(B)?

1. Clip
2. Post
3. Bayonet
4. Spring-loaded

2-33. When you use the fuseholder shown in figure 2-B(B), which connector should be connected to the power source?

1. Ground
2. Center
3. Inside
4. Outside

2-34. When you use the fuseholder shown in figure 2-B(B), which connector should be connected to the load?

1. Ground
2. Center
3. Inside
4. Outside

2-35. Which of the following methods will provide an ABSOLUTE determination as to whether or not a fuse is open?

1. A visual inspection
2. A check of the fuse indicator
3. A voltmeter check of the fuse
4. A thermometer check of the temperature of the fuse

2-36. A fuse is removed from a circuit, checked with an ohmmeter, and found to be shorted. What action should be taken?

1. Discard the fuse
2. Check the fuse with a voltmeter
3. Put the fuse back in the circuit
4. Return the fuse to the supply department

2-37. Which of the following methods should be used to check a .002 ampere fuse?

1. Use a megger and place a capacitor in parallel with the fuse
2. Use a megger and place a capacitor in series with the fuse
3. Use an ohmmeter and place a resistor in parallel with the fuse
4. Use an ohmmeter and place a resistor in series with the fuse

2-38. What should you use to remove a fuse from a clip-type fuseholder?

1. A scribe
2. A fusepuller
3. A screwdriver
4. A pair of pliers

2-39. Which of the following is a safety precaution to be observed when a fuse is checked?

1. Turn the power off and discharge the circuit before the fuse is removed
2. When you check a fuse with an ohmmeter, be careful to avoid short circuits
3. When you use a voltmeter to check a low current fuse, be careful to avoid opening the fuse by excessive current from the voltmeter
4. All of the above

YOU HAVE FOUND AN OPEN FUSE IN A PIECE OF EQUIPMENT AND HAVE REPAIRED THE CASUALTY. THE TECHNICAL MANUAL FOR THE EQUIPMENT SPECIFIES A FUSE CODED F02A125V3A. NO FUSES WITH THAT DESIGNATION ARE AVAILABLE. THE FOLLOWING FUSES ARE CARRIED BY THE SUPPLY SYSTEM.

A. F03D3R00A
B. 3AG3250V
C. F02A3R00B
D. AGC3125V
E. F02D3R00C
F. AGC5250V

Figure 2-C.—Fuse replacement problem.

IN ANSWERING QUESTIONS 2-40 THROUGH 2-46, USE THE INFORMATION IN FIGURE 2-C.

2-40. What fuse is a direct replacement?

1. A
2. C
3. D
4. E

2-41. What fuse is the best substitute?

1. A
2. B
3. C
4. F

2-42. What fuse is the second best in the fuseholder substitute?

1. A
2. C
3. E
4. F

2-43. What fuse is unacceptable because the physical size is incorrect?

1. A
2. C
3. E
4. F

2-44. What fuse is unacceptable because of the current rating?

1. B
2. D
3. E
4. F

2-45. What fuse is unacceptable because of the voltage rating?

1. A
2. B
3. C
4. D

2-46. What fuse is unacceptable because of the time-delay rating?

1. A
2. C
3. D
4. E

2-47. Before replacing a fuse, you should check for which of the following conditions?

1. Proper fit
2. Proper fuse
3. Both 1 and 2 above
4. Proper input voltage

2-48. Which of the following is NOT a safety precaution to be observed when a fuse is changed?

1. Be sure to "tag out" the fuseholder when you remove the fuse
2. Remove the power from a circuit before removing and replacing a fuse
3. Remove any corrosion from the fuseholder before replacing a fuse
4. Be sure the fuse fits properly in the fuseholder

2-49. When you perform preventive maintenance on fuses, which of the following is NOT a condition you should check?

1. Corrosion
2. Shorted fuse
3. Improper fit
4. Improper fuse

2-50. What is the total number of main components in a circuit breaker?

1. Five
2. Two
3. Three
4. Four

2-51. Which of the following is NOT a type of trip element for a circuit breaker?

1. Thermal
2. Magnetic
3. Mechanical
4. Thermal-magnetic

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 2-52 THROUGH 2-54, SELECT FROM COLUMN B THE TRIP ELEMENT THAT IS DESCRIBED BY THE ACTION LISTED IN COLUMN A. NOT ALL ITEMS IN COLUMN B WILL BE USED.

	A. ACTIONS		B. TRIP ELEMENTS
2-52.	An electromagnet is connected in series with the load	1.	Thermal
2-53.	A bimetallic strip is heated by the load current	2.	Magnetic
2-54.	A bimetallic strip is heated by the load current and electromagnet is connected in series with the load	3.	Mechanical
		4.	Thermal-magnetic

2-55. A circuit breaker that will trip even if the operating mechanism is held ON is known as what type of circuit breaker?

1. Standard
2. Emergency
3. Trip free
4. Nontrip free

2-56. What type of circuit breaker can be overridden if the operating mechanism is held ON?

1. Standard
2. Emergency
3. Trip free
4. Nontrip free

2-57. Which of the following is NOT a time-delay rating for a circuit breaker?

1. Long
2. Short
3. Standard
4. Instantaneous

2-58. Selective tripping is used to cause which of the following circuit breakers to trip when there is an overload?

1. The least expensive
2. The most accessible
3. The smallest current rating
4. The closest to the fault

2-59. Selective tripping is used to accomplish which of the following purposes?

1. To reduce wear and tear on circuit breakers
2. To isolate a faulty circuit without affecting other circuits
3. To simplify the task of resetting the circuit breaker
4. To enable the application of power to emergency circuits during an overload

THIS SPACE LEFT BLANK INTENTIONALLY.

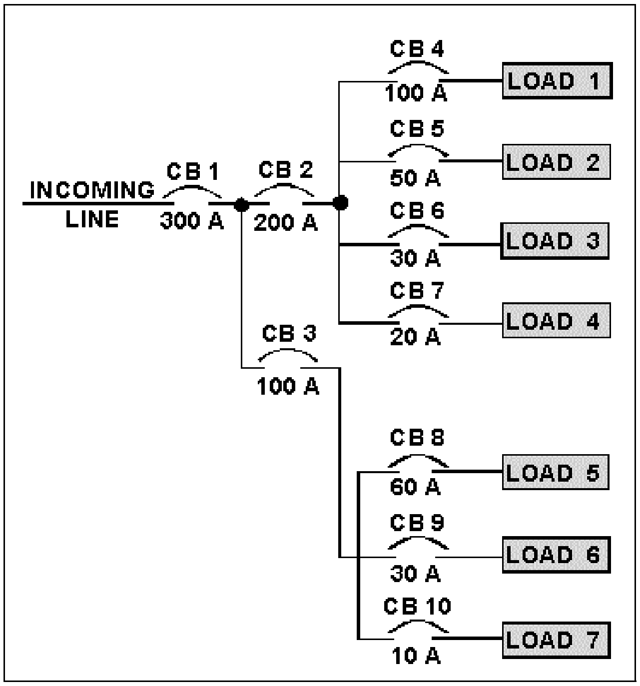


Figure 2-D.—Use of circuit breakers in a power distribution system.

IN ANSWERING QUESTIONS 2-60 THROUGH 2-62, REFER TO FIGURE 2-D.

2-60. Which of the following circuit breakers should have a long time delay?

1. CB1
2. CB2
3. CB3
4. CB4

2-61. Which of the following circuit breakers should have a short time delay?

1. CB1
2. CB2
3. CB5
4. CB4

2-62. Which of the following circuit breakers should have an instantaneous time delay?

1. CB1
2. CB2
3. CB3
4. CB4

2-63. The following actions must be taken prior to working on a circuit breaker. Arrange these items in the proper sequence, then select the choice below that lists the events in the proper sequence.

A. Tag the power switch
B. Obtain the approval of the electrical officer
C. Remove power to the circuit breaker
D. Check the applicable technical manual

1. A, B, C, D
2. C, B, D, A
3. D, B, C, A
4. B, A, D, C

2-64. Which of the following items is NOT checked during maintenance on a circuit breaker?

1. Input power voltage
2. Operating mechanism smoothness
3. Terminal tightness and corrosion
4. Contact surfaces for pitting

ASSIGNMENT 3

Textbook assignment: Chapter 3, "Circuit Control Devices," pages 3-1 through 3-39.

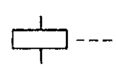
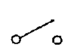
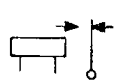
3-1. Circuit control devices should NOT be used for which of the following reasons?

1. To adjust the power level of a device
2. To remove power from a malfunctioning device
3. To apply power to a device when work is completed on it
4. To select the function or circuit desired within a device

3-2. Which of the following are types of circuit control devices?

1. Relays
2. Switches
3. Solenoids
4. All of the above

IN ANSWERING QUESTIONS 3-3 THROUGH 3-5, MATCH THE SCHEMATIC SYMBOL SHOWN IN COLUMN B TO THE DEVICE LISTED IN COLUMN A.

A. DEVICES		B. SCHEMATIC SYMBOLS	
3-3.	A switch	1.	
3-4.	A relay	2.	
3-5.	A solenoid	3.	
		4.	

3-6. Which of the following is a manual switch?

1. A light switch
2. A limit switch
3. A thermostat
4. A distributor

3-7. Which of the following is an automatic switch?

1. An ignition switch on a motor vehicle
2. A switch that turns on a light in a refrigerator
3. A channel selector on a television
4. A dial or push button on a telephone

3-8. Control or selection of one or more circuits is a function of which of the following switches?

1. A manual switch
2. An automatic switch
3. A multicontact switch
4. A single contact switch

THIS SPACE LEFT BLANK INTENTIONALLY.

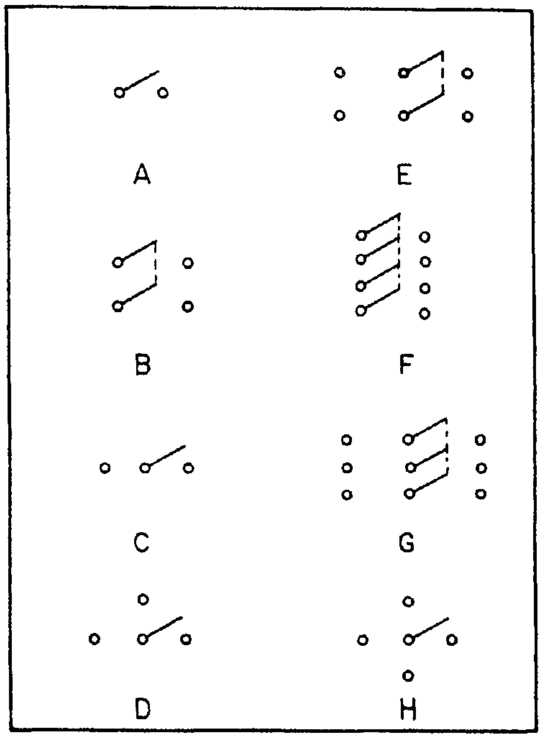


Figure 3-A.—Switch schematics.

IN ANSWERING QUESTIONS 3-9 THROUGH 3-15, REFER TO FIGURE 3-A. SELECT THE SYMBOL THAT REPRESENTS THE TYPE OF SWITCH STATED IN EACH QUESTION.

3-9. A single-pole, double-throw switch.

1. A
2. B
3. C
4. D

3-10. A double-pole, double-throw switch.

1. B
2. E
3. G
4. H

3-11. A single-pole, four-throw switch.

1. B
2. D
3. F
4. H

3-12. A double-pole, single-throw switch.

1. B
2. C
3. E
4. H

3-13. A single-pole, single-throw switch.

1. A
2. B
3. C
4. D

3-14. A three-pole, double-throw switch.

1. E
2. F
3. G
4. H

3-15. A single-pole, triple-throw switch.

1. B
2. D
3. E
4. G

THIS SPACE LEFT BLANK INTENTIONALLY.

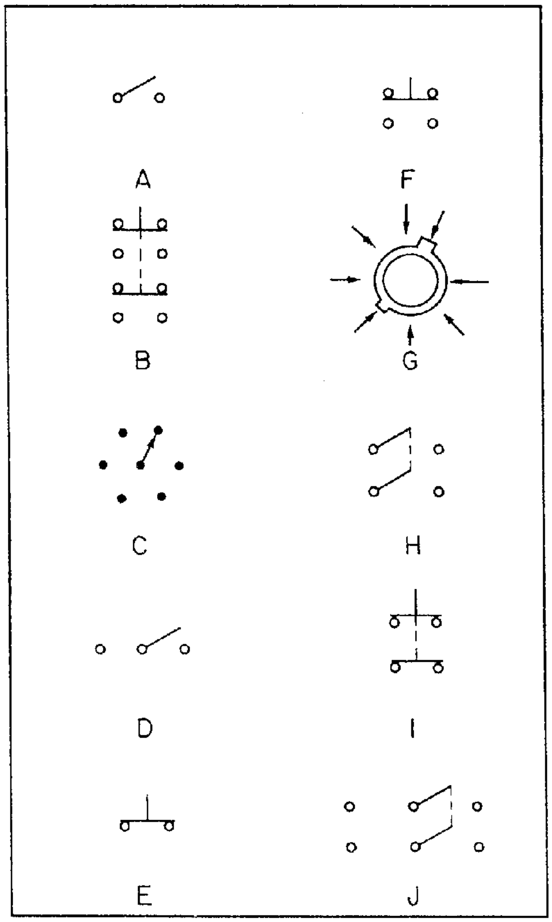


Figure 3-B.—Switch schematics.

IN ANSWERING QUESTIONS 3-16 THROUGH 3-24, REFER TO FIGURE 3-B. SELECT THE SYMBOL THAT REPRESENTS THE TYPE OF SWITCH STATED IN EACH QUESTION.

3-16. A single-pole, single-throw, double-break switch.

1. A
2. C
3. D
4. E

3-17. A double-pole, single-throw, double-break switch.

1. F
2. H
3. I
4. J

3-18. A single-pole, double-throw, single-break switch.

1. A
2. D
3. E
4. F

3-19. A single-pole, double-throw, double-break switch.

1. E
2. F
3. H
4. I

3-20. A double-pole, single-throw, single-break switch.

1. E
2. H
3. I
4. J

3-21. A double-pole, double-throw, double-break switch.

1. B
2. F
3. H
4. I

3-22. A double-pole, double-throw, single-break switch.

1. F
2. H
3. I
4. J

3-23. A rotary switch.

1. C
2. D
3. H
4. I

3-24. A wafer switch.

1. D
2. E
3. G
4. H

3-25. A switch actuator is described by which of the following terms?

1. Momentary
2. Two-position
3. Toggle
4. Four-position

3-26. What is the maximum number of different single-pole, single throw switch positions possible?

1. One
2. Two
3. Three
4. Four

3-27. What is the maximum number of different single-pole, double throw switch positions possible?

1. One
2. Two
3. Three
4. Four

3-28. Control of a circuit requiring a temporary actuation signal is provided by which of the following switches?

1. Momentary
2. Locked-in
3. Locked-out
4. Rotary

3-29. To guard against the accidental actuation of a circuit, which of the following types of switches are used?

1. Momentary
2. Locked-in
3. Locked-out
4. Rotary

3-30. To guard against the accidental turning off of a circuit, which of the following types of switches are used?

1. Momentary
2. Locked-in
3. Locked-out
4. Rotary

3-31. What is the common name for a accurate snap-acting switch?

1. Maxiswitch
2. Multiswitch
3. Miniswitch
4. Microswitch

3-32. Designation of switch current rating is based on which of the following current values?

1. Maximum
2. Minimum
3. Nominal
4. Average

3-33. Designation of switch voltage rating is based on which of the following voltage values?

1. Maximum
2. Minimum
3. Nominal
4. Average

3-34. Checking a switch with the circuit power NOT applied is accomplished by using which of the following meters?

1. Wattmeter
2. Frequency meter
3. Temperature meter
4. Ohmmeter

3-35. Checking a switch with the power applied is accomplished by using which of the following meters?

1. Megger
2. Ohmmeter
3. Wattmeter
4. Voltmeter

A THREE-POLE, DOUBLE-THROW, SINGLE-BREAK, THREE POSITION, ROCKER SWITCH IS FAULTY. THIS SWITCH HAS A MOMENTARY POSITION 3 AND IS LOCKED INTO POSITION 1. THE VOLTAGE AND CURRENT RATING FOR THE SWITCH ARE 230 VOLTS, 3 AMPERES. THE FOLLOWING SWITCHES ARE AVAILABLE.

	POLES	THROWS	BREAKS	NUMBER OF POSITIONS	MOMENTARY POSITION	LOCKED POSITION	ACTUATOR	RATING
A	4	1	1	2	---	---	TOGGLE	115V 3A
B	2	2	2	3	2	OUT 1	ROCKER	400V 1A
C	4	1	1	2	3	OUT 1	ROCKER	230V 3A
D	2	2	2	3	2	IN 1	ROCKER	115V 6A
E	3	3	2	3	1	IN 2	TOGGLE	400V 3A
F	3	2	1	3	3	IN 1	ROCKER	230V 5A
G	3	2	1	3	3	IN 1	TOGGLE	230V 3A
H	3	2	1	3	1	IN 2	ROCKER	230V 10A

Figure 3-C.—Switch replacement problem.

IN ANSWERING QUESTIONS 3-36 THROUGH 3-45, REFER TO FIGURE 3-C.

3-36. What switch is the best substitute?

1. C
2. F
3. G
4. H

3-37. What switch is the second best substitute?

1. C
2. F
3. G
4. H

3-38. Which of the following switches is unacceptable because of the number of poles?

1. A
2. D
3. F
4. G

3-39. Which of the following switches is unacceptable because of the number of throws?

1. A
2. B
3. D
4. E

3-40. Which of the following switches is unacceptable because of the number of breaks?

1. A
2. C
3. E
4. G

3-41. Which of the following switches is unacceptable because of the number of positions?

1. C
2. D
3. E
4. F

3-42. Which of the following switches is unacceptable because of the momentary position?

1. C
2. F
3. G
4. H

3-43. Which of the following switches has the wrong locked position?

1. B
2. D
3. F
4. G

3-44. Which of the following switches has an unacceptable voltage rating?

1. B
2. D
3. E
4. F

3-45. Which of the following switches has an unacceptable current rating?

1. B
2. C
3. D
4. F

3-46. When you perform preventive maintenance on a switch, which of the following items should be checked?

1. The terminals for corrosion
2. The physical condition of the switch
3. The switch operation for smooth and correct operation
4. All of the above

3-47. A solenoid is based upon which of the following principles?

1. A bimetallic strip bends when it is heated
2. A thermocouple produces a current when heated
3. A coil attracts a soft iron core when current flows in the coil
4. A soft iron core moving in a magnetic field creates a current

3-48. A solenoid is commonly used in which of the following devices?

1. A fuel quantity indicating system
2. A shipboard lighting system
3. A sound-powered telephone system
4. A starter for a motor vehicle

3-49. If a solenoid is not operating properly, which of the following items need NOT be checked?

1. Coil
2. Armature
3. Plunger
4. Energizing voltage

IN ANSWERING QUESTIONS 3-50 THROUGH 3-54, MATCH THE DEVICE(S) LISTED IN COLUMN B TO THE STATEMENTS IN COLUMN A. THE DEVICES IN COLUMN B MAY BE USED MORE THAN ONCE.

	A. STATEMENTS		B. DEVICES
3-50.	An electromagnetic device	1.	A switch only
3-51.	A device with a movable plunger	2.	A solenoid only
3-52.	A device with a fixed core	3.	A relay only
3-53.	A device classified as power or control	4.	A relay and a solenoid
3-54.	A device containing an armature		

3-55. If a relay is hermetically sealed with an opaque cover, which of the following methods should be used to determine whether the relay is operating?

1. Shake the relay and listen for loose parts
2. Place your finger on the cover and feel the relay contact movement
3. Remove the cover and visually observe the relay contacts when the relay is activated
4. Activate the relay and observe whether a metal object is attracted by the magnetic field

3-56. If a relay is NOT operating properly, which of the following items need NOT be checked?

1. The armature resistance
2. The terminal leads
3. The contact surfaces
4. The contact spacing

3-57. What should be used to clean the contacts of a relay?

1. Sandpaper
2. Emery cloth
3. A jeweler's file
4. A burnishing tool

3-58. What should be used to adjust contact clearances on a relay?

1. A point bender
2. A burnishing tool
3. A pair of pliers
4. A pair of hemostats

NONRESIDENT
TRAINING
COURSE
SEPTEMBER 1998

Navy Electricity and Electronics Training Series

Module 4—Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading

NAVEDTRA 14176

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Although the words “he,” “him,” and “his” are used sparingly in this course to enhance communication, they are not intended to be gender driven or to affront or discriminate against anyone.

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: To introduce the student to the subject of Electrical Conductors, Wiring Techniques, and Schematic Reading who needs such a background in accomplishing daily work and/or in preparing for further study.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and either the occupational or naval standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards*, NAVPERS 18068.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
AECS Steve Heartsfield

NAVSUP Logistics Tracking Number
0504-LP-026-8290

Sailor's Creed

"I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country's Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all."

TABLE OF CONTENTS

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7*, Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 3 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

Student Comments

Course Title: *NEETS Module 4*
Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading

NAVEDTRA: 14176 **Date**: ______________

We need some information about you:

Rate/Rank and Name: ______________ SSN: ______________ Command/Unit ______________

Street Address: ______________ City: ______________ State/FPO: ______________ Zip ______________

Your comments, suggestions, etc.:

Privacy Act Statement: Under authority of Title 5, USC 301, information regarding your military status is requested in processing your comments and in preparing a reply. This information will not be divulged without written authorization to anyone other than those within DOD for official use in determining performance.

NETPDTC 1550/41 (Rev 4-00)

CHAPTER 1

ELECTRICAL CONDUCTORS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the OCC-ECC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completing this chapter, you should be able to:

1. Recall the definitions of unit size, mil-foot, square mil, and circular mil and the mathematical equations and calculations for each.
2. Define specific resistance and recall the three factors used to calculate it in ohms.
3. Describe the proper use of the American Wire Gauge when making wire measurements.
4. Recall the factors required in selecting proper size wire.
5. State the advantages and disadvantages of copper or aluminum as conductors.
6. Define insulation resistance and dielectric strength including how the dielectric strength of an insulator is determined.
7. Identify the safety precautions to be taken when working with insulating materials.
8. Recall the most common insulators used for extremely high voltages.
9. State the type of conductor protection normally used for shipboard wiring.
10. Recall the design and use of coaxial cable.

ELECTRICAL CONDUCTORS

In the previous modules of this training series, you have learned about various circuit components. These components provide the majority of the operating characteristics of any electrical circuit. They are useless, however, if they are not connected together. Conductors are the means used to tie these components together.

Many factors determine the type of electrical conductor used to connect components. Some of these factors are the physical size of the conductor, its composition, and its electrical characteristics. Other factors that can determine the choice of a conductor are the weight, the cost, and the environment where the conductor will be used.

CONDUCTOR SIZES

To compare the resistance and size of one conductor with that of another, we need to establish a standard or unit size. A convenient unit of measurement of the diameter of a conductor is the mil (0.001, or one-thousandth of an inch). A convenient unit of conductor length is the foot. The standard unit of size in most cases is the MIL-FOOT. A wire will have a unit size if it has a diameter of 1 mil and a length of 1 foot.

SQUARE MIL

The square mil is a unit of measurement used to determine the cross-sectional area of a square or rectangular conductor (views A and B of figure 1-1). A square mil is defined as the area of a square, the sides of which are each 1 mil. To obtain the cross-sectional area of a square conductor, multiply the dimension of any side of the square by itself. For example, assume that you have a square conductor with a side dimension of 3 mils. Multiply 3 mils by itself (3 mils × 3 mils). This gives you a cross-sectional area of 9 square mils.

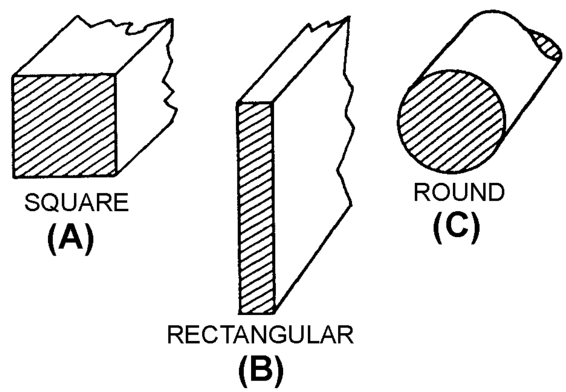


Figure 1-1.—Cross-sectional areas of conductors.

Q1. State the reason for the establishment of a "unit size" for conductors.

Q2. Calculate the diameter in MILS of a conductor that has a diameter of 0.375 inch.

Q3. Define a mil-foot.

To determine the cross-sectional area of a rectangular conductor, multiply the length times the width of the end face of the conductor (side is expressed in mils). For example, assume that one side of the rectangular cross-sectional area is 6 mils and the other side is 3 mils. Multiply 6 mils × 3 mils, which equals 18 square mils. Here is another example. Assume that a conductor is 3/8 inch thick and 4 inches wide. The 3/8 inch can be expressed in decimal form as 0.375 inch. Since 1 mil equals 0.001 inch, the thickness of the conductor will be 0.001 × 0.375, or 375 mils. Since the width is 4 inches and there are 1,000 mils per inch, the width will be 4 × 1,000, or 4,000 mils. To determine the cross-sectional area, multiply the length by the width; or 375 mils × 4,000 mils. The area will be 1,500,000 square mils.

Q4. Define a square mil as it relates to a square conductor.

CIRCULAR MIL

The circular mil is the standard unit of measurement of a round wire cross-sectional area (view C of figure 1-1). This unit of measurement is found in American and English wire tables. The diameter of a round conductor (wire) used to conduct electricity may be only a fraction of an inch. Therefore, it is convenient to express this diameter in mils to avoid using decimals. For example, the diameter of a wire is expressed as 25 mils instead of 0.025 inch. A circular mil is the area of a circle having a diameter of 1 mil, as shown in view B of figure 1-2. The area in circular mils of a round conductor is obtained by squaring the diameter, measured in mils. Thus, a wire having a diameter of 25 mils has an area of 25^2, or 625 circular mils. To determine the number of square mils in the same conductor, apply the conventional formula for determining the area of a circle ($A = \pi r^2$). In this formula, A (area) is the unknown and is equal to the cross-sectional area in square mils, π is the constant 3.14, and r is the radius of the circle, or half the diameter (D). Through substitution, A = 3.14, and $(12.5)^2$; therefore, 3.14 × 156.25 = 490.625

square mils. The cross-sectional area of the wire has 625 circular mils but only 490.625 square mils. Therefore, a circular mil represents a smaller unit of area than the square mil.

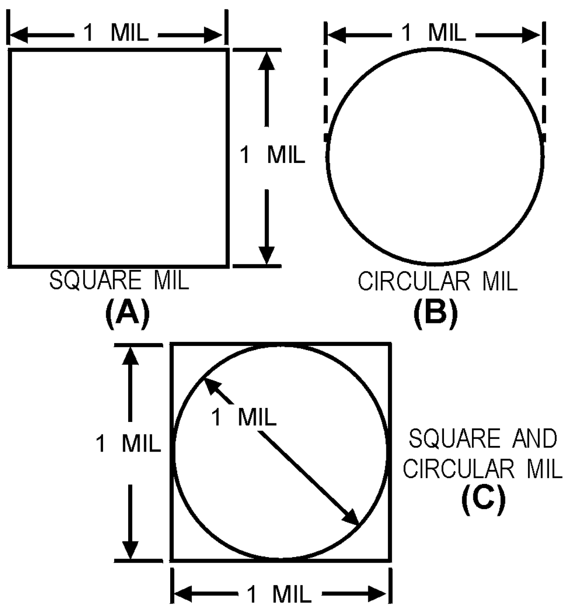


Figure 1-2.—A comparison of circular and square mils.

If a wire has a cross-sectional diameter of 1 mil, by definition, the circular mil area (CMA) is $A = D^2$, or $A = 1^2$, or A = 1 circular mil. To determine the square mil area of the same wire, apply the formula $A = \pi r^2$; therefore, $A = 3.14 \times (.5)^2$ (.5 representing half the diameter). When $A = 3.14 \times .25$, A = .7854 square mil. From this, it can be concluded that 1 circular mil is equal to. 7854 square mil. This becomes important when square (view A of figure 1-2) and round (view B) conductors are compared as in view C of figure 1-2.

When the square mil area is given, divide the area by 0.7854 to determine the circular mil area, or CMA. When the CMA is given, multiply the area by 0.7854 to determine the square mil area. For example,

Problem: A 12-gauge wire has a diameter of 80.81 mils. What is (1) its area in circular mils and (2) its area in square mils?

Solution:

(1) $A = D^2 = 80.81^2 = 6{,}530$ circular mils

(2) $A = 0.7854 \times 6{,}530 = 5{,}128.7$ square mils

Problem: A rectangular conductor is 1.5 inches wide and 0.25 inch thick. What is (1) its area in square mils and (2) in circular mils? What size of round conductor is necessary to carry the same current as the rectangular bar?

Solution:

(1) 1.5 inches = 1.5 inches × 1,000 mils per inch = 1,500 mils

0.25 inch = 0.25 inch × 1,000 mils per inch = 250 mils

$A = 1{,}500 \times 250 = 375{,}000$ square mils

(2) To carry the same current, the cross-sectional area of the round conductor must be equal. There are more circular mils than square mils in this area. Therefore:

$$A = \frac{375{,}000}{0.7854} = 477{,}000 \text{ circular mils}$$

A wire in its usual form is a single slender rod or filament of drawn metal. In large sizes, wire becomes difficult to handle. To increase its flexibility, it is stranded. Strands are usually single wires twisted together in sufficient numbers to make up the necessary cross-sectional area of the cable. The total area of stranded wire in circular mils is determined by multiplying the area in circular mils of one strand by the number of strands in the cable.

Q5. Define a circular mil.

Q6. What is the circular mil area of a 19-strand conductor if each strand is 0.004 inch?

CIRCULAR-MIL-FOOT

A circular-mil-foot (figure 1-3) is a unit of volume. It is a unit conductor 1foot in length and has a cross-sectional area of 1 circular mil. Because it is a unit conductor, the circular-mil-foot is useful in making comparisons between wires consisting of different metals. For example, a basis of comparison of the RESISTIVITY (to be discussed shortly) of various substances may be made by determining the resistance of a circular-mil-foot of each of the substances.

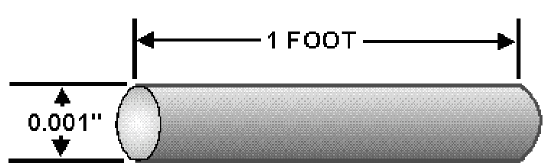


Figure 1-3.—Circular-mil-foot.

In working with square or rectangular conductors, such as ammeter shunts and bus bars, you may sometimes find it more convenient to use a different unit volume. A bus bar is a heavy copper strap or bar used to connect several circuits together. Bus bars are used when a large current capacity is required. Unit volume may be measured as the centimeter cube. Specific resistance, therefore, becomes the resistance offered by a cube-shaped conductor 1centimeter in length and 1 square centimeter in cross-sectional area. The unit of volume to be used is given in tables of specific resistances.

SPECIFIC RESISTANCE OR RESISTIVITY

Specific resistance, or resistivity, is the resistance in ohms offered by a unit volume (the circular-mil-foot or the centimeter cube) of a substance to the flow of electric current. Resistivity is the reciprocal of conductivity. A substance that has a high resistivity will have a low conductivity, and vice versa. Thus, the specific resistance of a substance is the resistance of a unit volume of that substance.

Many tables of specific resistance are based on the resistance in ohms of a volume of a substance 1 foot in length and 1 circular mil in cross-sectional area. The temperature at which the resistance measurement is made is also specified. If you know the kind of metal a conductor is made of, you can obtain the specific resistance of the metal from a table. The specific resistances of some common substances are given in table 1-1.

Table 1-1.—Specific Resistances of Common Substances

Substance	Specific resistance at 20ºC.	
	Centimeter cube (microhoms)	Circular-mil-foot (ohms)
Silver	1.629	9.8
Copper (drawn)	1.724	10.37
Gold	2.44	14.7
Aluminum	2.828	17.02
Carbon (amorphous)	3.8 to 4.1	……………………
Tungsten	5.51	33.2
Brass	7.0	42.1
Steel (soft)	15.9	95.8
Nichrome	109.0	660.0

The resistance of a conductor of a uniform cross section varies directly as the product of the length and the specific resistance of the conductor, and inversely as the cross-sectional area of the conductor. Therefore, you can calculate the resistance of a conductor if you know the length, cross-sectional area, and specific resistance of the substance. Expressed as an equation, the "R" (resistance in ohms) of a conductor is

$$R = \rho \frac{L}{A}$$

Where:

ρ = (Greek rho) the specific resistance in ohms per circular-mil-foot (refer to table 1-1)

L = length in feet

A = the cross-sectional area in circular mils

Problem:

What is the resistance of 1,000 feet of copper wire having a cross-sectional area of 10,400 circular mils (No. 10 wire) at a temperature of 20º C?

Solution:

The specific resistance of copper (table 1-1) is 10.37 ohms. Substituting the known values in the preceding equation, the resistance, R, is determined as

Given: $\rho = 10.37$ ohms

$L = 1{,}000$ ft

$A = 10{,}400$ circular mils

Solution:

$$R = \rho \frac{L}{A} = 10.37 \times \frac{1{,}000}{10{,}400}$$

$$= 1 \text{ ohm (approximately)}$$

If R, ρ, and A are known, the length (L) can be determined by a simple mathematical transposition. This has many valuable applications. For example, when locating a ground in a telephone line, you will use special test equipment. This equipment operates on the principle that the resistance of a line varies directly with its length. Thus, the distance between the test point and a fault can be computed accurately.

Q7. Define specific resistance.

Q8. List the three factors used to calculate resistance of a particular conductor in ohms.

WIRE SIZES

The most common method for measuring wire size in the Navy is by using the American Wire Gauge (AWG). An exception is aircraft wiring, which varies slightly in size and flexibility from AWG standards. For information concerning aircraft wire sizes, refer to the proper publications for specific aircraft. Only AWG wire sizes are used in the following discussion.

Wire is manufactured in sizes numbered according to the AWG tables. The various wires (solid or stranded) and the material they are made from (copper, aluminum, and so forth) are published by the National Bureau of Standards. An AWG table for copper wire is shown at table 1-2. The wire diameters become smaller as the gauge numbers become larger. Numbers are rounded off for convenience but are accurate for practical application. The largest wire size shown in the table is 0000 (read "4 naught"), and the smallest is number 40. Larger and smaller sizes are manufactured, but are not commonly used by the Navy. AWG tables show the diameter in mils, circular mil area, and area in square inches of AWG wire sizes. They also show the resistance (ohms) per thousand feet and per mile of wire sizes at specific temperatures. The last column shows the weight of the wire per thousand feet. An example of the use of table 1-2 is as follows.

Table 1-2.—Standard Solid Copper (American Wire Gauge)

Gage number	Diameter (mils)	Cross Section		Ohms per 1,000 ft		Ohms per mile 25ºC. = 77ºF.	Pounds per 1,000 ft.
		Circular mils	Square inches	25ºC. = 77ºF.	65ºC. = 149ºF.		
0000	460.0	212,000.0	0.166	0.0500	0.0577	0.264	641.0
000	410.0	168,000.0	.132	.0630	.0727	.333	508.0
00	365.0	133,000.0	.105	.0795	.0917	.420	403.0
0	325.0	106,000.0	.0829	.100	.116	.528	319.0
1	289.0	83,700.0	.0657	.126	.146	.665	253.0
2	258.0	66,400.0	.0521	.159	.184	.839	201.0
3	229.0	52,600.0	.0413	.201	.232	1.061	159.0
4	204.0	41,700.0	.0328	.253	.292	1.335	126.0
5	182.0	33,100.0	.0260	.319	.369	1.685	100.0
6	162.0	26,300.0	.0206	.403	.465	2.13	79.5
7	144.0	20,800.0	.0164	.508	.586	2.68	63.0
8	128.0	16,500.0	.0130	.641	.739	3.38	50.0
9	114.0	13,100.0	.0103	.808	.932	4.27	39.6
10	102.0	10,400.0	.00815	1.02	1.18	5.38	31.4
11	91.0	8,230.0	.00647	1.28	1.48	6.75	24.9
12	81.0	6,530.0	.00513	1.62	1.87	8.55	19.8
13	72.0	5,180.0	.00407	2.04	2.36	10.77	15.7
14	64.0	4,110.0	.00323	2.58	2.97	13.62	12.4
15	57.0	3,260.0	.00256	3.25	3.75	17.16	9.86
16	51.0	2,580.0	.00203	4.09	4.73	21.6	7.82
17	45.0	2,050.0	.00161	5.16	5.96	27.2	6.20
18	40.0	1,620.0	.00128	6.51	7.51	34.4	4.92
19	36.0	1,290.0	.00101	8.21	9.48	43.3	3.90
20	32.0	1,020.0	.000802	10.4	11.9	54.9	3.09
21	28.5	810.0	.000636	13.1	15.1	69.1	2.45
22	25.3	642.0	.000505	16.5	19.0	87.1	1.94
23	22.6	509.0	.000400	20.8	24.0	109.8	1.54
24	20.1	404.0	.000317	26.2	30.2	138.3	1.22
25	17.9	320.0	.000252	33.0	38.1	174.1	0.970
26	15.9	254.0	.000200	41.6	48.0	220.0	0.769
27	14.2	202.0	.000158	52.5	60.6	277.0	0.610
28	12.6	160.0	.000126	66.2	76.4	350.0	0.484
29	11.3	127.0	.0000995	83.4	96	440.0	0.384
30	10.0	101.0	.0000789	105.0	121.0	554.0	0.304
31	8.9	79.7	.0000626	133.0	153.0	702.0	0.241
32	8.0	63.2	.0000496	167.0	193.0	882.0	0.191
33	7.1	50.1	.0000394	211.0	243.0	1,114.0	0.152
34	6.3	39.8	.0000312	266.0	307.0	1,404.0	0.120
35	5.6	31.5	.0000248	335.0	387.0	1,769.0	0.0954
36	5.0	25.0	.0000196	423.0	488.0	2,230.0	0.0757
37	4.5	19.8	.0000156	533.0	616.0	2,810.0	0.0600
38	4.0	15.7	.0000123	673.0	776.0	3,550.0	0.0476
39	3.5	12.5	.0000098	848.0	979.0	4,480.0	0.0377
40	3.1	9.9	.0000078	1,070.0	1,230.0	5,650.0	0.0299

Problem: You are required to run 2,000 feet of AWG 20 solid copper wire for a new piece of equipment. The temperature where the wire is to be run is 25º C (77º F). How much resistance will the wire offer to current flow?

Solution: Under the gauge number column, find size AWG 20. Now read across the columns until you reach the "ohms per 1,000 feet for 25º C (77º F)" column. You will find that the wire will offer 10.4 ohms of resistance to current flow. Since we are using 2,000 feet of wire, multiply by 2.

10.4 ohms $\times$ 2 = 20.8 ohms

An American Standard Wire Gauge (figure 1-4) is used to measure wires ranging in size from number 0 to number 36. To use this gauge, insert the wire to be measured into the smallest slot that will just accommodate the bare wire. The gauge number on that slot indicates the wire size. The front part of the slot has parallel sides, and this is where the wire measurement is taken. It should not be confused with the larger semicircular opening at the rear of the slot. The rear opening simply permits the free movement of the wire all the way through the slot.

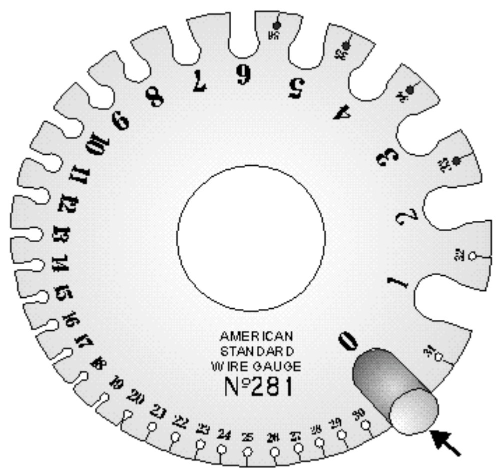


Figure 1-4.—Wire gauge.

Q9. Using table 1-2, determine the resistance of 1,500 feet of AWG 20 wire at 25º C.

Q10. When using an American Standard Wire Gauge to determine the size of a wire, where should you place the wire in the gauge to get the correct measurement?

STRANDED WIRES AND CABLES

A wire is a single slender rod or filament of drawn metal. This definition restricts the term to what would ordinarily be understood as "solid wire." The word "slender" is used because the length of a wire is usually large when compared to its diameter. If a wire is covered with insulation, it is an insulated wire. Although the term "wire" properly refers to the metal, it also includes the insulation.

A conductor is a wire suitable for carrying an electric current.

A stranded conductor is a conductor composed of a group of wires or of any combination of groups of wires. The wires in a stranded conductor are usually twisted together and not insulated from each other.

A cable is either a stranded conductor (single-conductor cable) or a combination of conductors insulated from one another (multiple-conductor cable). The term "cable" is a general one and usually applies only to the larger sizes of conductors. A small cable is more often called a stranded wire or cord (such as that used for an iron or a lamp cord). Cables may be bare or insulated. Insulated cables may be sheathed (covered) with lead, or protective armor. Figure 1-5 shows different types of wire and cable used in the Navy.

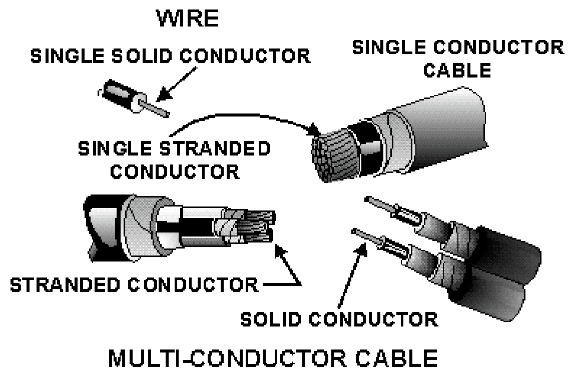


Figure 1-5.—Conductors.

Conductors are stranded mainly to increase their flexibility. The wire strands in cables are arranged in the following order:

The first layer of strands around the center conductor is made up of six conductors. The second layer is made up of 12 additional conductors. The third layer is made up of 18 additional conductors, and so on. Thus, standard cables are composed of 7, 19, and 37 strands, in continuing fixed increments. The overall flexibility can be increased by further stranding of the individual strands.

Figure 1-6 shows a typical cross section of a 37-strand cable. It also shows how the total circular-mil cross-sectional area of a stranded cable is determined.

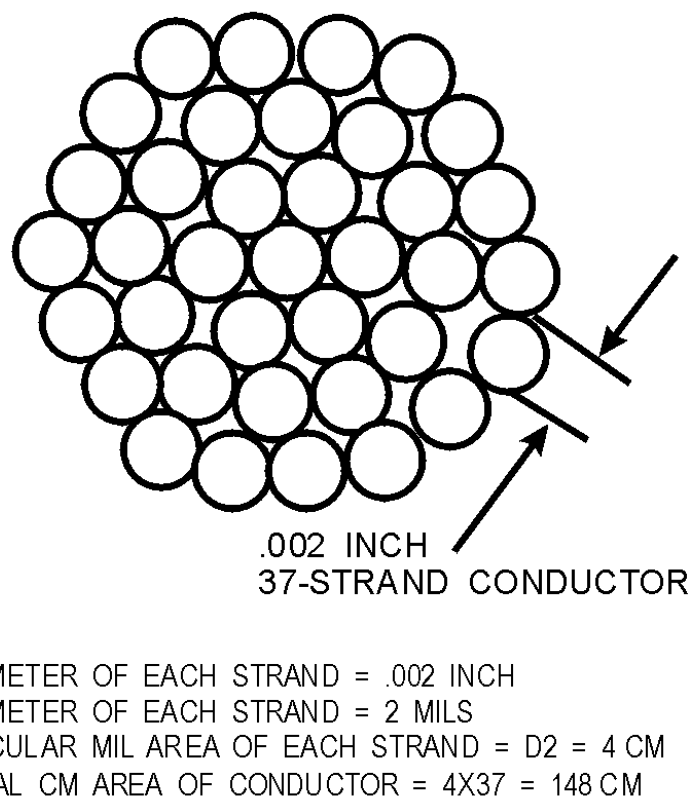


Figure 1-6.—Stranded conductor.

SELECTION OF WIRE SIZE

Several factors must be considered in selecting the size of wire to be used for transmitting and distributing electric power. These factors will be discussed throughout this section. Military specifications cover the installation of wiring in aircraft, ships, and electrical/electronic equipment. These specifications describe the technical requirements for material purchased from manufacturers by the Department of Defense. An important reason for having these specifications is to ensure uniformity of sizes to reduce the danger of fires caused by the improper selection of wire sizes. Wires can carry only a limited amount of current safely. If the current flowing through a wire exceeds the current-carrying capacity of the wire, excess heat is generated. This heat may be great enough to burn off the insulation around the wire and start a fire.

FACTORS GOVERNING THE CURRENT RATING

The current rating of a cable or wire indicates the current capacity that the wire or cable can safely carry continuously. If this limit, or current rating, is exceeded for a length of time, the heat generated may burn the insulation. The current rating of a wire is used to determine what size is needed for a given load, or current drain.

The factors that determine the current rating of a wire are the conductor size, the location of the wire in a circuit, the type of insulation, and the safe current rating. Another factor that will be discussed later in this chapter is the material the wire is made of. As you have already seen, these factors also affect the resistance in ohms of a wire-carrying current.

CONDUCTOR SIZE

An increase in the diameter, or cross section, of a wire conductor decreases its resistance and increases its capacity to carry current. An increase in the specific resistance of a conductor increases its resistance and decreases its capacity to carry current.

WIRE LOCATION

The location of a wire in a circuit determines the temperature under which it operates. A wire may be located in a conduit or laced with other wires in a cable. Because it is confined, the wire operates at a higher temperature than if it were open to the free air. The higher the temperature under which a wire is operating, the greater will be its resistance. Its capacity to carry current is also lowered. Note that, in each case, the resistance of a wire determines its current-carrying capacity. The greater the resistance, the more power it dissipates in the form of heat energy.

Conductors may also be installed in locations where the ambient (surrounding) temperature is relatively high. When this is the case, the heat generated by external sources is an important part of the total conductor heating. This heating factor will be explained further when we discuss temperature coefficient. We must understand how external heating influences how much current a conductor can carry. Each case has its own specific limitations. The maximum allowable operating temperature of insulated conductors is specified in tables. It varies with the type of conductor insulation being used.

INSULATION

The insulation of a wire does not affect the resistance of the wire. Resistance does, however, determine how much heat is needed to burn the insulation. As current flows through an insulated conductor, the limit of current that the conductor can withstand depends on how hot the conductor can get before it burns the insulation. Different types of insulation will burn at different temperatures. Therefore, the type of insulation used is the third factor that determines the current rating of a conductor. For instance, rubber insulation will begin deteriorating at relatively low temperatures, whereas varnished cloth insulation retains its insulating properties at higher temperatures. Other types of insulation are fluorinated ethylene propylene (FEP), silicone rubber, or extruded polytetrafluoroethylene. They are effective at still higher temperatures.

SAFE CURRENT RATINGS

The National Board of Fire Underwriters prepares tables showing the safe current ratings for sizes and types of conductors covered with various types of insulation. The allowable current-carrying capacities of single copper conductors in free air at a maximum room temperature of 30º C (86º F) are given in table 1-3. At ambient temperatures greater than 30º C, these conductors would have less current-carrying capacity.

Table 1-3.—Temperature Ratings and Current-Carrying Capacities (in Amperes) of Some Single Copper Conductors at Ambient Temperatures of 30ºC

Size	Moisture Resistant Rubber or Thermoplastic	Varnished Cambric or Heat Resistant Thermoplastic	Silicone Rubber or Fluorinated Ethylene Propylene (FEP)	Polytetra-Fluoroethylene
0000	300	385	510	850
000	260	330	430	725
00	225	285	370	605
0	195	245	325	545
1	165	210	280	450
2	140	180	240	390
3	120	155	210	335
4	105	135	180	285
6	80	100	135	210
8	55	70	100	115
10	40	55	75	110
12	25	40	55	80
14	20	30	45	60

Q11. List the four factors you should use to select wire for a specified current rating.

Q12. What are three types of nonmetallic insulating materials that can be used in a high-temperature environments?

Q13. State why it is important for you to consider the ambient (surrounding) temperature of a conductor when selecting wire size.

COPPER-VERSUS-ALUMINUM CONDUCTORS

Although silver is the best conductor, its cost limits its use to special circuits. Silver is used where a substance with high conductivity or low resistivity is needed.

The two most commonly used conductors are copper and aluminum. Each has positive and negative characteristics that affect its use under varying circumstances. A comparison of some of the characteristics of copper and aluminum is given in table 1-4.

Table 1-4.—Comparative Characteristics of Copper and Aluminum

CHARACTERISTICS	COPPER	ALUMINUM
Tensile strength (lb/in2).	55,000	25,000
Tensile strength for same conductivity (lb).	55,000	40,000
Weight for same conductivity (lb).	100	48
Cross section for same conductivity (C.M.).	100	160
Specific resistance (Ω/mil ft).		

Copper has a higher conductivity than aluminum. It is more ductile (can be drawn out). Copper has relatively high tensile strength (the greatest stress a substance can bear along its length without tearing apart). It can also be easily soldered. However, copper is more expensive and heavier than aluminum.

Although aluminum has only about 60 percent of the conductivity of copper, its lightness makes long spans possible. Its relatively large diameter for a given conductivity reduces corona. Corona is the discharge of electricity from the wire when it has a high potential. The discharge is greater when smaller diameter wire is used than when larger diameter wire is used. However, the relatively large size of aluminum for a given conductance does not permit the economical use of an insulation covering.

Q14. State two advantages of using aluminum wire for carrying electricity over long distances.

Q15. State four advantages of copper over aluminum as a conductor.

TEMPERATURE COEFFICIENT

The resistance of pure metals, such as silver, copper, and aluminum, increases as the temperature increases. However, the resistance of some alloys, such as constantan and manganin, changes very little as the temperature changes. Measuring instruments use these alloys because the resistance of the circuits must remain constant to get accurate measurements.

In table 1-1, the resistance of a circular-mil-foot of wire (the specific resistance) is given at a specific temperature, 20º C in this case. It is necessary to establish a standard temperature. As we stated earlier, the resistance of pure metals increases with an increase in temperature. Therefore, a true basis of comparison cannot be made unless the resistances of all the substances being compared are measured at the same temperature. The amount of increase in the resistance of a 1-ohm sample of the conductor per degree rise in temperature above 0º C is called the temperature coefficient of resistance. For copper, the value is approximately 0.00427 ohm.

A length of copper wire having a resistance of 50 ohms at an initial temperature of 0º C will have an increase in resistance of 50×0.00427, or 0.214 ohms. This applies to the entire length of wire and for each degree of temperature rise above 0º C. A 20º C increase in resistance is approximately 20×0.214, or 4.28 ohms. The total resistance at 20º C is $50 + 4.28$, or 54.28 ohms.

Q16. Define the temperature coefficient of resistance.

Q17. What happens to the resistance of copper when it is heated?

CONDUCTOR INSULATION

To be useful and safe, electric current must be forced to flow only where it is needed. It must be "channeled" from the power source to a useful load. In general, current-carrying conductors must not be allowed to come in contact with one another, their supporting hardware, or personnel working near them.

To accomplish this, conductors are coated or wrapped with various materials. These materials have such a high resistance that they are, for all practical purposes, nonconductors. Nonconductors are generally referred to as "insulators" or "insulating material."

Only the necessary minimum amount of insulation is applied to any particular type of conductor designed to do a particular job. This is done because of several factors. The expense, stiffening effect, and a variety of physical and electrical conditions under which the conductors are operated must be taken into account. Therefore, there are a variety of insulated conductors available to meet the requirements of any job.

Two fundamental properties of insulating materials (that is, rubber, glass, asbestos, or plastic) are insulation resistance and dielectric strength. These are two entirely different and distinct properties.

INSULATION RESISTANCE

Insulation resistance is the resistance to current leakage through the insulation materials. Insulation resistance can be measured with a megger without damaging the insulation. Information so obtained is a useful guide in appraising the general condition of insulation. Clean, dry insulation having cracks or other faults may show a high value of insulation resistance but would not be suitable for use.

DIELECTRIC STRENGTH

Dielectric strength is the ability of an insulator to withstand potential difference. It is usually expressed in terms of the voltage at which the insulation fails because of the electrostatic stress. Maximum dielectric strength values can be measured only by raising the voltage of a TEST SAMPLE until the insulation breaks down.

Q18. Compare the resistance of a conductor to that of an insulator.

Q19. State two fundamental properties of insulating materials.

Q20. Define insulation resistance.

Q21. Define dielectric strength.

Q22. How is the dielectric strength of an insulator determined?

TYPES OF INSULATION

The insulating materials discussed in the next paragraphs are commonly used in Navy electrical and electronic equipment.

Rubber

One of the most common types of insulation is rubber. The voltage that may be applied to a rubber-covered conductor is dependent on the thickness and the quality of the rubber covering. Other factors being equal, the thicker the insulation, the higher may be the applied voltage. Rubber insulation is normally used for low- or medium-range voltage. Figure 1-7 shows two types of rubber-covered wire. One is a two-conductor cable in which each stranded conductor is covered with rubber insulation; the other is a single, solid conductor. In each case, the rubber serves the same purpose: to confine the current to its conductor.

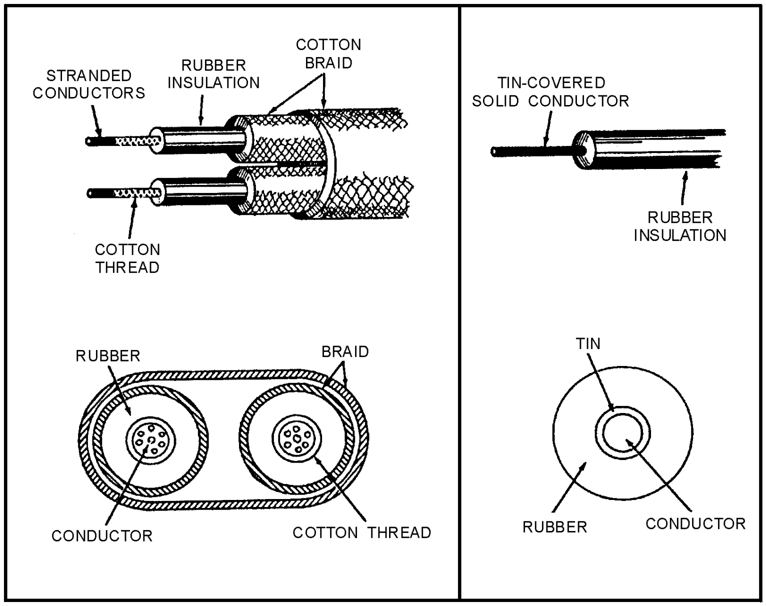


Figure 1-7.—Rubber insulation.

Referring to the enlarged cross-sectional view in figure 1-7, note that a thin coating of tin separates the copper conductor from the rubber insulation. If the thin coating of tin were not used, a chemical action would take place and the rubber would become soft and gummy where it makes contact with the copper. When small, solid, or stranded conductors are used, a winding of cotton threads is applied between the conductors and the rubber insulation.

CODE-GRADED RUBBER.—Code-graded rubber is the standard that the National Electrical Code (NEC) has adopted as the minimum requirements for rubber insulation as specified by Underwriters' Laboratories. In this code system, the letter R indicates the use of a rubber insulator. Type R signifies that the wire is rubber coated.

The NEC codes Type RH and Type RHH signify a rubber heat-resistant compound. Type RW signifies a rubber moisture-resistant compound. A Type RHW signifies a rubber heat- and moisture-resistant compound. Type RHW is approved for use in wet or dry locations at a maximum conductor temperature of 75º C. Neoprene, a low-voltage compound, is the one exception to Type RHW. Although not a rubber compound, neoprene meets the requirements of Underwriters' Laboratories and was designated Type RHW.

LATEX RUBBER.—Latex rubber is a high-grade compound consisting of 90 percent unmilled grainless rubber. There are two designations for this type of insulation: Type RUH and Type RUW. Type RUH (rubber unmilled heat-resistant) is used in dry locations when the conductor temperature does not exceed 75º C. Type RUW (rubber unmilled moisture-resistant) is used in wet locations when the conductor does not exceed 60º C.

SILICONE.—Silicone is a rubber compound that does not carry the "R" designator for many of its applications. An example of this is Type SA (silicone-asbestos). In Type SA, the insulator around the conductor is silicone rubber, but the outer covering must consist of heavy glass, asbestos-glass, or asbestos braiding treated with a heat, flame, and moisture-resistant compound.

Q23. What is the purpose of coating a copper conductor with tin when rubber insulation is used?

Plastics

Plastic is one of the more commonly used types of insulating material for electrical conductors. It has good insulating, flexibility, and moisture-resistant qualities. Although there are many types of plastic insulating materials, thermoplastic is one of the most common. With the use of thermoplastic, the conductor temperature can be higher than with some other types of insulating materials without damage to the insulating quality of the material. Plastic insulation is normally used for low- or medium-range voltage.

The designators used with thermoplastics are much like those used with rubber insulators. The following letters are used when dealing with NEC type designators for thermoplastics:

Letter	Designates
T	Thermoplastic
H	Heat-resistant
W	Moisture-resistant
A	Asbestos
N	Outer nylon jacket
M	Oil-resistant

For example, a NEC designator of Type THWN would indicate thermoplastic heat- and moisture-resistant with an outer nylon jacket.

Varnished Cambric

Varnished cambric insulation can withstand much higher temperatures than rubber insulation. Varnished cambric is cotton cloth that has been coated with an insulating varnish. Figure 1-8 shows a cable covered with varnished cambric insulation. The varnished cambric is in tape form and is wound around the conductor in layers. An oily compound is applied between each layer of the tape to prevent water from seeping through the insulation. It also acts as a lubricant between the layers of tape, so they will slide over each other when the cable is bent.

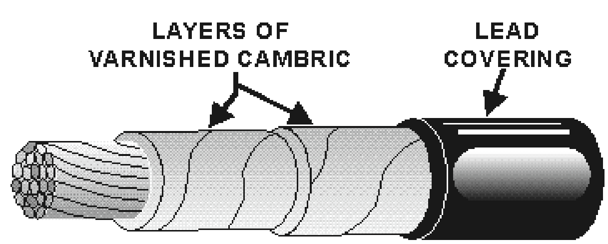


Figure 1-8.—Varnished cambric insulation.

Cambric insulation is used on extremely high-voltage conductors used in substations and powerhouses. It is also used in other locations subjected to high temperatures. In addition, it is used on the coils and leads of high-voltage generators. Transformer leads also use this insulation because it is unaffected by oils or grease and has high dielectric strength. Varnished cambric and paper insulation for cables are the two types of insulating materials most widely used at voltages above 15,000 volts. Such cable is always lead covered to keep out moisture.

Extruded Polytetrafluoroethylene

Extruded polytetrafluoroethylene is a high-temperature insulation used extensively in aircraft and equipment installations. It will not burn, but will vaporize when subjected to intense heat. Conductors for high temperatures use a nickel coating rather than tin or silver to prevent oxidation. Nickel-coated wire is more difficult to solder, but makes satisfactory connections with proper soldering techniques.

WARNING

Avoid breathing the vapors from extruded polytetrafluoroethylene insulation when it is heated. Symptoms of overexposure are dizziness or headaches. These symptoms disappear upon exposure to fresh air.

Q24. What safety precaution should you take when working with extruded polytetrafluoroethylene insulated wiring?

Fluorinated Ethylene Propylene (FEP)

FEP has properties similar to extruded polytetrafluoroethylene, but will melt at soldering temperatures. It is rated at 200º C and is, therefore, considered a high-temperature insulation. There are no known toxic vapors from FEP. Common-sense practice, however, requires that you provide adequate ventilation during any soldering operation.

Asbestos

Asbestos insulation was used extensively in the past for high-temperature insulation. Today, it is seldom used by the Navy. Many naval ships and aircraft, however, still contain asbestos-insulated wiring. Aboard ship, this is particularly true in galley and laundry equipment. The reason for discontinuing the use of asbestos as an insulator is that breathing asbestos fibers can produce severe lung damage. It can render you disabled or cause fatal fibrosis of the lungs. Asbestos is also a factor in the development of cancer in the gastrointestinal tract. Safety precautions concerning asbestos will be covered in more detail at the end of chapter 3.

WARNING

Avoid inhalation of asbestos fibers. Asbestos fibers have been found to cause severe lung damage (asbestosis) and cancer of the gastrointestinal tract. Follow Navy safety precautions when working with all asbestos products.

One type of asbestos-covered wire is shown in figure 1-9. It consists of stranded copper conductors covered with felted asbestos. The wire is, in turn, covered with asbestos braid. This type of wire is used in motion-picture projectors, arc lamps, spotlights, heating element leads, and so forth.

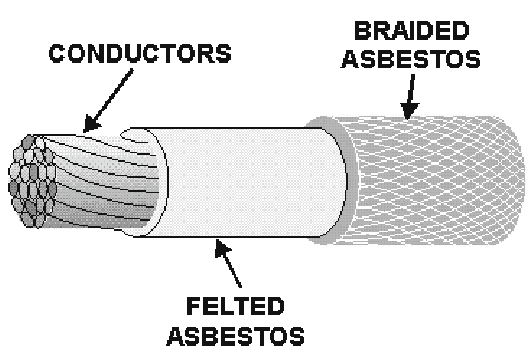


Figure 1-9.—Asbestos Insulation.

Another type of asbestos-covered cable is shown in figure 1-10. It is combination of asbestos and varnished cambric. This type of insulation serves as leads for motors and transformers that sometimes must operate in hot, damp locations. The varnished cambric covers the inner layer of felted asbestos. This prevents moisture from reaching the innermost layer of asbestos. Asbestos loses its insulating properties when it becomes wet. It will, in fact, become a conductor. Varnished cambric prevents this from happening because it resists moisture. Although this insulation will withstand some moisture, it should not be used on conductors that may at times be partially immersed in water. Under those circumstances, the insulation must be protected with an outer lead sheath.

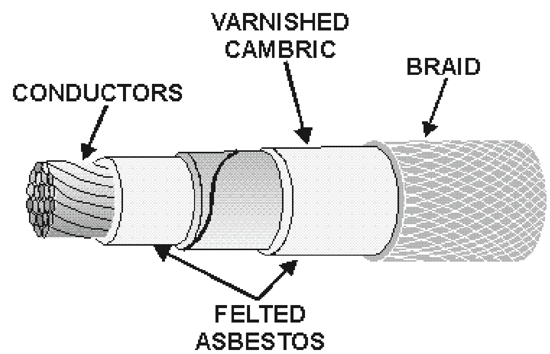


Figure 1-10.—Asbestos and varnished cambric insulation.

The NEC has designators for eight types of asbestos wire. The designators and a description of each are listed below.

Type A	Nonimpregnated asbestos without an asbestos braid
Type AA	Nonimpregnated asbestos with an outer asbestos braid or glass
Type AI	Impregnated asbestos without an asbestos braid
Type AIA	Impregnated asbestos with an outer asbestos braid or glass
Type AVA	Asbestos, varnish-cambric insulation with an outer asbestos braid or glass
Type AVL	Asbestos, varnish-cambric insulation with an outer asbestos braid covered with a lead sheath
Type AVB	Asbestos, varnish-cambric insulation with an outer flame-retardant cotton braid
Type SA	Silicone rubber insulated with outer heavy glass, asbestos-glass, or asbestos braid

Q25. State the reasons that the Navy is getting away from the use of asbestos insulation.

Q26. State what happens to the insulating characteristics of asbestos when it gets wet.

Paper

Paper has little insulation value alone. However, when impregnated with a high grade of mineral oil, it serves as a satisfactory insulation for extremely high-voltage cables. The oil has a high dielectric strength, and tends to prevent breakdown of the paper insulation. The paper must be thoroughly saturated with the oil. The thin paper tape is wrapped in many layers around the conductors, and then soaked with oil.

The three-conductor cable shown in figure 1-11 consists of paper insulation on each conductor. It has a spirally wrapped nonmagnetic metallic tape over the insulation. The space between conductors is filled with a suitable spacer to round out the cable. Another nonmagnetic metal tape is used to secure the entire cable. Over this, a lead sheath is applied. This type of cable is used on voltages from 10,000 volts to 35,000 volts.

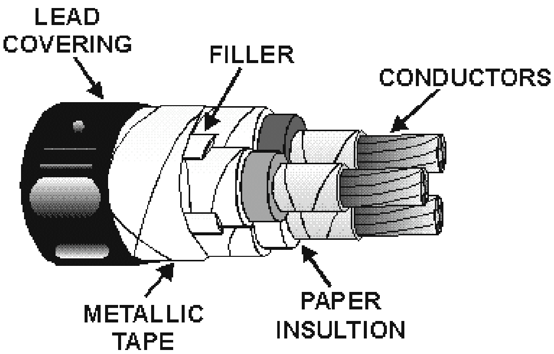


Figure 1-11.—Paper-insulated power cables.

Q27. What are the most common insulators used for extremely high voltages?

Silk and Cotton

In certain types of circuits (for example, communications circuits), a large number of conductors are needed, perhaps as many as several hundred. Figure 1-12 shows a cable containing many conductors. Each is insulated from the others by silk and cotton thread. Because the insulation in this type of cable is not subjected to high voltage, the use of thin layers of silk and cotton is satisfactory.

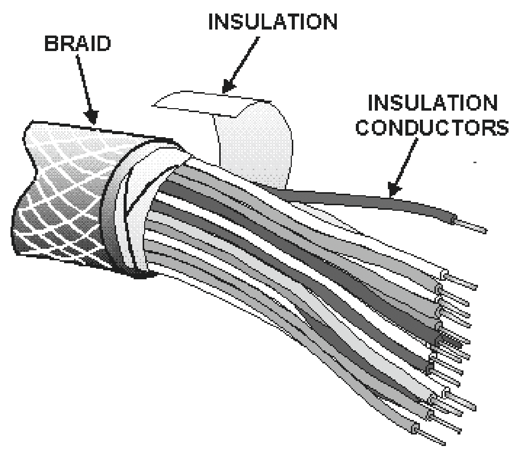


Figure 1-12.—Silk and cotton Insulation.

Silk and cotton insulation keeps the size of the cable small enough to be handled easily. The silk and cotton threads are wrapped around the individual conductors in reverse directions. The covering is then impregnated with a special wax compound.

Enamel

The wire used on the coils of meters, relays, small transformers, motor windings, and so forth, is called magnet wire. This wire is insulated with an enamel coating. The enamel is a synthetic compound of cellulose acetate (wood pulp and magnesium). In the manufacturing process, the bare wire is passed through a solution of hot enamel and then cooled. This process is repeated until the wire acquires from 6 to 10 coatings. Thickness for thickness, enamel has higher dielectric strength than rubber. It is not practical for large wires because of the expense and because the insulation is readily fractured when large wires are bent.

Figure 1-13 shows an enamel-coated wire. Enamel is the thinnest insulating coating that can be applied to wires. Hence, enamel-insulated magnet wire makes smaller coils. Enameled wire is sometimes covered with one or more layers of cotton to protect the enamel from nicks, cuts, or abrasions.

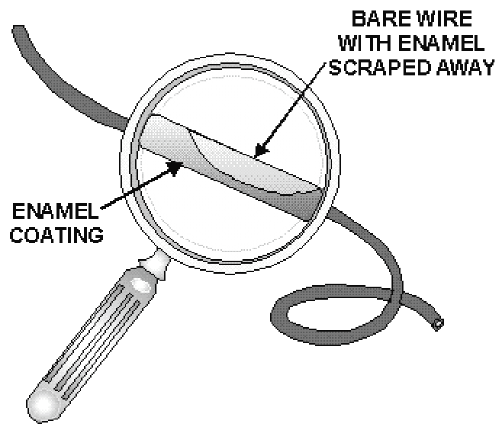


Figure 1-13.—Enamel Insulation.

Q28. What is the common name for enamel-insulated wire?

Mineral Insulated

Mineral-insulated (MI) cable was developed to meet the needs of a noncombustible, high heat-resistant, and water-resistant cable. MI cable has from one to seven electrical conductors. These conductors are insulated in a highly compressed mineral, normally magnesium oxide, and sealed in a liquidtight, gastight metallic tube, normally made of seamless copper (figure 1-14).

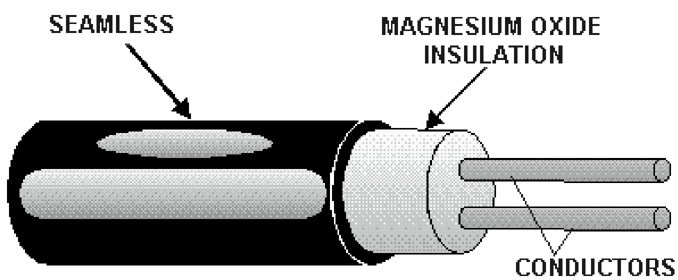


Figure 1-14.—Two-conductor mineral-insulated (MI) cable.

CONDUCTOR PROTECTION

Wires and cables are generally subject to abuse. The type and amount of abuse depends on how and where they are installed and the manner in which they are used. Cables buried directly in the ground must resist moisture, chemical action, and abrasion. Wires installed in buildings must be protected against mechanical injury and overloading. Wires strung on crossarms on poles must be kept far enough apart so that the wires do not touch. Snow, ice, and strong winds make it necessary to use conductors having high tensile strength and substantial frame structures.

Generally, except for overhead transmission lines, wires or cables are protected by some form of covering. The covering may be some type of insulator like rubber or plastic. Over this, additional layers of fibrous braid or tape may be used and then covered with a finish or saturated with a protective coating. If the wire or cable is installed where it is likely to receive rough treatment, a metallic coat should be added.

The materials used to make up the protection for a wire or cable are grouped into one of two categories: nonmetallic or metallic.

Q29. If a cable is installed where it receives rough treatment, what should be added?

NONMETALLIC PROTECTION

The category of nonmetallic protective coverings is divided into three areas. These areas are (1) according to the material used as the covering, (2) according to the saturant in which the covering was impregnated, and (3) according to the external finish on the wire or cable. These three areas reflect three different methods of protecting the wire or cable. These methods allow some wire or cable to be classified under more than one category. Most of the time, however, the wire or cable will be classified based upon the material used as the covering regardless of whether or not a saturant or finish is applied.

Many types of nonmetallic materials are used to protect wires and cables. Fibrous braid is by far the most common and will be discussed first.

Fibrous Braid

Fibrous braid is used extensively as a protective covering for cables. This braid is woven over the insulation to form a continuous covering without joints (figure 1-15). The braid is generally saturated with asphalt, paint, or varnish to give added protection against moisture, flame, weathering, oil, or acid. Additionally, the outside braid is often given a finish of stearin pitch and mica flakes, paint, wax, lacquer, or varnish depending on the environment where the cable is to be used.

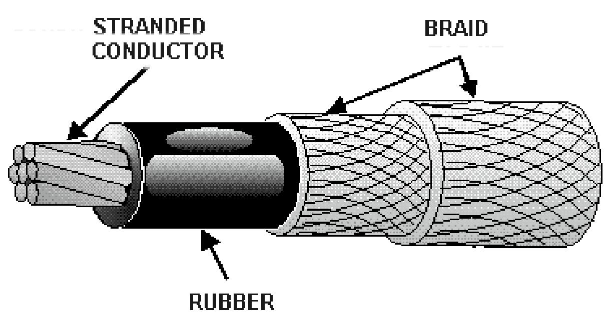


Figure 1-15.—Fibrous braid covering.

The most common type of fibrous braid is woven from light, standard, or heavy cotton yarn. Cotton yarn comes in different colors, which allows color-coding of the individual conductors. Cotton will not withstand all the possible environments in which a cable may be laid. Other materials currently being used to make fibrous braid are glazed cotton, seine twine or hawser cord, hemp, paper and cotton, jute, asbestos, silk, rayon, and fibrous glass. The choice of which material to use depends on the strength needed in the cable as well as how resistant it must be to its environment.

Fibrous Tape

Fibrous tape coverings are frequently used as a part of the protective covering of cables. The material of tape coverings is made into the tape before application to the cable. The material in yarns for braid covering is woven into fabric during the application to the cable. When tape covering is used, it is wrapped helically around the cable with each turn overlapping the previous turn.

The most common types of fibrous tape are rubber-filled cloth tape and a combination of cotton cloth and rubber. Except for duct tape, tape covering is never used as the outer covering on a cable. Tape coverings are used directly over the insulation of individual conductors and for the inner covering over the assembled conductors of a multiconductor cable. Frequently, tape coverings are used under the sheath of a lead-sheathed cable. Duct tape, which is made of heavy canvas webbing saturated with an asphalt compound, is often used over a lead-sheathed cable for protection against corrosion and mechanical injury.

Q30. How many categories of nonmetallic protective coverings are there?

Q31. What is the most common type of nonmetallic material used to protect wires and cables?

Q32. What are the most common types of fibrous tape?

Woven Covers

Woven covers, commonly called loom, are used when exceptional abrasion-resistant qualities are required. These covers are composed of thick, heavy, long-fibered cotton yarns woven around the cable in a circular loom, much like that used on a fire hose. They are not braids, although braid covering are also woven; they are designated differently.

Rubber and Synthetic Coverings

Rubber and synthetic coverings are not standardized. Different manufactures have their own special compounds designated by individual trade names. These compounds are different from the rubber compounds used to insulate cable. These compounds have been perfected not for insulation qualities but for resistance to abrasion, moisture, oil, gasoline, acids, earth solutions, and alkalies. None of these coverings will provide protection against all types of exposure. Each covering has its own particular limitations and qualifications.

Jute and Asphalt Coverings

Jute and asphalt coverings are commonly used as a cushion between cable insulation and metallic armor. Frequently, they are also used as a corrosive-resistant covering over a lead sheath or metallic armor. Jute and asphalt coverings consist of asphalt-impregnated jute yarn heli-wrapped around the cable or of alternate layers of asphalt-impregnated jute yarn. These coverings serve as a weatherproofing.

Unspun Felted Cotton

Unspun felted cotton is commonly used only in special classes of service. It is made as a solid felted covering for a cable.

Q33. What materials are commonly used as cushions between cable insulation and metallic armor?

METALLIC PROTECTION

Metallic protection is of two types: sheath or armor. As with all wires and cables, the type of protection needed will depend on the environment where the wire or cable will be used.

Metallic Sheath

Cables or wires that are continually subjected to water must be protected by a watertight cover. This watertight cover is either a continuous metal jacket or a rubber sheath molded around the cable.

Figure 1-16 is an example of a lead-sheathed (jacketed) cable used in power work. This cable is a standard three-conductor type. Each conductor is insulated and then wrapped in a layer of rubberized tape. The conductors are twisted together, and rope or fillers are added to form a round core. Over this is wrapped a second layer of tape called a serving. Finally, a lead sheath is molded around the cable.

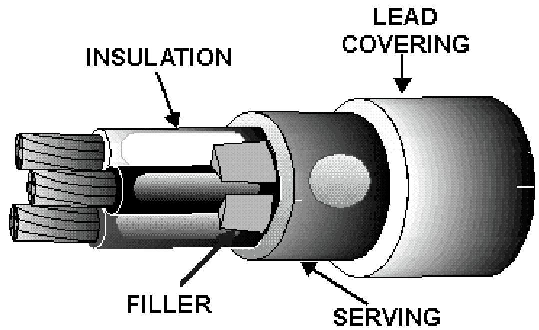


Figure 1-16.—Lead-sheathed cable.

Lead-sheathed cable is one of three types currently being used: alloy lead, pure lead, and reinforced lead. An alloy-lead sheath is much like a pure lead sheath but is manufactured with 2-percent tin. This alloy is more resistant to gouging and abrasion during and after installation. Reinforced lead sheath is used mainly for oil-filled cables where high internal pressures can be expected. Reinforced lead sheath consists of a double lead sheath. A thin tape of hard-drawn copper, bronze, or other elastic metal (preferably nonmagnetic) is wrapped around the inner sheath. This tape gives considerable additional strength and elasticity to the sheath, but must be protected from corrosion. For this reason, a second lead sheath is applied over the tape.

Metallic Armor

Metallic armor provides a tough protective covering for wires and cables. The type, thickness, and kind of metal used to make the armor depend on three factors: (1) the use of the conductors, (2) the environment where the conductors are to be used, and (3) the amount of rough treatment that is expected.

Figure 1-17 shows three examples of metallic armor cable: wire braid, steel tape, and wire armor.

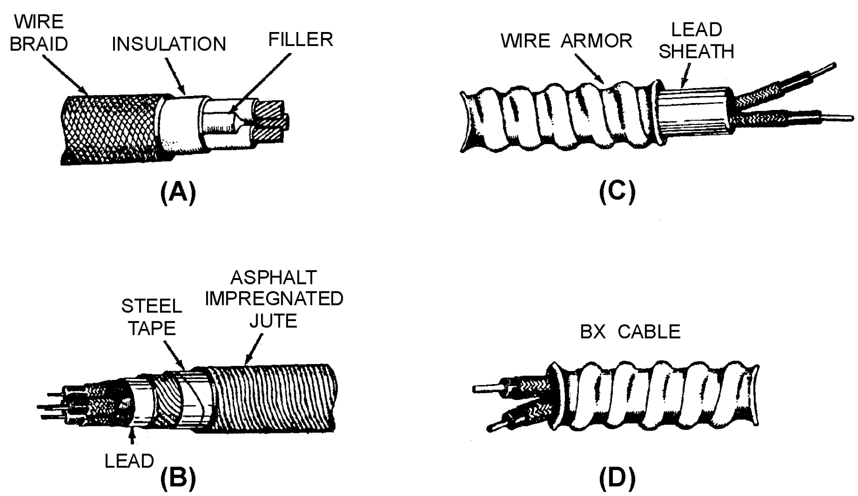


Figure 1-17.—Metallic armor cable.

WIRE-BRAID ARMOR.—Wire-braid armor (view A of figure 1-17), also known as basket-weave armor, is used when light and flexible protection is needed. Wire braid is constructed much like fibrous braid. The metal is woven directly over the cable as the outer covering. The metal used in this braid is galvanized steel, bronze, copper, or aluminum. Wire-braid armor is mainly for shipboard use.

STEEL TAPE.—A second type of metallic armor is steel tape. Steel tape covering (view B of figure 1-17) is wrapped around the cable and then covered with a serving of jute. There are two types of steel tape armor. The first is called interlocking armor. Interlocking armor is applied by wrapping the tape around the cable so that each turn is overlapped by the next and is locked in place. The second type is flat-band armor. Flat-band armor consists of two layers of steel tape. The first layer is wrapped around the cable but is not overlapped. The second layer is then wrapped around the cable covering the area that was not covered by the first layer.

WIRE ARMOR.—Wire armor is a layer of wound metal wire wrapped around the cable. Wire armor is usually made of galvanized steel and can be used over a lead sheath (see view C of figure 1-17). It can be used with the sheath as a buried cable where moisture is a concern, or without the sheath (view D of figure 1-17) when used in buildings.

Q34. What are the two types of metallic protection?

Q35. What are the three types of lead-sheathed cables?

Q36. What are the three examples of metallic armor cable that were discussed?

COAXIAL CABLE

Coaxial cable (figure 1-18) is defined as two concentric wires, cylindrical in shape, separated by a dielectric of some type. One wire is the center conductor and the other is the outer conductor. These conductors are covered by a protective jacket. The protective jacket is then covered by an outer protective armor.

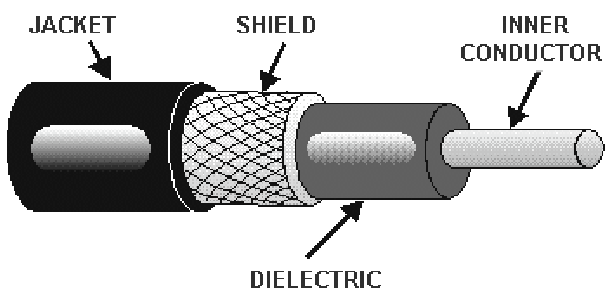


Figure 1-18.—Coaxial cable.

Coaxial cables are used as transmission lines and are constructed to provide protection against outside signal interference.

SUMMARY

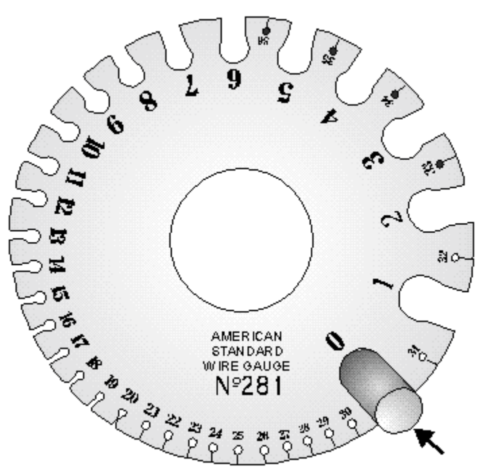
In this chapter you learned that conductors are the means for tying the various components of an electrical or electronic system together. Many factors determine the type of conductor to be used in a specific application. In order for you to compare the different types and sizes of conductors, we discussed the following factors:

Unit Size—The unit size of a conductor is the mil-foot. A mil-foot is a circular conductor 1foot long with a diameter of 1 mil (0.001 inch, or one-thousandth of an inch).

Conductor Sizes—The square mil and the circular mil are the units of measure used to determine the cross-sectional area of electrical conductors. The square mil, as it relates to a square conductor, is the cross-sectional area of a square conductor that has a side of 1 mil. The circular mil is the cross-sectional area of a circular conductor having a diameter of 1 mil. The circular mil area (CMA) of a conductor is computed by squaring the diameter of the circular conductor is mils. Thus, a wire having a diameter of 4 mils (0.004 inch) has a CMA of 4^2, or 16 circular mils. If the conductor is stranded, the CMA for a strand is computed, and the CMA for the conductor is computed by multiplying the CMA of the strand by the number of strands. The relationship of the square mil to the circular mil is determined by comparing the square mil area of a circular conductor having a diameter of 1 mil ($A = \pi r^2$) to the circular mil area of the same conductor (D^2). Therefore, there is 0.7854 square mil to 1 circular mil. There are more circular mils than square mils in a given area.

Specific Resistance—The specific resistance of a substance is the resistance in ohms offered by a unit volume (the circular-mil-foot) to the flow of electric current. The three factors that are used to calculate the specific resistance of a particular conductor are (1) its length, (2) its cross-sectional area, and (3) the specific resistance of a unit volume of the substance from which the conductor is made. The specific resistance for various sizes and lengths of standard solid copper wire can be determined by the use of tables.

Wire Gauge—A wire gauge is used to determine the American Standard Wire Gauge size of conductors. The measurement of a bare conductor is taken in the slot, not in the circular area at the bottom of the slot.



Selection of Wire Size—Four factors must be considered in selecting the proper wire size for a particular electrical circuit. These factors are (1) conductor size, (2) the material it's made of, (3) the location of the wire in the circuit, and (4) the type of insulation used. Some of the types of insulation used in a high-temperature environment are FEP, extruded polytetrafluoroethylene, and silicone rubber. The ambient (surrounding) temperature of a conductor is an important part of total conductor heating.

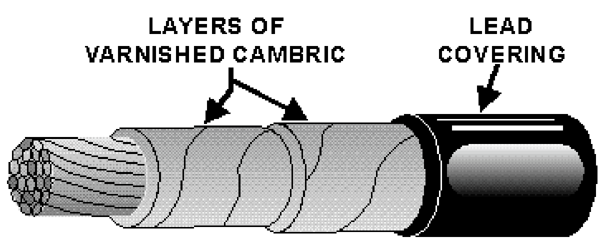
Copper-versus-Aluminum Conductors—The two most common metals used for electrical conductors are copper and aluminum. Some advantages of copper over aluminum as a conductor are that copper has higher conductivity, is more ductile, has a higher tensile strength, and can be easily soldered. Two advantages of aluminum wire for carrying electricity over long distances are its lightness and it reduces corona (the discharge of electricity from a wire at high potential).

Temperature Coefficient of Resistance—The temperature coefficient of resistance is the amount of increase in the resistance of a 1-ohm sample of a conductor per degree of temperature rise above 0º C. The resistance of copper and other pure metals increases with an increase in temperature.

Conductor Insulation—Insulators have a resistance that is so great that, for all practical purposes, they are nonconductors. Two fundamental properties of insulating materials are (1) insulation resistance and (2) the resistance to current leakage through the insulation. Dielectric strength is the ability of the insulation material to withstand potential difference. The dielectric strength of an insulator is determined by raising the voltage on a test sample until it breaks down.

Insulating Materials—Some common insulating materials have properties and safety precautions that should be remembered. These are:

- The purpose of coating a copper conductor with tin when rubber insulation is used is to prevent the insulation from deteriorating due to chemical action.
- When extruded polytetrafluoroethylene insulation is heated, caution should be observed not to breathe the vapors.
- The most commonly used insulating materials for extremely high-voltage conductors are varnished cambric and oil-impregnated paper.
- Magnet wire is the common name for enamel-insulated wire used in meters, relays, small transformers, motor windings, and so forth.
- The Navy is getting away from using asbestos insulation because asbestos fibers can cause lung disease and/or cancer.
- Asbestos insulation becomes a conductor when it gets wet.



Conductor Protection—There are several types of conductor protection in use. The type commonly used aboard Navy ships is wire-braid armor.

ANSWERS TO QUESTIONS Q1. THROUGH Q36.

A1. To allow comparisons between conductors of different sizes and resistance.

A2. 375 mils (move the decimal three places to the right).

A3. A circular conductor with a diameter of 1 mil and a length of 1 foot.

A4. The cross-sectional area of a square conductor with a side of 1 mil.

A5. The cross-sectional area of a circular conductor with a diameter of 1 mil.

A6. Circular mil area (CMA) = D^2 (in mils) $\times$ number of strands0.0004 inch = 4 mils (CMA) = $4^2 \times$ 19 (strands)(CMA) = 16×19 = 304 mils.

A7. The resistance of a unit volume of a substance.

A8. Length, cross-sectional area, and specific resistance of a unit volume of the substance from which the conductor is made.

A9. 1,000 ft = 10.4 ohms1,500 ft = 1.5×0.4 = 15.6 ohms

A10. In the parallel walled slot not the circular area.

A11. Conductor size, the material it is made of the location of the wire in a circuit, and the type of insulation used.

A12. FEP, extruded polytetrafluoroethylene, and silicone rubber.

A13. The heat surrounding the conductor is an important part of total conductor heating.

A14. It is light and reduces corona.

A15. It has higher conductivity, it is more ductile, it has relatively high tensile strength, and it can be easily soldered.

A16. The amount of increase in the resistance of a 1-ohm sample of the conductor per degree of temperature rise above 0º C

A17. It increases.

A18. Conductors have a very low resistance and insulators have a resistance that is so great that, for all practical purposes, they are nonconductors.

A19. Insulation resistance and dielectric strength.

A20. The resistance to current leakage through the insulation.

A21. The ability of the insulation material to withstand potential difference.

A22. By raising the voltage on a test sample until it breaks down.

A23. To prevent the rubber insulation from deteriorating due to chemical action.

A24. Avoid breathing the vapors when the insulation is heated.

A25. Breathing asbestos fibers can cause lung disease and/or cancer

A26. It will become a conductor.

A27. Varnished cambric and oil-impregnated paper.

A28. Magnet wire.

A29. Metallic coat.

A30. Three.

A31. Fibrous Braid.

A32. Rubber-filled cloth tape and a combination of cotton cloth and rubber.

A33. Jute and Asphalt coverings.

A34. Sheath and armor

A35. Alloy lead, pure lead, and reinforced lead.

A36. Wire braid, steel tape, and wire armor

CHAPTER 2

WIRING TECHNIQUES

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to:

1. State the basic requirements for any splice and terminal connection, including the preferred wire-stripping method.
2. State the reason the ends of the wire are clamped down after a Western Union splice has been made.
3. Explain the major advantage of the crimped terminal over the soldered terminal.
4. Name the two types of insulation commonly used for noninsulated splices and terminal lugs.
5. State an advantage of using preinsulated terminal lugs and the color code used for each.
6. Explain the procedures for crimping terminal lugs with a hand crimp tool.
7. Recall the physical description and operating procedures for the HT-900B/920B compressed air/nitrogen heating tool.
8. Recall the safety precautions for using the compressed air/nitrogen heating tool.
9. Recall the procedures, precautions, and tools associated with soldering.
10. Explain the procedures and precautions for tinning wire.
11. Recall the types of soldering irons and their uses.
12. State the purposes and required properties of flux.
13. State the purpose for lacing conductors.
14. Recall when double lacing of wire bundles is required.
15. Recall the requirements for using spot ties.

WIRING TECHNIQUES

This chapter will assist you in learning the basic skills of proper wiring techniques. It explains the different ways to terminate and splice electrical conductors. It also discusses various soldering techniques that will assist you in mastering the basic soldering skills. The chapter ends with a discussion of the procedure to be followed when you lace wire bundles within electrical and electronic equipment.

CONDUCTOR SPLICES AND TERMINAL CONNECTIONS

Conductor splices and connections are an essential part of any electrical circuit. When conductors join each other or connect to a load, splices or terminals must be used. Therefore, it is important that they be properly made. Any electrical circuit is only as good as its weakest link. The basic requirement of any splice or connection is that it be both mechanically and electrically as sound as the conductor or device with which it is used. Quality workmanship and materials must be used to ensure lasting electrical contact, physical strength, and insulation. The most common methods of making splices and connections in electrical cables is explained in the discussion that follows.

INSULATION REMOVAL

The preferred method of removing insulation is with a wire-stripping tool, if available. A sharp knife may also be used. Other typical wire strippers in use in the Navy are illustrated in figure 2-1. The hot-blade, rotary, and bench wire strippers (views A, B, and C, respectively) are usually found in shops where large wire bundles are made. When using any of these automatic wire strippers, follow the manufacturer's instructions for adjusting the machine; this avoids nicking, cutting, or otherwise damaging the conductors. The hand wire strippers are common hand tools found throughout the Navy. The hand wire strippers (view D of figure 2-1) are the ones you will most likely be using. Wire strippers vary in size according to wire size and can be ordered for any size needed.

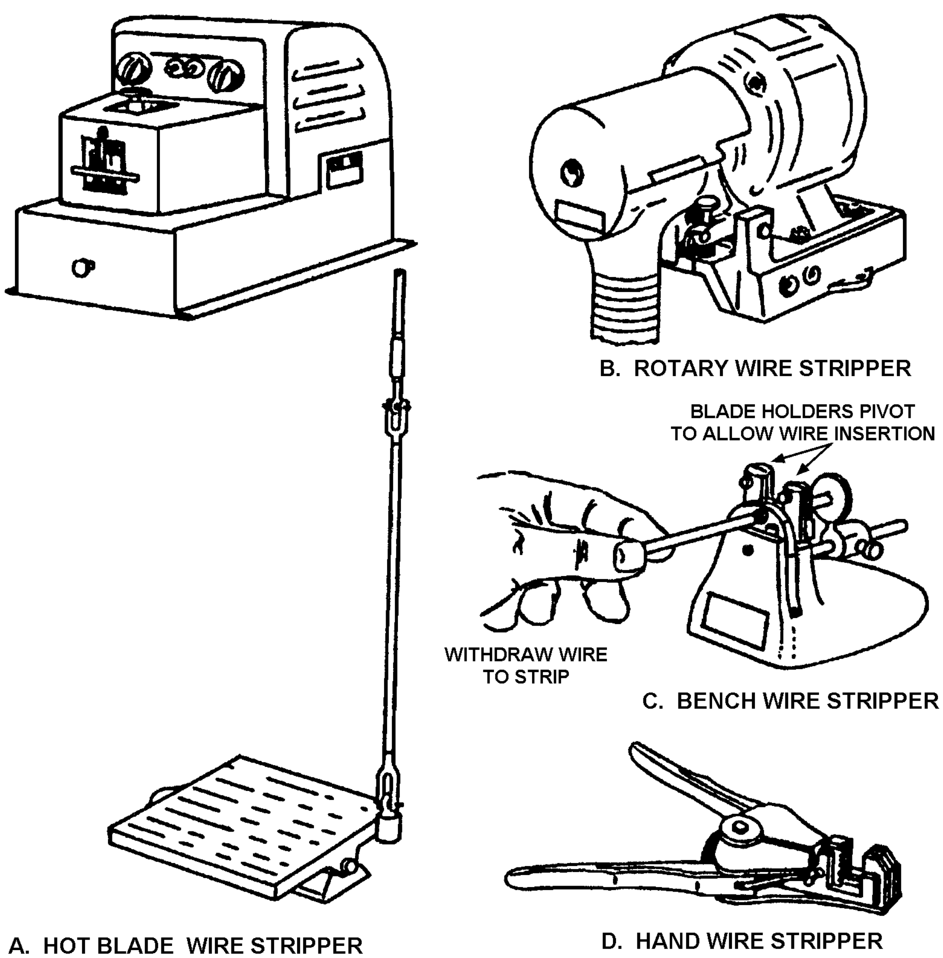


Figure 2-1.—Typical wire-stripping tools.

Hand Wire Stripper

The procedure for stripping wire with the hand wire stripper is as follows (refer to figure 2-2):

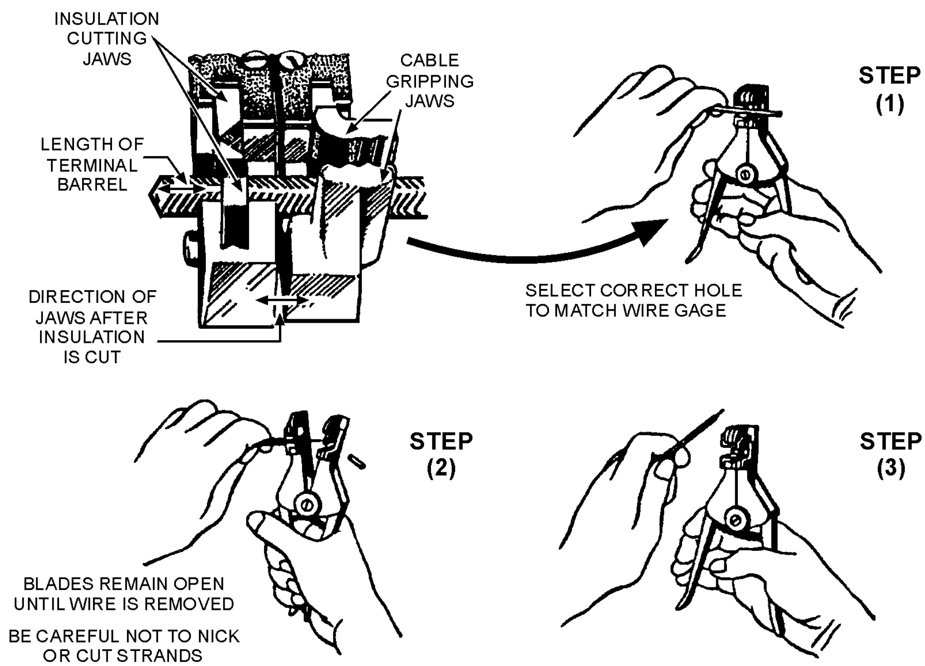


Figure 2-2.—Stripping wire with a hand stripper.

1. Insert the wire into the center of the correct cutting slot for the wire size to be stripped. The wire sizes are listed on the cutting jaws of the hand wire strippers beneath each slot.
2. After inserting the wire into the proper slot, close the handles together as far as they will go.
3. Slowly release the pressure on the handles so as not to allow the cutting blades to make contact with the stripped conductor. On some of the newer style hand wire strippers, the cutting jaws have a safety lock that helps prevent this from happening. Continue to release pressure until the gripper jaws release the stripped wire, then remove.

Knife Stripping

A sharp knife may be used to strip the insulation from a conductor. The procedure is much the same as for sharpening a pencil. The knife should be held at approximately a 60º angle to the conductor. Use extreme care when cutting through the insulation to avoid nicking or cutting the conductor. This procedure produces a taper on the cut insulation as shown in figure 2-3.

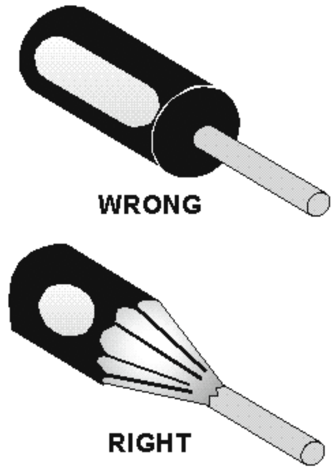


Figure 2-3.—Knife stripping.

Locally Made Hot-Blade Wire Stripper

If you are required to strip a large number of wires, you can use a locally made hot-blade stripper (figure 2-4) as follows:

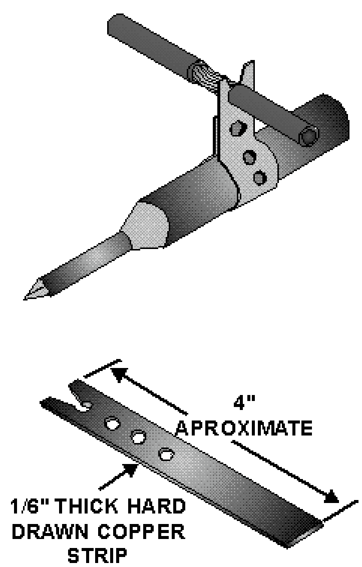


Figure 2-4.—Locally made hot-blade stripper.

1. In the end of a piece of copper strip, cut a sharp-edged "V." At the bottom of the "V," make a wire slot of suitable diameter for the size wire to be stripped.

2. Fasten the copper strip around the heating element of an electric soldering iron as shown in figure 2-4. The iron must be rated at 100 watts or greater in order to transfer enough heat to the copper strip to melt the wire insulation.

3. Lay the wire or cable to be stripped in the "V"; a clean channel will be melted in the insulation.

4. Remove the insulation with a slight pull.

General Wire-Stripping Instructions

When stripping wire with any of the tools mentioned, observe the following precautions:

1. Do not attempt to use a hot-blade stripper on wiring with glass braid or asbestos insulation. These insulators are highly heat resistant.

2. When using the hot-blade stripper, make sure the blades are clean. Clean the blades with a brass wire brush as necessary.

3. Make sure all stripping blades are sharp and free from nicks, dents, and so forth.

4. When using any type of wire stripper, hold the wire perpendicular to the cutting blades.

5. Make sure the insulation is clean-cut with no frayed or ragged edges; trim if necessary.

6. Make sure all insulation is removed from the stripped area. Some types of wire are supplied with a transparent layer between the conductor and the primary insulation. If this is present, remove it.

7. When the hand strippers are used to remove lengths of insulation longer than 3/4 inch, the stripping procedure must be done in two or more operations. The strippers will only strip about 3/4 inch at one time.

8. Retwist strands by hand, if necessary, to restore the natural lay and tightness of the strands.

9. Strip aluminum wires with a knife as described earlier. Aluminum wire should be stripped very carefully. Care should be taken not to nick the aluminum wire as the strands break very easily when nicked.

Q1. What are the basic requirements for any splice or terminal connection?

Q2. What is the preferred method for stripping wire?

Q3. What stripping tool would NOT be used to strip glass braid insulation?

Q4. What tool should be used to strip aluminum wire?

TYPES OF SPLICES

There are six commonly used types of splices. Each has advantages and disadvantages for use. Each splice will be discussed in the following section.

Western Union Splice

The Western Union splice joins small, solid conductors. Figure 2-5 shows the steps in making a Western Union splice.

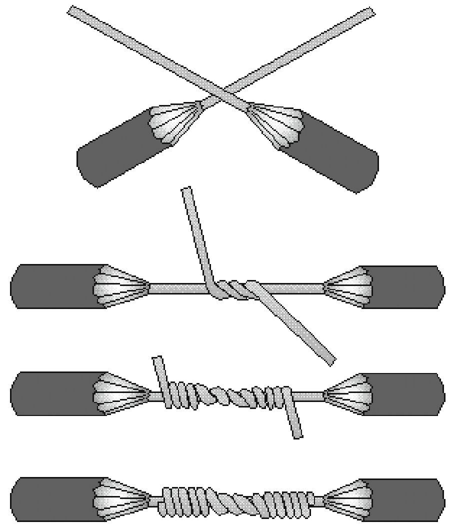


Figure 2-5.—Western Union splice.

1. Prepare the wires for splicing. Enough insulation is removed to make the splice. The conductor is cleaned.

2. Bring the wires to a crossed position and make a long twist or bend in each wire.

3. Wrap one end of the wire and then the other end four or five times around the straight portion of each wire.

4. Press the ends of the wires down as close as possible to the straight portion of the wire. This prevents the sharp ends from puncturing the tape covering that is wrapped over the splice. The various types of tape and their uses are discussed later in this chapter.

Staggering Splices

Joining small multiconductor cables often presents a problem. Each conductor must be spliced and taped. If the splices are directly opposite each other, the overall size of the joint becomes large and bulky. A smoother and less bulky joint can be made by staggering the splices.

Figure 2-6 shows how a two-conductor cable is joined to a similar size cable by using a Western Union splice and by staggering the splices. Care should be taken to ensure that a short wire from one side of the cable is spliced to a long wire, from the other side of the cable. The sharp ends are then clamped firmly down on the conductor. The figure shows a Western Union splice, but other types of splices work just as well.

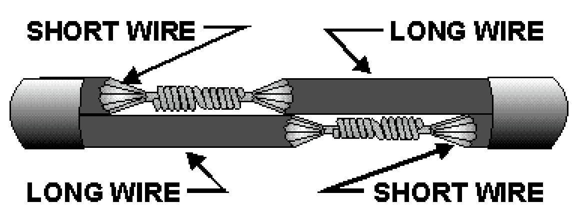


Figure 2-6.—Staggering splices.

Rattail Joint

A splice that is used in a junction box and for connecting branch circuits is the rattail joint (figure 2-7).

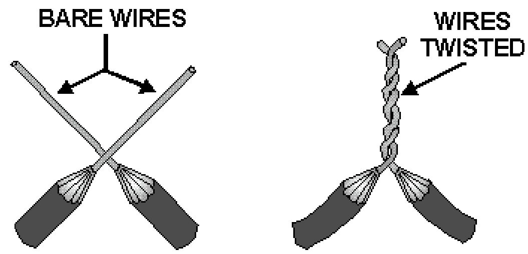


Figure 2-7.—Rattail joint.

Wiring that is installed in buildings is usually placed inside long lengths of steel or aluminum pipe called a conduit. Whenever branch or multiple circuits are needed, junction boxes are used to join the conduit.

To create a rattail joint, first strip the insulation off the ends of the conductors to be joined. You then twist the wires to form the rattail effect. This type of splice will not stand much stress.

Fixture Joint

The fixture joint is used to connect a small-diameter wire, such as in a lighting fixture, to a larger diameter wire used in a branch circuit. Like the rattail joint, the fixture joint will not stand much strain.

Figure 2-8 shows the steps in making a fixture joint. The first step is to remove the insulation and clean the wires to be joined. After the wires are prepared, the fixture wire is wrapped a few times around the branch wire. The end of the branch wire is then bent over the completed turns. The remainder of the bare fixture wire is then wrapped over the bent branch wire. Soldering and taping completes the job.

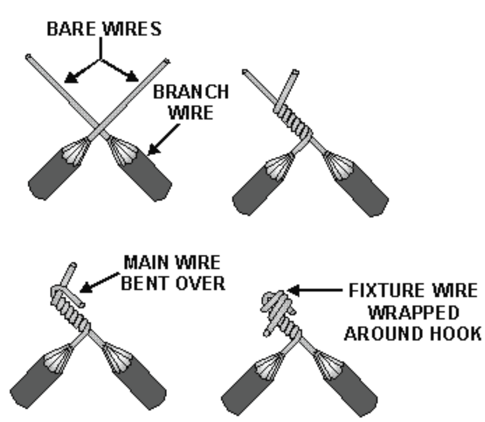


Figure 2-8.—Fixture joint.

Knotted Tap Joint

All the splices discussed up to this point are known as butted splices. Each was made by joining the free ends of the conductors together. Sometimes, however, it is necessary to join a branch conductor to a continuous wire called the main wire. Such a junction is called a tap joint.

The main wire, to which the branch wire is to be tapped, has about 1 inch of insulation removed. The branch wire is stripped of about 3 inches of insulation. The knotted tap is shown in figure 2-9.

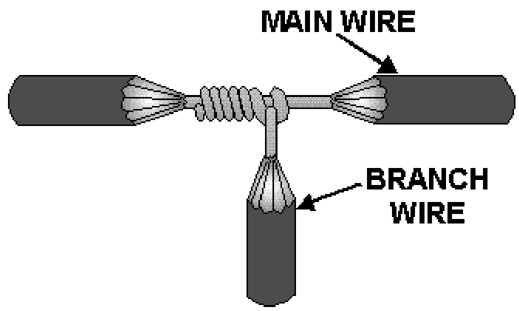


Figure 2-9.—Knotted tap joint.

The branch wire is laid behind the main wire. About three-fourths of the bare portion of the branch wire extends above the main wire. The branch wire is brought under the main wire, around itself, and then over the main wire to form a knot. The branch wire is then wrapped around the main conductor in short, tight turns; and the end is trimmed off.

The knotted tap is used where the splice is subject to strain or slippage. When there is no strain, the knot may be eliminated.

Wire Nut and Split Bolt Splices

The wire nut (view A of figure 2-10) is a device commonly used to replace the rattail joint splice. The wire nut is housed in plastic insulating material. To use the wire nut, place the two stripped conductors into the wire nut and twist the nut. In so doing, this will form a splice like the rattail joint and insulate itself by drawing the wire insulation into the wire nut insulation.

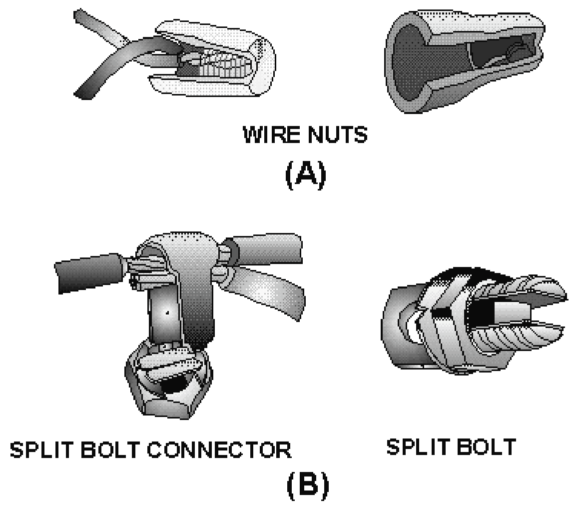


Figure 2-10.—Wire nut and split bolt splices.

The split bolt splice (view B of figure 2-10) is used extensively to join large conductors. In the illustration, it is shown replacing the knotted tap joint. The split bolt splice can also be used to replace the "butted" splices mentioned previously when using large conductors.

Q5. Why are the ends of the wire clamped down after a Western Union splice is made?

Q6. Why are splices staggered on multiconductor cables?

Q7. Where is the rattail joint normally used?

Q8. Which type of splice is used to splice a lighting fixture to a branch circuit?

SPLICE INSULATION

The splices we have discussed so far are usually insulated with tape. The following discussion will cover some characteristics of rubber, friction, and plastic insulation tapes.

Rubber Tape

Latex (rubber) tape is a splicing compound. It is used where the original insulation was rubber. The tape is applied to the splice with a light tension so that each layer presses tightly against the one beneath it. This pressure causes the rubber tape to blend into a solid mass. Upon completion, insulation similar to the original is restored.

In roll form, there is a layer of paper or treated cloth between each layer of rubber tape. This layer prevents the latex from fusing while still on the roll. The paper or cloth is peeled off and discarded before the tape is applied to the splice.

The rubber splicing tape should be applied smoothly and under tension so no air space exists between the layers. Start the first layer near the middle of the joint instead of the end. The diameter of the completed insulated joint should be somewhat greater than the overall diameter of the original wire, including the insulation.

WARNING

Some rubber tapes are made for special applications. These types are semiconducting and will pass electrical current, which presents a shock hazard. These types of tape are packaged similar to the latex rubber tape. Care should be taken to insulate splices only with latex rubber insulating tape.

Friction Tape

Putting rubber tape over the splice means that the insulation has been restored to a great degree. It is also necessary to restore the protective covering. Friction tape is used for this purpose. It also provides a minor degree of electrical insulation.

Friction tape is a cotton cloth that has been treated with a sticky rubber compound. It comes in rolls similar to rubber tape except that no paper or cloth separator is used. Friction tape is applied like rubber tape; however, it does not stretch.

The friction tape should be started slightly back on the original insulation. Wind the tape so that each turn overlaps the one before it. Extend the tape over onto the insulation at the other end of the splice. From this point, a second layer is wound back along the splice until the original starting point is reached. Cutting the tape and firmly pressing down the ends completes the job. When proper care is taken, the splice and insulation can take as much abuse as the rest of the original wire.

Plastic Electrical Tape

Plastic electrical tape has come into wide use in recent years. It has certain advantages over rubber and friction tape. For example, it can withstand higher voltages for a given thickness. Single thin layers of certain plastic tape will withstand several thousand volts without breaking down. However, to provide an extra margin of safety, several layers are usually wound over the splice. The extra layers of thin tape add very little bulk. The additional layers of plastic tape provide the added protection normally furnished by friction tape.

Plastic electrical tape usually has a certain amount of stretch so that it easily conforms to the contour of the splice.

Q9. Which of the splices discussed is NOT a butted splice?

Q10. Why is friction tape used in splicing?

TERMINAL LUGS

Since most cable wires are stranded, it is necessary to use terminal lugs to hold the strands together to aid in fastening the wires to terminal studs (see figure 2-11). The terminals used in electrical wiring are either of the soldered or crimped type. Terminals used in repair work must be of the size and type specified on the electrical wiring diagram for the particular equipment.

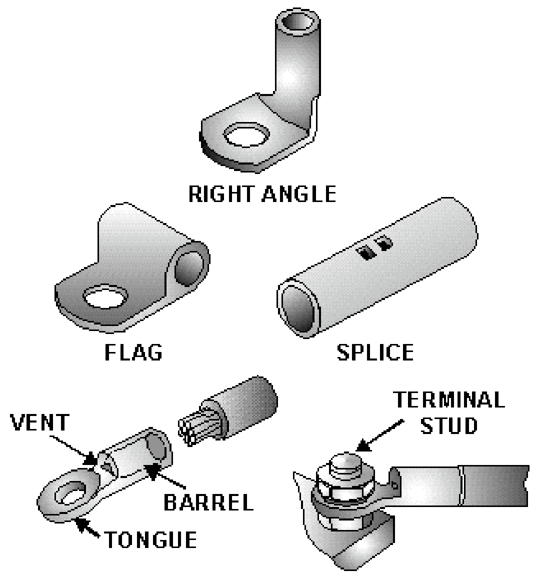


Figure 2-11.—Noninsulated terminal lugs and splices.

The increased use of crimp-on terminals is due to the limitations of soldered terminals. The quality of soldered connections depends mostly upon the operator's skill. Other factors, such as temperature, flux, cleanliness, oxides, and insulation damage due to heat, also add to defective connections. Solder-type connections are covered later in this chapter.

An advantage of the crimp-on solderless terminal lugs is that they require relatively little operator skill to use. Another advantage is that the only tool needed is the crimping tool. This allows terminal lugs to be applied with a minimum of time and effort. The connections are made rapidly, are clean, and uniform in construction. Because of the pressures exerted and the material used, the crimped connection or splice, properly made, is both mechanically and electrically sound. Some of the basic types of terminals are shown in figure 2-11. There are several variations of these basic types, such as the use of a slot instead of a terminal hole, three- and four-way splice-type connectors, and others.

Since the Navy uses both copper and aluminum wiring, both copper and aluminum terminals are necessary. Various size terminal or stud holes may be found for each of the different wire sizes. A further refinement of the solderless terminals and splices is the insulated type. The barrel of the terminal or splice is enclosed in an insulated material. The insulation is compressed along with the terminal barrel when it is crimped, but is not damaged in the process. This rids you of the need for taping or tying an insulating sleeve over the joint.

There are several different types of crimping tools used with copper terminals. However, you will normally be concerned only with wire sizes AWG (American Wire Gauge) 10 or smaller. For wire of these sizes, a small plier-type crimper is used to crimp on uninsulated terminals, as shown in figure 2-12. The small plier-type crimper has several sizes of notches for the different size terminals. Care should be used to select the correct size crimping tool for the particular terminal.

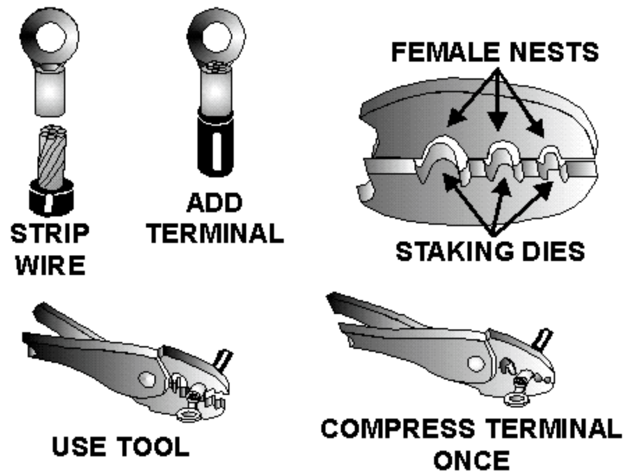


Figure 2-12.—Crimping small copper uninsulated terminals.

NONINSULATED TERMINAL AND SPLICE INSULATION

When noninsulated terminals and splices are used, some form of insulation must be used to cover the bare conductor. The two most common forms of insulator used for terminals and splices are transparent tubing (commonly called spaghetti) and heat-shrinkable tubing. If spaghetti is used, it must be tied with lacing twine, as illustrated in figure 2-13. Heat-shrinkable tubing is shrunk to the desirable size by applying dry heat. It is also a good way to insulate terminals and splices, as illustrated in figure 2-14. This tubing shrinks to approximately one-half its original diameter when heated with an electrical hot-air gun (figure 2-15). Here are the steps for using the hot-air gun:

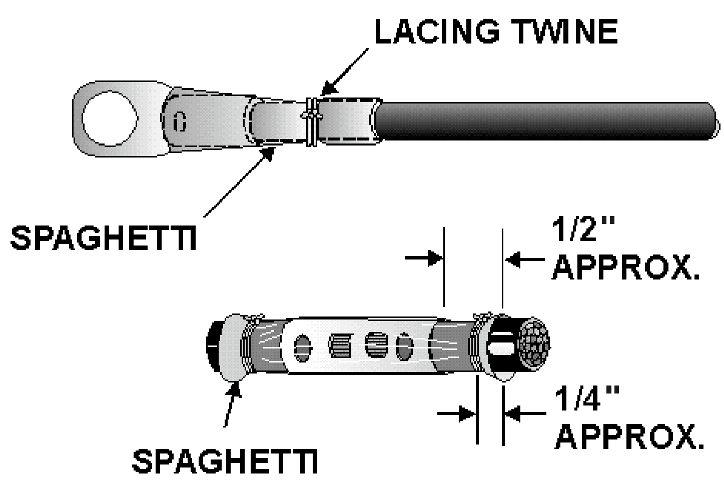


Figure 2-13.—Spaghetti tied with lacing twine.

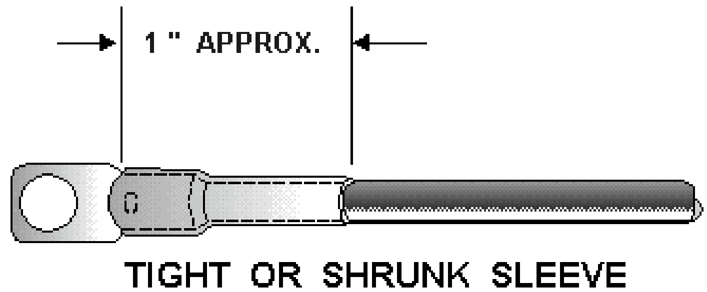


Figure 2-14.—Shrunken sleeve.

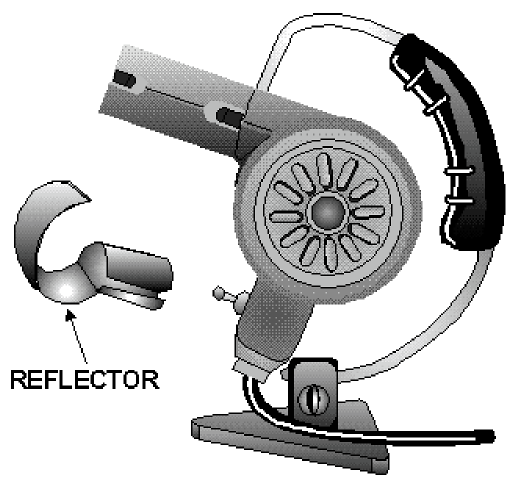


Figure 2-15.—Typical hot-air gun.

1. Hold the heat source 4 to 5 inches away from the wire. Apply a heat of 275º F to 300º F for about 30 seconds. Rotate the wire while applying the heat so that the heat is evenly distributed.

2. Remove the heat as soon as the tubing conforms to the shape of the wire. Allow the tubing to cool for at least 30 seconds before handling.

CAUTION

Do not apply heat higher than 300º F as this may damage the wire. Do not continue to apply heat after the tubing has shrunk onto the wire. Further application of heat will not cause additional shrinkage of the tubing.

COMPRESSED AIR/NITROGEN HEATING TOOL

The compressed air/nitrogen heating tool (figure 2-16) is a new tool in the fleet and was designed as a portable source of heat. This tool is safe for use around fueled aircraft because an open heating element is not required. The compressed air/nitrogen heating tool can be used on heat-shrinkable tubing.

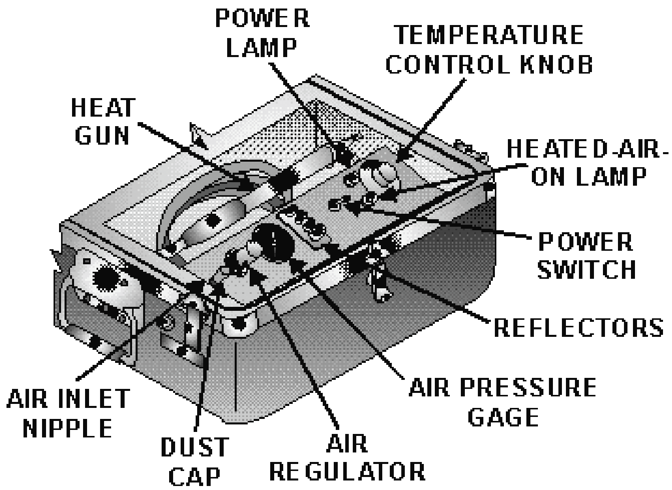


Figure 2-16.—Compressed air/nitrogen heating tool.

The compressed air/nitrogen heating tool comes in two styles: ac or dc electrical power supplies. The power requirements are listed in table 2-1 for both styles.

Table 2-1.—Compressed Air/Nitrogen Heating Tool Power Requirements

Electrical Power, HT-900B	115 VAC, 50-400 Hz, single-phase, 7 Amps
Electrical Power, HT-920B	220 VAC, 50-400 Hz, single-phase, 3.5 Amps
Heat gun output temperature	550-920°F (290-495°C)
Compressed air/nitrogen	80-200 psig, 4 SCFM (Dry and oil-free)

Refer to the operator's manual for safe operating procedures for the compressed air/nitrogen heating tool. A brief summary of these procedures follows:

1. Push down and fully turn the air regulator knob counterclockwise. This is to ensure that the air regulator is off.

2. Remove the dust cap from the air inlet nipple. The inlet nipple is what we connect the air or nitrogen source line to.

WARNING

If nitrogen is used, make sure that you are in a well-ventilated area. Using nitrogen is a poorly ventilated area. Using nitrogen is a poorly ventilated area can result in suffocation.

CAUTION

As noted in table 2-1, the compressed air/nitrogen source CANNOT be greater than 200 psig.

3. Attach the air/nitrogen hose to the inlet nipple, making sure there is a firm connection.

4. Once the air/nitrogen source is properly attached, push down and turn the air regulator knob clockwise until the pressure on the air pressure gauge indicates between 5 to 7 psig.

5. Plug in the power cord to an appropriate grounded power supply.

6. Set the power switch to the ON position. The power lamp and heated-air-on lamp will both illuminate. (If the lights do not come on, check the switch on the gun handle. The switch must be positioned toward the front of the handle.)

7. There is a 1-minute warm-up time. During this warm-up period, ensure that the indicated air pressure increases to 10 to 15 psig on the air-pressure gauge.

8. You can now adjust the temperature control knob to the desired temperature setting.

9. You can turn the air/nitrogen pressure off and on to the gun without powering down the module by using the switch mounted on the gun handle.

After you complete your task with the compressed air/nitrogen heating tool, use the following shutdown procedures:

1. Push down and fully turn the air regulator knob counterclockwise. Observe that the air pressure gauge indication drops to 0 psig and the heated air lamp goes out.

2. Position the switch on the heating gun toward the rear of the handle.

3. Place the power switch to the OFF position and observe that the power lamp goes out.

4. Allow the air/nitrogen to flow for a minimum of 1 minute to cool the heating gun. (This procedure is done to extend the life of the heating element.)

5. Disconnect the power connecter from the power source.

6. Turn off air/nitrogen source at place of origin and disconnect.

7. Disconnect the compressed air/nitrogen hose from the air inlet nipple and install the dust cap on the air inlet nipple.

Noninsulated Copper Terminals

The procedure for crimping a copper terminal (noninsulated) to a copper wire is as follows:

1. With a wire stripper, trim the insulation from the wire about one thirty-second of an inch longer than the length of the terminal barrel. When using a wire stripper, be sure to use the correct size stripping slot for the wire size used. Otherwise, all the insulation will not be removed or, if the slot is too small, the outside strands of the conductor will be nicked and consequently weakened. When a knife is used for stripping wire, care should be used to prevent nicking the strands. Slip the spaghetti or heat-shrinkable tubing over the wire and back far enough to be out of the way of the crimping operation.

2. Slip the terminal barrel over the bared wire end and up against the insulation. Make certain that all wire strands are inside the tubular barrel of the terminal.

3. Center the terminal barrel in the female nest of the plier jaws as shown in figure 2-12 so that the indentation formed by the staking die will be in the center of the barrel. Crimp until the pliers reach their stop or limit. This is necessary for a good mechanical and electrical connection.

4. Slip the tubular insulation down over the terminal barrel so that it extends a little beyond the barrel. Tie it in place if spaghetti is used. If heat-shrinkable tubing is used, shrink with a heat gun.

Q11. What is a major advantage of the crimped terminal over the soldered terminal?

Q12. What are the two types of insulation most commonly used for noninsulated splices and terminal lugs?

Q13. What is the maximum allowable temperature that should be used on heat-shrinkable tubing?

Q14. What is the maximum allowable source pressure that can be used with the compressor air/nitrogen heating tool?

ALUMINUM TERMINALS AND SPLICES

Terminals that are used with aluminum wire are made of aluminum. Proper crimping is more difficult with these terminals because of such factors as aluminum creep and softness. Aluminum wire has an undesirable characteristic called aluminum creep. Aluminum has the tendency to actually move away from the point where pressure is applied. This is not only true during the crimping operation but also takes place during temperature changes. The aluminum wire is softer than the terminal lugs and splice connectors and contracts faster than the connector when the temperature drops. This causes the wires to creep away from the crimped connections, which, in turn, causes loose connections. The softness of aluminum wire also makes it subject to being cut or nicked during stripping. You should be careful never to use an aluminum terminal with copper wire or a copper terminal with aluminum wire because of electrolysis. Electrolysis is the chemical action that takes place when an electric current passes through two dissimilar metals. This chemical action corrodes (eats away) the metal. Also, never use the aluminum crimping tool for crimping other than the aluminum terminals. Aluminum terminal lugs and splices are not insulated, so you must use spaghetti or heat-shrinkable tubing for insulation as discussed earlier.

The barrels of several styles of larger size aluminum terminal lugs are filled with a petroleum abrasive compound. This compound causes a grinding action during the crimping operation. This removes the oxide film from the aluminum. It also prevents the oxide film from reforming in the connection. All aluminum terminals and splices have an inspection hole to allow checking the depth of wire insertion. This hole is sealed with a removable plug, which also serves to hold in the oxide-inhibiting compound (figure 2-17).

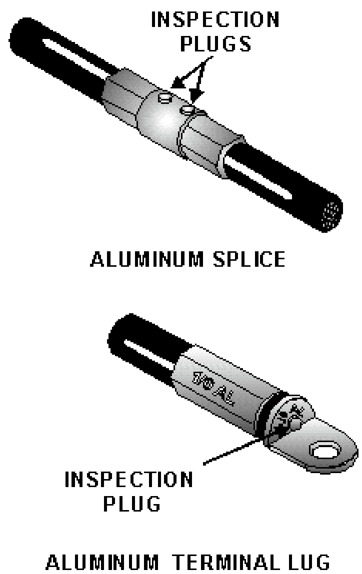


Figure 2-17.—Aluminum terminal lug and splice.

It is recommended that only power-operated crimping tools be used to install large aluminum terminal lugs and splices. (See view A of figure 2-18.)

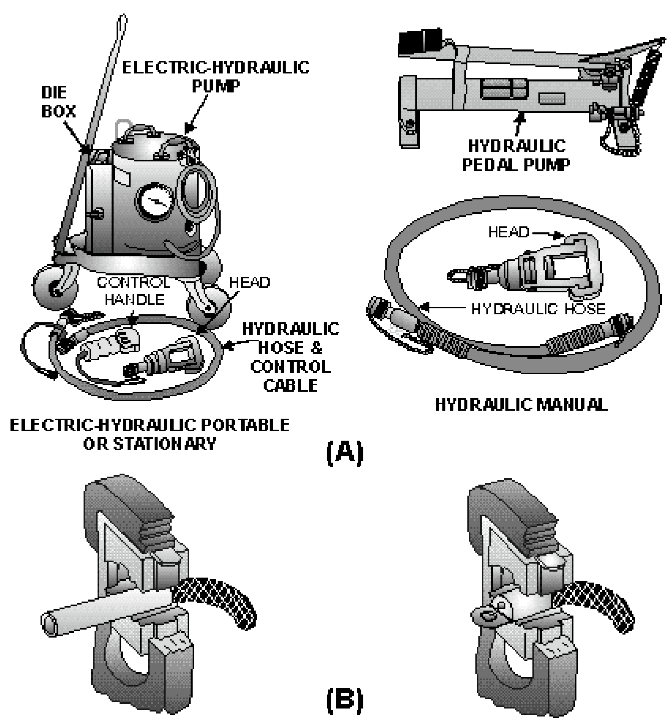


Figure 2-18.—Power crimping tools.

The steps used for crimping an aluminum terminal or splice to an aluminum wire (view B of figure 2-18) are as follows:

1. Carefully remove the conductor insulation. Do not cut or nick the aluminum conductors. Do not wire-brush or scrape the aluminum conductor (the compound in the terminal or splice barrel will clean it satisfactorily).

2. Remove the protective foil wrapping from the terminal or splice and check the amount of compound in the terminal barrel. It should be one-fourth to one-half full.

3. Slip the spaghetti or heat-shrinkable tubing over the wire and back far enough to be out of the way of the crimping operation. Insert the stripped conductor the full length of the terminal or splice barrel. While doing this, leave the plug over the inspection hole. This allows the compound to be forced in and around the strands.

4. Center the terminal lug or splice in the crimping tool.

5. Actuate the power crimping tool.

6. Wipe off the excess compound. Inspect the joint with a probe through the inspection hole. The end of the conductor should come to the edge of the inspection hole.

7. Slip the tubular insulation down over the terminal or splice barrel. Tie it in place if spaghetti is used. If using heat-shrinkable tubing, shrink with a heat gun.

Q15. Should aluminum wire be cleaned prior to installing an aluminum terminal lug or splice?

Q16. What tools should be used to install large aluminum terminal lugs and splices?

Q17. Why should a lockwasher never be used with an aluminum terminal?

Improper crimping procedures eventually cause terminal failure. Be especially careful of undercrimping, overcrimping, using wrong crimping tools, improper cleaning methods, and cutting or nicking the conductors. A loose contact allows an oxide film to form between the wire and the terminal. This results in increased resistance, and the resistance causes heat. The heat accelerates deterioration, and eventually a failure results.

PREINSULATED COPPER TERMINAL LUGS AND SPLICES

The use of preinsulated terminal lugs and splices has become the most common method for copper wire termination and splicing in recent years. It is by far the best and easiest method. There are many tools used for crimping terminal lugs and splices.

Hand, portable power, and stationary power tools are available for crimping terminal lugs. These tools crimp the barrel to the conductor and, at the same time, form the insulation support to the wire insulation.

The power tools, both stationary and portable, are usually found in large shops where wire bundles are made up. In the next paragraphs, we will discuss the more common hand-crimping tools you will most likely be using in your day-to-day work.

TERMINATING COPPER WIRE WITH PREINSULATED TERMINAL LUGS

Small-diameter copper wires are terminated with solderless, preinsulated copper terminal lugs. As shown in figure 2-19, the insulation is part of the terminal lug. It extends beyond the barrel so that it covers a portion of the wire insulation. This makes the use of spaghetti or heat-shrinkable tubing unnecessary. Preinsulated terminal lugs also have an insulation support (a metal reinforcing sleeve) beneath the insulation for extra supporting strength of the wire insulation. Some preinsulated terminals fit more than one size of wire. The insulation is color coded, and the range of wire sizes is marked on the tongue. This identifies the wire sizes that can be terminated with each of the terminal lug sizes. (See table 2-2.)

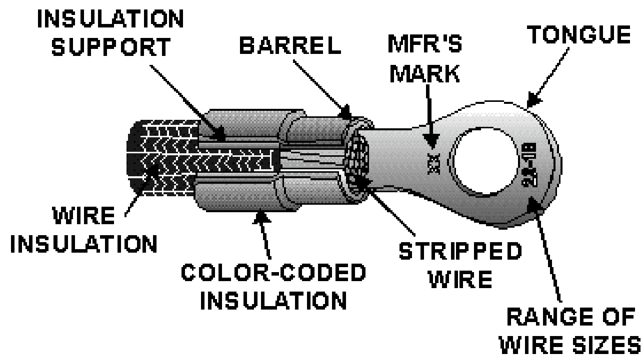


Figure 2-19.—Preinsulated straight copper terminal lug.

Table 2-2.—Color Coding of Copper Terminal Lug or Splice Insulation

Color of Terminal Lug or Splice Insulation	To Be Used on Wire Sizes
Yellow (Bright)	#26 - #24
Red	#22 - #20, #18
Blue	#16 - #14
Yellow (Dull)	#12 - #10

For crimping small copper terminal lugs, several hand-crimping tools can be used for wire sizes AWG 26 through 10 (figure 2-20). These hand-crimping tools have a self-locking ratchet, which prevents the tool from opening until the crimp is completed. Some of these tools have a color-coded selector knob to match the color-coded terminal lug or splice being used. Other tools have a replaceable set of dies for several wire sizes. The hand-crimping procedure for preinsulated copper terminal lugs in wire sizes No. 26 through No. 10 with the standard hand-crimp tool is as follows:

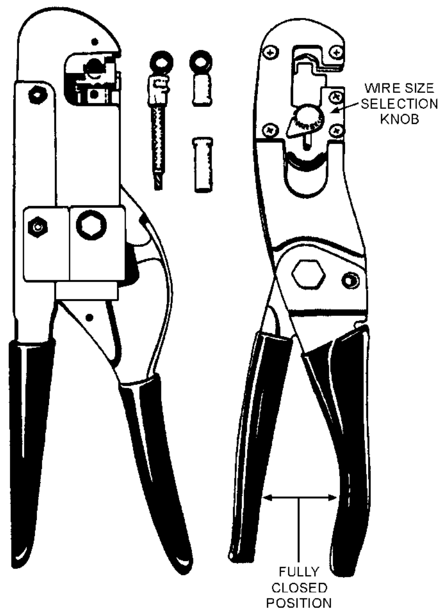


Figure 2-20.—Hand-crimping tools.

1. Strip the wire insulation using the recommended stripping procedures already discussed.
2. Ensure that the tool handles are fully open and the proper die set has been installed correctly.
3. Insert the terminal lug, tongue first, into the wire side of the hand tool barrel crimping jaws. Be certain the terminal lug barrel butts flush against the tool stop on the locator. See figure 2-21 for the correct insertion method.

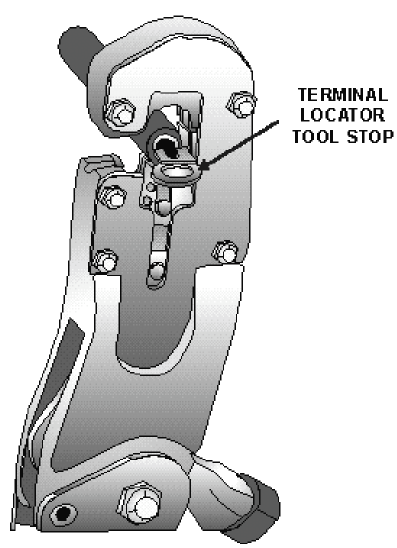


Figure 2-21.—Crimping tool with terminal lug inserted.

4. Squeeze the tool handles slowly until the tool jaws hold the terminal lug barrel firmly in place, but without denting it.

5. Insert the stripped wire into the terminal lug barrel until the wire insulation butts flush against the near end of the wire barrel. (See figure 2-22.)

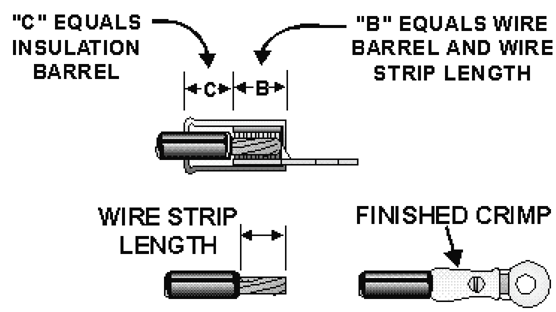


Figure 2-22.—Proper insertion of stripped wire in insulation terminal lug for crimping.

6. Squeeze the tool handles until the rachet releases.

7. Remove the completed assembly and examine it for the proper crimp in accordance with the following:

   a. Indent centered on the terminal lug barrel.

   b. Indent in line with the barrel.

   c. Terminal lug not cracked.

   d. Terminal lug insulation not cracked.

   e. Insulation grip crimped.

CAUTION

If not properly stripped, some of the smaller gauge, thin-wall wire insulation can be inadvertently inserted and crimped in the terminal wire barrels. This will cause a bad electrical connection. Do not use any connection that is found defective as a result of a visual inspection. Cut off the defective connection and remake using a new terminal lug.

PREINSULATED SPLICES

Preinsulated permanent copper splices are used to join small copper wire AWG sizes No. 26 through No. 10. A typical splice is shown in figure 2-23. Note that the splice preinsulation extends over the wire insulation. Each splice size can be used for more than one wire size. Splices are color coded in the same manner as preinsulated small copper terminal lugs (see table 2-2).

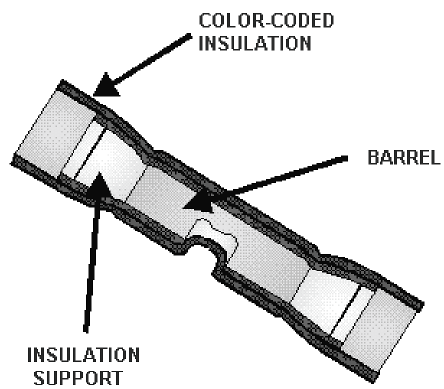


Figure 2-23.—Preinsulated copper splice.

Crimping Procedure for Splices.

Crimping small preinsulated copper splices in the No. 26 to No. 14 wire-size range can be accomplished with several recommended tools. In this section, we will discuss the basic crimping procedures.

1. Strip wire to length following one of the procedures already discussed.

2. With the tool handles fully open, set the wire size selector knob to the proper position for the wire size being crimped. Slide the terminal lug locator down below the die surface into the fully retracted position. (See figure 2-24.) Slide the splice locator back into the retracted position. Insert the splice into the tool so that the "locating shoulder" on the side of the splice to be crimped is in the space between the two crimping dies. The insulation barrel on this side of the splice should protrude from the "wire side" of the tool. (See figure 2-24.) Slide the splice locator into the fully extended position. Insert the splice into the stationary die so that the locator "finger" fits into the locator groove in the splice.

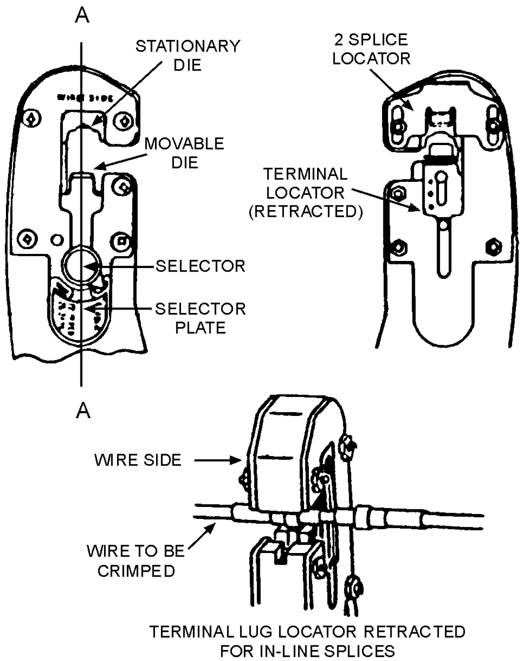


Figure 2-24.—Locating splice in crimping tool.

3. Squeeze the tool handles slowly until the tool jaws hold the spice barrel firmly in place, but without denting he barrel.

4. Insert the stripped wire into the splice barrel, which protrudes from the "wire side" of the splice, until the stripped end of wire butts against the stop in the center of the splice. This can be seen through the splice inspection window.

5. Crimp by closing the tool handles. The tool will not open until the full crimping cycle has been completed.

6. After crimping, check that the wire end is still visible through the splice inspection window.

7. Reverse the position of the splice in the crimping tool (or location of the crimping tool on the splice) and repeat steps 1 through 6 to crimp the wire into the other side of the splice.

If the correct tools are used and the proper procedures followed, crimp-on connections are more effective electrically, as well as mechanically, than soldered connections. A visual inspection is very important. It reveals oxidation, deterioration, overheating, and broken conductors. In some cases it may be necessary to check these connections with an ohmmeter. The proper resistance, for all practical purposes, should be zero. Any defective terminal should be removed and a new terminal crimped on.

Q18. What is the most common method of terminating and splicing wires?

Q19. Besides not having to insulate a noninsulated terminal, what other advantage is gained by using a preinsulated terminal lug?

Q20. Why are preinsulated terminal lugs and splices color coded?

SOLDERING

The following information will aid you in learning basic soldering skills. It should enable you to solder wires to electrical connectors, splices, and terminal lugs that we have discussed earlier in the chapter. Special skills and schooling are required for the soldering techniques used in printed circuit boards and microminiature component repair.

SOLDERING PROCESS

Cleanliness is essential for efficient, effective soldering. Solder will not adhere to dirty, greasy, or oxidized surfaces. Heated metals tend to oxidize rapidly. This is the reason the oxides, scale, and dirt must be removed by chemical or mechanical means. Grease or oil films can be removed with a suitable solvent. Connections to be soldered should be cleaned just prior to the actual soldering operation.

Items to be soldered should normally be "tinned" before making a mechanical connection. Tinning is the coating of the material to be soldered with a light coat of solder. When the surface has been properly cleaned, a thin, even coating of flux should be placed over the surface to be tinned. This will prevent oxidation while the part is being heated to soldering temperature. Rosin-core solder is usually preferred in electrical work. However, a separate rosin flux may be used instead. Separate rosin flux is frequently used when wires in cable fabrication are tinned.

Q21. Why must items to be soldered be cleaned just prior to the soldering process?

TINNING COPPER WIRE AND CABLE

Wires to be soldered to connectors should be stripped so that when the wire is placed in the barrel, there will be a gap of approximately 1/32 inch between the end of the barrel and the end of the insulation. This is done to prevent burning the insulation during the soldering process and to allow the wire to flex easier at a stress point. Before copper wires are soldered to connectors, the ends exposed by stripping are tinned to hold the strands solidly together. The tinning operation is satisfactory when the ends and sides of the wire strands are fused together with a coat of solder. Do not tin wires that are to be crimped to solderless terminals or splices.

Copper wires are usually tinned by dipping them into flux (view A of figure 2-25) and then into a solder bath (pot) (view B of the figure). In the field, copper wires can be tinned with a soldering iron and rosin-core solder. Tin the conductor for about half its exposed length. Tinning or solder on the wire above the barrel causes the wire to be stiff at the point where flexing takes place. This will result in the wire breaking.

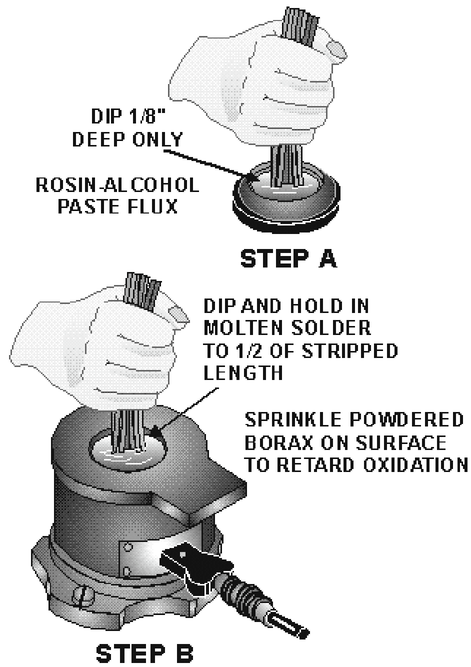


Figure 2-25.—Dip-tinning In a solder pot.

The flux used in tinning copper wire is a mixture of denatured alcohol and freshly ground rosin. This type of flux may be mixed just prior to use. A premixed paste flux may also be used. The solder used for terminal lugs, splices, and connectors is a mixture of 60-percent tin and 40-percent lead. Maintain the temperature of the solder bath (pot) between 450 and 500º F. This keeps the solder in a liquid state. Skim the surface of the solder pot, as necessary, with a metal spoon or blade. This keeps the solder clean and free from oxides, dirt, and so forth.

Dip-tin wires smaller than No. 8 in groups of 8 or 10. Dip-tin wires size No. 8 and larger individually. The procedure for dip-tinning is as follows:

1. Prepare the flux and solder as previously described.
2. Make sure the exposed end of the wire is clean and free from oil, grease, and dirt. Strands should be straight and parallel. Dirty wire should be restripped.
3. Grasp the wire firmly and dip it into the prepared flux to a depth of about 1/8 inch (see view A of figure 2-25).
4. Remove the wire and shake off the excess flux.
5. Immediately dip the wire into molten solder. Dip only half of the stripped conductor length into the solder (see view B of figure 2-25).
6. Turn the wire slowly in the solder bath until the wire is well tinned. Watch the solder fuse to the wire. Do not keep the wire in the bath longer than necessary.
7. Remove the excess solder by wiping the tinned conductor on a cloth.

WARNING

Do not shake off excess solder. It can cause serious burns if it contacts your skin. It can also cause short circuits in exposed electrical equipment that may be in the immediate area of the tinning operation.

CAUTION

Use only rosin flux or rosin-core solder for tinning copper wires to be used in electrical and electronics systems. Corrosive flux will cause damage. During the tinning operation, do not melt, scorch, or burn the insulation.

Q22. What does "tinning" mean in relationship to soldering?

Q23. Why should wire be stripped 1/32 inch longer than the depth of the solder barrel?

Q24. How much of the stripped length of a conductor should be tinned?

ALTERNATIVE DIP-TINNING PROCEDURE

If an electrically heated solder pot is not available, a small number of wires can be tinned using the following procedure (see figure 2-26):

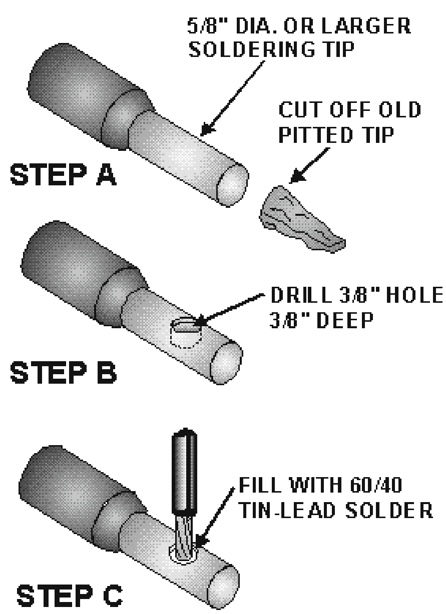


Figure 2-26.—Alternate dip-tinning method.

1. Cut off the beveled section of the tip of a discarded soldering iron tip.
2. Drill a hole (1/4- to 3/8-inch diameter) in the round part of the tip about two-thirds through.
3. Heat the iron and melt the rosin-core solder into the hole.
4. Tin the wires by dipping them into the molten solder one at a time.
5. Keep adding fresh rosin-core solder as the flux burns away.

PROCEDURE FOR TINNING COPPER WIRE WITH A SOLDERING IRON

In the field, wires smaller than size No. 10 can be tinned with a soldering iron and rosin-core solder as follows (see figure 2-27):

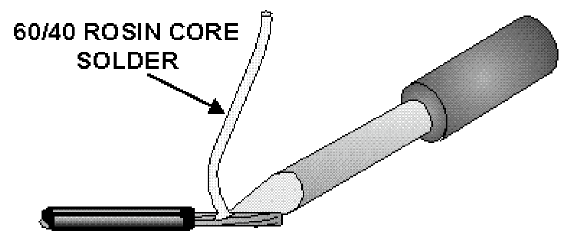


Figure 2-27.—Tinning wire with a soldering iron.

1. Select a soldering iron with the correct heat capacity for the wire size (see table 2-3). Make sure that the iron is clean and well tinned.

Table 2-3.—Approximate Soldering Iron Size for Tinning

Wire Size (AWG)	Soldering Iron Size (Heat Capacity)
#20 - #16	65 Watts
#14 & #12	100 Watts
#10 & #8	20 Watts

2. Start by holding the iron tip and solder together on the wire until the solder begins to flow.
3. Move the soldering iron to the opposite side of the wire and tin half of the exposed length of the conductor.

The tinned surfaces to be joined should be shaped, fitted, and then mechanically joined to make a good mechanical and electrical contact. The parts must be held still. Any motion between the parts while the solder is cooling usually results in a poor solder connection, commonly called a "fractured solder" joint.

Q25. What causes a "fractured solder" joint?

SOLDERING TOOLS

Many types of soldering tools are in use today. Some of the more common types are the soldering iron, soldering gun, resistance soldering set, and pencil iron. The following discussion will provide you with a working knowledge of these tools.

Soldering Irons

Some common types of hand soldering irons are shown in figure 2-28. All high-quality soldering irons operate in the temperature range of 500 to 600º F. Even the 25-watt midget irons produce this temperature. The important difference in iron sizes is not temperature, but thermal inertia. Thermal inertia is the capacity of the iron to generate and maintain a satisfactory soldering temperature while giving up heat to the joint to be soldered. Although it is not practical to solder large conductors with the 25-watt iron, this iron is quite suitable for replacing a half-watt resistor in an electronic circuit or soldering a miniature connector. One advantage of using a small iron for small work is that it is light and easy to handle and has a small tip that is easily used in close places. Even though its temperature is high enough, a midget iron does not have the thermal inertia to solder large conductors.

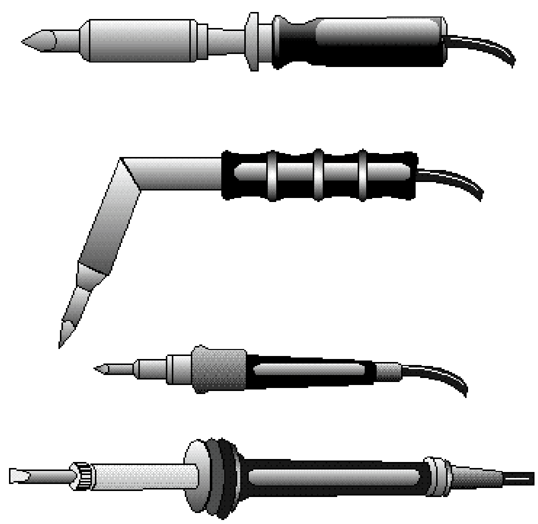


Figure 2-28.—Types of hand soldering Irons.

A well-designed iron is self-regulating. The resistance of its element increases with rising temperature. This limits the flow of current. Some common tip shapes of the soldering irons in use in the Navy are shown in figure 2-29.

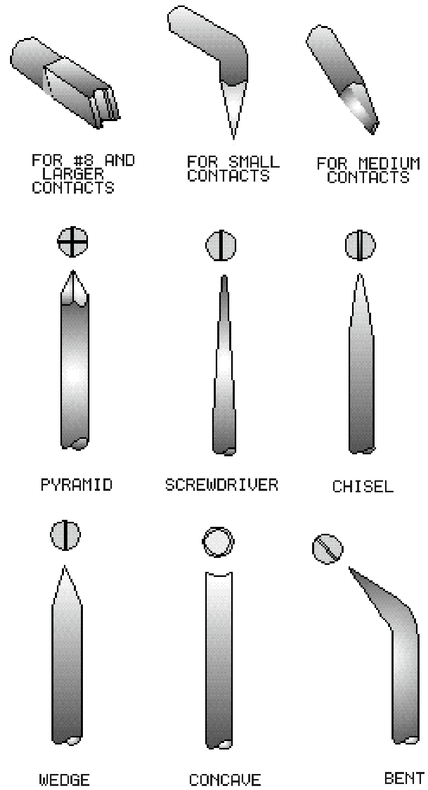


Figure 2-29.—Soldering iron tip shapes.

An iron should be tinned (the application of solder to the tip after the iron is heated) prior to soldering a component in a circuit. After extended use of an iron, the tip tends to become pitted due to oxidation. Pitting indicates the need for retinning. The tip is retinned after first filing the tip until it is smooth (see figure 2-30).

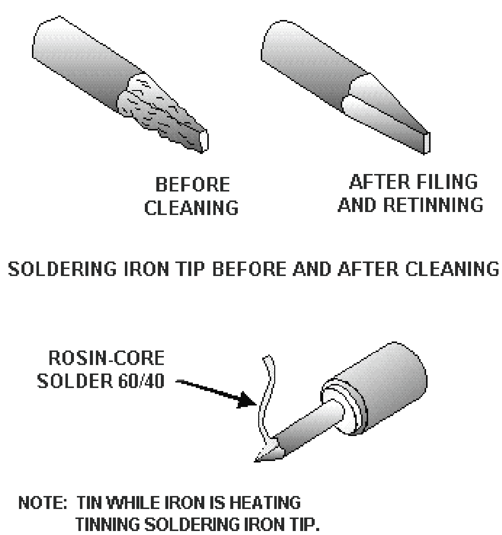


Figure 2-30.—Reconditioning pitted soldering iron tip.

Q26. Define thermal inertia.

Q27. Why are small-wattage soldering irons not used to solder large conductors?

Q28. State why a well-designed soldering iron is self-regulating.

Q29. What should be done to a soldering iron tip that is pitted?

Soldering Gun

The soldering gun (figure 2-31) has gained great popularity in recent years because it heats and cools rapidly. It is especially well adapted to maintenance and troubleshooting work where only a small part of the technician's time is spent actually soldering.

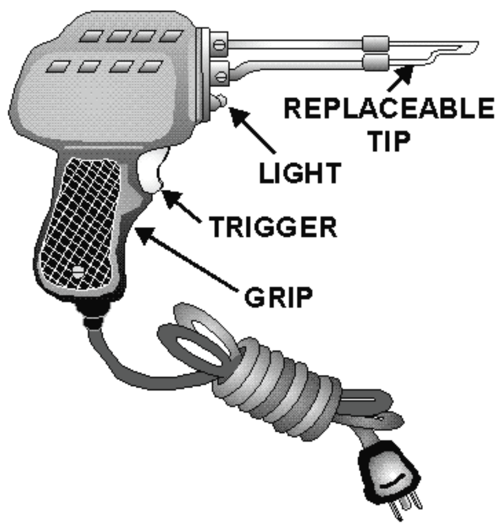


Figure 2-31.—Soldering gun.

A transformer in the soldering gun supplies approximately 1 volt at high current to a loop of copper, which acts as the soldering tip. It heats to soldering temperature in 3 to 5 seconds. However, it may overheat to the point of incandescence if left on over 30 seconds. This should be avoided because excess heat will burn the insulation off the wiring. The gun is operated by a finger switch. The gun heats only while the switch is pressed.

Since the gun normally operates only for short periods at a time, it is comparatively easy to keep clean and well tinned. Short operating time allows little oxidation to form. Because the tip is made of pure copper, it is likely to pit, due to the dissolving action of the solder.

The gun or iron should always be kept tinned to permit proper heat transfer to the connection to be soldered. Tinning also helps control the heat to prevent solder buildup on the tip. This control reduces the chance of the solder spilling over to nearby components and causing short circuits. Maintaining the proper tinning on the iron or gun, however, may be made easier by tinning with silver solder (a composition of silver, copper, and zinc). The temperature at which the bond is formed between the copper tip and the silver solder is much higher than with lead-tin solder. This tends to decrease the pitting action of the solder on the copper tip.

Overheating small or delicate wiring can easily occur when a soldering gun is used. For most jobs, even the LOW position of the trigger overheats the gun after 10 seconds. With practice, the heat can be controlled by pulsing the gun on and off with its trigger. The HIGH position is used only for fast heating and for soldering heavy connections.

When a soldering iron or gun is used, heating and cooling cycles tend to loosen the nuts or screws that hold the replaceable tips. When the nut on a gun becomes loose, the resistance of the tip connection increases. The temperature of the connection is increased, thus reducing the heat at the tip. Continued loosening may eventually cause an open circuit. Therefore, check and tighten the nut or screw, as needed.

CAUTION

Soldering guns should never be used to solder electronic components, such as resistors, capacitors, and transistors, because the heat generated can destroy the components. They should be used only on terminals, splices, and connectors (not the miniature type).

Q30. What happens if a soldering gun switch is pressed for periods longer than 30 seconds?

Q31. What causes the nuts or screws that hold the tips on soldering irons and guns to loosen?

Q32. A soldering gun should NOT be used on what components?

Resistance Soldering Set

A time-controlled resistance soldering set (figure 2-32) is now used at many maintenance activities. The set consists of a transformer that supplies 3 or 6 volts at a high current to stainless steel or carbon tips. The transformer is turned ON by a foot switch and OFF by an electronic timer. The timer can be adjusted for as long as 3 seconds soldering time. This set is especially useful for soldering cables to plugs and similar connectors; even the smallest types.

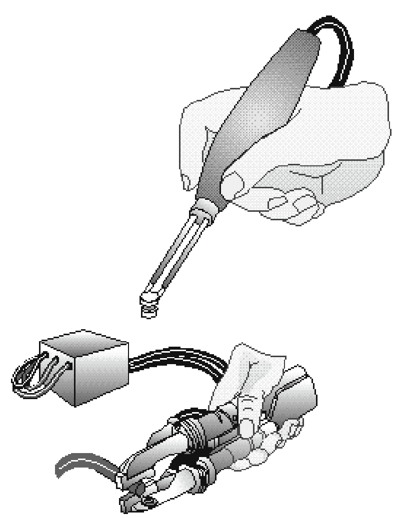


Figure 2-32.—Resistance soldering set

In use, the double-tip probes of the soldering unit are adjusted to straddle the connector cup (connector barrel) to be soldered. One pulse of current heats it for tinning. After the wire is inserted, a second pulse of current solders the connection and completes the job. Since the soldering tips are hot only during the brief period of actual soldering, burning of wire insulation and melting of connector inserts are greatly reduced.

The greatest difficulty with this device is keeping the probe tips free of rosin and corrosion. A cleaning block is mounted on the transformer case for this purpose. Some technicians prefer fine sandpaper for cleaning the double tips.

CAUTION

Do not use steel wool for cleaning tips. It is dangerous when used around electrical equipment because the strands can fall into the equipment and cause short circuits.

Q33. What is an advantage of using a resistance soldering iron when soldering wire to a connector?

Q34. Why is steel wool NEVER used as an abrasive to clean soldering tools?

Pencil Iron and Special Tips

An almost indispensable item is the pencil-type soldering iron with an assortment of tips (figure 2-33). Miniature soldering irons have a wattage rating of less than 40 watts. They are easy to use, and are recommended for soldering small components, such as miniature connectors.

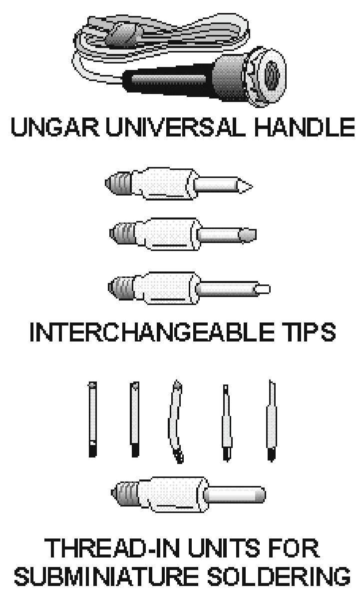


Figure 2-33.—Pencil iron with special tips.

One type of pencil iron is equipped with several different tips that range from one-fourth to one-half inch in size (diameter) and are of various shapes. This feature makes it adaptable to a variety of jobs. Unlike most tips that are held in place by setscrews, these tips have threads and screw into the barrel. This feature provides excellent contact with the heating element, thus improving heat transfer efficiency. "Antiseize" compound is generally applied to the threads of the tip each time a tip is installed into the iron. This allows the tip to be easily removed when another is to be inserted.

A special feature of this iron is the soldering pot that screws in like a tip and holds about a thimbleful of solder. It is useful for tinning the ends of a large number of wires.

The interchangeable tips are of various sizes and shapes for specific uses. Extra tips can be obtained and shaped to serve special purposes. The thread-in units are useful in soldering small items.

Another advantage of the pencil soldering iron is that it can be used as an improvised light source to inspect the completed work. Simply remove the soldering tip and insert a 120-volt, 6-watt, type 6S6 lamp bulb into the socket.

If leads, tabs, or small wires are bent against a board or terminal, slotted tips are provided to simultaneously melt the solder and straighten the leads.

If no suitable tip is available for a particular operation, an improvised tip can be made (see figure 2-34). Wrap a length of bare copper wire around one of the regular tips and bend the wire into the proper shape for the purpose. This method also serves to reduce thermal inertia when a larger iron must be used on small components.

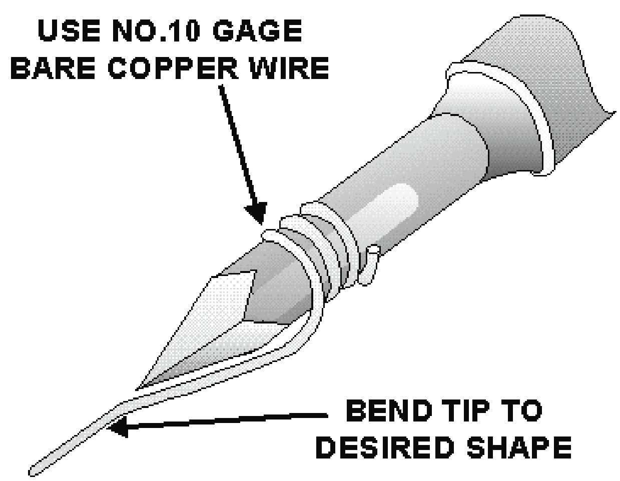


Figure 2-34.—Improvised tip.

Q35. Why should "antiseize" compound be used on the screw-in tips of the pencil iron?

Q36. If no suitable tip is available for a particular job, how may one be improvised?

SOLDER

Any discussion of soldering techniques should include an explanation of solder itself. Ordinary soft solder is a fusible alloy consisting chiefly of tin and lead. It is used to join two or more metals at temperatures below their melting point. In addition to tin and lead, soft solders occasionally contain varying amounts of antimony, bismuth, cadmium, or silver. These are added to change the melting point or physical properties of the alloy. Ordinary table salt has to be heated to 1,488º F before it melts. However, when a little water is added, it dissolves easily at room temperature. The action of molten solder on a metal like copper may be compared to the action of water on salt.

The solder bonds the connection by dissolving a small amount of the copper at temperatures quite below its melting point. Thus, the soldering process involves a metal solvent action between the solder and the metal being joined. A solder joint is therefore chemical in nature rather than purely physical. The bond is formed in part by chemical action and part by a physical bond.

The properties of a solder joint are different from those of the original solder. The solder is converted to a new and different alloy through the solvent action. Two metals soldered together behave like one solid metal. It is unlike two metals bolted, wired, or otherwise physically attached. These types of connections are still two pieces of metal. They are not even in direct contact due to an insulating film of oxide on the surfaces of the metals.

Temperature change does not affect the solder alloy. It withstands stress and strains without damaging the joint. An unsoldered connection eventually becomes loosened by small movements caused by temperature variations and by the gradual buildup of oxides on the metal surfaces.

To understand fully the alloy or solvent action on molten solder, look at the tin-lead fusion diagram shown in figure 2-35. This diagram shows that pure lead (point A) melts at 621º F. Point C shows the lowest melting point of the tin and lead alloy. The alloy at point C consists of 63-percent tin (SN63) and 37-percent lead. This is commonly called 63/37 solder. It has a melting point of 361º F. This type of solder, because of its very low melting point, is used in printed circuit boards and microminiature electronic repair. As you can see from the chart, the melting point of the alloy is lowered when tin is added to lead.

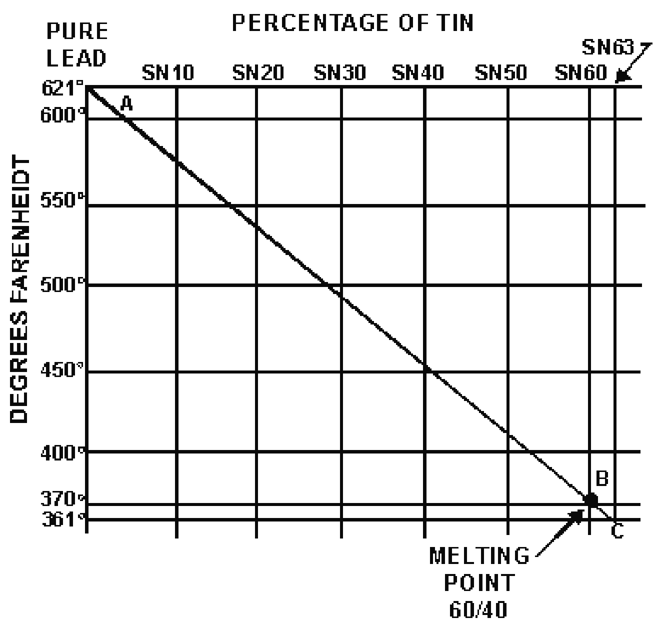


Figure 2-35.—Tin-lead fusion diagram.

The solder used to solder wires to electrical connectors, splices, and terminal lugs is a combination of 60-percent tin to 40-percent lead (60/40 solder). The melting point of 60/40 solder is 370º F, as shown at point B of the figure. Type 60/40 solder is less expensive than 63/37 solder and is suitable for all general uses.

Q37. What two metals are used to from soft solder?

Q38. Define the metal solvent action that takes place when copper conductors are soldered together.

Q39. What is the tin-lead alloy percentage of solder used for electrical connectors, splices, and terminal lugs?

FLUX

As you know, flux is a cleaning agent to remove oxidation during soldering. Heating a metal causes rapid oxidation. Oxidation prevents solder from reacting chemically with a metal. Flux cleans the metal by removing the oxide layer. This operation is shown in figure 2-36. As the iron is moved in the direction shown, the boiling flux floats away the oxide film. The molten solder following the iron then fuses rapidly with the clean surface of the metal.

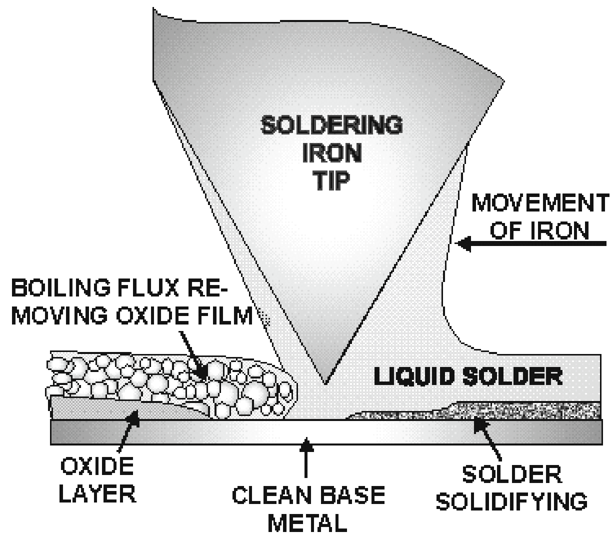


Figure 2-36.—Action of flux.

There are two classes of flux: corrosive and noncorrosive. Zinc chloride, hydrochloric acid, and sal ammoniac are corrosive fluxes. Corrosive flux should NEVER be used in electrical or electronic repair work. Use only rosin fluxes. Any flux remaining in the joint corrodes the connection and creates a defective circuit. Rosin is a noncorrosive flux and is available in paste, liquid, or powder form.

SOLVENTS

A solvent is used for cleaning and removing contaminants (oil, grease, dirt, and so forth) from the soldered connection. Solvents must be nonconductive and noncorrosive. Solvents must be used in a manner that keeps dissolved flux residue from "contact" surfaces, such as those in switches, potentiometers, or connectors. Ethyl and isopropyl alcohol are acceptable solvents.

WARNING

These cleaning solvents are highly flammable and may give off toxic vapors. Follow Navy safety precautions and take extreme care when using any flammable solvent.

Q40. What purpose does flux serve in the soldering process?

Q41. What type of flux must be used in all electrical and electronic soldering?

Q42. Why are solvents used in the soldering process?

Soldering Aids

Some type of heat shunt must be used in all soldering operations that involve heat-sensitive components. A typical heat shunt (figure 2-37) permits soldering the leads of component parts without overheating the part itself. The heat shunt should be attached carefully to prevent damage to the leads, terminals, or component parts. The shunt should be clipped to the lead, between the joint and the part being protected. As the joint is heated, the shunt absorbs the excess heat before it can reach the part and cause damage.

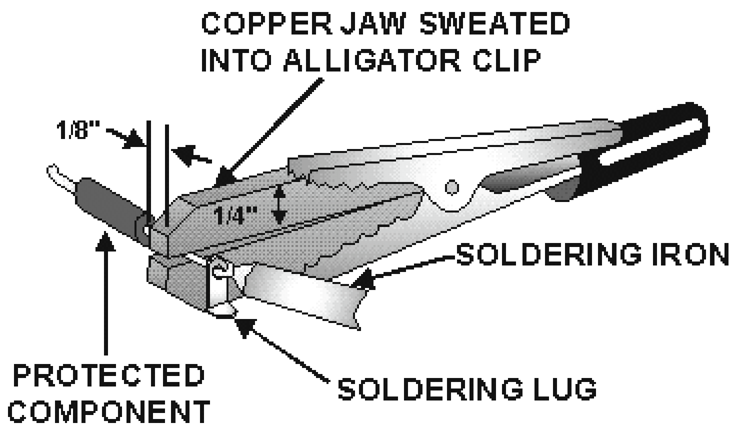


Figure 2-37.—Heat shunt.

A small piece of beeswax may be placed between the protected unit and the heat shunt. When the beeswax begins to melt, the temperature limit has been reached. The heat source should be removed immediately, but the shunt should be left in place.

Removing the shunt too soon permits the heat to flow from the melted solder into the component. The shunt should be allowed to remain in place until it cools to room temperature. A clip-on shunt is preferred because it requires positive action for removal. It does not require that the technician maintain pressure to hold it in place. This leaves both hands free to solder the connection.

Two safety devices are shown in figure 2-38. These devices prevent burns to the operator when the soldering iron is not in use for short periods of time.

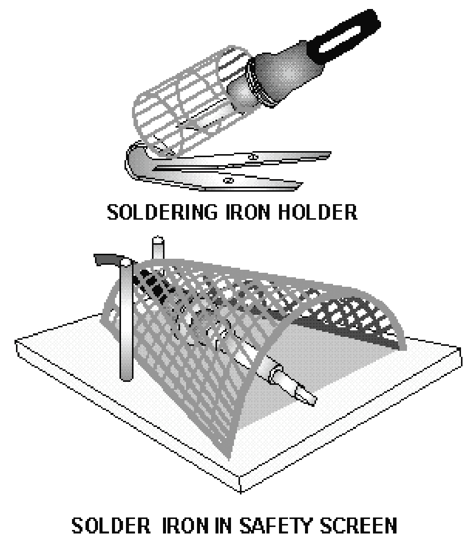


Figure 2-38.—Soldering iron safety devices.

Q43. What is the purpose of a heat shunt?

LACING CONDUCTORS

Conductors within equipment must be kept in place to present a neat appearance and aid in tracing the conductors when alterations or repairs are required. This is done by LACING the conductors into wire bundles called cables. An example of lacing is shown in figure 2-39. When conductors are properly laced, they support each other and form a neat, single cable.

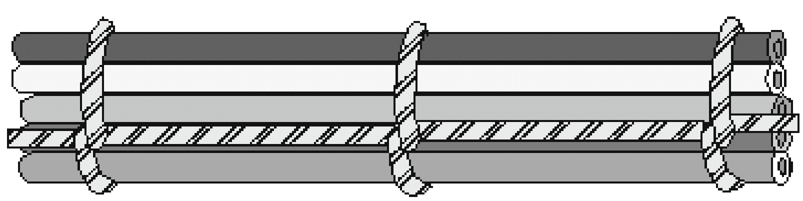


Figure 2-39.—Conductor lacing.

A narrow, flat tape should be used wherever possible for lacing and tying. This tape is not an adhesive type of tape. Round cord may also be used, but its use is not preferred because cord has a tendency to cut into wire insulation. Use cotton, linen, nylon, or glass fiber cord or tape, according to the temperature requirements. Cotton or linen cord or tape must be prewaxed to make it moisture and fungus resistant. Nylon cord or tape may be waxed or unwaxed; glass fiber cord or tape is usually not waxed.

The amount of flat tape or cord required to single lace a group of conductors is about two and one-half times the length of the longest conductor in the group. Twice this amount is required if the conductors are to be double laced.

Before lacing, lay the conductors out straight and parallel to each other. Do not twist them together because twisting makes conductor lacing and wire tracing difficult during troubleshooting.

Q44. Besides presenting a neat appearance and supporting each other, what is the other purpose for lacing conductors?

Q45. Why is flat tape preferred instead of round cord when wire bundles are laced?

Q46. What amount of flat tape or round cord is required to single lace a group of conductors?

A lacing shuttle on which the cord can be wound keeps the cord from fouling during the lacing operation. A shuttle similar to the one shown in figure 2-40 can easily be made from aluminum, brass, fiber, or plastic scrap. Rough edges of the material used for the shuttle should be filed smooth to prevent injury to the operator and damage to the cord. To fill the shuttle for a single lace, measure the cord, cut it, and wind it on the shuttle. For double lace, proceed as before, except double the length of the cord before you wind it on the shuttle. For double lace, start both ends of the cord or tape on the shuttle in order to leave a loop for starting the lace. This procedure is explained later in the chapter.

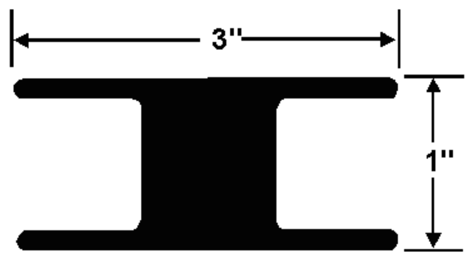


Figure 2-40.—Lacing shuttle.

Some equipment requires the use of twisted wires. One example is the use of "twisted pairs" for the ac filament leads of certain electron tube amplifiers to minimize radiation of their magnetic field. This prevents an annoying hum in the amplifier output. You should duplicate the original layout when relacing any wiring harness.

Lace or tie bundles tightly enough to prevent slipping, but not so tightly that the cord or tape cuts into or deforms the insulation. Be especially careful when lacing or tying coaxial cable. Coaxial cable is a conductor used primarily for radio-frequency transmission. It consists of a center conductor separated from an outer conductor (usually called a shield) by an insulating dielectric. The dielectric maintains a constant capacitance between the two conductors, which is very important in radio transmission. The dielectric is soft and deforms easily if tied too tightly or with the wrong type of tape.

CAUTION

Do not use round cord for lacing or tying coaxial cable or bundles that contain coaxial cable. Use only the approved military specification tape to lace or tie coaxial cables or bundles containing coaxial cables.

Q47. What is the purpose of a lacing shuttle?

Q48. When should wires be twisted prior to lacing?

Q49. What precautions should you take when tying bundles containing coaxial cables?

SINGLE LACE

Single lace can be started with a square knot and at least two marling hitches drawn tightly. Details of the square knot and marling hitch are shown in figure 2-41. Do not confuse the marling hitch with a half hitch. In the marling hitch, the end is passed over and under the strand, as shown in view A of the figure. After forming the marling hitches, draw them tightly against the square knot, as shown in view B. The lace consists of a series of marling hitches evenly spaced at 1/2-inch to 1-inch intervals along the length of the group of conductors, as shown in view C of the figure.

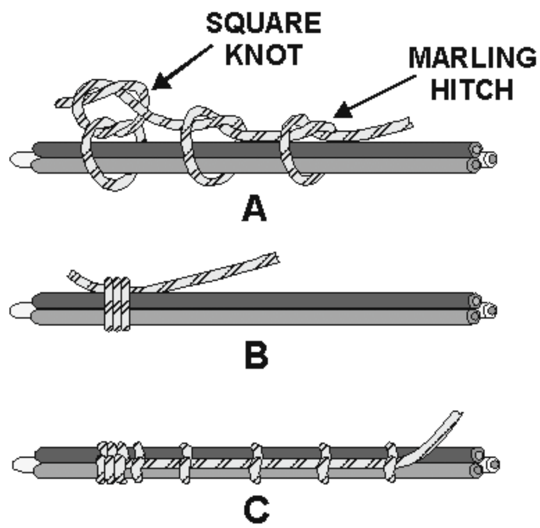


Figure 2-41.—Applying single lace.

When dividing conductors to form two or more branches, follow the procedure illustrated in figure 2-42. Bind the conductors with at least six turns between two marling hitches, and continue the lacing along one of the branches, as shown in view A. Start a new lacing along the other branch. To keep the bends in place, form them in the conductors before lacing. Always add an extra marling hitch just prior to a breakout as shown in view B.

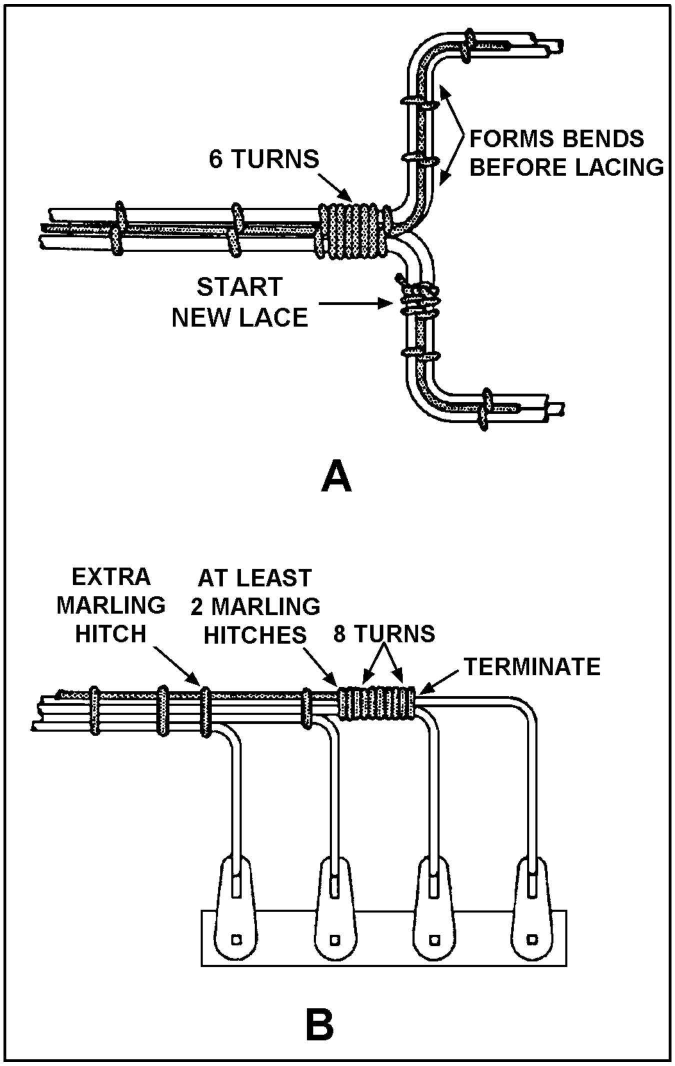


Figure 2-42.—Lacing branches and breakouts.

Double lace should be used on groups of conductors that are 1 inch or larger in total diameter. Either a single lace or a double lace may be used on groups of less than 1 inch.

Q50. How is the single lace started?

DOUBLE LACE

Double lace is applied in a manner similar to single lace, except that it is started with a telephone hitch and is double throughout the length of the lacing (figure 2-43). Both double and single lace may be ended by forming a loop from a separate length of cord and using it to pull the end of the lacing back underneath a serving of approximately eight turns (figure 2-44). An alternate method of ending the lacing is illustrated in figure 2-45. This method can also be used for either single- or double-cord lacing. Another method is by using a marling hitch as a lock stitch (figure 2-46) to prevent slippage. This procedure will also prevent unraveling should a break occur to the lacing.

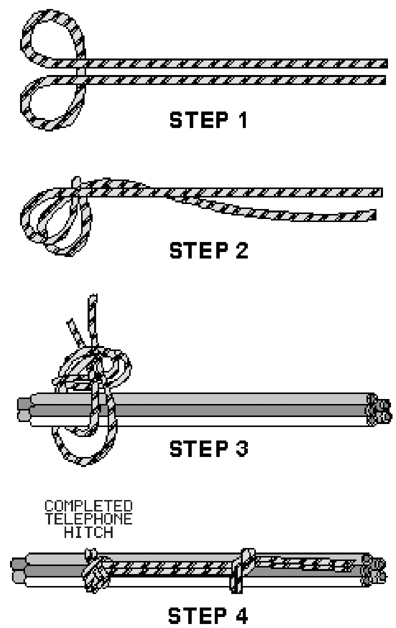


Figure 2-43.—Starting double lace.

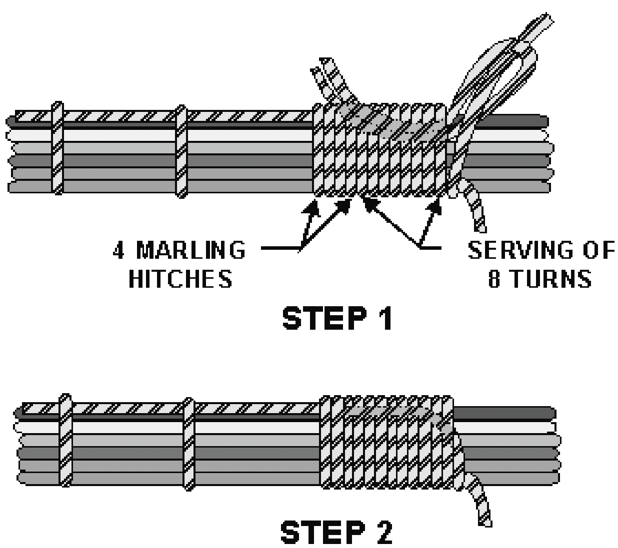


Figure 2-44.—Terminating double lace.

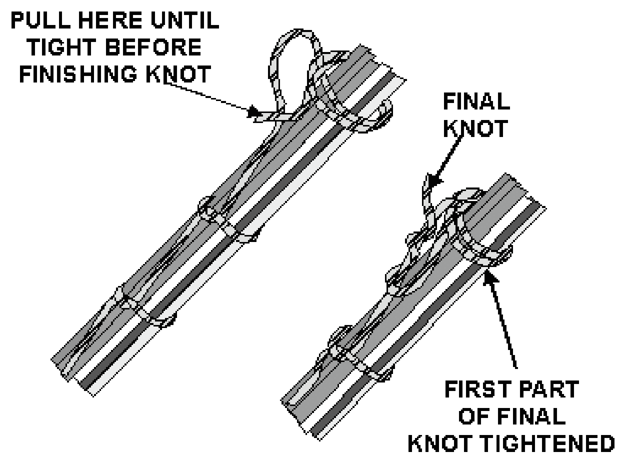


Figure 2-45.—Alternate method of terminating the lace.

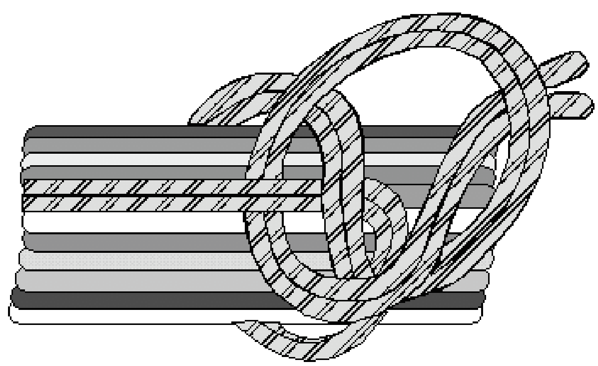


Figure 2-46.—Marling hitch as a lock stitch.

The spare conductors of a multiconductor cable should be laced separately, and then tied to active conductors of the cable with a few telephone hitches. When two or more cables enter an enclosure, each cable group should be laced separately. When groups are parallel to each other, they should be bound together at intervals with telephone hitches (figure 2-47).

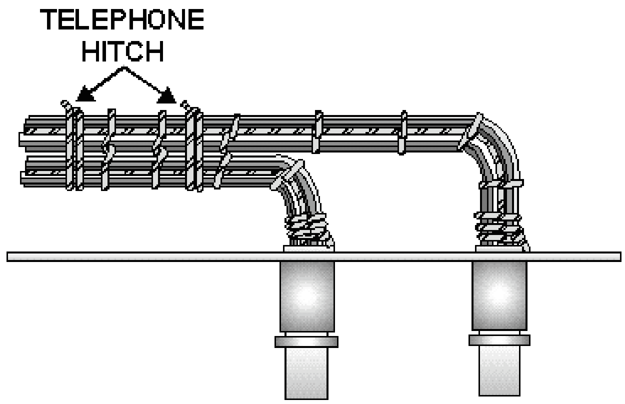


Figure 2-47.—Spot tying cable groups.

Q51. What size wire bundles require double lace?

Q52. How is the double lace started?

Q53. How are laced cable groups bound together?

SPOT TYING

When cable supports are used in equipment as shown in figure 2-48, spot ties are used to secure the conductor groups if the supports are more than 12 inches apart. The spot ties are made by wrapping the cord around the group as shown in figure 2-49. To finish the tie, use a clove hitch followed by a square knot with an extra loop. The free ends of the cord are then trimmed to a minimum of 3/8 inch.

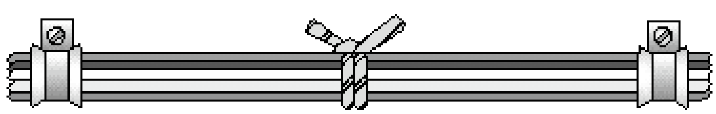


Figure 2-48.—Use of spot ties.

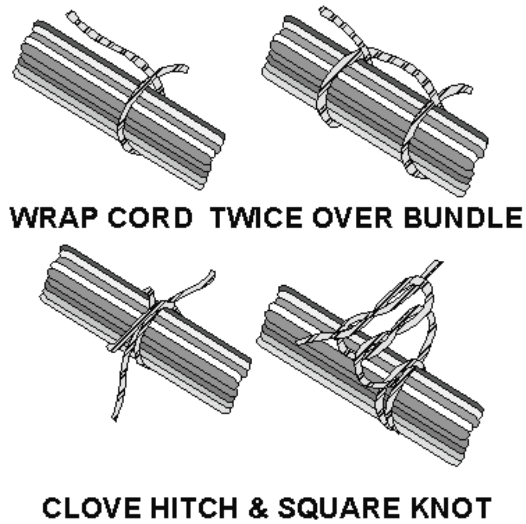


Figure 2-49.—Making spot ties.

SELF-CLINCHING CABLE STRAPS

Self-clinching cable straps are adjustable, lightweight, flat nylon straps. They have molded ribs or serrations on the inside surface to grip the wire. They may be used instead of individual cord ties for securing wire groups or bundles quickly. The straps are of two types: a plain cable strap and one that has a flat surface for identifying the cables.

CAUTION

Do not use nylon cable straps over wire bundles containing coaxial cable. Do not use straps in areas where failure of the strap would allow the strap to fall into movable parts.

Installing self-clinching cable straps is done with a Military Standard hand tool, as shown in figure 2-50. An illustration of the working parts of the tool is shown in figure 2-51. To use the tool, follow the manufacturer's instructions.

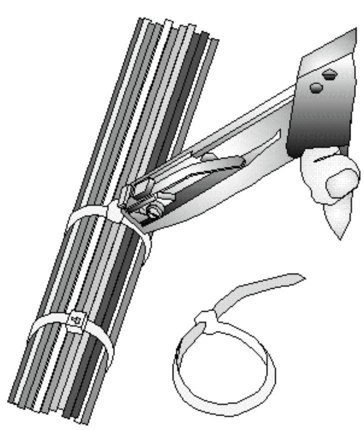


Figure 2-50.—Installing self-clinching cable straps.

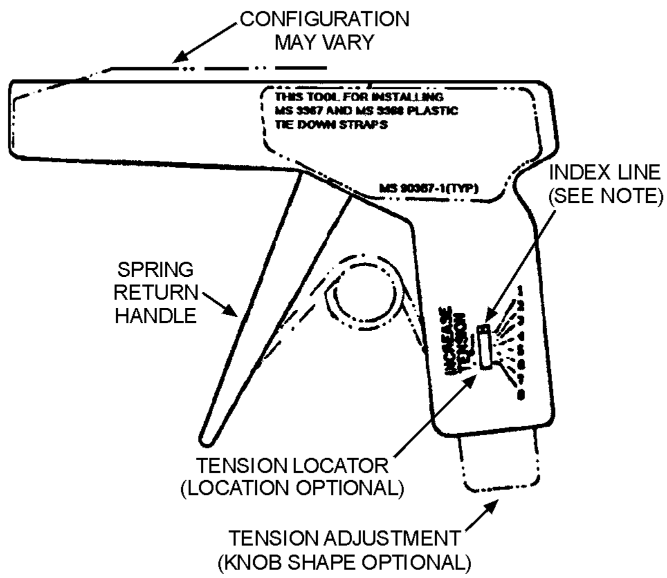


Figure 2-51.—Military Standard hand tool for self-clinching cable straps.

WARNING

Use proper tools and make sure the strap is cut flush with the eye of the strap. This prevents painful cuts and scratches caused by protruding strap ends. Do not use plastic cable straps in high-temperature areas (above 250º F).

HIGH-TEMPERATURE PRESSURE-SENSITIVE TAPE LACING

High-temperature, pressure-sensitive tape must be used to tie wire bundles in areas where the temperature may exceed 250º F. Install the tape as follows (figure 2-52):

1. Wrap the tape around the wire bundle three times, with a two-thirds overlap for each turn.
2. Heat-seal the loose tape end with the side of a soldering iron tip.

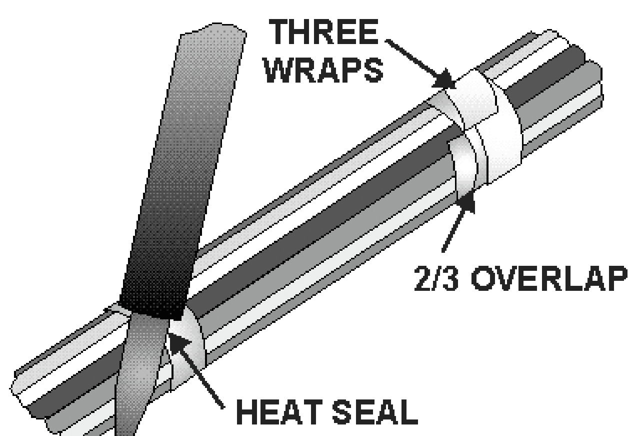


Figure 2-52.—Securing wire bundles in high-temperature areas.

WARNING

Insulation tape (including the glass fiber type) is highly flammable and should not be used in a high-temperature environment. Only insulation tape approved for high-temperature operation (suitable for continuous operation at 500º F) should be used in high-temperature environments.

Q54. When are spot ties used?

Q55. What is used to install self-clinching cable straps?

Q56. What is used to tie wire bundles in high-temperature areas?

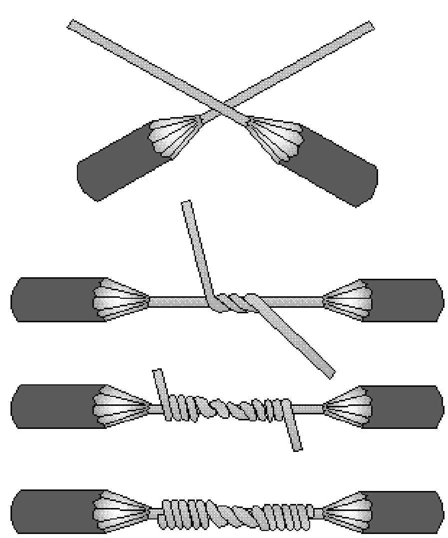
SUMMARY

In this chapter you have learned some of the basic skills required for proper wiring techniques. We have discussed conductor splices and terminal connections, basic soldering skills, and lacing and tying wire bundles.

The basic requirement for any splice or terminal connection is that it be both mechanically and electrically as strong as the conductor or device with which it is to be used.

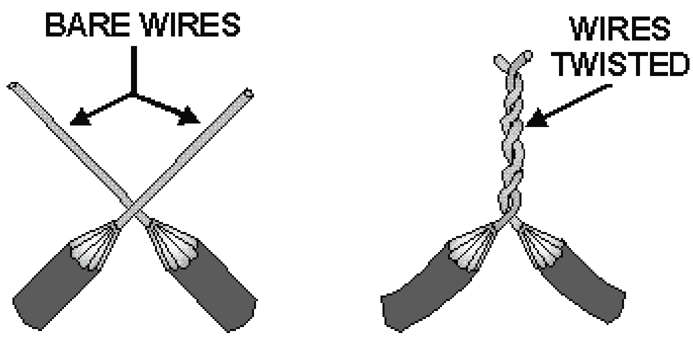
Insulation Removal—The first step in splicing or terminating electrical conductors is to remove the insulation. The preferred method for stripping wire is by use of a wire-stripping tool. The hot-blade stripper cannot be used on such insulation material as glass braid or asbestos. An alternate method for stripping copper wire is with a knife. A knife is the required tool to strip aluminum wire. Take extreme care when stripping aluminum wire. Knicking the strands will cause them to break easily.

Western Union Splice—A simple connection known as the Western Union splice is used to splice small, solid conductors together. After the splice is made, the ends of the wire are clamped down to prevent damage to the tape insulation.



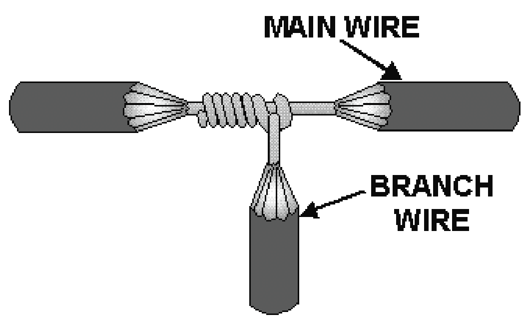
Staggered Splice—The staggered splice is used on multiconductor cables to prevent the joint from being bulky.

Rattail Joint—A splice that is used in a junction box and for connecting branch circuits; wiring is placed inside conduits.



Fixture Joint—When conductors of different sizes are to be spliced, such as fixture wires to a branch circuit, the fixture joint is used.

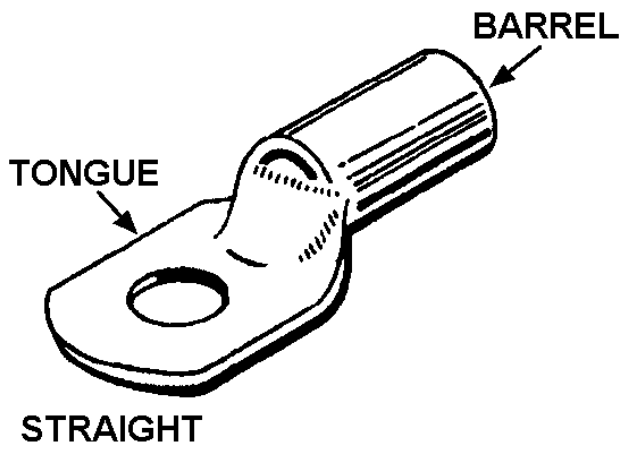
Knotted Tap Joint—This type of splice is used to splice a conductor to a continuous wire. It is not considered a "butted" splice as the ones previously discussed.



Splice Insulation—Rubber tape is an insulator for the type of splices we have discussed so far.

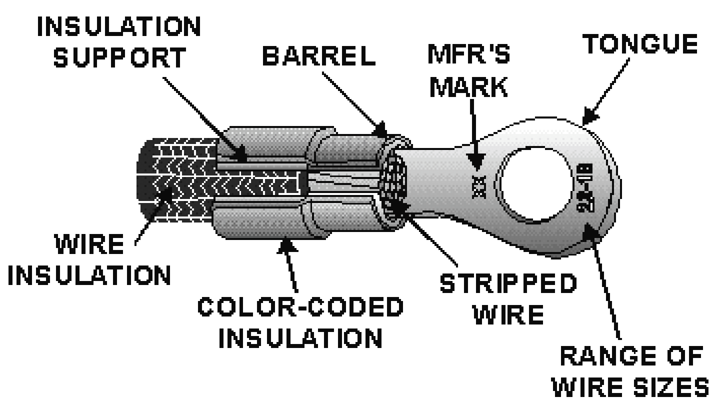
Friction Tape—It has very little insulating value but is used as a protective covering for the rubber tape. Another type of insulating tape is plastic electrical tape, which is quite expensive.

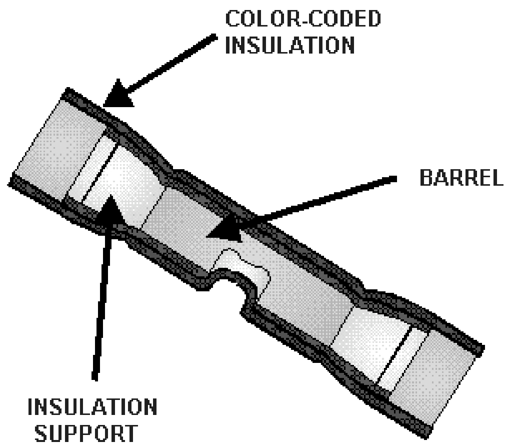
Terminal Lugs—The terminals used in electrical wiring are either of the soldered or crimped type. The advantage of using a crimped type of connection is that it requires very little operator skill, whereas the soldered connection is almost completely dependent on the skill of the operator. Some form of insulation must be used with noninsulated splices and terminal lugs. The types used are clear plastic tubing (spaghetti) and heat-shrinkable tubing. When a heat gun is used to shrink the heat-shrinkable tubing, the maximum allowable heat to be used is 300º F. When using the compressed air/nitrogen heating tool, the air/nitrogen source cannot be greater than 200 psig.



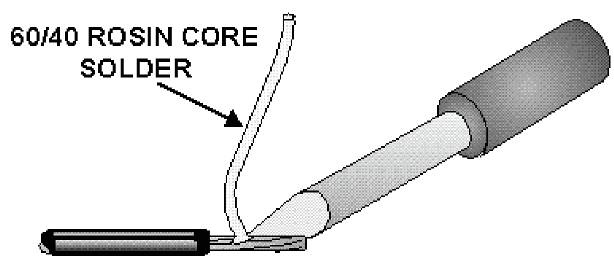
Aluminum Terminals and Splices—Aluminum terminals and splices are noninsulated and very difficult to use. Some of the things you should remember when working with aluminum wire are: (1) Never attempt to clean the aluminum wire. There is a petroleum abrasive compound in the terminal lug or splice that automatically cleans the wire. (2) The only tools that should be used for the crimping operation are the power crimping type. (3) Never use lock washers next to aluminum terminal lugs as they will gouge out the tinned area and increase deterioration.

Preinsulated Copper Terminal Lugs and Splices—The most common method of terminating and splicing copper wires is with the use of preinsulated terminal lugs and splices. Besides not having to insulate the terminal or splice after the crimping operation, the other advantage of this type is that it gives extra wire insulation support. Several types of crimping tools can be used for these types of terminals and splices. The tool varies with the size of the terminal or splice. Preinsulated terminal lugs and splices are color coded to indicate the wire size they are to be used with.





Soldering—The basic skills required to solder terminal lugs, splices, and electrical connectors are covered in this area. Prior to any soldering operation, the items to be soldered must be cleaned; they will not adhere to dirty, greasy, or oxidized surfaces. The next step is the "tinning" process. This process is accomplished by coating the material to be soldered with a bright coat of solder. The wire to be soldered must be stripped to 1/32 inch longer than the depth of the solder cup of the terminal, splice, or connector to which it is to be soldered. This is to prevent burning the insulation. It also allows the wire to flex at the stress point. When you tin the wire, it should be done to one-half of the stripped length. When soldering a connection, take precaution to prevent movement of the parts while the solder is cooling. A "fractured solder" joint will result if this precaution is not taken.



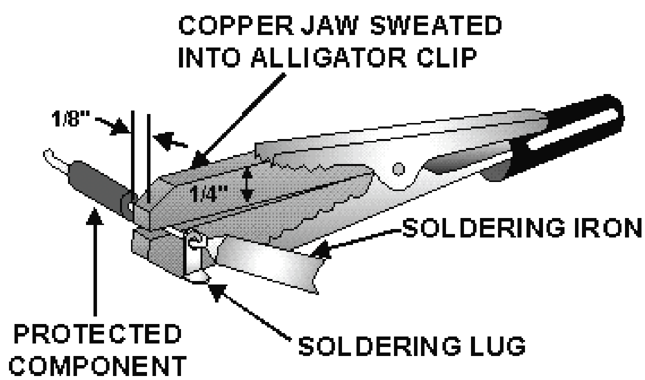
Soldering Tools—The important difference in soldering iron sizes is not the temperature (they all produce 500º F to 600º F), but the thermal inertia. Thermal inertia is the ability of soldering tools to maintain a satisfactory soldering temperature while giving up heat to the joint to be soldered. A well-designed soldering iron is self-regulating because its heating element increases with the rising temperature, thus inciting the current to a satisfactory level. When using a soldering gun, do not press the switch for periods longer than 30 seconds. Doing so will cause the tip to overheat to the point of incandescence. The nuts or screws that retain the tips on soldering irons and guns tend to loosen because of the continuous heating and cooling cycles. Therefore, they should be tightened periodically. You should <u>never</u> use a soldering gun on electronics components, such as resistors, capacitors, or transistors. An advantage of using a resistance soldering iron to solder a wire to a connector is that the soldering tips are only hot during the brief period of soldering the connection.

Solder—Ordinary soft solder is a fusible alloy of tin and lead used to join two or more metals at temperatures below their melting point. The metal solvent action that occurs when copper conductors are soldered together takes place because a small amount of the copper combines with the solder to form a new alloy. Therefore, the joint is one common metal. The tin-lead alloy used for general-purpose soldering is composed of 60-percent tin and 40-percent lead (60/40 solder).

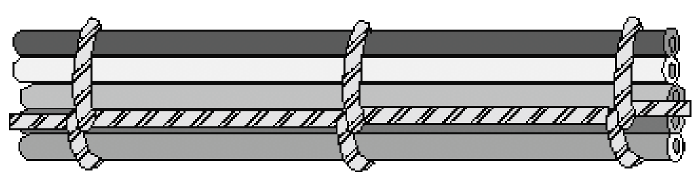
Flux—Flux is used in the soldering process to clean the metal by removing the oxide layer on the metal and to prevent further oxidation during the soldering process. Always use noncorrosive, nonconducting rosin fluxes when soldering electrical and electronic components.

Solvents—Solvents are used in the soldering process to remove contaminants from the surfaces to be soldered.

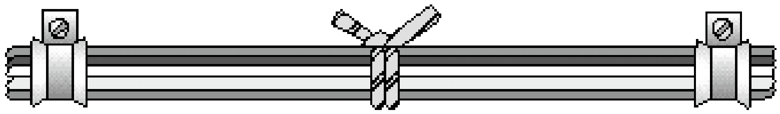
Soldering Aids—Use a heat shunt when you solder heat-sensitive components. It dissipates the heat, thereby preventing damage to the heat-sensitive component. Some type of soldering iron holder or guard should be used to prevent the operator from being burned.



Lacing Conductors—The purpose of lacing conductors is to present a neat appearance and to facilitate tracing the conductors when alterations or repairs are required. Flat tape is preferred for lacing instead of round cord. Cord has a tendency to cut into the wire insulation. The amount of flat tape or round cord required to lace a group of conductors is about two and one-half times the length of the longest conductor. A lacing shuttle is useful during the lacing operation to prevent the tape or cord from fouling. Wires should only be twisted prior to lacing if it is required, such as for filament leads in electron tube amplifiers. When lacing wire bundles containing coaxial cables, use the proper flat tape and do not tie the bundles too tightly. Never use round cord on coaxial cable. A single lace is started with a square knot and at least two marling Hitches. A double lace is required for wire bundles that are 1 inch or more in diameter. It is started with a telephone hitch. Cable groups are bound together by use of telephone hitch



Spot Ties—Spot ties are used when cable supports are used that are more than 12 inches apart.



Self-clinching Cable Straps—If self-clinching cable straps are used, they should be installed with the Military Standard hand tool designed for their use.

High-temperature Areas—When you are required to tie wire bundles in high-temperature operating areas, use only high-temperature, pressure-sensitive tape.

ANSWERS TO QUESTIONS Q1 THROUGH Q56.

A1. The connection must be both mechanically and electrically as strong as the conductor or device with which it is used

A2. By use of a wire-stripping tool

A3. Hot-blade stripper.

A4. Knife.

A5. To prevent damage to the tape insulation.

A6. To prevent the joint from being bulky.

A7. When wires are in conduit and a junction box is used.

A8. Fixture joint.

A9. Knotted tap joint.

A10. As a protective covering over the rubber tape.

A11. Requires relatively little operator skill to install.

A12. Spaghetti or heat-shrinkable tubing.

A13. 300º F

A14. 200 psig.

A15. No, it is done automatically by the petroleum abrasive compound that comes in the terminal or splices.

A16. Power-operated crimping tools.

A17. It gouges the terminal lug and causes deterioration.

A18. The use of preinsulated splices and terminal lugs.

A19. It has insulation support for extra supporting strength of the wire insulation.

A20. To identify wire sizes they are to be used on.

A21. Solder will not adhere to dirty, greasy, or oxidized surfaces.

A22. The coating of the material to be soldered with a light coat of solder.

A23. To prevent burning the insulation during the soldering process and to allow the wire to flex easier at a stress point.

A24. One-half the stripped length.

A25. Movement of the parts being soldered while the solder is cooling.

A26. The capacity of the soldering iron to generate and maintain a satisfactory soldering temperature while giving up heat to the joint being soldered.

A27. Although its temperature is as high as the larger irons, it does not have thermal inertia.

A28. The resistance of its heating element increases with rising temperature, thus limiting the current flow.

A29. File the tip until it is smooth and retin it.

A30. It will overheat and could burn the insulation of the wire being soldered.

A31. The heating and cooling cycles.

A32. Electronic components, such as resistors, capacitors, and transistors.

A33. The soldering tips are hot only during the brief period of soldering the connection, thus minimizing the chance of burning the wire insulation or connector inserts.

A34. The strands can fall into electrical equipment being worked on and cause short circuits.

A35. It enables the tip to be removed easily when another is to be inserted.

A36. Wrap a length of copper wire around one of the regular tips and bend to the proper shape for the purpose.

A37. Tin and lead.

A38. The solder dissolves a small amount of the copper, which combines with the solder forming a new alloy; therefore, the joint is one common metal.

A39. 60-percent tin and 40-percent lead (60/40 solder).

A40. It cleans the metal by removing the oxide layer and prevents further oxidation during the soldering.

A41. Noncorrosive, nonconductive rosin fluxes.

A42. To remove contaminants from soldered connections.

A43. To prevent damage to heat-sensitive components.

A44. To aid in tracing the conductors when alterations or repairs are required.

A45. Round cord has a tendency to cut into the wire insulation.

A46. Two and one-half times the length of the longest conductor in the group.

A47. To keep the tape or cord from fouling during the lacing operation.

A48. When required, such as for the filament leads in electron tube amplifiers.

A49. Do not tie too tightly and use the proper type of tape.

A50. With a square knot and at least two marling hitches drawn tightly.

A51. Bundles that are 1 inch or larger in diameter

A52. With a telephone hitch.

A53. They are bound together at intervals with telephone hitches.

A54. When wire bundles are supported by cable supports that are more than 12 inches apart.

A55. Military Standard hand tool.

A56. High-temperature, pressure-sensitive tape.

CHAPTER 3

SCHEMATIC READING

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to:

1. Recognize the marking system for cables to include shipboard and test equipment systems.
2. Recognize the marking system for wire to include aircraft and shipboard electronic equipment systems.
3. Recall the seven types of electrical diagrams and the functional design of each.
4. Recall basic safety practices and precautions for working around electrical and electronic systems.

SCHEMATIC READING

This chapter is divided into three subtopics—(1) cable and wire-marling systems, (2) electrical and electronic diagrams, and (3) safety precautions. First, we will discuss the systems used for marking cables and wires. We will then explain each of the types of diagrams you will encounter when troubleshooting, testing, repairing, or learning about circuit or system operation. Finally, we will briefly discuss safety practices relating to working around electrical and electronic systems.

CABLE- AND WIRE-MARKING SYSTEMS

Cables and wires are marked to give the technician a means of tracing them when troubleshooting and repairing electrical and electronic systems.

Numerous cable- and wire-marking systems are used in ships, aircraft, and equipment throughout the Navy. A few of these systems are briefly discussed here to acquaint you with how marking systems are used. For a specific system or equipment, you should refer to tile applicable technical manual.

CABLE-MARKING SYSTEMS

Two typical cable-marking systems you are likely to see are the (1) shipboard and (2) test equipment cable-marking systems.

Shipboard Cable-Marking Systems

Metal tags embossed with the cable markings are used to identify all permanently installed shipboard electrical cables. These cable tags (figure 3-1) are placed on cables close to each point of connection, and on both sides of decks, bulkheads, and other barriers to identify the cables. The markings on the cable tags identify cables for maintenance and circuit repairs. The tags show (1) the SERVICE LETTER, which identifies a particular electrical system, (2) the CIRCUIT LETTER or LETTERS, which identify a specific circuit within a particular system, and (3) the CABLE NUMBER, which identifies an individual cable in a specific circuit.

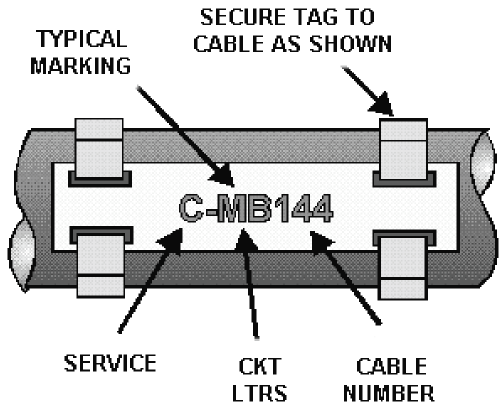


Figure 3-1.—Cable tag.

In figure 3-1, note that the cable is marked "C-MB144." The letter C denotes the service; in this case, the IC (interior communication) system. The letters MB denote the circuit; in this case, the engine-order circuit. The number 144 denotes cable number 144 of the MB circuit.

Q1. Why must cables and wires be identified?

Q2. Where would you find the wire identification system for a specific piece of equipment?

Q3. What does the cable number identify?

Test Equipment Cable-Marking Systems

View A of figure 3-2 shows apiece of test equipment that is used to check out electrical or electronic equipment or a system. It also shows the cables that are used to hook the tester to the equipment. The cables have metal or plastic tags at each end showing the cable number and the connector number.

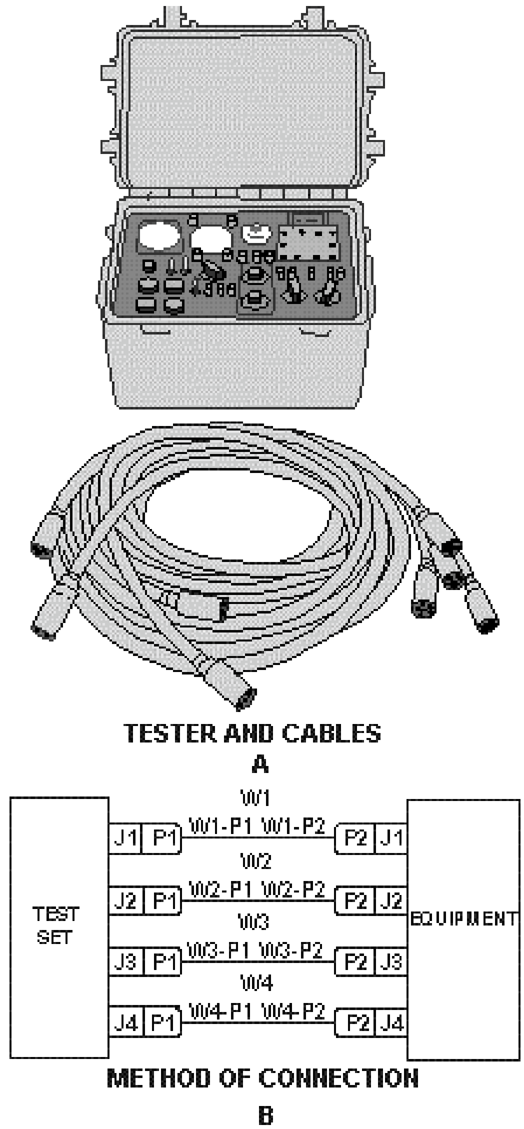


Figure 3-2.—Test equipment cable marking.

View B of figure 3-2 shows the method of connecting the tester to the piece of equipment to be tested. (For a specific tester, the technical manual supplied with the tester shows the method of connection.) The tester shown has four cables. These are numbered W1, W2, W3, and W4. Each cable has two connectors (plugs), one on each end, that are numbered P1 and P2. The cable tag on one end of the cable reads W1-P1, and the other end reads W1 -P2. As shown in the figure, W1-P1 is connected to the receptacle J1 on the tester. W1-P2 is then connected to receptacle J1 on the equipment to be tested. The same procedure is followed for connecting the remaining three cables. The hookup is then complete.

The shipboard and the test equipment cable markings just discussed are only two of many cable-marking systems you may encounter. There are too many systems to attempt to discuss them all. As stated earlier, you should study an equipment or installation technical manual before attempting repairs or connections.

WIRE-MARKING SYSTEMS

Wire-marking systems are used to identify wires in aircraft, shipboard electronic equipment, and power tool and appliance cables.

Aircraft Wire-Marking Systems

All aircraft wiring is identified on wiring diagrams exactly as the wire is marked in the aircraft. Each wire is coded by a combination of letters and numbers (figure 3-3) imprinted on the wire at prescribed intervals along the wire run.

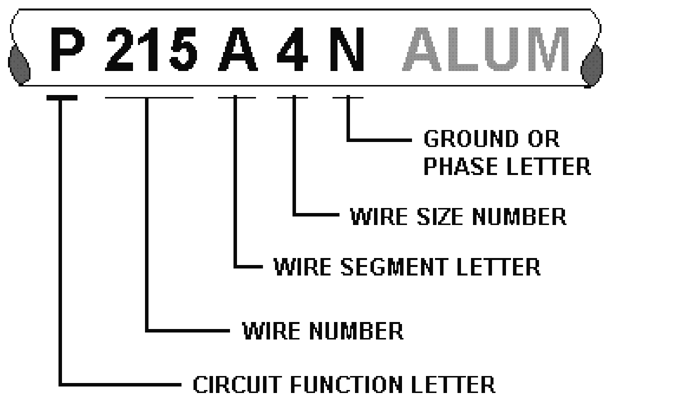


Figure 3-3.—Aircraft wire marking.

Look at figure 3-3. The circuit function letter (P in this example) identifies the basic function of the circuit concerned. The letter P indicates that the wire is in the dc power distribution system of the aircraft. The wire number, 215, indicates that it is the 215th wire in the dc distribution system. The wire segment letter (A) identifies the position of each wire segment of the circuit. The wire segments are lettered in alphabetical sequence and change each time the wire passes through a terminal or connector. For example, after the wire passes through the first terminal or connector, the segment letter A, as in this instance, would change to B.

The wire size number (4) is the AN wire size. AN wire sizes have more strands for flexibility and are slightly different in circular mil area than AWG (American Wire Gauge) wire sizes. The current-carrying capacity of each is almost the same. The last letter (N) is the ground or phase letter. The letter N identifies any wire that completes the circuit to the ground network of the aircraft.

In a 3-phase ac power distribution system, a phase letter (A, B, or C) is used as the last letter of the wire marking. If aluminum wire is used as the conductor, ALUMINUM or ALUM will be added as a suffix to the wire identification code.

Q4. If a wire passes through a connector what portion of the aircraft wire identification number changes?

Shipboard Electronic Equipment Wire-Marking Systems

The following explanation is an example of the type of conductor marking used in shipboard electronic equipment. These conductors may be contained in cables within the equipment. Cables within equipment are usually numbered by the manufacturer. These numbers will be found in the technical manual for the equipment. If the cables connect equipment between compartments on a ship, they will be marked by the shipboard cable-numbering system previously described.

On the conductor lead, at the end near the point of connection to a terminal post, spaghetti sleeving is used as a marking material and an insulator. The sleeving is marked with identifying numbers and letters and then slid over the conductor. The marking on the sleeving identifies the conductor connections both "to" and "from" by giving the following information (figure 3-4):

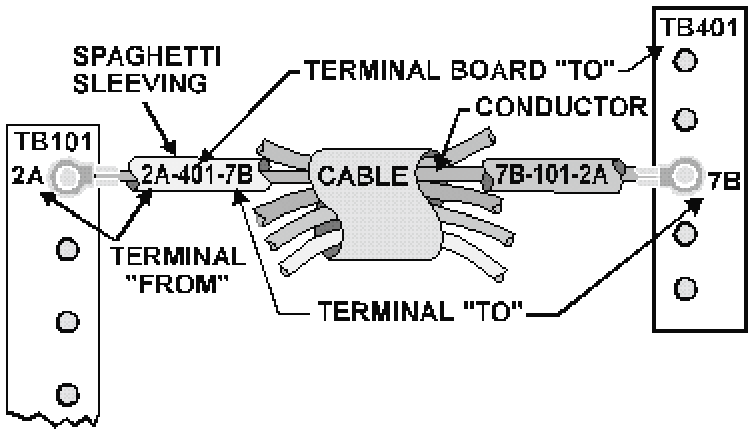


Figure 3-4.—Designating conductor marking between unlike terminals.

The terminal "from"

The terminal board "to"

The terminal "to."

These designations on the sleeving are separated by a dash. The order of the markings is such that the first set of numbers and letters reading from left to right is the designation corresponding to the terminal "from" which the conductor runs. Following this is the number of the terminal board "to" which the conductor runs. ("TB" is omitted when the sleeve is marked.) The third designation is the terminal "to" which the conductor runs.

For example, as shown in figure 3-4, the conductor is attached to terminal 2A of terminal board 101 (terminal "from" 2A on the spaghetti sleeving). The next designation on the sleeving is 401, indicating it is going "to" terminal board 401. The last designation is 7B, indicating it is attached "to" terminal 7B of TB 401. The spaghetti marking on the other end of the conductor is read the same way. The conductor is going "from" terminal 7B on terminal board 401 "to" terminal 2A on terminal board 101.

On occasion, it may be necessary to run conductors to units that have no terminal board numbers; for example, a junction box. In this case, an easily recognizable abbreviation may be used in place of the terminal board number on the spaghetti sleeving. The designation "JB2" indicates that the conductor is connected to junction box No. 2. A conductor to junction box No. 2 of a piece of equipment would be identified as shown in figure 3-5. In the same manner, a plug would be identified as "P." This P number would be substituted for the terminal board number marking on the sleeving.

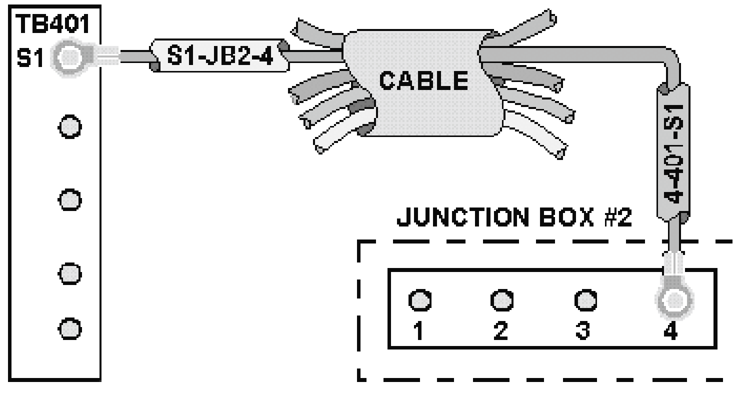


Figure 3-5.—Marking of conductors running to a junction box.

POWER TOOL AND APPLIANCE MARKING SYSTEMS

As with the wire- and cable-numbering systems discussed so far, there are many color-coding systems used in electrical and electronic applications. The color-coding system discussed here is the one used to code conductors for power tools and appliances.

An electrical power tool or appliance is required to have a three-wire cable. The conductors in the cable are color-coded black, white, and green. At shore bases or civilian facilities, one side of the electrical input is grounded. The grounded side is called the "common," and is color-coded white. The other side of the input is called the "line," or hot side, and is color coded "black". The green conductor is connected to ground and to the frame of the applicance or tool.

Aboard ship, neither side is grounded; therefore, both sides are considered the "fine," or both are hot. The black or the white conductor may be connected to either line, since there is no difference. The green conductor is connected to ground. Ground aboard ship is the ship's hull.

The purpose of the ground wire (green) is to prevent an electrical shock to the operator in case there is an electrical short to the frame of the appliance or tool.

Q5. What markings are found on spaghetti sleeving?

Q6. What is the purpose of the green conductor in a power tool or electrical appliance cable?

ELECTRICAL DIAGRAMS

It is absolutely essential that personnel in the electrical or electronic ratings be able to "read" (interpret) various types of electrical diagrams. Personnel working in these ratings commonly refer to all electrical diagrams as "schematics." This term is not correct, however. A schematic is a specific type of diagram with characteristics of its own and with a specific purpose. Each of the various diagrams discussed in this chapter has a specific purpose and distinguishing features that set it apart from the others. The diagrams discussed may be used for the following purposes:

- To learn a specific system operation
- To locate the components of a system

- To identify the components of a system
- To trace a circuit
- The troubleshoot equipment
- The repair equipment.

When you have completed this subject, you should be able to recognize the relationship between the various diagrams, their distinguishing features, and the purpose of each type of diagram. A continuing reference to the figures in the text should help you understand the subject matter more clearly.

We will use a simplified drawing of the electrical system of an automobile to explain the various electrical diagrams and how to "read" them.

PICTORIAL DIAGRAM

The simplest of all diagrams is the pictorial diagram. It shows a picture or sketch of the various components of a specific system and the wiring between these components. This simplified diagram provides the means to readily identify the components of a system, even if you are not familiar with their physical appearance. This type of diagram shows the various components without regard to their physical location, how the wiring is marked, or how the wiring is routed. It does, however, show you the sequence in which the components are connected.

Figure 3-6 is a pictorial diagram of an automobile starting and ignition system. If you are not already familiar with the components of this system, study the diagram. You should then be able to recognize the physical appearance of each component and its interconnections with the other components of the system.

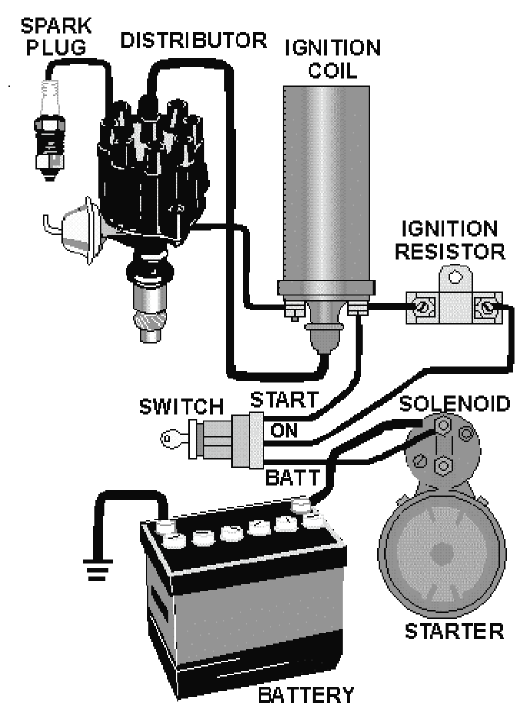


Figure 3-6.—Pictorial diagram of automotive starter and ignition systems.

ISOMETRIC DIAGRAM

The purpose of an isometric diagram is to assist you in locating a component within a system. If you do not know where to look for a component, the isometric diagram is of considerable value to you. This type of diagram shows you the outline of a ship, airplane, or piece of equipment. Within the outline are drawn the various components of a system in their respective locations. The isometric diagram also shows the interconnecting cable runs between these components.

Figure 3-7 is an isometric diagram of portions of the same automobile starting and lighting systems discussed in the pictorial diagram (figure 3-6). The battery, starter, and other components can now be seen, each in its actual location within the automobile.

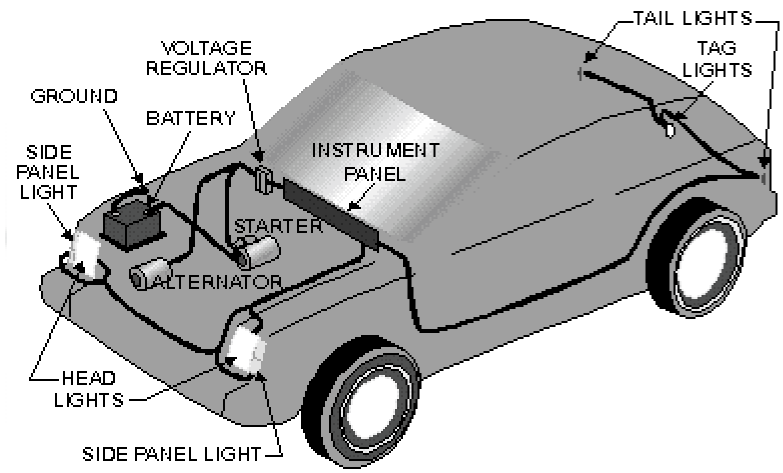


Figure 3-7.—Isometric diagram.

BLOCK DIAGRAM

A block diagram is used primarily to present a general description of a system and its functions. This type of diagram is generally used in conjunction with text material. A block diagram shows the major components of a system and the interconnections of these components. All components are shown in block form, and each block is labeled for identification purposes.

The block diagram shown in figure 3-8 is an illustration of an automobile's electrical power, starting, and ignition systems. It must be emphasized that the following explanation is primarily for the purpose of assisting you in learning to "read" or interpret a block diagram. The explanation of the functions of the automobile power, starting, and ignition systems is of secondary importance. By tracing from component to component in the block diagram and following the explanation, you are given a general description of the system functions. In addition, you should be able to understand the arrangement of the components in a block diagram.

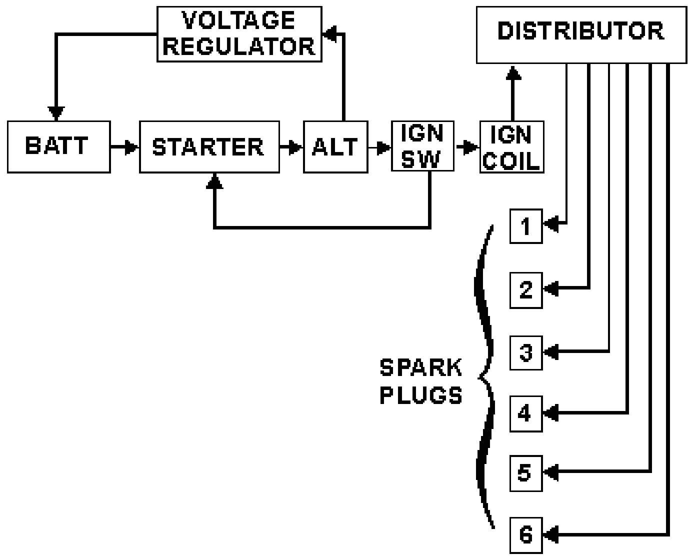


Figure 3-8.—Block diagram.

The battery is the initial source of power for the starter and ignition systems. The starter is turned by power from the battery when the ignition switch is turned to the START position. Power is also supplied, through the ignition switch, to the coil. From the coil, power is supplied to the distributor and finally to the spark plugs for ignition.

Once the engine is running, the starter is no longer required. The running engine acts as the prime mover for the alternator. (This is accomplished through a belt and pulley system attached to the engine's crankshaft.) The alternator now takes over as the power supplier for the ignition system. It supplies power through the ignition switch to the coil, from the coil to the distributor, and finally from the distributor to the spark plugs. At the same time, the alternator supplies power back through the voltage regulator to the battery for charging purposes. This completes the cycle until the engine is shut down and started again.

Note that the engine is not shown in the block diagram as the prime mover for the alternator. The engine is a mechanical rather than an electrical function. The illustrated block diagram is of the electrical system only. There are block diagrams that show strictly mechanical components or both mechanical and electrical components.

SINGLE-LINE DIAGRAM

The single-line diagram is used basically for the same purpose as the block diagram. When used with text material, it gives you a basic understanding of the functions of the components of a system.

There are two major differences between the single-line diagram and the block diagram. The first difference is that the single-line diagram uses symbols, rather than labeled blocks, to represent components. Second, the single-line diagram shows all components in a single line (figure 3-9). There are no interconnections shown for selected components as were shown on the block diagram (for example, alternator to voltage regulator and back to the battery). The single-line diagram is very simplified and should be used primarily to learn (in very broad terms) the function of each of the various components as a part of the total system.

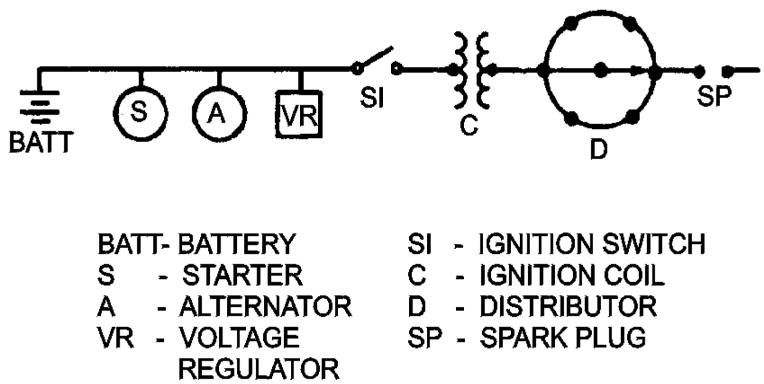


Figure 3-9.—Single-line diagram.

Q7. What type of electrical diagram is used to identify the components of a system?

Q8. What type of diagram is used to find the location of a component?

Q9. What types of diagrams are the most convenient from which to learn the basic Functions of a circuit?

SCHEMATIC DIAGRAM

The schematic diagram shows, by means of graphic symbols, the electrical connections and functions of a specific circuit arrangement. The schematic diagram is used to trace the circuit and its functions without regard to the actual physical size, shape, or location of the component devices or parts. The schematic diagram is the most useful of all the diagrams in learning overall system operation.

Figure 3-10 is a schematic diagram of an automobile electrical system. The automobile electrical system uses the frame of the automobile as a conductor. The frame is called the ground side. Figure 3-10 shows all the electrical components grounded on one side. The negative side of the battery is also grounded. Therefore, the frame is the negative conductor of the system. The opposite side of each of the components is connected through switches to the positive side of the battery. For the purpose of teaching schematic reading, we will discuss only the lighting system and engine instruments.

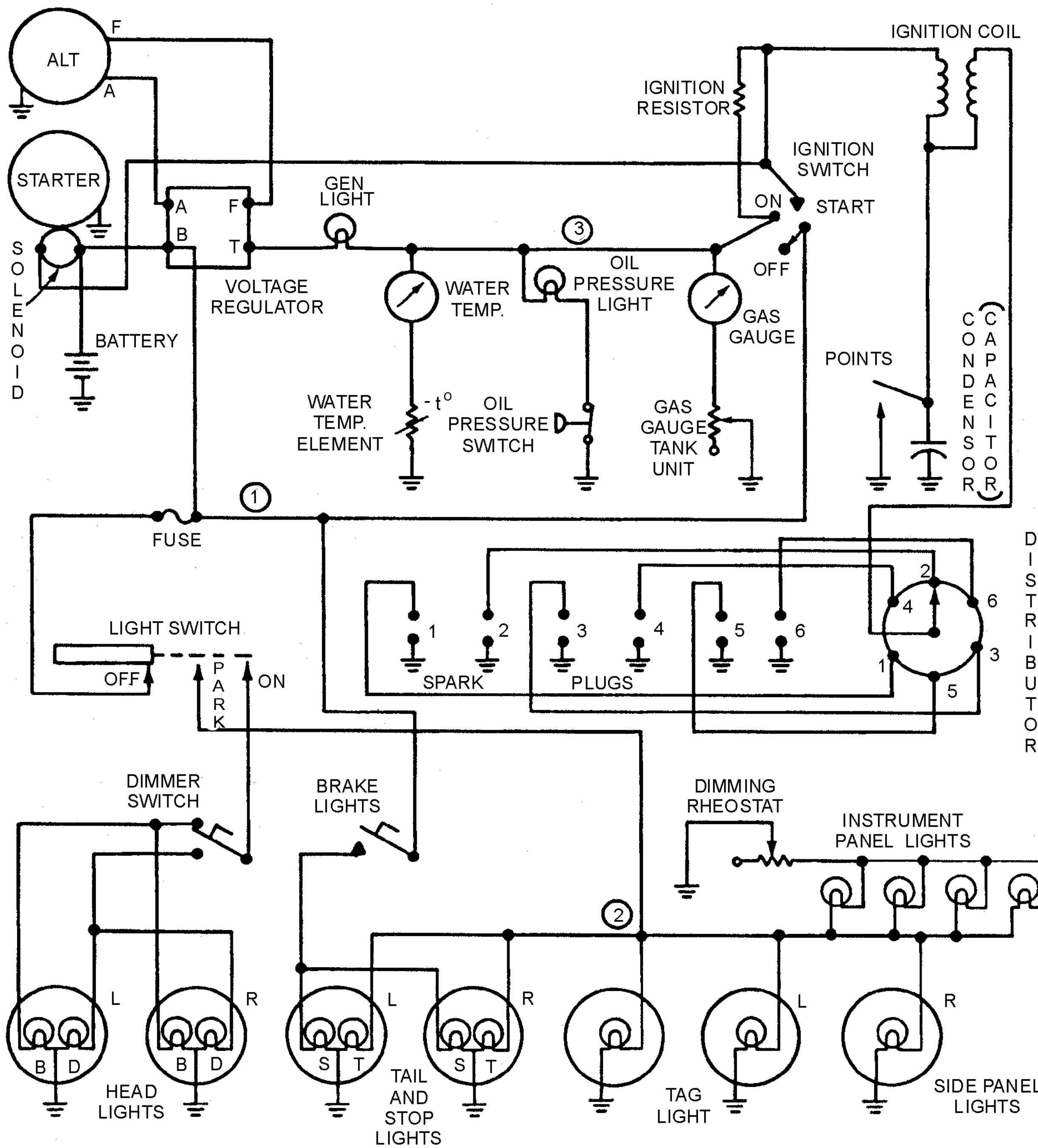


Figure 3-10—Schematic diagram.

The positive side of the 12-volt battery is connected to the starter solenoid, then to terminal B of the voltage regulator, and then down to point (1). (It should be noted that points (1), (2), (3), and so on, normally are not indicated on the schematic. They are shown here only to help you follow the diagram.) Therefore, if no faults are in the system, point (1) has a 12-volt positive potential at all times. This positive potential can be traced through the fuse to the OFF position of the light switch. The dashed line indicates the mechanical linkage of the switch. When the switch is pulled to the first position (park), +12 volts are applied to point (2). It can now be seen that the tail lights (T), the tag light, the side panel lights, and the instrument lights have +12 volts applied. The opposite side of each light is grounded. The instrument panel lights are grounded through the dimming rheostat. This completes the path for current flow from the negative side of the battery, through all the light bulbs (lamps), back to the positive side of the battery. If no faults exist, the lamps will light.

When the light switch is pulled to the next position (on), the bar on the switch contacts the "off," "park," and "on" contacts of the switch. The lights that were illuminated before are still on, and the + 12 volt potential is now applied to the bright (B) side of the headlights through the dimmer switch. Since the headlights are also grounded on one side, there is now a complete path for current flow, and the headlights also light. If the dimmer switch is actuated, the positive potential is switched from the bright filament to the dim filament of the headlights, and the lights dim.

The brake-light switch has +12 volts applied from point (1), directly to the stop lights (not fused). If the brake pedal is pressed, the switch is actuated, and the +12 volts are applied to both stop lights (S). Because one side of each light is tied to ground, there is a path for current flow, and the lights will light. If the dimming rheostat for the instrument lights is turned in the direction that increases the resistance, more voltage is dropped across the rheostat, less across the lights, and the lights will get dimmer.

The +12 volts at point (1) are also supplied to the OFF position of the ignition switch. When the ignition switch is turned on, the +12 volts are felt at point (3). This is a common point to all the engine instruments.

The gas gauge is a galvanometer with the dial graduated according to the amount of fuel in the tank. The gas gauge tank unit is a rheostat mechanically linked to a float in the gas tank. When the tank is full, the float rises to its highest level and positions the movable arm of the rheostat to a position of minimum resistance. This allows maximum current flow through the galvanometer, and the dial rests at the "full" mark on the gas gauge. As fuel is used by the engine, the float lowers, increasing the resistance of the rheostat to ground. This reduces the current through the galvanometer, and the dial shows a lesser amount of fuel.

The oil-pressure light gets its ground through a normally closed pressure switch. (When no pressure is applied, the switch is closed.) When the engine is started, the oil pressure increases and opens the switch. This turns the light off by removing the ground.

The water-temperature gauge is a galvanometer like the gas gauge, except its dial is graduated in degrees of temperature. The water-temperature element is a thermistor with a negative temperature coefficient. (A thermistor is a semiconductor device whose resistance varies with temperature.) When the engine is cold, the resistance of the thermistor is at a maximum. This reduces the current through the galvanometer, and a low temperature is indicated on the dial. As the water temperature of the engine increases, the resistance of the thermistor decreases. This allows more current to flow from ground through the galvanometer, and the temperature on the dial shows an increase.

On the voltage regulator shown, the "T" terminal is grounded anytime the alternator does not have an output. This gives the alternator light a ground and causes it to illuminate.

Q10. What type of diagram is the most useful in learning the overall operation of a system?

Q11. Refer to the schematic diagram in figure 3-10. If the ignition switch is placed in the ON position and all the engine instruments operate properly except the gas gauge, where would the fault probably be?

Q12. If the fuse shown on the schematic (figure 3-10) opens, what lights will operate?

WIRING DIAGRAM

A wiring diagram is a detailed diagram of each circuit installation showing all of the wiring, connectors, terminal boards, and electrical or electronic components of the circuit. It also identifies the wires by wire numbers or color coding. Wiring diagrams are necessary to troubleshoot and repair electrical or electronic circuits. The wiring diagram for an automobile is shown in figure 3-11. It shows all the electrical components and that the interconnecting wiring is color coded.

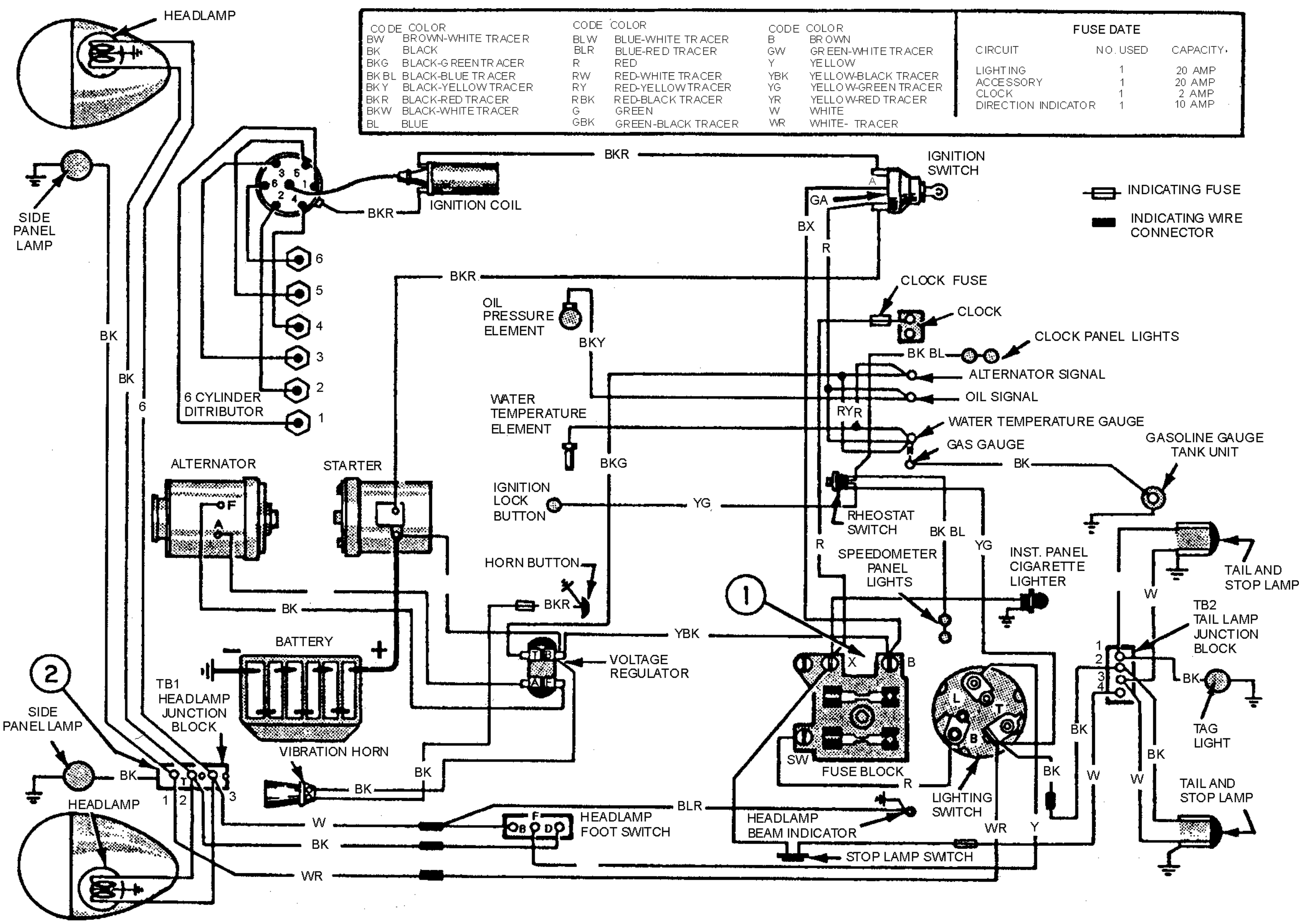


Figure 3-11.—Wiring diagram.

You should use the schematic diagram previously discussed to determine where the trouble might be in the circuit when a malfunction occurs. The schematic diagram does not show the terminals, connector points, and so forth, of the circuit. Therefore, you must go to the circuit wiring diagram to determine where to make the voltage or resistance checks in the circuit when troubleshooting. Following is an example of how to use a schematic diagram in conjunction with a wiring diagram to troubleshoot a circuit.

In the discussion of schematic diagrams, you will recall that when the light switch is pulled to the PARK position, the tail lights, side panel lights, tag light, and the instrument lights come on. Now, suppose that when the light switch is pulled to the PARK position all the lights come on, except the tag light. Referring to the schematic diagram (figure 3-10), you will recall that when the light switch is placed in the PARK position, +12 volts are applied to point (2). If all the lights come on except the tag light, then the fault must be between point (2) and the tag light ground.

On the schematic shown in figure 3-11, you can see that there are numerous connections to point (2). Point (2) on the wiring diagram is actually composed of three different functions: terminal 1 of TB 1 (the head lamp junction block), terminals 1 and 2 of TB2 (the tail lamp junction block), and the "T" terminal of the light switch; all correspond to point (2) on the schematic. The fault here is in the tag light, which normally receives its +12 volts from terminal 1 of TB2.

To use a voltmeter to find the fault, place the positive lead of the voltmeter to the ground terminal of the tag light and the negative lead to the frame. The voltmeter should read zero, because there should be no difference of potential between the two points. If the meter reads a voltage, the ground lead is either open or has a high-resistance connection. If the meter reads zero, as it should, you will have to go to another test point. In this case, place the positive voltmeter lead on the positive terminal of the tail light. If the voltmeter reads +12 volts, the light bulb is probably burned out or the light socket is defective. If the voltmeter reads zero, then the open is between terminal 1 of TB2 and the light.

TERMINAL DIAGRAM

A terminal diagram is useful when connecting wires to terminal boards, relays, switches, and other components of a circuit. Figure 3-12 shows two typical terminal diagrams. View A of the figure shows the wire numbers connected to each terminal of a terminal board. View B shows the different color codes of the wires that are connected to a relay.

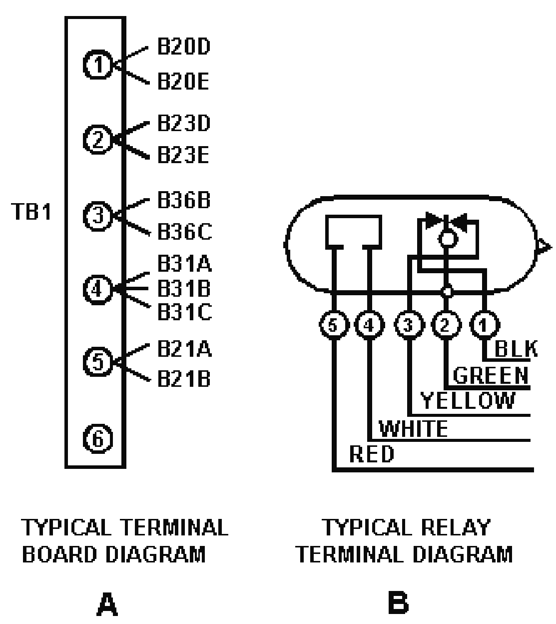


Figure 3-12.—Terminal diagrams.

This has been a brief overview of the use and interpretation of electrical diagrams. The diagrams used were selected because of their simplicity and ease of interpretation. Many diagrams you will encounter are far more complex. Start with the simpler diagrams you will be working with on the job. Your proficiency in using the more complex diagrams will increase with experience and study.

Q13. What type of diagram is the most detailed?

Q14. Why must a wiring diagram be used in conjunction with a schematic to troubleshoot a system?

Q15. What type of diagram would be most useful for wiring a relay into a circuit?

SAFETY

The Secretary of the Navy, in establishing a Department of the Navy safety program, stressed, "Safety is an inherent responsibility of command...." He further outlined that, "Assignment of safety responsibility at all echelons of command is a basic requirement." This means responsibility right down through the lowest rated personnel in the command. Most noncombat accidents can be prevented if all personnel cooperate in eliminating unsafe conditions and acts. To this end, each individual is responsible for understanding and applying safety rules, standards, and regulations in all activities. Safety consciousness will help prevent personal injury and damage to property.

Some safety precautions applicable to this module deal with fumes from synthetic insulation, breathing asbestos fibers, and working around/with electrical and electronic circuits and portable power tools.

SYNTHETIC INSULATION

Almost without exception, the fumes from synthetic materials, such as plastics in high-temperature environments, are objectionable from the standpoint of health and safety. Fluoroplastics (FEP and polytetrafluoroethylene) resist decomposition at higher temperature better than most other plastics.

Exposure to fumes when working with fluoroplastics may cause a temporary flu-like condition similar to the metal fume fever (or "foundryman's fever"). These symptoms are commonly called polymer fume fever. They do not ordinarily occur until several hours after exposure, and pass within 36 to 48 hours, even in the absence of treatment.

One of the largest uses of fluoroplastics is as a wire and cable insulation. When insulated wiring is installed, soldering is a routine fabricating procedure, as is the use of a heated element to remove insulation. In neither of these operations do the combined effects of temperature, quantity of resin, and exposure time produce toxic conditions of significance, as long as normal ventilation is maintained.

Any special practices or precautions that may be required should follow the same common sense rules that apply to all soldering jobs. Prolonged soldering in confined spaces with restricted air circulation will require some ventilation for personal comfort. The same is true for open shop areas where a number of personnel are engaged in soldering or hot-wire stripping. Normal ventilation for personal comfort usually provides adequate safety. However, it is recommended that a small duct fan or "elephant trunk" exhaust be used at the workbench during soldering or wire stripping to carry away any toxic vapors.

ASBESTOS

Although asbestos-free products have been developed, older products containing asbestos materials still exist and continue to be used in the Navy. One such product is asbestos insulation used on wiring in high-temperature areas aboard ships and in aircraft.

Because of the serious health hazards of asbestos exposure, the government has imposed strict occupational health and environmental protection standards for the control of asbestos. These standards must be strictly enforced and followed by all Navy personnel.

Asbestos is a general term used to describe several fibrous mineral silicates. Major uses of asbestos include asbestos cement products, floor tiles, fireproofing, high-temperature insulation, asbestos cloth, friction materials (such as brake linings and clutch facings), various gasket materials, and miscellaneous other products.

Inhaling asbestos fibers can produce disabling or fatal fibrosis of the lungs. Fibrosis of the lungs (asbestos) comes from inhaling asbestos fibers. Asbestos is a factor in the development of lung cancer as well as cancer of the gastrointestinal tract. It may take 20 to 40 years between initial exposure to asbestos and the appearance of a cancerous condition. Know where asbestos is in your environment and avoid or take precautions to prevent exposure.

ELECTRICAL OR ELECTRONIC CIRCUITS AND PORTABLE POWER TOOLS

When working on electrical or electronic circuits, you must observe certain general precautions. The following is a listing of common sense safety precautions that you must observe at all times:

- Remember that electrical and electronic circuits often have more than one source of power. Take time to study the schematics or wiring diagrams of the entire system to ensure that all power sources are deactivated.
- Remove all metal objects from your person.
- Use one hand when turning switches on or off. Keep the doors to switch and fuse boxes closed, except when working inside or replacing fuses.
- After first making certain that the circuit is dead, use a fuse puller (figure 3-13) to remove cartridge fuses.
-

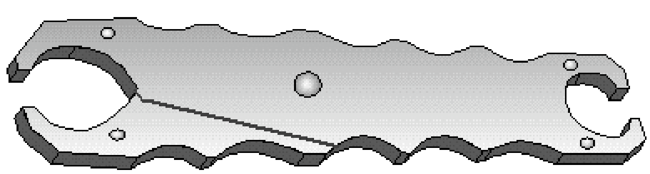


Figure 3-13.—Fuse puller.

●All supply switches or cutout switches from which power could possibly be fed should be secured in the OFF or OPEN (safety) position and tagged (figure 3-14). The tagging procedures must be done in accordance with the appropriate manual or instruction for your field of training.

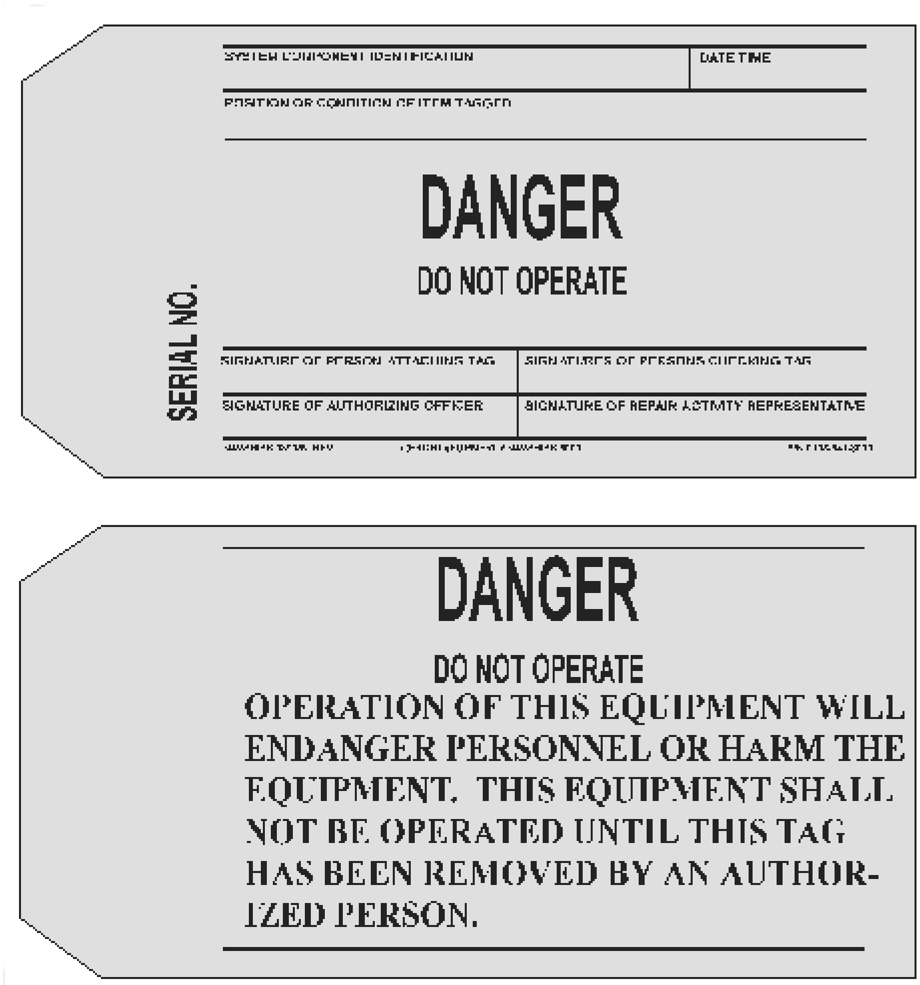


Figure 3-14.—DANGER tag.

- Keep clothing, hands, and feet dry if possible. When it is necessary to work in wet or damp locations, use a dry platform or wooden stool to sit or stand on, and place a rubber mat or other nonconductive material on top of the wood. Use insulated tools and molded insulated flashlights when you are required to work on exposed parts. In all instances, repairs on energized circuits must not be made with the primary power applied, except in an emergency, and then only after specific approval has been given by your commanding officer. When approval has been obtained to work on equipment with the power applied, keep one hand free at all times (BEHIND YOU OR IN YOUR POCKET).

- Never short out, tamper with, or block open an interlock switch.

- Keep clear of exposed equipment; when it is necessary to work on it, work with one hand as much as possible.

- Avoid reaching into enclosures, except when it is absolutely necessary. When reaching into an enclosure, use rubber blankets to prevent accidental contact with the enclosure.

- Make certain that equipment is properly grounded.

- Turn off the power before connecting alligator clips to any circuit.

- Never use your finger to test a "hot" line. Use approved voltmeters or other voltage-indicating devices.

High Voltage Precautions

In addition to observing the general precautions just discussed, you must observe the following additional precautions when working with high voltages:

- Do NOT work with high voltage by yourself; have another person (safety observer), qualified in first aid for electrical shock, present at all times. This individual, stationed nearby, should also know the circuits and location of the switches controlling the equipment, and should be given instructions to pull the switch immediately if anything unforeseen happens.

- Always be aware of the nearness of high-voltage lines or circuits. Use rubber gloves where applicable and stand on approved rubber matting. Not all so-called rubber mats are good insulators.

- Always discharge the high voltage from components or terminals by using a safety probe.

- Do NOT hold the test probe when circuits over 300 volts are tested.

Soldering Irons

When using a soldering iron, always keep in mind the following precautions and procedures:

- To avoid burns, ALWAYS ASSUME that a soldering iron is hot.

- Never rest a heated iron anywhere but on a metal surface or rack provided for this purpose. Faulty action on your part could result in fire, extensive equipment damage, and serious injuries.

- Never use an excessive amount of solder, since drippings may cause serious skin or eye burns.

- Do not swing an iron to remove excess solder. Bits of hot solder that are removed in this manner can cause serious skin or eye burns. Hot solder may also ignite combustible materials in the work area.

- When cleaning an iron, use a cleaning cloth, but DO NOT hold the cleaning cloth in your hand. Always place the cloth on a suitable surface and wipe the iron across it to prevent burning your hand.

- Hold small soldering jobs with pliers or a suitable clamping device to avoid burns. Never hold the work in your hand.

- Do not use an iron that has a frayed cord or damaged plug.

- Do not solder components unless the equipment is disconnected from the power supply circuit. Serious burns or death can result from contact with a high voltage.

- After completing the task requiring the use of soldering iron, disconnect the power cord from the receptacle and, when the iron has cooled, stow it in its assigned storage area.

Portable Electric Power Tools

Navy specifications for portable electric power tools require that the electric cord of each tool have a distinctively marked ground wire in addition to the conductors for supplying power to the tool. (Double-insulated portable electric tools obtained from sources qualified under the applicable military specification are exempt from this grounding requirement.) The end of the ground wire within the tool must be connected to the metal housing of the tool. The other end must be connected to a positive ground. For this ground connection, specifically designed ground-type plugs and receptacles, which automatically make this connection when the plug is inserted into the receptacle, must be used. These grounded-type receptacles must be installed for all power outlets. When installed, they must be used with the grounded-type plugs to ground portable tools and equipment. If grounded-type receptacles have not yet been installed, they must be installed as soon as possible. Portable tools not provided with the ground-type plug, and miscellaneous portable electric equipment that does not have a cord with a ground conductor and grounded plug, must be given a three-conductor cord with a standard Navy grounded-type plug. The ground wire must be connected to a positive ground.

Care must be exercised in connecting the plugs and cords. The grounding conductor of the cord must be connected to the ground contact of the plug at one end and to the metal equipment housing at the other end. The cord must be arranged so as not to create a tripping hazard. If the conductor connected to the metallic equipment housing is inadvertently connected to a line contact of the plug, a dangerous potential would be placed on the equipment casing. This could result in a fatal shock to the operator. If the cord is pulled loose from the plug, only a qualified electrician is authorized to repair it.

If the grounded-type plugs and receptacles have not been installed in the spaces where a portable tool is to be used, other types of plugs and receptacles may be used only if a separate ground wire is connected between the tool housing and a positive ground. When the tool cord does not include an extra wire for grounding, an additional insulated wire should be connected between the metal housing of the tool and ground. If the tool housing has two or more conducting parts that are not electrically connected, each part must be connected to the ground wire. Connection of the ground wire to the tool housing and to the ground must be by means of screws or bolts. The use of spring clips for either end of the grounding wire is prohibited.

When the ground connection is to be made by means other than a contact in the plug and receptacle, care must be taken to secure a good contact between the ground wire and the metal by scraping away paint from the metal to ensure a clean surface. The ground connection must be made before inserting the power supply connecting plug, and the plug must be pulled out before removing the ground connection. Frequent inspections of each of the connections of a portable electric tool must be made to ensure that the supply cord and its connections within the tool are suitably insulated and that the ground connection is intact.

The safety precautions just discussed are to protect you and your shipmates. Follow safety precautions to the letter. DO NOT TAKE CHANCES. Carelessness could cost you your life.

Q16. What safety precaution must you observe when soldering or hot-wire stripping fluoroplastic insulated wire?

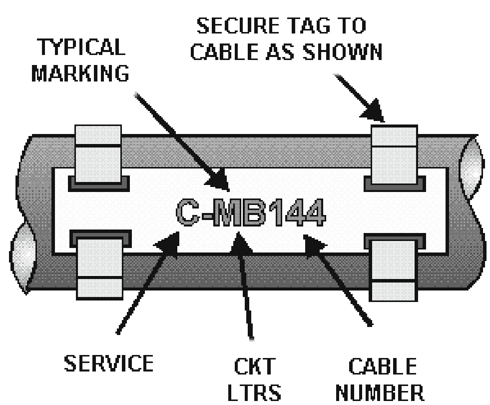
Q17. What must be used to test an activated circuit?

Q18. How should excess solder be removed from a hot soldering iron?

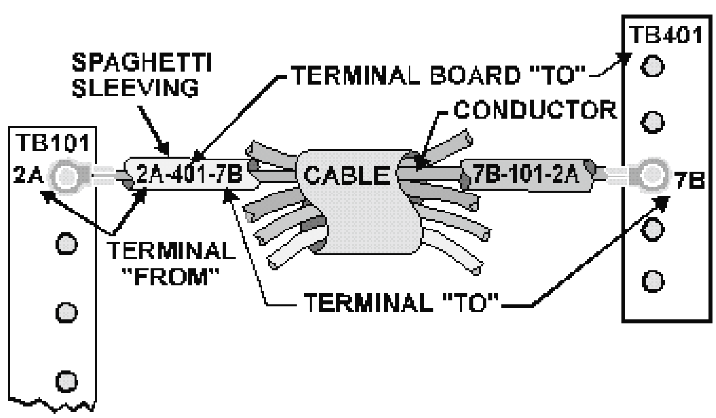
SUMMARY

In this chapter, we have discussed some typical cable- and wire-marking systems, electrical diagrams, and some basic safety precautions. A brief summary of these subjects follows:

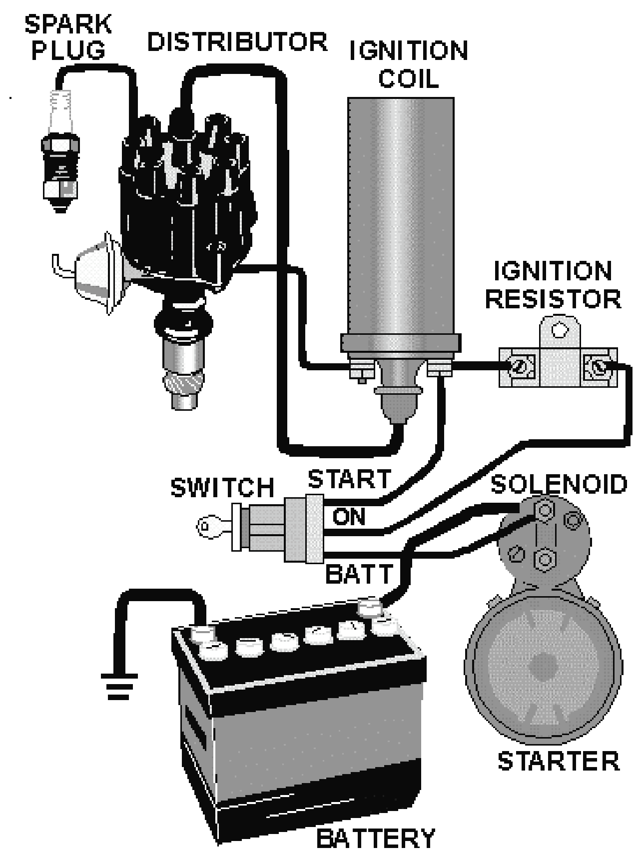
Cable- and Wire-Marking Systems—Cables and wires must be identified to provide the technician with a means of tracing them when troubleshooting and repairing electrical and electronic systems. The cable and wire-marking systems discussed in this chapter are typical systems. The number of systems used throughout the Navy is too numerous to discuss. For the cable or wire identification for a specific piece of equipment, consult the technical manual for that equipment. One wire identification system you will surely come in contact with is the color coding of wires used on electrical power tools and appliances. Remember, the purpose of the green conductor in a power tool or appliance cable is to prevent electrical shock to the operator in case there is an electrical short to the frame of the appliance or tool.



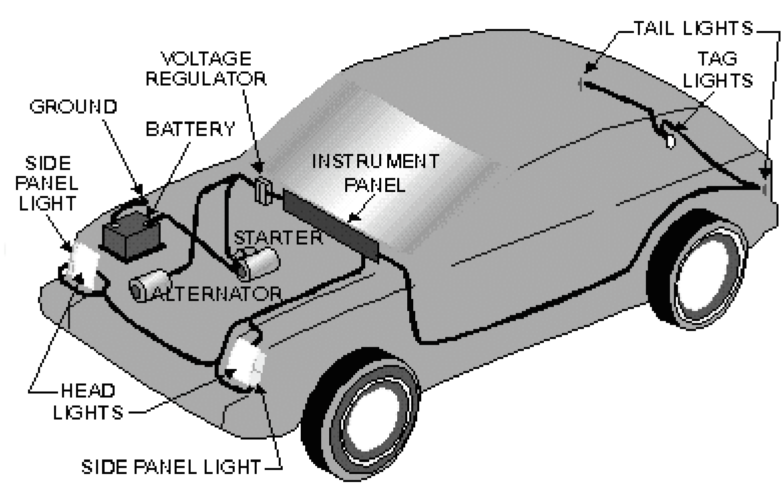
Electrical Diagrams—Examples of electrical diagrams you will be required to "read" (interpret) and their uses are as follows:



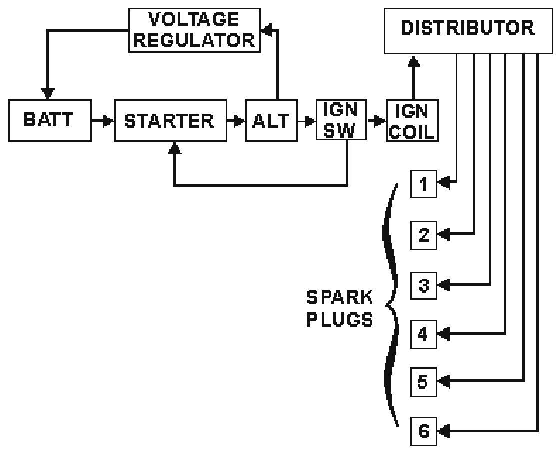
Pictorial Diagram—Shows a picture or sketch of the various components of a system and the wiring between the components. This diagram is used to identify the components of a system.



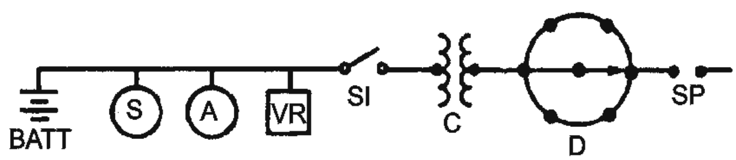
Isometric Diagram—Shows the outline of a ship, airplane, or piece of equipment. This diagram shows the components and the cable runs between the components. This diagram is used to locate components in a system.



Block Diagram—Shows the components in block form. Block diagrams are used in conjunction with text material. They are used to present a general description of a system and its functions.



Single-Line Diagram—Used for essentially the same purpose as the block diagram-to show the basic functions of a circuit.

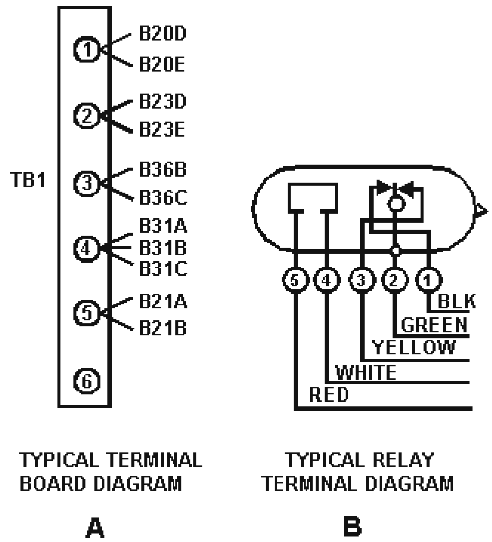


BATT- BATTERY
S - STARTER
A - ALTERNATOR
VR - VOLTAGE REGULATOR
SI - IGNITION SWITCH
C - IGNITION COIL
D - DISTRIBUTOR
SP - SPARK PLUG

Schematic Diagram—Shows, through graphic symbols, the electrical connections and functions of a specific circuit arrangement. It is used to trace the circuit without regard to the physical size, shape, or location of the component devices or parts. A schematic diagram shows the overall operation of a system. It is used during troubleshooting to identify possible circuit malfunction locations.

Wiring Diagram—Is a detailed diagram of each circuit installation showing all wiring, connectors, terminal boards, and the electrical or electronic components of the circuit. It also identifies the wire-by-wire numbers or color coding. This diagram must be used in conjunction with a schematic diagram to troubleshoot a system in order to find the test point for voltage and resistance checks.

Terminal Diagram—Is used in connecting wiring to terminal boards, relays, switches, and other components of a circuit.



Safety—All individuals are responsible for understanding and complying with safety standards and regulations established to prevent injury to themselves and others and damage to property and equipment.

Having safe working habits and adhering to safety precautions protects YOU and YOUR SHIPMATES. Follow safety precautions to the letter. DO NOT TAKE CHANCES. Carelessness could cost you your life.

ANSWERS TO QUESTIONS Q1. THROUGH Q18.

A1. To provide the technician with a means to trace the wires when troubleshooting and repairing electrical and electronic systems.

A2. In the technical manual for the equipment.

A3. Individual cable in a specific circuit.

A4. Wire segment letter.

A5. The conductor connections both "to" and "from."

A6. To prevent electrical shock to the operator in case there is an electrical short to the frame of the appliance or too.

A7. A pictorial diagram.

A8. An isometric diagram.

A9. Block or single-line diagram.

A10. A schematic diagram.

A11. Between point (3) and the gas gauge tank unit ground.

A12. Only the brake lights.

A13. Wiring diagram.

A14. To find the test points.

A15. Terminal diagram.

A16. Adequate ventilation.

A17. Approved meters or other indicating devices.

A18. By use of a cleaning cloth.

APPENDIX I

GLOSSARY

ALUMINUM CREEP—The movement of aluminum wire away from a point where pressure is applied. (2) The retreat of heated aluminum wire as it cools.

AMBIENT TEMPERATURE—The "surrounding temperature"—as the temperature surrounding a conductor in a compartment or within a piece of equipment.

AMERICAN WIRE GAUGE (AWG)— The standards adopted in the United States for the measurement of wire sizes.

ANTISEIZE COMPOUND—A silicon-based, high-temperature lubricant applied to threaded components to facilitate their removal after being subjected to rapid heating and cooling.

ASBESTOS—A fiber-like mineral, noncombustible and nonconductive, used as an insulating material.

BLOCK DIAGRAM—A diagram in which the major components of a piece of equipment or of a system are represented by squares, rectangles, or other geometric figures, and the normal flow of a signal or current is represented by lines.

BRAID—The weaving of metal or cloth material as an outer coating to a cable; prevents cable damage from moisture and rough treatment.

BRANCH—An individual current path in a parallel circuit.

BUS BAR—A heavy cooper strap or bar used to connect several circuits together when a large current-carrying capacity is required.

CABLE—Either a stranded conductor (single-conductor cable) or a combination of conductors insulated from one another (multiple-conductor cable). Small cable sizes are called stranded wire or cords.

CENTIMETER CUBE—A unit of volume for large rectangular or square conductors. The cross-sectional area equals 1 square centimeter with a length of 1 centimeter.

CIRCULAR MIL—The area of a circle having a diameter of 1 mil. The standard unit of measurement of wire cross-sectional area. One circular mil equals .7854 square miles.

CIRCULAR-MIL-FOOT—A unit of volume of a conductor having a cross-sectional area of 1 circular mil and a length of 1 foot.

COAXIAL CABLE—A cable made up of a center conductor separated from an outer conductor by a dielectric material. Normally used for radio-frequency transmission.

COMPRESSED AIR/NITROGEN HEATING TOOL—A portable source of heat for use with heat-shrinkable products.

CONDUIT—A tubular raceway, usually metal or plastic, for enclosing wires or cables.

CONDUCTANCE—The ability of a material to conduct or carry an electric current. It is the reciprocal of the resistance of the material.

CORONA—The discharge of electricity from a conductor with a high potential.

CURRENT RATING—The safe current-carrying capacity of a wire or cable on a continuous basis.

DIELECTRIC STRENGTH—The ability of an insulator to withstand a potential difference without breaking down. (Usually expressed in terms of voltage).

DUCTILE—Easily drawn out-as to form filaments or wires.

ELECTRICAL SYMBOLS—Graphic symbols used to illustrate the various electrical or electronic components of a circuit.

ELECTROLYSIS—The process of changing the chemical composition of a material by passing an electric current through it.

ELECTROSTATIC STRESS—The force exerted on an insulator by the voltage in a conductor.

ENAMEL—Synthetic compound of cellulose acetate (wood pulp and magnesium). Used to insulate wire in meters, relays, and motor windings.

EXTRUDED POLYTETRAFLUOROETHYLENE—A high-temperature insulation used extensively in aircraft and equipment installations. (Emits dangerous fumes when heated.)

FEP Fluorinated Ethylene Propylene)—A synthetic type of insulation.

FIBROUS BRAID—An outer covering used to protect the insulating material of a conductor. Commonly made from cotton, linen, silk, rayon, or fiberglass.

FLUX—A material that removes oxides from surfaces to be joined by soldering or welding.

GALVANOMETER—A meter used to measure small values of current by electromagnetic or electrodynamics means.

HEAT-SHRINKABLE TUBING—A plastic tube that, when heated, shrinks to encapsulate, protect, or insulate connections, splices, terminations, and other configurations.

HEAT SHUNT—A device (preferably a clip-on type) used to absorb heat and protect heat-sensitive components during soldering.

INSULATION—Materials used to coat or wrap conductors to prevent current leakage.

INSULATION RESISTANCE—The resistance offered by an insulating material to current leakage.

ISOMETRIC DIAGRAM—A diagram showing the outline of a ship or aircraft or equipment, and the location of equipment and cable runs.

JUNCTION BOX—A box with a cover for joining different runs of wire or cable and for providing space for the connection and branching of the enclosed conductors.

LACING SHUTTLE—A device upon which lacing may be wound to prevent fouling the tape or cord and facilitate the lacing process. (Usually made from brass, aluminum, fiber, or plastic.)

LEAD SHEATH—A continuous jacket of lead molded around a single-conductor or multiple-conductor cable. Generally used to ensure that conductors are protected from water or extensive moisture.

MAGNET WIRE—Wire coated with an enamel insulation and used in coils, relays, transformers, motor windings, and so forth.

METALLIC ARMOR—A protective covering for wires or cables. Made as a woven wire braid, metal tape, or interlocking metal cover. Made from steel, copper, bronze, or aluminum.

MIL—The diameter of a conductor equal to 1/1000th (.001) inch.

MIL-FOOT—A unit of measurement for conductors. (Diameter of 1 mil, 1 foot in length.)

MILITARY SPECIFICATIONS (MIL-SPEC)—Technical requirements and standards adopted by the Department of Defense, which are to be met by vendors selling materials to DOD.

MULTICONDUCTOR—More than one conductor.

NEGATIVE TEMPERATURE COEFFICIENT—The temperature coefficient expressing the amount or reduction in the value of a quantity, such as resistance for each degree of increase in temperature.

OXIDATION—The addition of atmospheric oxygen to metal to form rust, or to cause a breakdown in the internal construction of the metal.

PETROLEUM ABRASIVE COMPOUND—A compound that causes a grinding action during the crimping operation and removes the oxide film from the aluminum.

PICTORIAL DIAGRAM—A diagram showing pictorial sketches of the parts of a piece of equipment and the electrical connections between the parts.

POWER LOSS—The electrical power supplied to a circuit that does no work, usually dissipated as heat.

RECEPTIVITY—The reciprocal of conductivity. (See also SPECIFIC RESISTANCE.)

RHEOSTAT—(1) A resistor whose value can be varied. (2) A variable resistor used to adjust the current in a circuit.

RHO—Greek letter "rho" (ρ). Used in electricity and electronics to represent the specific resistance of a substance.

SCHEMATIC—A diagram that shows, in graphic symbols, the electrical connections and functions of a specific circuit arrangement. The schematic diagram makes tracing the circuit and its functions easier without regard to the physical size, shape, or location of the component device or parts.

SINGLE-LINE DIAGRAM—A diagram that shows, in single lines and graphic symbols, the course of an electric circuit or system of circuits and the component devices or parts used in the circuit(s).

SOLDERING—The joining of metals with a higher melting point than solder.

SPAGHETTI TUBING—(See TRANSPARENT TUBING.)

SPECIFIC RESISTANCE—The resistance, measured in ohms, of a unit volume of a substance to the flow of electric current. (The unit volume used is generally the Circular-Mil-Foot.)

SPLICE—A joint formed by connecting two or more conductors.

SQUARE MIL—The area of a square, the sides of which are each equal to 1 mil. One square mil is equal to 1.2732 circular mils.

STRANDED CONDUCTOR—A conductor composed of a group of wires. The wires in a stranded conductor are usually twisted together and not insulated from each other.

STRANDS—Fine metallic filaments twisted together to form a single wire.

TEMPERATURE COEFFICIENT OF RESISTANCE—The amount of increase in the resistance of a 1-ohm sample of a conductor per each degree of rise in temperature above 0º C.

TENSILE STRENGTH—The greatest stress a substance can withstand along its length without tearing apart.

TERMINAL—A point of connection for two or more conductors in an electrical circuit.

TERMINAL BOARD—(Also called a terminal strip.) An insulating base or slab equipped with terminals for connecting wiring.

TERMINAL DIAGRAM—A diagram of a switch, relay, terminal board, or other component showing the connections to the equipment.

TERMINAL LUG—A device attached to a conductor for connecting to a terminal.

THERMAL INERTIA—The capacity of a soldering iron to generate and maintain a satisfactory soldering temperature while giving up heat to the material being soldered.

THERMISTOR—A semiconductor device whose resistance varies with temperature.

THERMOPLASTIC—A synthetic mixture of rosins that is flexible and used as an insulating material. Generally used as an insulator for low- and medium-range voltages.

TINNING—The process of applying a thin coat of solder to materials prior to their being soldered (for example, application of a light coat of solder to the filaments of a conductor to hold the filaments in place prior to soldering the conductor).

TOXIC VAPORS—Vapors emitted by a substance that can do bodily harm.

TRANSPARENT TUBING—(Also known as spaghetti tubing.) A plastic tubing used for insulation and wire marking.

UNIT SIZE—The standards adopted to make comparisons between things of like value (for example, the unit size for conductors is the mil-foot).

VARNISHED CAMBRIC—Cotton cloth coated with insulation varnish. An insulation used on high-voltage conductors.

VOLTAGE DROP—The difference in voltage between two points. It is the result of the loss of electrical pressure as a current flows through a resistance.

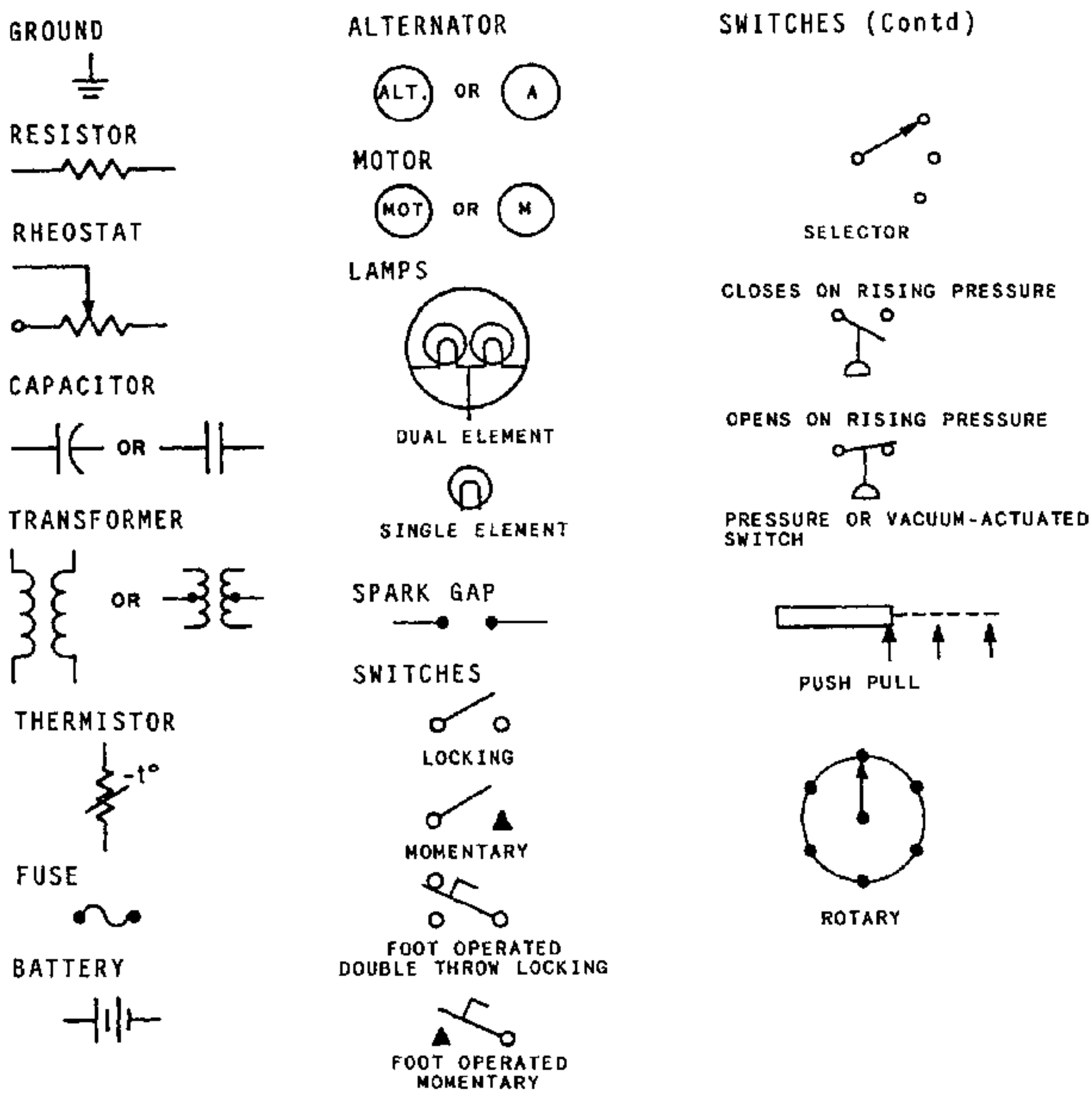
WIRE—A solid or stranded group of solid, cylindrical conductors having low resistance to current flow, with an associated insulation.

WIRE STRIPPERS—A tool used to strip insulation from wire.

WIRING DIAGRAM—A diagram that shows the connections for an installation or for its component devices or parts. The diagram may show internal or external connections, or both, and also show the details needed to make or trace the connections involved.

APPENDIX II

ELECTRICAL AND ELECTRONIC SYMBOLS



APPENDIX III

REFERENCES USED TO DEVELOP THE TRAINING MANUAL

Aircraft Electric and Electronic Wiring, NAVAIR 01-1A-505, Naval Air Systems Command, Washington, DC, June 1988.

Dictionary of Standard Terminal Designations for Electronic Equipment, NAVSEA 0967-LP-146-0010, Bureau of Ships, Washington DC, March 1954

Electronic Installation and Maintenance Book, NAVSEA 0967-LP-000-0110, Naval Sea Systems Command, Washington, DC, September 1977.

National Electrical Code, National Fire Protection Association, Battery March Park, Quincy, MD, 1990.

Navy Occupational Safety and Health (NAVOSH) Program Manual, SPAWARINST 5100.9C, Space and Naval Warfare Systems Command, June 1987.

Subsystem Design and Engineering Standards for Common Long Haul/Tactical Cable and Wire Communications, MIL-STD-188-112, Department of Defense, Washington, DC, August 1983.

MODULE 4 INDEX

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook Assignment: Chapter 1, “Electrical Conductors,” pages 1-1 through 1-27.
Chapter 2 “Wiring Techniques,” pages 2-1 through 2-23.

1-1. Why has a "unit size" for conductors been established?

1. To compare the size and resistance of one conductor with that of another
2. To establish a uniform style for conductors
3. To determine the requirements for conductors
4. To ensure all conductors are interchangeable

1-2. What is the decimal equivalent of one (1) mil?

1. 1.000 in.
2. 0.100 in.
3. 0.010 in.
4. 0.001 in.

1-3. If a conductor has a diameter of 1/4 inch, what is its diameter in mils?

1. 250.0 mil
2. 25.0 mil
3. 2.50 mil
4. 0.250 mil

1-4. What is the definition of a mil foot?

1. A conductor .001 foot in length with a diameter of .001 millimeter
2. A conductor 1 foot in length with a diameter of .001 foot
3. A conductor 1 foot in length with a diameter of 1 mil
4. A conductor .001 foot in length with a diameter of .001 inch

1-5. A square mil is defined as the area of a square, the sides of which are each equal in length to what dimension?

1. 1 mil-foot
2. 1 mil
3. 1.0 inch
4. .001 mil

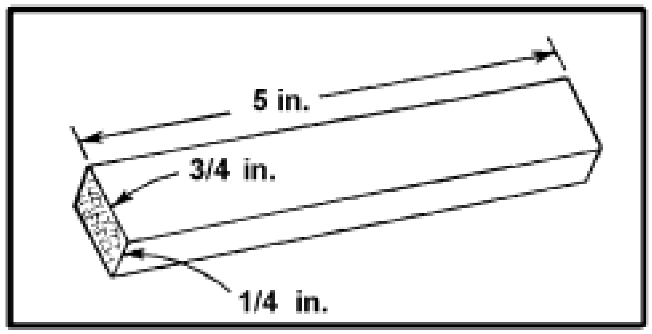


Figure 1A.—Rectangular Bar Conductor.

IN ANSWERING QUESTION 1-6, REFER TO FIGURE 1A.

1-6. What is the cross-section area of the illustrated rectangular bar conductor in square mils?

1. 937,500,000 sq. mil
2. 3,750,000 sq. mil
3. 1,250,000 sq. mil
4. 187,500 sq. mil

1-7. A circular mil is defined as the area of a circle having what dimension?

1. A radius of 1 mil
2. A diameter of 1 mil
3. A circumference of 1 mil
4. A chord of 1 mil

1-8. If you have a 12-strand conductor and each strand has a radius (one half the diameter) of 2/10 inch, what is the circular mil area of the conductor?

1. 1,920,000 cir. mil
2. 57,680 cir. mil
3. 48,000 cir. mil
4. 2,400 cir. mil

1-9. What is the square mil area for the conductor explained in question 1-8?

1. 244,344,097 sq. mil
2. 1,507,965 sq. mil
3. 61,115 sq. mil
4. 45,239 sq. mil

1-10. What is the definition of specific resistance?

1. The resistance of a length of conductor, at a given temperature, to voltage
2. The resistance of a cross-sectional area of a conductor, at a given temperature, to the flow of current
3. The resistance of a unit volume of a substance to the flow of current expressed in ohms

1-11. What factor(s) must be known to compute the resistance of a conductor?

1. The length of the conductor
2. The cross-sectional area of the conductor
3. The specific resistance of the substance of the conductor
4. Each of the above

THIS SPACE LEFT BLANK INTENTIONALLY.

Gage Number	Ohms per 1,000 ft.		Pounds per 1,000 ft.
	25 deg. C	65 deg. C	
6	.403	.465	79.5
7	.508	.586	63.0
8	.641	.739	50.0
9	.808	.932	39.6
10	1.020	1.180	31.4
11	1.280	1.480	24.9
12	1.620	1.870	19.8

Figure 1B

REFER TO THE FOLLOWING INFORMATION AND USE FIGURE 1B TO ANSWER QUESTION 1-12. ASSUME THE TEMPERTURE TO BE 25º C AND YOU HAVE TWO NEW COPPER CONDUCTORS TO RUN. RUN NO. 1 IS TO BE 2500 FEET LONG USING NO. 12 WIRE. RUN NO. 2 IS TO BE 6000 FEET LONG USING NO. 7 WIRE.

1-12. What is the approximate total resistance of (a) Run No. 1 and (b) Run No. 2?

1. 9.72 ohms (b) 1.27 ohms
2. 4.68 ohms (b) 3.52 ohms
3. 4.05 ohms (b) 3.05 ohms
4. 1.12 ohms (b) 4.98 ohms

1-13. When a wire gauge is used to determine the size of a wire, the measurement should be made in what part of the gauge?

1. In the slot
2. In the semicircular opening
3. Either 1 or 2 above, depending on wire size

THIS SPACE LEFT BLANK INTENTIONALLY.

A.	Conductor size
B.	Material of conductor
C.	Load requirement
D.	Wire ductility
E.	Type of insulation
F.	Location of wire in circuit
G.	Source voltage

Figure 1C

IN ANSWERING QUESTION 1-14, REFER TO FIGURE 1C.

1-14. Which of the following factors is used to determine the current rating of a wire?

1. B, C, E, and G
2. A, D, E, and G
3. B, D, E, and F
4. A, B, E, and F

1-15. What types of insulation are best suited for use in a high-temperature environment?

1. Cotton, polytetrafluoroethylene, and oiled paper
2. FEP, silicone rubber, and extruded polytetrafluoroethylene
3. Oiled paper, FEP, and rubber
4. Rubber. polytetrafluoroethylene, and silk

1-16. What is the "ambient" temperature of a conductor?

1. It is the normal temperature of a conductor through which current is flowing
2. It is the heat generated by external sources and affecting a conductor's temperature
3. It is the maximum heat a conductor can withstand according to its current rating
4. It is the temperature at which the insulation of a conductor begins to break down

1-17. Which of the following metals is the best conductor of current?

1. Aluminum
2. Copper
3. Silver
4. Manganin

A.	High cost
B.	High ductility
C.	High tensile strength
D.	Easily soldered
E.	Very light in weight
F.	Difficult to solder
G.	Reduces corona
H.	Generally uninsulated
I.	High conductivity

Figure 1D

FIGURE 1D LISTS SOME OF THE ADVANTAGES AND DISADVANTAGES OF VARIOUS CONDUCTORS. IN ANSWERING QUESTIONS 1-18 AND 1-19, REFER TO FIGURE 1D.

1-18. When electricity is carried over long distances, which of the following are advantages for using aluminum as the conductor as opposed to copper?

1. B and C
2. D and E
3. C and I
4. E and G

1-19. Which of the following are advantages for using copper as the conductor as opposed to aluminum?

1. B, D, and I
2. C, E, and G
3. B, E, and H
4. A, F, and I

1-20. Because its resistance changes very little with temperature changes, what conductor is best suited for use in measuring instruments?

1. Copper
2. Aluminum
3. Manganin
4. Silver

1-21. If a 75-ohm sample of copper wire at 0ºC is heated to 30ºC, what is the approximate total resistance? (The temperature coefficient of resistance of copper at 0ºC is 0.00427).

1. 0.32 ohm
2. 9.61 ohms
3. 65.39 ohms
4. 84.61 ohms

1-22. What definition best describes an insulating material?

1. A material that has a very high resistance
2. A material that has a very low resistance
3. A material that has a very high conductivity
4. A material that has a very low dielectric strength

1-23. The dielectric strength of an insulating material is a measurement of the material's ability to resist electrostatic stress caused by what factor?

1. Resistance
2. Current
3. Voltage
4. Chafing or friction

1-24. Insulation resistance can best be defined as the ability of an insulating material to resist what action?

1. Current leakage
2. Electrostatic stress
3. Breakdown by voltage
4. External factors acting upon the conductor

1-25. For a material to be a good insulator, what two properties are most important?

1. High dielectric strength and low insulation resistance
2. High dielectric strength and high insulation resistance
3. Low dielectric strength and high insulation resistance
4. Low dielectric strength and low insulation resistance

1-26. When rubber is used as the insulating material over a copper conductor, why is a thin coating of tin used between the two materials?

1. To decrease the electrostatic stress
2. To increase the insulation resistance of the rubber
3. To prevent a chemical action from taking place between the copper and rubber
4. To reduce the amount of insulating material required.

1-27. What is the NEC for a rubber heat-resistance compound?

1. RHH
2. RWH
3. RTW
4. RWT

1-28. Latex rubber is a high-grade compound consisting of what percentage of unmilled grainless rubber?

1. 70%
2. 80%
3. 90%
4. 95%

1-29. Plastic insulation is normally used for what levels of voltage?

1. Very high to high
2. High to medium
3. Medium to low
4. Low to very low

1-30. When dealing with NEC type designators for thermoplastics, the letter "W" stands for what type of insulation?

1. Oil-resistant
2. Moisture-resistant
3. Heat-resistant
4. Asbestos

1-31. When you work on synthetic insulated wiring, what safety precaution must be observed?

1. Wear protective goggles at all times
2. Avoid breathing the vapors when the insulation is heated
3. Wear a dust mask in confined spaces
4. Wear protective gloves if there are cuts or abrasions on your hands

1-32. What nonmetallic material is most commonly used to protect wires and cables?

1. Rubber
2. Jute and asphalt covering
3. Fibrous tape
4. Fibrous braid

1-33. What is the common name for woven covers?

1. Yarn
2. Loom
3. Fibrous tape
4. Unspun felted cotton

1-34. What percentage of tin is used in alloy-lead sheathing?

1. 6%
2. 2%
3. 8%
4. 4%

1-35. Why is the use of asbestos being discontinued as an insulating material in the Navy?

1. It breaks down rapidly with continued use
2. It is not as effective as other types of insulation
3. It has not proven suitable for a shipboard environment
4. It poses a health hazard to personnel

1-36. What happens to asbestos insulation when it gets wet?

1. Its insulation resistance becomes too high
2. It emits dangerous fumes
3. It acquires too great a dielectric strength
4. It becomes a conductor

1-37. What insulating materials are best suited for use with high voltage?

1. Thermoplastic and rubber
2. Varnished cambric and oil impregnated paper
3. Teflon and silk
4. Silk and cotton

1-38. What is the common name for enamel-insulated wire?

1. Winding wire
2. Motor wire
3. Magnet wire
4. Coil wire

1-39. What types of conductor protection are normally used for shipboard wiring?

1. Wire braid armor and nonmagnetic steel tape
2. Lead cable and Jute
3. Jute and nonmagnetic steel tape
4. Lead sheathing and rubberized tape

1-40. What are the basic requirements for a splice or terminal connection?

1. To be mechanically and electrically effective
2. To be preinsulated and nonconductive
3. To have minimum cost and maximum efficiency
4. To have circuit continuity and minimum cost

1-41. The preferred method for removing insulation from most types of insulated wire is by using what tool?

1. Razor blade
2. Electrician's pliers
3. Wire stripper
4. Knife

1-42. When a wire is insulated with glass braid or asbestos and requires stripping, which of the following tools should NOT be used?

1. Knife
2. Rotary wire stripper
3. Hand wire stripper
4. Hot-blade wire stripper

1-43. What is the preferred tool to use to strip aluminum wire?

1. Knife
2. Rotary wire stripper
3. Hand wire stripper
4. Hot-blade wire stripper

1-44. When a Western Union splice is used to connect two wires, why should the twisted ends of the wires be pressed down as close as possible to the straight portion of the wire?

1. To increase the strength of the splice
2. To prevent the wires from puncturing the tape covering
3. To minimize the resistance change in the circuit
4. To increase the dielectric strength of the insulation

1-45. When multiconductor cables are spliced, why are the splices staggered?

1. To prevent possible shorting between conductors
2. To increase the strength of the individual splices
3. To decrease insulated resistance
4. To reduce the overall size of the joint

1-46. When is a rattail joint normally used?

1. When a branch circuit is required and a junction box is used to join conduit
2. When a Western Union splice would be too bulky
3. When asbestos or glass braid is used as insulation
4. When the branch wire will be subjected to a heavy strain

1-47. If a fluorescent light is to be attached to a branch circuit, which of the following splices should normally be used?

1. Staggered splice
2. Knotted tap joint
3. Western Union splice
4. Fixture joint

1-48. When is a knotted tap joint normally used?

1. When a branch circuit is joined to a continuous wire (main wire)
2. When a Western Union splice would be too bulky
3. When a lighting fixture is joined to a branch circuit
4. When a wire nut is used to complete the joint

1-49. Which of the following splices is NOT butted?

1. Fixture joint
2. Rattail joint
3. Knotted tap joint
4. Western Union splice

1-50. Why is friction tape used over a splice?

1. To provide a protective covering over the rubber tape
2. To provide maximum insulation to the splice
3. To prevent shock when latex rubber is used
4. To reduce the amount of rubber tape required

1-51. Why would you use a crimped terminal instead of a soldered terminal?

1. Connections can be made more rapidly
2. Less operator skill is required
3. Connections are more uniform in construction
4. Each of the above

1-52. When noninsulated splices and terminal lugs are insulated, what types of insulation are most commonly used?

1. Rubber tape and friction tape
2. Spaghetti and heat-shrinkable tubing
3. Spaghetti and friction tape
4. Rubber tape and heat shrinkable tubing

1-53. When heat-shrinkable tubing is used, what is the maximum temperature to which the wire should be subjected?

1. 180ºF
2. 220ºF
3. 300ºF
4. 340ºF

1-54. When a large aluminum terminal lug or splice is installed, why is it NOT necessary to clean the aluminum wire?

1. It is done automatically by the tubing
2. The wire is cleaned by the abrasive compound in the lug or splice
3. Oxide film does not form on aluminum
4. The insulation used provides the necessary cleaning agent

1-55. When aluminum terminals lugs or splices are installed, which of the following tools is generally recommended for use?

1. Pliers
2. Power crimping tool
3. Hand crimping tool
4. Vise grips

1-56. Why is a lockwasher NOT used with an aluminum terminal?

1. The washer will reduce conductivity at the terminal
2. The washer will gouge the lug and cause deterioration
3. The washer will set up a corrosive action between dissimilar metals
4. The washer will increase resistance and heat causing eventual failure

1-57. The most common method of terminating and splicing wires is by using preinsulated terminal lugs and splices.

1. True
2. False

1-58. Which of the following is an advantage of using preinsulated splices and terminal lugs?

1. Heat shrinkable tubing is not required
2. Spaghetti is not required
3. They offer extra supporting strength to the wire insulation
4. Each of the above

1-59. Color codes are used on preinsulated terminal lugs and splices to indicate what information?

1. The resistance, in ohms, of the lugs and splices
2. The style of crimping tool to be used
3. The type of circuit in which they are to be used
4. The wire sizes on which they are to be used

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Wiring Techniques," pages 2-24 through 2-53.
Chapter 3, "Schematic Reading," pages 3-1 through 3-24.

2-1. Why must materials to be soldered be cleaned just prior to the soldering process?

1. To ensure the solder will adhere to the surface
2. To prevent the solder from becoming brittle from impurities and eventually failing
3. To prevent an uneven flow of solder to the surface
4. Each of the above

2-2. What is meant by the term "tinning"?

1. Removing the oxide coating of the material to be soldered
2. Preheating the material to be soldered to remove any impurities left from the stripped insulation
3. Coating the material to be soldered with a light coat of solder
4. Applying pure tin to the material to be soldered to ensure adherence of the solder

2-3. When a wire is soldered to a connector, why should the wire be stripped approximately 1/32 inch longer than the depth of the solder barrel?

1. To prevent burning the wire insulation
2. To allow the wire to flex more easily at stress points
3. Both 1 and 2 above
4. To prevent the flux from touching the insulation

2-4. When a wire has been properly stripped and is to be soldered to a connector, what total length of the exposed wire should be tinned?

1. One-third
2. One-half
3. Two-thirds
4. The entire exposed length

2-5. What action generally causes a fractured solder joint?

1. Movement of the soldered parts during the cooling process
2. Application of too much heat to the parts
3. Introduction of impurities to the joint from dirty solder or flux
4. Application of too much solder to the joint

2-6. What term defines the capacity of a soldering iron to generate and maintain a satisfactory soldering temperature while giving up heat to the joint being soldered?

1. Iron current flow
2. Thermal inertia
3. Resistance soldering
4. Self-regulating heat

THIS SPACE LEFT BLANK INTENTIONALLY.

2-7. Why should a small wattage soldering iron NOT be used to solder large conductors?

1. The current flow is limited
2. The iron cannot reach a high enough temperature
3. The iron cannot maintain a satisfactory soldering temperature while giving up heat to the conductor
4. The tip of a small wattage iron is too small for large conductors

2-8. Which of the following features BEST describes a well designed soldering iron?

1. It may be used for both large and miniature soldering jobs
2. It is light weight with an all-purpose tip
3. It can be automatically switched from a low wattage to a high wattage output
4. 4.It has a built-in self-regulating element

2-9. What should be done with the removable tip of a soldering gun after it becomes pitted?

1. Dip it in flux and continue to use it
2. Discard the tip and replace it
3. Grind the tip down to the next smaller size and reuse it
4. File the tip smooth and retin it

2-10. If, during the soldering process, the soldering gun switch is depressed for longer than 30 seconds, what danger exists?

1. The insulation of the wire may be burned
2. An oxide film will rapidly form on the conductor
3. The flux may ignite
4. The finger switch may be locked in the depressed position from the heat

2-11. What condition causes the nuts or screws which hold the tip of a soldering gun to loosen?

1. The trigger is depressed for too long a period
2. The gun is pulsed too often
3. The heating and cooling cycle loosens them
4. The gun is used for soldering items beyond its capacity

2-12. Which of the following electronic components should NOT be installed or removed by the use of a soldering gun?

1. Transistors
2. Resistors
3. Capacitors
4. Each of the above

2-13. Why are resistance soldering irons safer for electrical equipment components than other soldering irons or guns?

1. The current flow is very low
2. The tips are hot only during the brief period of actual soldering
3. The transformer provides a high voltage for a measured period of time
4. The tips are made from highly conductive ferrous iron which heat and cool very rapidly

2-14. For which of the following reasons is antisieze compound used with a pencil iron equipped with removable tips?

1. To allow the tip to be easily removed
2. To prevent the tip form loosening during repeated soldering operations
3. To minimize the number of times the tip must be retinned
4. Each of the above

2-15. If you do not have a suitable tip for desoldering, how can one be improvised?

1. File an available tip down to the desired size
2. Bend a piece of wire to the desired shape and insert the ends of the wire into the barrel
3. Bend a piece of copper wire to the desired shape after wrapping it around a regular tip
4. File a piece of round stock, preferably steel, to the desired shape and insert it in the barrel

2-16. What are the two metals most often used to form soft solder?

1. Lead and antimony
2. Tin and lead
3. Bismuth and tin
4. Tin and cadmium

2-17. What chemical or physical change causes a joint of soldered copper conductors to become one common metal?

1. A physical change takes place as the solder flows between the molecules of copper joining them together when cooled
2. A physical change takes place as both metals displace one another
3. A chemical change takes place as the copper is dissolved into the solder thereby forming an alloyed metal
4. A chemical change takes place when the additional materials added to the solder are heated causing a gluing effect between the solder and the copper

2-18. When you solder electrical connectors, splices, and terminal lugs, what type of solder should you use?

1. 65/35 solder
2. 63/37 solder
3. 60/40 solder
4. 57/43 solder

2-19. Why is flux used in the soldering process?

1. It dilutes the molten solder and allows it to flow
2. It acts as a cleaning agent to remove oxide
3. It acts as the bonding agent between the solder and metal
4. It forms a conductive bond between the metal and the solder

2-20. When electrical and electronic components are soldered, what type of flux must be used?

1. Hydrochloric acid
2. Sal ammoniac
3. Zinc chloride
4. Rosin

2-21. What two properties must a solvent have?

1. Noncorrosive-nonconductive
2. Corrosive-conductive
3. Noncorrosive-conductive
4. Corrosive-nonconductive

2-22. Why are solvents used in the soldering process?

1. To remove the flux from the metal surface being soldered
2. To remove contaminants from the soldered connection
3. To dilute the flux and allow it to flow freely
4. To improve the conductivity of the flux

2-23. Why are heat shunts used in the soldering process?

1. To conduct heat from the component being soldered back to the iron
2. To increase the temperature of the soldering iron or gun
3. To prevent damage to adjacent heat-sensitive components
4. To decrease the temperature to the conductor being soldered

2-24. For which of the following reasons are conductors laced together?

1. To present a neat appearance
2. To help support each other
3. To aid in tracing conductors
4. Each of the above

2-25. Although it may be used, why is the use of round cord discouraged for lacing conductors?

1. It is bulkier than the flat type
2. It is more difficult to handle
3. It is not fungus resistant
4. It has a tendency to cut into wire insulation

2-26. If you are preparing to single lace conductors, what total length must the lacing be in relationship to the longest conductor?

1. One and one-half times the length
2. Twice the length
3. Two and one-half times the length
4. Five times the length

2-27. Why is a lacing shuttle used when conductors are laced in bundles?

1. It helps prevent the conductors from twisting together
2. It helps prevent the cord or tape from fouling
3. It keeps the "lay" of the cord or tape
4. It ensures that hitches are evenly spaced

2-28. Under certain circumstances, it is permissable to twist conductors together prior to lacing.

1. True
2. False

2-29. When coaxial cables are laced, the use of round cord is prohibited. What additional precaution must be observed?

1. Coaxial cables may not be laced with other conductors
2. Bundles containing coaxial cables must be double laced
3. Half hitches must be used in place of marling hitches
4. Coaxial cables must not be tied so tightly as to deform the dielectric

2-30. How should a single lace be started?

1. With a square knot and two marling hitches
2. With a marling hitch and a telephone hitch
3. With a telephone hitch and two half hitches
4. With a square knot and two half hitches

2-31. Under which of the following conditions should a double lace be used?

1. Three coaxial cables form the bundle
2. A maximum of six conductors form the bundle
3. The bundle is larger than one inch in diameter
4. The bundle exceeds 10 feet in length

2-32. How should a double lace be started?

1. With a square knot
2. With a half hitch
3. With a marling hitch
4. With a telephone hitch

2-33. How should laced cable groups that run parallel to each other be bound together?

1. With marling hitches
2. With telephone hitches
3. With square knots
4. With half hitches

2-34. What tool or technique should be used to install self-clinching cable straps?

1. Military standard hand tool
2. Circle snips
3. Electrician's pliers
4. Hand installation

2-35. If a bundle of conductors passes through a very high-temperature area, what restraint should be used to tie the bundle?

1. High-temperature pressure-sensitive tape
2. Flat glass fiber tape
3. Self-clinching cable straps
4. Double lacing

2-36. Why do cables and wires require identification?

1. To assist the technician in troubleshooting a circuit
2. To assist the technician in making repairs
3. To permit the tracing of a circuit
4. Each of the above

2-37. Of the following publications, which should be used to determine the wire identification system for a specific piece of equipment?

1. The damage control manual
2. The technical manual for the equipment
3. The maintenance material management manual
4. The illustrated parts breakdown of the equipment

2-38. What is the purpose of the green conductor in a power tool or electric appliance cable?

1. To complete the circuit
2. To act as the "hot" lead
3. To prevent electrical shock to the operator
4. To prevent the motor of the unit from overloading

A.	Schematic diagram
B.	Single-line diagram
C.	Wiring diagram
D.	Block diagram
E.	Isometric diagram
F.	Pictorial diagram
G.	Terminal diagram

Figure 2A.—Types of diagrams.

IN ANSWERING QUESTIONS 2-39 THROUGH 2-45, REFER TO FIGURE 2A.

2-39. Which of the following diagrams is primarily used to identify the components of a system?

1. A
2. C
3. D
4. F

2-40. Which of the following diagrams is primarily used to locate the components of a system?

1. B
2. D
3. E
4. G

2-41. What two diagrams are used in conjunction with text materials to explain basic functions of a circuit?

1. B and D
2. C and G
3. E and F
4. G and A

2-42. Which of the following diagrams is primarily used to explain the overall operation of a system?

1. A
2. B
3. C
4. G

2-43. What diagram must be used in conjunction with a schematic to troubleshoot a system?

1. F
2. E
3. D
4. C

2-44. What diagram shows the most details of a system?

1. A
2. C
3. F
4. G

2-45. If you are required to wire a relay into a circuit, what diagram would be most useful?

1. G
2. F
3. C
4. A

THIS SPACE LEFT BLANK INTENTIONALLY.

THIS PAGE LEFT BLANK INTENTIONALLY.

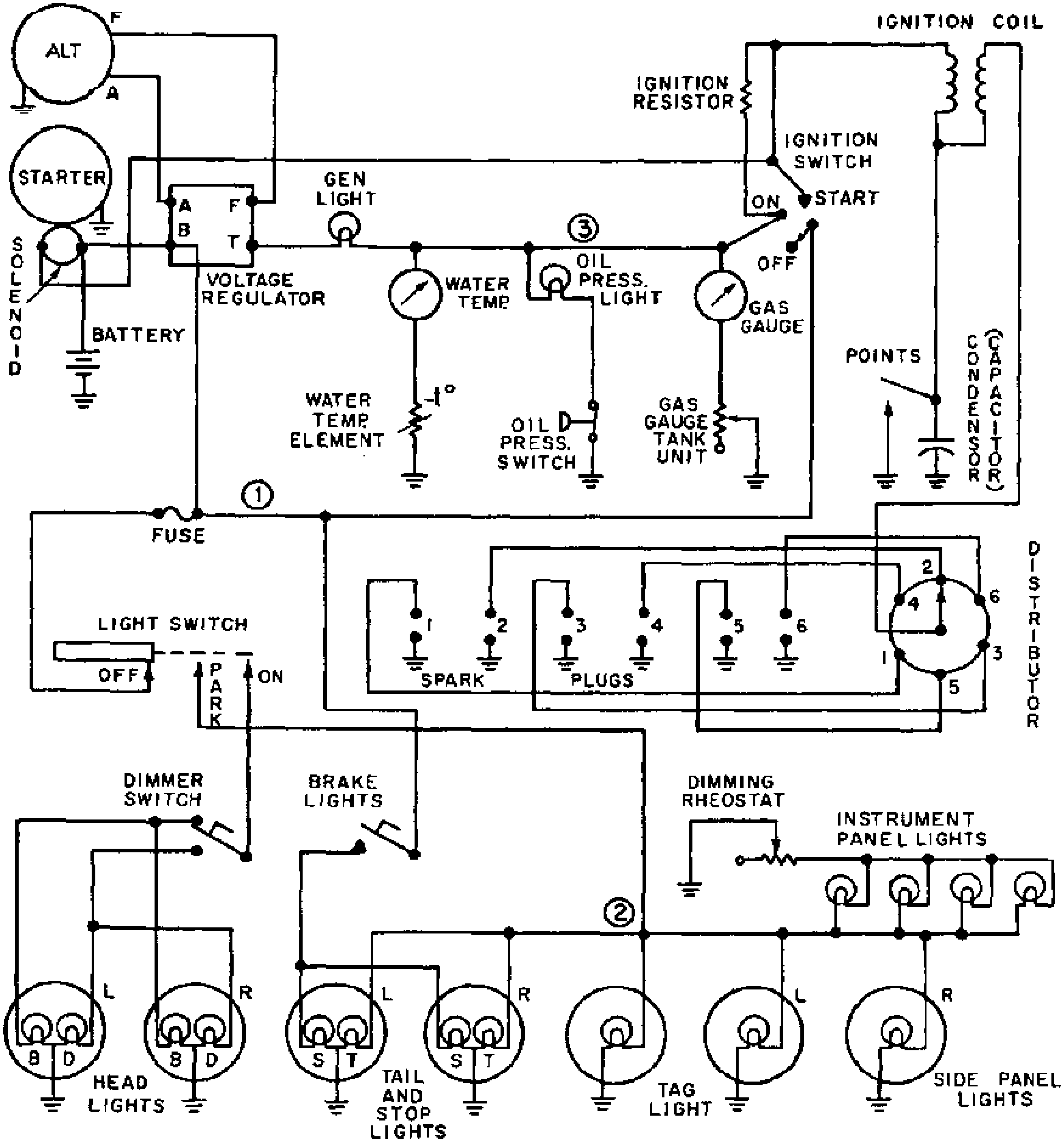


Figure 2B.—Schematic Diagram.

IN ANSWERING QUESTIONS 2-46 AND 2-47, REFER TO FIGURE 2B.

2-46. If the headlights operate normally in the bright position but do not light in the dim position, what would be the probable cause?

1. The dimmer switch is defective
2. The light switch is defective
3. A fuse is open
4. The ground to the headlights is open

2-47. Which of the following faults could cause the left tail light to be inoperative while the other lights operate normally?

1. The light switch is defective
2. The bulb is defective
3. A fuse is blown
4. There is no voltage to point 2

2-48. When you solder or hot-wire strip fluoroplastic insulated wire, which of the following safety precautions should be observed?

1. Wear a safety mask at all times
2. Wear protective gloves
3. Maintain good ventilation to carry off the fumes
4. Do not allow the resin to touch the insulation

2-49. If a circuit has power restored to it, what meter may be used to test the circuit?

1. An ohmmeter
2. A wattmeter
3. A megohmmeter
4. A voltmeter

2-50. If excess solder adheres to the tip of a soldering iron, how should you remove it?

1. Flow flux over the tip
2. Wipe it off on a clean cloth
3. Dip the tip in water
4. Shake it off

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 5—Introduction to Generators and Motors

NAVEDTRA 14177

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Sailor’s Creed

“I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country’s Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all.”

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: This course introduces the student to the subject of Generators and Motors. It provides a background for accomplishing daily work and/or in preparing for further study.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and the occupational standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards*, NAVPERS 18068, found on line at https://buperscd.technology.navy.mil /bup_updt/upd_CD/BUPERS/enlistedManOpen.htm.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
FTCS(SS) Steven F. Reith

Reviewed for accuracy by ETC Scott Collie
March 2003
Corrections were made to Assignments

NAVSUP Logistics Tracking Number
0504-LP-026-8300

TABLE OF CONTENTS

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

you may submit your answers as soon as you complete an assignment, and
you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 2 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

CHAPTER 1

DIRECT CURRENT GENERATORS

LEARNING OBJECTIVES

Upon completion of the chapter you will be able to:

1. State the principle by which generators convert mechanical energy to electrical energy.
2. State the rule to be applied when you determine the direction of induced emf in a coil.
3. State the purpose of slip rings.
4. State the reason why no emf is induced in a rotating coil as it passes through a neutral plane.
5. State what component causes a generator to produce direct current rather than alternating current.
6. Identify the point at which the brush contact should change from one commutator segment to the next.
7. State how field strength can be varied in a dc generator.
8. Describe the cause of sparking between brushes and commutator.
9. State what is meant by "armature reaction."
10. State the purpose of interpoles.
11. Explain the effect of motor reaction in a dc generator.
12. Explain the causes of armature losses.
13. List the types of armatures used in dc generators.
14. State the three classifications of dc generators.
15. State the term that applies to voltage variation from no-load to full-load conditions and how it is expressed as a percentage.
16. State the term that describes the use of two or more generators to supply a common load.
17. State the purpose of a dc generator that has been modified to function as an amplidyne.

INTRODUCTION

A generator is a machine that converts mechanical energy into electrical energy by using the principle of magnetic induction. This principle is explained as follows:

Whenever a conductor is moved within a magnetic field in such a way that the conductor cuts across magnetic lines of flux, voltage is generated in the conductor.

The AMOUNT of voltage generated depends on (1) the strength of the magnetic field, (2) the angle at which the conductor cuts the magnetic field, (3) the speed at which the conductor is moved, and (4) the length of the conductor within the magnetic field.

The POLARITY of the voltage depends on the direction of the magnetic lines of flux and the direction of movement of the conductor. To determine the direction of current in a given situation, the LEFT-HAND RULE FOR GENERATORS is used. This rule is explained in the following manner.

Extend the thumb, forefinger, and middle finger of your left hand at right angles to one another, as shown in figure 1-1. Point your thumb in the direction the conductor is being moved. Point your forefinger in the direction of magnetic flux (from north to south). Your middle finger will then point in the direction of current flow in an external circuit to which the voltage is applied.

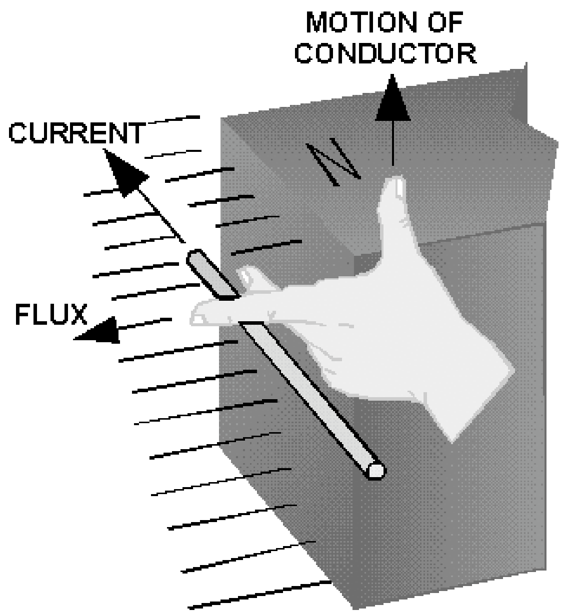


Figure 1-1.—Left-hand rule for generators.

THE ELEMENTARY GENERATOR

The simplest elementary generator that can be built is an ac generator. Basic generating principles are most easily explained through the use of the elementary ac generator. For this reason, the ac generator will be discussed first. The dc generator will be discussed later.

An elementary generator (fig. 1-2) consists of a wire loop placed so that it can be rotated in a stationary magnetic field. This will produce an induced emf in the loop. Sliding contacts (brushes) connect the loop to an external circuit load in order to pick up or use the induced emf.

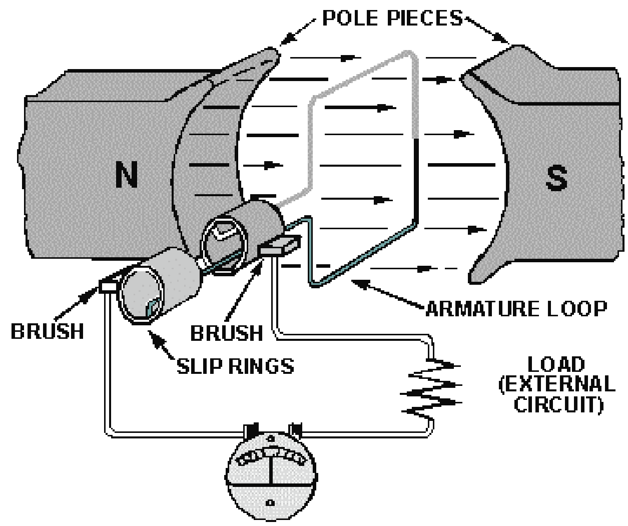


Figure 1-2.—The elementary generator.

The pole pieces (marked N and S) provide the magnetic field. The pole pieces are shaped and positioned as shown to concentrate the magnetic field as close as possible to the wire loop. The loop of wire that rotates through the field is called the ARMATURE. The ends of the armature loop are connected to rings called SLIP RINGS. They rotate with the armature. The brushes, usually made of carbon, with wires attached to them, ride against the rings. The generated voltage appears across these brushes.

The elementary generator produces a voltage in the following manner (fig. 1-3). The armature loop is rotated in a clockwise direction. The initial or starting point is shown at position A. (This will be considered the zero-degree position.) At 0º the armature loop is perpendicular to the magnetic field. The black and white conductors of the loop are moving parallel to the field. The instant the conductors are moving parallel to the magnetic field, they do not cut any lines of flux. Therefore, no emf is induced in the conductors, and the meter at position A indicates zero. This position is called the NEUTRAL PLANE. As the armature loop rotates from position A (0º) to position B (90º), the conductors cut through more and more lines of flux, at a continually increasing angle. At 90º they are cutting through a maximum number of lines of flux and at maximum angle. The result is that between 0º and 90º , the induced emf in the conductors builds up from zero to a maximum value. Observe that from 0º to 90º , the black conductor cuts DOWN through the field. At the same time the white conductor cuts UP through the field. The induced emfs in the conductors are series-adding. This means the resultant voltage across the brushes (the terminal voltage) is the sum of the two induced voltages. The meter at position B reads maximum value. As the armature loop continues rotating from 90º (position B) to 180º (position C), the conductors which were cutting through a maximum number of lines of flux at position B now cut through fewer lines. They are again moving parallel to the magnetic field at position C. They no longer cut through any lines of flux. As the armature rotates from 90º to 180º , the induced voltage will decrease to zero in the same manner that it increased during the rotation from 0º to 90º . The meter again reads zero. From 0º to 180º the conductors of the armature loop have been moving in the same direction through the magnetic field. Therefore, the polarity of the induced voltage has remained the same. This is shown by points A through C on the graph. As the loop rotates beyond 180º (position C), through 270º (position D), and back to the initial or starting point (position A), the direction of the cutting action of the conductors through the magnetic field reverses. Now the black conductor cuts UP through the field while the white conductor cuts DOWN through the field. As a result, the polarity of the induced voltage reverses. Following the sequence shown by graph points C, D, and back to A, the voltage will be in the direction opposite to that shown from points A, B, and C. The terminal voltage will be the same as it was from A to C except that the polarity is reversed (as shown by the meter deflection at position D). The voltage output waveform for the complete revolution of the loop is shown on the graph in figure 1-3.

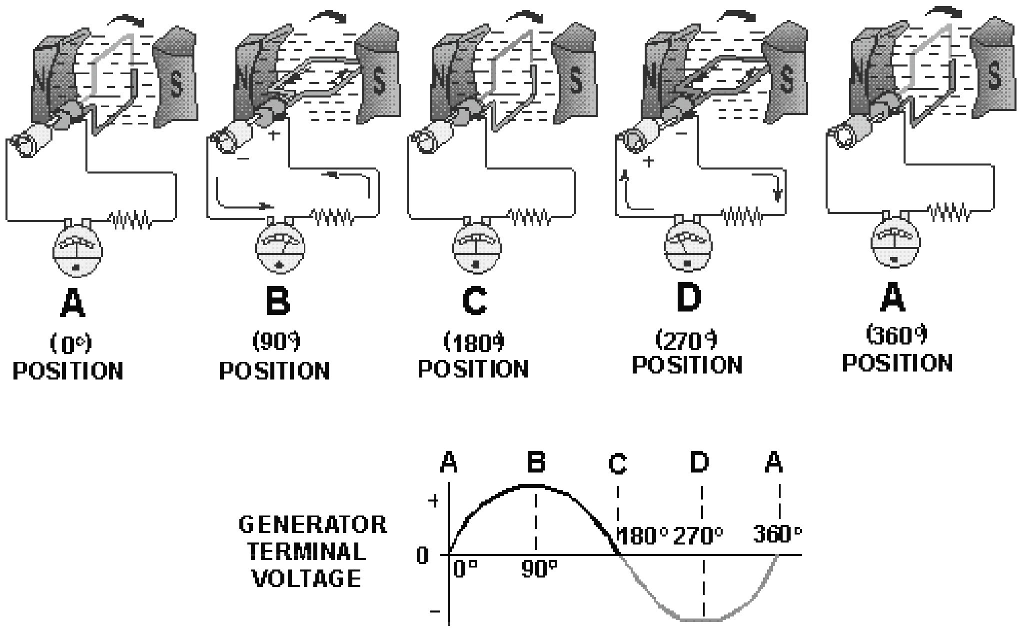


Figure 1-3.—Output voltage of an elementary generator during one revolution.

Q1. Generators convert mechanical motion to electrical energy using what principle?

Q2. What rule should you use to determine the direction of induced emf in a coil?

Q3. What is the purpose of the slip rings?

Q4. Why is no emf induced in a rotating coil when it passes through the neutral plane?

THE ELEMENTARY DC GENERATOR

A single-loop generator with each terminal connected to a segment of a two-segment metal ring is shown in figure 1-4. The two segments of the split metal ring are insulated from each other. This forms a simple COMMUTATOR. The commutator in a dc generator replaces the slip rings of the ac generator. This is the main difference in their construction. The commutator mechanically reverses the armature loop connections to the external circuit. This occurs at the same instant that the polarity of the voltage in the armature loop reverses. Through this process the commutator changes the generated ac voltage to a pulsating dc voltage as shown in the graph of figure 1-4. This action is known as commutation. Commutation is described in detail later in this chapter.

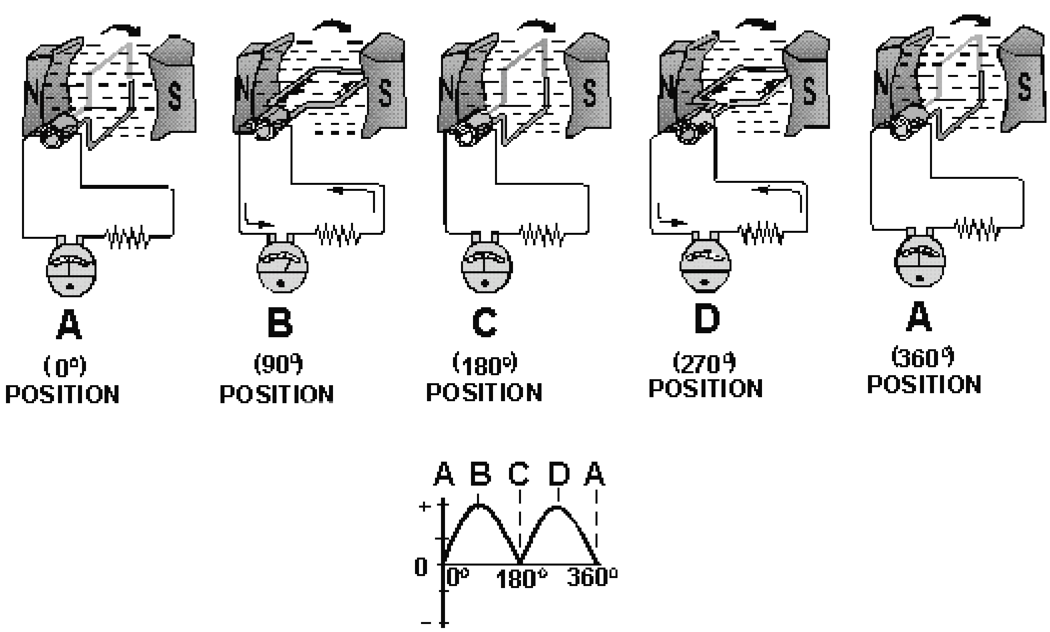


Figure 1-4.—Effects of commutation.

For the remainder of this discussion, refer to figure 1-4, parts A through D. This will help you in following the step-by-step description of the operation of a dc generator. When the armature loop rotates clockwise from position A to position B, a voltage is induced in the armature loop which causes a current in a direction that deflects the meter to the right. Current flows through loop, out of the negative brush, through the meter and the load, and back through the positive brush to the loop. Voltage reaches its maximum value at point B on the graph for reasons explained earlier. The generated voltage and the current fall to zero at position C. At this instant each brush makes contact with both segments of the commutator. As the armature loop rotates to position D, a voltage is again induced in the loop. In this case, however, the voltage is of opposite polarity.

The voltages induced in the two sides of the coil at position D are in the reverse direction to that of the voltages shown at position B. Note that the current is flowing from the black side to the white side in position B and from the white side to the black side in position D. However, because the segments of the commutator have rotated with the loop and are contacted by opposite brushes, the direction of current flow through the brushes and the meter remains the same as at position B. The voltage developed across the brushes is pulsating and unidirectional (in one direction only). It varies twice during each revolution between zero and maximum. This variation is called RIPPLE.

A pulsating voltage, such as that produced in the preceding description, is unsuitable for most applications. Therefore, in practical generators more armature loops (coils) and more commutator segments are used to produce an output voltage waveform with less ripple.

Q5. What component causes a generator to produce dc voltage rather than ac voltage at its output terminals?

Q6. At what point should brush contact change from one commutator segment to the next?

Q7. An elementary, single coil, dc generator will have an output voltage with how many pulsations per revolution?

EFFECTS OF ADDING ADDITIONAL COILS AND POLES

The effects of additional coils may be illustrated by the addition of a second coil to the armature. The commutator must now be divided into four parts since there are four coil ends (see fig. 1-5). The coil is rotated in a clockwise direction from the position shown. The voltage induced in the white coil, DECREASES FOR THE NEXT 90º of rotation (from maximum to zero). The voltage induced in the black coil INCREASES from zero to maximum at the same time. Since there are four segments in the commutator, a new segment passes each brush every 90º instead of every 180º . This allows the brush to switch from the white coil to the black coil at the instant the voltages in the two coils are equal. The brush remains in contact with the black coil as its induced voltage increases to maximum, level B in the graph. It then decreases to level A, 90º later. At this point, the brush will contact the white coil again.

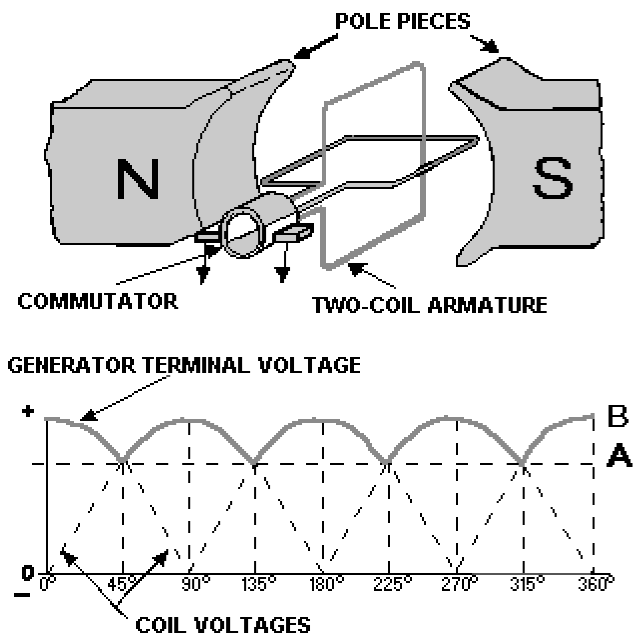


Figure 1-5.—Effects of additional coils.

The graph in figure 1-5 shows the ripple effect of the voltage when two armature coils are used. Since there are now four commutator segments in the commutator and only two brushes, the voltage cannot fall any lower than at point A. Therefore, the ripple is limited to the rise and fall between points A and B on the graph. By adding more armature coils, the ripple effect can be further reduced. Decreasing ripple in this way increases the effective voltage of the output.

NOTE: Effective voltage is the equivalent level of dc voltage, which will cause the same average current through a given resistance. By using additional armature coils, the voltage across the brushes is not allowed to fall to as low a level between peaks. Compare the graphs in figure 1-4 and 1-5. Notice that the ripple has been reduced. Practical generators use many armature coils. They also use more than one pair of magnetic poles. The additional magnetic poles have the same effect on ripple as did the additional armature coils. In addition, the increased number of poles provides a stronger magnetic field (greater number of flux lines). This, in turn, allows an increase in output voltage because the coils cut more lines of flux per revolution.

Q8. How many commutator segments are required in a two-coil generator?

ELECTROMAGNETIC POLES

Nearly all practical generators use electromagnetic poles instead of the permanent magnets used in our elementary generator. The electromagnetic field poles consist of coils of insulated copper wire wound on soft iron cores, as shown in figure 1-6. The main advantages of using electromagnetic poles are (1) increased field strength and (2) a means of controlling the strength of the fields. By varying the input voltage, the field strength is varied. By varying the field strength, the output voltage of the generator can be controlled.

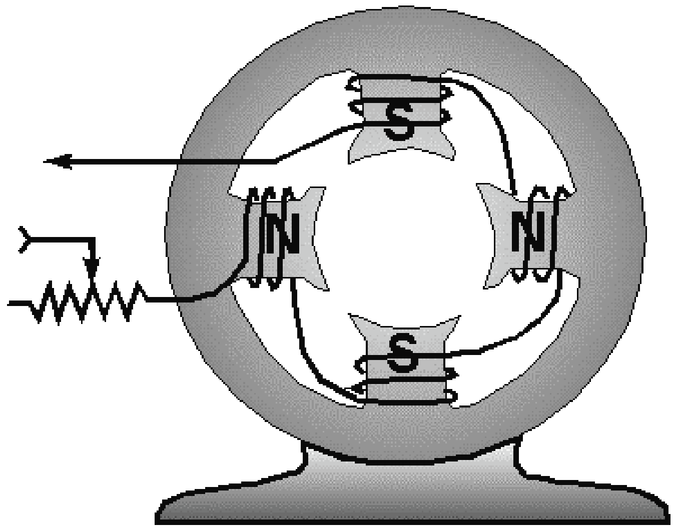


Figure 1-6.—Four-pole generator (without armature).

Q9. How can field strength be varied in a practical dc generator?

COMMUTATION

Commutation is the process by which a dc voltage output is taken from an armature that has an ac voltage induced in it. You should remember from our discussion of the elementary dc generator that the commutator mechanically reverses the armature loop connections to the external circuit. This occurs at the same instant that the voltage polarity in the armature loop reverses. A dc voltage is applied to the load because the output connections are reversed as each commutator segment passes under a brush. The segments are insulated from each other.

In figure 1-7, commutation occurs simultaneously in the two coils that are briefly short-circuited by the brushes. Coil B is short-circuited by the negative brush. Coil Y, the opposite coil, is short-circuited by the positive brush. The brushes are positioned on the commutator so that each coil is short-circuited as it moves through its own electrical neutral plane. As you have seen previously, there is no voltage generated in the coil at that time. Therefore, no sparking can occur between the commutator and the brush. Sparking between the brushes and the commutator is an indication of improper commutation. Improper brush placement is the main cause of improper commutation.

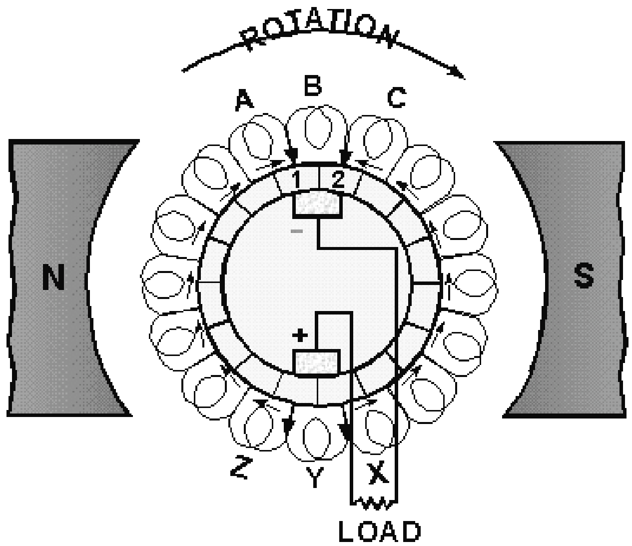


Figure 1-7.—Commutation of a dc generator.

Q10. What causes sparking between the brushes and the commutator?

ARMATURE REACTION

From previous study, you know that all current-carrying conductors produce magnetic fields. The magnetic field produced by current in the armature of a dc generator affects the flux pattern and distorts the main field. This distortion causes a shift in the neutral plane, which affects commutation. This change in the neutral plane and the reaction of the magnetic field is called ARMATURE REACTION.

You know that for proper commutation, the coil short-circuited by the brushes must be in the neutral plane. Consider the operation of a simple two-pole dc generator, shown in figure 1-8. View A of the figure shows the field poles and the main magnetic field. The armature is shown in a simplified view in views B and C with the cross section of its coil represented as little circles. The symbols within the circles represent arrows. The dot represents the point of the arrow coming toward you, and the cross represents the tail, or feathered end, going away from you. When the armature rotates clockwise, the sides of the coil to the left will have current flowing toward you, as indicated by the dot. The side of the coil to the right will have current flowing away from you, as indicated by the cross. The field generated around each side of the coil is shown in view B of figure 1-8. This field increases in strength for each wire in the armature coil, and sets up a magnetic field almost perpendicular to the main field.

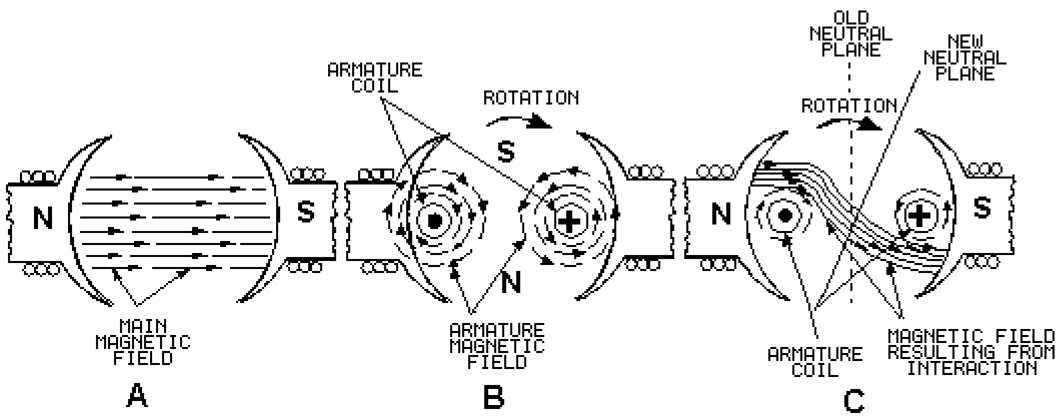


Figure 1-8.—Armature reaction.

Now you have two fields — the main field, view A, and the field around the armature coil, view B. View C of figure 1-8 shows how the armature field distorts the main field and how the neutral plane is shifted in the direction of rotation. If the brushes remain in the old neutral plane, they will be short-circuiting coils that have voltage induced in them. Consequently, there will be arcing between the brushes and commutator.

To prevent arcing, the brushes must be shifted to the new neutral plane.

Q11. What is armature reaction?

COMPENSATING WINDINGS AND INTERPOLES

Shifting the brushes to the advanced position (the new neutral plane) does not completely solve the problems of armature reaction. The effect of armature reaction varies with the load current. Therefore, each time the load current varies, the neutral plane shifts. This means the brush position must be changed each time the load current varies.

In small generators, the effects of armature reaction are reduced by actually mechanically shifting the position of the brushes. The practice of shifting the brush position for each current variation is not practiced except in small generators. In larger generators, other means are taken to eliminate armature reaction. COMPENSATING WINDINGS or INTERPOLES are used for this purpose (fig. 1-9). The compensating windings consist of a series of coils embedded in slots in the pole faces. These coils are connected in series with the armature. The series-connected compensating windings produce a magnetic field, which varies directly with armature current. Because the compensating windings are wound to produce a field that opposes the magnetic field of the armature, they tend to cancel the effects of the armature magnetic field. The neutral plane will remain stationary and in its original position for all values of armature current. Because of this, once the brushes have been set correctly, they do not have to be moved again.

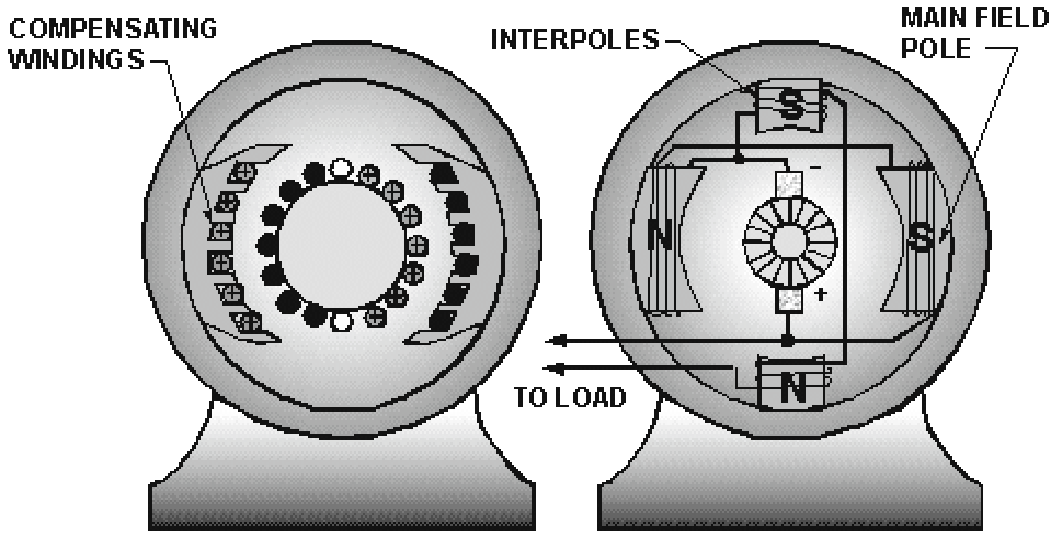


Figure 1-9.—Compensating windings and interpoles.

Another way to reduce the effects of armature reaction is to place small auxiliary poles called "interpoles" between the main field poles. The interpoles have a few turns of large wire and are connected in series with the armature. Interpoles are wound and placed so that each interpole has the same magnetic polarity as the main pole ahead of it, in the direction of rotation. The field generated by the interpoles produces the same effect as the compensating winding. This field, in effect, cancels the armature reaction for all values of load current by causing a shift in the neutral plane opposite to the shift caused by armature reaction. The amount of shift caused by the interpoles will equal the shift caused by armature reaction since both shifts are a result of armature current.

Q12. What is the purpose of interpoles?

MOTOR REACTION IN A GENERATOR

When a generator delivers current to a load, the armature current creates a magnetic force that opposes the rotation of the armature. This is called MOTOR REACTION. A single armature conductor is represented in figure 1-10, view A. When the conductor is stationary, no voltage is generated and no current flows. Therefore, no force acts on the conductor. When the conductor is moved downward (fig. 1-10, view B) and the circuit is completed through an external load, current flows through the conductor in the direction indicated. This sets up lines of flux around the conductor in a clockwise direction.

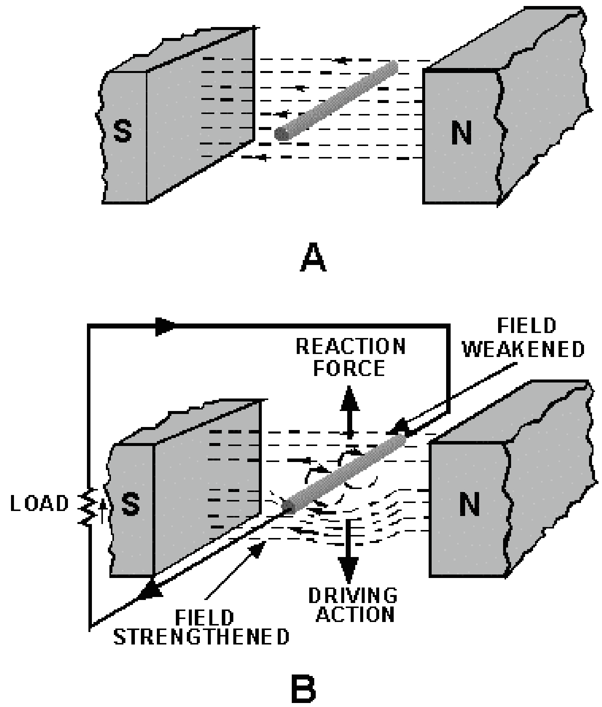


Figure 1-10.—Motor reaction in a generator.

The interaction between the conductor field and the main field of the generator weakens the field above the conductor and strengthens the field below the conductor. The main field consists of lines that now act like stretched rubber bands. Thus, an upward reaction force is produced that acts in opposition to the downward driving force applied to the armature conductor. If the current in the conductor increases, the reaction force increases. Therefore, more force must be applied to the conductor to keep it moving.

With no armature current, there is no magnetic (motor) reaction. Therefore, the force required to turn the armature is low. As the armature current increases, the reaction of each armature conductor against rotation increases. The actual force in a generator is multiplied by the number of conductors in the armature. The driving force required to maintain the generator armature speed must be increased to overcome the motor reaction. The force applied to turn the armature must overcome the motor reaction force in all dc generators. The device that provides the turning force applied to the armature is called the PRIME MOVER. The prime mover may be an electric motor, a gasoline engine, a steam turbine, or any other mechanical device that provides turning force.

Q13. What is the effect of motor reaction in a dc generator?

ARMATURE LOSSES

In dc generators, as in most electrical devices, certain forces act to decrease the efficiency. These forces, as they affect the armature, are considered as losses and may be defined as follows:

1. I^2R, or copper loss in the winding
2. Eddy current loss in the core
3. Hysteresis loss (a sort of magnetic friction)

Copper Losses

The power lost in the form of heat in the armature winding of a generator is known as COPPER LOSS. Heat is generated any time current flows in a conductor. Copper loss is an I^2R loss, which increases as current increases. The amount of heat generated is also proportional to the resistance of the conductor. The resistance of the conductor varies directly with its length and inversely with its cross-sectional area. Copper loss is minimized in armature windings by using large diameter wire.

Q14. What causes copper losses?

Eddy Current Losses

The core of a generator armature is made from soft iron, which is a conducting material with desirable magnetic characteristics. Any conductor will have currents induced in it when it is rotated in a magnetic field. These currents that are induced in the generator armature core are called EDDY CURRENTS. The power dissipated in the form of heat, as a result of the eddy currents, is considered a loss.

Eddy currents, just like any other electrical currents, are affected by the resistance of the material in which the currents flow. The resistance of any material is inversely proportional to its cross-sectional area. Figure 1-11, view A, shows the eddy currents induced in an armature core that is a solid piece of soft iron. Figure 1-11, view B, shows a soft iron core of the same size, but made up of several small pieces insulated from each other. This process is called lamination. The currents in each piece of the laminated core are considerably less than in the solid core because the resistance of the pieces is much higher. (Resistance is inversely proportional to cross-sectional area.) The currents in the individual pieces of the laminated core are so small that the sum of the individual currents is much less than the total of eddy currents in the solid iron core.

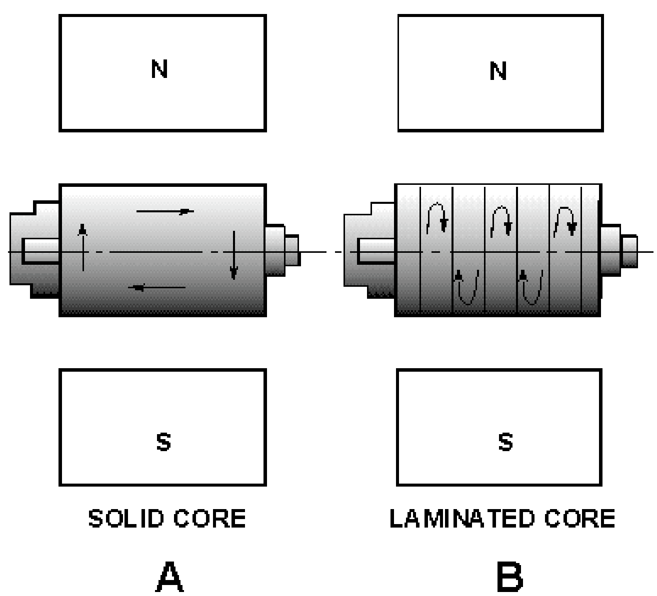


Figure 1-11.—Eddy currents in dc generator armature cores.

As you can see, eddy current losses are kept low when the core material is made up of many thin sheets of metal. Laminations in a small generator armature may be as thin as 1/64 inch. The laminations are insulated from each other by a thin coat of lacquer or, in some instances, simply by the oxidation of the surfaces. Oxidation is caused by contact with the air while the laminations are being annealed. The insulation value need not be high because the voltages induced are very small.

Most generators use armatures with laminated cores to reduce eddy current losses.

Q15. How can eddy current be reduced?

Hysteresis Losses

Hysteresis loss is a heat loss caused by the magnetic properties of the armature. When an armature core is in a magnetic field, the magnetic particles of the core tend to line up with the magnetic field. When the armature core is rotating, its magnetic field keeps changing direction. The continuous movement of the magnetic particles, as they try to align themselves with the magnetic field, produces molecular friction. This, in turn, produces heat. This heat is transmitted to the armature windings. The heat causes armature resistances to increase.

To compensate for hysteresis losses, heat-treated silicon steel laminations are used in most dc generator armatures. After the steel has been formed to the proper shape, the laminations are heated and allowed to cool. This annealing process reduces the hysteresis loss to a low value.

THE PRACTICAL DC GENERATOR

The actual construction and operation of a practical dc generator differs somewhat from our elementary generators. The differences are in the construction of the armature, the manner in which the armature is wound, and the method of developing the main field.

A generator that has only one or two armature loops has high ripple voltage. This results in too little current to be of any practical use. To increase the amount of current output, a number of loops of wire are used. These additional loops do away with most of the ripple. The loops of wire, called windings, are evenly spaced around the armature so that the distance between each winding is the same.

The commutator in a practical generator is also different. It has several segments instead of two or four, as in our elementary generators. The number of segments must equal the number of armature coils.

GRAMME-RING ARMATURE

The diagram of a GRAMME-RING armature is shown in figure 1-12, view A. Each coil is connected to two commutator segments as shown. One end of coil 1goes to segment A, and the other end of coil 1 goes to segment B. One end of coil 2 goes to segment C, and the other end of coil 2 goes to segment B. The rest of the coils are connected in a like manner, in series, around the armature. To complete the series arrangement, coil 8 connects to segment A. Therefore, each coil is in series with every other coil.

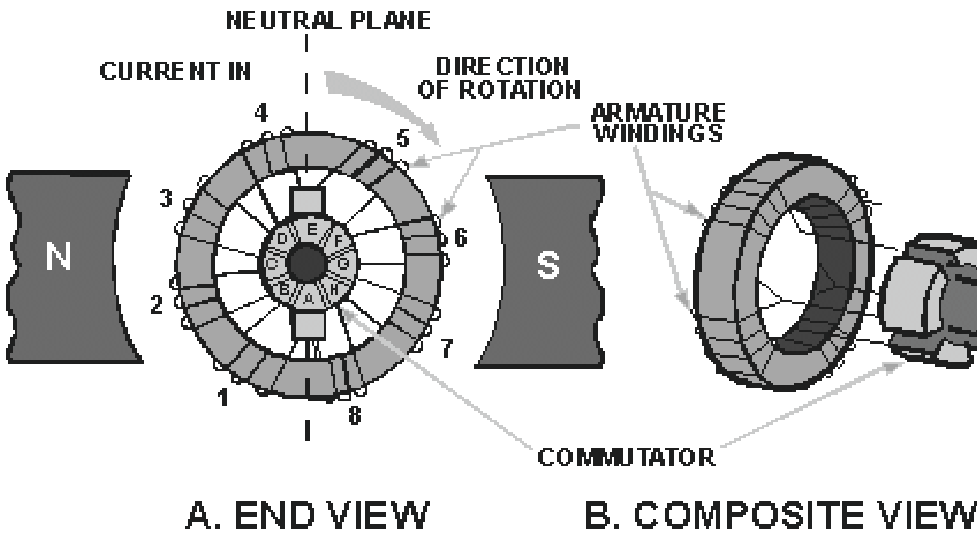


Figure 1-12.—Gramme-ring armature.

Figure 1-12, view B shows a composite view of a Gramme-ring armature. It illustrates more graphically the physical relationship of the coils and commutator locations.

The windings of a Gramme-ring armature are placed on an iron ring. A disadvantage of this arrangement is that the windings located on the inner side of the iron ring cut few lines of flux. Therefore, they have little, if any, voltage induced in them. For this reason, the Gramme-ring armature is not widely used.

DRUM-TYPE ARMATURE

A drum-type armature is shown in figure 1-13. The armature windings are placed in slots cut in a drum-shaped iron core. Each winding completely surrounds the core so that the entire length of the conductor cuts the main magnetic field. Therefore, the total voltage induced in the armature is greater than in the Gramme-ring. You can see that the drum-type armature is much more efficient than the Gramme-ring. This accounts for the almost universal use of the drum-type armature in modem dc generators.

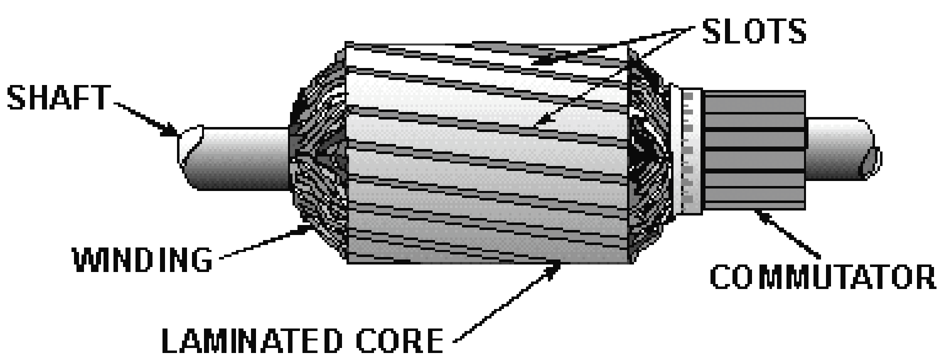


Figure 1-13.—Drum-type armature.

Drum-type armatures are wound with either of two types of windings — the LAP WINDING or the WAVE WINDING.

The lap winding is illustrated in figure 1-14, view A This type of winding is used in dc generators designed for high-current applications. The windings are connected to provide several parallel paths for current in the armature. For this reason, lap-wound armatures used in dc generators require several pairs of poles and brushes.

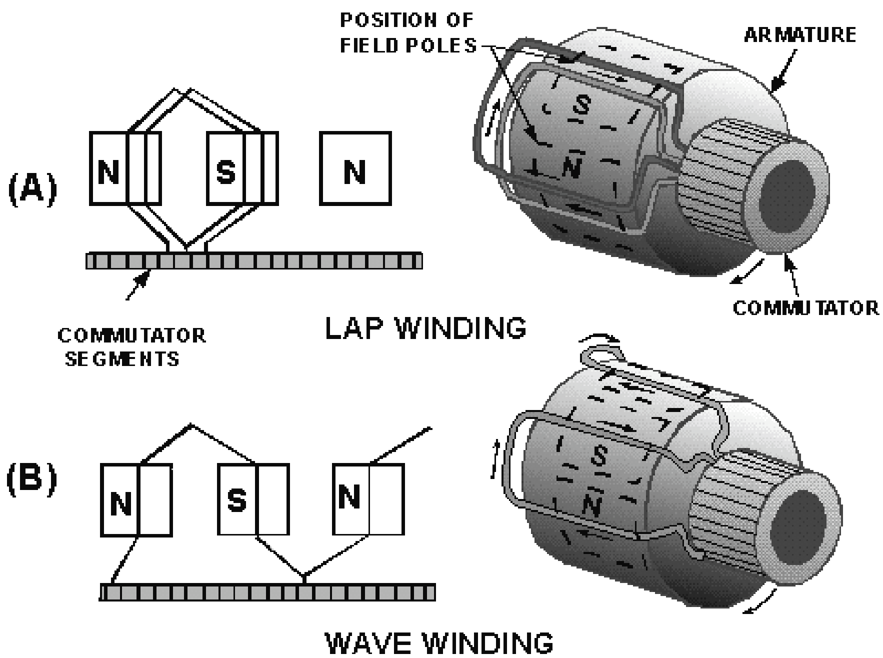


Figure 1-14.—Types of windings used on drum-type armatures.

Figure 1-14, view B, shows a wave winding on a drum-type armature. This type of winding is used in dc generators employed in high-voltage applications. Notice that the two ends of each coil are connected to commutator segments separated by the distance between poles. This configuration allows the series addition of the voltages in all the windings between brushes. This type of winding only requires one pair of brushes. In practice, a practical generator may have several pairs to improve commutation.

Q16. Why are drum-type armatures preferred over the Gramme-ring armature in modern dc generators?

Q17. Lap windings are used in generators designed for what type of application?

FIELD EXCITATION

When a dc voltage is applied to the field windings of a dc generator, current flows through the windings and sets up a steady magnetic field. This is called FIELD EXCITATION.

This excitation voltage can be produced by the generator itself or it can be supplied by an outside source, such as a battery. A generator that supplies its own field excitation is called a SELF-EXCITED GENERATOR. Self-excitation is possible only if the field pole pieces have retained a slight amount of permanent magnetism, called RESIDUAL MAGNETISM. When the generator starts rotating, the weak residual magnetism causes a small voltage to be generated in the armature. This small voltage applied to the field coils causes a small field current. Although small, this field current strengthens the magnetic field and allows the armature to generate a higher voltage. The higher voltage increases the field strength, and so on. This process continues until the output voltage reaches the rated output of the generator.

CLASSIFICATION OF GENERATORS

Self-excited generators are classed according to the type of field connection they use. There are three general types of field connections — SERIES-WOUND, SHUNT-WOUND (parallel), and COMPOUND-WOUND. Compound-wound generators are further classified as cumulative-compound and differential-compound. These last two classifications are not discussed in this chapter.

Series-Wound Generator

In the series-wound generator, shown in figure 1-15, the field windings are connected in series with the armature. Current that flows in the armature flows through the external circuit and through the field windings. The external circuit connected to the generator is called the load circuit.

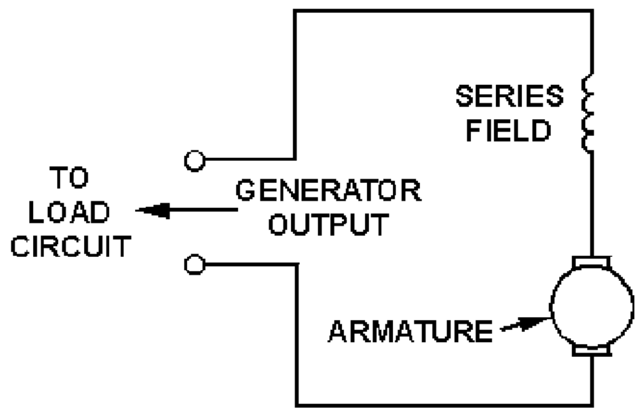


Figure 1-15.—Series-wound generator.

A series-wound generator uses very low resistance field coils, which consist of a few turns of large diameter wire.

The voltage output increases as the load circuit starts drawing more current. Under low-load current conditions, the current that flows in the load and through the generator is small. Since small current means that a small magnetic field is set up by the field poles, only a small voltage is induced in the armature. If the resistance of the load decreases, the load current increases. Under this condition, more current flows through the field. This increases the magnetic field and increases the output voltage. A series-wound dc generator has the characteristic that the output voltage varies with load current. This is undesirable in most applications. For this reason, this type of generator is rarely used in everyday practice.

The series-wound generator has provided an easy method to introduce you to the subject of self-excited generators.

Shunt-Wound Generators

In a shunt-wound generator, like the one shown in figure 1-16, the field coils consist of many turns of small wire. They are connected in parallel with the load. In other words, they are connected across the output voltage of the armature.

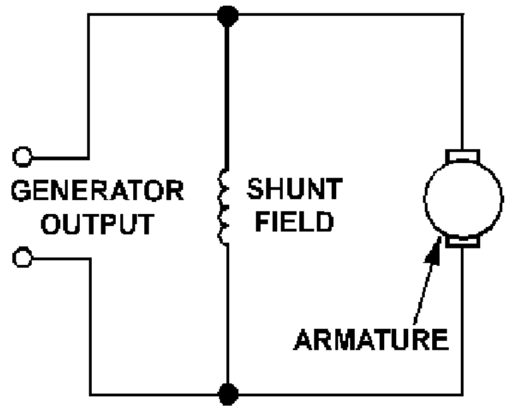


Figure 1-16.—Shunt-wound generator.

Current in the field windings of a shunt-wound generator is independent of the load current (currents in parallel branches are independent of each other). Since field current, and therefore field strength, is not affected by load current, the output voltage remains more nearly constant than does the output voltage of the series-wound generator.

In actual use, the output voltage in a dc shunt-wound generator varies inversely as load current varies. The output voltage decreases as load current increases because the voltage drop across the armature resistance increases (E = IR).

In a series-wound generator, output voltage varies directly with load current. In the shunt-wound generator, output voltage varies inversely with load current. A combination of the two types can overcome the disadvantages of both. This combination of windings is called the compound-wound dc generator.

Compound-Wound Generators

Compound-wound generators have a series-field winding in addition to a shunt-field winding, as shown in figure 1-17. The shunt and series windings are wound on the same pole pieces.

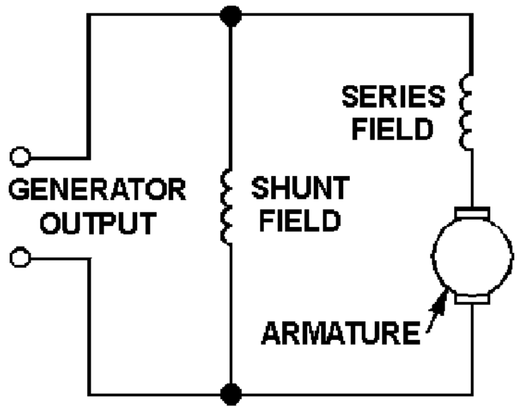


Figure 1-17.—Compound-wound generator.

In the compound-wound generator when load current increases, the armature voltage decreases just as in the shunt-wound generator. This causes the voltage applied to the shunt-field winding to decrease, which results in a decrease in the magnetic field. This same increase in load current, since it flows through the series winding, causes an increase in the magnetic field produced by that winding.

By proportioning the two fields so that the decrease in the shunt field is just compensated by the increase in the series field, the output voltage remains constant. This is shown in figure 1-18, which shows the voltage characteristics of the series-, shunt-, and compound-wound generators. As you can see, by proportioning the effects of the two fields (series and shunt), a compound-wound generator provides a constant output voltage under varying load conditions. Actual curves are seldom, if ever, as perfect as shown.

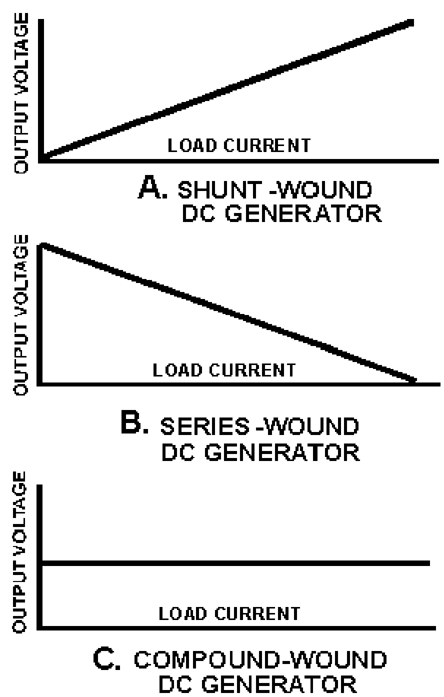


Figure 1-18.—Voltage output characteristics of the series-, shunt-, and compound-wound dc generators.

Q18. What are the three classifications of dc generators?

Q19. What is the main disadvantage of series generators?

GENERATOR CONSTRUCTION

Figure 1-19, views A through E, shows the component parts of dc generators. Figure 1-20 shows the entire generator with the component parts installed. The cutaway drawing helps you to see the physical relationship of the components to each other.

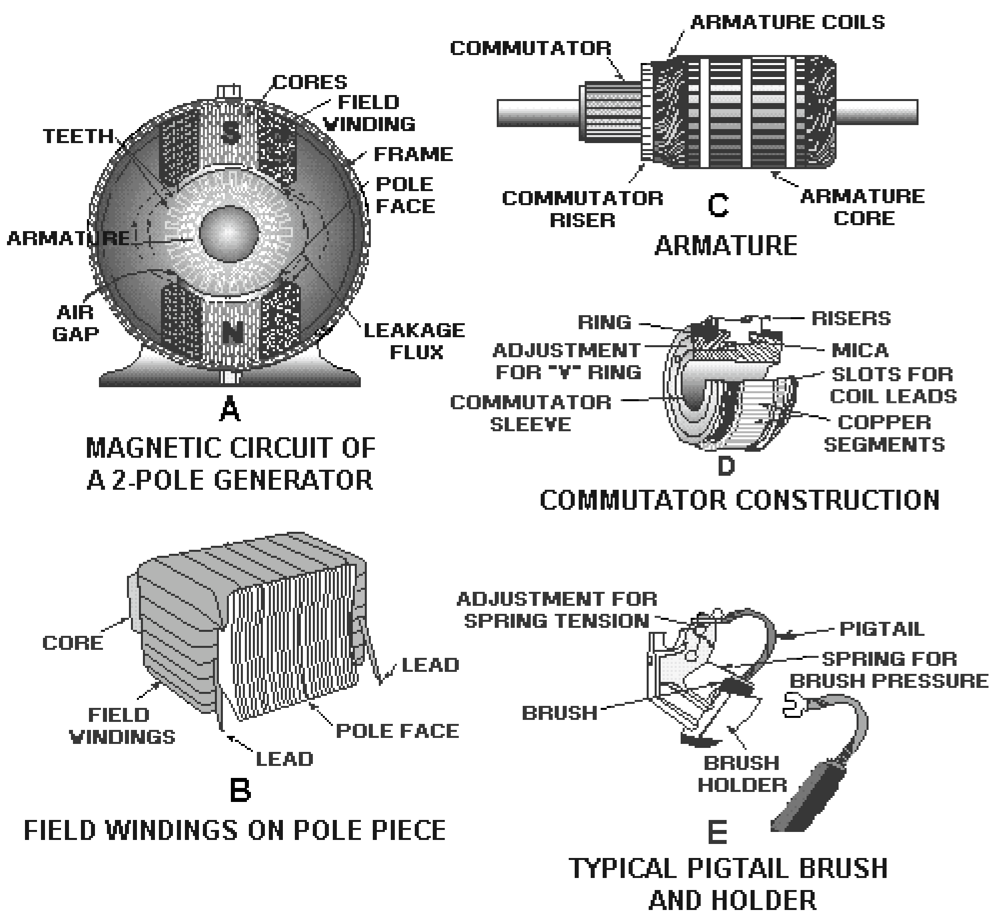


Figure 1-19.—Components of a dc generator.

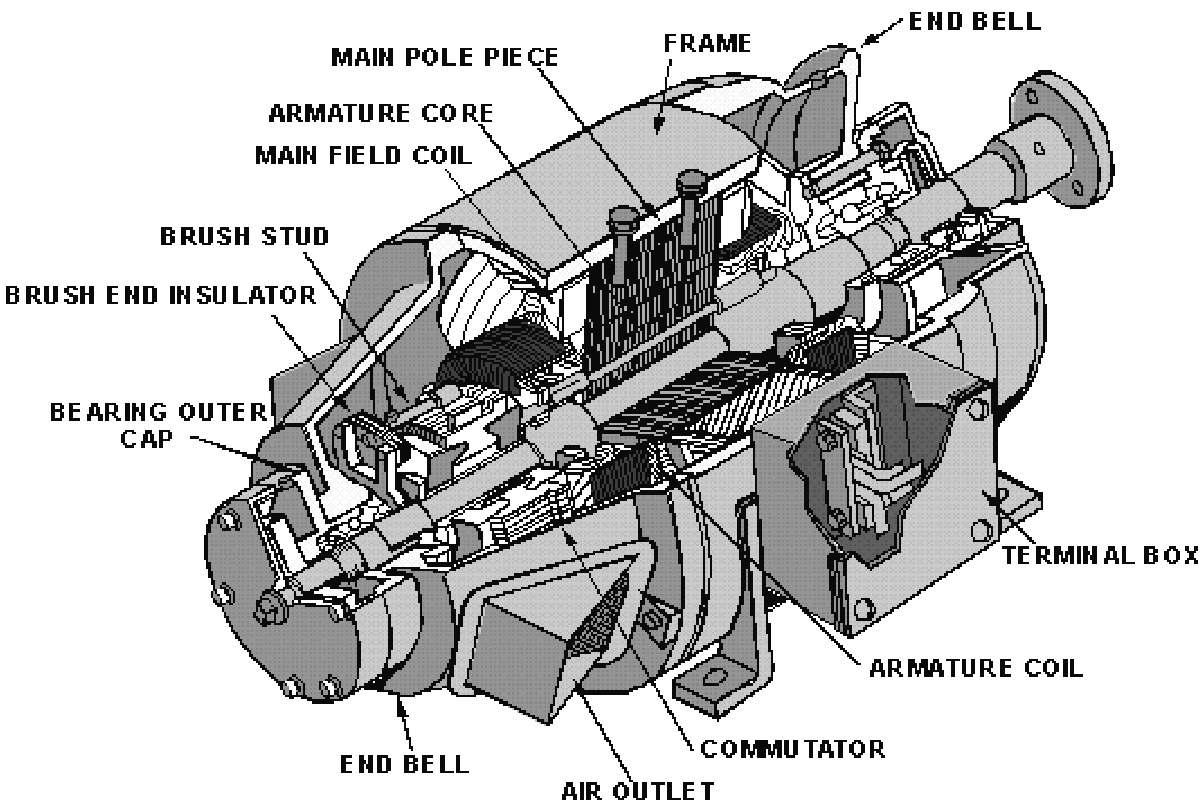


Figure 1-20.—Construction of a dc generator (cutaway drawing).

VOLTAGE REGULATION

The regulation of a generator refers to the VOLTAGE CHANGE that takes place when the load changes. It is usually expressed as the change in voltage from a no-load condition to a full-load condition, and is expressed as a percentage of full-load. It is expressed in the following formula:

$$\text{Percent of regulation} = \frac{(E_{nL} - E_{fL})}{E_{fL}} \times 100$$

where E_{nL} is the no-load terminal voltage and E_{fL} is the full-load terminal voltage of the generator. For example, to calculate the percent of regulation of a generator with a no-load voltage of 462 volts and a full-load voltage of 440 volts

Given:

- No-load voltage 462 V
- Full-load voltage 440 V

Solution:

$$\text{Percent of regulation} = \frac{(E_{nL} - E_{fL})}{E_{fL}} \times 100$$

$$\text{Percent of regulation} = \frac{(462V - 440V)}{440V} \times 100$$

$$\text{Percent of regulation} = \frac{22V}{440V} \times 100$$

$$\text{Percent of regulation} = .05 \times 100$$

$$\text{Regulation} = 5\%$$

NOTE: The lower the percent of regulation, the better the generator. In the above example, the 5% regulation represented a 22-volt change from no load to full load. A 1% change would represent a change of 4.4 volts, which, of course, would be better.

Q20. What term applies to the voltage variation from no-load to full-load conditions and is expressed as a percentage?

VOLTAGE CONTROL

Voltage control is either (1) manual or (2) automatic. In most cases the process involves changing the resistance of the field circuit. By changing the field circuit resistance, the field current is controlled. Controlling the field current permits control of the output voltage. The major difference between the various voltage control systems is merely the method by which the field circuit resistance and the current are controlled.

VOLTAGE REGULATION should not be confused with VOLTAGE CONTROL. As described previously, voltage regulation is an internal action occurring within the generator whenever the load changes. Voltage control is an imposed action, usually through an external adjustment, for the purpose of increasing or decreasing terminal voltage.

Manual Voltage Control

The hand-operated field rheostat, shown in figure 1-21, is a typical example of manual voltage control. The field rheostat is connected in series with the shunt field circuit. This provides the simplest method of controlling the terminal voltage of a dc generator.

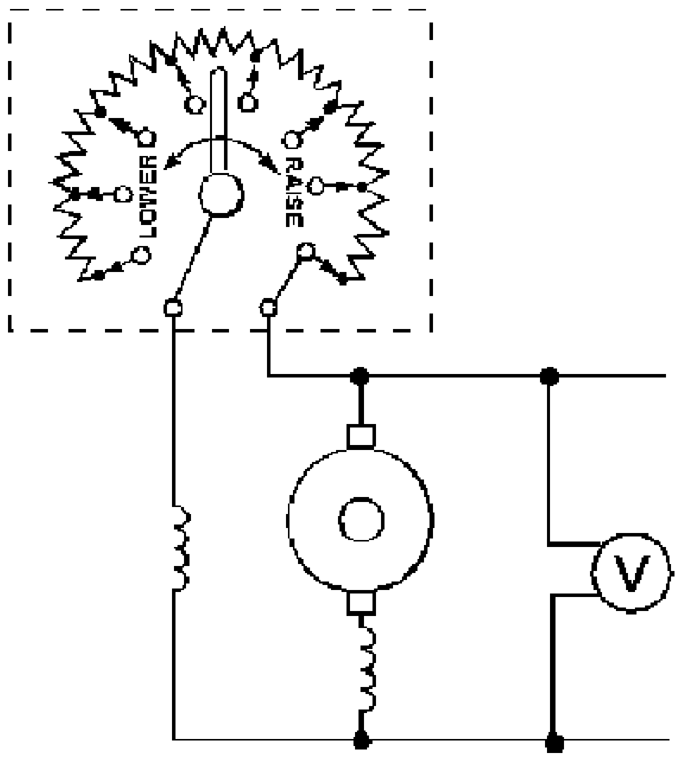


Figure 1-21.—Hand-operated field rheostat.

This type of field rheostat contains tapped resistors with leads to a multiterminal switch. The arm of the switch may be rotated to make contact with the various resistor taps. This varies the amount of resistance in the field circuit. Rotating the arm in the direction of the LOWER arrow (counterclockwise) increases the resistance and lowers the output voltage. Rotating the arm in the direction of the RAISE arrow (clockwise) decreases the resistance and increases the output voltage.

Most field rheostats for generators use resistors of alloy wire. They have a high specific resistance and a low temperature coefficient. These alloys include copper, nickel, manganese, and chromium. They are marked under trade names such as Nichrome, Advance, Manganin, and so forth. Some very large generators use cast-iron grids in place of rheostats, and motor-operated switching mechanisms to provide voltage control.

Automatic Voltage Control

Automatic voltage control may be used where load current variations exceed the built-in ability of the generator to regulate itself. An automatic voltage control device "senses" changes in output voltage and causes a change in field resistance to keep output voltage constant.

The actual circuitry involved in automatic voltage control will not be covered in this chapter. Whichever control method is used, the range over which voltage can be changed is a design characteristic of the generator. The voltage can be controlled only within the design limits.

PARALLEL OPERATION OF GENERATORS

When two or more generators are supplying a common load, they are said to be operating in parallel. The purpose of connecting generators in parallel is simply to provide more current than a single generator is capable of providing. The generators may be physically located quite a distance apart. However, they are connected to the common load through the power distribution system.

There are several reasons for operating generators in parallel. The number of generators used may be selected in accordance with the load demand. By operating each generator as nearly as possible to its rated capacity, maximum efficiency is achieved. A disabled or faulty generator may be taken off-line and replaced without interrupting normal operations.

Q21. What term applies to the use of two or more generators to supply a common load?

AMPLIDYNES

Amplidynes are special-purpose dc generators. They supply large dc currents, precisely controlled, to the large dc motors used to drive heavy physical loads, such as gun turrets and missile launchers.

The amplidyne is really a motor and a generator. It consists of a constant-speed ac motor (the prime mover) mechanically coupled to a dc generator, which is wired to function as a high-gain amplifier (an amplifier is a device in which a small input voltage can control a large current source). For instance, in a normal dc generator, a small dc voltage applied to the field windings is able to control the output of the generator. In a typical generator, a change in voltage from 0-volt dc to 3-volts dc applied to the field winding may cause the generator output to vary from 0-volt dc to 300-volts dc. If the 3 volts applied to the field winding is considered an input, and the 300 volts taken from the brushes is an output, there is a gain of 100. Gain is expressed as the ratio of output to input:

$$\text{Gain} = \frac{\text{output}}{\text{input}}$$

In this case 300 V ÷ 3 V = 100. This means that the 3 volts output is 100 times larger than the input.

The following paragraphs explain how gain is achieved in a typical dc generator and how the modifications making the generator an amplidyne increase the gain to as high as 10,000.

The schematic diagram in figure 1-22 shows a separately excited dc generator. Because of the 10-volt controlling voltage, 10 amperes of current will flow through the 1-ohm field winding. This draws 100 watts of input power (P = IE).

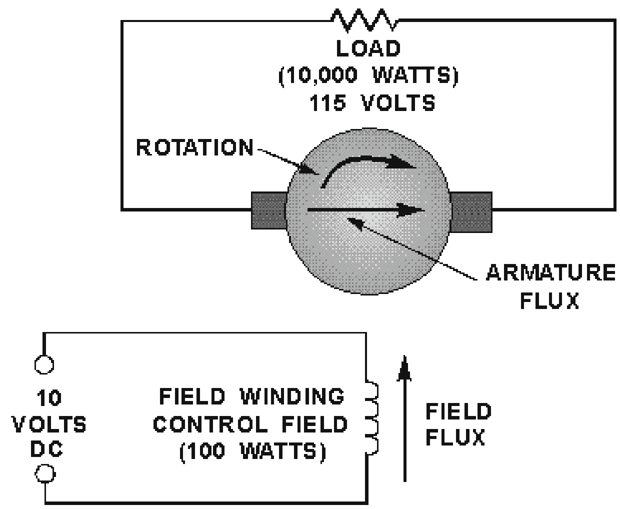


Figure 1-22.—Ordinary dc generator.

Assume that the characteristics of this generator enable it to produce approximately 87 amperes of armature current at 115 volts at the output terminals. This represents an output power of approximately 10,000 watts (P = IE). You can see that the power gain of this generator is 100. In effect, 100 watts controls 10,000 watts.

An amplidyne is a special type of dc generator. The following changes, for explanation purposes, will convert the typical dc generator above into an amplidyne.

The first step is to short the brushes together, as shown in figure 1-23. This removes nearly all of the resistance in the armature circuit.

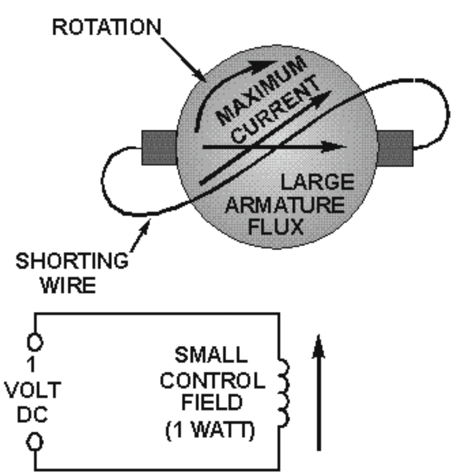


Figure 1-23.—Brushes shorted in a dc generator.

Because of the very low resistance in the armature circuit, a much lower control-field flux produces full-load armature current (full-load current in the armature is still about 87 amperes). The smaller control field now requires a control voltage of only 1 volt and an input power of 1 watt (1 volt across 1 ohm causes 1 ampere of current, which produces 1 watt of input power).

The next step is to add another set of brushes. These now become the output brushes of the amplidyne. They are placed against the commutator in a position perpendicular to the original brushes, as shown in figure 1-24. The previously shorted brushes are now called the "quadrature" brushes. This is because they are in quadrature (perpendicular) to the output brushes. The output brushes are in line with the armature flux. Therefore, they pick off the voltage induced in the armature windings at this point. The voltage at the output will be the same as in the original generator, 115 volts in our example.

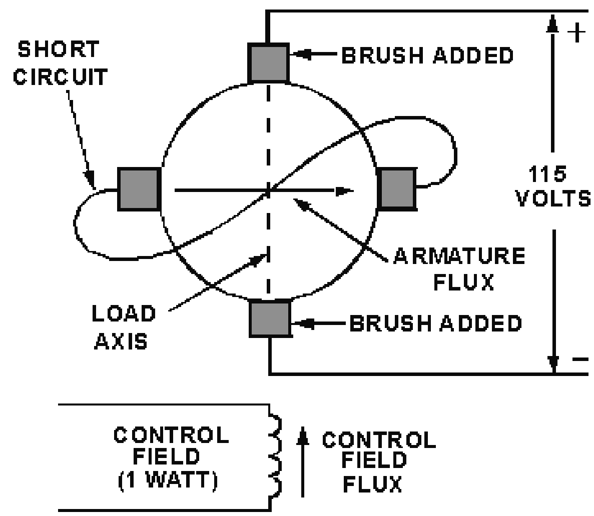


Figure 1-24.—Amplidyne load brushes.

As you have seen, the original generator produced a 10,000-watt output with a 100-watt input. The amplidyne produces the same 10,000-watt output with only a 1-watt input. This represents a gain of 10,000. The gain of the original generator has been greatly increased.

As previously stated, an amplidyne is used to provide large dc currents. The primary use of an amplidyne is in the positioning of heavy loads through the use of synchro/servo systems. Synchro/servo systems will be studied in a later module.

Assume that a very large turning force is required to rotate a heavy object, such as an antenna, to a very precise position. A low-power, relatively weak voltage representing the amount of antenna rotation required can be used to control the field winding of an amplidyne. Because of the amplidyne's ability to amplify, its output can be used to drive a powerful motor, which turns the heavy object (antenna). When the source of the input voltage senses the correct movement of the object, it drops the voltage to zero. The field is no longer strong enough to allow an output voltage to be developed, so the motor ceases to drive the object (antenna).

The above is an oversimplification and is not meant to describe a functioning system. The intent is to show a typical sequence of events between the demand for movement and the movement itself. It is meant to strengthen the idea that with the amplidyne, something large and heavy can be controlled very precisely by something very small, almost insignificant.

Q22. What is the purpose of a dc generator that has been modified to function as an amplidyne?

Q23. What is the formula used to determine the gain of an amplifying device?

Q24. What are the two inputs to an amplidyne?

SAFETY PRECAUTIONS

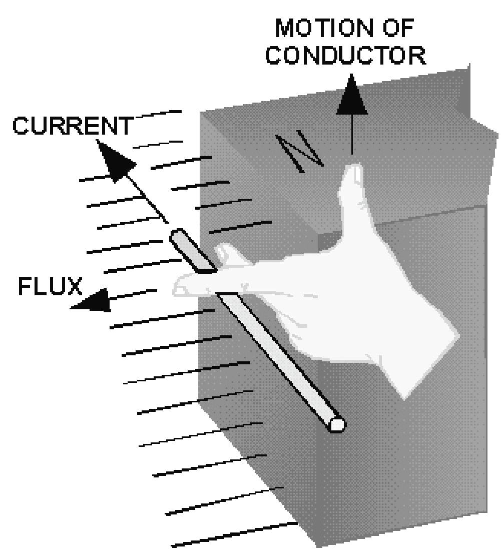
You must always observe safety precautions when working around electrical equipment to avoid injury to personnel and damage to equipment. Electrical equipment frequently has accessories that require separate sources of power. Lighting fixtures, heaters, externally powered temperature detectors, and alarm systems are examples of accessories whose terminals must be deenergized. When working on dc generators, you must check to ensure that all such circuits have been de-energized and tagged before you attempt any maintenance or repair work. You must also use the greatest care when working on or near the output terminals of dc generators.

SUMMARY

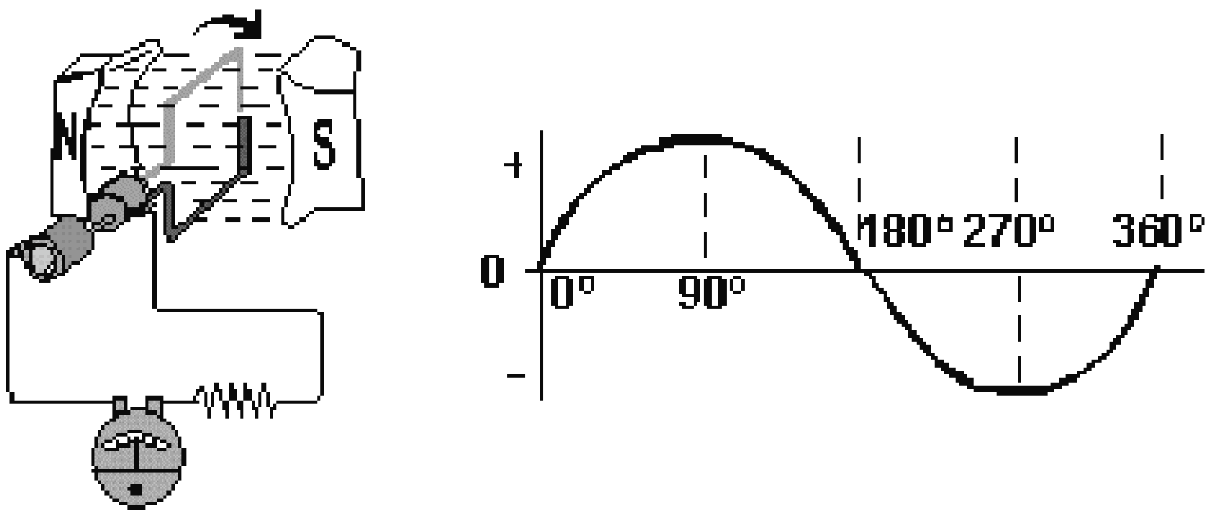
This chapter introduced you to the basic principles concerning direct current generators. The different types of dc generators and their characteristics were covered. The following information provides a summary of the major subjects of the chapter for your review.

MAGNETIC INDUCTION takes place when a conductor is moved in a magnetic field in such a way that it cuts flux lines, and a voltage (emf) is induced in the conductor.

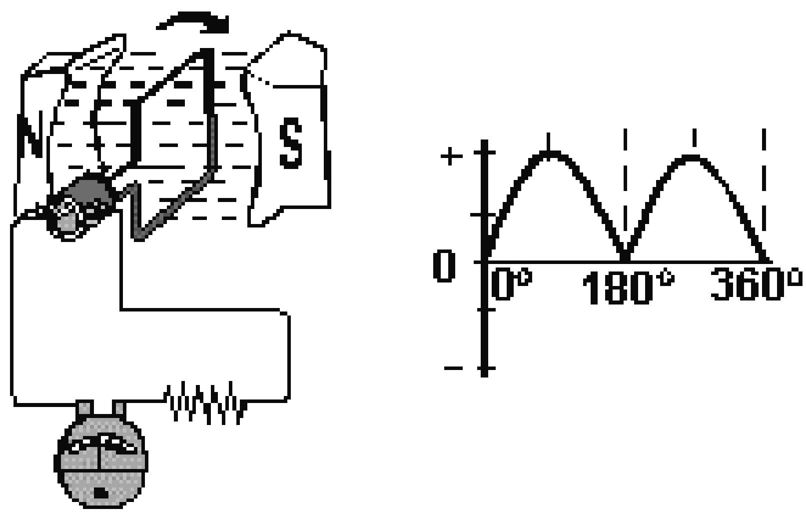
THE LEFT-HAND RULE FOR GENERATORS states that when the thumb, forefinger, and middle finger of the left hand are extended at right angles to each other so that the thumb indicates the direction of movement of the conductor in the magnetic field, and the forefinger points in the direction of the flux lines (north to south), the middle finger shows the direction of induced emf in the conductor.



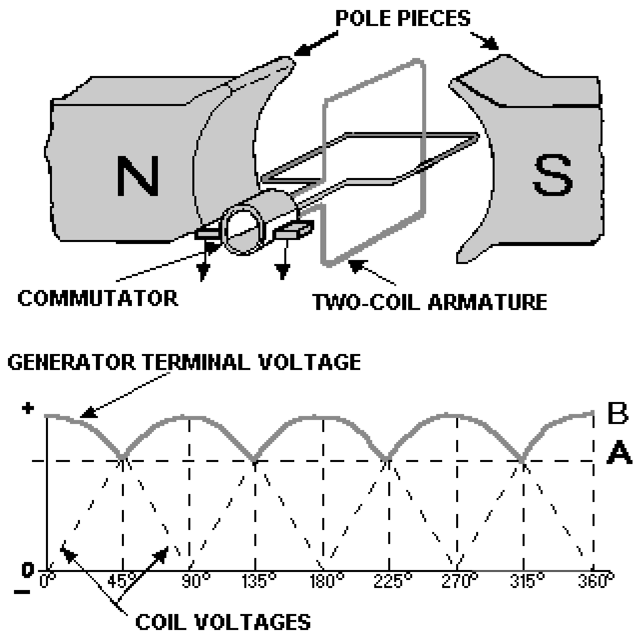
AN **ELEMENTARY GENERATOR** consists of a single coil rotated in a magnetic field. It produces an ac voltage.



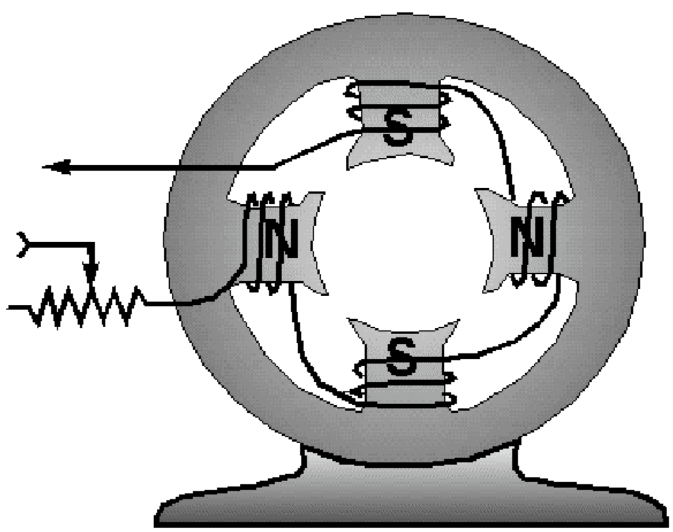
A **BASIC DC GENERATOR** results when you replace the slip rings of an elementary generator with a two-piece commutator, changing the output voltage to pulsating dc.



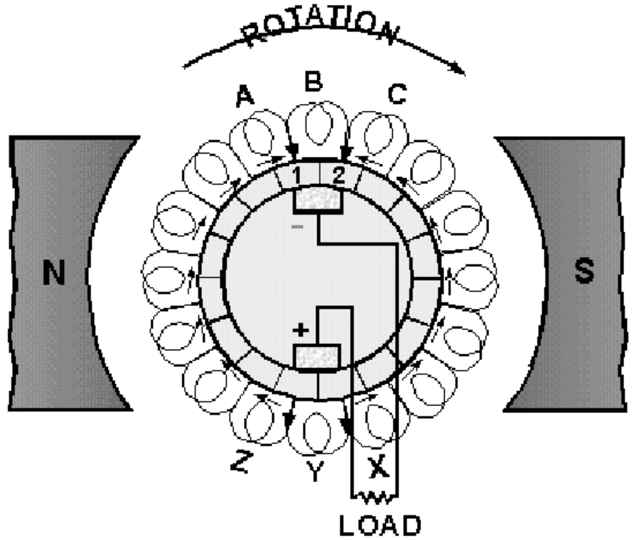
A **MULTIPLE COIL ARMATURE** (adding coils to the armature) decreases the ripple voltage in the output of a dc generator, and increases the output voltage.



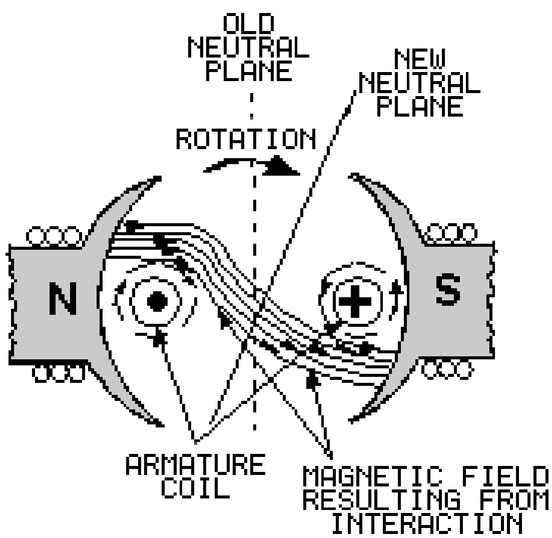
A **MULTIPOLE GENERATOR** is the result of adding more field poles to a dc generator. They have much the same effect as adding coils to the armature. In practical generators, the poles are electromagnets.



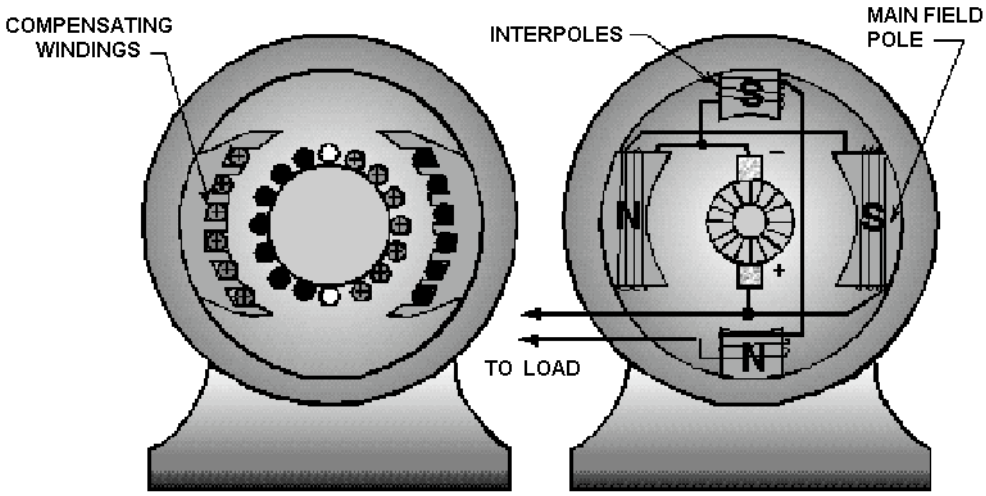
COMMUTATION is the process used to get direct current from a generator. The coil connections to the load must be reversed as the coil passes through the neutral plane. The brushes must be positioned so that commutation is accomplished without brush sparking.



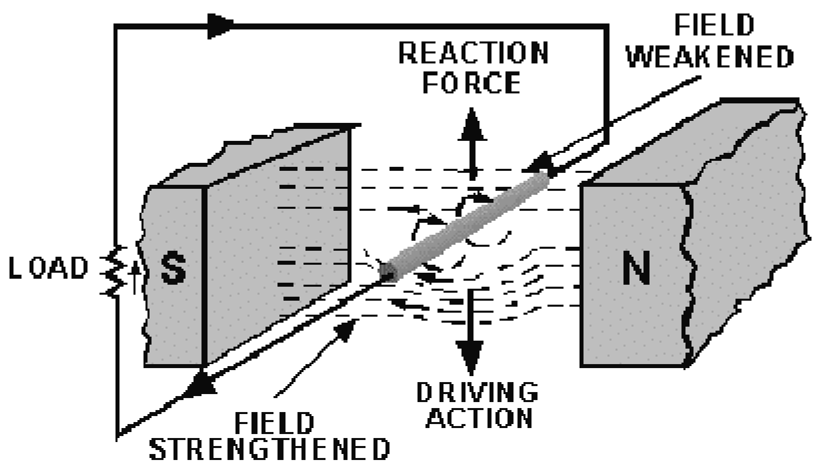
ARMATURE REACTION takes place when armature current causes the armature to become an electromagnet. The armature field disturbs the field from the pole pieces. This results in a shift of the neutral plane in the direction of rotation.



COMPENSATING WINDINGS AND INTERPOLES are used to counteract the effects of armature reaction. They are supplied by armature current and shift the neutral plane back to its original position.



MOTOR REACTION is caused by the magnetic field that is set up in the armature. It tends to oppose the rotation of the armature, due to the attraction and repulsion forces between the armature field and the main field.

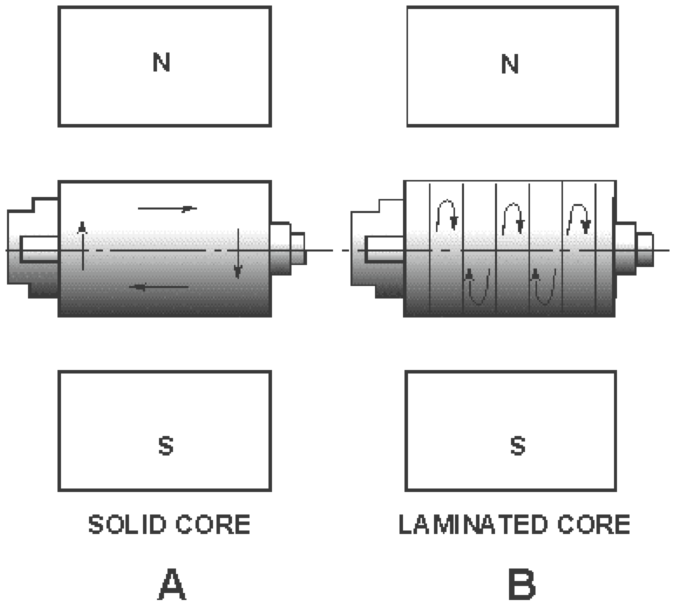


ARMATURE LOSSES in dc generator armatures affect the outputs. These losses are as follows:

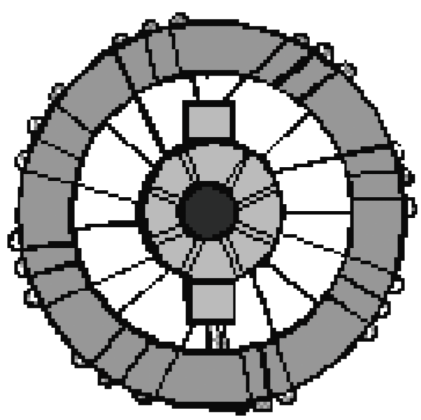
1. Copper losses are simply 1^2R (heat) losses caused by current flowing through the resistance of the armature windings.

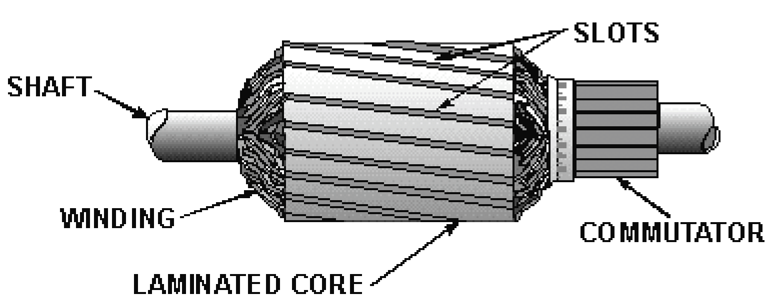
2. Eddy currents are induced in core material and cause heat.

3. Hysteresis losses occur due to the rapidly changing magnetic fields in the armature, resulting in heat.



ARMATURE TYPES used in dc generators are the Gramme-ring (seldom used) and the drum-type, used in most applications.



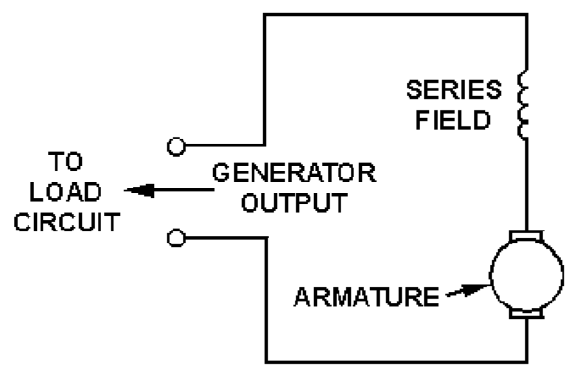


FIELD EXCITATION is the voltage applied to the main field windings. The current in the field coils determines the strength and the direction of the magnetic field.

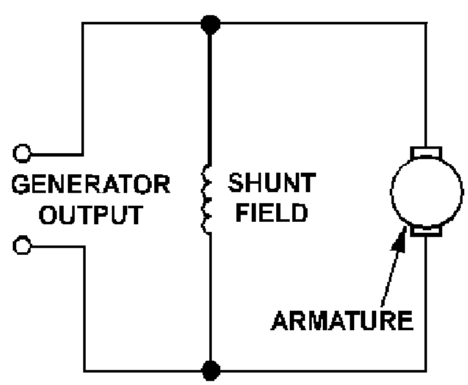
SEPARATELY EXCITED GENERATORS receive current for field coils from an outside source such, as a battery or another dc generator.

SELF-EXCITED GENERATORS use their own output voltage to energize field coils.

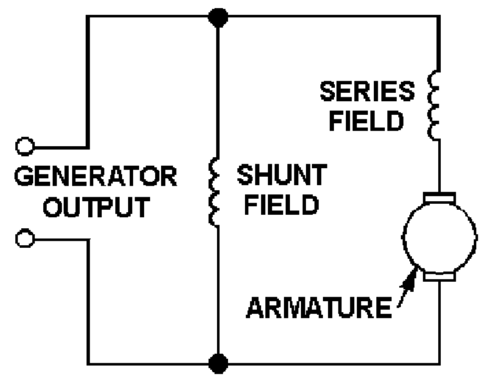
SERIES-WOUND DC GENERATORS have field windings and armature windings connected in series. Outputs vary directly with load currents. Series-wound generators have few practical applications.



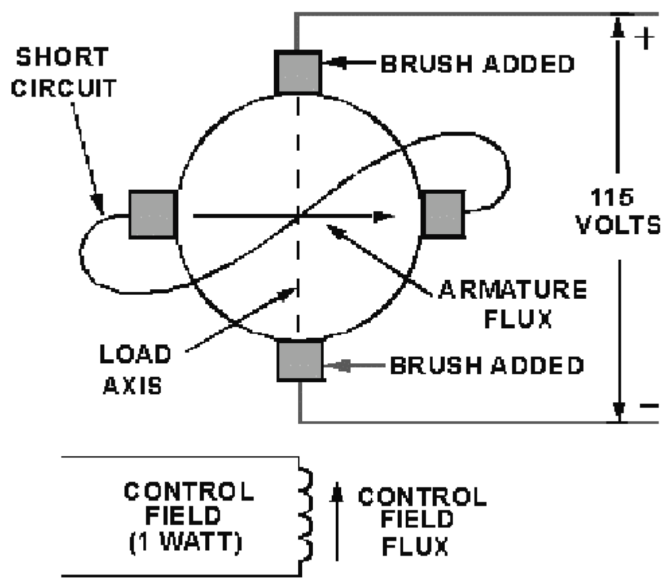
SHUNT-WOUND DC GENERATORS have field windings and armature windings connected in parallel (shunt). The output varies inversely with load current.



COMPOUND-WOUND DC GENERATORS have both series field windings and shunt field windings. These generators combine the characteristics of series and shunt generators. The output voltage remains relatively constant for all values of load current within the design of the generator. Compound generators are used in many applications because of the relatively constant voltage.



AMPLIDYNES are dc generators that are designed to act as high-gain amplifiers. By short-circuiting the brushes in a normal dc generator and adding another set of brushes perpendicular to the original ones, an amplidyne is formed. Its power output may be up to 10,000 times larger than the power input to its control windings.



ANSWERS TO QUESTIONS Q1. THROUGH Q24.

A1. Magnetic induction.

A2. The left-hand rule for generators.

A3. To conduct the currents induced in the armature to an external load.

A4. No flux lines are cut.

A5. A commutator

A6. The point at which the voltage is zero across the two segments.

A7. Two.

A8. Four

A9. By varying the input voltage to the field coils.

A10. Improper commutation.

A11. Distortion of the main field due to the effects of armature current.

A12. To counter act armature reaction.

A13. A force which causes opposition to applied turning force.

A14. Resistance in the armature coils, which increases with temperature.

A15. By laminating the core material.

A16. Drum-type armatures are more efficient, because flux lines are cut by both sides of each coil.

A17. Higher load currents are possible.

A18. Series-wound, shunt-wound, and compound-wound.

A19. Output voltage varies as the load varies.

A20. Voltage regulation.

A21. Parallel operation.

A22. It can serve as a power amplifier.

A23. Gain = output ÷ input.

A24. The mechanical force applied to turn the amplidyne, and the electrical input signal.

CHAPTER 2

DIRECT CURRENT MOTORS

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. State the factors that determine the direction of rotation in a dc motor.
2. State the right-hand rule for motors.
3. Describe the main differences and similarities between a dc generator and a dc motor.
4. Describe the cause and effect of counter emf in a dc motor.
5. Explain the term "load" as it pertains to an electric motor.
6. List the advantages and disadvantages of the different types of dc motors.
7. Compare the types of armatures and uses for each.
8. Discuss the means of controlling the speed and direction of a dc motor.
9. Describe the effect of armature reaction in a dc motor.
10. Explain the need for a starting resistor in a dc motor.

INTRODUCTION

The dc motor is a mechanical workhorse, that can be used in many different ways. Many large pieces of equipment depend on a dc motor for their power to move. The speed and direction of rotation of a dc motor are easily controlled. This makes it especially useful for operating equipment, such as winches, cranes, and missile launchers, which must move in different directions and at varying speeds.

PRINCIPLES OF OPERATION

The operation of a dc motor is based on the following principle:

A current-carrying conductor placed in a magnetic field, perpendicular to the lines of flux, tends to move in a direction perpendicular to the magnetic lines of flux.

There is a definite relationship between the direction of the magnetic field, the direction of current in the conductor, and the direction in which the conductor tends to move. This relationship is best explained by using the RIGHT-HAND RULE FOR MOTORS (fig. 2-1).

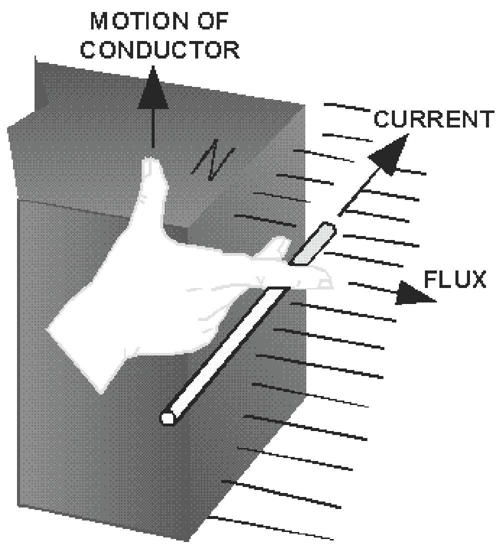


Figure 2-1.—Right-hand rule for motors.

To find the direction of motion of a conductor, extend the thumb, forefinger, and middle finger of your right hand so they are at right angles to each other. If the forefinger is pointed in the direction of magnetic flux (north to south) and the middle finger is pointed in the direction of current flow in the conductor, the thumb will point in the direction the conductor will move.

Stated very simply, a dc motor rotates as a result of two magnetic fields interacting with each other. The armature of a dc motor acts like an electromagnet when current flows through its coils. Since the armature is located within the magnetic field of the field poles, these two magnetic fields interact. Like magnetic poles repel each other, and unlike magnetic poles attract each other. As in the dc generator, the dc motor has field poles that are stationary and an armature that turns on bearings in the space between the field poles. The armature of a dc motor has windings on it just like the armature of a dc generator. These windings are also connected to commutator segments. A dc motor consists of the same components as a dc generator. In fact, most dc generators can be made to act as motors, and vice versa.

Look at the simple dc motor shown in figure 2-2. It has two field poles, one a north pole and one a south pole. The magnetic lines of force extend across the opening between the poles from north to south.

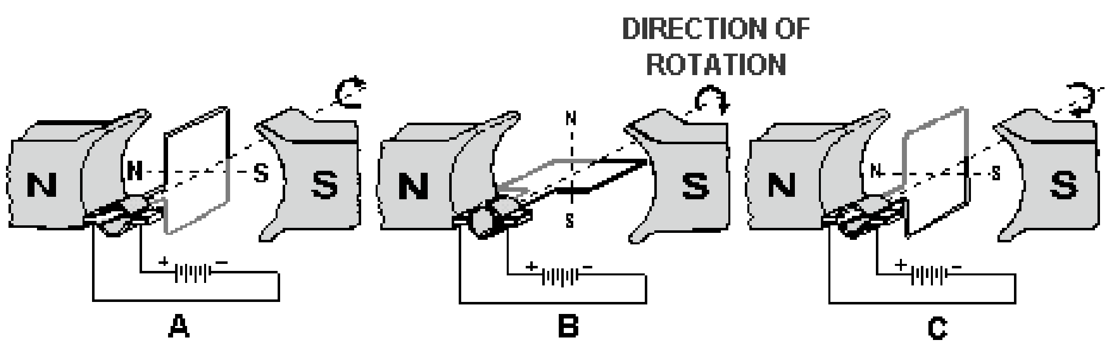


Figure 2-2.—Dc motor armature rotation.

The armature in this simple dc motor is a single loop of wire, just as in the simple armature you studied at the beginning of the chapter on dc generators. The loop of wire in the dc motor, however, has current flowing through it from an external source. This current causes a magnetic field to be produced. This field is indicated by the dotted line through the loops. The loop (armature) field is both attracted and repelled by the field from the field poles. Since the current through the loop goes around in the direction of the arrows, the north pole of the armature is at the upper left, and the south pole of the armature is at the lower right, as shown in figure 2-2, (view A). Of course, as the loop (armature) turns, these magnetic poles turn with it. Now, as shown in the illustrations, the north armature pole is repelled from the north field pole and attracted to the right by the south field pole. Likewise, the south armature pole is repelled from the south field pole and is attracted to the left by the north field pole. This action causes the armature to turn in a clockwise direction, as shown in figure 2-2 (view B).

After the loop has turned far enough so that its north pole is exactly opposite the south field pole, the brushes advance to the next segments. This changes the direction of current flow through the armature loop. Also, it changes the polarity of the armature field, as shown in figure 2-2 (view C). The magnetic fields again repel and attract each other, and the armature continues to turn.

In this simple motor, the momentum of the rotating armature carries the armature past the position where the unlike poles are exactly lined up. However, if these fields are exactly lined up when the armature current is turned on, there is no momentum to start the armature moving. In this case, the motor would not rotate. It would be necessary to give a motor like this a spin to start it. This disadvantage does not exist when there are more turns on the armature, because there is more than one armature field. No two armature fields could be exactly aligned with the field from the field poles at the same time.

Q1. What factors determine the direction of rotation in a dc motor?

Q2. The right-hand rule for motors is used to find the relationship between what motor characteristics?

Q3. What are the differences between the components of a dc generator and a dc motor?

COUNTER EMF

While a dc motor is running, it acts somewhat like a dc generator. There is a magnetic field from the field poles, and a loop of wire is turning and cutting this magnetic field. For the moment, disregard the fact that there is current flowing through the loop of wire from the battery. As the loop sides cut the magnetic field, a voltage is induced in them, the same as it was in the loop sides of the dc generator. This induced voltage causes current to flow in the loop.

Now, consider the relative direction between this current and the current that causes the motor to run. First, check the direction the current flows as a result of the generator action taking place (view A of fig. 2-2). (Apply the left-hand rule for generators which was discussed in the last chapter.) Using the left hand, hold it so that the forefinger points in the direction of the magnetic field (north to south) and the thumb points in the direction that the black side of the armature moves (up). Your middle finger then points out of the paper (toward you), showing the direction of current flow caused by the generator action in the black half of the armature. This is in the direction opposite to that of the battery current. Since this generator-action voltage is opposite that of the battery, it is called "counter emf." (The letters emf stand for electromotive force, which is another name for voltage.) The two currents are flowing in opposite directions. This proves that the battery voltage and the counter emf are opposite in polarity.

At the beginning of this discussion, we disregarded armature current while explaining how counter emf was generated. Then, we showed that normal armature current flowed opposite to the current created by the counter emf. We talked about two opposite currents that flow at the same time. However, this is a bit oversimplified, as you may already suspect. Actually, only one current flows. Because the counter emf can never become as large as the applied voltage, and because they are of opposite polarity as we have seen, the counter emf effectively cancels part of the armature voltage. The single current that flows is armature current, but it is greatly reduced because of the counter emf.

In a dc motor, there is always a counter emf developed. This counter emf cannot be equal to or greater than the applied battery voltage; if it were, the motor would not run. The counter emf is always a little less. However, the counter emf opposes the applied voltage enough to keep the armature current from the battery to a fairly low value. If there were no such thing as counter emf, much more current would flow through the armature, and the motor would run much faster. However, there is no way to avoid the counter emf.

Q4. What causes counter emf in a dc motor?

Q5. What motor characteristic is affected by counter emf?

MOTOR LOADS

Motors are used to turn mechanical devices, such as water pumps, grinding wheels, fan blades, and circular saws. For example, when a motor is turning a water pump, the water pump is the load. The water pump is the mechanical device that the motor must move. This is the definition of a motor load.

As with electrical loads, the mechanical load connected to a dc motor affects many electrical quantities. Such things as the power drawn from the line, amount of current, speed, efficiency, etc., are all partially controlled by the size of the load. The physical and electrical characteristics of the motor must be matched to the requirements of the load if the work is to be done without the possibility of damage to either the load or the motor.

Q6. What is the load on a dc motor?

PRACTICAL DC MOTORS

As you have seen, dc motors are electrically identical to dc generators. In fact, the same dc machine may be driven mechanically to generate a voltage, or it may be driven electrically to move a mechanical load. While this is not normally done, it does point out the similarities between the two machines. These similarities will be used in the remainder of this chapter to introduce you to practical dc motors. You will immediately recognize series, shunt, and compound types of motors as being directly related to their generator counterparts.

SERIES DC MOTOR

In a series dc motor, the field is connected in series with the armature. The field is wound with a few turns of large wire, because it must carry full armature current. The circuit for a series dc motor is shown in figure 2-3.

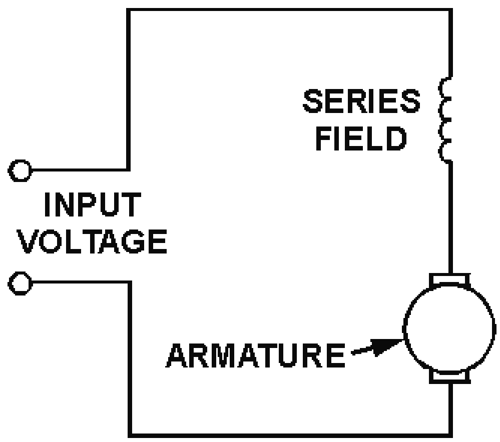


Figure 2-3.—Series-wound dc motor.

This type of motor develops a very large amount of turning force, called torque, from a standstill. Because of this characteristic, the series dc motor can be used to operate small electric appliances, portable electric tools, cranes, winches, hoists, and the like.

Another characteristic is that the speed varies widely between no-load and full-load. Series motors cannot be used where a relatively constant speed is required under conditions of varying load.

A major disadvantage of the series motor is related to the speed characteristic mentioned in the last paragraph. The speed of a series motor with no load connected to it increases to the point where the motor may become damaged. Usually, either the bearings are damaged or the windings fly out of the slots in the armature. There is a danger to both equipment and personnel. Some load must ALWAYS be connected to a series motor before you turn it on. This precaution is primarily for large motors. Small motors, such as those used in electric hand drills, have enough internal friction to load themselves.

A final advantage of series motors is that they can be operated by using either an ac or dc power source. This will be covered in the chapter on ac motors.

Q7. What is the main disadvantage of a series motor?

Q8. What is the main advantage of a series motor?

SHUNT MOTOR

A shunt motor is connected in the same way as a shunt generator. The field windings are connected in parallel (shunt) with the armature windings. The circuit for a shunt motor is shown in figure 2-4.

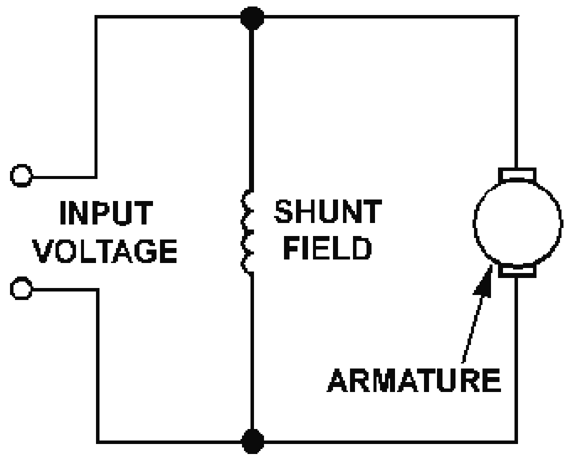


Figure 2-4.—Shunt-wound dc motor.

Once you adjust the speed of a dc shunt motor, the speed remains relatively constant even under changing load conditions. One reason for this is that the field flux remains constant. A constant voltage across the field makes the field independent of variations in the armature circuit.

If the load on the motor is increased, the motor tends to slow down. When this happens, the counter emf generated in the armature decreases. This causes a corresponding decrease in the opposition to battery current flow through the armature. Armature current increases, causing the motor to speed up. The conditions that established the original speed are reestablished, and the original speed is maintained.

Conversely, if the motor load is decreased, the motor tends to increase speed; counter emf increases, armature current decreases, and the speed decreases.

In each case, all of this happens so rapidly that any actual change in speed is slight. There is instantaneous tendency to change rather than a large fluctuation in speed.

Q9. What advantage does a shunt motor have over a series motor?

COMPOUND MOTOR

A compound motor has two field windings, as shown in figure 2-5. One is a shunt field connected in parallel with the armature; the other is a series field that is connected in series with the armature. The shunt field gives this type of motor the constant speed advantage of a regular shunt motor. The series field gives it the advantage of being able to develop a large torque when the motor is started under a heavy load. It should not be a surprise that the compound motor has both shunt- and series-motor characteristics.

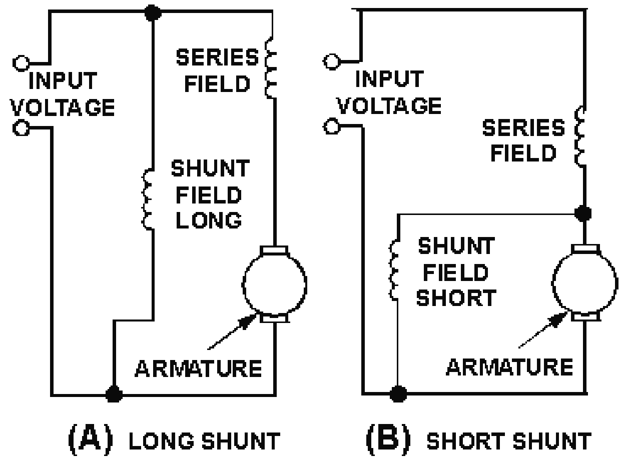


Figure 2-5.—Compound-wound dc motor.

When the shunt field is connected in parallel with the series field and armature, it is called a "long shunt" as shown in figure 2-5, (view A). Otherwise, it is called a "short shunt", as shown in figure 2-5, (view B).

TYPES OF ARMATURES

As with dc generators, dc motors can be constructed using one of two types of armatures. A brief review of the Gramme-ring and drum-wound armatures is necessary to emphasize the similarities between dc generators and dc motors.

GRAMME-RING ARMATURE

The Gramme-ring armature is constructed by winding an insulated wire around a soft-iron ring (fig. 2-6). Eight equally spaced connections are made to the winding. Each of these is connected to a commutator segment. The brushes touch only the top and bottom segments. There are two parallel paths for current to follow — one up the left side and one up the right side. These paths are completed through the top brush back to the positive lead of the battery.

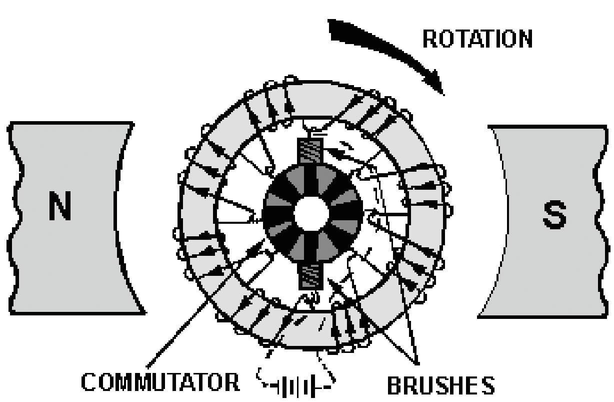


Figure 2-6.—Gramme-ring armature.

To check the direction of rotation of this armature, you should use the right-hand rule for motors. Hold your thumb, forefinger, and middle finger at right angles. Point your forefinger in the direction of field flux; in this case, from left to right. Now turn your wrist so that your middle finger points in the direction that the current flows in the winding on the outside of the ring. Note that current flows into the page (away from you) in the left-hand windings and out of the page (toward you) in the right-hand windings. Your thumb now points in the direction that the winding will move.

The Gramme-ring armature is seldom used in modem dc motors. The windings on the inside of the ring are shielded from magnetic flux, which causes this type of armature to be inefficient. The Gramme-ring armature is discussed primarily to help you better understand the drum-wound armature.

DRUM-WOUND ARMATURE

The drum-wound armature is generally used in ac motors. It is identical to the drum winding discussed in the chapter on dc generators.

If the drum-wound armature were cut in half, an end view at the cut would resemble the drawing in figure 2-7, (view A), Figure 2-7, (view B) is a side view of the armature and pole pieces. Notice that the length of each conductor is positioned parallel to the faces of the pole pieces. Therefore, each conductor of the armature can cut the maximum flux of the motor field. The inefficiency of the Gramme-ring armature is overcome by this positioning.

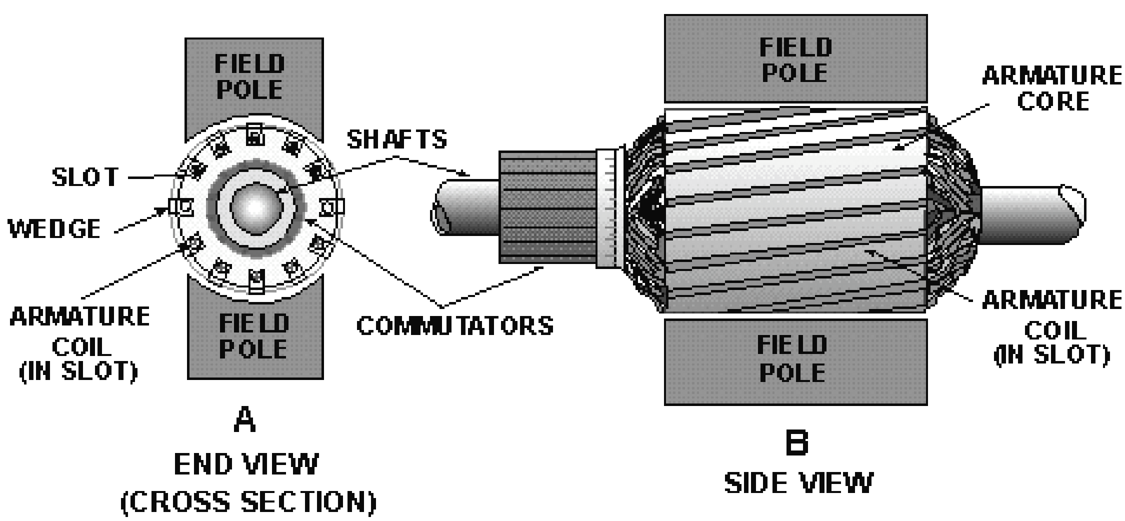


Figure 2-7.—Drum-type armature.

The direction of current flow is marked in each conductor in figure 2-7, (view A) as though the armature were turning in a magnetic field. The dots show that current is flowing toward you on the left side, and the crosses show that the current is flowing away from you on the right side.

Strips of insulation are inserted in the slots to keep windings in place when the armature spins. These are shown as wedges in figure 2-7, (view A).

Q10. Why is the Gramme-ring armature not more widely used?

Q11. How is the disadvantage of the Gramme-ring armature overcome in the drum-wound armature?

DIRECTION OF ROTATION

The direction of rotation of a dc motor depends on the direction of the magnetic field and the direction of current flow in the armature. If either the direction of the field or the direction of current flow through the armature is reversed, the rotation of the motor will reverse. However, if both of these factors are reversed at the same time, the motor will continue rotating in the same direction. In actual practice, the field excitation voltage is reversed in order to reverse motor direction.

Ordinarily, a motor is set up to do a particular job that requires a fixed direction of rotation. However, there are times when it is necessary to change the direction of rotation, such as a drive motor for a gun turret or missile launcher. Each of these must be able to move in both directions. Remember, the connections of either the armature or the field must be reversed, but not both. In such applications, the proper connections are brought out to a reversing switch.

Q12. In a dc motor that must be able to rotate in both directions, how is the direction changed?

MOTOR SPEED

A motor whose speed can be controlled is called a variable-speed motor; dc motors are variable-speed motors. The speed of a dc motor is changed by changing the current in the field or by changing the current in the armature.

When the field current is decreased, the field flux is reduced, and the counter emf decreases. This permits more armature current. Therefore, the motor speeds up. When the field current is increased, the field flux is increased. More counter emf is developed, which opposes the armature current. The armature current then decreases, and the motor slows down.

When the voltage applied to the armature is decreased, the armature current is decreased, and the motor again slows down. When the armature voltage and current are both increased, the motor speeds up.

In a shunt motor, speed is usually controlled by a rheostat connected in series with the field windings, as shown in figure 2-8. When the resistance of the rheostat is increased, the current through the field winding is decreased. The decreased flux momentarily decreases the counter emf. The motor then speeds up, and the increase in counter emf keeps the armature current the same. In a similar manner, a decrease in rheostat resistance increases the current flow through the field windings and causes the motor to slow down.

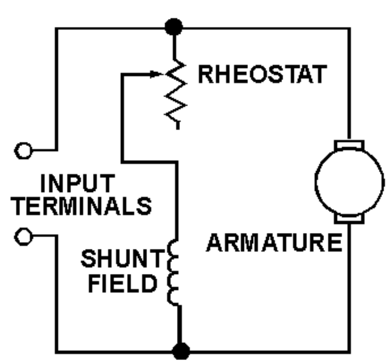


Figure 2-8.—Controlling motor speed.

In a series motor, the rheostat speed control may be connected either in parallel or in series with the armature windings. In either case, moving the rheostat in a direction that lowers the voltage across the armature lowers the current through the armature and slows the motor. Moving the rheostat in a direction that increases the voltage and current through the armature increases motor speed.

Q13. What is the effect on motor speed if the field current is increased?

ARMATURE REACTION

You will remember that the subject of armature reaction was covered in the previous chapter on dc generators. The reasons for armature reaction and the methods of compensating for its effects are basically the same for dc motors as for dc generators.

Figure 2-9 reiterates for you the distorting effect that the armature field has on the flux between the pole pieces. Notice, however, that the effect has shifted the neutral plane backward, against the direction of rotation. This is different from a dc generator, where the neutral plane shifted forward in the direction of rotation.

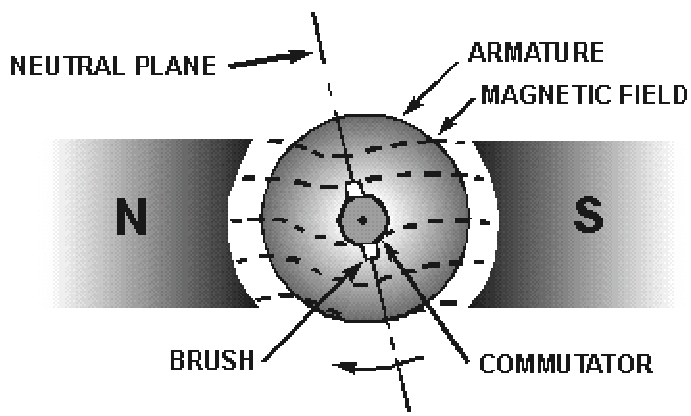


Figure 2-9.—Armature reaction.

As before, the brushes must be shifted to the new neutral plane. As shown in figure 2-9, the shift is counterclockwise. Again, the proper location is reached when there is no sparking from the brushes.

Q14. Armature reaction in a dc motor causes a shift of the neutral plane in which direction?

Compensating windings and interpoles, two more "old" subjects, cancel armature reaction in dc motors. Shifting brushes reduces sparking, but it also makes the field less effective. Canceling armature reaction eliminates the need to shift brushes in the first place.

Compensating windings and interpoles are as important in motors as they are in generators. Compensating windings are relatively expensive; therefore, most large dc motors depend on interpoles to correct armature reaction. Compensating windings are the same in motors as they are in generators. Interpoles, however, are slightly different. The difference is that in a generator the interpole has the same polarity as the main pole AHEAD of it in the direction of rotation. In a motor the interpole has the same polarity as the main pole FOLLOWING it.

The interpole coil in a motor is connected to carry the armature current the same as in a generator. As the load varies, the interpole flux varies, and commutation is automatically corrected as the load changes. It is not necessary to shift the brushes when there is an increase or decrease in load. The brushes are located on the no-load neutral plane. They remain in that position for all conditions of load.

Q15. What current flows in the interpole windings?

The dc motor is reversed by reversing the direction of the current in the armature. When the armature current is reversed, the current through the interpole is also reversed. Therefore, the interpole still has the proper polarity to provide automatic commutation.

MANUAL AND AUTOMATIC STARTERS

Because the dc resistance of most motor armatures is low (0.05 to 0.5 ohm), and because the counter emf does not exist until the armature begins to turn, it is necessary to use an external starting resistance in series with the armature of a dc motor to keep the initial armature current to a safe value. As the armature begins to turn, counter emf increases; and, since the counter emf opposes the applied voltage, the armature current is reduced. The external resistance in series with the armature is decreased or eliminated as the motor comes up to normal speed and full voltage is applied across the armature.

Controlling the starting resistance in a dc motor is accomplished either manually, by an operator, or by any of several automatic devices. The automatic devices are usually just switches controlled by motor speed sensors. Automatic starters are not covered in detail in this module.

Q16. What is the purpose of starting resistors?

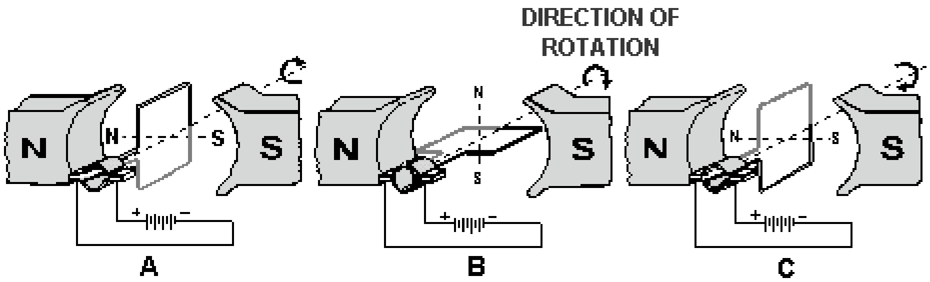
SUMMARY

This chapter presented the operating principles and characteristics of direct-current motors. The following information provides a summary of the main subjects for review.

The main **PRINCIPLE OF A DC MOTOR** is that current flow through the armature coil causes the armature to act as a magnet. The armature poles are attracted to field poles of opposite polarity, causing the armature to rotate.

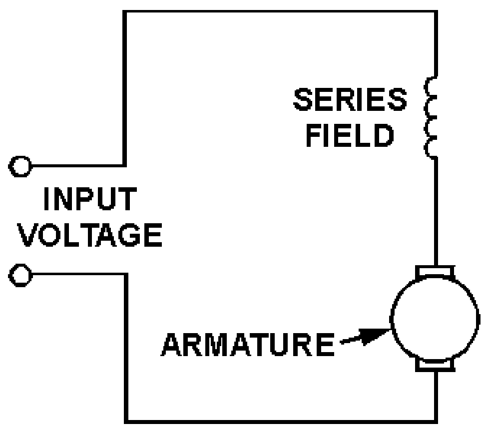
The **CONSTRUCTION** of a dc motor is almost identical to that of a dc generator, both physically and electrically. In fact, most dc generators can be made to act as dc motors, and vice versa.

COMMUTATION IN A DC MOTOR is the process of reversing armature current at the moment when unlike poles of the armature and field are facing each other, thereby reversing the polarity of the armature field. Like poles of the armature and field then repel each other, causing armature rotation to continue.

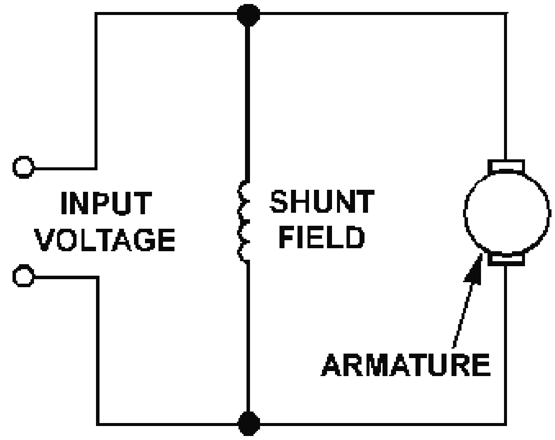


COUNTER-ELECTROMOTIVE FORCE is generated in a dc motor as armature coils cut the field flux. This emf opposes the applied voltage, and limits the flow of armature current.

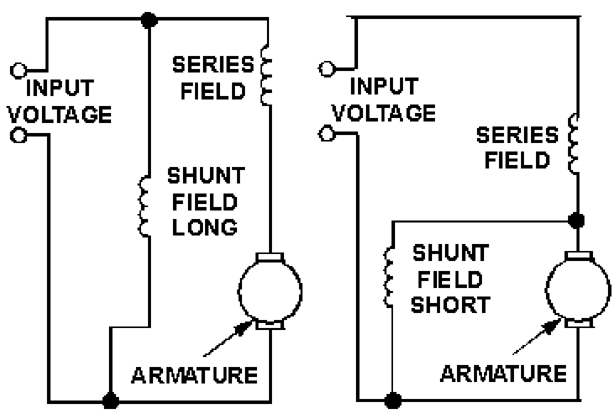
In **SERIES MOTORS**, the field windings are connected in series with the armature coil. The field strength varies with changes in armature current. When its speed is reduced by a load, the series motor develops greater torque. Its starting torque is greater than other types of dc motors. Its speed varies widely between full-load and no-load. Unloaded operation of large machines is dangerous.



In **SHUNT MOTORS**, the field windings are connected in parallel (shunt) across the armature coil. The field strength is independent of the armature current. Shunt-motor speed varies only slightly with changes in load, and the starting torque is less than that of other types of dc motors.



In **COMPOUND MOTORS**, one set of field windings is connected in series with the armature, and one set is connected in parallel. The speed and torque characteristics are a combination of the desirable characteristics of both series and shunt motors.

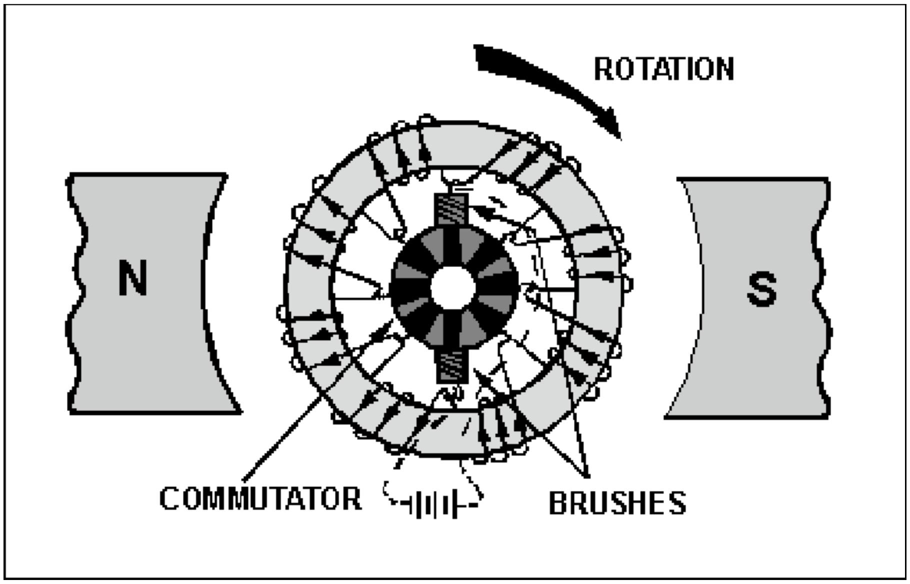


(A) LONG SHUNT (B) SHORT SHUNT

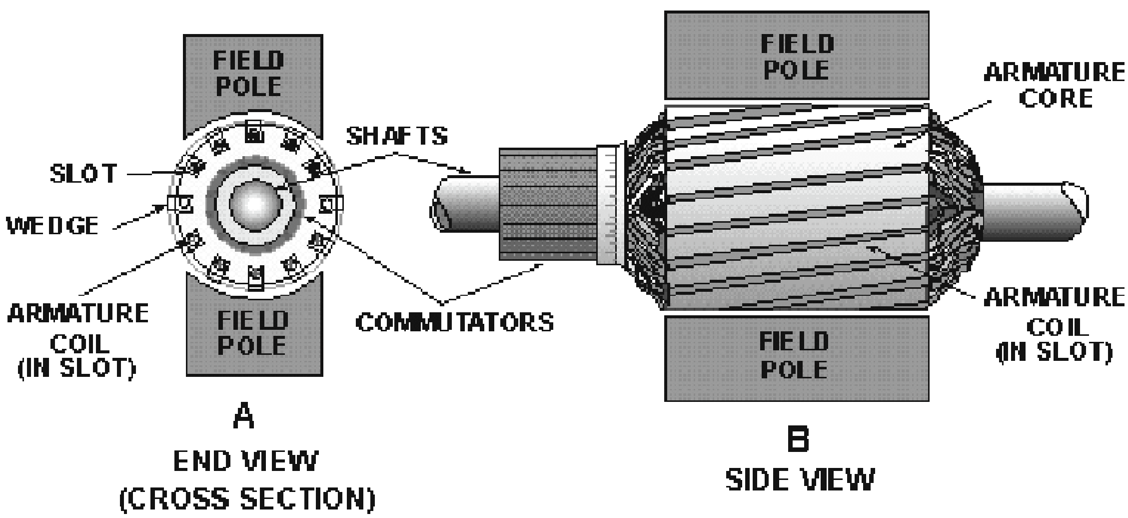
LOAD on a motor is the physical object to be moved by the motor.

DC MOTOR ARMATURES are of two types. They are the Gramme-ring and the drum-wound types.

THE GRAMME-RING ARMATURE is inefficient since part of each armature coil is prevented from cutting flux lines. Gramme-ring wound armatures are seldom used for this reason.

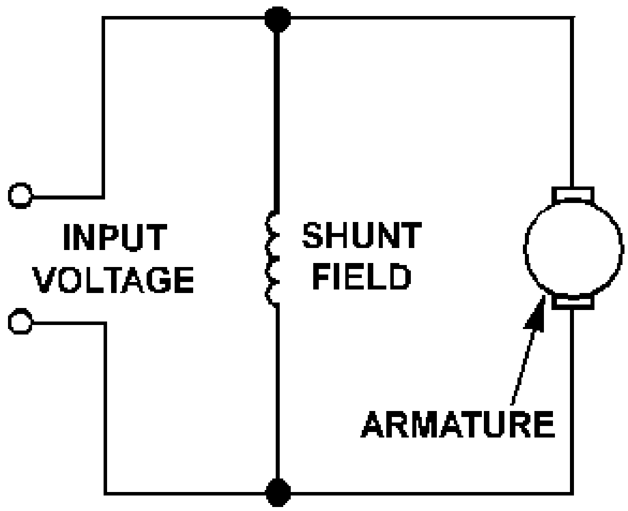


THE DRUM-WOUND ARMATURE consists of coils actually wound around the armature core so that all coil surfaces are exposed to the magnetic field. Nearly all dc motors have drum-wound armatures.



MOTOR REVERSAL in a dc motor can be accomplished by reversing the field connections or by reversing the armature connections. If both are reversed, rotation will continue in the original direction.

SPEED CONTROL IN A DC MOTOR is maintained by varying the resistance either in series with the field coil or in series with the armature coil. Increasing shunt-field circuit resistance increases motor speed. Increasing the armature circuit resistance decreases motor speed.



ARMATURE REACTION is the distortion of the main field in a motor by the armature field. This causes the neutral plane to be shifted in the direction opposite to that of armature rotation. Interpoles and compensating windings are used to reduce the effect of armature reaction on motor operation.

STARTING RESISTORS are necessary since the dc resistance of a motor armature is very low. Excessive current will flow when dc voltage is first applied unless current is limited in some way. Adding resistance in series with the armature windings reduces initial current. It may then be removed after counter emf has been built up.

ANSWERS TO QUESTIONS Q1. THROUGH Q16.

A1. Direction of armature current, and direction of magnetic flux in field.

A2. Direction of conductor movement (rotation), direction of flux, and the direction of current flow.

A3. There are no differences.

A4. Generator action.

A5. Speed.

A6. The device to be driven by the motor.

A7. It must have a load connected to avoid damage from excess speed.

A8. High torque (turning force) at low speed.

A9. It maintains a constant speed under varying loads.

A10. Only outside of coils cut flux (inefficient).

A11. By winding the armature in a way that places the entire coil where it is exposed to maximum flux.

A12. By reversing either field or armature connections.

A13. Motor will slow down.

A14. Opposite the rotation.

A15. Armature current.

A16. To limit armature current until counter emf builds up.

CHAPTER 3

ALTERNATING CURRENT GENERATORS

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to:

1. Describe the principle of magnetic induction as it applies to ac generators.
2. Describe the differences between the two basic types of ac generators.
3. List the advantages and disadvantages of the two types of ac generators.
4. Describe exciter generators within alternators; discuss construction and purpose.
5. Compare the types of rotors used in ac generators, and the applications of each type to different prime movers.
6. Explain the factors that determine the maximum power output of an ac generator, and the effect of these factors in rating generators.
7. Explain the operation of multiphase ac generators and compare with single-phase.
8. Describe the relationships between the individual output and resultant vectorial sum voltages in multiphase generators.
9. Explain, using diagrams, the different methods of connecting three-phase alternators and transformers.
10. List the factors that determine the frequency and voltage of the alternator output.
11. Explain the terms voltage control and voltage regulation in ac generators, and list the factors that affect each quantity.
12. Describe the purpose and procedure of parallel generator operation.

INTRODUCTION

Most of the electrical power used aboard Navy ships and aircraft as well as in civilian applications is ac. As a result, the ac generator is the most important means of producing electrical power. Ac generators, generally called alternators, vary greatly in size depending upon the load to which they supply power. For example, the alternators in use at hydroelectric plants, such as Hoover Dam, are tremendous in size, generating thousands of kilowatts at very high voltage levels. Another example is the alternator in a typical automobile, which is very small by comparison. It weighs only a few pounds and produces between 100 and 200 watts of power, usually at a potential of 12 volts.

Many of the terms and principles covered in this chapter will be familiar to you. They are the same as those covered in the chapter on dc generators. You are encouraged to refer back, as needed, and to refer to any other source that will help you master the subject of this chapter. No one source meets the complete needs of everyone.

BASIC AC GENERATORS

Regardless of size, all electrical generators, whether dc or ac, depend upon the principle of magnetic induction. An emf is induced in a coil as a result of (1) a coil cutting through a magnetic field, or (2) a magnetic field cutting through a coil. As long as there is relative motion between a conductor and a magnetic field, a voltage will be induced in the conductor. That part of a generator that produces the magnetic field is called the field. That part in which the voltage is induced is called the armature. For relative motion to take place between the conductor and the magnetic field, all generators must have two mechanical parts — a rotor and a stator. The <u>ROT</u>or is the part that <u>ROT</u>ates; the <u>STAT</u>or is the part that remains <u>STAT</u>ionary. In a dc generator, the armature is always the rotor. In alternators, the armature may be either the rotor or stator.

Q1. Magnetic induction occurs when there is relative motion between what two elements?

ROTATING-ARMATURE ALTERNATORS

The rotating-armature alternator is similar in construction to the dc generator in that the armature rotates in a stationary magnetic field as shown in figure 3-1, view A. In the dc generator, the emf generated in the armature windings is converted from ac to dc by means of the commutator. In the alternator, the generated ac is brought to the load unchanged by means of slip rings. The rotating armature is found only in alternators of low power rating and generally is not used to supply electric power in large quantities.

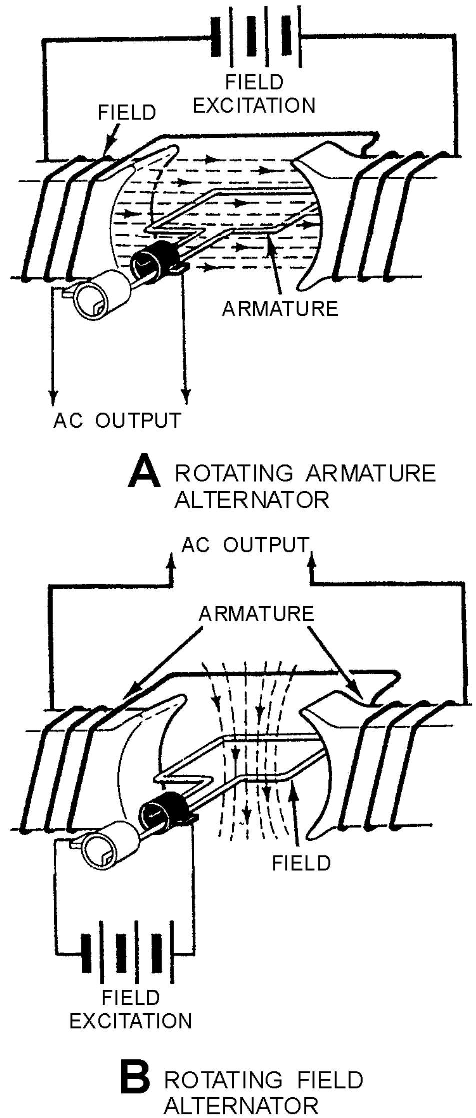


Figure 3-1.—Types of ac generators.

ROTATING-FIELD ALTERNATORS

The rotating-field alternator has a stationary armature winding and a rotating-field winding as shown in figure 3-1, view B The advantage of having a stationary armature winding is that the generated voltage can be connected directly to the load.

A rotating armature requires slip rings and brushes to conduct the current from the armature to the load. The armature, brushes, and slip rings are difficult to insulate, and arc-overs and short circuits can result at high voltages. For this reason, high-voltage alternators are usually of the rotating-field type. Since the voltage applied to the rotating field is low voltage dc, the problem of high voltage arc-over at the slip rings does not exist.

The stationary armature, or stator, of this type of alternator holds the windings that are cut by the rotating magnetic field. The voltage generated in the armature as a result of this cutting action is the ac power that will be applied to the load.

The stators of all rotating-field alternators are about the same. The stator consists of a laminated iron core with the armature windings embedded in this core as shown in figure 3-2. The core is secured to the stator frame.

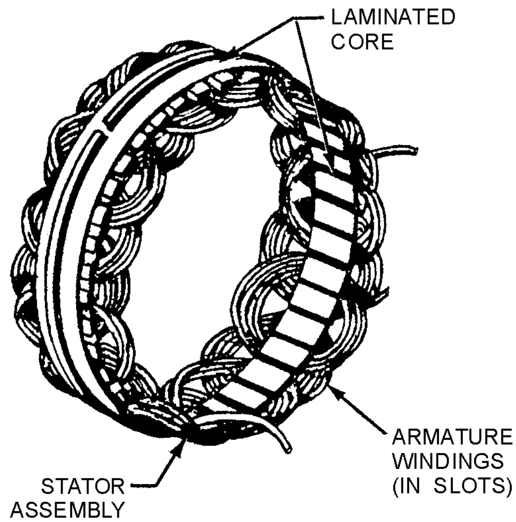


Figure 3-2.—Stationary armature windings.

Q2. What is the part of an alternator in which the output voltage is generated?

Q3. What are the two basic types of alternators?

Q4. What is the main advantage of the rotating field alternator?

PRACTICAL ALTERNATORS

The alternators described so far in this chapter are ELEMENTARY in nature; they are seldom used except as examples to aid in understanding practical alternators.

The remainder of this chapter will relate the principles of the elementary alternator to the alternators actually in use in the civilian community, as well as aboard Navy ships and aircraft. The following paragraphs in this chapter will introduce such concepts as prime movers, field excitation, armature characteristics and limitations, single-phase and polyphase alternators, controls, regulation, and parallel operation.

FUNCTIONS OF ALTERNATOR COMPONENTS

A typical rotating-field ac generator consists of an alternator and a smaller dc generator built into a single unit. The output of the alternator section supplies alternating voltage to the load. The only purpose for the dc exciter generator is to supply the direct current required to maintain the alternator field. This dc generator is referred to as the exciter. A typical alternator is shown in figure 3-3, view A; figure 3-3, view B, is a simplified schematic of the generator.

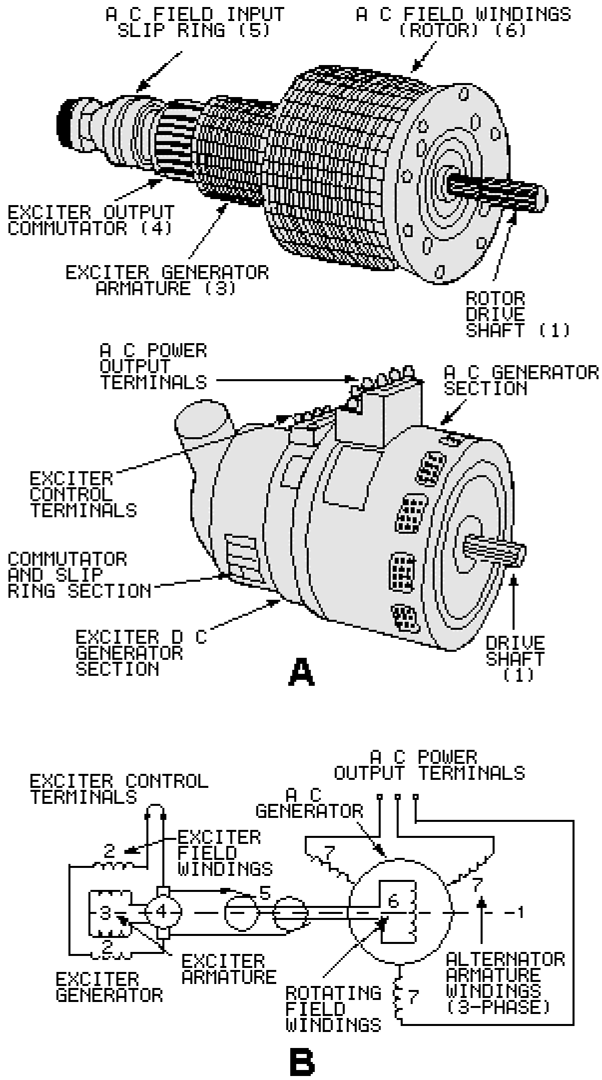


Figure 3-3.—Ac generator pictorial and schematic drawings.

The exciter is a dc, shunt-wound, self-excited generator. The exciter shunt field (2) creates an area of intense magnetic flux between its poles. When the exciter armature (3) is rotated in the exciter-field flux, voltage is induced in the exciter armature windings. The output from the exciter commutator (4) is connected through brushes and slip rings (5) to the alternator field. Since this is direct current already converted by the exciter commutator, the current always flows in one direction through the alternator field (6). Thus, a fixed-polarity magnetic field is maintained at all times in the alternator field windings. When the alternator field is rotated, its magnetic flux is passed through and across the alternator armature windings (7).

The armature is wound for a three-phase output, which will be covered later in this chapter. Remember, a voltage is induced in a conductor if it is stationary and a magnetic field is passed across the conductor, the same as if the field is stationary and the conductor is moved. The alternating voltage in the ac generator armature windings is connected through fixed terminals to the ac load.

Q5. Most large alternators have a small dc generator built into them. What is its purpose?

PRIME MOVERS

All generators, large and small, ac and dc, require a source of mechanical power to turn their rotors. This source of mechanical energy is called a prime mover.

Prime movers are divided into two classes for generators-high-speed and low-speed. Steam and gas turbines are high-speed prime movers, while internal-combustion engines, water, and electric motors are considered low-speed prime movers.

The type of prime mover plays an important part in the design of alternators since the speed at which the rotor is turned determines certain characteristics of alternator construction and operation.

ALTERNATOR ROTORS

There are two types of rotors used in rotating-field alternators. They are called the turbine-driven and salient-pole rotors.

As you may have guessed, the turbine-driven rotor shown in figure 3-4, view A, is used when the prime mover is a high-speed turbine. The windings in the turbine-driven rotor are arranged to form two or four distinct poles. The windings are firmly embedded in slots to withstand the tremendous centrifugal forces encountered at high speeds.

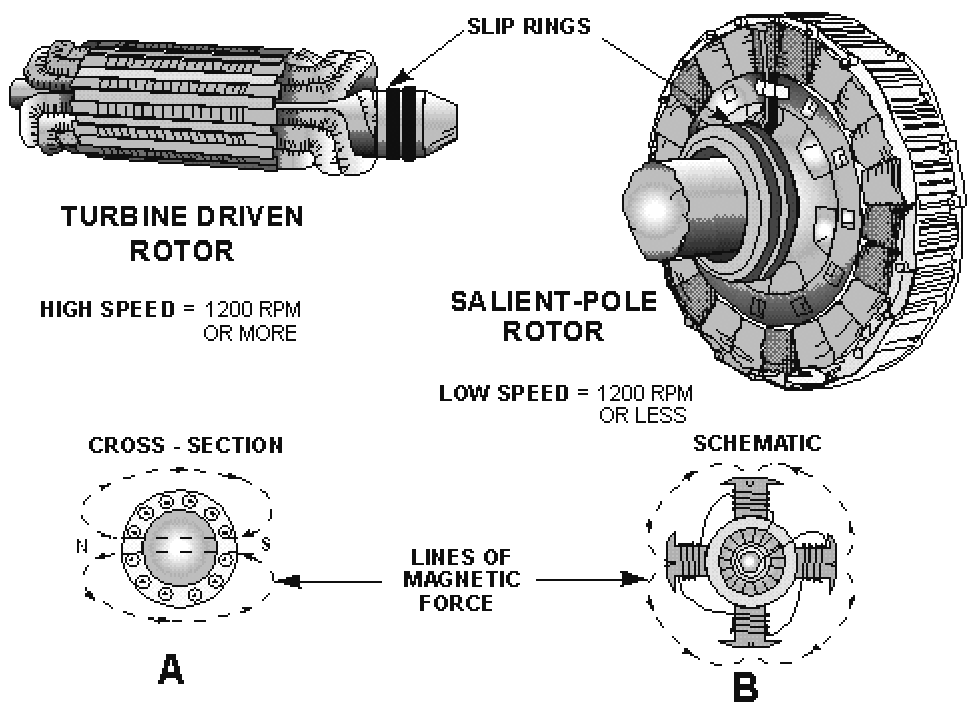


Figure 3-4.—Types of rotors used in alternators.

The salient-pole rotor shown in figure 3-4, view B, is used in low-speed alternators. The salient-pole rotor often consists of several separately wound pole pieces, bolted to the frame of the rotor.

If you could compare the physical size of the two types of rotors with the same electrical characteristics, you would see that the salient-pole rotor would have a greater diameter. At the same number of revolutions per minute, it has a greater centrifugal force than does the turbine-driven rotor. To reduce this force to a safe level so that the windings will not be thrown out of the machine, the salient pole is used only in low-speed designs.

ALTERNATOR CHARACTERISTICS AND LIMITATIONS

Alternators are rated according to the voltage they are designed to produce and the maximum current they are capable of providing. The maximum current that can be supplied by an alternator depends upon the maximum heating loss that can be sustained in the armature. This heating loss (which is an I^2R power loss) acts to heat the conductors, and if excessive, destroys the insulation. Thus, alternators are rated in terms of this current and in terms of the voltage output — the alternator rating in small units is in volt-amperes; in large units it is kilovolt-amperes.

When an alternator leaves the factory, it is already destined to do a very specific job. The speed at which it is designed to rotate, the voltage it will produce, the current limits, and other operating characteristics are built in. This information is usually stamped on a nameplate on the case so that the user will know the limitations.

Q6. How are alternators usually rated?

Q7. What type of prime mover requires a specially designed high-speed alternator?

Q8. Salient-pole rotors may be used in alternators driven by what types of prime movers?

SINGLE-PHASE ALTERNATORS

A generator that produces a single, continuously alternating voltage is known as a SINGLE-PHASE alternator. All of the alternators that have been discussed so far fit this definition. The stator (armature) windings are connected in series. The individual voltages, therefore, add to produce a single-phase ac voltage. Figure 3-5 shows a basic alternator with its single-phase output voltage.

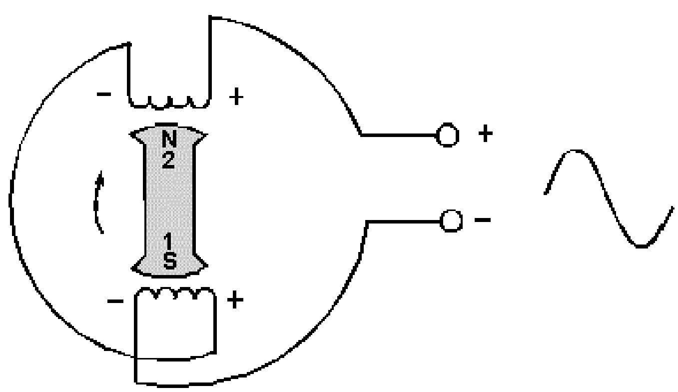


Figure 3-5.—Single-phase alternator.

The definition of phase as you learned it in studying ac circuits may not help too much right here. Remember, "out of phase" meant "out of time."

Now, it may be easier to think of the word *phase* as meaning voltage as in single voltage. The need for a modified definition of phase in this usage will be easier to see as we go along.

Single-phase alternators are found in many applications. They are most often used when the loads being driven are relatively light. The reason for this will be more apparent as we get into multiphase alternators (also called polyphase).

Power that is used in homes, shops, and ships to operate portable tools and small appliances is single-phase power. Single-phase power alternators always generate single-phase power. However, all single-phase power does not come from single-phase alternators. This will sound more reasonable to you as we get into the next subjects.

Q9. What does the term single phase indicate?

Q10. In single-phase alternators, in order for the voltages induced in all the armature windings to add together for a single output, how must the windings be connected?

TWO-PHASE ALTERNATORS

Two phase implies two voltages if we apply our new definition of phase. And, it's that simple. A two-phase alternator is designed to produce two completely separate voltages. Each voltage, by itself, may be considered as a single-phase voltage. Each is generated completely independent of the other. Certain advantages are gained. These and the mechanics of generation will be covered in the following paragraphs.

Generation of Two-Phase Power

Figure 3-6 shows a simplified two-pole, two-phase alternator. Note that the windings of the two phases are physically at right angles (90º) to each other. You would expect the outputs of each phase to be 90º apart, which they are. The graph shows the two phases to be 90º apart, with A leading B. Note that by using our original definition of phase (from previous modules), we could say that A and B are 90º out of phase. There will always be 90º between the phases of a two-phase alternator. This is by design.

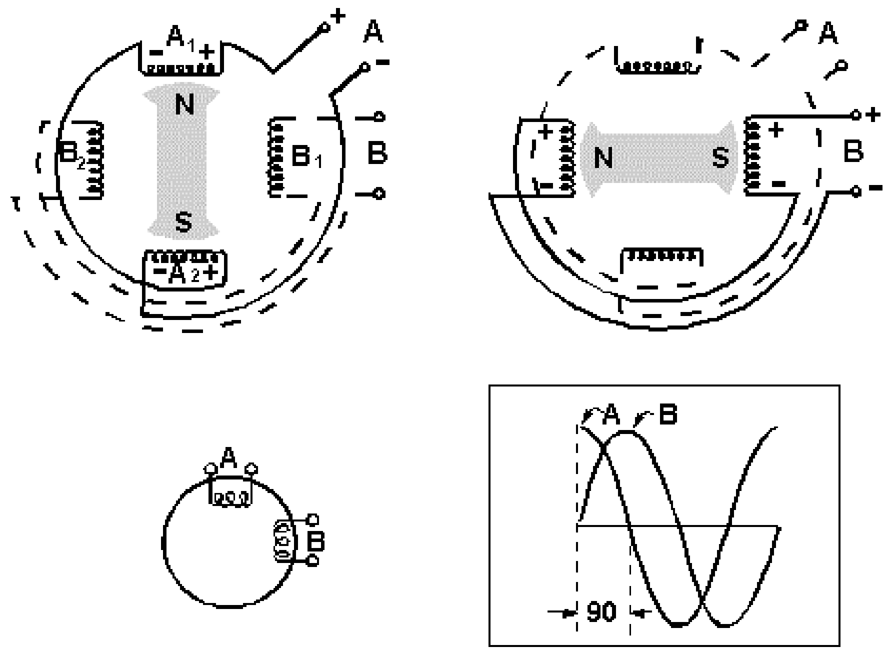


Figure 3-6.—Two-phase alternator.

Now, let's go back and see the similarities and differences between our original (single-phase) alternators and this new one (two-phase). Note that the principles applied are not new. This alternator works the same as the others we have discussed.

The stator in figure 3-6 consists of two single-phase windings completely separated from each other. Each winding is made up of two windings that are connected in series so that their voltages add. The rotor is identical to that used in the single-phase alternator. In the left-hand schematic, the rotor poles are opposite all the windings of phase A. Therefore, the voltage induced in phase A is maximum, and the voltage induced in phase B is zero. As the rotor continues rotating counterclockwise, it moves away from the A windings and approaches the B windings. As a result, the voltage induced in phase A decreases from its maximum value, and the voltage induced in phase B increases from zero. In the right-hand schematic, the rotor poles are opposite the windings of phase B. Now the voltage induced in phase B is maximum, whereas the voltage induced in phase A has dropped to zero. Notice that a 90-degree rotation of the rotor corresponds to one-quarter of a cycle, or 90 electrical degrees. The waveform picture shows the voltages induced in phase A and B for one cycle. The two voltages are 90º out of phase. Notice that the two outputs, A and B, are independent of each other. Each output is a single-phase voltage, just as if the other did not exist.

The obvious advantage, so far, is that we have two separate output voltages. There is some saving in having one set of bearings, one rotor, one housing, and so on, to do the work of two. There is the disadvantage of having twice as many stator coils, which require a larger and more complex stator.

The large schematic in figure 3-7 shows four separate wires brought out from the A and B stator windings. This is the same as in figure 3-6. Notice, however, that the dotted wire now connects one end of B1 to one end of A2. The effect of making this connection is to provide a new output voltage. This sine-wave voltage, C in the picture, is larger than either A or B. It is the result of adding the instantaneous values of phase A and phase B. For this reason it appears exactly half way between A and B. Therefore, C must lag A by 45º and lead B by 45º, as shown in the small vector diagram.

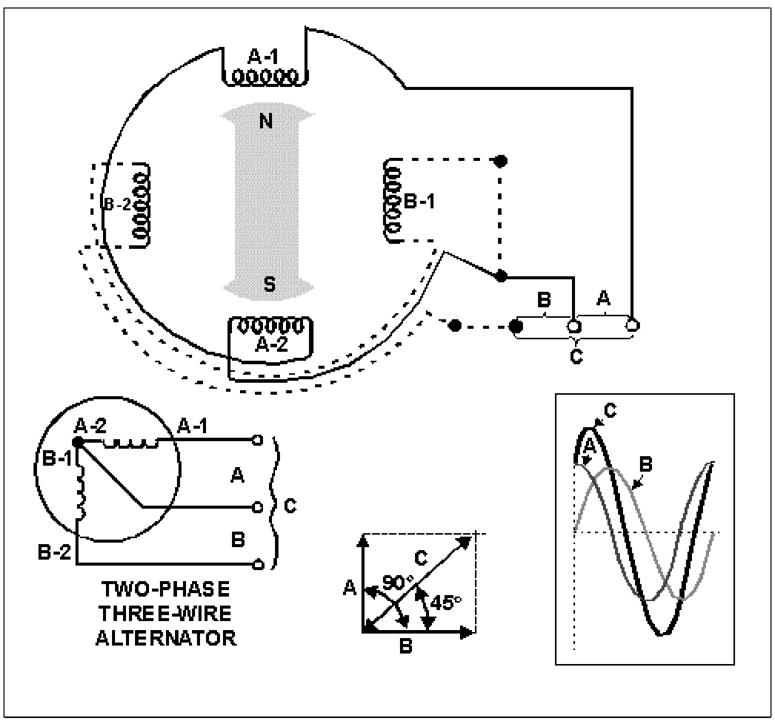


Figure 3-7.—Connections of a two-phase, three-wire alternator output.

Now, look at the smaller schematic diagram in figure 3-7. Only three connections have been brought out from the stator. Electrically, this is the same as the large diagram above it. Instead of being connected at the output terminals, the B1-A2 connection was made internally when the stator was wired. A two-phase alternator connected in this manner is called a two-phase, three-wire alternator.

The three-wire connection makes possible three different load connections: A and B (across each phase), and C (across both phases). The output at C is always 1.414 times the voltage of either phase. These multiple outputs are additional advantages of the two-phase alternator over the single-phase type.

Now, you can understand why single-phase power doesn't always come from single-phase alternators. It can be generated by two-phase alternators as well as other multiphase (polyphase) alternators, as you will soon see.

The two-phase alternator discussed in the preceding paragraphs is seldom seen in actual use. However, the operation of polyphase alternators is more easily explained using two phases than three phases. The three-phase alternator, which will be covered next, is by far the most common of all alternators in use today, both in military and civilian applications.

Q11. What determines the phase relationship between the voltages in a two-phase ac generator?

Q12. How many voltage outputs are available from a two-phase three-wire alternator?

Q13. What is the relationship of the voltage at C in figure 3-7 to the voltages at A and B?

THREE-PHASE ALTERNATOR

The three-phase alternator, as the name implies, has three single-phase windings spaced such that the voltage induced in any one phase is displaced by 120º from the other two. A schematic diagram of a three-phase stator showing all the coils becomes complex, and it is difficult to see what is actually happening. The simplified schematic of figure 3-8, view A, shows all the windings of each phase lumped together as one winding. The rotor is omitted for simplicity. The voltage waveforms generated across each phase are drawn on a graph, phase-displaced 120º from each other. The three-phase alternator as shown in this schematic is made up of three single-phase alternators whose generated voltages are out of phase by 120º . The three phases are independent of each other.

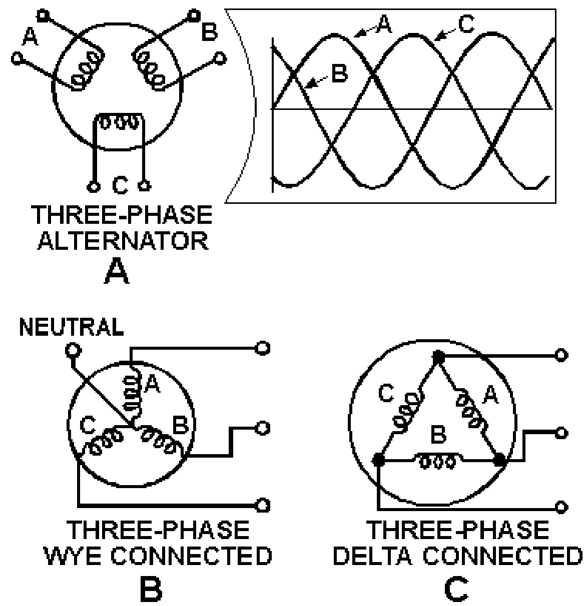


Figure 3-8.—Three-phase alternator connections.

Rather than having six leads coming out of the three-phase alternator, the same leads from each phase may be connected together to form a wye (Y) connection, as shown in figure 3-8, view B. It is called a wye connection because, without the neutral, the windings appear as the letter Y, in this case sideways or upside down.

The neutral connection is brought out to a terminal when a single-phase load must be supplied. Single-phase voltage is available from neutral to A, neutral to B, and neutral to C.

In a three-phase, Y-connected alternator, the total voltage, or line voltage, across any two of the three line leads is the vector sum of the individual phase voltages. Each line voltage is 1.73 times one of the phase voltages. Because the windings form only one path for current flow between phases, the line and phase currents are the same (equal).

A three-phase stator can also be connected so that the phases are connected end-to-end; it is now delta connected (fig. 3-8, view C). (Delta because it looks like the Greek letter delta, Δ.) In the delta connection, line voltages are equal to phase voltages, but each line current is equal to 1.73 times the phase current. Both the wye and the delta connections are used in alternators.

The majority of all alternators in use in the Navy today are three-phase machines. They are much more efficient than either two-phase or single-phase alternators.

Three-Phase Connections

The stator coils of three-phase alternators may be joined together in either wye or delta connections, as shown in figure 3-9. With these connections only three wires come out of the alternator. This allows convenient connection to three-phase motors or power distribution transformers. It is necessary to use three-phase transformers or their electrical equivalent with this type of system.

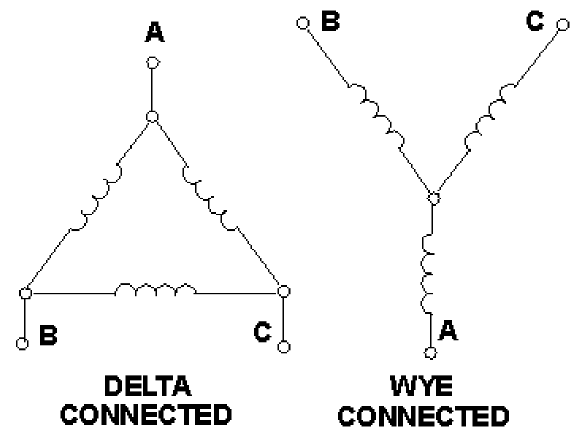


Figure 3-9.—Three-phase alternator or transformer connections.

A three-phase transformer may be made up of three, single-phase transformers connected in delta, wye, or a combination of both. If both the primary and secondary are connected in wye, the transformer is called a wye-wye. If both windings are connected in delta, the transformer is called a delta-delta.

Figure 3-10 shows single-phase transformers connected delta-delta for operation in a three-phase system. You will note that the transformer windings are not angled to illustrate the typical delta (Δ) as has been done with alternator windings. Physically, each transformer in the diagram stands alone. There is no angular relationship between the windings of the individual transformers. However, if you follow the connections, you will see that they form an electrical delta. The primary windings, for example, are connected to each other to form a closed loop. Each of these junctions is fed with a phase voltage from a three-phase alternator. The alternator may be connected either delta or wye depending on load and voltage requirements, and the design of the system.

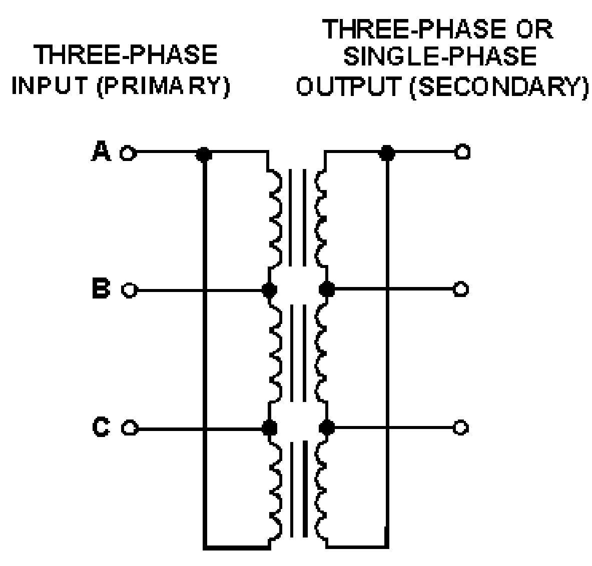


Figure 3-10.—Three single-phase transformers connected delta-delta.

Figure 3-11 shows three single-phase transformers connected wye-wye. Again, note that the transformer windings are not angled. Electrically, a Y is formed by the connections. The lower connections of each winding are shorted together. These form the common point of the wye. The opposite end of each winding is isolated. These ends form the arms of the wye.

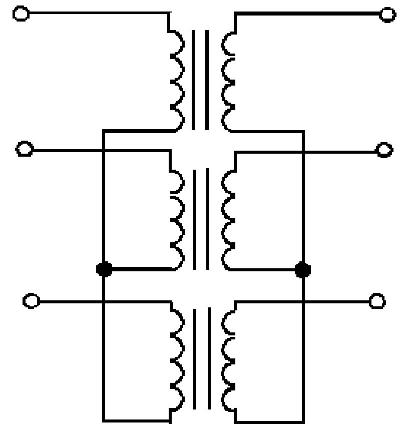


Figure 3-11.—Three single-phase transformers connected wye-wye.

The ac power on most ships is distributed by a three-phase, three-wire, 450-volt system. The single-phase transformers step the voltage down to 117 volts. These transformers are connected delta-delta as in figure 3-10. With a delta-delta configuration, the load may be a three-phase device connected to all phases; or, it may be a single-phase device connected to only one phase.

At this point, it is important to remember that such a distribution system includes everything between the alternator and the load. Because of the many choices that three-phase systems provide, care must be taken to ensure that any change of connections does not provide the load with the wrong voltage or the wrong phase.

Q14. In a three-phase alternator, what is the phase relationship between the individual output voltages?

Q15. What are the two methods of connecting the outputs from a three-phase alternator to the load?

Q16. Ships' generators produce 450-volt, three-phase, ac power; however, most equipment uses 117-volt, single-phase power What transformers and connections are used to convert 450-volt, three-phase power to 117-volt, single-phase power?

FREQUENCY

The output frequency of alternator voltage depends upon the speed of rotation of the rotor and the number of poles. The faster the speed, the higher the frequency. The lower the speed, the lower the frequency. The more poles there are on the rotor, the higher the frequency is for a given speed. When a rotor has rotated through an angle such that two adjacent rotor poles (a north and a south pole) have passed one winding, the voltage induced in that winding will have varied through one complete cycle. For a given frequency, the more pairs of poles there are, the lower the speed of rotation. This principle is illustrated in figure 3-12; a two-pole generator must rotate at four times the speed of an eight-pole generator to produce the same frequency of generated voltage. The frequency of any ac generator in hertz (Hz), which is the number of cycles per second, is related to the number of poles and the speed of rotation, as expressed by the equation

$$F = \frac{NP}{120}$$

where P is the number of poles, N is the speed of rotation in revolutions per minute (rpm), and 120 is a constant to allow for the conversion of minutes to seconds and from poles to pairs of poles. For example, a 2-pole, 3600-rpm alternator has a frequency of 60 Hz; determined as follows:

$$\frac{2 \times 3600}{120} = 60\text{Hz}$$

A 4-pole, 1800-rpm generator also has a frequency of 60 Hz. A 6-pole, 500-rpm generator has a frequency of

$$\frac{6 \times 500}{120} = 25\text{Hz}$$

A 12-pole, 4000-rpm generator has a frequency of

$$\frac{12 \times 4000}{120} = 400\text{Hz}$$

Q17. What two factors determine the frequency of the output voltage of an alternator?

Q18. What is the frequency of the output voltage of an alternator with four poles that is rotated at 3600 rpm?

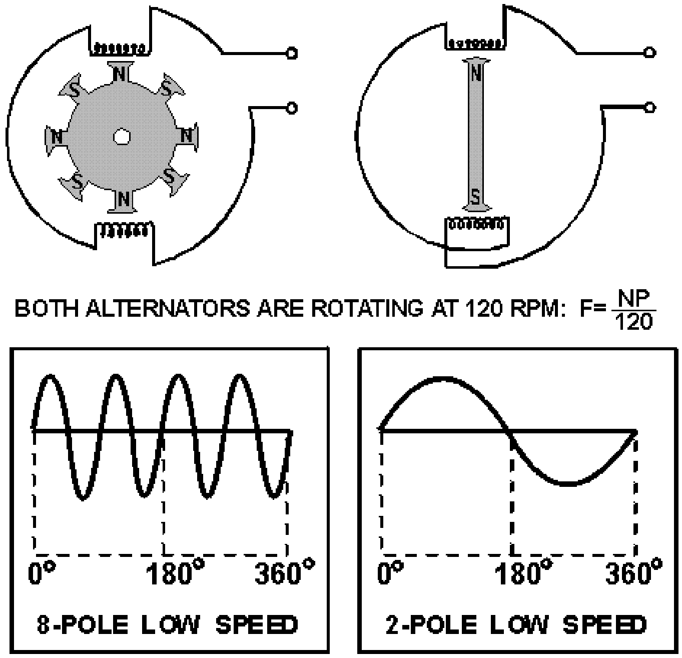


Figure 3-12.—Frequency regulation.

VOLTAGE REGULATION

As we have seen before, when the load on a generator is changed, the terminal voltage varies. The amount of variation depends on the design of the generator.

The voltage regulation of an alternator is the change of voltage from full load to no load, expressed as a percentage of full-load volts, when the speed and dc field current are held constant.

$$\frac{E_{nL} - E_{fL}}{E_{fL}} \times 100 = \text{Percent of regulation}$$

Assume the no-load voltage of an alternator is 250 volts and the full-load voltage is 220 volts. The percent of regulation is

$$\frac{250 - 220}{220} \times 100 = 13.6\%$$

Remember, the lower the percent of regulation, the better it is in most applications.

Q19. The variation in output voltage as the load changes is referred to as what? How is it expressed?

PRINCIPLES OF AC VOLTAGE CONTROL

In an alternator, an alternating voltage is induced in the armature windings when magnetic fields of alternating polarity are passed across these windings. The amount of voltage induced in the windings depends mainly on three things: (1) the number of conductors in series per winding, (2) the speed (alternator rpm) at which the magnetic field cuts the winding, and (3) the strength of the magnetic field. Any of these three factors could be used to control the amount of voltage induced in the alternator windings.

The number of windings, of course, is fixed when the alternator is manufactured. Also, if the output frequency is required to be of a constant value, then the speed of the rotating field must be held constant. This prevents the use of the alternator rpm as a means of controlling the voltage output. Thus, the only practical method for obtaining voltage control is to control the strength of the rotating magnetic field. The strength of this electromagnetic field may be varied by changing the amount of current flowing through the field coil. This is accomplished by varying the amount of voltage applied across the field cod.

Q20. How is output voltage controlled in practical alternators?

PARALLEL OPERATION OF ALTERNATORS

Alternators are connected in parallel to (1) increase the output capacity of a system beyond that of a single unit, (2) serve as additional reserve power for expected demands, or (3) permit shutting down one machine and cutting in a standby machine without interrupting power distribution. When alternators are of sufficient size, and are operating at different frequencies and terminal voltages, severe damage may result if they are suddenly connected to each other through a common bus. To avoid this, the machines must be synchronized as closely as possible before connecting them together. This may be accomplished by connecting one generator to the bus (referred to as bus generator), and then synchronizing the other (incoming generator) to it before closing the incoming generator's main power contactor. The generators are synchronized when the following conditions are set:

1. Equal terminal voltages. This is obtained by adjustment of the incoming generator's field strength.
2. Equal frequency. This is obtained by adjustment of the incoming generator's prime-mover speed.
3. Phase voltages in proper phase relation. The procedure for synchronizing generators is not discussed in this chapter. At this point, it is enough for you to know that the above must be accomplished to prevent damage to the machines.

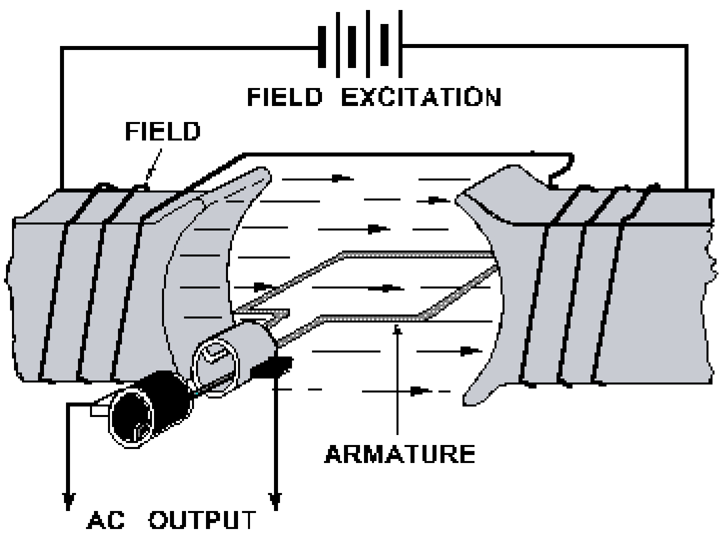
Q21. What generator characteristics must be considered when alternators are synchronized for parallel operation?

SUMMARY

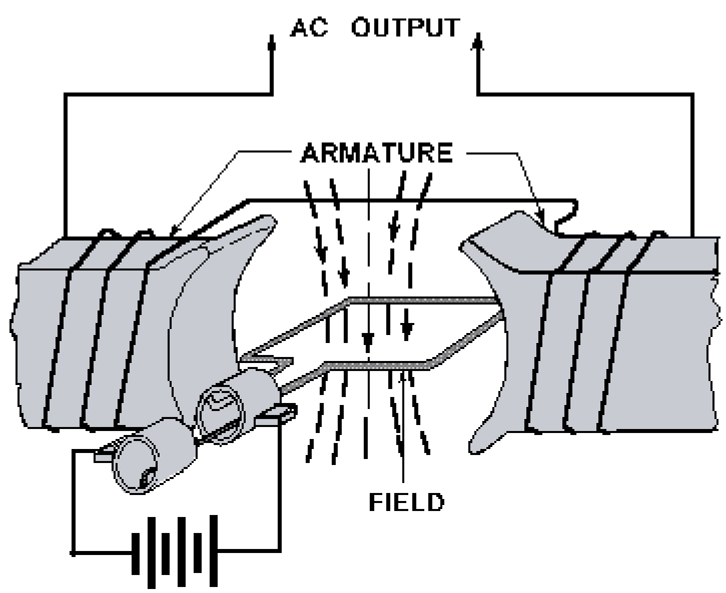
This chapter has presented an introduction to the subject of alternators. You have studied the characteristics and applications of different types. The following information provides a summary of the chapter for your review.

MAGNETIC INDUCTION is the process of inducing an emf in a coil whenever the coil is placed in a magnetic field and motion exists between the coil and the magnetic lines of flux. This is true if either the coil or the magnetic field moves, as long as the coil is caused to cut across magnetic flux lines.

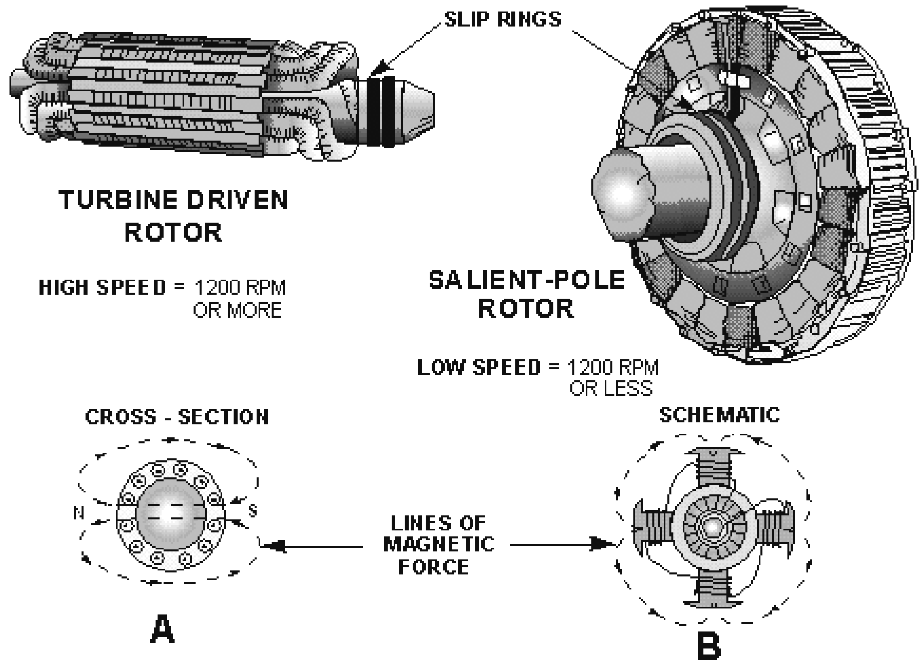
The **ROTATING ARMATURE-ALTERNATOR** is essentially a loop rotating through a stationary magnetic fealties cutting action of the loop through the magnetic field generates ac in the loop. This ac is removed from the loop by means of slip rings and applied to an external load.



The **ROTATING-FIELD ALTERNATOR** has a stationary armature and a rotating field. High voltages can be generated in the armature and applied to the load directly, without the need of slip rings and brushes. The low dc voltage is applied to the rotor field by means of slip rings, but this does not introduce any insulation problems.



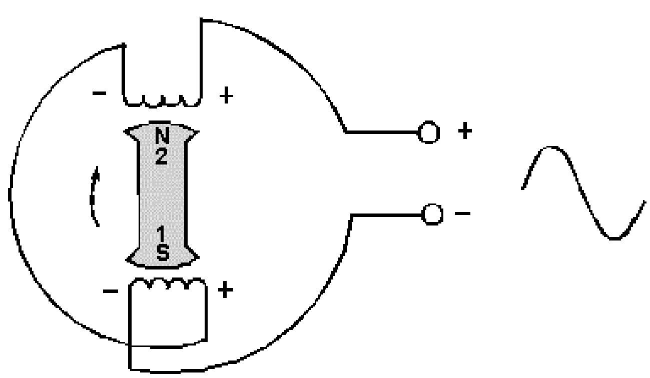
ROTOR CONSTRUCTION in alternators may be either of two types. The salient-pole rotor is used in slower speed alternators. The turbine driven-type is wound in a manner to allow high-speed use without flying apart.



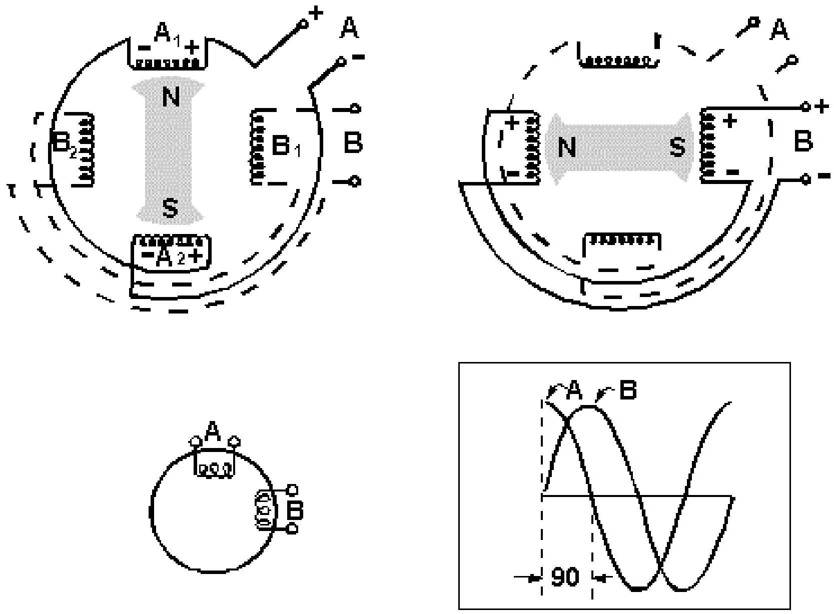
GENERATOR RATINGS are dependent on the amount of current they are capable of providing at full output voltage; this rating is expressed as the product of the voltage times the current. A 10-volt alternator capable of supplying 10 amperes of current would be rated at 100 volt-amperes. Larger alternators are rated in kilovolt-amperes.

EXCITER GENERATORS are small dc generators built into alternators to provide excitation current to field windings. These dc generators are called exciters.

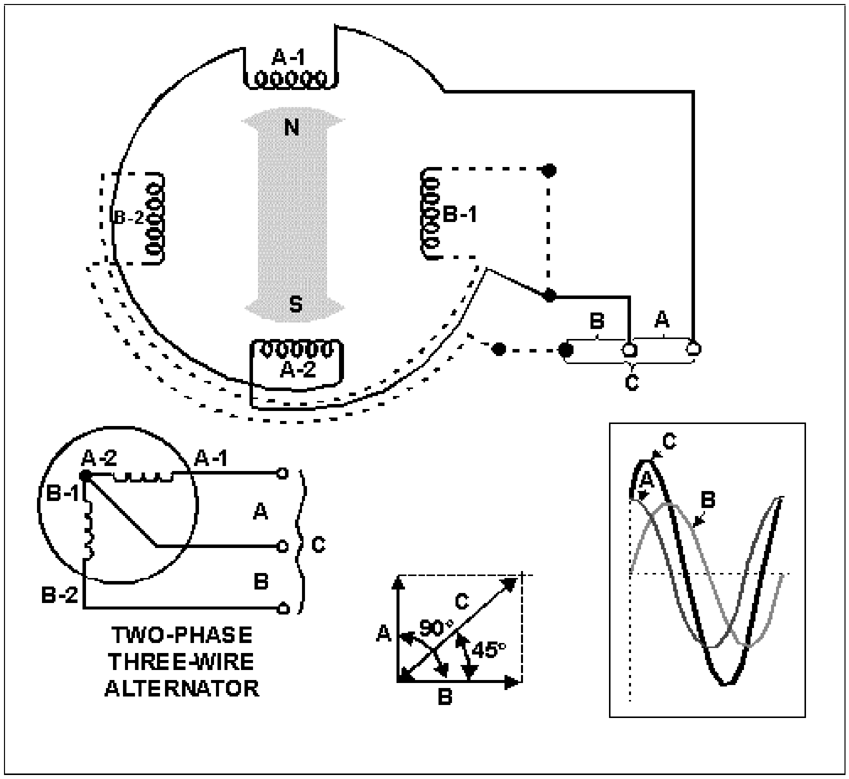
The **SINGLE-PHASE ALTERNATOR** has an armature that consists of a number of windings placed symmetrically around the stator and connected in series. The voltages generated in each winding add to produce the total voltage across the two output terminals.



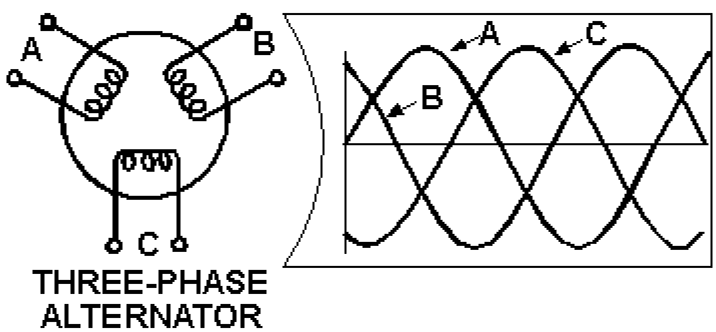
A **TWO-PHASE ALTERNATOR** consists of two phases whose windings are so placed around the stator that the voltages generated in them are 90° out of phase.



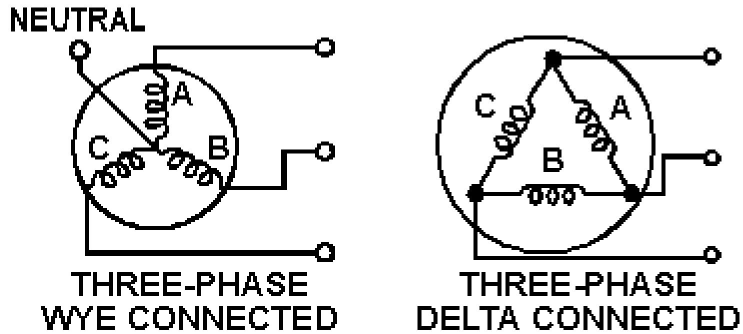
TWO-PHASE ALTERNATOR CONNECTIONS may be modified so that the output of a two-phase alternator is in a three-wire manner, which actually provides three outputs, two induced phase voltages, plus a vectorial sum voltage.



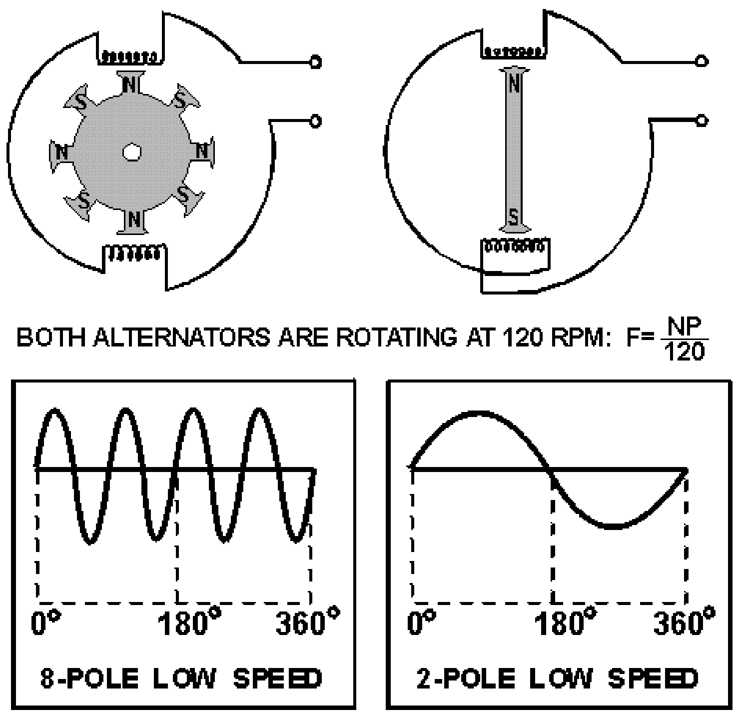
In **THREE-PHASE ALTERNATORS** the windings have voltages generated in them which are 120° out of phase. Three-phase alternators are most often used to generate ac power.



THREE-PHASE ALTERNATOR CONNECTIONS may be delta or wye connections depending on the application. The ac power aboard ship is usually taken from the ship's generators through delta connections, for the convenience of step-down transformers.



ALTERNATOR FREQUENCY depends upon the speed of rotation and the number of pairs of rotor poles.



VOLTAGE REGULATION is the change in output voltage of an alternator under varying load conditions.

VOLTAGE CONTROL in alternators is accomplished by varying the current in the field windings, much as in dc generators.

ANSWERS TO QUESTIONS Q1. THROUGH Q21.

A1. A conductor and a magnetic field.

A2. Armature.

A3. Rotating armature and rotating field.

A4. Output voltage is taken directly from the armature (not through brushes or slip rings).

A5. To provide dc current for the rotating field.

A6. Kilovolt-amperes (volt amperes).

A7. Steam turbine.

A8. Internal combustion engines, water force and electric motors.

A9. One voltage (one output).

A10. In series.

A11. Placement of armature coils.

A12. Three.

A13. C is 1.414 times greater than A or B.

A14. Each phase is displaced 120º from the other two.

A15. Wye and Delta.

A16. Three single-phase, delta-delta, step-down transformers.

A17. Speed of rotation and number of poles.

A18. 120 Hz.

A19. Voltage regulation. As a percentage.

A20. By varying the voltage applied to the field windings.

A21. Output voltage, frequency, and phase relationships.

CHAPTER 4

ALTERNATING CURRENT MOTORS

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. List three basic types of ac motors and describe the characteristics of each type.
2. Describe the characteristics of a series motor that enable it to be used as a universal motor.
3. Explain the relationships of the individual phases of multiphase voltages as they produce rotating magnetic fields in ac motors.
4. Describe the placement of stator windings in two-phase, ac motors using rotating fields.
5. List the similarities and differences between the stator windings of two-phase and three-phase ac motors.
6. State the primary application of synchronous motors, and explain the characteristics that make them suitable for that application.
7. Describe the features that make the ac induction motor the most widely used of electric motors.
8. Describe the difference between the rotating field of multiphase motors and the "apparent" rotating field of single-phase motors.
9. Explain the operation of split-phase windings in single-phase ac induction motors.
10. Describe the effects of shaded poles in single-phase, ac induction motors.

INTRODUCTION

Most of the power-generating systems, ashore and afloat, produce ac. For this reason a majority of the motors used throughout the Navy are designed to operate on ac. There are other advantages in the use of ac motors besides the wide availability of ac power. In general, ac motors cost less than dc motors. Some types of ac motors do not use brushes and commutators. This eliminates many problems of maintenance and wear. It also eliminates the problem of dangerous sparking.

An ac motor is particularly well suited for constant-speed applications. This is because its speed is determined by the frequency of the ac voltage applied to the motor terminals.

The dc motor is better suited than an ac motor for some uses, such as those that require variable-speeds. An ac motor can also be made with variable speed characteristics but only within certain limits.

Industry builds ac motors in different sizes, shapes, and ratings for many different types of jobs. These motors are designed for use with either polyphase or single-phase power systems. It is not possible here to cover all aspects of the subject of ac motors. Only the principles of the most commonly used types are dealt with in this chapter.

In this chapter, ac motors will be divided into (1) series, (2) synchronous, and (3) induction motors. Single-phase and polyphase motors will be discussed. Synchronous motors, for purposes of this chapter, may be considered as polyphase motors, of constant speed, whose rotors are energized with dc voltage. Induction motors, single-phase or polyphase, whose rotors are energized by induction, are the most commonly used ac motor. The series ac motor, in a sense, is a familiar type of motor. It is very similar to the dc motor that was covered in chapter 2 and will serve as a bridge between the old and the new.

Q1. What are the three basic types of ac motors?

SERIES AC MOTOR

A series ac motor is the same electrically as a dc series motor. Refer to figure 4-1 and use the left-hand rule for the polarity of coils. You can see that the instantaneous magnetic polarities of the armature and field oppose each other, and motor action results. Now, reverse the current by reversing the polarity of the input. Note that the field magnetic polarity still opposes the armature magnetic polarity. This is because the reversal effects both the armature and the field. The ac input causes these reversals to take place continuously.

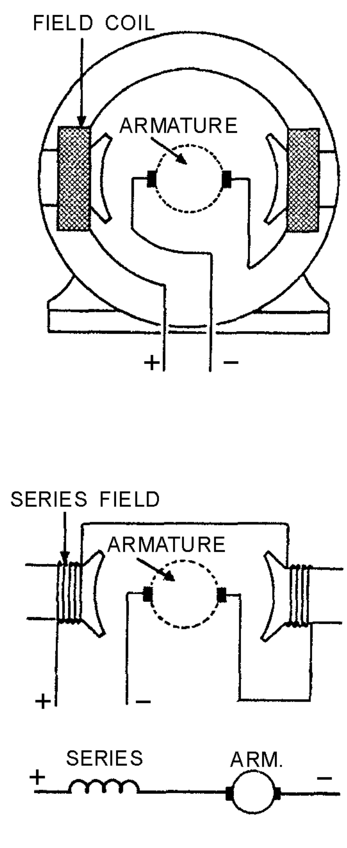


Figure 4-1.—Series ac motor.

The construction of the ac series motor differs slightly from the dc series motor. Special metals, laminations, and windings are used. They reduce losses caused by eddy currents, hysteresis, and high reactance. Dc power can be used to drive an ac series motor efficiently, but the opposite is not true.

The characteristics of a series ac motor are similar to those of a series dc motor. It is a varying-speed machine. It has low speeds for large loads and high speeds for light loads. The starting torque is very high. Series motors are used for driving fans, electric drills, and other small appliances. Since the series ac motor has the same general characteristics as the series dc motor, a series motor has been designed that can operate both on ac and dc. This ac/dc motor is called a universal motor. It finds wide use in small electric appliances. Universal motors operate at lower efficiency than either the ac or dc series motor. They are built in small sizes only. Universal motors do not operate on polyphase ac power.

Q2. Series motors are generally used to operate what type of equipment?

Q3. Why are series motors sometimes called universal motors?

ROTATING MAGNETIC FIELDS

The principle of rotating magnetic fields is the key to the operation of most ac motors. Both synchronous and induction types of motors rely on rotating magnetic fields in their stators to cause their rotors to turn.

The idea is simple. A magnetic field in a stator can be made to rotate electrically, around and around. Another magnetic field in the rotor can be made to chase it by being attracted and repelled by the stator field. Because the rotor is free to turn, it follows the rotating magnetic field in the stator. Let's see how it is done.

Rotating magnetic fields may be set up in two-phase or three-phase machines. To establish a rotating magnetic field in a motor stator, the number of pole pairs must be the same as (or a multiple of) the number of phases in the applied voltage. The poles must then be displaced from each other by an angle equal to the phase angle between the individual phases of the applied voltage.

Q4. What determines the number of field poles required to establish a rotating magnetic field in a multiphase motor stator?

TWO-PHASE ROTATING MAGNETIC FIELD

A rotating magnetic field is probably most easily seen in a two-phase stator. The stator of a two-phase induction motor is made up of two windings (or a multiple of two). They are placed at right angles to each other around the stator. The simplified drawing in figure 4-2 illustrates a two-phase stator.

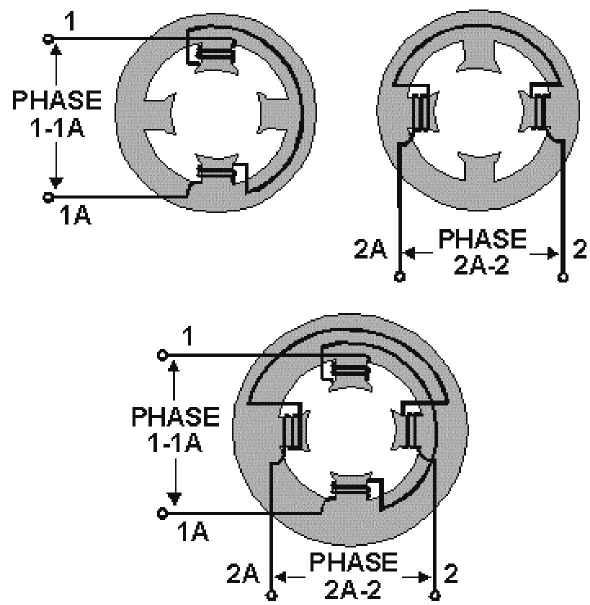


Figure 4-2.—Two-phase motor stator.

If the voltages applied to phases 1-1A and 2-2A are 90º out of phase, the currents that flow in the phases are displaced from each other by 90º . Since the magnetic fields generated in the coils are in phase with their respective currents, the magnetic fields are also 90º out of phase with each other. These two out-of-phase magnetic fields, whose coil axes are at right angles to each other, add together at every instant during their cycle. They produce a resultant field that rotates one revolution for each cycle of ac.

To analyze the rotating magnetic field in a two-phase stator, refer to figure 4-3. The arrow represents the rotor. For each point set up on the voltage chart, consider that current flows in a direction that will cause the magnetic polarity indicated at each pole piece. Note that from one point to the next, the polarities are rotating from one pole to the next in a clockwise manner. One complete cycle of input voltage produces a 360-degree rotation of the pole polarities. Let's see how this result is obtained.

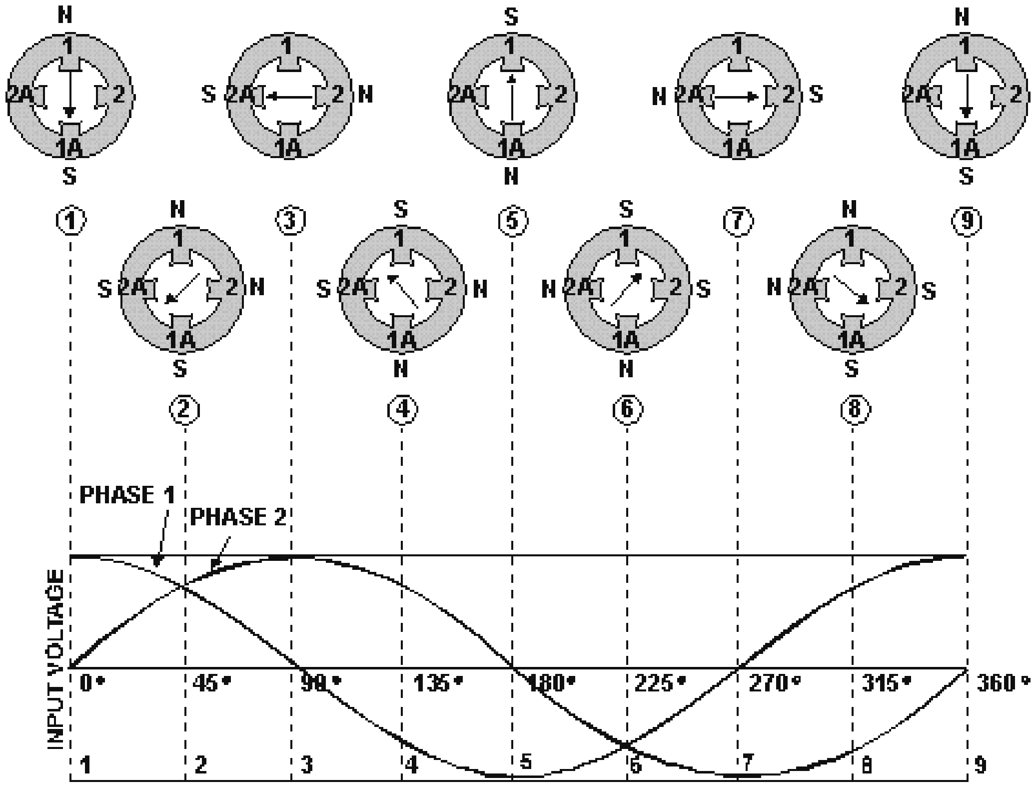


Figure 4-3.—Two-phase rotating field.

The waveforms in figure 4-3 are of the two input phases, displaced 90° because of the way they were generated in a two-phase alternator. The waveforms are numbered to match their associated phase. Although not shown in this figure, the windings for the poles 1-1A and 2-2A would be as shown in the previous figure. At position 1, the current flow and magnetic field in winding 1-1A is at maximum (because the phase voltage is maximum). The current flow and magnetic field in winding 2-2A is zero (because the phase voltage is zero). The resultant magnetic field is therefore in the direction of the 1-1A axis. At the 45-degree point (position 2), the resultant magnetic field lies midway between windings 1-1A and 2-2A. The coil currents and magnetic fields are equal in strength. At 90° (position 3), the magnetic field in winding 1-1A is zero. The magnetic field in winding 2-2A is at maximum. Now the resultant magnetic field lies along the axis of the 2-2A winding as shown. The resultant magnetic field has rotated clockwise through 90° to get from position 1 to position 3. When the two-phase voltages have completed one full cycle (position 9), the resultant magnetic field has rotated through 360° . Thus, by placing two windings at right angles to each other and exciting these windings with voltages 90° out of phase, a rotating magnetic field results.

Two-phase motors are rarely used except in special-purpose equipment. They are discussed here to aid in understanding rotating fields. You will, however, encounter many single-phase and three-phase motors.

Q5. What is the angular displacement between field poles in a two-phase motor stator?

THREE-PHASE ROTATING FIELDS

The three-phase induction motor also operates on the principle of a rotating magnetic field. The following discussion shows how the stator windings can be connected to a three-phase ac input and have a resultant magnetic field that rotates.

Figure 4-4, views A-C show the individual windings for each phase. Figure 4-4, view D, shows how the three phases are tied together in a Y-connected stator. The dot in each diagram indicates the common point of the Y-connection. You can see that the individual phase windings are equally spaced around the stator. This places the windings 120º apart.

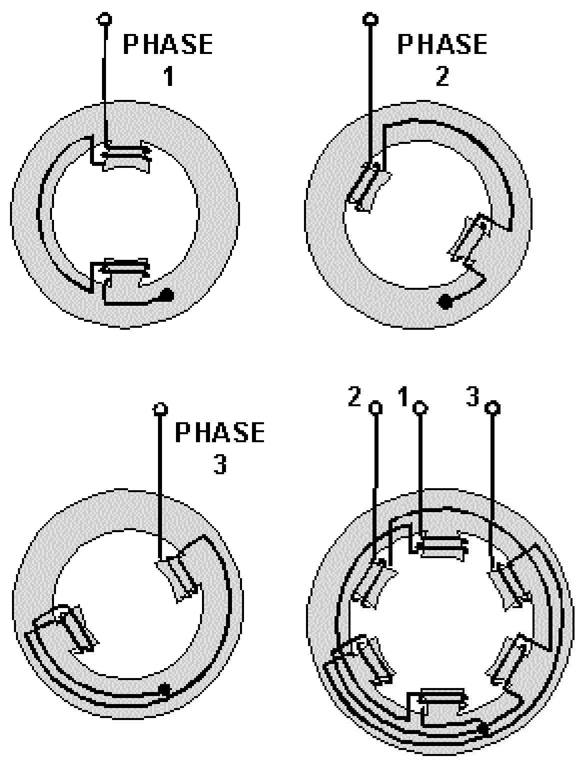


Figure 4-4.—Three-phase, Y-connected stator.

The three-phase input voltage to the stator of figure 4-4 is shown in the graph of figure 4-5. Use the left-hand rule for determining the electromagnetic polarity of the poles at any given instant. In applying the rule to the coils in figure 4-4, consider that current flows toward the terminal numbers for positive voltages, and away from the terminal numbers for negative voltages.

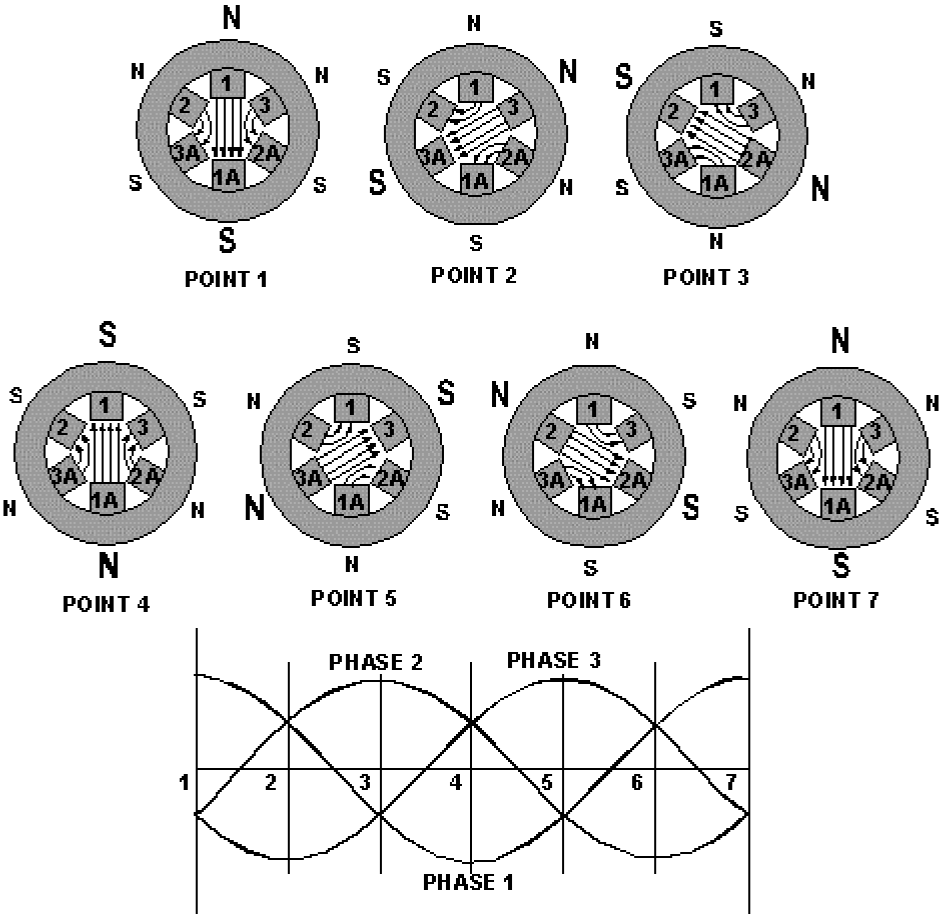


Figure 4-5.—Three-phase rotating-field polarities and input voltages.

The results of this analysis are shown for voltage points 1 through 7 in figure 4-5. At point 1, the magnetic field in coils 1-1A is maximum with polarities as shown. At the same time, negative voltages are being felt in the 2-2A and 3-3A windings. These create weaker magnetic fields, which tend to aid the 1-1A field. At point 2, maximum negative voltage is being felt in the 3-3A windings. This creates a strong magnetic field which, in turn, is aided by the weaker fields in 1-1A and 2-2A. As each point on the voltage graph is analyzed, it can be seen that the resultant magnetic field is rotating in a clockwise direction. When the three-phase voltage completes one full cycle (point 7), the magnetic field has rotated through 360º .

Q6. What is the major difference between a two-phase and a three-phase stator?

ROTOR BEHAVIOR IN A ROTATING FIELD

For purposes of explaining rotor movement, let's assume that we can place a bar magnet in the center of the stator diagrams of figure 4-5. We'll mount this magnet so that it is free to rotate in this area. Let's also assume that the bar magnet is aligned so that at point 1 its south pole is opposite the large N of the stator field.

You can see that this alignment is natural. Unlike poles attract, and the two fields are aligned so that they are attracting. Now, go from point 1 through point 7. As before, the stator field rotates clockwise. The bar magnet, free to move, will follow the stator field, because the attraction between the two fields continues to exist. A shaft running through the pivot point of the bar magnet would rotate at the same speed as the rotating field. This speed is known as synchronous speed. The shaft represents the shaft of an operating motor to which the load is attached.

Remember, this explanation is an oversimplification. It is meant to show how a rotating field can cause mechanical rotation of a shaft. Such an arrangement would work, but it is not used. There are limitations to a permanent magnet rotor. Practical motors use other methods, as we shall see in the next paragraphs.

SYNCHRONOUS MOTORS

The construction of the synchronous motors is essentially the same as the construction of the salient-pole alternator. In fact, such an alternator may be run as an ac motor. It is similar to the drawing in figure 4-6. Synchronous motors have the characteristic of constant speed between no load and full load. They are capable of correcting the low power factor of an inductive load when they are operated under certain conditions. They are often used to drive dc generators. Synchronous motors are designed in sizes up to thousands of horsepower. They may be designed as either single-phase or multiphase machines. The discussion that follows is based on a three-phase design.

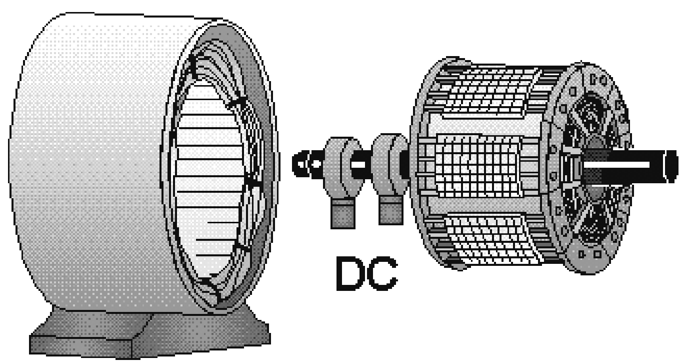


Figure 4-6.—Revolving-field synchronous motor.

To understand how the synchronous motor works, assume that the application of three-phase ac power to the stator causes a rotating magnetic field to be set up around the rotor. The rotor is energized with dc (it acts like a bar magnet). The strong rotating magnetic field attracts the strong rotor field activated by the dc. This results in a strong turning force on the rotor shaft. The rotor is therefore able to turn a load as it rotates in step with the rotating magnetic field.

It works this way once it's started. However, one of the disadvantages of a synchronous motor is that it cannot be started from a standstill by applying three-phase ac power to the stator. When ac is applied to the stator, a high-speed rotating magnetic field appears immediately. This rotating field rushes past the rotor poles so quickly that the rotor does not have a chance to get started. In effect, the rotor is repelled first in one direction and then the other. A synchronous motor in its purest form has no starting torque. It has torque only when it is running at synchronous speed.

A squirrel-cage type of winding is added to the rotor of a synchronous motor to cause it to start. The squirrel cage is shown as the outer part of the rotor in figure 4-7. It is so named because it is shaped and looks something like a turnable squirrel cage. Simply, the windings are heavy copper bars shorted together by copper rings. A low voltage is induced in these shorted windings by the rotating three-phase stator field. Because of the short circuit, a relatively large current flows in the squirrel cage. This causes a magnetic field that interacts with the rotating field of the stator. Because of the interaction, the rotor begins to turn, following the stator field; the motor starts. We will run into squirrel cages again in other applications, where they will be covered in more detail.

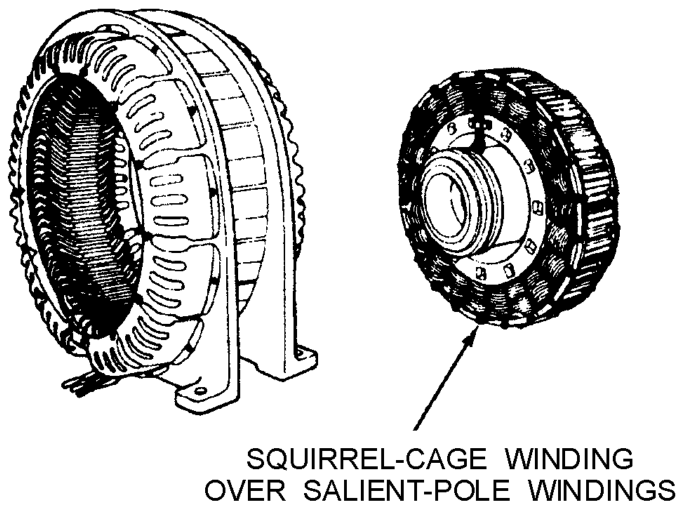


Figure 4-7.—Self-starting synchronous ac motor.

To start a practical synchronous motor, the stator is energized, but the dc supply to the rotor field is not energized. The squirrel-cage windings bring the rotor to near synchronous speed. At that point, the dc field is energized. This locks the rotor in step with the rotating stator field. Full torque is developed, and the load is driven. A mechanical switching device that operates on centrifugal force is often used to apply dc to the rotor as synchronous speed is reached.

The practical synchronous motor has the disadvantage of requiring a dc exciter voltage for the rotor. This voltage may be obtained either externally or internally, depending on the design of the motor.

Q7. What requirement is the synchronous motor specifically designed to meet?

INDUCTION MOTORS

The induction motor is the most commonly used type of ac motor. Its simple, rugged construction costs relatively little to manufacture. The induction motor has a rotor that is not connected to an external source of voltage. The induction motor derives its name from the fact that ac voltages are induced in the rotor circuit by the rotating magnetic field of the stator. In many ways, induction in this motor is similar to the induction between the primary and secondary windings of a transformer.

Large motors and permanently mounted motors that drive loads at fairly constant speed are often induction motors. Examples are found in washing machines, refrigerator compressors, bench grinders, and table saws.

The stator construction of the three-phase induction motor and the three-phase synchronous motor are almost identical. However, their rotors are completely different (see fig. 4-8). The induction rotor is made of a laminated cylinder with slots in its surface. The windings in these slots are one of two types (shown in fig. 4-9). The most common is the squirrel-cage winding. This entire winding is made up of heavy copper bars connected together at each end by a metal ring made of copper or brass. No insulation is required between the core and the bars. This is because of the very low voltages generated in the rotor bars. The other type of winding contains actual coils placed in the rotor slots. The rotor is then called a wound rotor.

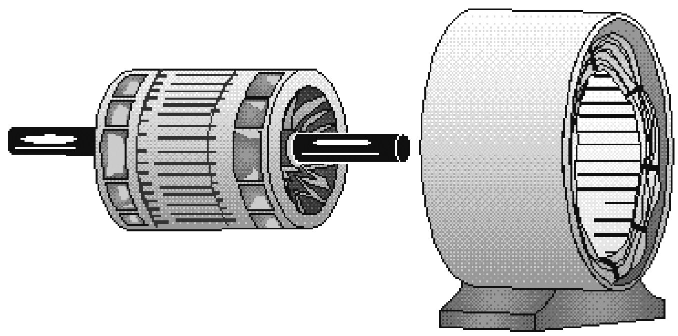


Figure 4-8.—Induction motor.

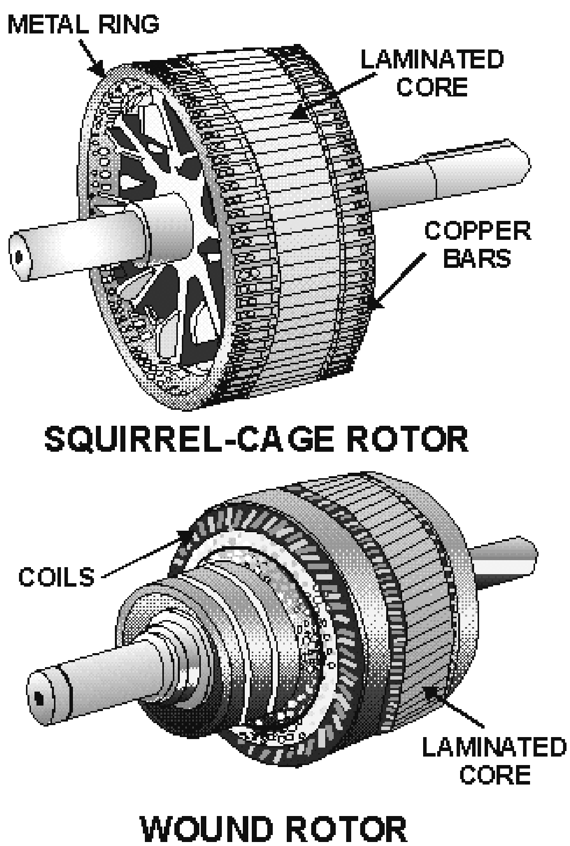


Figure 4-9.—Types of ac induction motor rotors.

Regardless of the type of rotor used, the basic principle is the same. The rotating magnetic field generated in the stator induces a magnetic field in the rotor. The two fields interact and cause the rotor to turn. To obtain maximum interaction between the fields, the air gap between the rotor and stator is very small.

As you know from Lenz's law, any induced emf tries to oppose the changing field that induces it. In the case of an induction motor, the changing field is the motion of the resultant stator field. A force is exerted on the rotor by the induced emf and the resultant magnetic field. This force tends to cancel the relative motion between the rotor and the stator field. The rotor, as a result, moves in the same direction as the rotating stator field.

It is impossible for the rotor of an induction motor to turn at the same speed as the rotating magnetic field. If the speeds were the same, there would be no relative motion between the stator and rotor fields; without relative motion there would be no induced voltage in the rotor. In order for relative motion to exist between the two, the rotor must rotate at a speed slower than that of the rotating magnetic field. The. difference between the speed of the rotating stator field and the rotor speed is called slip. The smaller the slip, the closer the rotor speed approaches the stator field speed.

The speed of the rotor depends upon the torque requirements of the load. The bigger the load, the stronger the turning force needed to rotate the rotor. The turning force can increase only if the rotor-induced emf increases. This emf can increase only if the magnetic field cuts through the rotor at a faster rate. To increase the relative speed between the field and rotor, the rotor must slow down. Therefore, for heavier loads the induction motor turns slower than for lighter loads. You can see from the previous statement that slip is directly proportional to the load on the motor. Actually only a slight change in speed is necessary to produce the usual current changes required for normal changes in load. This is because the rotor windings have such a low resistance. As a result, induction motors are called constant-speed motors.

Q8. Why is the ac induction motor used more often than other types?

Q9. The speed of the rotor is always somewhat less than the speed of the rotating field. What is the difference called?

Q10. What determines the amount of slip in an induction motor?

SINGLE-PHASE INDUCTION MOTORS

There are probably more single-phase ac induction motors in use today than the total of all the other types put together.

It is logical that the least expensive, lowest maintenance type of ac motor should be used most often. The single-phase ac induction motor fits that description.

Unlike polyphase induction motors, the stator field in the single-phase motor does not rotate. Instead it simply alternates polarity between poles as the ac voltage changes polarity.

Voltage is induced in the rotor as a result of magnetic induction, and a magnetic field is produced around the rotor. This field will always be in opposition to the stator field (Lenz's law applies). The interaction between the rotor and stator fields will not produce rotation, however. The interaction is shown by the double-ended arrow in figure 4-10, view A. Because this force is across the rotor and through the pole pieces, there is no rotary motion, just a push and/or pull along this line.

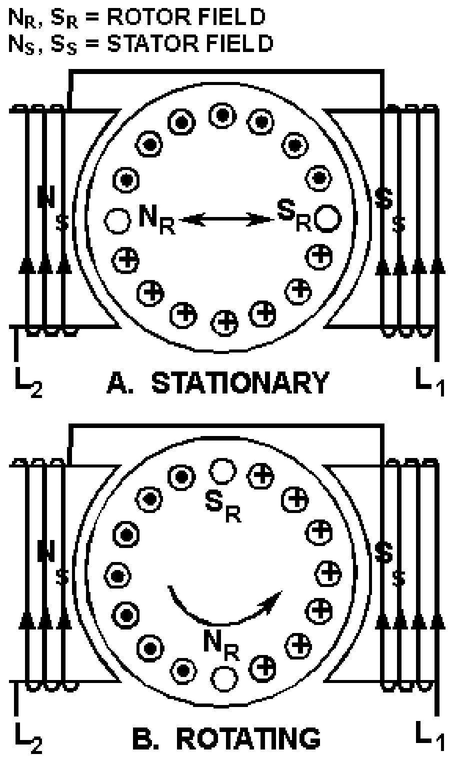


Figure 4-10.—Rotor currents in a single-phase ac induction motor.

Now, if the rotor is rotated by some outside force (a twist of your hand, or something), the push-pull along the line in figure 4-10, view A, is disturbed. Look at the fields as shown in figure 4-10, view B. At this instant the south pole on the rotor is being attracted by the left-hand pole. The north rotor pole is being attracted to the right-hand pole. All of this is a result of the rotor being rotated 90º by the outside force. The pull that now exists between the two fields becomes a rotary force, turning the rotor toward magnetic correspondence with the stator. Because the two fields continuously alternate, they will never actually line up, and the rotor will continue to turn once started. It remains for us to learn practical methods of getting the rotor to start.

There are several types of single-phase induction motors in use today. Basically they are identical except for the means of starting. In this chapter we will discuss the split-phase and shaded-pole motors; so named because of the methods employed to get them started. Once they are up to operating speed, all single-phase induction motors operate the same.

Q11. What type of ac motor is most widely used?

Split-Phase Induction Motors

One type of induction motor, which incorporates a starting device, is called a split-phase induction motor. Split-phase motors are designed to use inductance, capacitance, or resistance to develop a starting torque. The principles are those that you learned in your study of alternating current.

CAPACITOR-START.—The first type of split-phase induction motor that will be covered is the capacitor-start type. Figure 4-11 shows a simplified schematic of a typical capacitor-start motor. The stator consists of the main winding and a starting winding (auxiliary). The starting winding is connected in parallel with the main winding and is placed physically at right angles to it. A 90-degree electrical phase difference between the two windings is obtained by connecting the auxiliary winding in series with a capacitor and starting switch. When the motor is first energized, the starting switch is closed. This places the capacitor in series with the auxiliary winding. The capacitor is of such value that the auxiliary circuit is effectively a resistive-capacitive circuit (referred to as capacitive reactance and expressed as X_C). In this circuit the current leads the line voltage by about 45° (because X_C about equals R). The main winding has enough resistance-inductance (referred to as inductive reactance and expressed as X_L) to cause the current to lag the line voltage by about 45° (because X_L about equals R). The currents in each winding are therefore 90° out of phase - so are the magnetic fields that are generated. The effect is that the two windings act like a two-phase stator and produce the rotating field required to start the motor.

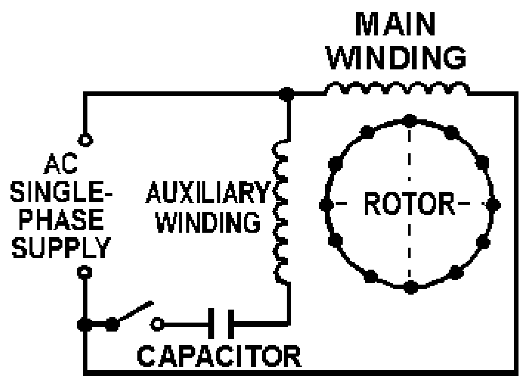


Figure 4-11.—Capacitor-start, ac induction motor.

When nearly full speed is obtained, a centrifugal device (the starting switch) cuts out the starting winding. The motor then runs as a plain single-phase induction motor. Since the auxiliary winding is only a light winding, the motor does not develop sufficient torque to start heavy loads. Split-phase motors, therefore, come only in small sizes.

RESISTANCE-START.—Another type of split-phase induction motor is the resistance-start motor. This motor also has a starting winding (shown in fig. 4-12) in addition to the main winding. It is switched in and out of the circuit just as it was in the capacitor-start motor. The starting winding is positioned at right angles to the main winding. The electrical phase shift between the currents in the two windings is obtained by making the impedance of the windings unequal. The main winding has a high inductance and a low resistance. The current, therefore, lags the voltage by a large angle. The starting winding is designed to have a fairly low inductance and a high resistance. Here the current lags the voltage by a smaller angle. For example, suppose the current in the main winding lags the voltage by 70º. The current in the auxiliary winding lags the voltage by 40º. The currents are, therefore, out of phase by 30º. The magnetic fields are out of phase by the same amount. Although the ideal angular phase difference is 90º for maximum starting torque, the 30-degree phase difference still generates a rotating field. This supplies enough torque to start the motor. When the motor comes up to speed, a speed-controlled switch disconnects the starting winding from the line, and the motor continues to run as an induction motor. The starting torque is not as great as it is in the capacitor-start.

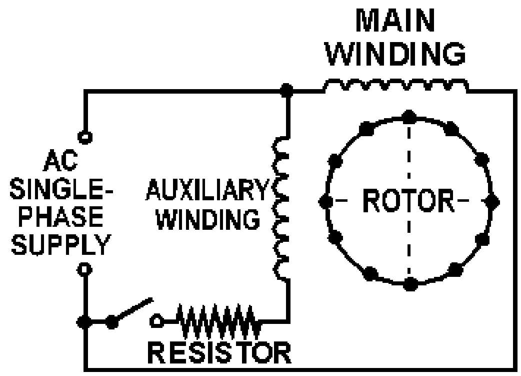


Figure 4-12.—Resistance-start ac induction motor.

Q12. How do split-phase induction motors become self-starting?

Shaded-Pole Induction Motors

The shaded-pole induction motor is another single-phase motor. It uses a unique method to start the rotor turning. The effect of a moving magnetic field is produced by constructing the stator in a special way. This motor has projecting pole pieces just like some dc motors. In addition, portions of the pole piece surfaces are surrounded by a copper strap called a shading coil. A pole piece with the strap in place is shown in figure 4-13. The strap causes the field to move back and forth across the face of the pole piece. Note the numbered sequence and points on the magnetization curve in the figure. As the alternating stator field starts increasing from zero (1), the lines of force expand across the face of the pole piece and cut through the strap. A voltage is induced in the strap. The current that results generates a field that opposes the cutting action (and decreases the strength) of the main field. This produces the following actions: As the field increases from zero to a maximum at 90º , a large portion of the magnetic lines of force are concentrated in the unshaded portion of the pole (1). At 90º the field reaches its maximum value. Since the lines of force have stopped expanding, no emf is induced in the strap, and no opposing magnetic field is generated. As a result, the main field is uniformly distributed across the pole (2). From 90º to 180º , the main field starts decreasing or collapsing inward. The field generated in the strap opposes the collapsing field. The effect is to concentrate the lines of force in the shaded portion of the pole face (3). You can see that from 0º to 180º , the main field has shifted across the pole face from the unshaded to the shaded portion. From 180º to 360º , the main field goes through the same change as it did from 0º to 180º ; however, it is now in the opposite direction (4). The direction of the field does not affect the way the shaded pole works. The motion of the field is the same during the second half-cycle as it was during the first half of the cycle.

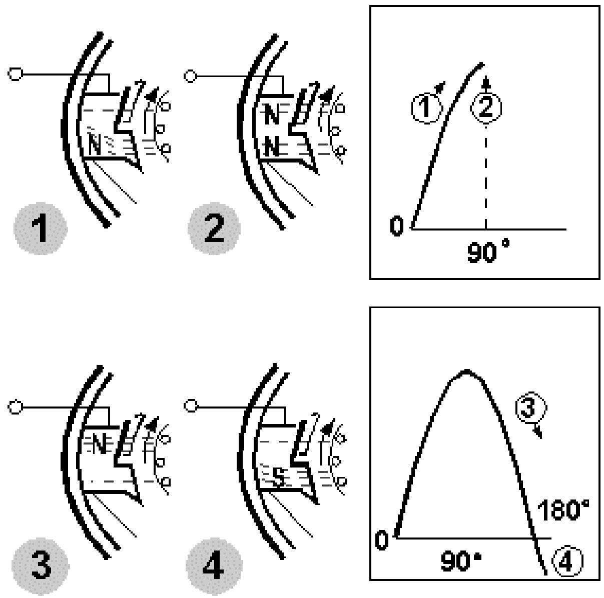


Figure 4-13.—Shaded poles as used in shaded-pole ac induction motors.

The motion of the field back and forth between shaded and unshaded portions produces a weak torque to start the motor. Because of the weak starting torque, shaded-pole motors are built only in small sizes. They drive such devices as fans, clocks, blowers, and electric razors.

Q13. Why are shaded-pole motors used to drive only very small devices?

Speed of Single-Phase Induction Motors

The speed of induction motors is dependent on motor design. The synchronous speed (the speed at which the stator field rotates) is determined by the frequency of the input ac power and the number of poles in the stator. The greater the number of poles, the slower the synchronous speed. The higher the frequency of applied voltage, the higher the synchronous speed. Remember, however, that neither frequency nor number of poles are variables. They are both fixed by the manufacturer.

The relationship between poles, frequency, and synchronous speed is as follows:

$$n\ (rpm) = \frac{120f}{p}$$

where n is the synchronous speed in rpm, f is the frequency of applied voltage in hertz, and p is the number of poles in the stator.

Let's use an example of a 4-pole motor, built to operate on 60 hertz. The synchronous speed is determined as follows:

$$n = \frac{120f}{p}$$

$$n = \frac{120 \times 60}{4}$$

$$n = 1800 \text{ rpm}$$

Common synchronous speeds for 60-hertz motors are 3600, 1800, 1200, and 900 rpm, depending on the number of poles in the original design.

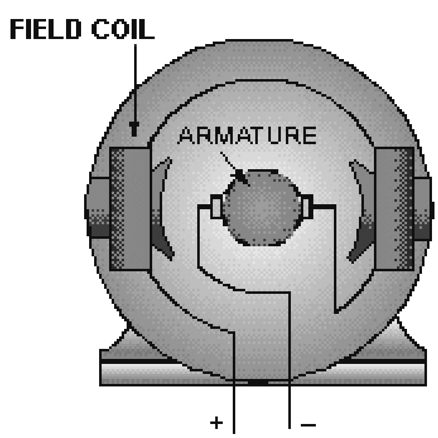
As we have seen before, the rotor is never able to reach synchronous speed. If it did, there would be no voltage induced in the rotor. No torque would be developed. The motor would not operate. The difference between rotor speed and synchronous speed is called slip. The difference between these two speeds is not great. For example, a rotor speed of 3400 to 3500 rpm can be expected from a synchronous speed of 3600 rpm.

SUMMARY

This chapter introduced you to the basic principles concerning ac motors. While many variations of types exist, the three types presented provide you with background for further study if you require more extensive knowledge of the subject. The following information provides a summary of the major subjects of this chapter for your review.

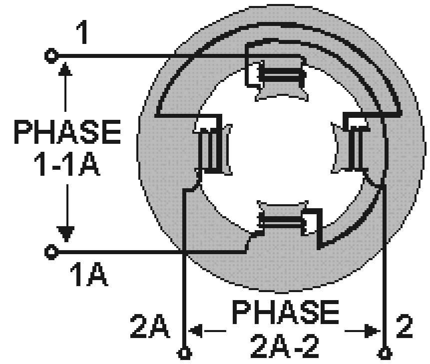
The three **AC MOTOR TYPES** presented are the series, synchronous, and induction ac motors.

AC SERIES MOTORS are nearly identical to the dc series motors. Special construction techniques allow ac series motors to be used as **UNIVERSAL MOTORS**, operating on either ac or dc power.

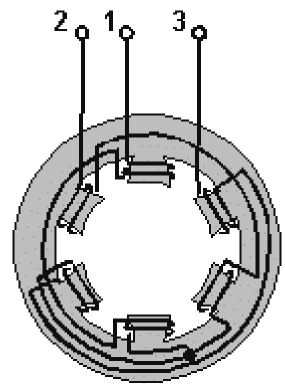


ROTATING FIELDS are developed by applying multiphase voltages to stator windings, which consist of multiple field coils. This rotating magnetic field causes the rotor to be pushed and pulled because of interaction between it and the rotor's own field.

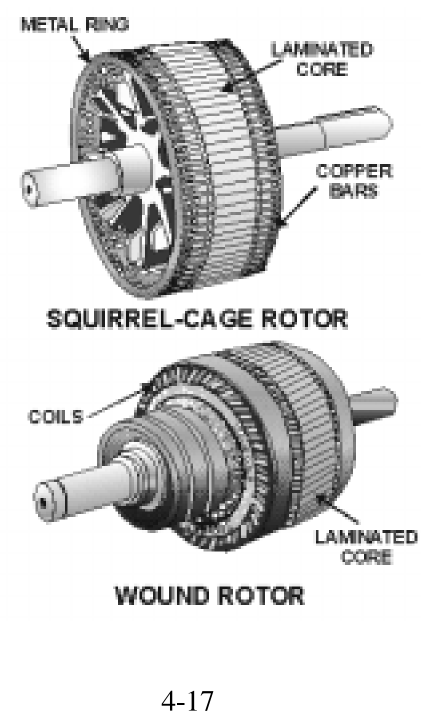
TWO-PHASE ROTATING FIELDS require two pairs of field coils displaced by 90º . They must be energized by voltages that also have a phase displacement of 90º .



THREE-PHASE ROTATING FIELDS require three pairs of windings 120º apart, energized by voltages that also have a 120-degree phase displacement.



SYNCHRONOUS MOTORS are specifically designed to maintain constant speed, with the rotor synchronous to the rotating field. Synchronous motors require modification (such as squirrel-cage windings) to be self-starting.



INDUCTION MOTORS are the most commonly used of all electric motors due to their simplicity and low cost. Induction motors may be single-phase or multiphase. They do not require electrical rotor connection. Split-phase motors with special starting windings, and shaded-pole motors, are types of single-phase induction motors.

SYNCHRONOUS SPEED is the speed of stator field rotation. It is determined by the number of poles and the frequency of the input voltage. Thus, for a given motor, synchronous speed is constant.

SLIP is the difference between actual rotor speed and the synchronous speed in induction motors. Slip must exist for there to be torque at the rotor shaft.

ANSWERS TO QUESTIONS Q1. THROUGH Q13.

A1. Series, synchronous, induction.

A2. To power small appliances.

A3. They operate on either ac or dc.

A4. The number of phases in the applied voltage.

A5. 90º .

A6. Number and location of field poles.

A7. Constant speed required by some loads.

A8. They are simple and inexpensive to make.

A9. Slip.

A10. Load.

A11. Single-phase induction motor.

A12. By using combinations of inductance and capacitance to apply out-of phase currents in starting windings.

A13. They have very weak starting torques.

APPENDIX I

GLOSSARY

AMPLIDYNE—A special dc generator in which a small dc voltage applied to field windings controls a large output voltage from the generator. In effect, an amplidyne is a rotary amplifier, oftentimes producing gain in the order of 10,000.

ARMATURE—The windings in which the output voltage is generated in a generator or in which input current creates a magnetic field that interacts with the main field in a motor. Note: Armature is often used as being identical with ROTOR. This usage is correct only part of the time. See the text and the entries under ROTOR and STATOR in this Glossary.

ARMATURE LOSSES—Copper losses, eddy current losses, hysteresis losses which act to decrease the efficiency of armatures.

ARMATURE REACTION—The effect, in a dc generator, of current in the annature creating a magnetic field that distorts the main field and causing a shift in the neutral plane.

BRUSHES—Sliding contacts, usually carbon, that make electrical connection to the rotating part of a motor or generator.

COMMUTATION—The act of a commutator in converting generator output from an ac voltage to a dc voltage.

COMMUTATOR—A mechanical device that reverses armature connections in motors and generators at the proper instant so that current continues to flow in only one direction. In effect, the commutator changes ac to dc.

COMPENSATING WINDINGS—Windings embedded in slots in pole pieces, connected in series with the armature, whose magnetic field opposes the armature field and cancels armature reaction.

COMPOUND-WOUND MOTORS AND GENERATORS—Machines that have a series field in addition to a shunt field. Such machines have characteristics of both series- and shunt-wound machines.

CAPACITOR-START MOTOR—A type of single-phase, ac induction motor in which a starting winding and a capacitor are placed in series to start the motor. The values of Xc and R are such that the main-winding and starting-winding currents are nearly 90 degrees apart, and starting torque is produced as in a two-phase motor.

COUNTER EMF—The voltage generated within a coil by a moving magnetic field cutting across the coil itself. This voltage is in opposition (counter) to the moving field that created it. Counter emf is present in every motor, generator, transformer, or other inductance winding, whenever an alternating current flows.

DELTA—A 3-phase connection in which windings are connected end-to-end, forming a closed loop that resembles the Greek letter Delta. A separate phase wire is then connected to each of the three junctions.

DRUM-TYPE ARMATURE—An efficient, popular type of armature designed so that the entire length of the winding is cutting the field at all times. Most wound armatures are of this type.

EDDY CURRENTS—Currents induced in the body of a conducting mass by a variation in magnetic flux.

FIELD—The electromagnet that furnishes the magnetic field that interacts with the armature in motors and generators.

FIELD EXCITATION—The creation of a steady magnetic field within the field windings by applying a dc voltage either from the generator itself or from an external source.

GENERATOR—A machine that converts mechanical energy to electrical energy by applying the principal of magnetic induction. A machine that produces ac or dc voltage, depending on the original design.

GRAMME-RING ARMATURE—An inefficient type of armature winding in which many of the turns are shielded from the field by its own iron ring.

INDUCTION MOTOR—A simple, rugged, ac motor with desirable characteristics. The rotor is energized by transformer action (induction) from the stator. More induction motors are used than any other type.

INTERPOLES—Small auxiliary poles placed between main field poles, whose magnetic field opposes the armature field and cancels armature reaction. Interpoles accomplish the same thing as compensating windings.

LAP WINDING—An armature winding in which opposite ends of each coil are connected to adjoining segments of the commutator so that the windings overlap.

LEFT-HAND RULE FOR GENERATORS—A representation of the relationships between motion, magnetic force, and resultant current in the generation of a voltage. The thumb, forefinger, and middle finger of the left hand are extended at right angles to each other. The thumb should point in the direction the conductor moves. The forefinger should point in the direction of magnetic flux from north to south. The middle finger will then point in the direction the generated voltage forces current to flow. Any of three quantities may be found if the other two are known.

MAGNETIC INDUCTION—The generation of a voltage in a circuit by causing relative motion between a magnetic field and the circuit. The relative motion can be the result of physical movement or the rise and fall of a magnetic field created by a changing current.

MOTOR—A machine that converts electrical energy to mechanical energy. It is activated by ac or dc voltage, depending on the design.

MOTOR LOAD—Any device driven by a motor. Typical loads are drills, saws, water pumps, rotating antennas, generators, etc. The speed and power capabilities of a motor must be matched to the speed and power requirements of the motor load.

MOTOR REACTION—The force created by generator armature current that tends to oppose normal rotation of the armature.

MOTOR STARTERS—Large resistive devices placed in series with dc motor armatures to prevent the armature from drawing excessive current until armature speed develops counter emf. The resistance is gradually removed from the circuit either automatically or manually as motor speed increases.

MULTIPHASE—See polyphase.

POLE PIECES—The shaped magnetic material upon which the stator windings of motors and generators are mounted or wound.

POLE—The sections of a field magnet where the flux lines are concentrated; also where they enter and leave the magnet.

POLYPHASE—Term that describes systems or units of a system that are activated by or which generate separate out-of-phase voltages. Typical polyphase systems are 2-phase and 3-phase whose voltages are 90- and 120-degrees out of phase, respectively. This term means the same as MULTIPHASE.

PRIME MOVER—The source of the turning force applied to the rotor of a generator. This may be an electric motor, a gasoline engine, steam turbine, etc.

ROTATING FIELD—The magnetic field in a multiphase ac motor that is the result of field windings being energized by out-of-phase voltages. In effect, the magnetic field is made to rotate electrically rather than mechanically.

ROTOR—The revolving part of a rotating electrical machine. The rotor may be either the field or the armature, depending on the design of the machine.

SELF-EXCITED GENERATORS—Dc generators in which the generator output is fed to the field to produce field excitation.

SERIES-WOUND MOTORS AND GENERATORS—Machines in which the armature and field windings are connected in series with each other.

SHUNT-WOUND MOTORS AND GENERATORS—Machines in which the armature and field windings are connected in parallel (shunt) with each other.

SLIP—The difference between rotor speed and synchronous speed in an ac induction motor. The rotor will always be slower than the synchronous speed by the amount of slip, otherwise, no voltage would be induced in the rotor.

SLIP RINGS—Contacts that are mounted on the shaft of a motor or generator to which the rotor windings are connected, and against which the brushes ride.

SQUIRREL-CAGE WINDINGS—A type of rotor winding in which heavy conductors are imbedded in the rotor body. The conductors are shorted together at the ends by continuous rings. No insulation is required between the windings and the core. This type of winding is rugged, easily manufactured, and practically maintenance free. It is widely applied in ac induction motors. Physically, it appears as a rotating squirrel-cage, thus the name.

STATOR—The stationary part of a rotating electrical machine. The stator may be either the field or the armature, depending on the design of the machine.

SYNCHRONOUS MOTOR—An ac motor whose rotor is activated by dc. It is characterized by constant speed and requires squirrel-cage windings or some other method to be self-starting.

SYNCHRONOUS SPEED—The speed at which the rotating field in an ac motor revolves. This speed is a function of the number of poles in the field and the frequency of the applied voltage.

VOLTAGE REGULATION—A measure of the ability of a generator to maintain a constant output voltage from no-load to full-load operation. Expressed as a percentage of full-load voltage, the better the regulation, the lower the percent.

WAVE WINDING—An armature winding in which the two ends of each coil are connected to commutator segments separated by the distance between poles. The winding goes successively under each main pole before reaching the starting point again.

WYE (Y)—A 3-phase connection in which one end of each phase winding is connected to a common point. Each free end is connected to a separate phase wire. The diagram of this connection often resembles the letter Y.

Assignment Questions

Information: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: "Introduction to Generators and Motors," pages 1-1 through 4-18.

1-1. In generators, what principle is used to convert mechanical motion to electrical energy?

1. Atomic reaction
2. Electrical attraction
3. Magnetic repulsion
4. Magnetic induction

1-2. When you use the left-hand rule for generators, what is indicated by the middle finger?

1. Direction of flux
2. Direction of motion
3. Direction of current flow
4. Direction of the magnetic field

1-3. The output voltage of an elementary generator is coupled from the armature to the brushes by what devices?

1. Slip rings
2. Interpoles
3. Terminals
4. Pigtails

1-4. An elementary generator consists of a single coil rotating in a magnetic field. Why is NO voltage induced in the coil as it passes through the neutral plane?

1. Flux lines are too dense
2. Flux lines are not being cut
3. Flux lines are not present
4. Flux lines are being cut in the wrong direction

1-5. What components cause(s) a generator to produce a dc voltage instead of an ac voltage at its output?

1. The brushes
2. The armature
3. The slip rings
4. The commutator

1-6. When two adjacent segments of the commutator on a single-loop dc generator come in contact with the brush at the same time, which of the following conditions will occur?

1. The output voltage will be zero
2. The output voltage will be maximum negative
3. The output voltage will be maximum positive

1-7. In an elementary, single-coil, dc generator with one pair of poles, what is the maximum number of pulsations produced in one revolution?

1. One
2. Two
3. Three
4. Four

1-8. If an elementary dc generator has a two-coil armature and four field poles, what is the total number of segments required in the commutator?

1. 8
2. 2
3. 16
4. 4

1-9. How can you vary the strength of the magnetic field in a dc generator?

1. By varying the armature current
2. By varying the speed of armature rotation
3. By varying the voltage applied to the electromagnetic field coils
4. By varying the polarity of the field poles

1-10. Under which of the following conditions does sparking occur between the brushes and the commutator?

1. When operating under normal conditions
2. When there is improper commutation
3. When there is an excessive load current
4. When commutation is in the neutral plane

1-11. Distortion of the main field by interaction with the armature field defines what term?

1. Commutation
2. Mutual reaction
3. Armature reaction
4. Mutual induction

1-12. Distortion of the main field by interaction with the armature field can be compensated for by the use of

1. slip rings
2. interpoles
3. a commutator
4. special brushes

1-13. Motor reaction in a dc generator is a physical force caused by the magnetic interaction between the armature and the field. What effect, if any, does this force have on the operation of the generator?

1. It tends to oppose the rotation of the armature
2. It tends to aid the rotation of the armature
3. It causes the generator to vibrate
4. None

1-14. In dc generators, copper losses are caused by which of the following factors?

1. Reluctance in the field poles
2. Resistance in the armature winding
3. Reactance in the armature and field windings
4. All of the above

1-15. Eddy currents in armature cores are kept low by which of the following actions?

1. Using powdered iron as a core material
2. Limiting armature current
3. Insulating the core
4. Laminating the iron in the core

1-16. What makes the drum-type armature more efficient than the Gramme-ring armature?

1. The drum-type armature has more windings than the Gramme-ring armature
2. The drum-type armature can be rotated faster than the Gramme-ring armature
3. The drum-type armature coils are fully exposed to the magnetic field, while the Gramme-ring armature coils are only partially exposed to the magnetic field
4. The drum-type armature has a laminated core, while the Gramme-ring armature has a solid core

1-17. What type of dc generator application best utilizes the features of the lap-wound armature?

1. High-voltage
2. High-current
3. High-speed
4. Variable-speed

1-18. Which of the following is NOT a major classification of dc generators?

1. Compound-wound
2. Series-wound
3. Shunt-wound
4. Lap-wound

1-19. What characteristic of series-wound generators makes them unsuitable for most applications?

1. They require external field excitation
2. The output voltage varies as the speed varies
3. They are not capable of supplying heavy loads
4. The output voltage varies as the load current varies

1-20. As the load current of a dc generator varies from no-load to full-load, the variation in output voltage is expressed as a percent of the full-load voltage. What term applies to this expression?

1. Gain
2. Voltage control
3. Voltage regulation
4. Load limit

1-21. When two or more generators are used to supply a common load, what term is applied to this method of operation?

1. Series
2. Compound
3. Split-load
4. Parallel

1-22. What special-purpose dc generator is used as a high-gain power amplifier?

1. Lap-wound
2. Shunt-wound
3. Amplidyne
4. Compound-connected

1-23. The gain of an amplifying device can be determined by which of the following formulas?

1. GAIN = INPUT + OUTPUT
2. GAIN = INPUT × OUTPUT
3. GAIN = OUTPUT − INPUT
4. GAIN = OUTPUT ÷ INPUT

1-24. The maximum gain possible from an amplidyne is approximately

1. 100
2. 5,000
3. 10,000
4. 50,000

1-25. What determines the direction of rotation of a dc motor?

1. The type of armature
2. The method of excitation
3. The number of armature coils
4. The polarity of armature current and direction of magnetic flux

1-26. When you use the right-hand rule for motors, what quantity is indicated by the extended forefinger?

1. Direction of flux north to south
2. Direction of flux south to north
3. Direction of current
4. Direction of motion

1-27. Which, if any, of the following situations is a major electrical difference between a dc motor and a dc generator?

1. The armatures are different
2. The shunt connections are different
3. The dc generator requires a commutator, the dc motor does not
4. None of the above

1-28. In a dc motor, what causes counter emf?

1. Improper commutation
2. Armature reaction
3. Generator action
4. Excessive speed

1-29. In a dc motor, how, if at all, does counter emf affect speed?

1. It causes the speed to increase
2. It causes the speed to decrease
3. It causes rapid fluctuations of the speed
4. It does not affect speed

1-30. What is the load on a dc motor?

1. The field current
2. The armature current
3. The mechanical device the motor moves
4. The total current drawn from the source

1-31. When a series dc motor is operated without a load, which of the following conditions occurs?

1. The armature draws excessive current
2. The voltage requirement increases
3. The armature will not turn
4. The armature speeds out of control

1-32. A dc series motor is best suited for which of the following applications?

1. Steady load, low torque
2. Variable load, low torque
3. Steady load, high torque
4. Variable load, high torque

1-33. What is the main advantage of a shunt motor over a series motor?

1. A shunt motor develops higher torque at lower speeds than a series motor
2. A shunt motor can be operated at higher speeds than a series motor
3. A shunt motor draws less current from the source than a series motor
4. A shunt motor maintains a more constant speed under varying load conditions than a series motor

1-34. How can the direction of rotation be changed in a dc motor?

1. Only by reversing the field connections
2. Only by reversing the armature connections
3. By reversing both the armature connections and the field connections
4. 4.By reversing either the armature connections or the field connections

1-35. When the voltage applied to the armature of a dc shunt motor is decreased, what happens to the motor speed?

1. It becomes uncontrollable
2. It decreases
3. It increases
4. The motor stops

1-36. In a dc motor, the neutral plane shifts in what direction as the result of armature reaction?

1. Clockwise
2. Counterclockwise
3. In the direction of rotation
4. Opposite the direction of rotation

1-37. The current in the interpoles of a dc motor is the same as the

1. armature current
2. field current
3. total load current
4. eddy current

1-38. In a dc motor, what is the purpose of the resistor placed in series with the armature?

1. To counteract armature reaction
2. To limit armature current
3. To increase field strength
4. To prevent overspeeding

1-39. Magnetic induction in an alternator is a result of relative motion between what two elements?

1. The rotor and the armature
2. The armature and the field
3. The field and the stator
4. The rotor and the field

1-40. Voltage is induced in what part of an alternator?

1. The commutator
2. The brushes
3. The armature
4. The field

1-41. What are the two basic types of alternators?

1. Multiphase and polyphase
2. Alternating current and direct current
3. Rotating field and rotating armature
4. Series-wound and shunt-wound

1-42. Which of the following alternator types is most widely used?

1. Shunt-wound
2. Rotating-armature
3. Series-wound
4. Rotating-field

1-43. The purpose of the exciter in an alternator is to

1. provide dc field excitation
2. compensate for armature losses
3. compensate for counter emf
4. counteract armature reaction

1-44. An alternator using a gas turbine as a prime mover should have what type of rotor?

1. Turbine-driven
2. Salient-pole
3. Armature
4. Geared

1-45. In alternators with low-speed prime movers, only what type of rotor may be used?

1. Geared
2. Armature
3. Salient-pole
4. Turbine-driven

1-46. Alternators are rated using which of the following terms?

1. Volts
2. Watts
3. Amperes
4. Volt-amperes

1-47. What does the term single-phase mean relative to single-phase alternators?

1. All output voltages are in phase with each other
2. The voltage and current are in phase
3. The phase angle is constant
4. Only one voltage is produced

1-48. In a single-phase alternator with multiple armature windings, how must the windings be connected?

1. Series
2. Parallel
3. Wye
4. Delta

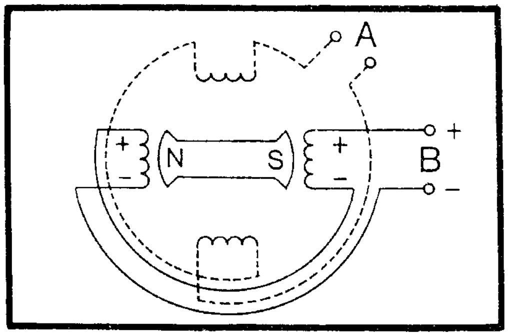


Figure 1A.—Two-phase alternator.

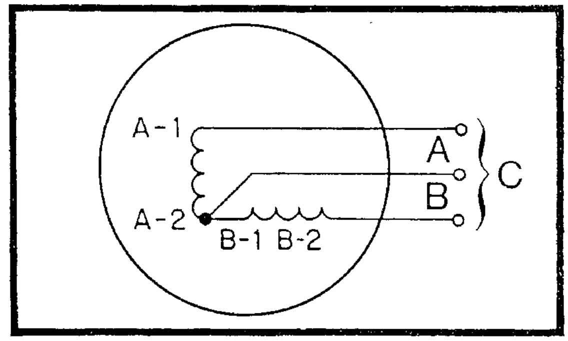


Figure 1B.—Connections for two-phase, three-wire alternator output.

IN ANSWERING QUESTION 1-49, REFER TO FIGURE 1A.

1-49. What is the phase relationship between voltages A and B?

1. In phase
2. 45° out of phase
3. 90° out of phase
4. 180° out of phase

1-50. A two-phase, three-wire alternator has what maximum number of output voltages available?

1. One
2. Two
3. Three
4. Four

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTION 1-51, REFER TO FIGURE 1B.

1-51. What is the relative amplitude of the voltage at output C as compared to A and B?

1. C is .707 times A or B
2. C is equal to the difference between and B
3. C is 1.414 times A or B
4. C is twice the sum of A and B

1-52. What determines the phase relationship between the individual output voltages in a multiphase alternator?

1. The speed of rotation
2. The number of field poles
3. The method of connecting the terminals
4. The placement of the armature coils

1-53. What is the phase relationship between the output voltages of a three-phase alternator?

1. In phase
2. 60° out of phase
3. 90° out of phase
4. 120° out of phase

1-54. The ac power aboard ship is usually distributed as what voltage?

1. 115-volt, three-phase
2. 115-volt, single-phase
3. 230-volt, single-phase
4. 450-volt, three-phase

1-55. The output frequency of an alternator is determined by what two factors?

1. The number of poles and the number of phases
2. The number of poles and the speed of rotation
3. The speed of rotation and the volt-ampere rating
4. The number of phases and the volt-ampere rating

1-56. A four-pole, single-phase alternator rotating at 1800 RPM rpm will produce what output frequency?

1. 60 Hz
2. 400 Hz
3. 1800 Hz
4. 3600 Hz

1-57. Which of the following is the correct formula for determining the percent of regulation of an alternator?

1. $\frac{E_{NL} - E_{FL}}{E_{FL}} \times 100 = \%$
2. $\frac{E_{NL} \times E_{FL}}{100} = \%$
3. $E_{NL} - E_{FL} \times 100 = \%$
4. $\frac{E_{NL}}{100} \times E_{FL} = \%$

1-58. In most alternators, the output voltage is controlled by adjusting the

1. rotor speed
2. field voltage
3. armature resistance
4. electric load

1-59. When alternators are to be operated in parallel, which of the following alternator characteristics must be considered?

1. Voltage
2. Frequency
3. Phase relationship
4. All the above

1-60. Which of the following motors is/are types of ac motor?

1. Series
2. Synchronous
3. Induction
4. All of the above

1-61. Which of the following types of motors is widely used to power small appliances?

1. Universal
2. Synchronous
3. Polyphase
4. Compound

1-62. A universal motor is a special type of

1. synchronous motor
2. series motor
3. parallel motor
4. polyphase motor

1-63. The number of pole pairs required to establish a rotating magnetic field in a multiphase motor stator is determined by which of the following factors?

1. The magnitude of the voltage
2. The magnitude of the current
3. The number of phases
4. The size of the motor

1-64. In a two-phase motor stator, what is the angular displacement between the field poles?

1. 0º
2. 90º
3. 180º
4. 360º

1-65. Adjacent phase windings of a 3-phase motor stator are what total number of degrees apart?

1. 30º
2. 90º
3. 120º
4. 180º

1-66. Which of the following types of motors has a constant speed from no load to full load?

1. Series
2. Synchronous
3. Induction
4. Universal

1-67. What type of ac motor is the simplest and least expensive to manufacture?

1. Induction
2. Series
3. Synchronous
4. Two-phase

1-68. What term applies to the difference between the speed of the rotating stator field and the rotor speed?

1. Slip
2. Synchronous
3. Rotor error
4. Torque

1-69. The speed of the rotor of an induction motor depends upon which of the following factors?

1. The method of connecting the load
2. The dc voltage applied to the rotor
3. The torque requirements of the load
4. The current in the rotor

1-70. What type of ac motor is most widely used?

1. Series
2. Universal
3. Synchronous
4. Single-phase induction

1-71. What type of ac motor uses a combination of inductance and capacitance to apply out-of-phase currents to the start windings?

1. Three-phase
2. Series
3. Synchronous
4. Split-phase induction

1-72. Why are shaded-pole motors built only in small sizes?

1. They have weak starting torque
2. They are expensive in large sizes
3. They are unidirectional
4. They require large starting current

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 6—Introduction to Electronic Emission, Tubes, and Power Supplies

NAVEDTRA 14178

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Sailor's Creed

"I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country's Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all."

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: This course introduces the student to Electronic Emissions, Tubes, and Power Supplies. It provides a background for accomplishing daily work and/or preparing for further study.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and the occupational standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards*, NAVPERS 18068, found on line at https://buperscd.technology.navy.mil /bup_updt/upd_CD/BUPERS/enlistedManOpen.htm.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
ETC Allen F. Carney

Reviewed for accuracy by ETC Scott Collie
March 2003
Corrections were made to the Assignments

NAVSUP Logistics Tracking Number
0504-LP-026-8310

TABLE OF CONTENTS

ASSIGNMENT QUESTIONS follow Appendix I.

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

you may submit your answers as soon as you complete an assignment, and
you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 5 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

CHAPTER 1

INTRODUCTION TO ELECTRON TUBES

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the OCC/ECC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completion of this chapter, you will be able to:

1. State the principle of thermionic emission and the Edison Effect and give the reasons for electron movement in vacuum tubes.
2. Identify the schematic representation for the various electron tubes and their elements.
3. Explain how the diode, triode, tetrode, and pentode electron tubes are constructed, the purpose of the various elements of the tube, and the theory of operation associated with each tube.
4. State the advantages, disadvantages, and limitations of the various types of electron tubes.
5. Describe amplification in the electron tube, the classes of amplification, and how amplification is obtained.
6. Explain biasing and the effect of bias in the electron tube circuit.
7. Describe the effects the physical structure of a tube has on electron tube operation and name the four most important tube constants that affect efficient tube operation.
8. Describe, through the use of a characteristic curve, the operating parameters of the electron tube.

INTRODUCTION TO ELECTRON TUBES

In previous study you have learned that current flows in the conductor of a completed circuit when a voltage is present. You learned that current and voltage always obey certain laws. In electronics, the laws still apply. You will use them continuously in working with electronic circuits.

One basic difference in electronic circuits that will at first seem to violate the basic laws is that electrons flow across a gap, a break in the circuit in which there appears to be no conductor. A large part of the field of electronics and the entire field of electron tubes are concerned with the flow and control of these electrons "across the gap." The following paragraphs will explain this interesting phenomenon.

THERMIONIC EMISSION

You will remember that metallic conductors contain many free electrons, which at any given instant are not bound to atoms. These free electrons are in continuous motion. The higher the temperature of the conductor, the more agitated are the free electrons, and the faster they move. A temperature can be reached where some of the free electrons become so agitated that they actually escape from the conductor. They "boil" from the conductor's surface. The process is similar to steam leaving the surface of boiling water.

Heating a conductor to a temperature sufficiently high causing the conductor to give off electrons is called **THERMIONIC EMISSION**. The idea of electrons leaving the surface is shown in figure 1-1.

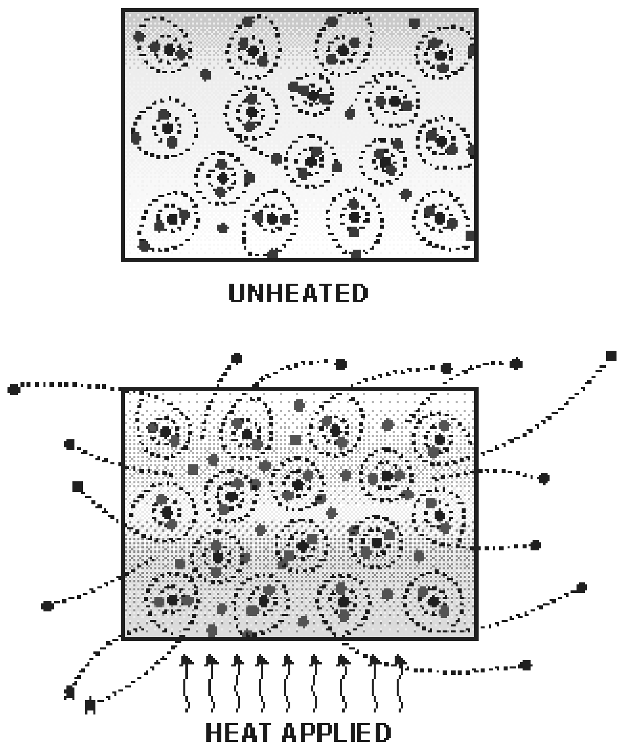


Figure 1-1.—Thermionic emission.

Thomas Edison discovered the principle of thermionic emission as he looked for ways to keep soot from clouding his incandescent light bulb. Edison placed a metal plate inside his bulb along with the normal filament. He left a gap, a space, between the filament and the plate. He then placed a battery in series between the plate and the filament, with the positive side toward the plate and the negative side toward the filament. This circuit is shown in figure 1-2.

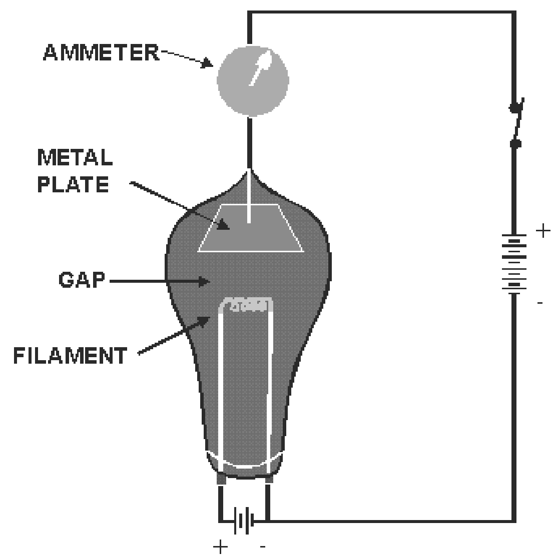


Figure 1-2.—Edison's experimental circuit.

When Edison connected the filament battery and allowed the filament to heat until it glowed, he discovered that the ammeter in the filament-plate circuit had deflected and remained deflected. He reasoned that an electrical current must be flowing in the circuit—**EVEN ACROSS THE GAP** between the filament and plate.

Edison could not explain exactly what was happening. At that time, he probably knew less about what makes up an electric circuit than you do now. Because it did not eliminate the soot problem, he did little with this discovery. However, he did patent the incandescent light bulb and made it available to the scientific community.

Let's analyze the circuit in figure 1-2. You probably already have a good idea of how the circuit works. The heated filament causes electrons to boil from its surface. The battery in the filament-plate circuit places a **POSITIVE** charge on the plate (because the plate is connected to the positive side of the battery). The electrons (negative charge) that boil from the filament are attracted to the positively charged plate. They continue through the ammeter, the battery, and back to the filament. You can see that electron flow across the space between filament and plate is actually an application of a basic law you already know—**UNLIKE CHARGES ATTRACT**.

Remember, Edison's bulb had a vacuum so the filament would glow without burning. Also, the space between the filament and plate was relatively small. The electrons emitted from the filament did not have far to go to reach the plate. Thus, the positive charge on the plate was able to attract the negative electrons.

The key to this explanation is that the electrons were floating free of the hot filament. It would have taken hundreds of volts, probably, to move electrons across the space if they had to be forcibly pulled from a cold filament. Such an action would destroy the filament and the flow would cease.

The application of thermionic emission that Edison made in causing electrons to flow across the space between the filament and the plate has become known as the **EDISON EFFECT**. It is fairly simple and extremely important. Practically everything that follows will be related in some way to the Edison effect. Be sure you have a good understanding of it before you go on.

Q1. How can a sheet of copper be made to emit electrons thermionically?

Q2. Why do electrons cross the gap in a vacuum tube?

THE DIODE TUBE

The diode vacuum tube we are about to study is really Edison's old incandescent bulb with the plate in it. Diode means two elements or two electrodes, and refers to the two parts within the glass container that make up the tube. We have called them *filament* and *plate*. More formally, they are called **CATHODE** and **PLATE**, respectively. Sometimes the filament is called a **HEATER**, for obvious reasons-more on this later.

Within a few years after the discovery of the Edison effect, scientists had learned a great deal more than Edison knew at the time of his discovery. By the early 1900s, J.J. Thomson in England had discovered the electron. Marconi, in Italy and England, had demonstrated the wireless, which was to become the radio. The theoretical knowledge of the nature of electricity and things electrical was increasing at a rapid rate.

J.A. Fleming, an English scientist, was trying to improve on Marconi's relatively crude wireless receiver when his mind went back to Edison's earlier work. His subsequent experiments resulted in what became known as the **FLEMING VALVE** (the diode), the first major step on the way to electronics.

OPERATION OF THE DIODE TUBE

Before learning about Fleming's valve, the forerunner of the modern diode, let's look at Edison's original circuit. This time, however, we'll draw it as a schematic diagram, using the symbol for a diode instead of a cartoon-like picture. The schematic is shown in figure 1-3.

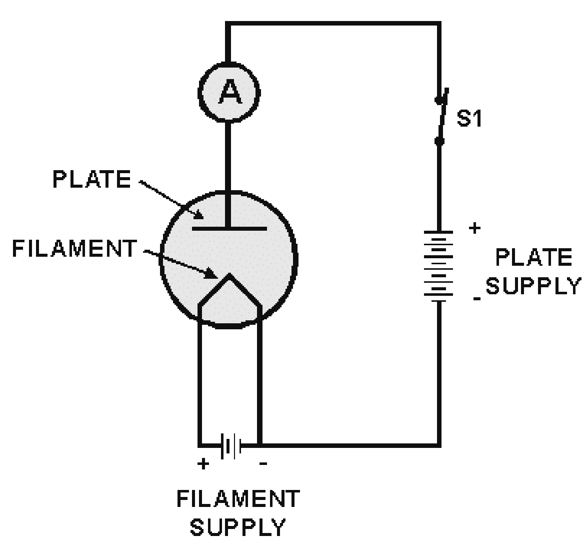


Figure 1-3.—Schematic of Edison's experimental circuit.

Note that this is really two series circuits. The filament battery and the filament itself form a series circuit. This circuit is known as the *filament circuit*.

The path of the second series circuit is from one side of the filament, across the space to the plate, through the ammeter and battery, then back to the filament. This circuit is known as the plate circuit.

You will note that a part of the filament circuit is also common to the plate circuit. This part enables the electrons boiled from the filament to return to the filament. No electron could flow anywhere if this return path were not completed. The electron flow measured by the ammeter is known as plate current.

The voltage applied between the filament and plate is known as plate voltage. You will become familiar with these terms and with others that are commonly used with diodes and diode circuits as we progress.

Diode Operation with a Positive Plate

Fleming started with a two-element tube (diode) similar to Edison's and at first duplicated Edison's experiment. The results are worth repeating here. Look at figure 1-3 again.

With the plate **POSITIVE** relative to the filament, the filament hot, and the circuit completed as shown, the ammeter detected a current flowing in the plate circuit. Because current is the same in all parts of a series circuit, we know that the same current must flow across the space between filament and plate. We know now that the electrons boiled from the heated filament are **NEGATIVE** and are attracted to the **POSITIVE** plate because **UNLIKE CHARGES ATTRACT**.

Diode Operation with a Negative Plate

Fleming's next step was to use a similar circuit but to reverse the plate battery. The circuit is shown in figure 1-4.

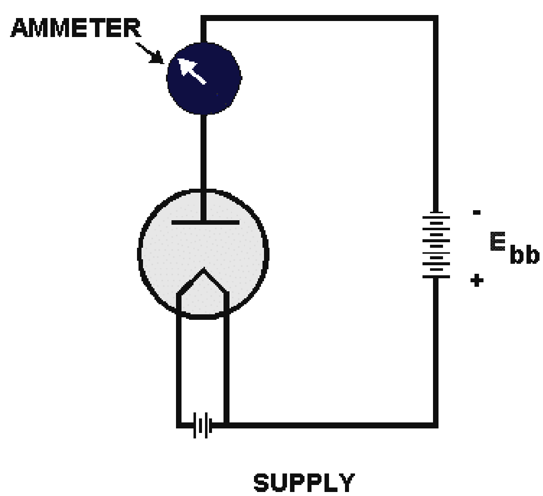


Figure 1-4.—Diode with a negative plate.

With the plate **NEGATIVE** relative to the filament, the filament hot, and the circuit completed as shown, the ammeter indicated that **ZERO** current was flowing in the plate circuit.

Fleming found that the **NEGATIVE** charge on the plate, relative to the filament, **CUT OFF** the flow of plate current as effectively as if a **VALVE** were used to stop the flow of water in a pipe.

You have all of the facts available that Fleming had. Can you give an explanation of why the diode cuts off current when the plate is negative?

Let's put the facts together. The filament is hot and electrons boil from its surface. Because the filament is the only heated element in the diode, it is the **ONLY** source of electrons within the space between filament and plate. However, because the plate is **NEGATIVE** and the electrons are **NEGATIVE**, the electrons are repelled back to the filament. Remember that **LIKE CHARGES REPEL**. If electrons cannot flow across the space, then no electrons can flow anywhere in the plate circuit. The ammeter therefore indicates **ZERO**.

It might seem to you that electrons flow from the negative plate to the positive filament under these conditions. This is **NOT** the case. Remember that it takes a heated element to emit electrons and that the filament is the only heated element in the diode. The plate is cold. Therefore, electrons cannot leave the plate, and plate-to-filament current cannot exist.

The following is a summary of diode operation as we have covered it to this point:

Assume that all parts of the circuit are operable and connected.

- PLATE CURRENT FLOWS WHEN THE PLATE IS POSITIVE.
- PLATE CURRENT IS CUT OFF WHEN THE PLATE IS NEGATIVE.
- PLATE CURRENT FLOWS ONLY IN ONE DIRECTION-FROM THE FILAMENT TO THE PLATE.

Measuring Diode Voltages

As you know, it is impossible to have a voltage at one point, because voltage is defined as a **DIFFERENCE** of **POTENTIAL** between two points. In our explanation above we referred to plate voltage. To be exactly right, we should refer to plate voltage as the **VOLTAGE BETWEEN PLATE** and **FILAMENT**. Plate voltages, and others that you will learn about soon, are often referred to as if they appear at one point. This should not confuse you if you remember your definition of voltage and realize that voltage is always measured between two points. M1 and M2 in figure 1-5 measure plate voltage and filament voltage, respectively.

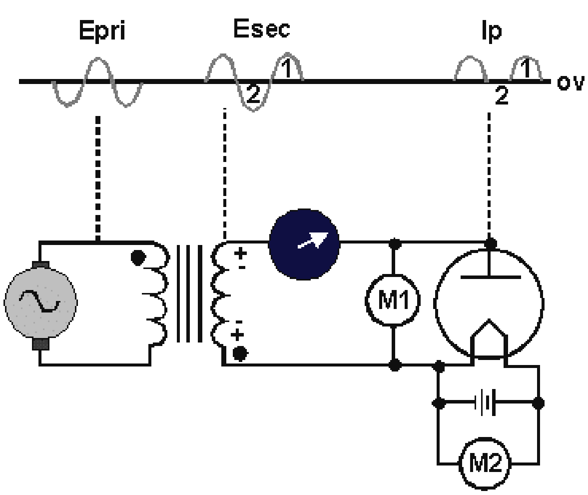


Figure 1-5.—Alternating voltage on the plate.

The reference point in diode and other tube circuits is usually a common point between the individual circuits within the tube. The reference point (common) in figure 1-5 is the conductor between the bottom of the transformer secondary and the negative side of the filament battery. Note that one side of each voltmeter is connected to this point.

Q3. Name the two series circuits that exist in a diode circuit.

Q4. Before a diode will conduct, the cathode must be what polarity relative to the plate?

Diode Operation with an Alternating Voltage on the Plate

After experimenting with a positive plate and a negative plate, Fleming replaced the direct voltage of the battery with an alternating voltage. In our explanation, we'll use a transformer as the source of alternating voltage. The circuit is shown in figure 1-5.

Note that the only real difference in this circuit from the previous ones is the transformer. The transformer secondary is connected in series with the plate circuit—where the plate battery was previously.

Remember from your study of transformers that the secondary (output) of a transformer always produces an alternating voltage. The secondary voltage is a sine wave as shown in the figure.

You'll remember that the sine wave is a visual picture, a graph of the change in alternating voltage as it builds from zero to a maximum value (positive) and then drops to zero again as it decreases to its minimum value (negative) in the cycle.

Assume that the polarity across the secondary during the first half-cycle of the input ac voltage is as shown in the figure. During this entire first half-cycle period, the plate's polarity will be **POSITIVE**. Under this condition, plate current flows, as shown by the ammeter.

The plate current will rise and fall because the voltage on the plate is rising and falling. Remember that current in a given circuit is directly proportional to voltage.

During the second half-cycle period, plate's polarity will be **NEGATIVE**. Under this condition, for this entire period, the diode will not conduct. If our ammeter could respond rapidly, it would drop to zero. The plate-current waveform (I_p) in figure 1-5 shows zero current during this period.

Here is a summary of effects of applying alternating voltage to the plate of the diode:

1. Diode plate current flows during the positive half-cycle. It changes value as the plate voltage rises and falls.

2. The diode cuts off plate current during the entire period of the negative half-cycle.

3. Diode plate current flows in **PULSES** because the diode cuts off half the time.

4. Diode plate current can flow in only one direction. It is always a direct current. (In this case **PULSATING DC**—one that flows in pulses.)

5. In effect, the diode has caused an alternating voltage to produce a direct current.

The ability to obtain direct current from an ac source is very important and one function of a diode that you will see again and again wherever you work in electronics.

The circuits that we have discussed up to this point were chosen to show the general concepts discovered by Edison and Fleming. They are not practical because they do no useful work. For now, only the concepts are important. Practical circuitry will be presented later in this chapter as you learn specific points about the construction, limitations, and other characteristics of modern diode tubes.

Q5. An ac voltage is applied across a diode. The tube will conduct when what alternation of ac is applied to the plate?

Q6. What would be the output of the circuit described in question 5?

DIODE CONSTRUCTION

Diode tubes in present use are descendants of Fleming's valve. There is a family resemblance, but many changes have been made from the original. Diodes are both smaller and larger, less powerful and more powerful, and above all, more efficient and more reliable. The search for greater efficiency and reliability has resulted in many physical changes, a few of which will be covered in the next paragraphs.

Most of what is said here about construction and materials will be true of all electron tubes, not just diodes.

Filaments

Modern filaments in **ALL** tubes last longer, emit greater amounts of electrons for a given size, and many operate at a lower temperature than in the early days. Most improvements have resulted from the use of new materials and from better quality control during manufacture.

Three materials that are commonly used as filaments are tungsten, thoriated tungsten, and oxide-coated metals.

Tungsten has great durability but requires large amounts of power for efficient thermionic emission. Thoriated-tungsten filaments are made of tungsten with a very thin coat of thorium, which makes a much better emitter of electrons than just tungsten. Oxide-coated filaments are made of metal, such as nickel, coated with a mixture of barium and strontium oxides. The oxide coat, in turn, is coated with a one-molecule-thick layer of metal barium and strontium. Oxide coating produces great emission efficiency and long life at relatively low heat.

A major advance in electronics was the elimination of batteries as power sources for tubes. Except in electronic devices designed to be operated away from the ac power source, alternating current is used to heat filaments.

Voltage may be supplied by a separate filament transformer or it may be taken from a filament winding that is part of a power transformer. The actual voltage may vary from 1 volt up and depends on the design of the tube. Common filament voltages are 5.0, 6.3, and 12.6 volts ac. Filaments may be connected in series with other tube filaments or may be in parallel with each other. This is determined by the equipment designer.

Cathodes

As was mentioned previously, a more formal name for the electron-emitting element in a tube is the **CATHODE**.

Cathodes in all tubes, not just diodes, are of two general types, either directly heated or indirectly heated. Each has its advantages and disadvantages.

DIRECTLY HEATED.—The filament that has been discussed so far is the directly heated cathode. Directly heated cathodes are fairly efficient and are capable of emitting large amounts of electrons. Figure 1-6 shows this type and its schematic symbol.

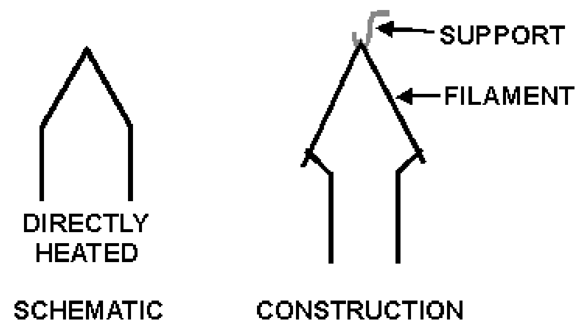


Figure 1-6.—Cathode schematic representation.

An added advantage of this type of filament is the rapidity with which it reaches electron-emitting temperature. Because this is almost instantaneous, many pieces of electronic equipment that must be turned on at infrequent intervals and be instantly usable have directly heated cathode tubes.

There are disadvantages. Because of its construction, parts of the filament are closer to the plate than other parts. This results in unequal emission and a loss of efficiency. Another disadvantage occurs when dc is used to heat a filament. The filament represents a resistance. When current flows through this resistance, a voltage drop occurs. The result is that one side of the resistance, or filament, is more negative than the other side. The negative side of the filament will emit more electrons than the positive side; which, again, is less efficient than if the filament has equal emission across its entire surface.

When ac is the source of filament power, it causes a small increase and decrease of temperature as it rises and falls. This causes a small increase and decrease of emitted electrons. This effect is not too important in many diode circuits, but it is undesirable in other tube circuits.

INDIRECTLY HEATED.—Figure 1-7 shows this type of cathode and its schematic symbol. Indirectly heated cathodes are always composed of oxide-coated material. The cathode is a cylinder, a kind of sleeve, that encloses the twisted wire filament. The only function of the filament is to heat the cathode. The filament is often called a heater when used in this manner.

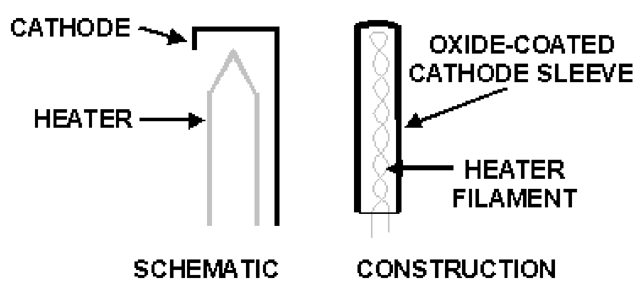


Figure 1-7.—Indirectly heated cathode schematic.

Some schematics do not show heaters and heater connections. Heaters, of course, are still present in the tubes, but their appearance in a schematic adds little to understanding the circuit. The heater is not considered to be an active element. For example, a tube with an indirectly heated cathode and a plate is still called a diode, even though it might seem that there are three elements in the tube.

Because indirectly heated cathodes are relatively large, they take longer to heat to electron-emitting temperature. Once up to temperature, however, they do not respond to the small variations in heater temperature caused by ac fluctuations. Because of the inherent advantages, most tubes in use today have indirectly heated cathodes.

Q7. Besides tungsten, what other materials are used for cathodes in vacuum tubes?

Q8. What is the advantage of directly heated cathodes?

Plates

Edison's plate was just that-a plate, a flat piece of metal. Plates are no longer flat but are designed in many different shapes. Figure 1-8 shows two diodes, one with a directly heated cathode, the other with an indirectly heated cathode. Each plate is cut away to show the internal position of elements and the plate shapes.

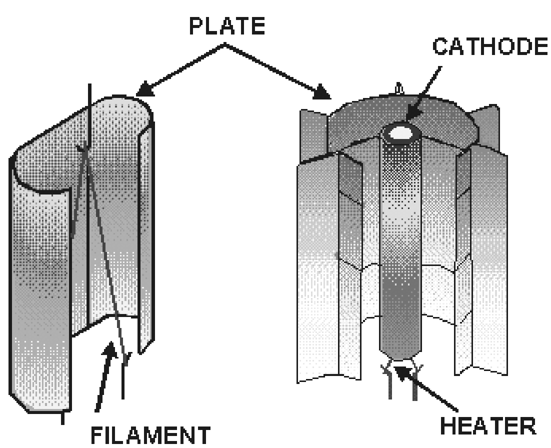


Figure 1-8.—Cutaway view of plate construction.

Plates must be able to hold up under the stress of heat created by the flow of plate currents and the closeness of hot cathodes. They need to be strong enough to withstand mechanical shocks produced by vibration and handling.

Some typical materials used for electron tube plates are tungsten, molybdenum, graphite, nickel, tantalum, and copper.

Tube Bases

The base shown in figure 1-9 has two functions. First, it serves as the mounting for tube elements. Second, it serves as the terminal points for the electrical connections to the tube elements. This is accomplished by molding or otherwise bringing pins (or prongs) through the base. The internal ends of these pins are connected to tube elements. The pins themselves are male connections.

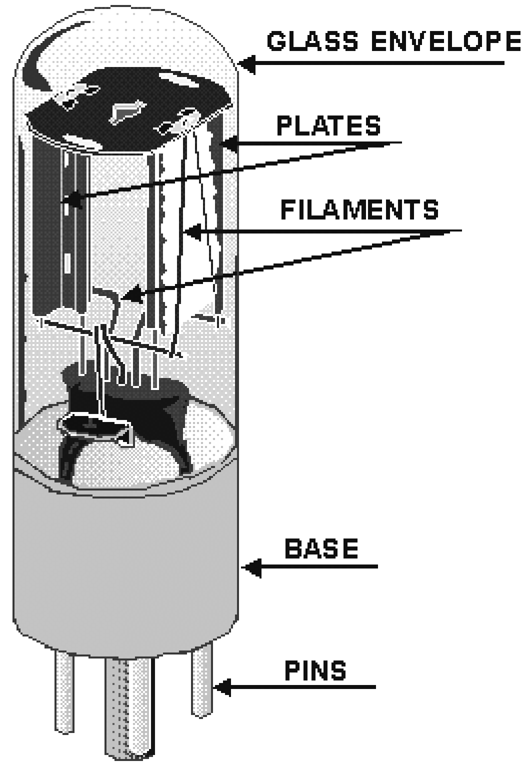


Figure 1-9.—Diode construction.

The base must be mechanically strong and made of an insulating material to prevent the tube elements from shorting.

Because they require relatively frequent replacement, most tubes are designed to plug into sockets permanently mounted in the equipment. Tube pins and sockets are so designed that tubes cannot be plugged in incorrectly.

Tube sockets must make secure mechanical and electrical contact with tube pins, must insulate pins from each other, and must provide terminals to which circuit components and conductors are connected.

Each element of a tube is connected to a pin in its base. To trace a circuit easily and efficiently, you must match elements with their pin numbers. This information is available in tube manuals and equipment schematics. Figure 1-10 shows these numbers on one example of a diode symbol. You will also note the designation V1 beside the tube. Electron tubes are often identified in schematic diagrams by the letter V and a number.

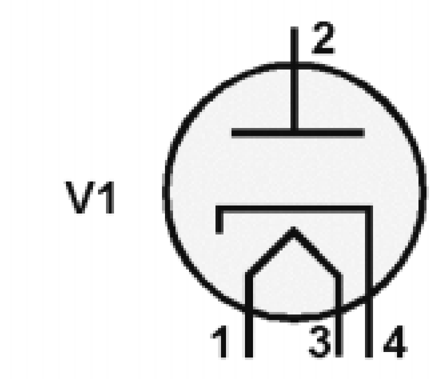


Figure 1-10.—Identification of tube elements.

Now, to use the information in the symbol, you need to know the system used to number tube pins and socket connections.

Figure 1-11 shows five common pin configurations as viewed from the bottom of each tube or socket. This is important. In every case, pins and pin connections on sockets are numbered in a clockwise direction—**WHEN VIEWED FROM THE BOTTOM**.

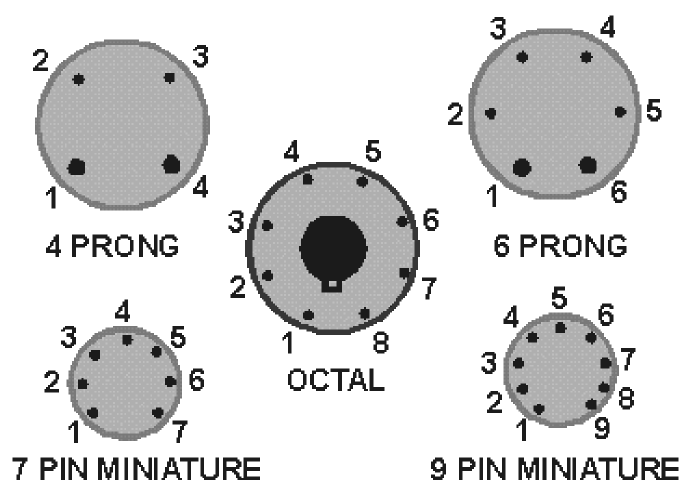


Figure 1-11.—Pin Identification; all tubes are viewed from the bottom.

In each of the five pictures in figure 1-11, there is an easily identified point from which to start numbering. In the 4-prong and 6-prong tubes, the point is between the two larger prongs. In the octal tube, the point is directly down from the keyway in the center of the tube. In the 7-pin and 9-pin miniatures, the point is identified by the larger distance between pins.

Q9. Name two functions of the base of a vacuum tube.

The Envelope

The envelope of a tube may be made of ceramic, metal, or glass. Its major purpose is to keep the vacuum in and the atmosphere out. The main reason for this is that the heated filament would burn up in the atmosphere. There are other reasons for providing a vacuum, but the important thing is to realize that a tube with a leaky envelope will not function properly.

The silver spot you will sometimes see on the inside surface of the glass envelope of a vacuum tube is normal. It was caused by the "flashing" of a chemical during the manufacture of the tube. Burning the chemical, called the **GETTER**, helps to produce a better vacuum and eliminates any remaining gases.

ELECTRICAL PARAMETERS OF DIODES

Thousands of different tubes exist. While many of them are similar and even interchangeable, many have unique characteristics. The differences in materials, dimensions, and other physical characteristics, such as we have just covered, result in differing electrical characteristics.

The electrical parameters of a diode, and any tube, are specific. In the process of discussing these parameters, we will state exact values. Voltages will be increased and decreased and the effects measured. Limiting factors and quantities will be explored and defined. The discussion will be based on simplified and experimental circuits.

It is important for you to realize that practically all of the parameters, limitations, definitions, abbreviations, and so on that we will cover in these next paragraphs will apply directly to the more complex tubes and circuits you will study later. Diode parameters are the foundation for all that follows.

Symbols

You have learned to use letters and letter combinations to abbreviate or symbolize electrical quantities. (The letters E, I, and R are examples.) We will continue this practice in referring to tube quantities. You should be aware that other publications may use different abbreviations. Many attempts have been made to standardize such abbreviations, inside the Navy and out. None have succeeded completely.

Table 1-1 lists electron-tube symbols used in the remainder of this chapter. The right-hand column shows equivalent symbols that you may find in **OTHER** texts and courses.

Table 1-1.—Symbols for Tube Parameters

SYMBOLS THIS TEXT	MEANING	OTHER TEXTS
E_p	PLATE VOLTAGE, D.C. VALUE	
E_{bb}	PLATE SUPPLY VOLTAGE, D.C.	B+
E_c	GRID BIAS VOLTAGE, D.C. VALUE	E_g
E_{cc}	GRID BIAS SUPPLY VOLTAGE, D.C.	C-
e_b	INSTANTANEOUS PLATE VOLTAGE	
e_c	INSTANTANEOUS GRID VOLTAGE	
e_g	A.C. COMPONENT OF GRID VOLTAGE	
e_p	A.C. COMPONENT OF PLATE VOLTAGE (ANODE)	
I_p	D.C. PLATE CURRENT	
R_p	D.C. PLATE RESISTANCE	
R_g	GRID RESISTANCE	
R_k	CATHODE RESISTANCE	
R_L	LOAD RESISTANCE	

Plate Voltage-Plate Current Characteristic

You know that a positive voltage on the diode plate allows current to flow in the plate circuit. Each diode, depending on the physical and electrical characteristics designed into the diode, is able to pass an exact amount of current for each specific plate voltage (more voltage, more current-at least to a point).

The plate voltage-plate current characteristic for a given diode is a measure of exactly how much plate voltage controls how much plate current. This is often called the E_p - I_p characteristic.

The E_p - I_p characteristic for a given diode, is determined by design engineers using mathematical analysis and laboratory experiment. You, as a technician, will never need to do this. However, you will use the results obtained by the engineers. You will also use your knowledge of the diode as you analyze equipment malfunction.

Assume that we have the circuit in figure 1-12. (The filament has the proper voltage-even though it isn't shown on the diagram.) Our purpose is to determine just how a changing voltage on the plate changes (or controls) the plate current. The method is as follows:

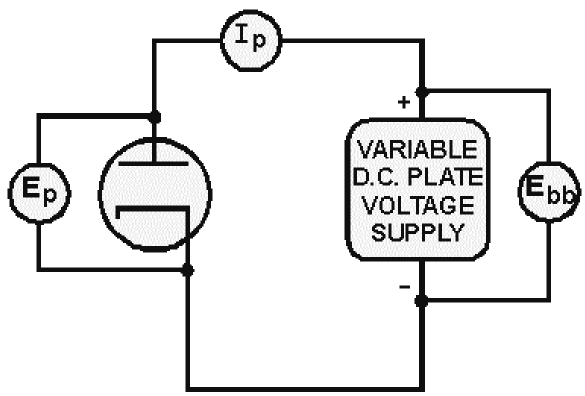


Figure 1-12.—Determining diode plate characteristic.

1. Starting with zero volts from our variable dc voltage source, increase the plate voltage (E_p) in steps of 50 volts until you reach 400 volts.

2. At a each 50-volt step, measure the milliamperes of plate current (I_p) that flow through the meter. Record the I_p meter readings, step by step, so that you may analyze the results.

Assume that table 1-2 shows our results. While we could use the table, a more normal procedure is to plot a graph of the values. Such a graph is called an **E_p - I_p CURVE** and is shown in figure 1-13. Each tube has its own E_p - I_p curve, which is available in commercial tube manuals and in many equipment technical manuals. Each curve will be different in some respects from every other curve. The shapes, however, will be similar.

Table 1-2.—E_p - I_p Values Obtained by Experiment

E_p	0	50	100	150	200	250	300	350	400
I_p	0	.002	.005	.010	.020	.030	.040	.042	.045

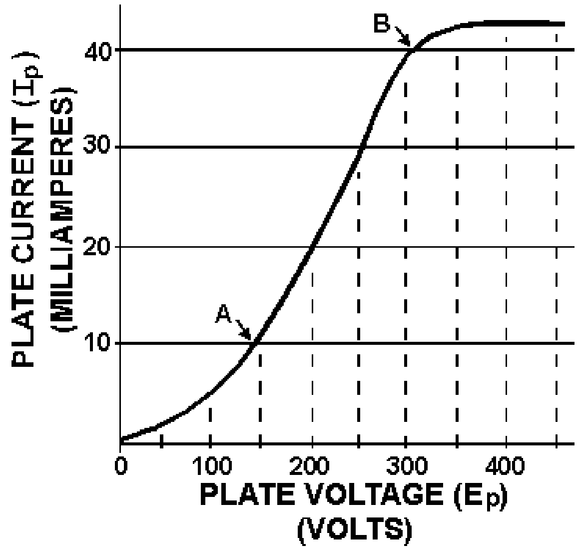


Figure 1-13.—E_p - I_p characteristic curve.

The E_p - I_p curve in figure 1-13, although just an example, is typical of real plate characteristic curves. You may learn certain characteristics that apply to both diodes and other tubes by studying it.

First, look at the part of the curve to the left of point A. Because it is not a straight line, it is referred to as **NONLINEAR**. Note that a change of 150 volts (0-150) caused a change of 10 mA of plate current (0-10). In comparison with the straight-line part of the curve, between points A and B, this is a relatively small change in current. The smaller the change in current, the flatter the curve.

In explaining this **NONLINEAR** portion of the curve, let's go back just a bit to electron emission. The electrons emitted by a cathode form a cloud around the cathode. This cloud is called the **SPACE CHARGE**. The closer the space charge is to the cathode, the more densely packed it is with electrons. In our example, the lower plate voltages (0-150 volts) over this part of the curve exert a pull on only the outer fringe of the space charge where there are few electrons. This results in relatively few electrons flowing to the plate.

Now look at the center portion of the curve between A and B. This is known as the **LINEAR** portion because it is nearly **a STRAIGHT LINE**. Over this portion, a change of 50 volts E_p causes a change of 10 mA I_p.

The reason for the increased change in plate current for a given change of plate voltage also has to do with the space charge. With a higher plate voltage (over 150 volts), the attraction from the plate begins to influence the **DENSER** part of the space charge that has greater numbers of electrons. Therefore, a higher current flows for a given voltage than in the nonlinear part. The curve becomes steeper. In our example, this linearity continues to about 300 volts, point B.

Lastly, let's look at the top portion of the curve. The plate current plotted here is produced by the higher plate voltages. However, the amount of current change for a given voltage change is greatly reduced. The reason for this again involves the space charge. At about 300 volts, almost all of the electrons in the space charge are flowing to the plate. A higher voltage cannot attract more electrons because the cathode cannot produce any more. The point where all (or almost all) available electrons are being drawn to the plate is called **PLATE SATURATION** or just **SATURATION**. This is one of the limiting factors of every tube.

You can see from the analysis that the most consistent control of plate current takes place over the linear portion of the E_p - I_p curve. In most applications, electron tubes are operated in this linear portion of the characteristic curve.

Plate Resistance (R_p)

One tube parameter that can be calculated from values on the E_p - I_p curve is known as plate resistance, abbreviated as R_p. In a properly designed electron tube, there is no physical resistor between cathode and plate; that is, the electrons do not pass through a resistor in arriving at the plate. You may have wondered, however, why the variable dc voltage source of figure 1-12 didn't blow a fuse. Doesn't the plate circuit appear to be a short circuit-a circuit without a load to limit the current?

The fact is, there is a very real, effective **RESISTANCE** between cathode and plate. It is not lumped in a resistor, but the circuit may be analyzed as if it is. The plate resistance of a given tube, R_p, can be calculated by applying Ohm's law to the values of E_p and I_p. Figure 1-14 is a typical diode E_p - I_p curve. The plate resistance has been figured for R_p under three different conditions, as follows:

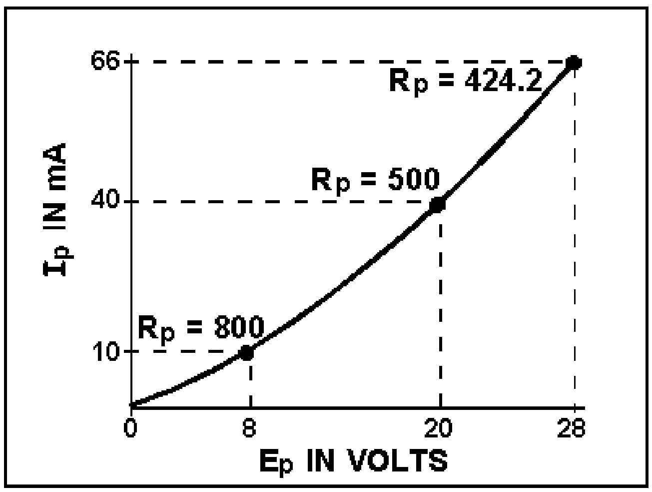


Figure 1-14.—The E_p - I characteristic curve for a diode.

Remember that 1 mA = .001 ampere; therefore 40 mA =.040 ampere.

Solution:

$$R_P = \frac{E_P}{I_P}$$

$$R_P = \frac{20\text{volts}}{.040\text{ amperes}}$$

$$R_P = 500\text{ ohms}$$

The other two indicated values of R_p were figured in the same way.

You should note that there is very little difference in plate resistance when the E_p and I_p values are taken from the linear portions of curves. Check this out with values taken from the linear portion of figure 1-13.

R_p (with a capital R) is the effective resistance offered to direct current.

PLATE RESISTANCE IN GAS DIODES.—Gas diodes are a type of tube that we have not yet discussed. They are mentioned here only because of their plate-resistance characteristic.

Instead of a high-vacuum environment, some tubes have small amounts of gas introduced in the envelope vacuum during manufacture. Argon, neon, helium, or mercury vapor are commonly used.

When a certain minimum voltage is placed on the plate, the gas molecules in the envelope ionize. This happens by a process that will be explained when gas diodes are studied. The positive ions tend to cancel some of the effects of the space charge that influence plate resistance in a vacuum tube. This canceling reduces internal plate resistance to a relatively low, constant value. In applications that require a large plate current, the low plate resistance of a gas-filled diode has an efficiency that cannot be approached by a high-vacuum diode.

This and other characteristics of gas tubes will be covered later.

Q10. Vacuum tubes are designed to operate in what portion of the E_p - I_p curve?

Q11. What value can be calculated from the values found on an E_p - I_p curve?

Plate Dissipation

When electrons are attracted from the space charge to the plate, they are accelerated by the attraction. Their gain in speed gives them energy that causes them to strike the plate with a considerable force. As the electrons strike the plate, this energy is converted to heat. The plate must be able to withstand the associated increase in temperature. The maximum amount of power (watts) that a given plate can safely dissipate (as heat) is called the **PLATE DISSIPATION** rating.

To find the amount of plate dissipation for a given tube under a particular set of plate conditions, use the following equation:

$$\text{Plate dissipation} = E_p \times I_p$$

For our current problem,

$$\text{Plate dissipation} = 20 \text{ volts} \times .040 \text{ amperes}$$
$$\text{Plate dissipation} = .8 \text{ watt}$$

This is a relatively small wattage. It's probable that the plate of our example diode is not overheating. A tube manual could tell us for sure.

Plate dissipation is a circuit loss that must be made good by the power source in a circuit. In our example, this is the plate voltage supply.

Peak Current Rating

The maximum instantaneous current that a tube can pass in the normal direction (cathode to plate) without damage is called the **PEAK CURRENT RATING**. Peak current rating is determined by the amount of electrons available from the cathode and the length of time plate current flows.

Peak Voltage Rating

This is the maximum instantaneous voltage that can be applied to a tube in the normal direction without a breakdown.

Peak Inverse Voltage Rating

This is the maximum voltage that can be applied to a tube in the reverse direction (plate negative with respect to the cathode)-exceeding this will cause arc-over from the plate to the cathode and will damage the tube. **PIV**, as this is sometimes abbreviated, becomes very important in the rectifier circuit to be discussed as a later major subject.

Transit Time

Things that happen in electricity and electronics are often explained as if they happen instantaneously. As fast as electricity acts, however, the truth is that cause and effect are separated by a certain amount of time.

Each tube has a factor called **TRANSIT TIME**, which is the time required for an individual electron to move from the cathode to the plate. In certain applications involving high-frequency voltages, transit time places a limitation on tubes. We will explain this limitation when we discuss the circuits it affects.

Summary of Diode Parameters and Limitations

You should now have a basic understanding of diodes, many of their characteristics, and some of their limitations. One of the more important concepts that you should now understand is that most of these characteristics influence each other. For example, practically all plate characteristics are interrelated. Change one and the others change. Another example is heater voltage. Every tube parameter affected by the cathode depends on proper heater voltage. Interrelationships such as these make electronics both fascinating and, at times, frustrating.

Many of the limiting factors that we have discussed are the same ones found in other electrical devices such as motors, stoves, toasters, and so on. Heating and overheating, insulation breakdown, and excessive voltage and current are all limitations that you have noted before.

The point is that you can and should apply just about everything you have learned about electricity to electron tubes. Little is new except the environment.

Q12. A large negative voltage is applied to the plate of a diode, and a large positive voltage is applied to the cathode. If the tube conducts, what tube parameter has been exceeded?

THE TRIODE

Diode electron tubes can be used as rectifiers, switches, and in many other useful applications. They are still used in Fleming's original application in some radio circuits. You will learn more of these applications in other *NEETS* modules and later will see the diode in several pieces of electronic equipment.

As with all inventions, Fleming's diode was immediately the subject of much experimentation and many attempts at improvement. An American experimenter, Dr. Lee De Forest, added another active element to the diode in 1906. He was trying to improve the radio application of Fleming's diode. His new tube was eventually called a triode.

DeForest's triode was not very successful as a radio "detector." (Detectors will be studied in a later *NEETS* module.) However, in 1912, De Forest discovered that his original triode could **AMPLIFY** or magnify very weak electrical impulses. It is because of the triode's ability to amplify that De Forest is honored as one of the great radio pioneers.

The immediate application of the triode amplifier was in telephone and radio. Both fields were limited because electrical impulses (signals) became weaker and weaker as the distance from the signal source increased. The triode, along with other developments of the time, made long-distance communications possible. Looking back, we can now see that the amplifying tube was the real beginning of modern electronics and influenced everything that followed. Let's find out more about the idea of amplification and how it is done in the triode.

You are already familiar with a type of amplification. In a previous *NEETS* module, step-up transformers were discussed. You should remember that an input voltage applied to the primary of a step-up transformer is increased in amplitude at the secondary by a factor determined by the step-up turns ratio.

For example, if 5 volts were applied to the primary of a 1:3 step-up transformer, the secondary would produce 15 volts. In other words, the input voltage was amplified by a factor of 3. When applied to electronic circuits, these primary and secondary voltages are more often called signals, or input and output signal, respectively. In electronics, the amplitude of an input signal must sometimes be increased many times-often, hundreds or thousands of times!

Because of size and design limitations, transformers are usually not practical for use in electronics as amplifiers.

DeForest's first experiment with the diode was to place an additional metal plate between the cathode and plate. He then placed an ac signal on the metal plate. When the circuit was energized, De Forest found that the ammeter stayed on zero regardless of the polarity of the input signal.

What was happening was that the new element was blocking (or shadowing) the plate. Any electrons attempting to reach the plate from the cathode would hit the new element instead. As the circuit didn't work, it was back to the drawing board.

In his next attempt, De Forest decided to change the element between the cathode and the plate. Instead of a solid metal plate, he used a wire mesh. This would allow electrons to flow from the cathode, **THROUGH THE WIRE MESH**, to the plate. This tube circuit is shown in figure 1-15. In view (A) you see De Forest's circuit with 0 volts applied to the third element, (today called a control grid or occasionally just the grid). Under these conditions, assume that the ammeter reads 5 milliamperes. With no voltage applied to the grid, the grid has little effect on the electron stream. For all practical purposes, the control grid is not there. Most electrons flow through the open mesh. The tube functions as a diode.

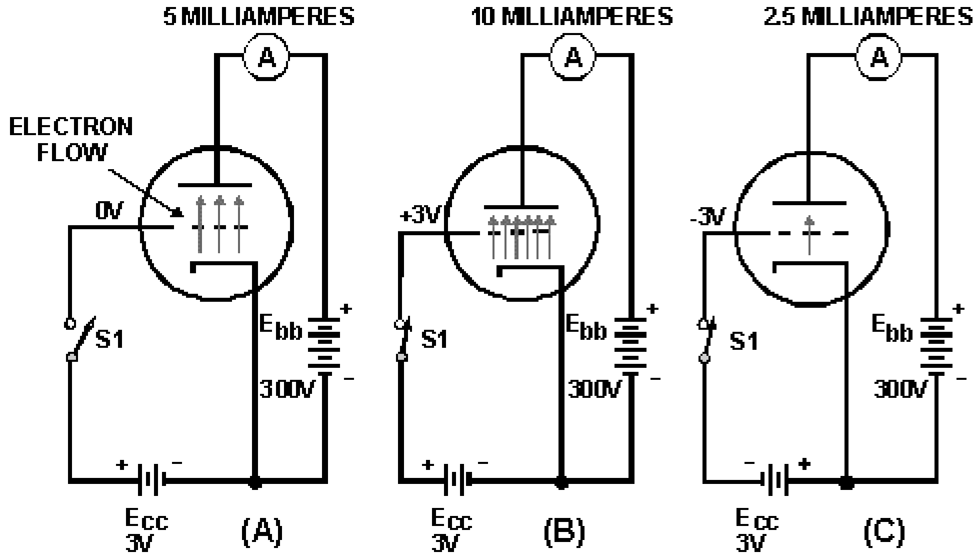


Figure 1-15.—DeForest's experiment.

In view (B), you see De Forest's tube with +3 volts applied to the control grid. When De Forest applied this voltage, he found that plate current, I_p, increased by a large amount. (We'll say it doubled to simplify the explanation.) You already know that the only way to double the plate current in a diode is to increase the plate voltage by a large amount. Yet, De Forest had doubled plate current by applying only 3 volts positive to the control grid!

The reason for this is fairly easy to understand. It's the old principle of "opposites attract." When the control grid was made positive, electrons surrounding the cathode (negative charges) were attracted to the grid. But remember, the grid is a metal mesh. Most of the electrons, instead of striking the grid wires, were propelled through the holes in the mesh. Once they had passed the grid, they were attracted to the positive charge in the plate.

You might wonder why the grid would make that much difference. After all, the plate has 300 volts on it, while the grid only has 3 volts on it. Surely the plate would have a greater effect on current flow than a grid with only one one-hundredth the attractive potential of the plate. But remember, in your study of capacitors you discovered that opposites attract because of electrostatic lines of force, and that the strength of electrostatic lines of force decreased with distance. In his tube, DeForest had placed the grid very close to the cathode. Therefore, it had a greater effect on current flow from the cathode than did the plate, which was placed at a much greater distance from the cathode. For this reason, De Forest was able to double the current flow through the tube with only +3 volts applied to the grid.

DeForest had certainly hit on something. Now the problem was to find out what would happen when a negative potential was applied to the grid. This is shown in view (C) of figure 1-15. When De Forest applied -3 volts to the grid, he found that plate current decreased to half of what it was when the grid had no voltage applied. The reason for this is found in the principle of "likes repel." The negatively charged grid simply repelled some of the electrons back toward the cathode. In this manner, the attractive effect of the plate was decreased, and less current flowed to the plate.

Now De Forest was getting somewhere. Using his new tube (which he called a triode because it had 3 elements in it), he was able to control relatively large changes of current with very small voltages. But! was it amplification? Remember, amplification is the process of taking a small signal and increasing its amplitude. In De Forest's circuit, the small input signal was 3 volts dc. What De Forest got for an output was a variation in plate current of 7.5 milliamperes. Instead of amplification, De Forest had obtained "conversion," or in other words, converted a signal voltage to a current variation. This wasn't exactly what he had in mind. As it stood, the circuit wasn't very useful. Obviously, something was needed. After examining the circuit, De Forest discovered the answer—Ohm's law. Remember E = I × R? De Forest wanted a voltage change, not a current change. The answer was simple:

If you want		and you have		simply add
E	=	I	×	R

In other words, run the plate current variation (caused by the voltage on the grid) through a resistor, and cause a varying voltage drop across the resistor. This is shown in figure 1-16.

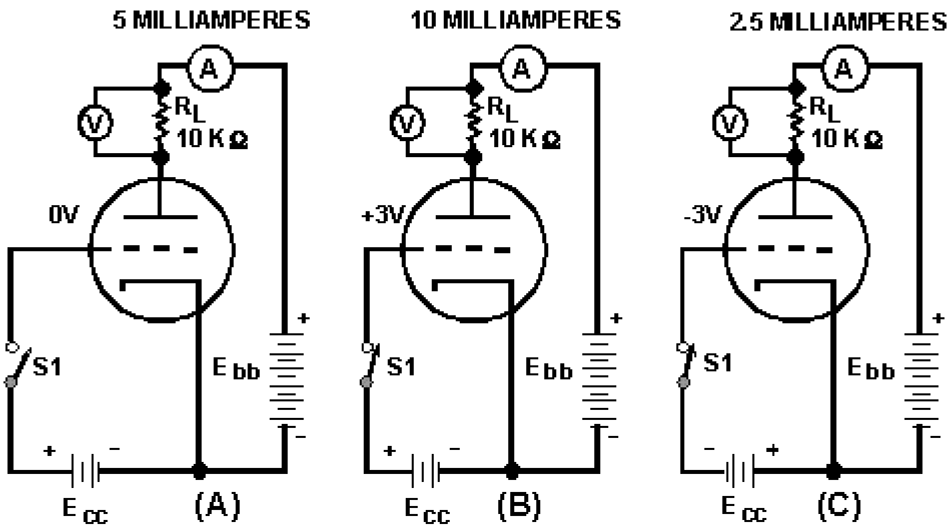


Figure 1-16.—Operation of the plate load resistor.

The circuit is identical to the one in figure 1-15 except that now a resistor (called a plate-load resistor, R_L) has been added to the plate circuit, and a voltmeter has been added to measure the voltage drop across R_L.

In view (A) of figure 1-16, the control grid is at 0 volts. Once again 5 milliamperes flow in the plate circuit. Now, the 5 milliamperes must flow through R_L. The voltage drop is equal to:

$E = I \times R$

$E = (5 \times 10^{-3} \text{ amperes}) \times (10 \times 10^{3} \text{ ohms})$

$E = (5 \times 10^{-3}) \times (10 \times 10^{3})$

$E = 5 \times 10$

$E = 50$ volts

Thus the voltage drop across the plate-load resistor, R_L, is 50 volts when no voltage is applied to the grid. In view (B) of the figure, +3 volts is applied to the control grid. Once again plate current increases to 10 milliamperes. The voltage drop across R_L is

$E = I \times R$

$E = (10 \times 10^{-3}$ amperes$) \times (10 \times 10^{3}$ ohms$)$

$E = (10 \times 10^{-3}) \times (10 \times 10^{3})$

$E = 10 \times 10$

$E = 100$ volts

By applying +3 volts to the grid, the voltage drop across R_L was increased by 50 volts (from the original 50 volts to 100 volts). In view (C), -3 volts has once again been applied to the control grid. Once again plate current decreases to 2.5 milliamperes, and the voltage drop across R_L drops to 25 volts.

We have caused the voltage across R_L to vary by varying the grid voltage; but is it amplification? Well, let's take a look at it. The grid voltage, or input signal, varies from +3 to -3 volts, or 6 volts. The voltage drop across R_L varies from 25 volts to 100 volts, or 75 volts. In other words, the triode has caused a 6-volt input signal (varying) to be outputted as a signal that varies by 75 volts. That's amplification!

Q13. What is the primary difference between a diode and a triode?

Q14. Why does the grid have a greater effect than the plate on electron flow through a vacuum tube?

Q15. What component is used in a triode amplifier to convert variation in current flow to voltage variation?

Let's summarize what you have learned so far:

- A relatively small change in voltage on the grid causes a relatively large change in plate current.
- By adding a plate-load resistor in series with the plate circuit, the changing plate current causes a changing voltage drop in the plate circuit.
- Therefore, the small voltage change on the grid causes a large change of voltage in the plate circuit.
- By this process, the small input signal on the grid has been amplified to a large output signal voltage in the plate circuit.

We'll leave De Forest at this point. He showed that the control grid can, in fact, **CONTROL** plate current. He also showed that the changing plate current can create a changing plate voltage. To some degree, his changing voltages and currents also changed the world.

INTRODUCTION TO GRID BIAS

We purposely left out several features of practical triode circuits from the circuits we just discussed. We did so to present the idea of grid control more simply. One of these features is grid bias.

Let's take another look at the circuit in figure 1-15(B). We found that the positive charge on the grid caused more plate current to flow. However, when the grid becomes positive, it begins to act like a small plate. It draws a few electrons from the space charge. These electrons flow from the cathode across the gap to the positive grid, and back through the external grid circuit to the cathode. This flow is known as *grid current*. In some tube applications, grid current is desired. In others it is relatively harmless, while in some, grid current causes problems and must be eliminated.

Most amplifier circuits are designed to operate with the grid **NEGATIVE** relative to the cathode. The voltage that causes this is called a **BIAS VOLTAGE**. The symbol for the bias supply is E_{cc}. One effect of bias (there are several other very important ones) is to reduce or eliminate grid current. Let's see how it works.

GRID BIAS is a steady, direct voltage that is placed at some point in the external circuit between the grid and the cathode. It may be in the cathode leg or the grid leg as shown in figure 1-17. It is always in series with the input signal voltage. In each of the circuits in figure 1-17, E_{cc} makes the grid negative with respect to the cathode because of the negative terminal being connected toward the grid and the positive terminal being connected toward the cathode. With identical components, each circuit would provide the same bias.

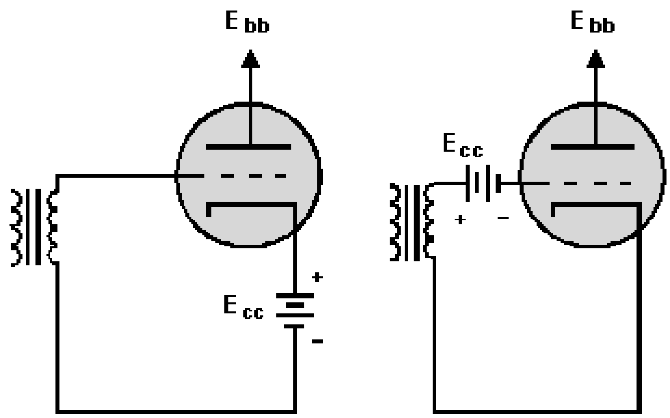
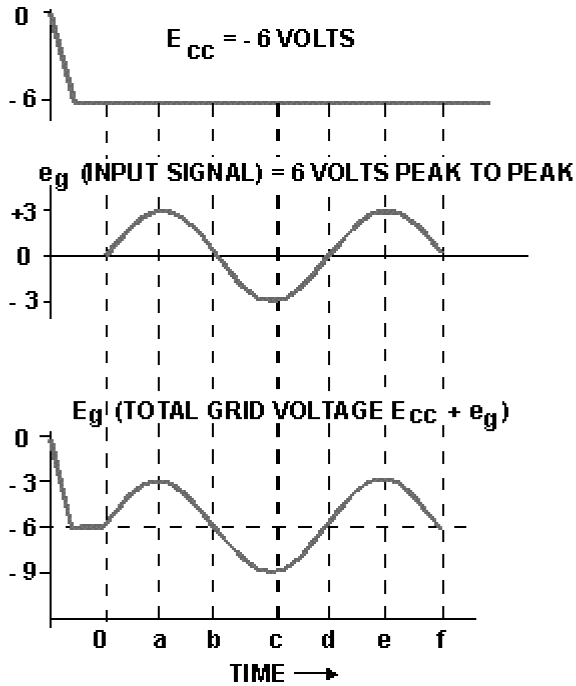


Figure 1-17.—Basic biasing of a triode.

Battery bias is practically never used in modern circuits. Because of its simplicity, however, we will use it in analyzing the effects of bias. We will present other, more practical methods later.

Let's assume that the bias voltage in figure 1-17 is -6 volts. Let's also assume that the peak-to-peak signal voltage from the transformer is 6 volts. Each of these voltage waveforms is shown in figure 1-18. From past experience you know that voltages in series **ADD**. Figure 1-18 has a table of the instantaneous values of the two voltages added together. The waveforms are drawn from these values.



TIME	0	a	b	c	d	e	f
BIAS (E_C)	- 6	- 6	- 6	- 6	- 6	- 6	- 6
SIGNAL (e_g)	0	+3	0	- 3	0	+3	0
TOTAL GRID VOLTS (E_g)	- 6	- 3	- 6	- 9	- 6	- 3	- 6

Figure 1-18.—Typical grid waveforms.

Because the bias voltage is more negative than the signal voltage is positive, the resultant voltage (bias plus signal), E_g, is **ALWAYS** negative. The signal, in this case, makes the grid voltage go either **MORE** or **LESS NEGATIVE**, (-9 to -3) but cannot drive it positive.

Under these circumstances, the negative grid always repels electrons from the space charge. The grid cannot draw current. Any problems associated with grid current are eliminated, because grid current cannot flow to a negative grid.

You have probably already realized that the negative bias also reduces plate current flow. (Negative charge on grid-less plate current, right?) The trick here is for the circuit designer to choose a bias and an input signal that, when added together, do not allow the grid to become positive nor to become negative enough to stop plate current.

Tube biasing is very important. You will learn much more about it shortly. From this brief introduction, you should have learned that grid bias

- is a steady, direct voltage that in most cases makes the grid negative with respect to the cathode;
- is in series with the signal voltage between grid and cathode;

- acts to reduce or eliminate grid current;
- acts to reduce plate current from what it would be if no bias existed;
- is produced in other ways than just by a battery; and
- is important for reasons other than those just studied.

OPERATION OF THE TRIODE

The circuit in figure 1-19 brings together all of the essential components of a triode amplifier. Before analyzing the circuit, however, we need to define the term **QUIESCENT**.

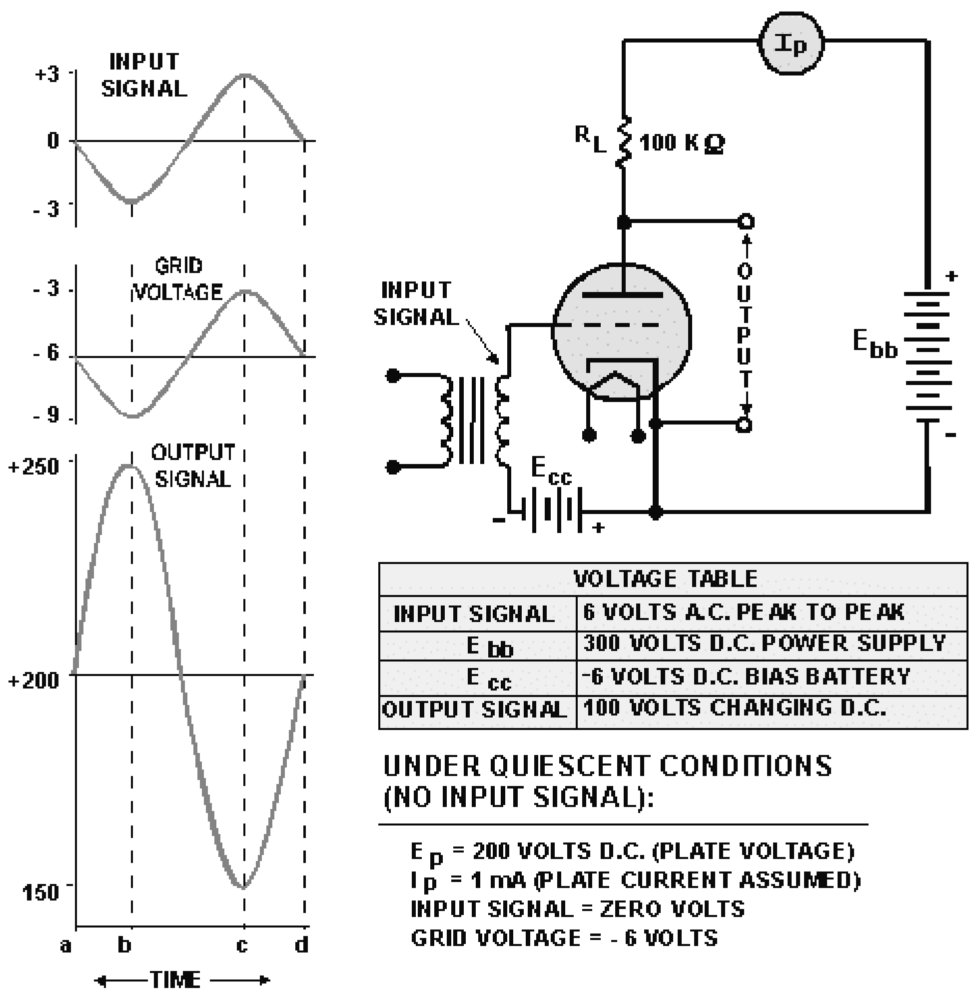


Figure 1-19.—Triode operation.

The term *quiescent* identifies the condition of a circuit with **NO INPUT SIGNAL** applied. With a given tube, bias supply, and plate supply, an exact amount of plate current will flow with no signal on the grid. This amount is known as the quiescent value of plate current. The quiescent value of plate voltage is the voltage between cathode and plate when quiescent current flows.

Simply, quiescent describes circuit conditions when the tube is not amplifying. The tube has no output signal and is in a kind of standby, waiting condition. Now let's go on to figure 1-19. With no input signal, under quiescent conditions, assume that 1 milliampere of current flows through the tube, cathode to plate. This current (I_p) will flow through R_L (load resistor) to the positive terminal of the battery. The current flowing through R_L causes a voltage drop (IR) across R_L equal to:

$$E = I_P \times R_L$$
$$E = 1 \times 10^3 \text{ amperes} \times 100 \times 10^3 \text{ohms}$$
$$E = 100 \text{ volts}$$

Subtracting the voltage dropped across the plate-load resistor from the source voltage of 300 volts gives you 200 volts (300 volts - 100 volts). Thus, the plate voltage (E_p) is at 200 volts. The quiescent conditions for the circuit are:

$$\text{grid voltage} = -6 \text{ volts}$$
$$\text{plate voltage } (E_P) = +200 \text{ volts}$$

These values are shown on the waveforms as time a in figure 1-19.

You should notice that even though the grid is more negative (-6 volts) than the cathode, the tube in the circuit is still conducting, but not as heavily as it would if the grid were at zero volts.

Now look at the input signal from the transformer secondary. For ease of explanation, we will consider only three points of the ac sine wave input: point b, the maximum negative excursion; point c, the maximum positive excursion; and point d, the zero reference or null point of the signal. At time b, the input signal at the grid will be at its most negative value (-3 volts). This will cause the grid to go to -9 volts (-6 volts + -3 volts). This is shown at time b on the grid voltage waveform. The increased negative voltage on the control grid will decrease the electrostatic attraction between the plate and the cathode. Conduction through the tube (I_p) will decrease. Assume that it drops to .5 milliamperes.

The decrease in plate current will cause the voltage drop across the plate-load resistor (R_L) to also decrease from 100 volts, as explained by Ohm's law:

$$E = I_P \times R_L$$
$$E = .0005 \text{ amperes} \times 100{,}000 \text{ ohms}$$

or

$$E = (.5 \times 10^{-3} \text{ amperes}) \times (100 \times 10^3 \text{ohms})$$
$$E = 50 \text{ volts}$$

Plate voltage will then rise +250 volts.

$$E_P = 300 \text{ volts} - 50 \text{ volts} = 250 \text{ volts}$$

This is shown on the output signal waveform at time b.

At time c, the input has reached its maximum positive value of +3 volts. This will decrease grid voltage to -3 volts (-6 volts + 3 volts). This is shown on the grid voltage waveform at time c. This in turn will increase the electrostatic force between the plate and cathode. More electrons will then flow from the cathode, through the grid, to the plate. Assume that the plate current in this case will increase to 1.5 milliamperes. This will cause plate voltage (E_b) to decrease to 150 volts as shown below.

$$E = I_P \times R_L$$
$$E = (1.5 \times 10^{-3} \text{ amperes}) \times (100 \times 10^3 \text{ohms})$$
$$E = 150 \text{ volts}$$
$$E_P = 300 \text{ volts} - 150 \text{ volts}$$
$$E_P = 150 \text{ volts}$$

This is shown on the output waveform at time c.

At time d, the input signal voltage decreases back to zero volts. The grid will return to the quiescent state of -6 volts, and conduction through the tube will again be at 1 milliampere. The plate will return to its quiescent voltage of +200 volts (shown at time d on the output waveform).

As you can see, varying the grid by only 6 volts has caused the output of the triode to vary by 100 volts. The input signal voltage has been amplified (or increased) by a factor of 16.6. This factor is an expression of amplifier **VOLTAGE GAIN** and is calculated by dividing the output signal voltage by the input signal voltage.

Before going on to the next section, there is one more thing of which you should be aware. Look again at the waveforms of figure 1-19. Notice that the output voltage of the amplifier is 180º out of phase with the input voltage. You will find that this polarity inversion is a characteristic of any amplifier in which the output is taken between the cathode and the plate. This is normal and should not confuse you when you troubleshoot or work with this type of circuit.

Q16. Why is the control grid of a triode amplifier negatively biased?

Q17. For a circuit to be considered to be in the quiescent condition, what normal operating voltage must be zero?

Q18. A triode amplifier similar to the one shown in figure 1-19 has an E_{bb} -350 volts dc. The plate-load resistor is 50 kΩ. Under quiescent conditions, 1.5 milliamperes of current conducts through the tube. What will be the plate voltage (E_p) under quiescent conditions?

Q19. A 2-volt, peak-to-peak, ac input signal is applied to the input of the circuit described in Q18. When the signal is at its maximum positive value, 2.5 milliamperes flows through the tube. When the input is at its maximum negative value, conduction through the tube decreases to .5 milliamperes.

a. What is the peak-to-peak voltage of the output signal?

b. What is the phase relationship between the input and output signals?

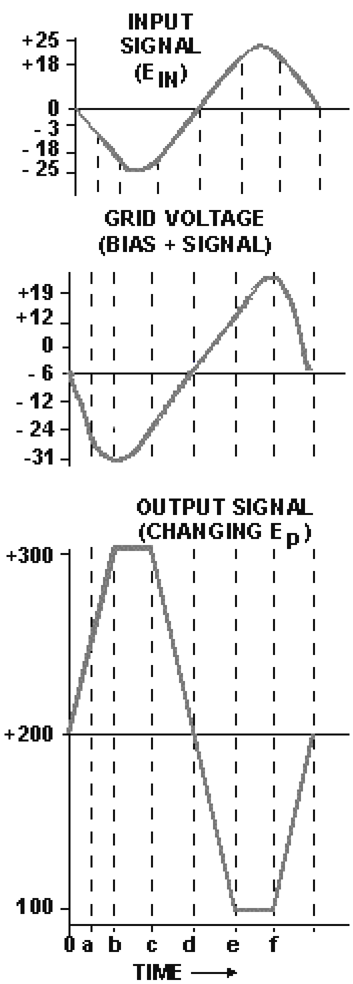
FACTORS AFFECTING TRIODE OPERATION

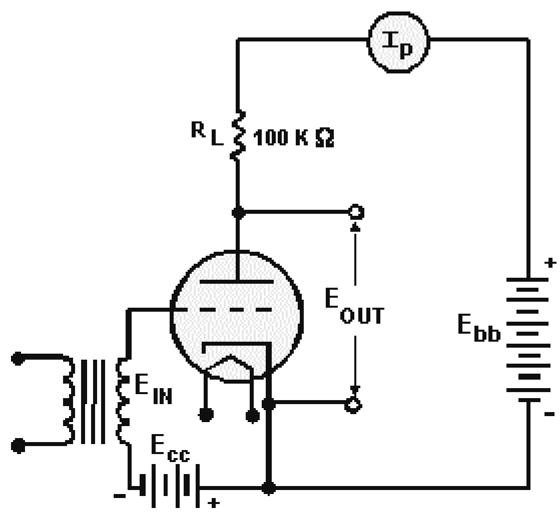
The triode circuit you have just studied is a fairly simple affair. In actual application, triode circuits are a bit more complex. There are two reasons for this. The first has to do with the triodes ability to amplify and perform other functions. Triodes come in many different types. Each of these types has different internal characteristics and different capabilities. Because of this, each triode circuit must be designed to accommodate the triodes special characteristics. The second reason for the increase in complexity has to do with **DISTORTION**. Distortion occurs in a tube circuit any time the output waveform is not a faithful reproduction of the input waveform.

Polarity inversion and voltage gain of the output waveform are not included in this definition of distortion. Some circuits are designed to distort the output. The reason and methods for this deliberate distortion will be covered in a later *NEETS* module. For the most part, however, we desire that circuits eliminate or reduce distortion.

Because the grid is close to the cathode, small changes in grid voltage have large effects on the conduction of triodes. If a large enough input signal is placed on the grid, a triode may be driven into either plate-current cutoff or plate-current saturation. When this occurs, the tube is said to be **OVERDRIVEN**. Overdriving is considered to be a form of **DISTORTION**.

Look at time zero (0) in the waveforms of figure 1-20. The input signal (E_{in}) is at zero volts. Grid voltage equals the bias voltage (-6 volts), and one milliampere of current is flowing through the tube (quiescent state). Plate voltage (E_p) is 200 volts.





VOLTAGE TABLE	
E_{IN} INPUT SIGNAL	50 VOLTS PEAK TO PEAK
E_{bb} PLATE SUPPLY	+ 300 VOLTS D.C.
E_{cc} BIAS SUPPLY	-6 VOLTS D.C.
E_{OUT} OUTPUT SIGNAL	200 VOLTS PEAK TO PEAK
E_p AT CUTOFF	+ 300 VOLTS D.C.
E_p AT SATURATION	+ 100 VOLTS D.C.

Figure 1-20.—Overdriven triode.

On the negative half of the input signal, the grid voltage is made more negative. This reduces plate current which, in turn, reduces the voltage drop across R_L. The voltage between cathode and the plate is thereby increased. You can see these relationships by following time "a" through the three waveforms.

Now, let's assume that this particular triode cuts plate current flow off when the grid reaches -24 volts. This point is reached at time b when E_{in} is -18 and the bias is -6 (-18 and -6 = -24). Plate current remains cut off for as long as the grid is at -24 volts or greater.

With zero current flowing in the plate circuit, there is no voltage drop across R_L. The entire plate-supply voltage, E_{bb} (300 volts), appears as plate voltage between the cathode and the plate. This is shown at time b in the output signal waveform. Between time b and time c, the grid voltage is greater than -24 volts. The plate current remains cutoff, and the plate voltage remains at +300. The output waveform between time b and time c cannot follow the input because the plate voltage cannot increase above +300 volts. The output waveform is "flattopped." This condition is known as **AMPLITUDE DISTORTION**.

When the grid voltage becomes less negative than -24 volts, after time c, the tube starts conducting, and the circuit again produces an output.

Between time c and time d, the circuit continues to operate without distortion. At time e, however, the output waveform is again distorted and remains distorted until time f. Let's see what happened.

Remember that every cathode is able to emit just so many electrons. When that maximum number is being emitted, the tube is said to be at **SATURATION** or **PLATE SATURATION**. Saturation is reached in a triode when the voltages on the grid and plate combine to draw all the electrons from the space charge.

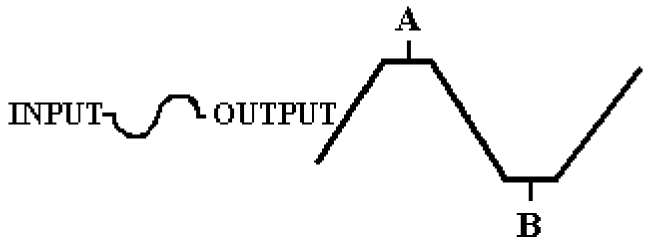
Now, as our grid becomes less negative (between time c and time d), and actually becomes positive (between time d and time e), the plate current increases, the voltage across R_L increases, and the plate voltage decreases.

Apparently when the grid voltage reached +12 volts at time e, the plate current reached saturation. Maximum plate current (at saturation) results in maximum voltage across R_L and minimum plate voltage. Any grid voltage higher than +12 volts cannot cause further changes in the output. Therefore, between time e and time f, the plate voltage remains at +100 volts and the waveform is distorted. This is also **AMPLITUDE DISTORTION**.

This has been an explanation of one cycle of an input signal that overdrives the tube. You should notice that, using the same circuit, a 50-volt peak-to-peak input signal caused a vastly different output from that caused by the 6-volt peak-to-peak input signal. The 6-volt peak-to-peak signal did not overdrive the tube. When the input signal was increased to 50-volts peak-to-peak, the tube was forced into cutoff when the grid was driven to -24 volts, and into saturation when the grid was driven to +12 volts (the grid voltage plus the signal voltage.) During these periods, the tube could not respond to the input signal. In other words, the output was distorted. A method commonly used to partially overcome distortion is to vary the bias voltage on the grid. The point at which the tube goes into cutoff or saturation can then be controlled.

For this reason tube biasing is of great importance in most tube circuits.

Q20. The waveforms shown below are the input and output of an overdriven triode.



a. Distortion A at the output is the result of what condition?
b. Distortion B at the output is the result of what condition?

TYPES OF BIASING

There are two main classes of biasing—**FIXED** and **SELF**. In a tube circuit that uses fixed bias, the grid-bias voltage is supplied from a power source external to the circuit. You are already familiar with battery bias, which is one form of fixed bias. When fixed bias is used in a circuit, it can be represented as either a battery (fig. 1-21, view A), or as a conductor connected to $-E_{cc}$ (fig. 1-21, view B). Fixed bias is rarely used in electronics today. Therefore, we will not discuss it further.

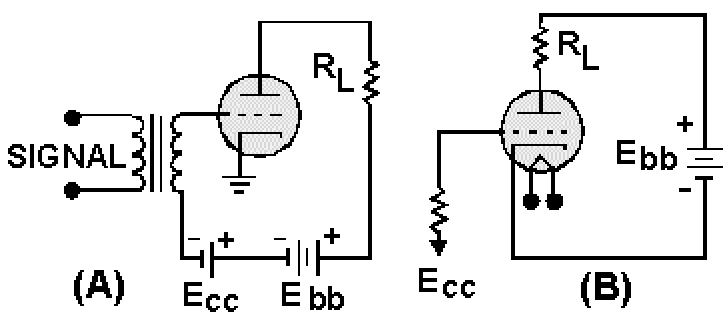


Figure 1-21.—Fixed bias: A. Battery B. Conductor

In circuits using self-bias, the bias voltage is developed across a resistor in the cathode or grid circuit by tube current. There are two main methods of self-bias: *cathode biasing* and *grid-leak biasing*.

Cathode Bias

In circuits using cathode bias, the cathode is made to go positive relative to the grid. The effect of this is the same as making the grid negative relative to the cathode. Because the biasing resistor is in the cathode leg of the circuit, the method is called **CATHODE BIASING** or **CATHODE BIAS**. A triode circuit using cathode bias is shown in figure 1-22.

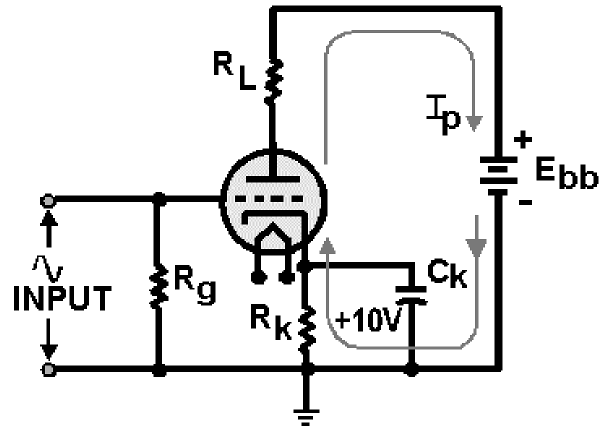


Figure 1-22.—Cathode bias.

The only difference between the illustrated circuit and the one used to demonstrate triode operation is the elimination of the battery, E_{cc}, and the addition of circuit components R_k, the cathode-biasing resistor; C_k, the cathode ac-bypass capacitor; and a grid resistor (whose purpose will be explained later).

When the tube conducts, current flows from the battery through R_k to the cathode, through the tube to the plate, and through R_L to the positive terminal of the battery. The current flowing through R_k will cause a voltage drop across R_k. The bottom of R_k goes negative while the top goes positive. This positive voltage at the top of R_k makes the cathode positive relative to the grid.

You may wonder what purpose C_k serves in this circuit. C_k serves as an **AC BYPASS**. Without C_k, the bias voltage will vary with ac input signals. This is particularly troublesome in the higher frequencies like those found in radio receivers. R_k, the cathode-biasing resistor, is used to develop the biasing voltage on the cathode.

The input signal will be developed across R_g. You will read more about the circuit component later in this chapter. Cathode-biasing voltage is developed in the following manner.

As we mentioned earlier, the bias voltage will vary with the input unless C_k, the cathode bypass capacitor, is used.

To understand how the bias voltage will vary with an ac input signal, disregard C_k for the moment and refer to figure 1-22 again.

Notice that under quiescent conditions, the voltage drop at the top of R_k is +10 volts. Now let's apply the positive-going signal illustrated to the left of the tube. When the positive signal is applied, conduction through the tube will increase. The only trouble is that current through R_k will also increase. This will increase the voltage drop across R_k, and the cathode voltage will now be greater than +10 volts. Remember, at this time the plate is going negative due to increased conduction through the tube. The combination of the negative-going plate and the positive-going cathode will decrease the electrostatic attraction across the tube and lower the conduction of the tube. This will reduce the gain of the tube.

When the negative-going signal is applied, conduction through the tube decreases. Current through R_k decreases and the voltage drop across R_k decreases. This causes the cathode to go more negative, which tends to increase conduction through the tube. A negative-going signal is amplified by decreasing plate current and allowing the plate to go positive (remember the 180º inversion.) Thus, increasing conduction on the negative half-cycle decreases the gain of that half-cycle. The overall effect of allowing cathode biasing to follow the input signal is to decrease the gain of the circuit with ac inputs.

This problem can be overcome by installing C_k. The purpose of C_k is to maintain the cathode bias voltage at a constant level. In common usage, the action of C_k is referred to as "bypassing the ac signal to ground."

The action of C_k will be explained using figure 1-23. View A shows the circuit under quiescent conditions. With some conduction through the tube, the cathode and the tops of R_k and C_k are at +10 volts.

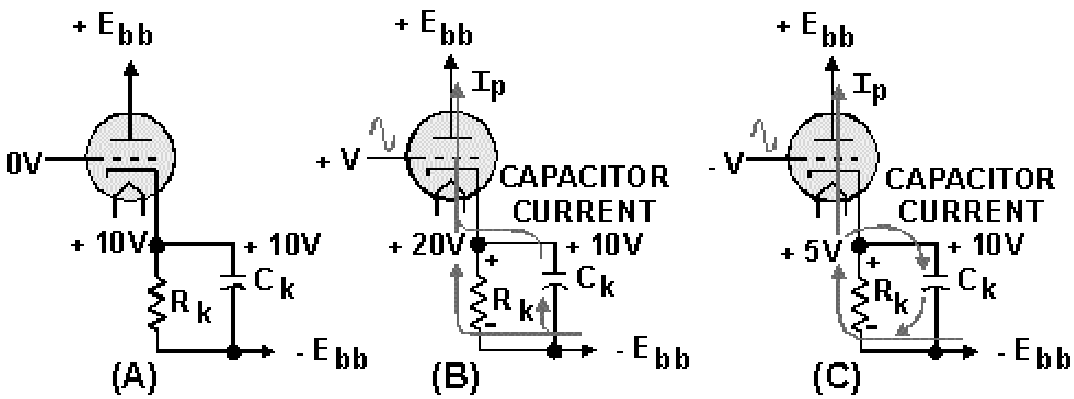


Figure 1-23.—Effect of the bypass capacitor.

In view B, the positive-going signal is applied to the grid. This causes increased conduction through the tube, which attempts to drive the cathode to +20 volts. But notice that the top of C_k is still at +10 volts (remember capacitors oppose a change in voltage). The top plate of C_k is, in effect, 10 volts negative in relation to the top of R_k. The only way that C_k can follow the signal on the top of R_k (+20 volts) is to charge through the tube back to the source, from the source to the lower plate of C_k. When C_k charges through the tube, it acts as the source of current for the cathode. This causes the cathode to remain at +10 volts while the capacitor is charging.

View C of the figure shows the same signal. Under these conditions, conduction through R_k will decrease. This will cause a decrease in current flow through R_k. Decreased current means decreased voltage drop. The top of R_k will try to go to +5 volts. C_k must now go more negative to follow the top of R_k. To do this, current must flow from C_k through R_k, to the top plate of C_k. This discharging of C_k will increase current flow through R_k and increase the voltage drop across R_k, forcing the top to go more positive. Remember, the voltage drop is due to current flow through the resistor. (The resistor could care less if the current is caused by conduction or capacitor action.) Thus, the cathode stays at +10 volts throughout the capacitor-charge cycle.

There is one point that we should make. C_k and R_k are in parallel. You learned from previous study that voltage in a parallel circuit is constant. Thus, it would seem impossible to have the top of R_k at one voltage while the top plate of C_k is at another. Remember, in electronics nothing happens instantaneously. There is always some time lag that may be measured in millionths or billionths of seconds. The action of C_k and R_k that was just described takes place within this time lag. To clarify the explanation, the voltages used at the components R_k and C_k were exaggerated. Long before a 10-volt differential could exist between the tops of R_k and C_k, C_k will act to eliminate this voltage differential.

The capacitor, then, can be said to regulate the current flow through the bias resistor. This action is considered as **BYPASSING** or eliminating the effect of the ac input signal in the cathode. For all practical purposes, you can assume that ac flows through the capacitor to ground. But, remember, ac only appears to flow across a capacitor. In reality the ac signal is shunted around the capacitor.

There are two disadvantages associated with cathode biasing. To maintain bias voltage continuously, current must flow through the tube, and plate voltage will never be able to reach the maximum value of the source voltage. This, in turn, limits the maximum positive output for a negative input signal (remember the 180º inversion). In addition, maximum plate voltage is decreased by the amount of cathode-biasing voltage. What this means is that you can't get something for nothing. If the cathode is biased at +20 volts, this voltage must be subtracted from the plate voltage. As an example, consider a triode with a 10,000 ohm plate resistor and a +300 volts dc source voltage. If a current of 2 milliamperes flows through the tube under quiescent conditions, 20 volts are dropped across the plate-load resistor. The maximum plate voltage is then 300 volts - 20 volts = 280 volts dc. Now, consider the 20-volt dropped across the cathode resistor. Plate voltage becomes 280 volts - 20 volts = 260 volts. To understand this a little more thoroughly, look at figure 1-24. In view A, the source voltage is 300 volts dc. There are two ways that this voltage can be looked at; either the plate is at +300 volts and the cathode is at 0 volts (ground), or the plate is at +150 volts and the cathode is at -150 volts. In electronics, it is common practice to assume that the plate is at +300 volts while the cathode is at 0 volts. To simplify this discussion, we will assume that the plate is at +150 volts, and the cathode is at -150 volts. The potential difference between the plate and the cathode is 300 volts. If a plate-load resistor is installed, as shown in view B, 20 volts are dropped by R_L. The potential difference between the plate and the cathode is now 280 volts. In view C, R_k has now been placed in the same circuit. R_k drops 20 volts. Therefore, the effect of cathode biasing is to reduce the maximum positive signal that the circuit can produce. In this case, the maximum positive signal has been reduced by 20 volts. Despite these disadvantages, cathode biasing has two main advantages. It is simple and economical.

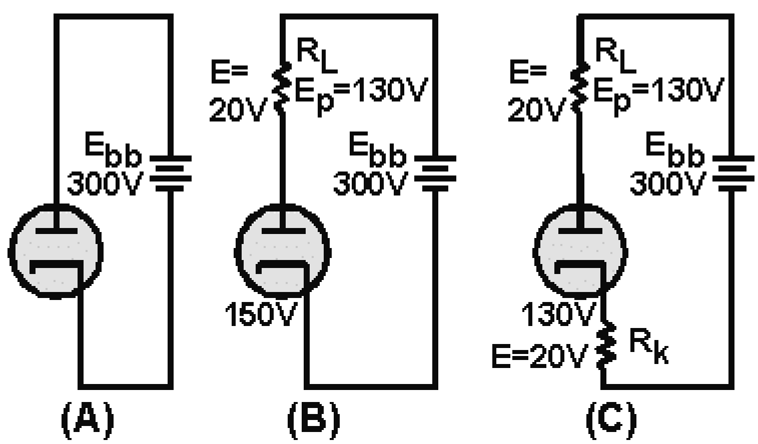


Figure 1-24.—Loss due to cathode biasing.

Grid-Leak Biasing

The second type of self-biasing to be discussed is **GRID-LEAK BIAS**. As the name implies, bias voltage is developed in the grid leg portion of the circuit. Bias voltage in this type of biasing is derived by allowing the positive input signal to draw grid current through a circuit made up of a resistor and a capacitor. There are two types of grid-leak bias commonly in use: **SHUNT TYPE** and **SERIES TYPE**. Because shunt type grid-leak biasing is the simplest, we will discuss it first. Figure 1-25 depicts a simplified triode circuit using the shunt-type grid-leak biasing. Before we begin the explanation of shunt grid-leak biasing, there is one thing you should bear in mind. Because the bias is derived from the positive input signal through capacitive action, the input signal must go through several positive alternations before the final operating bias voltage is achieved. We will explain why this is so in the following discussion.

View A of figure 1-25 shows the circuit under quiescent conditions. You will notice that the circuit is similar to the one we used to explain the action of a triode. The only additions are the grid resistor, R_g coupling capacitor, C_c, and resistance rgk. Resistance rgk doesn't exist as a physical component, but it is used to represent the internal tube resistance between the triode's cathode and grid. Electrically, rgk is quite small, about 500 ohms. Under quiescent conditions, some conduction occurs through the tube. Some electrons will strike the wires of the grid, and a small amount of **GRID CURRENT** will flow through R_g to ground. This will cause the right-hand plate of C_c to go slightly negative. This slight negative charge will, in turn, keep the grid of the tube slightly negative. This limits the number of electrons that strike the grid wires.

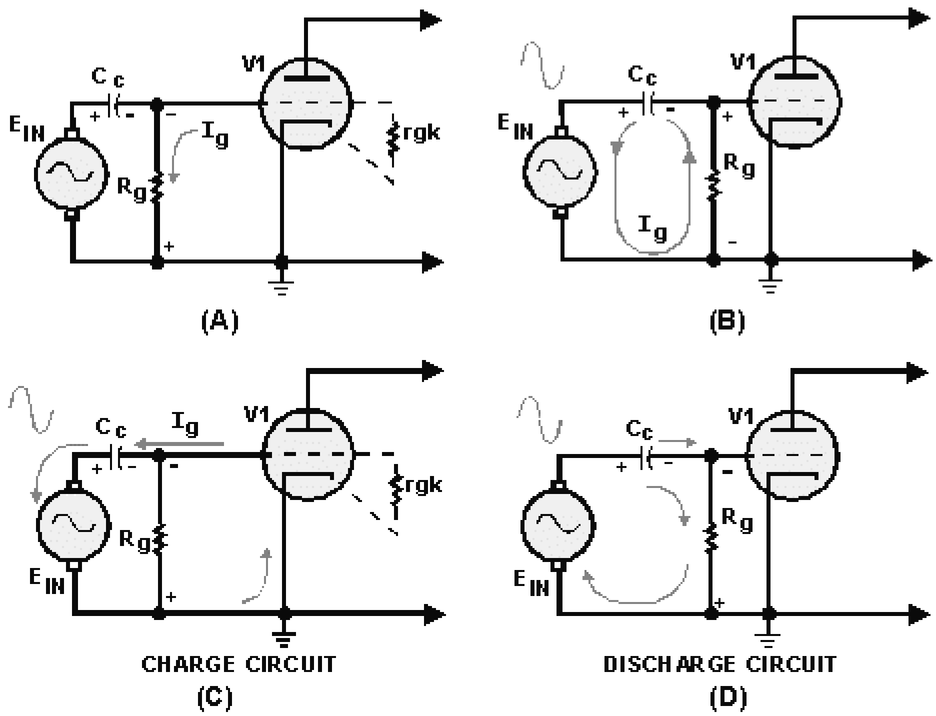


Figure 1-25.—Shunt grid-leak biasing.

In view B of the figure, the first positive alternation of a series of ac alternations, E_{in} is applied to the circuit. The positive-going voltage causes the left-hand plate of C_c to go positive. The left-hand plate must lose electrons to go positive. These electrons leave the left-hand plate of C_c and travel to the input source where they will be coupled to ground. From ground, current flows through R_g causing a negative (bottom) to positive (top) voltage drop across R_g. In effect, the ac signal has been coupled across the capacitor. Because of this, capacitors are said to pass the ac signal while blocking dc. (In reality, the ac signal is coupled around the capacitor.) In view C of the figure, the positive-going voltage at the top of R_g will be coupled to the grid causing the grid to go positive. The positively charged grid will attract electrons from the electron stream in the tube. Grid current will flow from the grid to the right-hand plate of C_c. This will cause the right-hand plate to go negative. (Electrostatic repulsion from the right-hand plate of C_c will force electrons from the left-hand plate of C_c, causing it to go positive.) The electrons will flow through the signal source, to ground, from ground to the cathode, from the cathode to the grid, and finally to the right-hand plate of C_c. This is the biasing charge cycle. You may wonder why the charge current went through the tube rather than through R_g. When the grid goes positive in response to the positive-going input signal, electrostatic attraction between the grid and cathode increases. This, in turn, reduces the resistance (rgk) between the grid and cathode. Current always follows the path of least resistance. Thus, the capacitor charge path is through the tube and not through R_g.

When the first negative alternation is applied to the circuit (view D), the left-hand plate of C_c must go negative. To do this, electrons are drawn from the right-hand plate. The electrons travel from the right-hand plate of C_c, through R_g causing a voltage drop negative (top) to positive (bottom), from the bottom of R_g, through the source, to the left-hand plate of C_c. C_c will discharge for the duration of the negative alternation. **BUT C_c can only discharge through R_g, which is a high-resistance path, compared to the charge path.** Remember from your study of capacitors that **RC** time constants and the rate of discharge increase with the size of **R**. C_c can therefore charge through the low resistance of rgk to its maximum negative value during the positive half-cycle. Because C_c discharges through R_g (the high resistance path), it cannot completely discharge during the duration of the negative half-cycle. As a result, at the completion of the negative alternation, C_c still retains part of the negative charge it gained during the positive alternation. When the next positive alternation starts, the right-hand plate of C_c will be more negative than when the first positive alternation started.

During the next cycle, the same process will be repeated, with C_c charging on the positive alternation and discharging a lesser amount during the negative alternation. Therefore, at the end of the second cycle, C_c will have an even larger negative charge than it did after the first cycle. You might think that the charge on C_c will continue to increase until the tube is forced into cutoff. This is not the case. As the negative charge on the right-hand plate of C_c forces the grid more negative, electrostatic attraction between the grid and cathode decreases. This, in effect, increases the resistance (rgk) between the cathode and the grid, until rgk becomes, in effect, the same size as R_g. At this point, charge and discharge of C_c will equal one another and the grid will remain at some negative, steady voltage. What has happened in this circuit is that C_c and R_g, through the use of unequal charge and discharge paths, have acted to change the ac input to a negative dc voltage. The extent of the bias on the grid will depend on three things: the amplitude of the input, the frequency of the input, and the size of R_g and C_c. This type of biasing has the advantage of being directly related to the amplitude of the input signal. If the amplitude increases, biasing increases in step with it. The main limiting factor is the amount of distortion that you may be willing to tolerate. Distortion occurs during the positive alternation when the grid draws current. Current drawn from the electron stream by the grid never reaches the plate; therefore the negative-going output is *not* a faithful reproduction of the input, while the positive-going output (during the negative input cycle) *will be* a faithful reproduction of the input. This is similar to the situation shown in the flattopped portion of the output signal in figure 1-20.

The **SERIES GRID-LEAK BIAS** circuit shown in figure 1-26 operates similarly to the shunt grid-leak circuit. When the first positive alternation is applied to the left-hand plate of the grid capacitor, C_g, the left-hand plate must lose electrons to go positive with the input. Electrons will leave the left-hand plate and flow through R_g, causing a negative (left-hand side) to positive (right-hand side) voltage drop. From the right-hand side of R_g, the electrons will flow to the right-hand plate of C_g. The positive voltage developed at the right-hand side of R_g will be coupled to the grid. As the grid goes positive, it will draw current, causing C_g to start to charge through the low resistance path of the tube. During the negative alternation of the input, C_g will discharge through the high resistance path of R_g. Once again it will not be completely discharged at the end of the negative alternation, and the capacitor will continue on its way toward charge equilibrium.

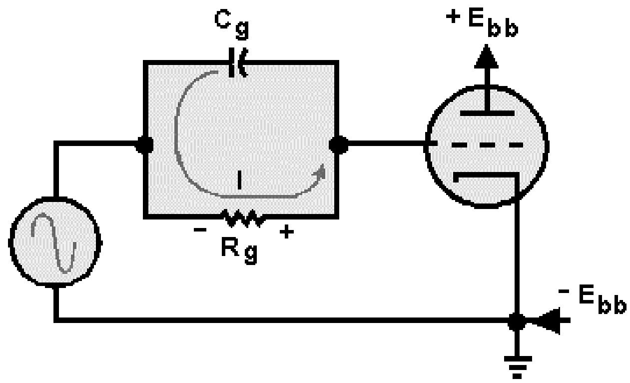


Figure 1-26.—Series grid-leak biasing.

In summary, grid-leak bias causes the grid to draw current when the input signal goes positive. This grid current (which is a negative charge) is stored by the coupling capacitor (C_c,) which will keep the grid at some negative potential. It is this potential that biases the tube.

Q21. What type of bias requires constant current flow through the cathode circuit of a triode?

Q22. When a circuit uses cathode biasing, the input signal can cause variations in the biasing level How is this problem eliminated?

Q23. In a circuit using grid-leak biasing, the coupling capacitor (C_c) charges through a low resistance path. What resistance is used in this charge path?

Q24. Grid-leak biasing in effect rectifies the input ac signal. What feature of the circuit is used to accomplish this rectification?

OPERATING CLASSIFICATIONS OF TUBE AMPLIFIERS

While the discussion of amplifiers will be covered in detail in later *NEETS* modules, some discussion of the classes of operation of an amplifier is needed at this point. This is because their operation class is directly determined by the bias voltage of the tube.

The classification of amplifiers by operation is based on the percentage of the time that the tube conducts when an input signal is applied. Under this system amplifiers may be divided into four main classes: A, AB, B, and C.

CLASS A OPERATION

An amplifier biased into Class A operation, is one in which conduction through the tube occurs throughout the duration of the input signal. Such an amplifier is shown in figure 1-27, view A. This is the same type of circuit with which you are already familiar. Notice when you compare the input to the output that the tube is always conducting, and that the entire input signal is reproduced at the output.

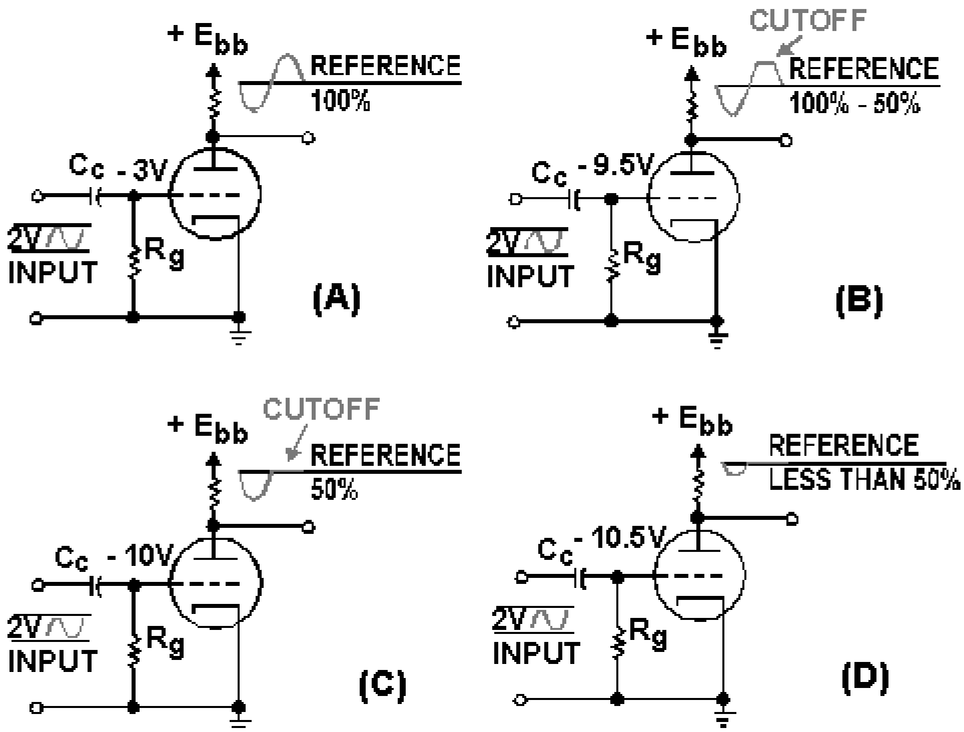


Figure 1-27.—Classes of amplifier operation.

CLASS AB OPERATION

The Class AB amplifier is one in which the tube conducts for more than half, but less than the entire input cycle.

View B of figure 1-27 depicts an amplifier biased into **CLASS AB** operation. Notice that in this application, grid bias has been increased to -9 volts. We will assume that the tube reaches cutoff when the voltage on the grid is -10 volts. Under these conditions, when the input reaches -10 volts, the tube will cut off and stay cut off until the input goes above -10 volts. The tube conducts during the entire duration of the positive alternation and part of the negative alternation. If you remember back in the discussion of distortion, we pointed out that this represents distortion. In some amplifiers, faithful reproduction of the input is not an important requirement. Class AB amplifiers are used only where this distortion can be tolerated.

CLASS B OPERATION

A **CLASS B** biased amplifier is one in which the tube will conduct for only half of the input signal duration. This is done by simply biasing the amplifier at cutoff. View C of figure 1-27 depicts a class B biased amplifier.

As you can see, the tube conducts on the positive alternations. As soon as the input signal voltage reaches 0 volts, the tube cuts off. The tube will remain cut off until the input signal voltage climbs above zero volts on the next positive alternation. Because the tube conducts during the entire positive alternation, but not on the negative alternation, the tube conducts for only half the input cycle duration.

CLASS C

CLASS C amplifiers are biased below cutoff, so that the tube will conduct for less than half of the input signal cycle duration. View D of figure 1-27 depicts a Class C amplifier. Notice that the tube is biased one-half volt below cutoff. The tube will only conduct on that part of the positive alternation that is above +.5 volts. Therefore, the tube conducts for less than one-half cycle of the input. Again, this class can be applied only where severe distortion can be tolerated.

TUBE CONSTANTS

In the discussion of triodes, we only considered the effects of the external circuit on the passage of current through the tube. The behavior of the electron stream in a conducting tube is also influenced by the physical structure of the tube. The effects that the physical structure of a tube has on the tube's operation are collectively called **TUBE CONSTANTS**. Four of the most important of these tube constants are: **TRANSIENT TIME, INTERELECTRODE CAPACITANCE**, **TRANSCONDUCTANCE**, and **AMPLIFICATION FACTOR.**

TRANSIT TIME

Unlike electron flow in a conductor, electrons in a vacuum tube do not move at the speed of light. Their velocity is determined by the potential difference between the plate and the cathode. The amount of time the electrons take to travel from the cathode to the plate is called **TRANSIT TIME**. As a result of this time difference, the appearance of a signal at the end of a tube is not followed instantaneously by a change in current flow in the tube. Under normal conditions, the effect of this small time lag between the input signal and a change in tube current is unnoticed. However, at frequencies such as those used in radar equipment, this is not the case. Transit time at these frequencies has a very marked effect on tube operation. It is a major factor that limits the use of a given tube at higher frequencies.

Q25. Match each amplifier characteristic listed below with its class of amplification.

a. Current flows through the tube for one-half cycle.

b. Current flows through the tube for less than one-half cycle.

c. Current flows through the tube for the entire cycle.

MU AND TRANSCONDUCTANCE

In your study of triodes so far, you have seen that the output of a triode circuit is developed across the tube. The output is caused by the voltage dropped across R_L due to current flow from tube conduction. In all the demonstrations of gain, we assumed that R_L was held constant and current through the tube was varied. In this manner we achieved a voltage gain. If the resistance of R_L is changed by the designer, the gain of a triode circuit can be either increased or decreased. This is fairly easy to understand. Assume that a circuit is composed of a triode with a plate-load resistor of 100 kohms. If a +2 volt signal causes 2 additional milliamperes to conduct through the tube, the voltage drop across R_L (the output) will be:

$$E = I \times R$$
$$E = I_P \times R_L$$
$$E = (2 \times 10^{-3} \text{ amperes}) \times (100 \times 10^3 \text{ohms})$$
$$E = 2 \times 100 \text{ volts}$$
$$E = 200 \text{ volts}$$

$$\text{Gain} = \frac{\text{output}}{\text{input}}$$
$$\text{Gain} = \frac{200 \text{ volts}}{2 \text{ volts}}$$
$$\text{Gain} = 100$$

Thus, the gain of the circuit is 100. If the plate-load resistor is reduced to 50 kohms and the input is kept at +2 volts, the gain will be reduced to:

$$E = (2 \times 10^{-3} \text{ amperes}) \times (50 \times 10^3 \text{ohms})$$
$$E = 2 \times 50 \text{ volts}$$
$$E = 100 \text{ volts}$$

$$\text{Gain} = \frac{100 \text{ volts}}{2 \text{ volts}}$$
$$\text{Gain} = 50$$

As you can see, voltage gain depends on both the tube characteristics and the external circuit design.

The voltage gain is a measure of circuit efficiency, not tube efficiency.

The actual characteristics of a tube are measured by two factors: mu(μ) or **AMPLIFICATION FACTOR**; and **TRANSCONDUCTANCE** or g_m. The amplification factor (represented as μ) of a tube is equal to the ratio of a change in plate voltage to the change in grid voltage required to cause the same change in plate current. This is expressed mathematically as

$$\mu = \frac{\Delta E_P}{\Delta E_g}$$

While this may sound complicated, it really isn't. Look at figure 1-28. Here you see in view A a triode with a +1 volt input signal. At this grid voltage, current through the tube is at 1 milliampere. If the input voltage is raised to +3 volts, current through the tube increases to 2 milliamperes. The change in E_g (ΔE_g) is then 2 volts. This is shown in view B. Suppose that the grid voltage is returned to +1 volt, and the plate voltage is increased until the ammeter in view C reads 2 milliamperes of plate current. At this point plate voltage is measured. Plate voltage had to be increased by 100 volts (350-250) to get the same change in plate current (1 mA). The change in plate voltages (ΔE_p) is then 100 volts. The amplification factor (μ) of just the tube is then equal to

$$\mu = \frac{\Delta E_p}{\Delta E_g}$$

$$\mu = \frac{100 \text{ volts}}{2 \text{ volts}}$$

$$\mu = 50$$

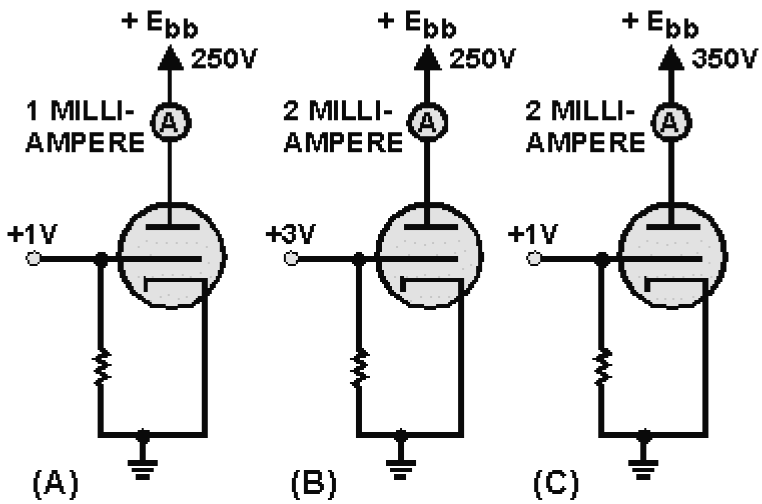


Figure 1-28.—Obtaining gain and transconductance.

As you can see, mu is a measure of the ability of a tube to amplify. By comparing the mu of two different types of tubes, you can get an idea of their efficiency. For example, assume you have two different tubes, one with a mu of 50, and the other with a mu of 100. If you place each tube in a circuit whose input varies by 2 volts, you can expect the following changes in plate voltage.

Tube 1:

$$\mu = \frac{\Delta E_p}{\Delta E_g}$$

$$\Delta E_p = \mu \times \Delta E_g$$

$$\Delta E_p = 50 \times 2 \text{ volts}$$

$$\Delta E_p = 100 \text{ volts}$$

Tube 2:

$$\mu = \mu \times \frac{\Delta E_p}{\Delta E_g}$$

$$\Delta E_p = \mu \times \Delta E_g$$

$$\Delta E_p = 100 \times 2 \text{ volts}$$

$$\Delta E_p = 200 \text{ volts}$$

Thus, you can expect twice the change in plate voltage from tube 2 as from tube 1 for the same input voltage. Therefore, tube 2 will have twice the gain of tube 1.

Transconductance

Transconductance is a measure of the change in plate current to a change in grid voltage, with plate voltage held constant. The unit for conductance is the mho (siemens), pronounced "moe." Transconductance is normally expressed in either micromhos or millimhos. Mathematically, transconductance is expressed by the formula:

$$gm = \frac{\Delta I_p}{\Delta E_g}$$

Examine figure 1-28, views A and B, again. In view A, the input voltage is +1 volt. At +1 volt E_g, the plate current is equal to 1 milliampere, with a plate voltage of 250 volts. In view B, the input voltage (E_g) is raised to +3 volts. ΔE_g, as before, is equal to 2 volts. This increase in grid voltage causes plate current to increase to 2 milliamperes. The change in plate current (ΔI_p) is then equal to 1 milliampere. Thus, transconductance (g_m) is equal to:

$$gm = \frac{\Delta I_p}{\Delta E_g}$$

$$gm = \frac{1 \text{ milliampere}}{2 \text{ volts}}$$

$$gm = .5 \text{ millimho}$$
$$\text{or } 500\ \mu\text{mhos}$$

Remember that the voltage gain of a circuit is measured by the ratio of the change in plate voltage to the change in grid voltage. Because plate voltage is developed across a resistor, the more current varies with a given input signal, the greater will be the output ($E = I \times R$). If you have two tubes, one with a gm of 500 mhos and the other with a gm of 500 µmhos, you can estimate the relative gain of these two tubes. Assume that the circuit in which you wish to use one of these tubes has a load resistor of 100 kohms and that ΔE_g will be 2 volts. The voltage gain of these two circuits will be:

Tube 1:

$$gm = \frac{\Delta I_p}{\Delta E_g}$$

$$\Delta I_p = \Delta E_g \times gm$$

$$\Delta I_p = 2 \times .5 \text{ millmho}$$

$$\Delta I_p = 1 \text{milliampere}$$

$$\Delta E_p = \Delta I_p \times R_L$$

$$\Delta E_p = (1 \times 10^{-3} \text{ ampere}) \times (100 \times 10^3 \text{ ohms})$$

$$\Delta E_p = 100 \text{ volts}$$

$$\text{Gain} = \frac{\text{output}}{\text{input}}$$

$$\text{Gain} = \frac{100 \text{ volts}}{2 \text{ volts}} = 50$$

Tube 2:

$$gm = \frac{\Delta I_p}{\Delta E_g}$$

$$\Delta I_p = \Delta E_g \times gm$$

$$\Delta I_p = 2 \text{ volts} \times 5 \text{ millmho}$$

$$\Delta I_p = 10 \text{ milliampere}$$

$$\Delta E_p = \Delta I_p \times R_L$$

$$\Delta E_p = (10 \times 10^{-3} \text{ ampere}) \times (100 \times 10^3 \text{ ohms})$$

$$\Delta E_p = 1000 \text{ volts}$$

$$\text{Gain} = \frac{\text{output}}{\text{input}}$$

$$\text{Gain} = \frac{1000 \text{ volts}}{2 \text{ volts}} = 500$$

As you can see, tube 2 is 10 times the amplifier that tube 1 is.

Q26. The plate voltage of a tube will vary 126 volts when a 3-volt ac signal is applied to the control grid. What is the gain of this tube?

Q27. If the mu of a tube is 85 and the signal at the control grid is 4 volts ac, the plate voltage will vary by what amount?

Q28. Transconductance is a measure of the relationship between what two factors?

Q29. A tube has a transconductance of 800 mhos and a load resistor of 50 kohms. When control grid voltage varies by 6 volts, the plate voltage will vary by what amount?

INTERELECTRODE CAPACITANCE

As you know, capacitance exists when two pieces of metal are separated by a dielectric.

You should also remember from your studies that a vacuum has a dielectric constant of 1. As the elements of the triode are made of metal and are separated by a dielectric, capacitance exists between them. This capacitance is called *interelectrode capacitance*, and is schematically represented in figure 1-29.

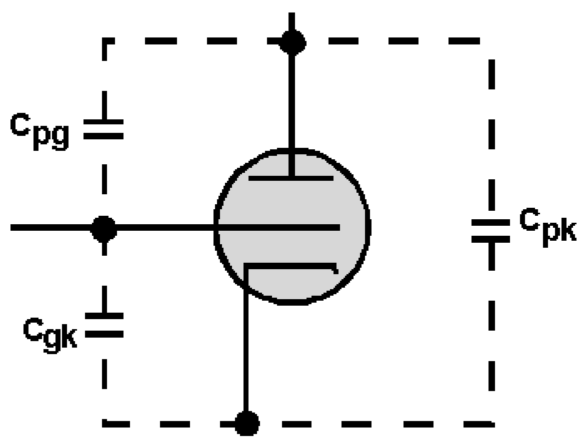


Figure 1-29.—Schematic representation of interelectrode capacitance.

Notice that there are three interelectrode capacitances involved in a triode. The capacitance between the plate and grid, designated C_{pg}, is the largest, because of the relatively large area of the plate, and therefore has the greatest effect on triode operation. The grid-to-cathode capacitance is designated C_{kg}.

The total capacitance across the tube is designated C_{pk}.

As we said earlier, C_{pg} has the greatest effect on the tube operation. This is because this capacitance will couple part of the ac signal from the plate back to the grid of the tube. The process of coupling the output of a circuit back to the input is called **FEEDBACK**. This feedback affects the gain of the stage. It may be desirable in some applications. In others, the effects must be neutralized. The effects of C_{pk} are greater at higher frequencies where X_c is lower.

DEVELOPMENT OF THE TETRODE

Interelectrode capacitance cannot be eliminated from vacuum tubes, but it can be reduced. The easiest method found to reduce interelectrode capacitance is to split the capacitance between the grid and plate (C_{pg}) into two capacitors connected in series. This is done by placing an extra grid, called the **SCREEN GRID**, between the control grid and the plate. This is shown in figure 1-30.

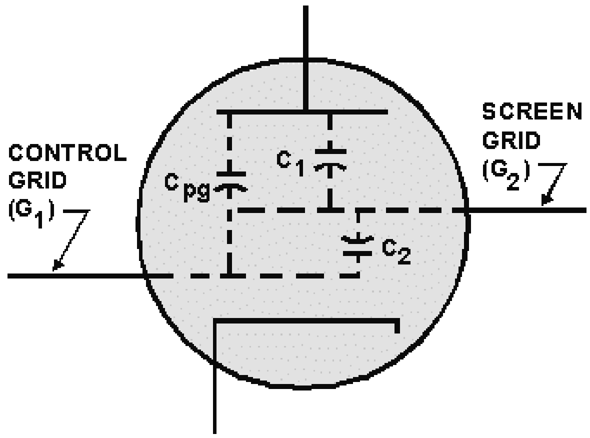


Figure 1-30.—Effect of the screen-grid on Interelectrode capacitance.

Remember from your study of capacitance that connecting capacitors in series reduces the total capacitance to a value smaller than either of the capacitors. This is mathematically summed up as follows:

$$CT = \frac{C1 \times C2}{C1 + C2}$$

The addition of the screen grid has the effect of splitting C_{pg} into two capacitances (C1 and C2) connected in series. Therefore, the total interelectrode capacitance between the control grid and the plate is greatly reduced.

OPERATION OF THE BASIC TETRODE CIRCUIT

Figure 1-31 depicts a basic tetrode circuit. While the circuit may look complicated, it isn't. You are already familiar with most of the circuit. Only three components have been added: the screen grid, the screen grid dropping resistor, and the screen grid bypass capacitor (C_{sg}).

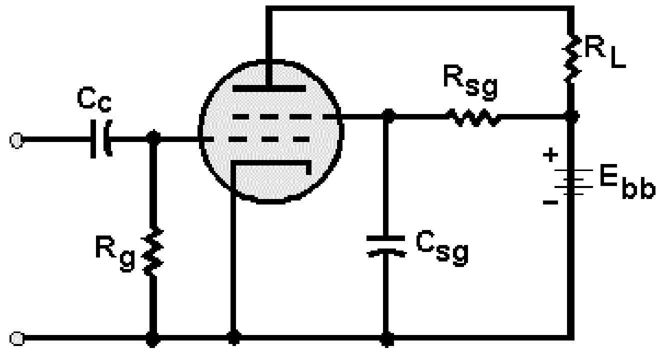


Figure 1-31.—Basic tetrode circuit.

The problem now is: at what voltage and polarity should the screen grid be operated? If the screen grid were operated at a potential that would make it negative in relation to the control grid, it would act as a negative screen between the plate and control grid. As a result, gain would be reduced. If the screen grid were operated at plate potential, it would draw current from the electron stream when the tube conducts. Because of this, the value of R_{sg} is normally selected to cause the screen grid to be positive in relation to the control grid, but not as positive as the plate.

Despite this precaution, the screen grid still draws some current from the electron stream. Any signal applied to the control grid will appear at the screen grid inverted by 180º from the input signal. This is undesirable, as it reduces the gain of the tube. Consider the effect if the control grid goes positive. Conduction through the tube increases, and since the screen grid is in the electron stream, it will draw some current. This causes the screen grid to go toward negative potential (less positive). The effect then, is to place a negative-going electrode between the plate and positive-going control grid. The plate becomes partially screened by the negative-going screen grid, and again, gain will decrease. Because the signal at the screen grid is always 180º out of phase with the control grid, its effect will always be to oppose the effect of the control grid.

To overcome this, a bypass capacitor (C_{sg}) is connected between the screen grid and ground. The addition of this capacitor will shunt, or pass, the ac variations on the screen grid to ground while maintaining a steady dc potential on the screen grid. In other words, C_{sg} moves all of the undesired effects mentioned in the previous paragraph.

One very useful characteristic of the tetrode tube is the relationship between the plate and screen grid. The screen grid will lessen the effect that a decreasing plate potential (negative-going signal) has on conduction through the tube. In a triode, when the grid goes positive, the plate goes negative. This decreases electrostatic attraction across the tube and tends to decrease the gain somewhat.

In a tetrode, the screen grid has the ability to isolate the effect that ac variations on the plate have on the electron stream.

The positively charged screen grid will accelerate electrons across the tube even though the plate is negative going. As long as the plate remains positive in relation to the cathode, it will draw off these accelerated electrons. As a result, conduction through the tube when the plate is going negative will not be decreased. This is another big advantage of screen-grid tubes.

TETRODE CHARACTERISTICS

Because the screen grid is in the electron stream, it will always draw some current. The current drawn by the screen grid will be lost to the plate. This means that the transconductance of a tetrode, which is based on the amount of plate current versus control-grid voltage, will be lower in tetrodes than in triodes. The formula for transconductance of a triode,

$$gm = \frac{\Delta I_p}{\Delta E_g}$$

must be adjusted for screen-grid current, and becomes

$$gm = \frac{\Delta I_p - \Delta I_{sg}}{\Delta E_g}$$

As you can see, the transconductance for a tetrode can never be as high as that of a triode of similar construction.

While lowered transconductance in a tetrode is an undesirable characteristic, it is not the reason that tetrodes have found little acceptance in electronics. The factor that severely limits the operation of tetrodes is **SECONDARY EMISSION**.

Because the screen grid is positively charged, electrons traveling from the cathode to the plate are accelerated. Electrons are accelerated to such an extent that they dislodge electrons from the plate when they strike it. This is similar to the manner in which a high-velocity rifle bullet fired into a pile of sawdust throws sawdust about. Some of these electrons are fired back into the tube, where they tend to accumulate between the screen grid and the plate. This effect is most pronounced when the signal at the control grid is going positive and conduction through the tube is increasing. The plate in this situation is going negative in answer to the control-grid signal. This causes the electrons accumulating between the plate and screen grid to be attracted to the screen grid. The current that is drawn by the screen grid is lost to the plate and gain is decreased. Gain is also decreased in another way. The negative charge accumulated by secondary emission causes some of the electrons (from the cathode) to be repelled from the plate, which further reduces gain.

Another undesirable characteristic of tetrodes associated with secondary emission is that the outputs are **NOISY**. What this means is that small sporadic signals appear on the main output signal, as shown in figure 1-32. When electrons are knocked from the plate, they represent losses of plate current and corresponding positive pulses on the output. Electrons falling back to the plate represent increases in plate current and cause negative-going pulses to appear in the output.

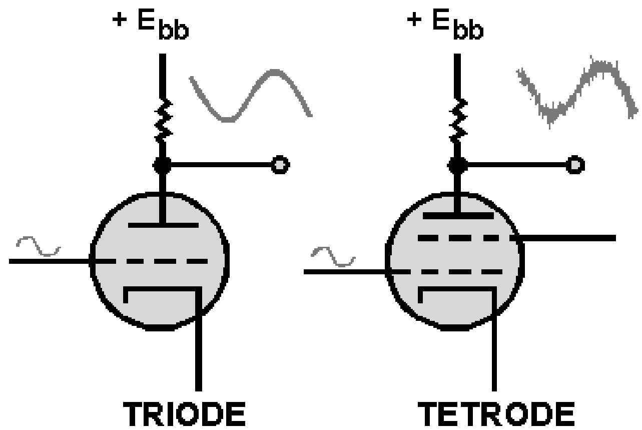


Figure 1-32.—Noise in a tetrode circuit.

For these reasons tetrodes are only used in very specialized applications of electronics.

Q30. How does the addition of a screen grid in a tetrode reduce interelectrode capacitance?

Q31. What undesirable effect does the screen grid in a tetrode create?

THE PENTODE

The problem of secondary emission associated with the screen grid of a tetrode has been reduced by—you guessed it, the addition of another grid.

This third grid, called a **SUPPRESSOR GRID**, is placed between the screen grid and the plate. The suppressor grid is normally connected either internally or externally to the cathode and bears the same charge as the cathode. This is shown in figure 1-33. Because of its negative potential (relative to individual electrons), any electrons that are emitted by the plate, through secondary emission, are repelled back toward the plate.

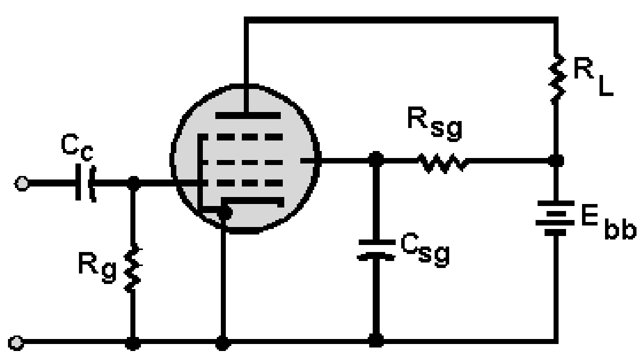


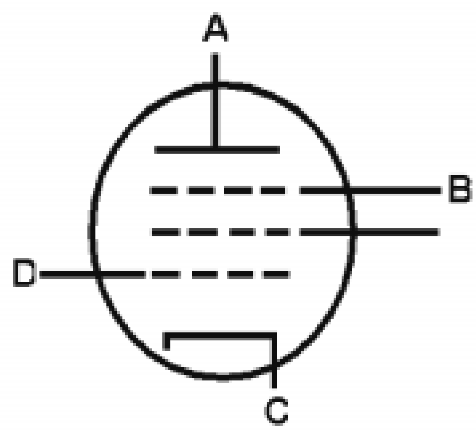
Figure 1-33.—Basic pentode circuit.

You might think that a grid with a negative potential placed close to the plate would interfere with the electron stream. However, this is not the case. Because the suppressor grid is negatively charged, it will not draw grid current. Additionally, the wide spacing within the mesh of the suppressor and its location between two positive elements of the tube ensures that the suppressor grid's effect on the electron stream will be minimum. Only the electrons emitted by secondary emission from the plate are affected by the suppressor grid.

Because pentodes do not suffer from secondary emission, they have replaced the tetrode in most applications.

Q32. The suppressor grid is added to a tetrode to reduce what undesirable characteristic of tetrode operation?

Q33. On the diagram below, name the elements of the vacuum tube and their potentials relative to dc ground.

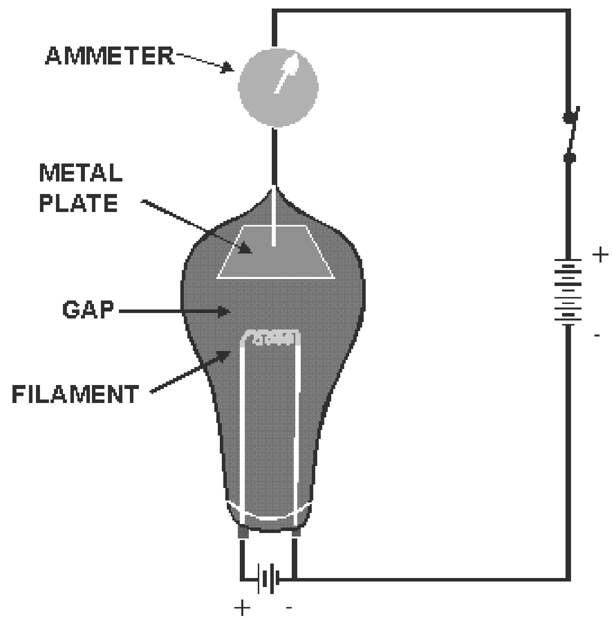


SUMMARY

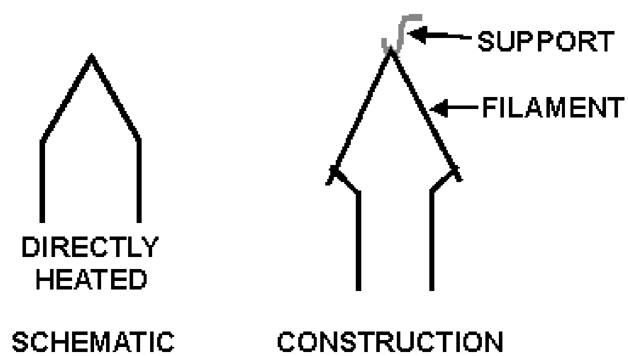
This chapter has introduced you to the four basic types of vacuum tubes. The following is a summary of the main points of the chapter.

THERMIONIC EMISSION is caused when metallic substances are heated to high temperatures. Electrons liberated by thermonic emission provide the conduction currents of vacuum tubes.

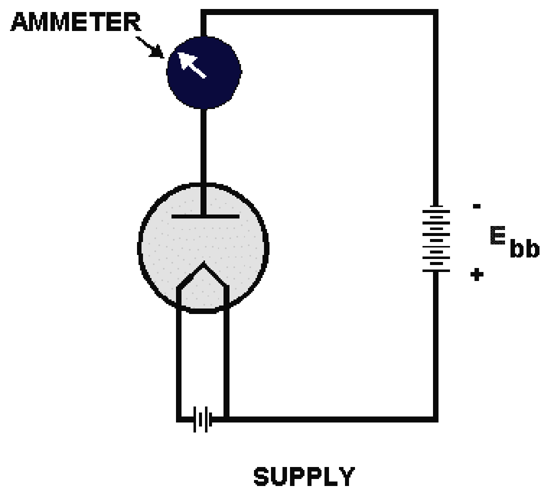
A **DIODE VACUUM TUBE** is composed of two elements: the cathode and the plate.



The **CATHODE** is the electron-emitting element of a tube. Cathodes are usually composed of special materials that are heated either directly or indirectly.



DIODE OPERATION depends upon current flow through the tube. Because the cathode is the only electron-emitting element in the tube, current can only flow in one direction, from the cathode to the plate. For current to flow, the plate must be positive relative to the cathode. When the plate is negative relative to the cathode, current cannot flow within the tube.

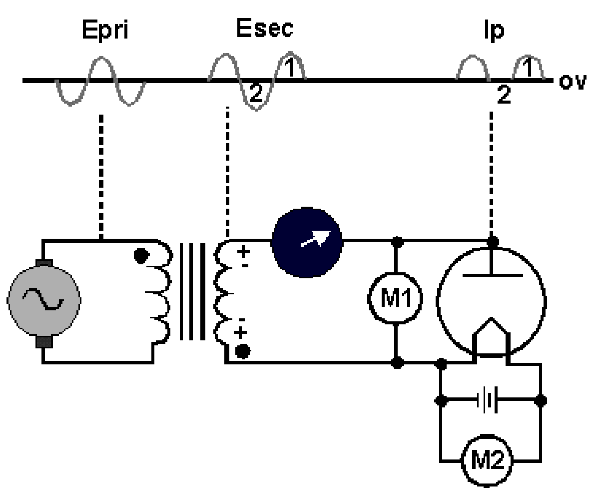


The **CHARACTERISTIC CURVE** for an electron tube is a graphic plot of plate current (I_p) versus plate voltage (E_p). From this, dc plate resistance can be computed by the formula:

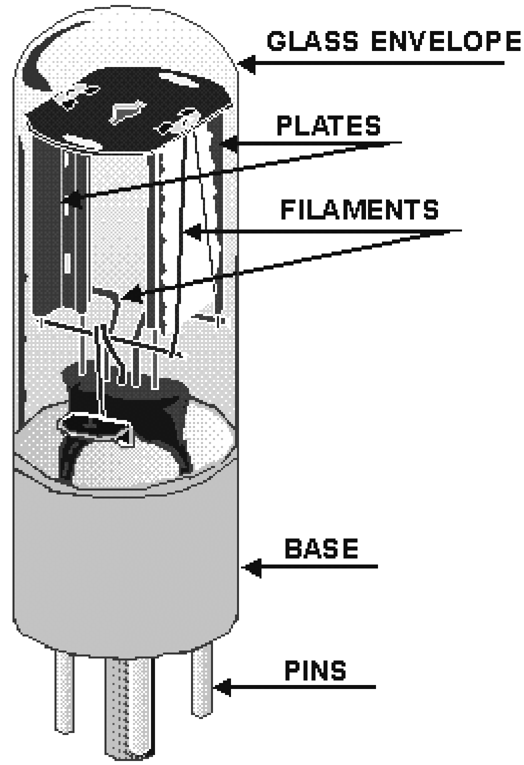
$$R_p = \frac{E_p}{I_p}$$

FACTORS THAT LIMIT VACUUM TUBE OPERATION are plate dissipation, maximum average current, maximum peak-plate current, and peak-inverse voltage.

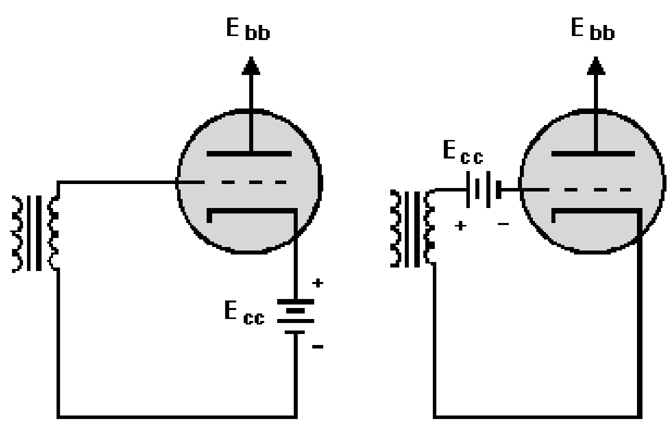
DIODE RECTIFIERS take advantage of the fact that diodes will conduct in only one direction. When ac voltages are applied to diodes, conduction occurs only on the alternation that makes the plate positive relative to the cathode. Because of this, the output current consists of one polarity. Because it flows in pulses rather than continuously, it is called pulsating dc.



DIODE CONSTRUCTION is the basic construction plan of most vacuum tubes. The tube is constructed of the following parts: filament and/or cathodes, plates, envelope, and base.

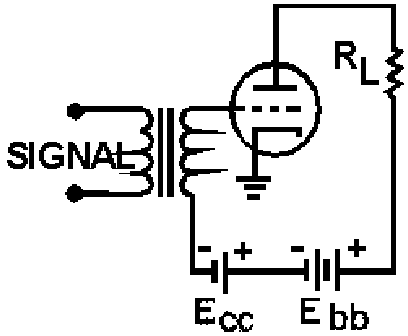


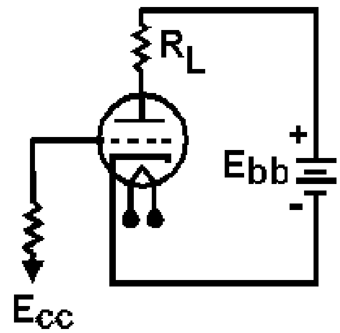
A **TRIODE** is basically a diode with a control grid mounted between the plate and the cathode. The control grid gives the triode the ability to amplify signals.

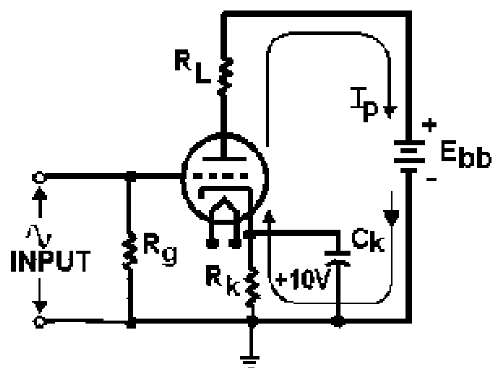


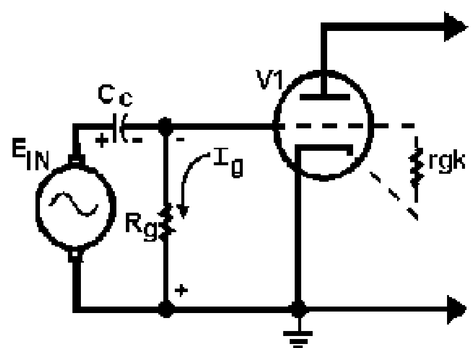
The **OPERATION OF A TRIODE** depends on the ability of the control grid to either increase or decrease conduction through the tube in response to an ac input signal. The output voltage is developed across the tube between the cathode and plate because of the voltage drop across the plate-load resistor changing as the plate current responds to the input signal.

TUBE BIASING is the process of placing a dc voltage, usually negative, on the grid. Bias has several functions in circuit design. Biasing may be divided into two types: fixed and self. Tubes using fixed bias have a dc voltage applied to their control grids from an external source such as a battery. Self-biasing voltages, on the other hand, are derived from current conducting through the tube. The most common types of self-biasing are cathode biasing and grid-leak biasing.





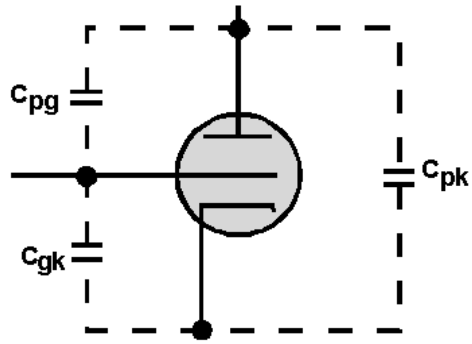




The **CLASS OF OPERATION OF AN AMPLIFIER** is determined by the bias applied to a triode. An amplifier operating as class A conducts continually through the duration of the input cycle. Class AB operation occurs when the amplifier conducts for more than half but less than the entire duration of the input cycle. A class B amplifier conducts for only 50% of the input cycle. The class C amplifier conducts for less than half of the input cycle.

TRANSIT TIME is the time required for electrons emitted by the cathode to reach the plate. Because transit time in a vacuum tube is considerably less than the speed of light, vacuum tube operation is affected at high frequencies.

INTERELECTRODE CAPACITANCE is created by the naturally occurring capacitance between elements in a vacuum tube. One effect of interelectrode capacitance is to feed back a portion of the output of a triode to the input. This effect is a prime-limiting factor in applying triodes. It is a major reason why triodes are seldom used—especially at the higher frequencies.



MU AND TRANSCONDUCTANCE are measures of tube efficiency. Mu (μ), or amplification factor, is a measure of the amount that plate voltage varies in relation to variation of the input voltage. Mathematically, mu (μ) is expressed as

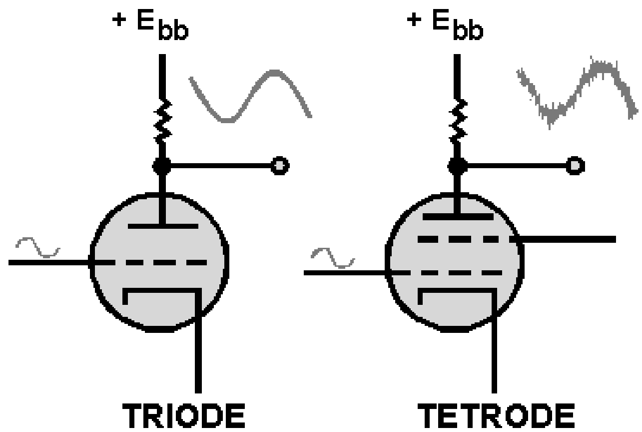
$$\mu = \frac{\Delta E_p}{\Delta E_g}$$

TRANSCONDUCTANCE, on the other hand, is a measure of the amount of variation of plate current caused by a variation of the input signal. Mathematically, it is expressed as:

$$gm = \frac{\Delta I_p}{E_g}$$

TETRODES were developed to compensate for the effects of interelectrode capacitance. Placing a positively charged screen grid between the control grid and plate has the effect of adding a capacitor in series with the capacitance that exists between the control grid and plate. This reduces total capacitance below the value of either capacitor as shown by applying the formula:

$$CT = \frac{C1 \times C2}{C1 + C2}$$



SECONDARY EMISSION of electrons from the plate is caused by the acceleration of electrons by the screen grid. This causes the performance of a tetrode to be degraded. In addition to reduced amplitude, the output signals become noisy.

PENTODES do not suffer from the effects of secondary emission. This is because a negatively charged suppression grid placed between the screen grid and plate forces any electrons emitted back to the plate.

ANSWERS TO QUESTIONS Q1. THROUGH Q33.

A1. By heating it.

A2. Because the negatively charged electrons are attracted to the positively charged plate.

A3. Filament and plate.

A4. Negative.

A5. Positive.

A6. Pulsating dc.

A7. Thoriated-tungsten and oxide-coated metals.

A8. They reach operating temperatures quickly.

A9. It serves as a mounting for the tube elements and as the terminal connection to the circuit.

A10. The linear portion.

A11. Plate resistant R_p.

A12. Peak Inverse Voltage (PIV).

A13. The triode contains a third element called the control rid.

A14. Because it is closer to the cathode.

A15. *A plate load resistor R_L*

A16. *To prevent them from drawing grid current.*

A17. *The input signal*

A18. *+275 volts.*

A19.

- *a.* *100 volts.*
- *b.* *180º out of phase.*

A20.

- *a.* *Cutoff.*
- *b.* *Saturation.*

A21. *Cathode biasing.*

A22. *Through the use of a bypass capacitor*

A23. *rkg, the cathode to grid resistance.*

A24. *Unequal charge and discharge paths of the coupling capacitor C_c.*

A25.

- *a.* *Class B.*
- *b.* *Class C*
- *c.* *Class A.*

A26. *42.*

A27. *340 volts.*

A28. *The changes in plate current and grid voltage.*

A29. *240 volts.*

A30. *The interelectrode capacitance (cpg) is divided between two series capacitances; thus, cpg is greatly reduced.*

A31. *Secondary emission, and noise.*

A32. *Secondary emission.*

A33.

- *a.* *Plate, positive.*
- *b.* *Suppressor grid, negative.*
- *c.* *Cathode, can be negative, positive, or at dc ground potential, depending on biasing type.*
- *d.* *Control grid, negative.*

CHAPTER 2

SPECIAL-PURPOSE TUBES

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to:

1. Determine the number and type of individual tubes contained within the signal envelope of a multi-unit tube.
2. Explain the function and operating principle of the beam power tube and the pentode tube.
3. State the difference between the capabilities of conventional tubes and variable-mu tubes.
4. Describe the construction of uhf tubes, and explain the effects that ultra-high frequencies have on conventional-tube operation.
5. Explain the operation of gas-filled diodes, thyratrons, and cold-cathode tubes.
6. Explain the operating principles behind cathode-ray tubes, and the manner in which these tubes present visual display of electronic signals.

INTRODUCTION TO SPECIAL-PURPOSE TUBES

Because of their great versatility, the four basic tube types (diode, triode, tetrode, and pentode) covered in chapter 1 have been used in the majority of electronic circuits. However, these types of tubes do have limits, size, frequency, and power handling capabilities.

Special-purpose tubes are designed to operate or perform functions beyond the capabilities of the basic tube types discussed in chapter 1. The special-purpose tubes covered in this chapter will include multi-unit, multi-electrode, beam power, power pentode, variable-mu, uhf, cold cathode, thyratrons, and cathode-ray tubes.

MULTI-UNIT AND MULTI-ELECTRODE TUBES

One of the probems associated with electron tubes is that they are bulky. The size of an electron tube circuit can be decreased by enclosing more than one tube within a single envelope, as mentioned in chapter 1. There is a large variety of tubes that can be combined into this grouping of "specialty tubes" called **MULTI-UNIT** tubes. Figure 2-1 illustrates the schematic symbols of a few of the possible combinations found in multi-unit tubes.

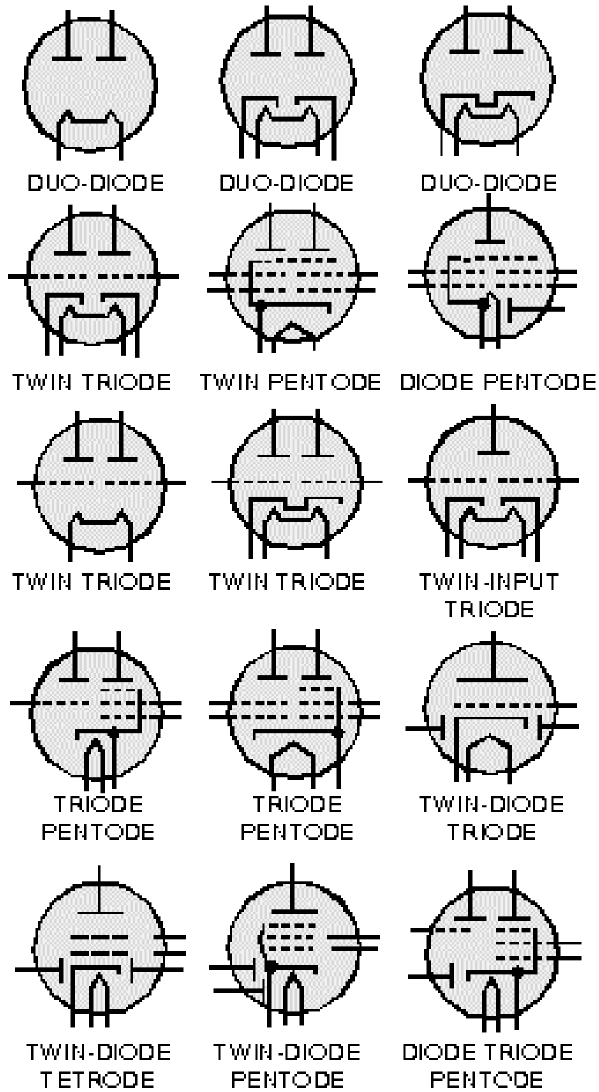


Figure 2-1.—Typical multi-unit tube symbols.

An important point to remember when dealing with multi-unit tubes is that each unit is capable of operating as a separate tube. But, how it operates, either as a single tube or as a multi-unit tube, is determined by the external circuit wiring. When you analyze the schematic of a circuit, simply treat each portion of a multi-unit tube as a single tube, as shown in figure 2-2.

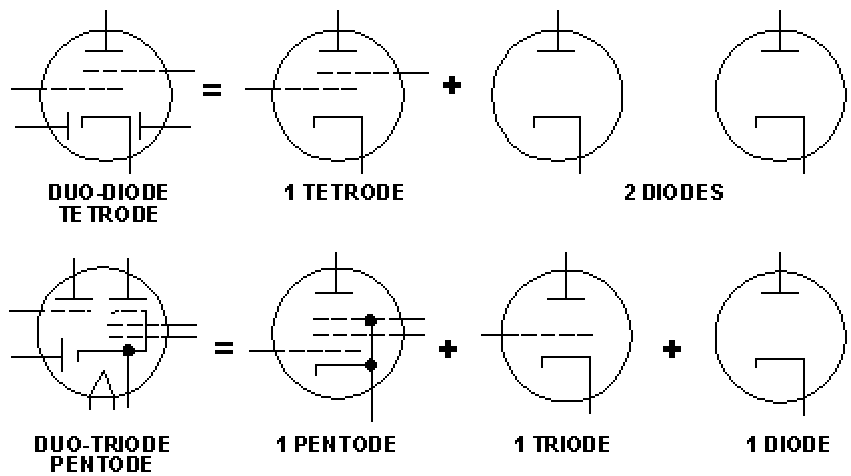


Figure 2-2.—Multi-unit tube Identification.

Another type of special-purpose tube is the **MULTI-ELECTRODE** tube. In some applications, tubes require more than the three grids found in conventional tubes. In some cases, up to seven grids may be used. These types of tubes are called multi-electrode tubes and are normally classified according to the number of grids they contain. An example of this is illustrated in figure 2-3. Here, you see a tube with five grids; hence, its name is "pentagrid." The application of these tube types is beyond the scope of this module, but because multi-electrode tubes have been commonly used you should be aware of their existence.

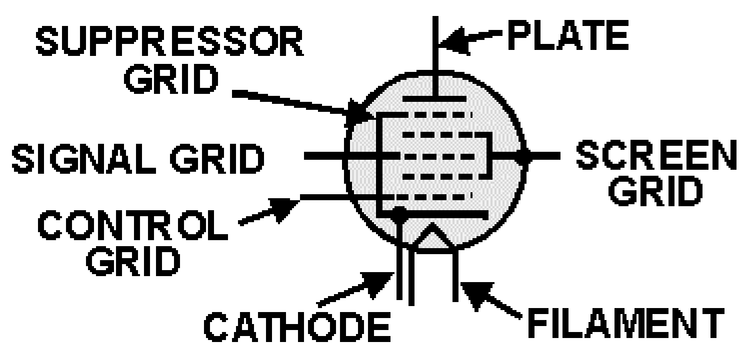


Figure 2-3.—Pentagrid multi-electrode tube.

BEAM POWER AND POWER PENTODE TUBES

The tube types you studied in the first chapter have one serious drawback; namely, they are not suitable as power amplifiers. Because of high-plate resistance and internal construction, tubes such as the triode, tetrode, and pentode are used only as voltage amplifiers. When power amplification is required (high-current requirements), special-purpose tubes called **POWER PENTODES** and **BEAM POWER** tubes are used.

Figure 2-4 shows the arrangement of grids in a conventional pentode. The small circles depict cross sections of the grids. Notice that each grid is offset, or staggered, from the grid directly behind it. This arrangement of grids permits each grid to be exposed to the electron stream flowing from cathode to plate.

In this way, each grid will have maximum effect on the electron stream. There are two undesirable effects associated with the staggered grid arrangement that make it unsuitable for use in power amplifiers.

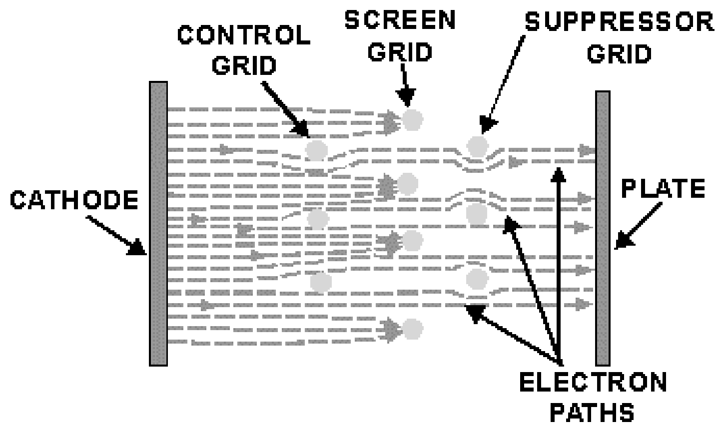


Figure 2-4.—Electron flow in a conventional pentode.

First, no direct path exists between the cathode and the plate. Electrons leaving the cathode must run an obstacle course around the grid wires to reach the plate. Some of these electrons are deflected by the grid and scattered and, thus, never reach the plate. Second, some electrons strike the grid wires and are removed from the electron stream as grid current. Because of these two undesirable effects, the amount of plate current that can flow through the tube is greatly reduced. Because of this loss of electrons from the stream, conventional tetrodes and pentodes are not suitable for power amplifiers. Therefore, a special class of tubes has been developed to overcome this problem—the **BEAM POWER TUBES** and **POWER PENTODE TUBES**.

Figure 2-5 shows the cross section of the power pentode. Notice that there is no staggered grid arrangement. Instead, each grid wire is directly in line with the grid in front of and behind it. The screen and suppressor grids are shielded from the electron stream by the control grid. Because the screen grid is "shielded" by the control grid, it can draw little grid current from the electron stream.

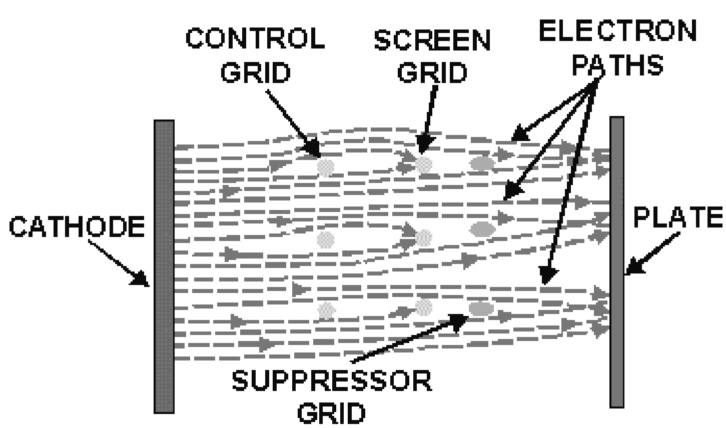


Figure 2-5.—Electron flow in a power pentode.

This arrangement of grids offers few obstacles to electron flow. Electrons will flow in "sheets" between the grid wires to the plate. The effect is to allow more of the electrons leaving the cathode to reach the plate. Thus, the tube has the advantage of high power output and high efficiency. An added advantage to this type of grid arrangement is high-power sensitivity. This means that the tube can respond to much smaller input signals than the conventional electron tube. The reason for this is obvious; many more electrons reach the plate from the cathode. Therefore, large plate currents can be obtained from relatively weak input signals.

Another type of power amplifier tube that is similar to the power pentode is the **BEAM POWER TUBE**. Beam power tubes have the same grid arrangement as the power pentodes. In addition, they use a set of beam-forming plates to force the electron stream into concentrated beams. Figure 2-6 depicts the internal construction of a beam-forming tube and its schematic representation. Notice that the beam-forming plates surround the grids and their supporting structures and are internally connected to the cathode. This internal connection ensures that the beam-forming plates are at the same negative potential as the cathode. Electrons that are emitted from the sides of the cathode are repelled from the grid supports and into the electron stream by the negative charge on the beam-forming plates. Electrons pass to the plate through the spaces between the beam-forming plates and, by doing so, are concentrated into beams. Because the beam-forming plates are at a negative potential, any electrons emitted by secondary emission are repelled back to the plate. The effect of the beam-forming plates is to increase the number of electrons in the electron stream by forcing stray electrons emitted from the sides of the cathode away from the grid supports and into the electron beam. Electrons that are deflected from the grid wires are also forced into the beam. This increases the total current flowing to the plate. For this reason, both beam-forming and power pentodes are suitable for use as power amplifiers.

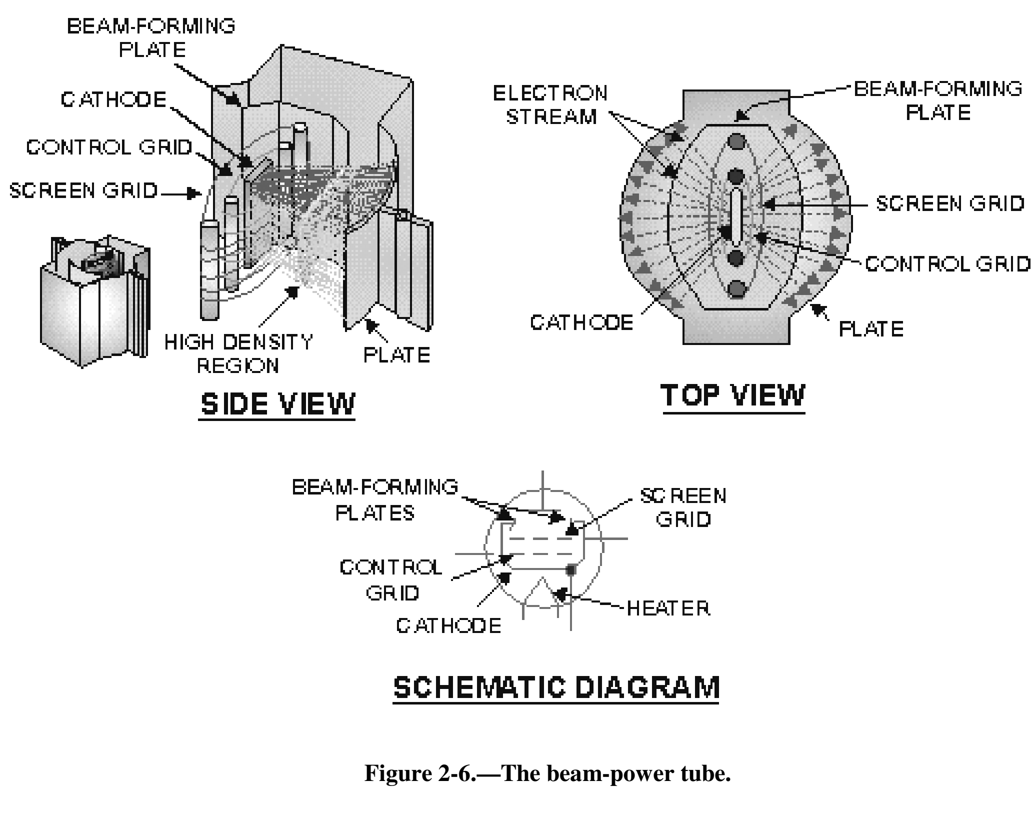


Figure 2-6.—The beam-power tube.

VARIABLE-MU TUBES

In most electron-tube circuits, the operating level of a tube is determined by the level of bias. When a negative-bias voltage is applied to the control grid of a tube, with no input signal, the conduction through the tube is reduced; thus the damage to the tube is minimized. There is one drawback to this. Because the control grid is already negatively charged by the bias voltage, the negative alternation of a large input signal will drive the tube into cutoff long before the positive alternation can drive the tube into saturation. Once the negative alternation reaches a certain level (determined by the bias voltage and tube characteristics), the tube simply cuts off. For this reason, conventional tubes, which you previously studied, are called **SHARP-CUTOFF TUBES**. Because of this sharp cutoff, the range of amplification of the conventional tube is limited by the bias voltage and tube characteristics. Once this range is exceeded, the output becomes distorted due to cutoff.

In most applications, the sharp cutoff feature of conventional electron tubes causes no problems. However, in some applications electron tubes are required to amplify relatively large input signals without distortion. For this reason, the variable-mu tube was developed. **VARIABLE-MU TUBES** have the ability to reduce their mu, or (μ), as the input signal gets larger. As the mu (μ) decreases, the likelihood that the tube will be driven into cutoff decreases. (For an amplifier, this may appear to be self-defeating, but it isn't.) The idea is to amplify large input signals as much as possible without causing the tube to cutoff or create distortion. Because of their ability to avoid being driven into cutoff, variable-mu tubes are called **REMOTE-CUTOFF TUBES**. You should be aware, however, that a variable-mu tube can be driven into cutoff, but the amplitude of the input signal required to do so is considerably greater than in conventional sharp-cutoff tubes.

The key to the ability of a variable-mu tube to decrease gain with an increase in the amplitude of the input lies in its grid construction.

To understand how the unique grid construction of a variable-mu tube works, we will first examine the grid operation of a conventional tube during cutoff. Look at figure 2-7. Here, you see a diagram of a conventional sharp-cutoff triode with zero volts applied to the control grid. In view A, the majority of the electrostatic lines of force leave the positive plate (+) and travel unhindered between the evenly spaced grid wires to the negative cathode (-). Electrons emitted by the cathode travel along these lines from the cathode, through the grid spacings, to the plate.

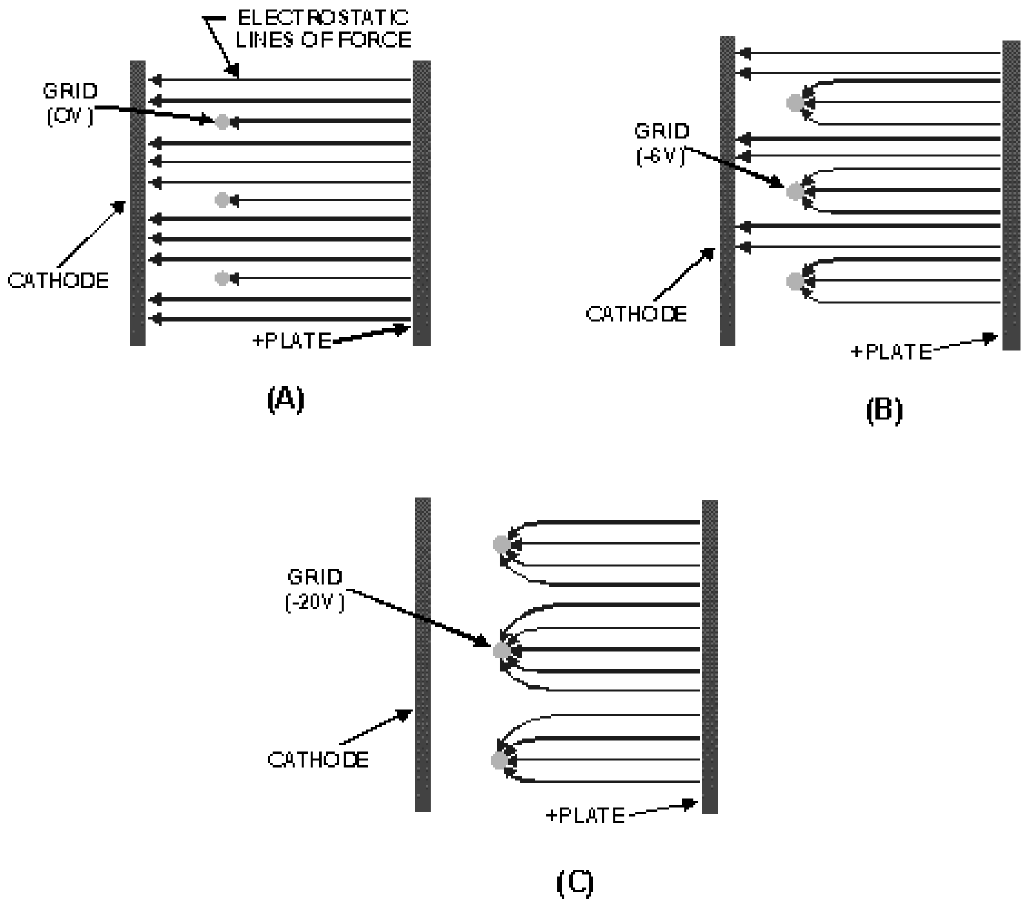


Figure 2-7.—Cutoff in a conventional tube.

In view B, a bias voltage of -6 volts is applied to the grid. As you can see, some of the electrostatic lines of force are attracted to the negatively charged grid wires while the rest pass through the grid spacings. Because there are fewer lines of force reaching the cathode, there are fewer paths for electrons to use to reach the plate. As a result, conduction through the tube is decreased.

In view C, the negative potential of the grid has been raised to -20 volts, which drives the tube into cutoff. All of the electrostatic lines of force terminate at the negatively charged grid, instead of continuing on to the cathode. The electrons emitted by the cathode will not feel the electrostatic attractive force of the positively charged plate. Under these conditions, current cannot flow through the tube.

Now look at figure 2-8. Here you see a diagram of a variable-mu, or remote-cutoff, tube. The only difference between the remote-cutoff tube depicted and the sharp-cutoff tube is in the grid wire spacing. In the conventional sharp-cutoff tube, the grid wires are evenly spaced, while in the remote-cutoff tube the grid wires in the middle of the grid are placed relatively far apart. This is shown in view A.

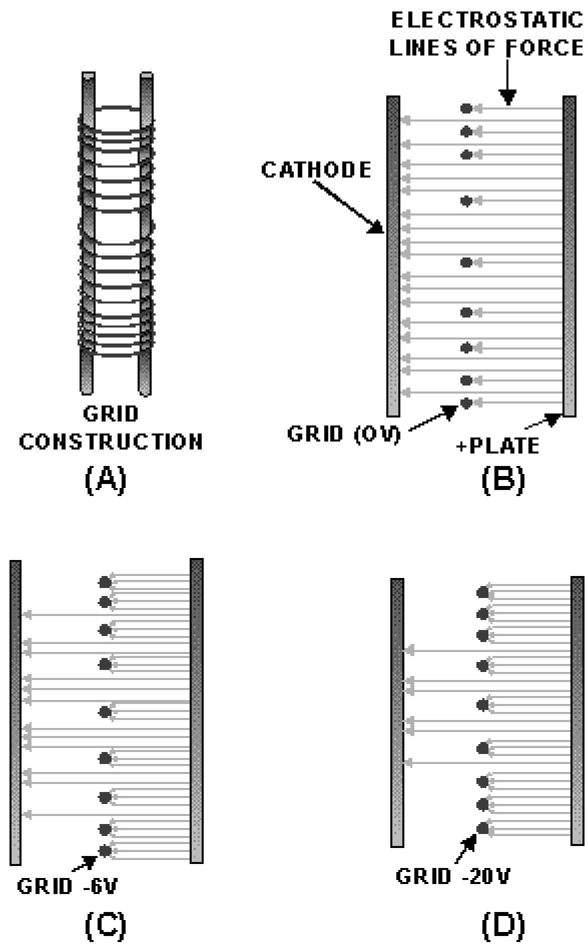


Figure 2-8.—Grid operation in a remote-cutoff tube.

In view B, the control grid is at zero potential (0 volts). Just as in the sharp-cutoff tube, electrons leave the cathode and travel along the lines of electrostatic attraction, through the spaces between the grid wires to the plate. In view C, a bias voltage of -6 volts is applied to the grid. Because of the close spacing of the grid wires at the ends of the grid, electrostatic lines of force at the ends are effectively terminated. The lines of force can only pass between the widely spaced grid wires closer to the center of the grid.

In view D, the same negative potential -20 volts) is applied to the grid that caused the conventional sharp-cutoff tube discussed earlier to go into cutoff. This voltage is high enough to terminate most of the electrostatic lines of force on the grid wire. But, because of the wide spacing between the center grid wires, some electrostatic lines of force are still able to pass between the center grid wires and reach the cathode. Conduction will still occur in the tube, but at a reduced level. If the grid is driven even more negative, lines of force will be blocked from reaching the cathode, except at the very center of the grid. As you can see, the remote-cutoff tube, by its ability to reduce gain (conduction), handles large signals without going into cutoff. A variable-mu tube such as a 6SK7 with -3 volts applied to the grid will have a transconductance of about 2000 (μ) mhos. If the grid is driven to -35 volts, the transconductance of the tube will decrease to 10 (μ) mhos. This same increase in negative-grid voltage would have driven a conventional tube into cutoff long before the grid reached -35 volts.

Q1. What is the major difference in grid construction between power pentodes and conventional pentodes?

Q2. Beam-forming tubes and power tubes are similar except that power pentodes lack what element?

Q3. What effect does the shielding of the screen grid by the control grid have on plate current in beam-forming tetrodes?

Q4. What effect does a large negative input signal applied to a variable-mu tube have on

a. conduction through the control grid, and

b. gain of the tube?

Q5. Identify the type of electron tube(s) that would be most suitable for the following applications.

a. Power amplifier

b. Voltage amplifier with small signal inputs

c. Low distortion amplifiers for use with large signal inputs

SPECIAL UHF TUBES

In the earlier discussion of conventional-electron tubes, you learned some of the limitations of tubes. One of these limitations was that the conventional tube was not able to operate (amplify) at extremely high frequencies such as those used in radar equipment. Even at frequencies lower than those used in radar equipment, problems occur. For example, at ultrahigh frequencies (300 MHz to 3000 MHz), transit time effects make the operation of a conventional-electron tube impossible. For this reason, the special ultrahigh frequency tubes were developed to operate within this frequency range.

Before we discuss the way in which special uhf tubes counter the effects of transit time, you should understand the manner in which transit time affects conventional tubes.

LIMITATION OF TRANSIT TIME

We will explain the limitation of transit time by using figure 2-9. In view A, the positive-going alternation of a uhf ac signal is applied to the grid of a conventional-triode tube. The first positive-going alternation reduces the negative bias on the grid, and electrons start to move toward the grid. Since the input is an ultrahigh frequency signal, the majority of the electrons cannot pass the grid before the input signal progresses to the negative alternation. The electrons that have not yet passed the grid are either stopped or repelled back toward the cathode. This is shown in view B. Before these electrons can move very far, the second positive alteration reaches the grid, and causes even more electrons to move from the cathode (view C). At the same time, the electrons that were repelled from the grid toward the cathode by the first negative alternation feel the effect of the positive-going grid. These electrons reverse direction and again move toward the grid. Because these electrons had to first reverse direction, they are now moving slower than the electrons that are attracted from the cathode by the second positive alteration. The result is that the electrons from the cathode catch up to the slower moving electrons and the two groups combine (view C). This action is called **BUNCHING**.

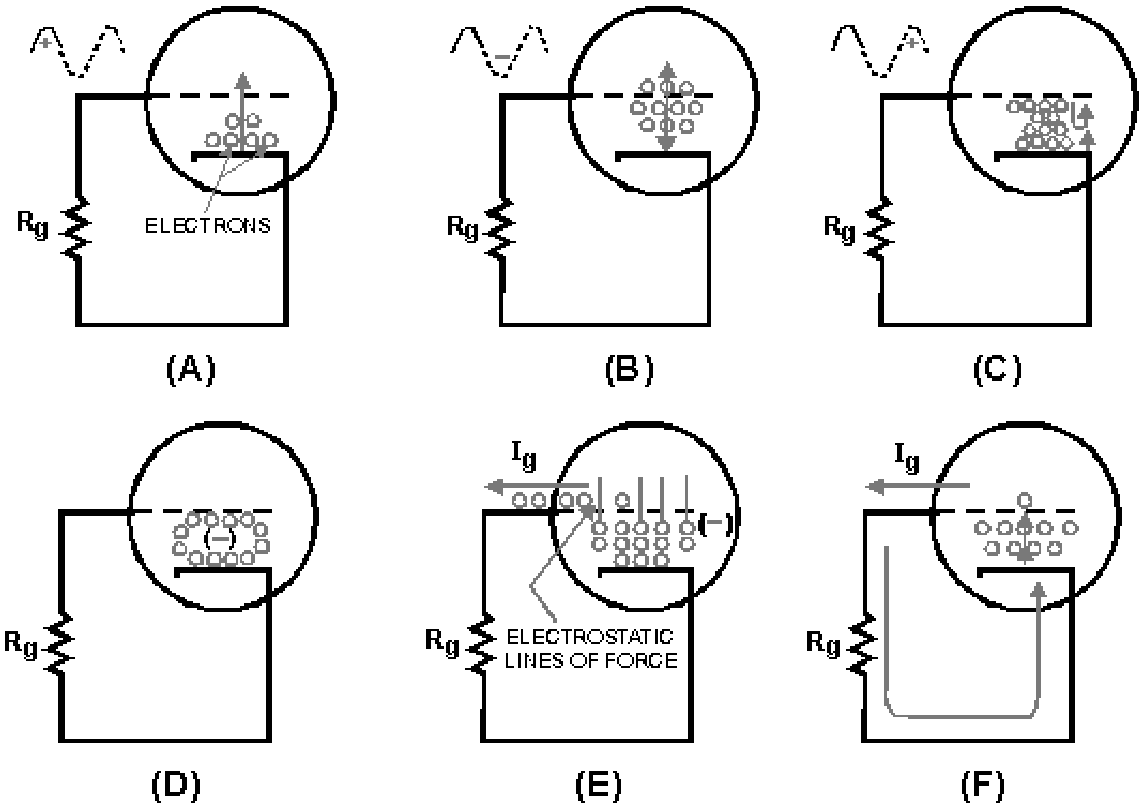


Figure 2-9.—Effect of transit time at ultrahigh frequencies.

In effect, the area between the grid and cathode becomes highly negatively charged, as shown in view D. This negative charge is surrounded by an electrostatic field. The electrostatic field cuts the grid and repels electrons that are present in the grid. As electrons are forced from the grid, the grid tries to go positive. Unfortunately, this tendency toward a positive charge attracts electrons from the mass or bunched charge. Thus, as an electron is forced from the grid; it is replaced by another from the massed charge. Electrons forced from the grid represent grid current (I_g), as shown in view E. The grid current flows from the grid through R_g, to the cathode, from the cathode, to the massed charged, and back again to the grid, The movement of current in this manner is, in effect, a path for current flow from the cathode to the grid. Because current flows between the cathode and grid, the resistance (rgk) between these elements is lowered to the point of a short circuit. The grid, in effect, is short circuited to the cathode and ceases to function; and this, in turn, lowers tube efficiency dramatically. This is shown in view F of figure 2-9.

Transit time may be decreased by reducing the spacing between electrodes or by increasing the electrode voltages, which in turn increases electron velocity through the tube.

The problem with the last solution is that the tube does not present an infinite resistance to current flow. If the operating voltage is raised to an operating potential that is too high, arcing (arc over) occurs between the cathode and the plate and, most likely, will destroy the tube. For this reason, the effects of transit time are reduced in uhf tubes by placing the tube elements very close together.

UHF TUBE TYPES

Uhf tubes have very small electrodes placed close together and often are manufactured without socket bases. By reducing all the physical dimensions of the tube by the same scale, the interelectrode-capacitance and transit time effects are reduced, without reducing the amplification capability of the tube. A disadvantage to this type of tube construction is that the power-handling capability of these tubes is also reduced due to the close placement of the tube elements.

Uhf tubes are placed in three broad categories based on their shape and/or construction; **ACORN**, **DOORKNOB**, and **PLANAR** tubes.

Acorn and Doorknob Tubes

ACORN TUBES, as shown in figure 2-10, are available for use as diodes, triodes, or pentodes. Acorns are very small tubes that have closely spaced electrodes and no bases. The tubes are connected to their circuits by short wire pins that are sealed in the glass or ceramic envelope. Because of their small size, acorn tubes are usually used in low-power uhf circuits.

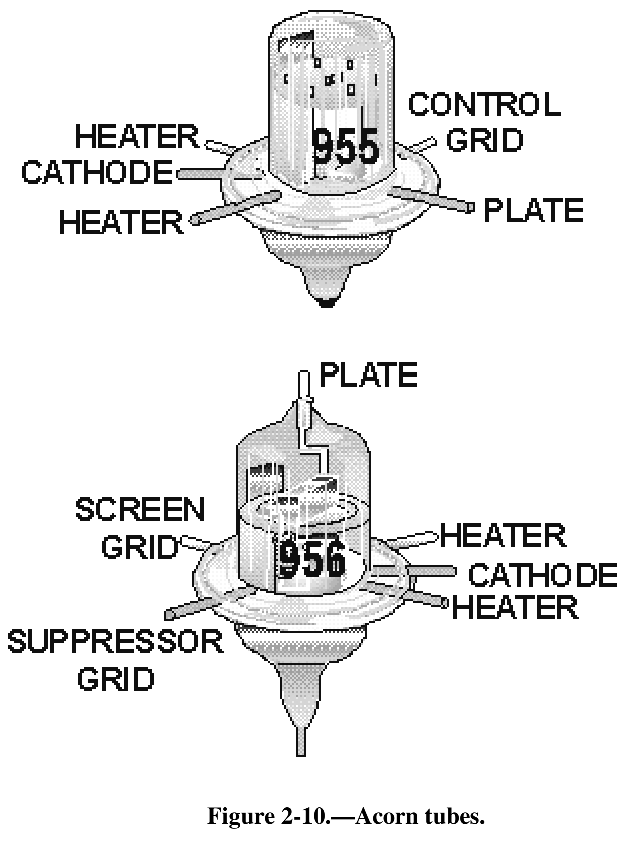


Figure 2-10.—Acorn tubes.

The **DOORKNOB TUBE** is an enlarged version of the acorn tube. Because of its larger physical size, it can be operated at higher power than the acorn tube.

Planar Tubes

PLANAR TUBES are so named because of their construction. The ordinary (conventional) tube you studied earlier uses concentric construction. This means that each element (cathode, grid, and plate) is cylindrical in shape. The grid is placed over the cathode, and the plate, which is the largest cylinder, is placed over the grid. The result is a tube composed of concentric cylinders like the one shown in figure 2-11. Thus, the name concentric tubes.

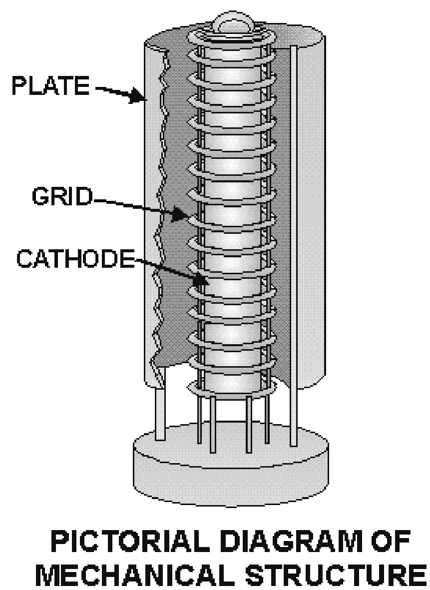


Figure 2-11.—Concentric construction of a conventional tube.

At ultrahigh frequencies, the problems of producing small tube elements while reducing the spacing between elements become very difficult. Not only are the elements hard to keep parallel with each other during the manufacturing process, but they also have a tendency to warp and sag under normal operating conditions. Since these elements are already as close together as possible, any reduction in element spacing can cause arcing. Therefore, a new type of tube was developed to prevent arcing or element sagging in conventional tubes. This tube is known as the planar tube.

Planar tubes are electron tubes in which the cathode, plate, and grids are mounted parallel to each other. Their physical construction greatly resembles a schematic diagram of a normal tube, as shown in figure 2-12.

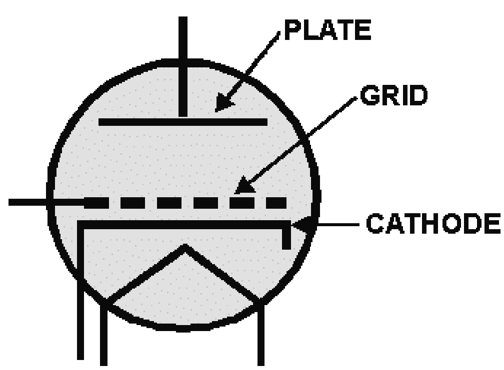


Figure 2-12.—Resemblance of a planar tube to a schematic diagram.

A typical planar tube is depicted in figure 2-13. Notice that the tube elements are mounted close to each other and are parallel to one another. The oxide coating of the cathode is applied to the top surface only. Therefore, the emitting surface of the cathode is parallel to the plate and the grid.

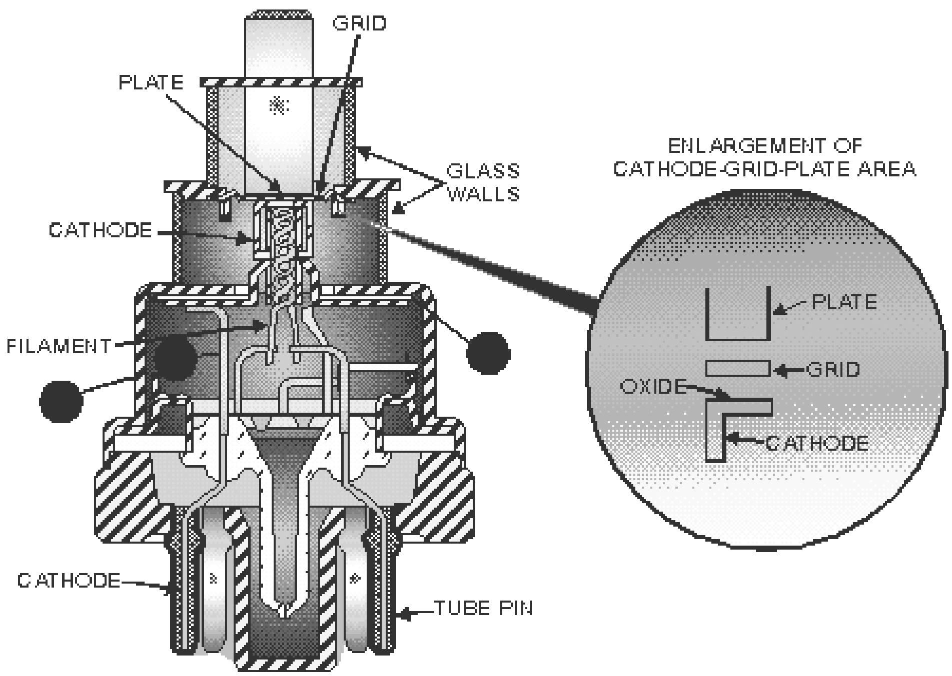


Figure 2-13.—Internal structure of a typical planar tube.

The plate of the tube consists of a cylindrical stud. This stud-plate construction has two purposes. Its flat lower surface serves as a parallel plate, and its external upper end serves as the external-plate connection from the tube to the circuit.

Because of its construction, the planar tube cannot use the ladder-type grid, with which you are familiar. Instead, the grid, formed into a circle, is composed of a wire mesh similar to that of a common screen door.

The cathode structure is manufactured in two parts. Point A of figure 2-13 is the metallic shell of the tube and is used to couple (or connect) unwanted radio frequency signals from the cathode to ground. This connection is not, however, a direct coupling. The wafer at point C of figure 2-13 is composed of mica, which serves as a dielectric. The lower extension of the cathode serves as one plate of the capacitor, while the other plate is formed from the flattened upper portion of the cathode connector ring. The cathode has a direct connection to the tube pin through the connector labeled point B. You might think that this is a rather complicated method to connect the cathode to a circuit, but it serves a purpose. At high frequencies, the wiring of a circuit can pick up radio frequency signals and retransmit them. If the wiring involved happens to be the wiring used to supply dc voltages to the circuit, all the tubes in the circuit will receive the signal. The result will be massive distortion throughout the circuit.

The problem can be eliminated by isolating the dc and radio frequency circuits from each other. In planar tubes, this is fairly simple. The point A ring is grounded. Any rf signals that the cathode may pick up through tube conduction are grounded or shorted to ground through the capacitive coupling with the point A shell. In other words, the point A shell (capacitive ground) serves the same function as the bypass capacitor in a cathode-biased circuit. Because the capacitor will not pass dc, bias voltages can be applied to the cathode through the tube pins.

Notice the external shape of the planar tube in figure 2-13. The tube is composed of five sections, or cylinders. As you go from the top to the bottom, each cylinder increases in diameter. Because of this piled cylinder construction, the tube resembles a lighthouse, and is therefore known as a **LIGHTHOUSE TUBE**.

Another type of planar tube is shown in figure 2-14. This type of tube, because of its external appearance, is called an **OILCAN TUBE**. The major difference between it and the lighthouse tube is the addition of cooling fins to allow it to handle more power than the lighthouse tube. Because of their planar construction, both types of tubes are capable of handling large amounts of power at uhf frequencies.

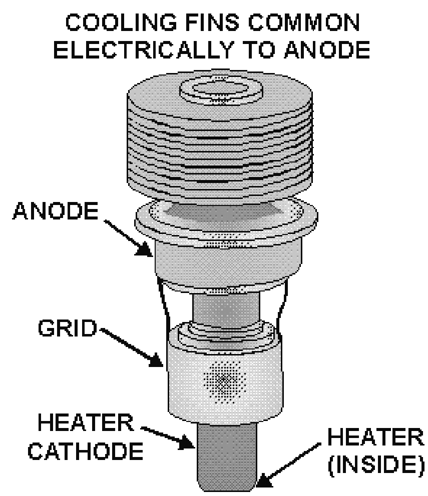


Figure 2-14.—Oilcan planar tube.

Q6. What effect does transit time have on a conventional triode operated at uhf frequencies?

Q7. How do uhf tubes counter the effects of transit time?

Q8. Why can acorn and doorknob tubes NOT handle large amounts of power?

Q9. What type of uhf tube was developed to handle large amounts of power?

GAS-FILLED TUBES

You know that great effort is made to produce a perfect vacuum within electron tubes. But, even the best vacuum pumps and getters cannot remove all of the air molecules. However, the chances of an electron hitting a molecule in a near-vacuum are very slim because of the great distance between the molecules, compared to the size of the electron. An electron can pass between two molecules of air inside the tube as easily as a pea could pass through a circle with a diameter equal to that of the earth!

In some tubes, the air is removed and replaced with an inert gas at a reduced pressure. The gases used include mercury vapor, neon, argon, and nitrogen. Gas-filled tubes, as they are called, have certain electrical characteristics that are advantageous in some circuits. They are capable of carrying much more current than high-vacuum tubes, and they tend to maintain a constant **IR** drop across their terminals within a limited range of currents. The principle of operation of the gas-filled tube involves the process called ionization.

ELECTRICAL CONDUCTION IN GAS DIODES

An operating gas-filled tube has molecules, ions, and free electrons present within the envelope. In a gas-filled diode, the electron stream from the hot cathode encounters gas molecules on its way to the plate. When an electron collides with a gas molecule, the energy transmitted by the collision may cause the molecule to release an electron. This second electron then may join the original stream of electrons and is capable of freeing other electrons. This process, which is cumulative, is a form of ionization. The free electrons, greatly increased in quantity by ionization, continue to the plate of the diode. The molecule which has lost an electron is called an ion and bears a positive charge. The positive ions drift toward the negative cathode and during their journey attract additional electrons from the cathode.

The velocity of the electrons traveling toward the plate varies directly with the plate voltage. If the plate voltage is very low, the gas-filled diode acts almost like an ordinary diode except that the electron stream is slowed to a certain extent by the gas molecules. These slower-moving electrons do not have enough energy to cause ionization when they hit the gas atoms. After the plate voltage is raised to the proper level of conduction, the electrons have enough energy to cause ionization when they hit the gas molecules. The plate potential at which ionization occurs is known as the **IONIZATION POINT**, or **FIRING POTENTIAL**, of a gas tube. If the plate voltage is reduced after ionization, it can be allowed to go several volts below the firing potential before ionization (and hence, high-plate current) win cease. The value of the plate voltage (E_p) at which ionization stops is called the **DEIONIZATION POTENTIAL**, or **EXTINCTION POTENTIAL**. The firing point is always at a higher plate potential than the deionization point.

GAS TRIODE

The point at which the gas ionizes can be controlled more accurately by inserting a grid into the gas diode. A negative voltage on the grid can prevent electrons from going to the plate, even when the plate voltage is above the normal firing point. If the negative-grid voltage is reduced to a point where a few electrons are allowed through the grid, ionization takes place. The grid immediately loses control, because the positive ions gather about the grid wires and neutralize the grid's negative charge. The gas triode then acts as a diode. If the grid is made much more negative in an effort to control the plate current, the only effect is that more ions collect about the grid wires—tube continues to conduct as a diode. Only by removing the plate potential or reducing it to the point where the electrons do not have enough energy to produce ionization will tube conduction and the production of positive ions stop. Only after the production of positive ions is stopped will the grid be able to regain control.

Such gas-filled triodes are known as **THYRATRONS**. Thyratrons are used in circuits where current flow in the thyratrons output circuit is possible only when a certain amount of voltage is present on the thyratrons grid. The flow of plate current persists even after the initiating grid voltage is no longer present at the grid, and it can be stopped only by removing or lowering the plate potential. The symbols for the gas-filled diode, the voltage regulator, and the thyratron are the same as those for high-vacuum tubes except that a dot is placed within the envelope circle to signify the presence of gas. Some examples of gas-filled tube schematic symbols are shown in figure 2-15.

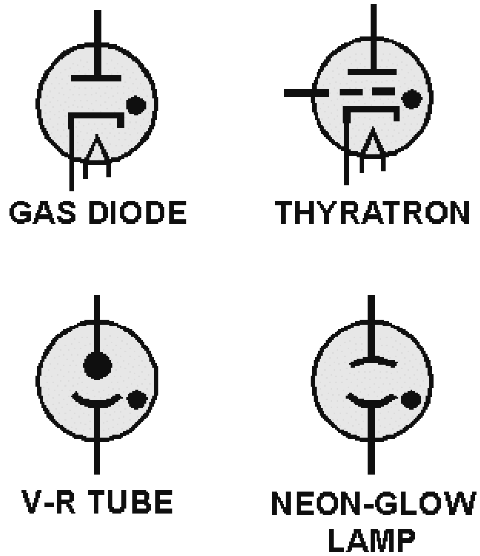


Figure 2-15.—Schematic diagram of gas-filled tubes.

Before leaving this section, you should be aware of one precaution associated with mercury-vapor tubes. The mercury vapor is not placed in the tube as a vapor; instead a small amount of liquid mercury is placed in the tube before it is sealed. When the liquid mercury comes in contact with the hot filament, the mercury vaporizes.

To ensure that the mercury has vaporized sufficiently, the filament voltage must be applied to mercury-vapor tubes for at least 30 seconds before the plate voltage is applied. If vaporization is incomplete, only partial ionization is possible. Under these conditions, the application of plate voltage results in a relatively high voltage drop across the tube (remember $E = I \times R$), and the positive ions present are accelerated to a high velocity in the direction of the cathode. As the ions strike the cathode, they tear away particles of the emitting surface, usually causing permanent damage to the cathode and the tube. When the mercury is completely vaporized, the action of the gas is such that the voltage drop across the tube can never rise above the ionization potential (about 15 volts). At this low potential, positive-ion bombardment of the cathode does not result in damage to the emitting surface.

Generally, when gas-filled tubes are in the state of ionization, they are illuminated internally by a soft, blue glow. This glow is brightest in the space between the electrodes and of lesser intensity throughout the remainder of the tube envelope. This glow is normal and must not be confused with the glow present in high-vacuum tubes when gases are present. A high-vacuum tube with a bluish glow is gassy and should be replaced. The ionization of these gases will distort the output of the tube and may cause the tube to operate with much higher plate current than it can carry safely.

COLD-CATHODE TUBES

The cold-cathode, gas-filled tube differs from the other types of gas-filled tubes in that it lacks filaments. Thus, its name "**COLD-CATHODE TUBE**." In the tubes covered in this text thus far, thermionic emission was used to send electrons from the cathode to the plate. This conduction of electrons can be caused in another manner. If the potential between the plate and the cathode is raised to the point where tube resistance is overcome, current will flow from the cathode whether it is heated or not. In most applications in electronics, this method is not used because it is not as efficient as thermionic emission. There are two applications where cold-cathode emission is used. The first application you are already familiar with, although you may not be aware of it. Every time you look at a neon sign you are watching a cold-cathode tube in operation. Thus, the first application of cold-cathode tubes is for visual display. You are also familiar with the reason for this visual display. In the ***NEETS*** module on matter and energy, we explained that when energy is fed into an atom (neon in this case), electrons are moved, or promoted, to higher orbits. When they fall back, they release the energy that originally lifted them to their higher orbits. The energy is in the form of light. Cold-cathode tubes are also used as **VOLTAGE REGULATORS**. Because voltage regulators will be dealt with extensively in the next chapter, we will not cover their operation now. At this point, you only need to understand that a cold-cathode tube has the ability to maintain a constant voltage drop across the tube despite changes of current flow through the tube. The tube does this by changing resistance as current flow varies.

Examine figure 2-16. Here you see a cold-cathode tube connected to a variable voltage source. The variable resistor rkp does not exist as a physical component, but is used to represent the resistance between the cathode and the plate. Most cold-cathode tubes have a firing point (ionizing voltage) at about 115 volts. Thus, the tube in view A of the figure is below the firing point. Because the tube lacks thermionic emission capabilities, no current will flow and the tube will have a resistance (rkp) near infinity. The potential difference between the plate and ground under these conditions will be equal to the source (E_{bb}) voltage, as shown on the voltmeter.

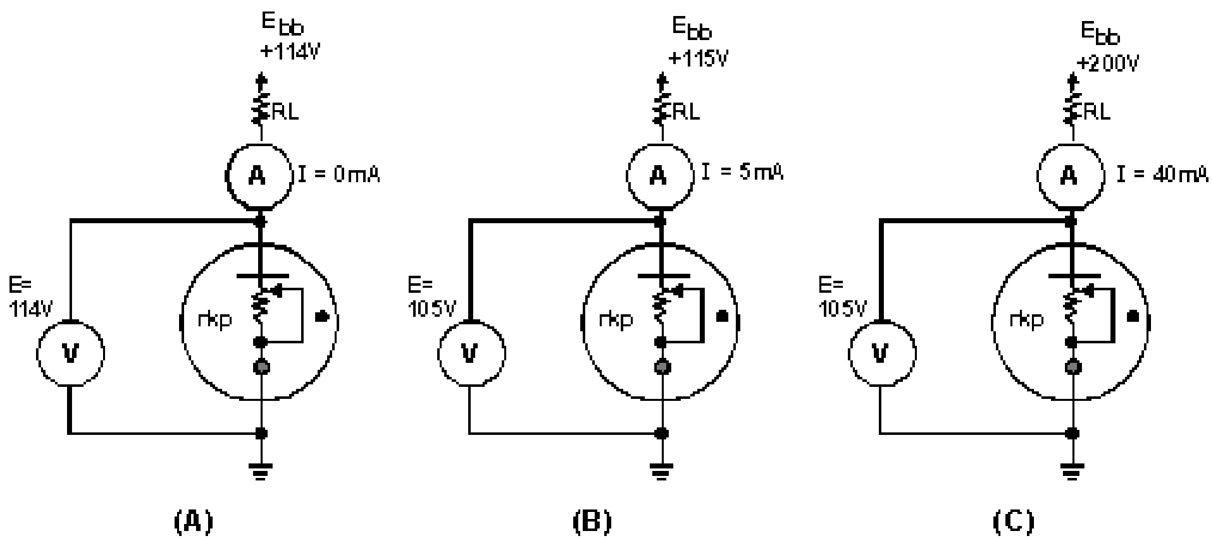


Figure 2-16.—Cold-cathode tube operation.

In view B, the source voltage has been raised to the firing point of 115 volts. This causes the gas to ionize and 5 milliamperes of current will flow through the tube. Because the tube represents a resistance (rkp), voltage will be dropped across the tube; in this case, 105 volts. The plate-load resistor (R_L) will drop the remaining 10 volts. The resistance of the tube at this time will be equal to:

$$rkp = \frac{E}{I}$$

$$rkp = \frac{105 \text{ volts}}{.005 \text{ amperes}}$$

$$rkp = 21 \text{ kohms}$$

In view C, the source voltage has been raised to 200 volts. This will cause more gas in the tube to ionize and 40 milliamperes of current to flow through the tube. The increased ionization will lower the resistance of the tube (rkp). Thus, the tube will still drop 105 volts. The tube's resistance (rkp) at this time will be equal to:

$$rkp = \frac{E}{I}$$

$$rkp = \frac{105 \text{ volts}}{40 \text{ milliamperes}}$$

$$rkp = 2.625 \text{ kohms}$$

As you can see, increasing the current flow will cause more ionization in the tube and a corresponding decrease in the tube's resistance. Because of this, the tube will always have a constant voltage drop between its plate and cathode throughout its operating range.

Q10. What are two advantages that gas-filled tubes have over conventional electron tubes?

Q11. Once ionization has occurred in a thyratron, what control does the control grid have over the tube's operation?

Q12. What precautions should be exercised when using mercury-vapor thyratrons?

Q13. Cold-cathode tubes can be used as voltage regulators because of what characteristic?

THE CATHODE-RAY TUBE (CRT)

Although you may not be aware of this fact, the **CATHODE-RAY TUBE** shown in figure 2-17 is, in all probability, the one tube with which you are most familiar. Before you started your study of electronics, you probably referred to cathode-ray tubes as picture tubes. The cathode-ray tube (**CRT**) and the picture tube of a television set are one and the same.

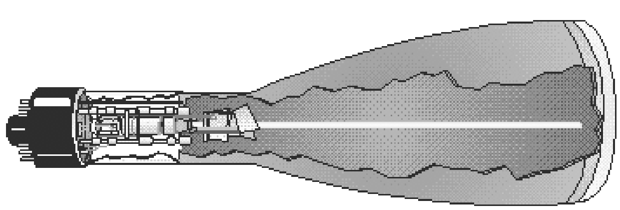


Figure 2-17.—Cutaway view of a typical CRT.

Cathode-ray tubes are used in more applications than just television. They can be considered as the heart of the many types of information.

Cathode-ray tubes have one function that cannot be duplicated by any other tube or transistor; namely, they have the ability to convert electronic signals to visual displays, such as pictures, radar sweeps, or electronic wave forms.

All CRT's have three main elements: an electron gun, a deflection system, and a screen. The electron gun provides an electron beam, which is a highly concentrated stream of electrons. The deflection system positions the electron beam on the screen, and the screen displays a small spot of light at the point where the electron beam strikes it.

THE ELECTRON GUN

The **ELECTRON GUN** is roughly equivalent to the cathodes of conventional tubes. The cathode of the electron gun in the CRT is required not only to emit electrons, but also to concentrate emitted electrons into a tight beam. In the electron tubes that you have studied, the cathode was cylindrical and emitted electrons in all directions along its entire length. This type of cathode is not suitable for producing a highly concentrated electron-beam. The cathode of the CRT consists of a small diameter nickel cap. The closed end of the cap is coated with emitting material. This is shown in figure 2-18. Because of this type of construction, electrons can only be emitted in one direction. Notice that the emitted electrons shown in figure 2-18 are leaving the cathode at different angles. If these electrons were allowed to strike the screen, the whole screen would glow. Since the object of the electron gun is to concentrate the electrons into a tight beam, a special grid must be used. This special grid is in the form of a solid metal cap with a small hole in the center. The grid is placed over the emitting surface of the cathode and charged negatively in relation to the cathode. The dotted lines represent the direction of cathode emitted electron repulsion, as shown in figure 2-19. Since all emitted electrons leave the cathode (point C), their paths can be identified. An electron attempting to travel from point C to point B (downward) will instead follow the path from point C to point E to point P. Consider an electron leaving from C in the direction of point A (upward). Its path will be curved from point C to point P by electrostatic repulsion. These curving electron paths are due to the negative potential of the grid coupled with the high positive potential of the anode. The potential of the anode attracts electrons out of the cathode-grid area past point P toward the screen. The grid potential may be varied to control the number of electrons allowed to go through the control-grid opening. Since the brightness or intensity of the display depends on the number of electrons that strike the screen, the control grid is used to control the brightness of the CRT.

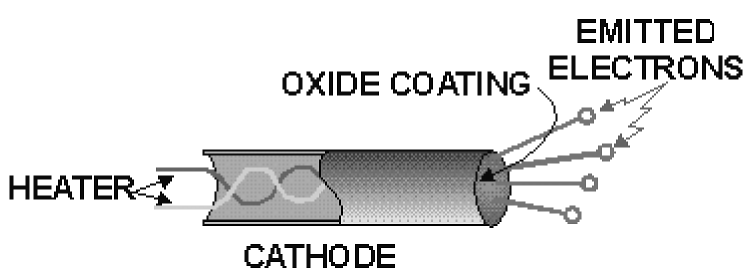


Figure 2-18.—CRT cathode.

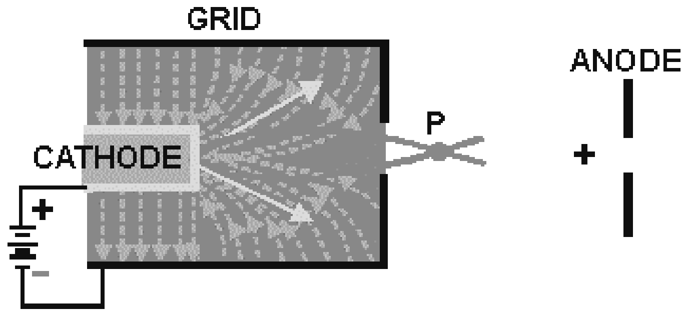


Figure 2-19.—Operation of the CRT grid.

The proper name, **BRIGHTNESS CONTROL**, is given to the potentiometer used to vary the potential applied to the control grid. The control grid actually serves as an electron lens. It is this electronic lens that you adjust when you turn up the brightness control on your TV set. Notice that the effect of the grid is to focus the electron beam at point P in figure 2-19.

After passing point P, the electrons start to spread out, or diverge, again. Therefore, it becomes necessary to provide some additional focusing to force the electrons into a tight beam again. This is done by two additional positively-charged electrodes as shown in figure 2-20. The first electrode is commonly called the **FOCUSING ANODE**. Generally, the focusing anode is charged a few hundred volts positive with respect to the cathode. Electrons emitted by the cathode are attracted to the focusing anode. This is the reason that they travel through the small hole in the grid. The second electrode, called the **ACCELERATING ANODE**, is charged several thousand volts positive in relation to the cathode. Any electrons approaching the focusing anode will feel the larger electrostatic pull of the accelerating anode and will be bent through the opening in the focusing anode and will travel into the area labeled D. You might think that once an electron is in this region, it is simply attracted to the accelerating anode and that is the end of it. This does not happen. Because the accelerating anode is cylindrical in shape, the electrostatic field radiating from it is equal in all directions. Thus, an electron is pulled in all directions at once, forcing the electron to travel down the center of the tube. Then, the electron is accelerated into the accelerating anode. Once it passes the mid-point (point E), it feels the electrostatic attraction from the front wall of the accelerating anode, which causes it to move faster toward the front. Once the electron reaches point F, equal electrostatic attraction on either side of the opening squeezes it through the small opening in the front of the anode. From there, it is joined by millions of other electrons and travels in a tight beam until it strikes the screen (point S).

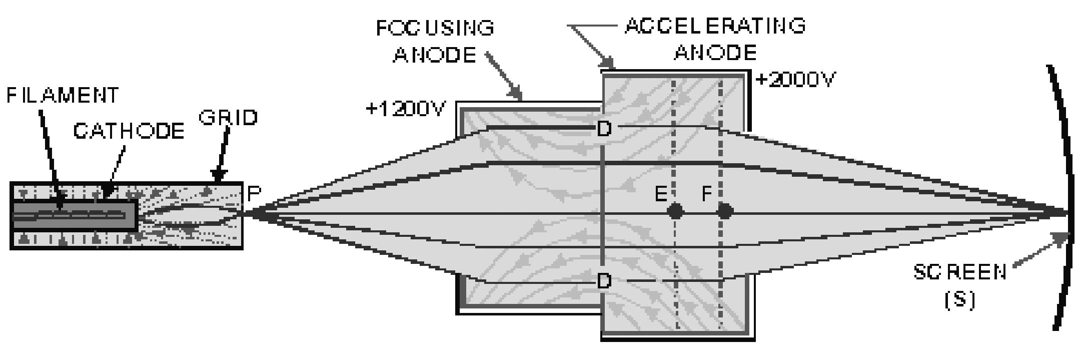


Figure 2-20.—Electron-beam formation In a CRT.

THE CRT SCREEN

The inside of the large end of a CRT is coated with a fluorescent material that gives off light when struck by electrons. This coating is necessary because the electron beam itself is invisible. The material used to convert the electrons' energy into visible light is a **PHOSPHOR**. Many different types of phosphor materials are used to provide different colored displays and displays that have different lengths of **PERSISTENCE** (duration of display).

In one way, the CRT screen is similar to a tetrode vacuum tube. Both suffer from the effects of secondary emission. In order to reach the screen, electrons from the cathode are accelerated to relatively high velocities. When these electrons strike the screen, they dislodge other electrons from the material of the screen. If these secondary emission electrons are allowed to accumulate, they will form a negatively-charged barrier between the screen and the electron beam, causing a distorted image on the CRT screen. The method used to control secondary emission, which you are already familiar with, i.e., a suppressor grid, is not practical in CRT's. Instead, a special coating called an **AQUADAG COATING** is applied to the inside of the tube as shown in figure 2-21. This coating is composed of a conductive material, such as graphite, and has the same high-positive potential applied to it that is applied to the accelerating anode. This allows the aquadag to perform two functions. First, since the aquadag coating is positive, it attracts the secondary emitted electrons and removes them. Second, because the aquadag is operated at a high-positive potential and is mounted in front of the accelerating anode, it aids in the acceleration of electrons toward the screen.

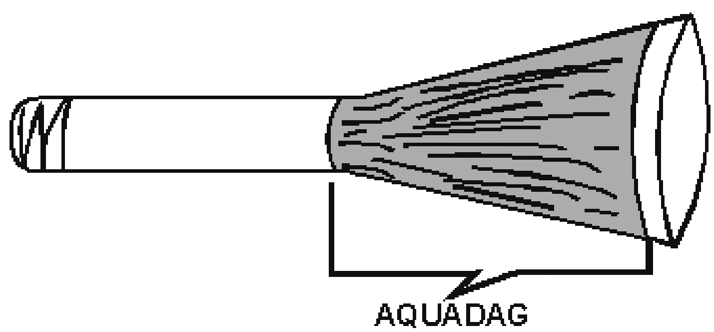


Figure 2-21.—Aquadag coating in a CRT.

Before going on, let's review what you have already learned about CRT operation.

1. Electrons are emitted from a specially constructed cathode and move toward the front of the CRT.

2. The number of electrons that leave the area of the cathode is determined by the cap-shaped grid. In addition, the grid concentrates the emitted electrons into a beam.

3. The electron beam is focused and accelerated toward the screen by two electrodes: the focusing anode and the acceleration anode.

4. The electron beam strikes the screen and causes a bright spot to appear at the point of impact.

5. Any electrons released by secondary emission are removed from the tube by the aquadag coating.

DEFLECTION

At this point, you have a bright spot in the center of the CRT screen as shown in figure 2-22. Having watched TV, you know that a TV picture consists of more than just a bright spot in the center of the picture tube. Obviously, something is necessary to produce the picture. That something is called **DEFLECTION**. For the CRT to work properly, the spot must be moved to various positions on the screen. In your TV set for example, the spot is moved horizontally across the CRT face to form a series of tightly packed lines. As each line is displayed, or traced, the electron beam is moved vertically to trace the next line as shown in figure 2-23. This process starts at the top of the tube and ends when the last line is traced at the bottom of the CRT screen. Because the beam is swept very quickly across the CRT and the phosphor continues to glow for a short time after the beam has moved on, you do not see a series of lines, but a continuous picture.

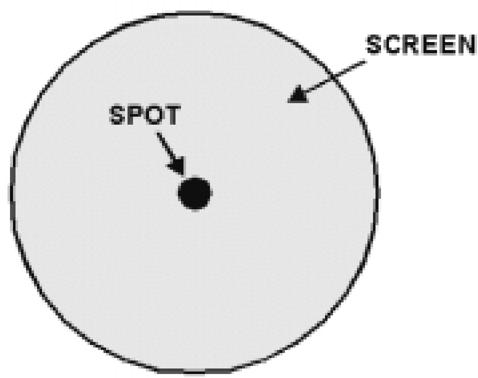


Figure 2-22.—Impact of an electron beam on a CRT screen.

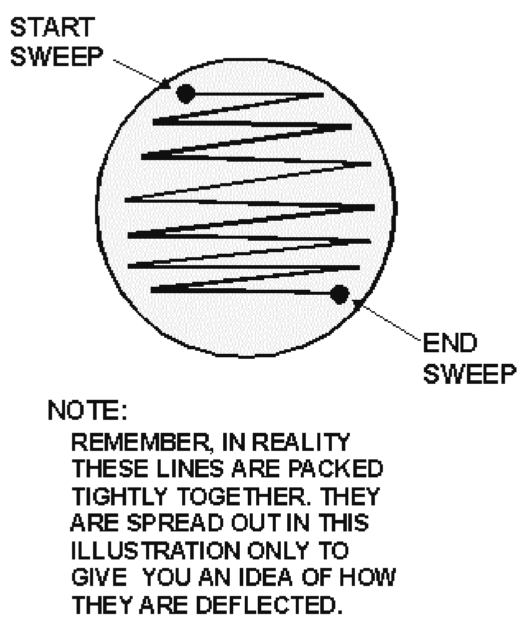


Figure 2-23.—Deflection of an electron beam across a TV screen.

These same principles also apply to the CRT used in your use of your major tool: the **OSCILLOSCOPE**. Remember, the unique function of a CRT is to convert electronic (and electrical) signals to a visual display. This function of a CRT is used in the oscilloscope to show the waveform of an electronic signal. To help you understand better how an oscilloscope works, we will discuss the type of deflection used in oscilloscopes. Bear in mind that the following discussion is only about deflection; we will cover the actual operation of an oscilloscope in a later ***NEETS*** module that deals specifically with test equipment.

Electrostatic Deflection

As you should know, there are two ways to move an electron (and thus an electron beam): either with a magnetic or with an electrostatic field. Because of this, there are three possible ways to move or deflect an electron beam in a CRT: magnetically, electromagnetically, and electrostatically. All three ways are used in electronics. In general, though, electrostatic and electromagnetic deflection are used most often. Your TV set, for example, uses electromagnetic deflection, while much of the test equipment in the Navy uses electrostatic deflection.

ELECTROSTATIC DEFLECTION uses principles you are already familiar with. Namely, opposites attract, and likes repel. Look at figure 2-24, view A. Here you see an electron traveling between two charged plates, H_1 and H_2. As you can see, before the electron reaches the charged plates, called **DEFLECTION PLATES**, its flight path is toward the center of the screen. In view B, the electron has reached the area of the deflection plates and is attracted toward the positive plate, H_2, while being repelled from the negative plate, H_1. As a result, the electron is deflected to the right on the inside of the screen. You, the viewer, will see the spot of light on the left side of the CRT face (remember, you are on the opposite side of the CRT screen). This is shown in view C.

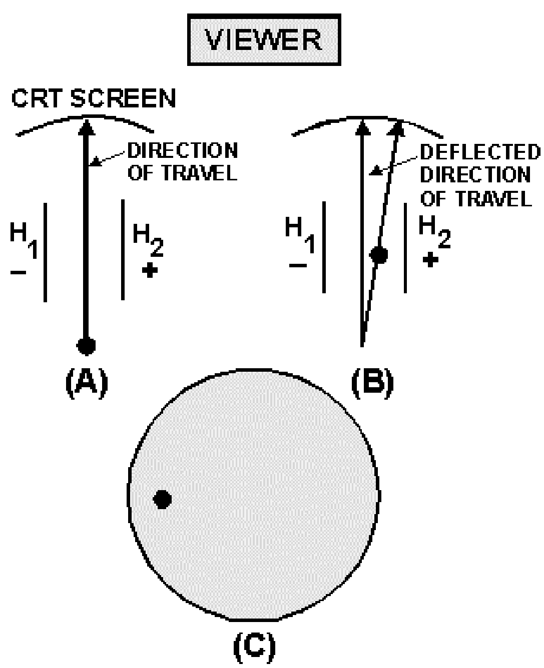


Figure 2-24.—Deflection in a CRT.

A spot of light on the left-hand side of the CRT screen, however, is no more useful than a spot of light in the center of the screen. To be useful, this spot will have to be converted to a bright line, called a sweep, across the face of the CRT screen. We will explain the manner in which this is done by using figure 2-25. In view A, five electrons are emitted in sequence, 1 through 5, by the electron gun. The right deflection plate, H_2, has a large positive potential on it while the left plate, H_1 has a large negative potential on it. Thus, when electron 1 reaches the area of the deflection plates, it is attracted to the right plate while being repelled from the left plate. In view B, electron 2 has reached the area of the deflection plates. However, before it arrives, R1 and R2 are adjusted to make the right plate less positive and the left plate less negative. Electron 2 will still be deflected to the right but not as much as electron 1. In view C, electron 3 has reached the area of the deflection plates. Before it gets there, R1 and R2 are adjusted to the mid-point. As a result, both plates have 0 volts applied to them. Electron 3 is not deflected and simply travels to the center of the CRT screen. In view D, electron 4 has reached the area of the deflection plates. Notice that R1 and R2 have been adjusted to make the right plate negative and the left plate positive. As a result, electron 4 will be deflected to the left. Finally, in view E, the left plate is at its maximum positive value. Electron 5 will be deflected to the extreme left. What you see when you are facing the CRT is a bright luminous line, as shown in view E. While this description dealt with only five electrons, in reality the horizontal line across a CRT face is composed of millions of electrons. Instead of seeing five bright spots in a line, you will see only a solid bright line.

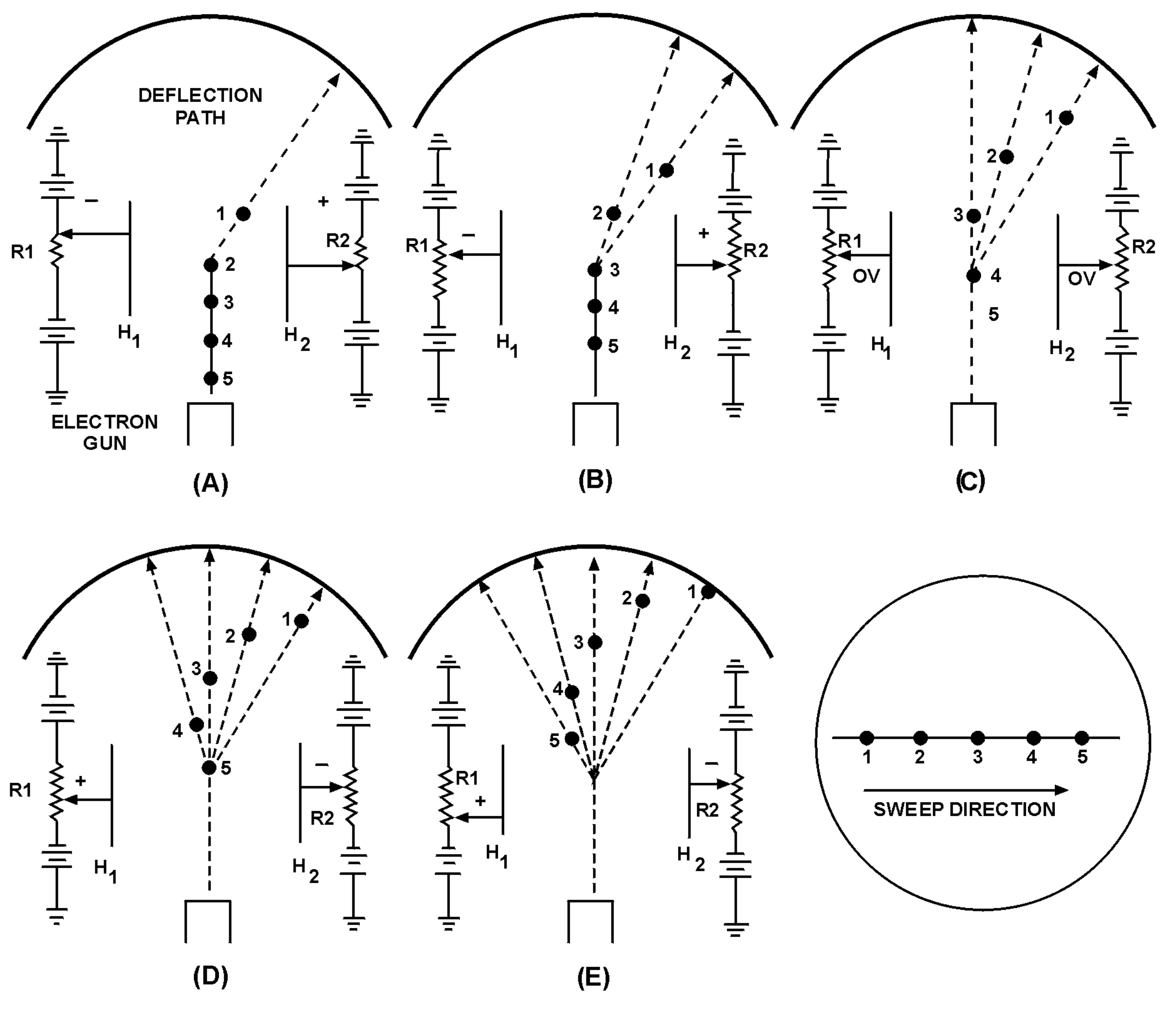


Figure 2-25.—Horizontal deflection.

In summary, the horizontal line displayed on a CRT or on the face of a television tube is made by sweeping a stream of electrons rapidly across the face of the CRT. This sweeping action, or scanning, is performed by rapidly varying the voltage potential on the deflection plates as the electron stream passes.

Vertical Deflection

As we mentioned earlier, a CRT can be used to graphically and visually plot an electronic signal, such as a sine wave. This is done by using a second set of deflection plates called **VERTICAL-DEFLECTION PLATES**. Examine figure 2-26. You are looking at the front view (facing the screen) of a CRT, back into the tube at the deflection plates. In normal usage, the horizontal plates sweep a straight line of electrons across the screen from left to right while the signal to be displayed is applied to the vertical deflection plates. A circuit of this type is shown in figure 2-27. We will use this figure to explain how a sine wave is displayed. First, however, you need to understand what is happening in view A. The box on the left of the CRT labeled **HORIZONTAL-DEFLECTION CIRCUITS** is an electronic circuit that will duplicate the actions of R1 and R2 used earlier in making up a horizontal line. How it works will be discussed in a later ***NEETS*** module. Notice T1; the output of this transformer is applied to the vertical-deflection plates. The signals applied to the vertical plates are 180º out of phase with each other. Thus, when one plate is attracting the electron beam, the other will be repelling the electron beam. Because you are only concerned with what happens inside the CRT, this circuitry will be eliminated and only the CRT and its deflection plates will be shown, as in view B.

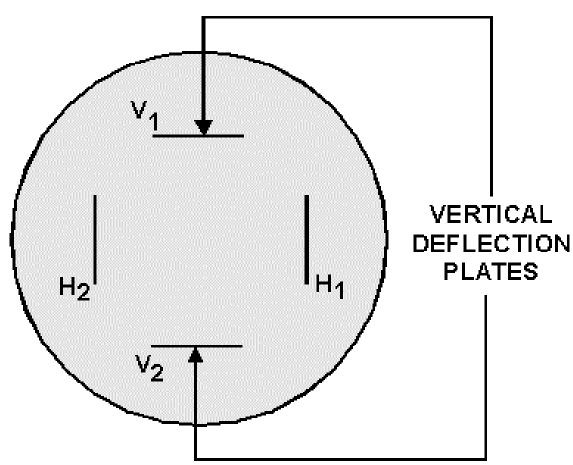


Figure 2-26.—Arrangement of deflection plates in a CRT, front view.

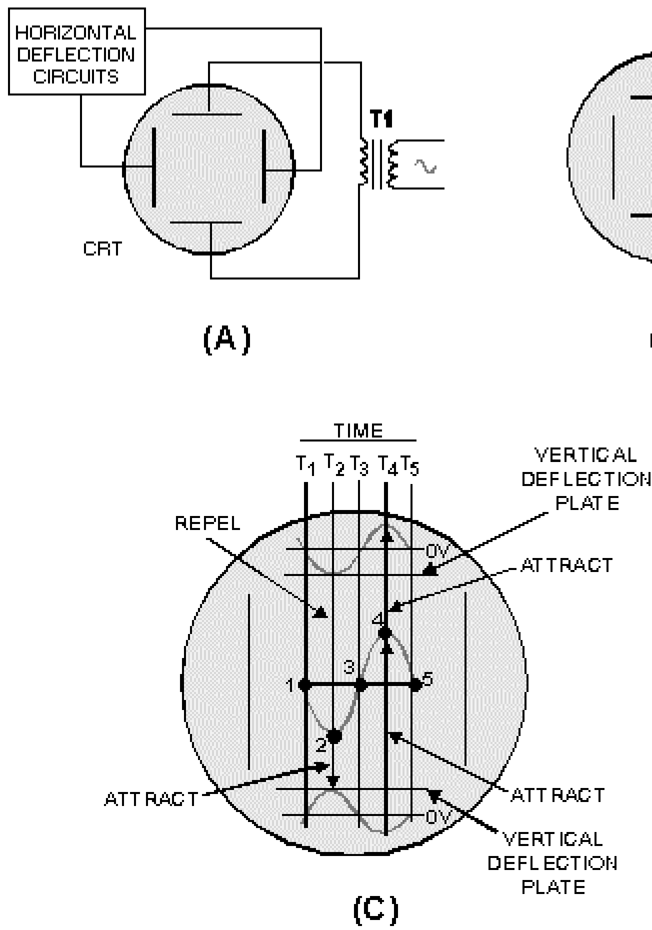


Figure 2-27.—Vertical deflection in a CRT.

Now look at view C. While this illustration looks complicated, don't let it worry you. You have already analyzed more complicated diagrams. The sine wave in the center of the screen is the signal that will be displayed as a result of the two 180º out-of-phase sine waves applied to the vertical-deflection plates. The five spots on the center sine wave represent the five electrons used to explain horizontal deflection. Only now these electrons will be deflected both vertically and horizontally. Time lines T1 through T5 represent the time when each like-numbered electron reaches the area of the deflection plates. Because you already know how the electron beam is swept or deflected horizontally, we will not discuss horizontal deflection. Just remember that from T1 to T5, the electron beam will be continuously moved from your left to your right. Now that you know where everything is on the illustration, you are ready to discover how a sine wave is displayed on a CRT.

At time 1 (T1), the sine waves applied to both vertical-deflection plates are at their null points, or zero volts. As a result, electron 1 is not vertically deflected and strikes the CRT at its vertical center. At time 2 (T2), the sine wave applied to the top plate is at its maximum negative value. This repels electron 2 toward the bottom of the CRT. At the same time, the sine wave applied to the bottom plate is at the most positive value, causing electron 2 to be attracted even further toward the bottom of the CRT. Remember, the beam is also being moved to the left. As a result, electron 2 strikes the CRT face to the right of and below electron 1. At time 3 (T3), both sine waves applied to the vertical-deflection plates are again at the null point, or zero volts. Therefore, there is no vertical deflection and electron 3 strikes the CRT face in the center of the vertical axis. Because the electron beam is still moving horizontally, electron 3 will appear to the right of and above electron 2. At time 4 (T4), the sine wave applied to the top vertical-deflection plate is at its maximum positive value. This attracts electron 4 toward the top deflection plate. The upward deflection of electron 4 is increased by the negative-going sine wave (at time 4) applied to the bottom deflection plate. This negative voltage repels electron 4 upward. Thus, electron 4 strikes the CRT face to the right of and above electron 3. Finally, at time 5 (T5) both input sine waves are again at zero volts. As a result, electron 5 is not deflected vertically, only horizontally. (Remember, the beam is continually moving from right to left.)

While this discussion is only concerned with five electrons, vertical scanning, or deflection, involves millions of electrons in a continuous electron beam. Instead of seeing five spots on the CRT screen, you will actually see a visual presentation of the sine wave input. This was, as you remember, described earlier as the unique feature of the CRT. You may have wondered why so much space in this chapter was taken up with the discussion of the CRT. There are two reasons for this. First, the field of electronics is in a constant state of evolution. Transistors replaced most vacuum tubes. Transistors are being replaced by integrated circuits (ICs). As you progress in your career in electronics, you will find that the equipment you work on will follow this evolution, from transistors to IC chips. Of all the tubes discussed in this text, the CRT is the least likely to be replaced in the near future. Thus, in all probability, whether your career in electronics lasts for only the time you spend with this text or 20 years, the CRT will be your constant companion and co-worker.

The second reason for this rather extensive coverage of the CRT is that, while the CRT has a unique ability, it operates exactly like all the tubes previously discussed.

SUMMARY OF THE CRT

This summary will not only review the CRT, but will also point out the similarities between the CRT and other tubes.

Look at figure 2-28. Here you see both a schematic diagram and a pictorial representation of a CRT. Each element is identified by a circuit number. We will review briefly the function of each element in a CRT and its similarity to elements in conventional tubes. This summary will help you tie together everything you have learned about the CRT and electron tubes in general.

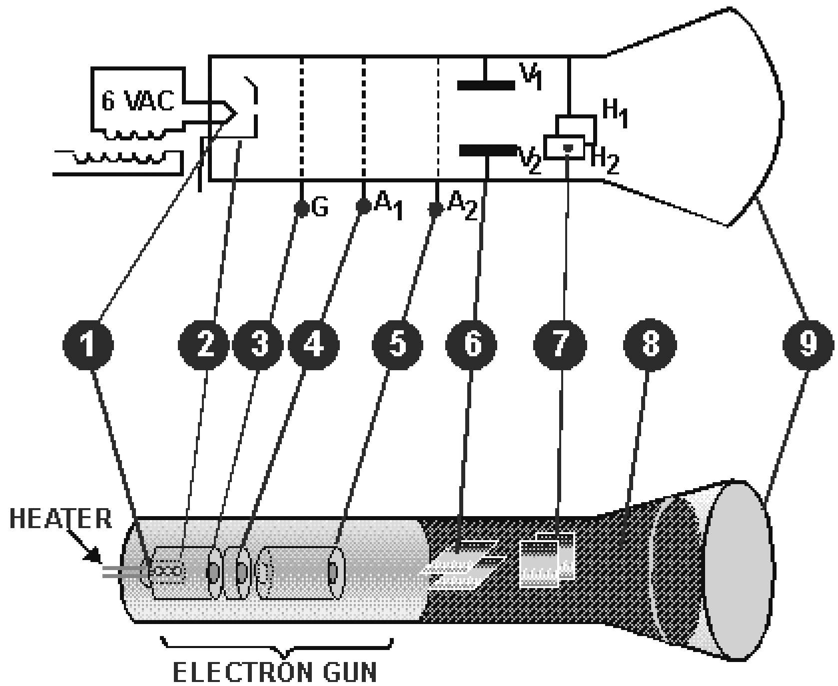


Figure 2-28.—Summary of the CRT.

1. The Heater—serves as the source of heat for the cathode in both the CRT and indirectly heated tubes.

2. The Cathode—serves as the source of thermionically emitted electrons in both the CRT and conventional tubes. The major difference is that in the CRT, the cathode is circular in shape and the outer surface is coated to ensure that electron emission is roughly unidirectional.

3. The Control Grid-in both the CRT and conventional vacuum tubes, the control grid controls the number of electrons that will be fired across "the gap." The major difference is in the physical construction. Conventional tubes use a wire-mesh ladder-type grid, while the CRT uses a cap-like grid.

4. The Focusing Anode—in the CRT, this anode serves a dual purpose of attracting electrons from the area of the control grid and focusing the electrons into a beam. Its function of attracting electrons from the area of the grid is similar to the action of the plate in a conventional tube. The focusing action of the anode is similar to that performed by beam-forming plates in the beam-forming tetrode. Bear in mind, though, that beam-forming plates are negatively charged and repel electrons into electron sheets, while the focusing anode is positively charged and attracts electrons into beam.

5. The Accelerating Anode—in the CRT, this anode is used to accelerate the electrons toward the front of the tube. Its action is similar to the screen grid of tetrodes and pentodes. But remember, while the screen grid in conventional tubes accelerates electrons toward the plate, its primary purpose is to reduce interelectrode capacitance, NOT accelerate electrons.

6. The Vertical-Deflection Plates—in the CRT, these plates move the electron beam up and down the screen. The input signal is usually applied to these plates. While no equivalent element is

found in conventional tubes, the principle employed (electrostatic attraction and repulsion) forms the heart of all vacuum tube operation.

7. The Horizontal-Deflection Plates—in the CRT, these plates move the electron beam by electrostatic attraction and repulsion, horizontally across the CRT screen. In most equipment using the CRT, including television sets, electronic signals are supplied to these plates to trace or paint a horizontal line.

8. The Aquadag Coating—in the CRT, this coating performs the same function as the suppressor grid in conventional tubes; namely, eliminating the effects of secondary emission. In conventional tubes, the suppressor grid is negatively charged and repels secondary emission electrons back to the plate. In the CRT, the aquadag is positively charged and attracts secondary emission away from the screen.

9. The Screen—also called the face, is a unique element of the CRT. When struck by electrons, the phosphor coating becomes luminous, or glows, thus enabling the tube to visually present electronic signals.

From this comparison of the CRT and other types of electron tubes, one fact should be clear. Almost all tubes, no matter what their function, operate on two principles: electrostatic attraction and repulsion, and thermionic emission. By keeping these two principles in mind, you should be able to analyze any type of tube operation.

Q14. What is the unique ability of the CRT?

Q15. What are the three main parts of CRT?

Q16. What term is used for the ability of a spot on a CRT screen to continue to glow after the electron bean has struck it and moved away?

Q17. The electron beam in a CRT is made to sweep from left to right across the screen. What tube element causes this sweeping motion?

Q18. In applications where electronic waveforms are displayed on a CRT screen, the input signal is normally applied to what CRT element?

SAFETY

There are certain safety precautions you should follow when you work with or handle the special tubes covered in this chapter. We will examine these tubes and their associated precautions in the following sections.

ELECTRON TUBES

The average electron tube is a rugged device capable of withstanding the shocks and knocks of everyday usage and handling. However, they are not indestructible. You should remember that most electron tubes contain a near vacuum enclosed by a glass envelope. Because of this, the glass is under constant stress from atmospheric pressure. Any undue stress, such as striking the envelope against a hard surface, may cause the envelope to shatter, resulting in an **IMPLOSION**.

An implosion is just the opposite of an explosion. When the glass envelope of an electron tube shatters, the outside atmosphere rushes into the tube to fill the vacuum. As the air rushes into the tube, it carries glass fragments of the envelope with it. Once these fragments reach the center of the tube, they continue outward with considerable velocity. The result is similar to an explosion, in that the immediate area surrounding the electron tube is filled with fast-moving glass fragments. You, as a nearby object, may find yourself the target for many of these glass fragments. For this reason you should handle all electron tubes with care.

CATHODE-RAY TUBES (CRTS)

Since most electron tubes are small, the possibility of them being a safety hazard is usually very small. There are two exceptions to this: CRT's and radioactive tubes.

The glass envelope of a CRT encloses a high vacuum. Because of its large volume and surface area, the force exerted on a CRT by atmospheric pressure is considerable. The total force on a 10-inch CRT may exceed 4,000 pounds. Over 1000 pounds is exerted on the CRT face alone.

When a CRT is broken, a large implosion usually occurs. Almost two tons of force hurl glass fragments toward the center of the tube. At the same time, the electron gun is normally thrown forward inside the tube. The face, because of its size, tends to move very slowly toward the center of the tube. This presents one of the main hazards of a broken CRT. The electron gun passes through the center of the tube with considerable force. It continues until it strikes the CRT face. The impact from the electron gun normally breaks the CRT face into many small fragments, which are hurled outward. The face is coated with a chemical coating that is extremely toxic. If you are unfortunate enough to experience an accidental implosion of a CRT and are nicked by one of these fragments, seek immediate medical aid. As you can see, improper handling of a CRT can be very hazardous to your health.

The CRT is, in essence, a tiny fragmentation bomb. The major difference between a CRT and a bomb is that a bomb is designed to explode; a CRT is not. As long as you handle a CRT properly, it represents no danger to you. Only when you mishandle it do you risk the danger of being pelted with an electron gun and toxic glass fragments. When handling a CRT, you should take the following precautions:

1. Avoid scratching or striking the surface of the CRT.
2. Do not use excessive force when you remove or replace a CRT's deflection yoke or socket.
3. Do not try to remove an electromagnetic-type CRT from its yoke until you have discharged the high voltage from the CRT's anode connector (hole).
4. Never hold the CRT by its neck.
5. Always set the CRT with its face down on a thick piece of felt, rubber, or smooth cloth.
6. Always handle the CRT gently. Rough handling or a sharp blow on the service bench can displace the electrodes within the tube, causing faulty operation.
7. Wear safety glasses and protective gloves.

One additional handling procedure you should be aware of is how to dispose of a CRT properly. When you replace a CRT, you cannot simply throw the old CRT over the side of the ship, or place it in the nearest dumpster. When thrown over the side of a ship, a CRT will float; if it washes ashore, it is dangerous to persons who may come in contact with it. A CRT thrown in a dumpster represents a hidden booby trap. Therefore, always render the CRT harmless before you dispose of it. This is a fairly simple procedure, as outlined below.

Note: Be sure to wear safety goggles.

- Place the CRT face down in an empty carton and cover its side and back with protective material.

- Carefully break off the plastic locating pin from the base (fig. 2-29) by crushing the locating pin with a pair of pliers.

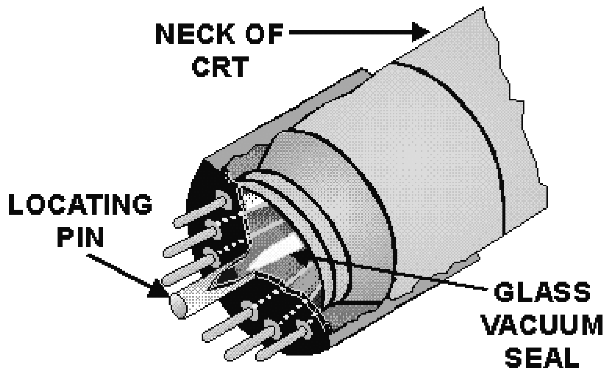


Figure 2-29.—Cathode-ray tube base structure.

- Brush the broken plastic from the pin off the CRT base.

- Look into the hole in the base where the locator pin was. You will see the glass extension of the CRT called the vacuum seal. Grasp the vacuum seal near the end with the pliers and crush it.

This may sound a little risky but it isn't. The vacuum seal can be crushed without shattering the tube. Once the seal has been crushed, air will rush into the tube and eliminate the vacuum.

RADIOACTIVE ELECTRON TUBES

Another type of tube that can prove hazardous to you, if you handle it improperly, is the radioactive tube.

These tubes contain radioactive material and are used as voltage-regulator, gas-switching, and cold-cathode, gas-rectifier tubes. Some of these tubes have dangerous radioactive intensity levels. Radioactive tubes are marked according to military specifications.

Radioactive material is added to a tube to aid in ionization. The radioactive material emits relatively slowly moving particles. This should not worry you because the glass envelope is thick enough to keep these particles inside the tube. Therefore, proper handling is nothing more than ensuring that the envelope remains unbroken. If these tubes are broken and the radioactive material is exposed, or escapes from the confines of the electron tube, the radioactive material becomes a potential hazard.

The concentration of radioactivity in an average collection of electron tubes in a maintenance shop does not approach a dangerous level, and the hazards of injury from exposure are slight. However, at major supply points, the storage of large quantities of radioactive electron tubes in a relatively small area may create a hazard. For this reason, personnel working with equipment using electron tubes containing radioactive material, or in areas where a large quantity of radioactive tubes are stored, should read and become thoroughly familiar with the safety practices contained in *Radiation, Health, and Protection Manual*, NAVMED P-5055. Strict compliance with the prescribed safety precautions and procedures of this manual will help to prevent accidents, and to maintain a safe working environment which is conducive to good health.

The clean-up procedures listed below are based on NAVMED P-5055. Your ship or station may have additional procedures that you should follow. Be sure you are aware of your command's policy concerning decontamination procedures before you begin working on equipment containing radioactive tubes. Some important instructions and precautions for the proper handling of radioactive tubes are listed below:

1. Do not remove radioactive tubes from their carton until you are ready to install them.
2. When you remove a tube containing a radioactive material from equipment, place it immediately in an appropriate carton to prevent possible breakage.
3. Never carry a radioactive tube in a manner that may cause it to break.
4. If a radioactive tube that you are handling or removing breaks, notify the proper authority and obtain the services of qualified radiological personnel immediately.
5. Isolate the immediate area of exposure to protect other personnel from possible contamination and exposure.
6. Follow the established procedures set forth in NAVMED P-5055.
7. Do not permit contaminated material to come in contact with any part of your body.
8. Take care to avoid breathing any vapor or dust that may be released by tube breakage.
9. Wear rubber or plastic gloves at all times during cleanup and decontamination procedures.
10. Use forceps to remove large fragments of a broken radioactive tube. Remove the remaining small particles with a designated vacuum cleaner, using an approved disposal collection bag. If a vacuum cleaner is not designated, use a wet cloth to wipe the affected area. In this case, be sure to make one stroke at a time. **DO NOT** use a back-and-forth motion. After each stroke, fold the cloth in half, always holding one clean side and using the other for the new stroke. Dispose of the cloth in the manner stated in instruction 14 below.
11. Do not bring food or drink into the contaminated area or near any radioactive material.
12. Immediately after leaving a contaminated area, if you handled radioactive material in any way, remove all of your clothing. Also wash your hands and arms thoroughly with soap and water, and rinse with clean water.
13. Notify a medical officer immediately if you sustain a wound from a sharp radioactive object. If a medical officer cannot reach the scene immediately, stimulated mild bleeding by applying pressure about the wound and by using suction bulbs. **DO NOT USE YOUR MOUTH** if the wound is a puncture-type wound. If the opening is small, make an incision to promote free bleeding, and to make the wound easier to clean and flush.
14. When you clean a contaminated area, seal all debris, cleaning cloths, and collection bags in a container such as a plastic bag, heavy wax paper, or glass jar, and place them in a steel can until they can be disposed of according to existing instructions.
15. Use soap and water to decontaminate all tools and implements you used to remove the radioactive substance. Monitor the tools and implements for radiation with an authorized radiac set to ensure that they are not contaminated.

As you can see, the cleanup that results from breaking a radioactive tube is a long and complicated procedure. You can avoid this by simply ensuring that you <u>don't break the tube</u>.

CONVENTIONAL TUBES

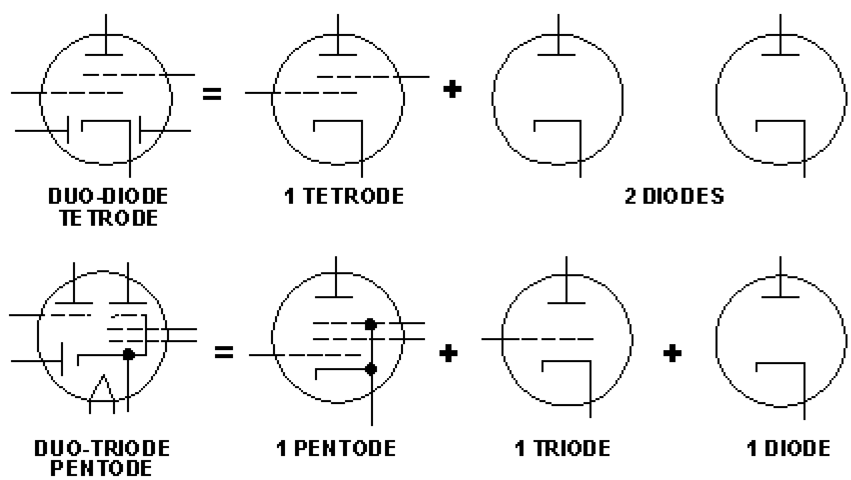
While conventional tubes present few safety problems, beyond broken glass and the possibility of cutting yourself, there is one precaution you must know. Namely, <u>electron tubes are hot</u>. The filaments of some tubes may operate at several thousand degrees. As a result, the envelopes can become very hot. When you work on electron tube equipment, always deenergize the equipment and allow the tubes sufficient time to cool before you remove them. If this is impossible, use special tube pullers which the Navy stocks for this purpose. <u>Never attempt to remove a hot tube from its socket with your bare fingers</u>.

SUMMARY

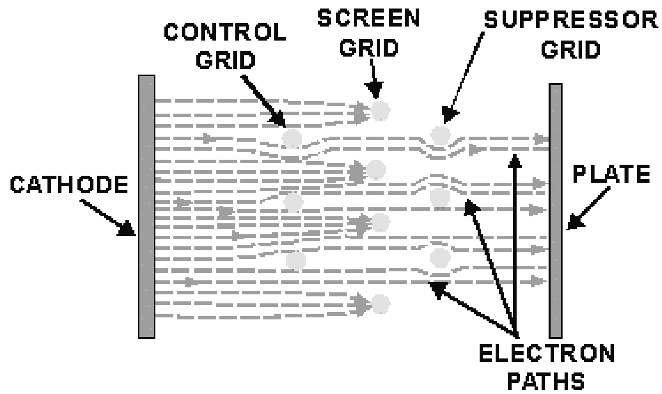
The following summary covers the main points of this chapter. Study it to be sure you understand the material before you begin the next chapter.

MULTI-UNIT TUBES were developed to reduce the size of vacuum tube circuits. Incorporating more than one tube in the same envelope allowed the size of a vacuum tube circuit to be reduced considerably. While a single envelope may contain two or more tubes, these tubes are independent of each other.

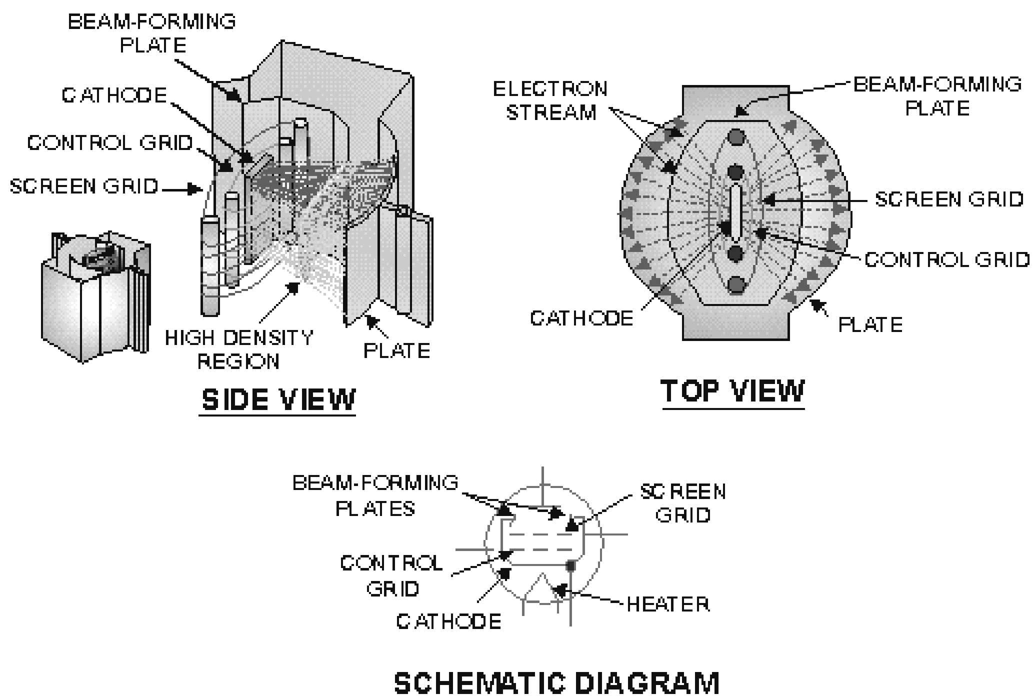
MULTI-ELECTRODE TUBES were developed to extend the capability of conventional tubes. In some cases, a multi-element tube may contain up to seven grids. These types of tubes are normally classified by the number of grids they contain.



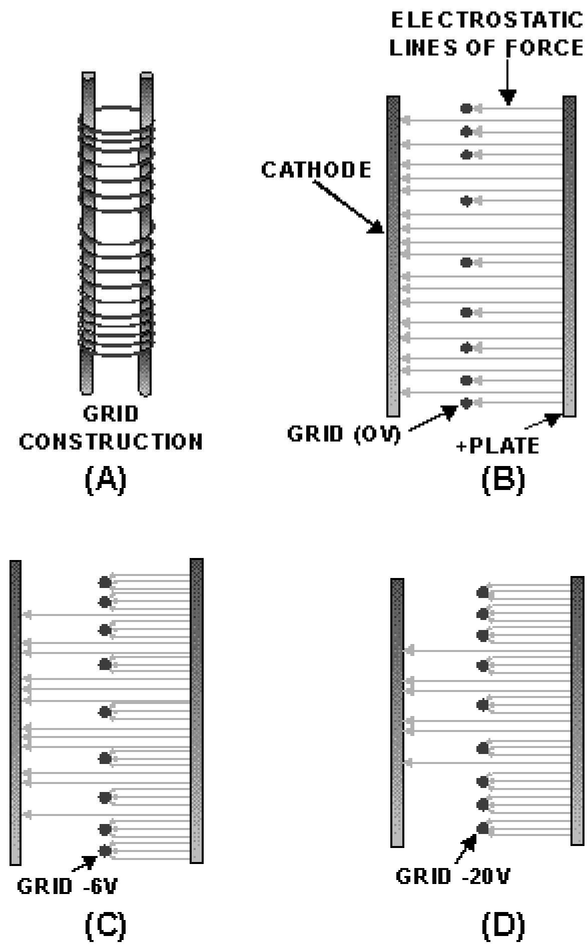
POWER PENTODES are used as current or power amplifiers. Power pentodes use in-line grid arrangements. In this manner, more electrons can reach the plate from the cathode. In effect, this lowers plate resistance and allows maximum conduction through the tube.



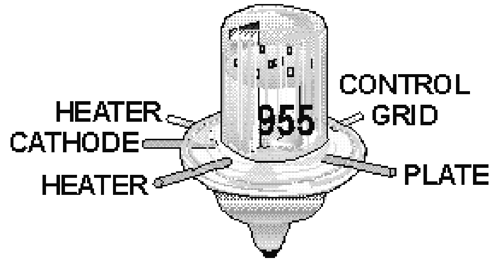
BEAM-POWER TUBES are also used as power amplifiers. In addition to the in-line grid arrangement, beam-power tubes use a set of negatively charged beam-forming plates. The beam-forming plates force electrons that would normally be deflected from the plate back into the electron steam and, thus, add to the number of electrons the tube can use for power amplification.

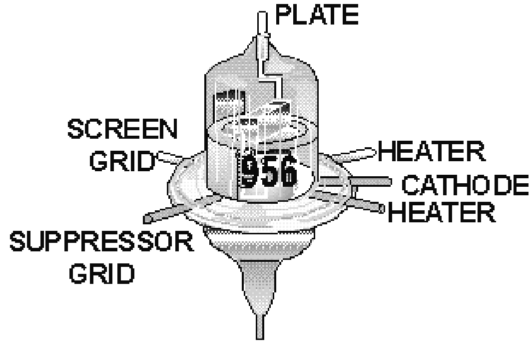


VARIABLE-MU (μ) TUBES or REMOTE-CUTOFF TUBES were developed to extend the amplification range of electron tubes by avoiding the possibility of driving the tube into cutoff. This is done by uneven spacing of the grid wires.

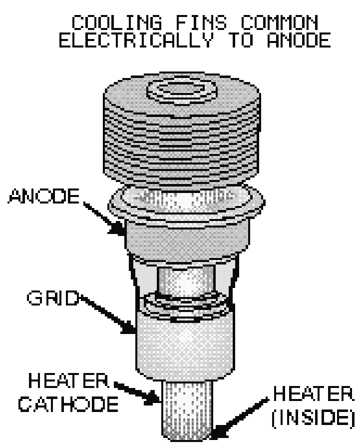


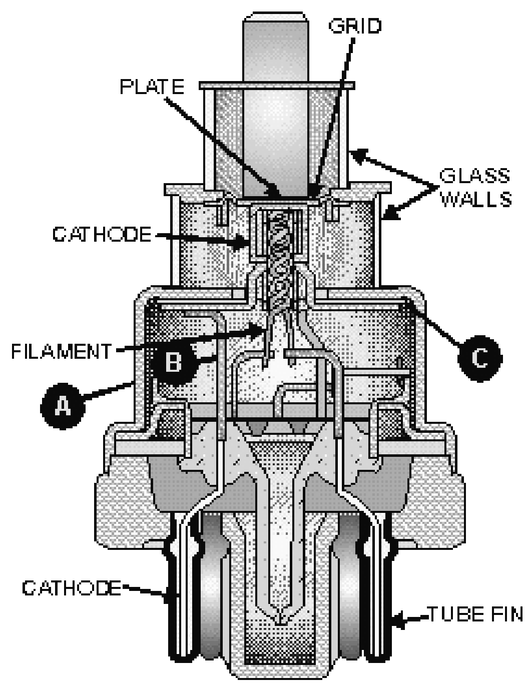
UHF TUBES are special-purpose tubes designed to operate at ultrahigh frequencies between 300 MHz and 3000 MHz with minimum effect from transit time limitations. Among these are acorn tubes, and doorknob tubes, lighthouse tubes, and oilcan tubes.



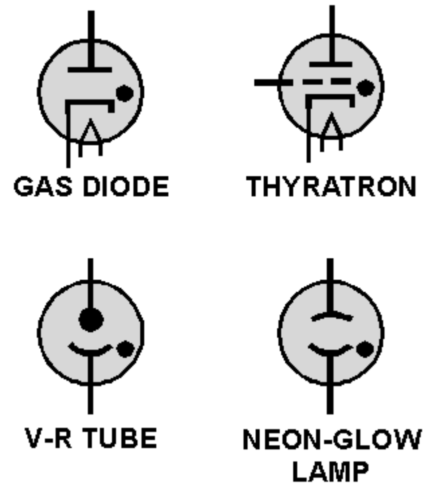


PLANAR TUBES have their plates and grids mounted parallel to each other. Because of their planar construction, they can handle large amounts of power at uhf frequencies.

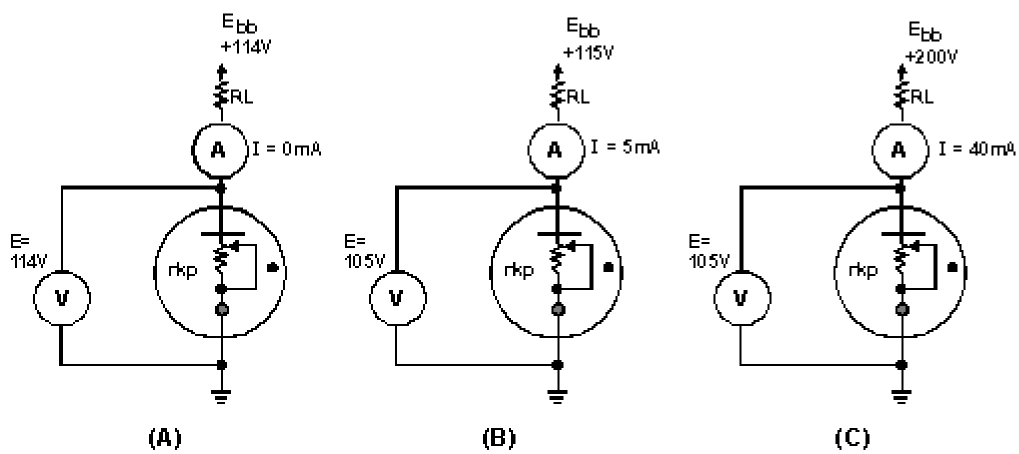




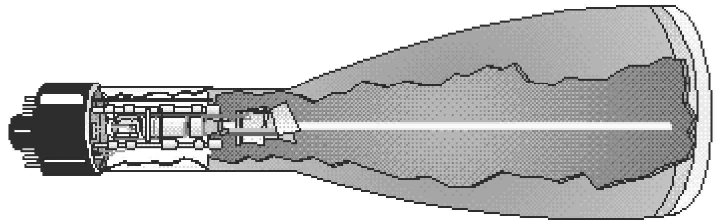
GAS-FILLED TUBES contain a small amount of gas that ionizes and reduces the internal resistance of the tubes. Because of this, gas-filled tubes can handle relatively large amounts of power while maintaining a constant voltage drop across the tube.



COLD-CATHODE TUBES lack heaters or filaments and, therefore, do not use thermionic emission. Instead, a voltage potential applied across the tube causes the internal gas to ionize. Once ionization has occurred, the voltage drop across the tube remains constant, regardless of increased conduction.



The **CRT** is a special-purpose tube that has the unique ability to visually display electronic signals. The CRT uses the principles of electrostatic attraction, repulsion, and fluorescence. Because of its unique ability, the CRT makes up the heart of many types of test equipment that you will become familiar with during your career in electronics.



ANSWERS TO QUESTIONS Q1. THROUGH Q18.

A1. Conventional pentodes have a staggered grid arrangement, while power pentodes have a shielded grid arrangement.

A2. Beam-forming plates.

A3. By increasing the number of electrons that reach the plate, plate current is increased.

A4. A large negative voltage causes conduction to occur only at the center of the grid

A5. Decreases gain.

a. Power pentode or beam-forming tetrode.

b. Conventional tube.

c. Variable-mu tube.

A6. It causes the control grid to short to the cathode.

A7. By reducing the spacing between tube elements.

A8. The close spacing of tube elements allows for the ready formation of arcs or short circuits.

A9. Planar

A10.

a. They can carry more current.

b. They maintain a constant IR drop across the tube.

A11. None.

A12. The filament's voltage should be applied to the tube at least 30 seconds before attempting to operate the tube.

A13. They have the ability to maintain a constant voltage drop across the tube despite changes in current flow.

A14. To visually display electronic signals.

A15.

a. Electron gun.

b. Deflection system

c. Screen.

A16. Persistence.

A17. The horizontal-deflection plate.

A18. The vertical-deflection plate.

CHAPTER 3

POWER SUPPLIES

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Identify the various sections of a power supply.
2. State the purpose of each section of a power supply.
3. Describe the operation of the power supply from both the whole unit standpoint and from the subunit standpoint.
4. Describe the purpose of the various types of rectifier circuits used in power supplies.
5. Describe the purpose of the various types of filter circuits used in power supplies.
6. Describe the operation of the various voltage and current regulators in a power supply.
7. Trace the flow of ac and dc in a power supply, from the ac input to the dc output on a schematic diagram.
8. Identify faulty components through visual checks.
9. Identify problems within specific areas of a power supply by using a logical isolation method of troubleshooting.
10. Apply safety precautions when working with electronic power supplies.

INTRODUCTION

In the early part of this century when electronics was first introduced, most electronic equipment was powered by batteries. While the use of batteries allowed the equipment to be portable (to some degree), it also placed several limitations on how the equipment could be used. Because of their general inefficiency, batteries had to be either replaced frequently or, if they were rechargeable, kept near a battery charger. Thus, the advantage of having portable equipment was more than offset by the need to replace or recharge the batteries frequently.

Users of electronic equipment needed a power supply that was reliable, convenient, and cost effective. Since batteries failed to satisfy these requirements, the "electronic power supply" was developed.

In today's Navy, all electronic equipment, both ashore and on board ship, require some type of power supply. Therefore, this chapter is of extreme importance to you. We will discuss the sections and individual components of the power supply and their purposes within the power supply. We will also discuss troubleshooting of each section and its components.

THE BASIC POWER SUPPLY

Figure 3-1 shows the block diagram of the basic power supply. Most power supplies are made up of four basic sections: a **TRANSFORMER**, a **RECTIFIER**, a **FILTER**, and a **REGULATOR**.

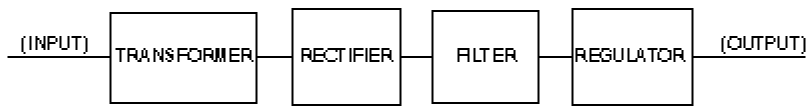


Figure 3-1.—Block diagram of a basic power supply.

As you can see, the first section is the **TRANSFORMER**. The transformer serves two primary purposes: (1) to step up or step down the input line voltage to the desired level and (2) to couple this voltage to the rectifier section. The **RECTIFIER** section converts the ac signal to a pulsating dc voltage. However, you will see later in this chapter that the pulsating dc voltage is not desirable. For this reason, a **FILTER** section is used to convert the pulsating dc voltage to filtered dc voltage. The final section, the **REGULATOR**, does just what the name implies. It maintains the output of the power supply at a constant level in spite of large changes in load current or in input line voltage. Depending upon the design of the equipment, the output of the regulator will maintain a constant dc voltage within certain limits.

Now that you know what each section does, let's trace a signal through the power supply and see what changes are made to the input signal. In figure 3-2, the input signal of 120 volts ac is applied to the primary of the transformer, which has a turns ratio of 1:3. We can calculate the output by multiplying the input voltage by the ratio of turns in the secondary winding to turns in the primary winding. Therefore, the output voltage of our example is: 120 volts ac × 3, or 360 volts ac. Depending on the type of rectifier used (full-wave or half-wave), the output from the rectifier will be a portion of the input. Figure 3-2 shows the ripple waveform associated with a full-wave rectifier. The filter section contains a network of resistors, capacitors, or inductors that controls the rise and fall time of the varying signal so that the signal remains at a more constant dc level. You will see this more clearly in the discussion of the actual filter circuits. You can see that the output of the filter is at a 180-volt dc level with an ac **RIPPLE** voltage riding on it. (Ripple voltage is a small ac voltage riding at some dc voltage level. Normally, ripple voltage is an unwanted ac voltage created by the filter section of a power supply.) This signal now goes to the regulator where it will be maintained at approximately 180 volts dc to the load.

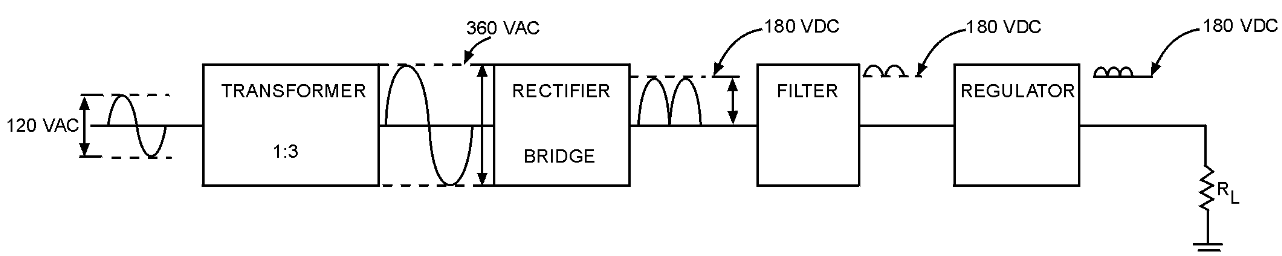


Figure 3-2.—Block diagram of a power supply.

Q1. What are the four basic sections to a power supply?

Q2. What is the purpose of the regulator?

THE TRANSFORMER

The transformer has several purposes: In addition to coupling the input ac signal to the power supply, it also isolates the electronic power supply from the external power source and either steps up or steps down the ac voltage to the desired level. Additionally, most input transformers have separate step-down windings to supply filament voltages to both power supply tubes and the tubes in the external equipment (load). Such a transformer is shown in figure 3-3. Because the input transformer is located in the power supply and is the ultimate source of power for both the load and the power supply, it is called the **POWER TRANSFORMER**. Notice that the transformer has the ability to deliver both 6.3 and 5 volts ac filament voltages to the electron tubes. The High-voltage winding is a 1:3 step-up winding and delivers 360 volts ac to the rectifier. This transformer also has what is called a *center tap*. This center tap provides the capability of developing two high-voltage outputs from one transformer.

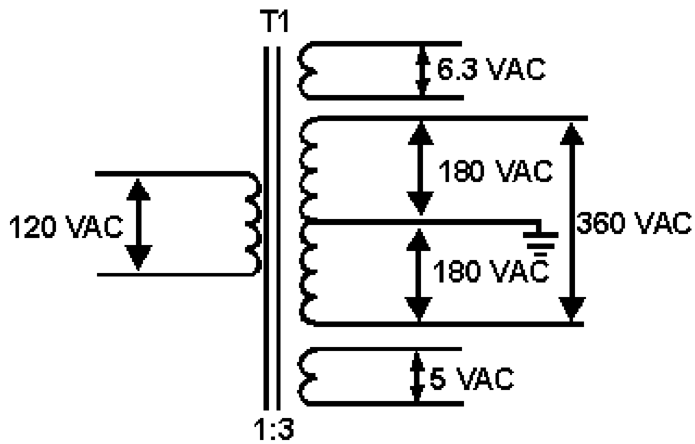


Figure 3-3.—Typical power transformer.

Q3. What are the purposes of the transformer in a power supply?

Q4. For what are the low voltage windings in a transformer used?

Q5. For what is the center tap on a transformer used?

RECTIFIERS

From previous discussions, you know that rectification is the changing of an ac voltage to a pulsating dc voltage. Now let's discuss the process of rectification.

Since a diode vacuum tube will pass current in only one direction, it is ideally suited for converting alternating current to direct current. If an ac voltage is applied to a diode, the diode will conduct **ONLY DURING THE POSITIVE ALTERNATION OF VOLTAGE** when the plate of the diode is made positive with respect to the cathode.

Figure 3-4 shows a diode connected across the 120-volt ac line. During the positive alternation of the source voltage, the sine wave applied to the tube makes the plate positive with respect to the cathode. At this time the diode conducts and plate current flows from the negative supply lead, through the milliammeter, through the tube, and to the positive supply lead. This is indicated by the shaded area of the output waveform. This current exists during the entire period of time that the plate is positive with respect to the cathode (for the first 180 degrees of the input sine wave).

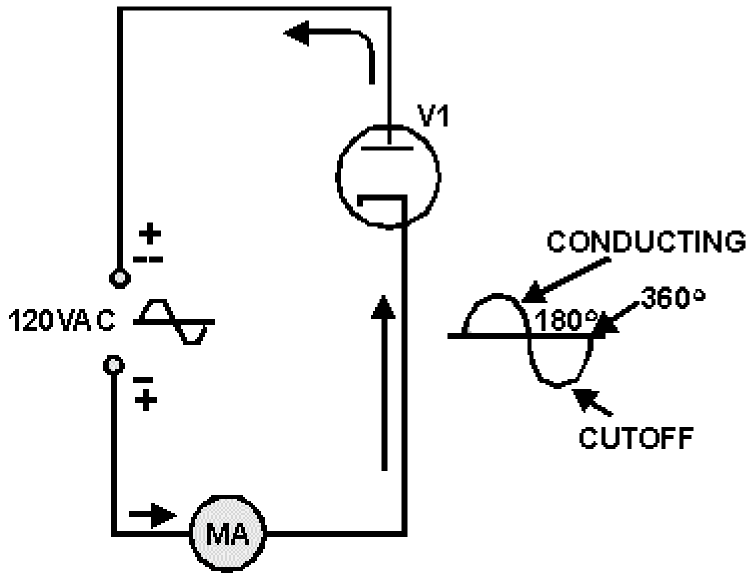


Figure 3-4.—Simple diode rectifier.

During the negative alternation of plate voltage (dotted polarity signs), the plate is driven negative and the tube cannot conduct. When conditions prevent the tube from conducting, the tube is said to be in **CUTOFF**. This is indicated by the dotted waveform. The tube will be in cutoff and no current will flow for the entire negative alternation.

For each 360-degree cycle of input voltage, the tube conducts for 180 degrees and is in cutoff for 180 degrees. The circuit current therefore has the appearance of a series of positive pulses, as shown by the shaded areas. Notice that although the current is in the form of pulses, the current always flows through the circuit in **THE SAME DIRECTION**. Current that flows in pulses in the same direction is called **PULSATING DC**. The diode has thus **RECTIFIED** the input voltage. Although the principle of rectification applies to all rectifier circuits, some rectifiers are more efficient than others. For this reason, we will explain the three rectifier circuits most commonly used in electronics today-the half-wave, full-wave, and bridge.

A Practical Half-Wave Rectifier

Figure 3-5 is a diagram of a complete half-wave rectifier circuit. For the diode to be used as a rectifier, it must be connected in series with a load device (R_L for this circuit), through which the direct current flows. Because Navy electronic equipment requires various input voltages, it is necessary to have a rectified voltage that is greater (or smaller in some cases) than the source voltage. The rectifier plate circuit is supplied power from a step-up (or step-down) transformer. Notice that the transformer has the two secondary windings mentioned earlier. The lower winding supplies high voltage to the plate and cathode of the diode, and the upper winding supplies a low ac voltage to the filaments of the diode. Notice also that the cathode of the diode is connected to the secondary winding of the transformer through the load resistor (R_L). Any current flowing through the tube also flows through the load resistor, causing a voltage to be developed across it. The magnitude of the voltage developed across the load resistor is directly proportional to the amount of current flowing through it (Ohm's law: $E = IR$).

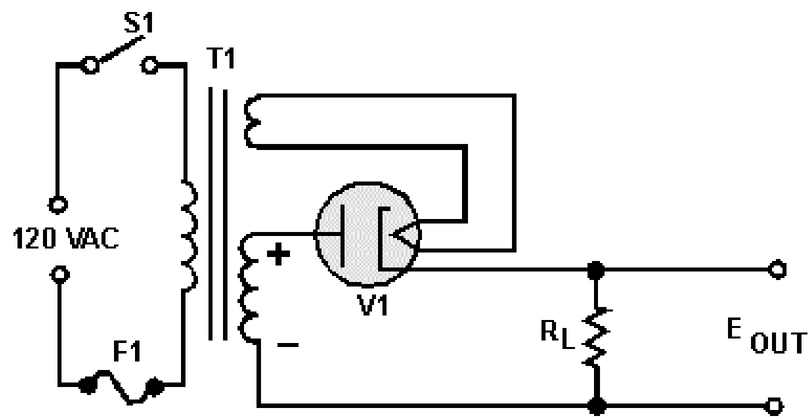


Figure 3-5.—Half-wave rectifier circuit.

You will better understand the operation of the half-wave rectifier circuit if it is redrawn in the form of a simplified series circuit. As you can see in figure 3-6, the diode (V1) and load resistor (R_L) are in series with the secondary winding of the transformer. During the positive alternation of the input, as the voltage in the secondary winding increases, the current through diode (V1) and load resistor (R_L) increases. Since the diode tube and the load resistor form a series circuit, the same current flows through both the tube and the resistor. This current produces a voltage drop across the tube and the load resistor, which have polarities as shown. Since the plate resistance of the tube is only about 500 ohms and the resistance of the load resistor is 10,000 ohms, approximately 95 percent of the applied 425 volts is dropped across the load resistor (425 × .95 = 404 V) and 5 percent (425 × .05 = 21 V) across the tube.

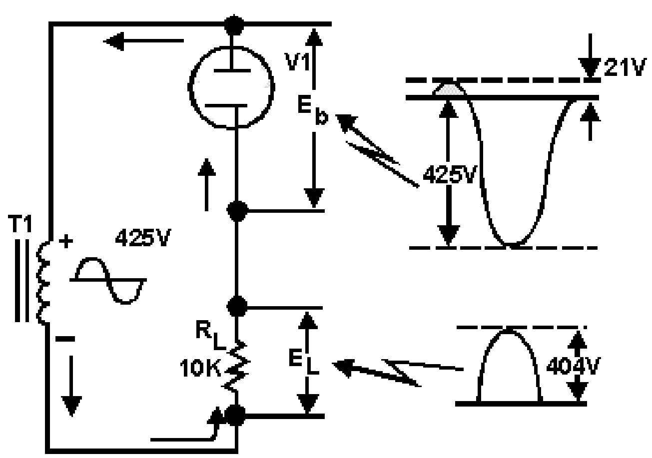


Figure 3-6.—Simplified half-wave rectifier circuit and waveforms.

During the negative half of the alternation of input voltage, the tube cannot conduct and no current flows in the circuit. Since there is no current flow through R_L, the load voltage remains at zero volts throughout the negative alternation. During this time the entire negative alternation is felt across the tube. The reason for this is derived from Kirchhoff's law, which states:

$$E_L + E_b = E_a$$

The sum of the load voltage and diode voltage equals the applied voltage.

Since a half-wave rectifier conducts once for each full cycle of input voltage, the frequency of the pulses is the same as the frequency of the input sine wave. The output pulse frequency is called **RIPPLE FREQUENCY**. If the rectifier circuit is supplied power from a 60-hertz ac line voltage, 60 pulses of load current will occur each second. Therefore, **THE RIPPLE FREQUENCY OF A HALF-WAVE RECTIFIER IS THE SAME AS THE LINE FREQUENCY**.

If a series of current pulses like those obtained from a half-wave rectifier is applied to a load resistance, an average amount of power will be dissipated over a given period of time. This average dc power is determined by the amplitude of the pulses and the time delay between pulses. The higher the peak amplitude of the pulses or the less the time between pulses, the greater the average dc power supplied to the load. To determine average dc voltage (E_{avg}), it is necessary to know the average value of the pulses and the peak value of load voltage. This is illustrated in figure 3-7.

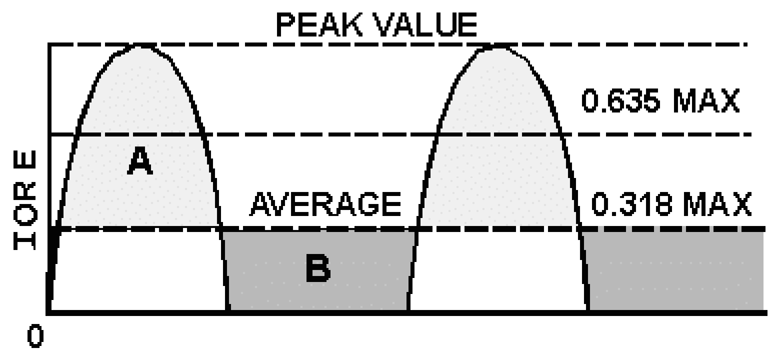


Figure 3-7.—Peak and average values for a half-wave rectifier.

Since current and voltage waveforms in a half-wave rectifier circuit are essentially half sine waves, we can develop a conversion factor. The formula for average value was discussed earlier in *NEETS*, module 2. By now you should know that the average value for a full sine wave is .637 times its peak or maximum value. Therefore, if you want the average value of a half-wave rectifier output, you should multiply half the value of .637 (.318) times the peak or maximum voltage, as expressed in the following equation:

E_{avg} (the average load voltage) = $.318 \times E_{max}$

Where:

E_{max} = The peak value of the load voltage pulse

In most applications the drop across the rectifier tube is small compared to the load voltage, so we can assume E_{max} in our equation to be the same as the peak value of the input sine wave.

Since the load current has the same wave shape as the load voltage, we can modify the equation so that it applies to the load current. Thus,

I_{avg} (the average load current) = $.318 \times I_{max}$

Where:

I_{max} = The peak load current

If a line is drawn through the rectified waveform at a point that is 0.318 of the distance from zero to maximum, the waveform will be divided so that area A is equal to area B (fig. 3-7). Therefore, current or voltage pulses with a value of .318 of the peak value have the same effect on the load as a steady voltage or current.

The half-wave rectifier uses the transformer during only one-half of the cycle. Therefore, for any given size transformer, less power is developed than if the transformer were used on both halves of the cycle. In other words, to obtain large amounts of power, the half-wave transformer must be relatively large in comparison to what it would have to be if both halves of the cycle were used. This disadvantage limits the use of the half-wave rectifier to applications that require a very small current drain. The half-wave rectifier is widely used for commercial ac and dc radio receivers and other applications where inexpensive voltage supplies will suffice. As you can see from your study on half-wave rectifiers, this type of circuit placed many limitations on electronic equipment. For this reason another type of rectifier circuit had to be developed. One of the factors that had to be considered was how to use the full output from the transformer to obtain the highest average voltage and current. Thus, the **FULL-WAVE** rectifier was developed.

Q6. Does a rectifier tube conduct on the positive or negative alternation of the input signal?

Q7. What term is used to describe the period when the diode is not conducting?

Q8. Current that flows in pulses in the same direction is called _____.

Q9. For a diode to act as a rectifier, should it be connected in series or parallel with the load?

Q10. What is the Ripple frequency of a half-wave rectifier if the input frequency is 60 Hz?

Q11. What is the equation for determining average voltage in a half-wave rectifier?

The Conventional Full-Wave Rectifier

A full-wave rectifier is a device that has two or more diodes arranged so the load current flows in the same direction during each half cycle of the ac supply.

A schematic diagram of a simple full-wave rectifier is shown in figure 3-8. The transformer supplies the source voltage for two rectifier tubes (V1 and V2). This power transformer has a **CENTER-TAPPED** high-voltage secondary winding that is divided into two equal parts (W1 and W2). W1 provides the source voltage for V1 and the other winding (W2) provides the source voltage for V2. The connections to the diodes are arranged so that the diodes conduct on alternate half cycles.

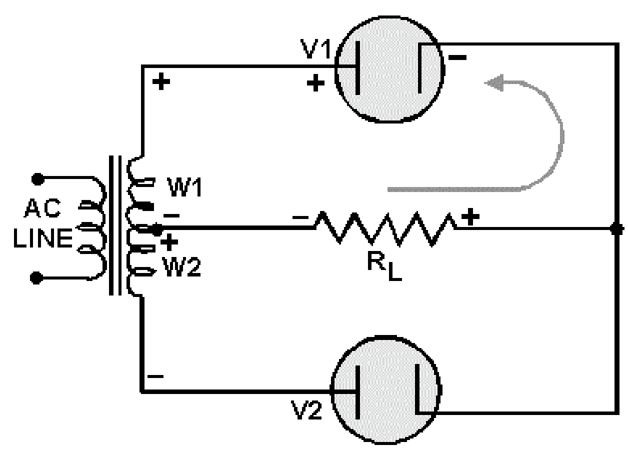


Figure 3-8.—Simple full-wave rectifier (first alternation).

During one alternation of the secondary voltage, the polarities will be as shown in figure 3-8. The source for diode V2 is the voltage induced into the lower half of the transformer secondary winding (W2). At the specific instant of time shown in the figure, the plate voltage on V2 is negative, and V2 cannot conduct.

Throughout the period of time during which the plate of V2 is negative, the plate of V1 is positive. This is illustrated by the polarity signs across W1, which indicate the source for V1. Since the plate of V1 is positive, it conducts, causing current to flow through the load resistor in the direction shown by the arrow.

Figure 3-9 shows the next half cycle of secondary voltage. As you can see, the polarities across W1 and W2 are reversed. During this alternation the plate of V1 is driven negative and V1 cannot conduct.

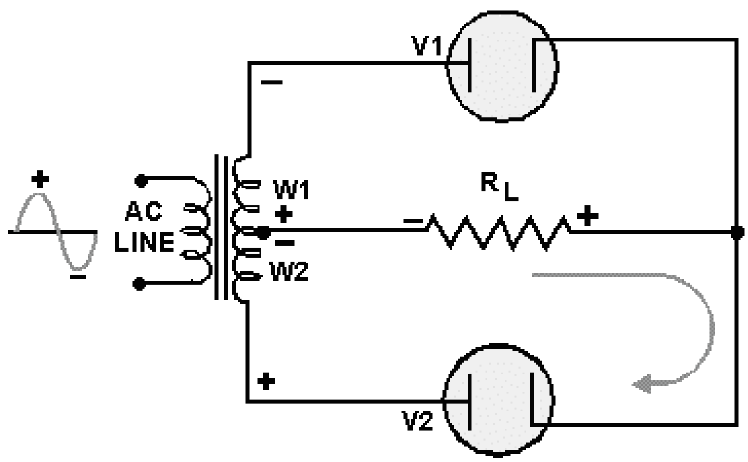


Figure 3-9.—Simple full-wave recliner (second alternation).

For the period of time that V1 is negative, the plate of V2 is positive, permitting V2 to conduct. Notice that the plate current of V2 passes through the load resistor in the same direction as did the plate current of V1. In this circuit arrangement, a pulse of load current flows during each alternation of the input cycle. Since both alternations of the input voltage cycle are used, the circuit is called a **FULL-WAVE RECTIFIER.**

Now that you have a basic understanding of how a full-wave rectifier works, let's cover in detail a practical full-wave rectifier and its waveforms.

A Practical Full-Wave Rectifier

A practical full-wave rectifier circuit is shown in figure 3-10. It uses two diodes (V1 and V2) and a center-tapped transformer (T1). When the center tap is grounded, the voltages at the opposite ends of the secondary windings are 180 degrees out of phase with each other. Thus, when the voltage at point A is positive with respect to ground, the voltage at point B is negative with respect to ground. Let's examine the operation of the circuit during one complete cycle.

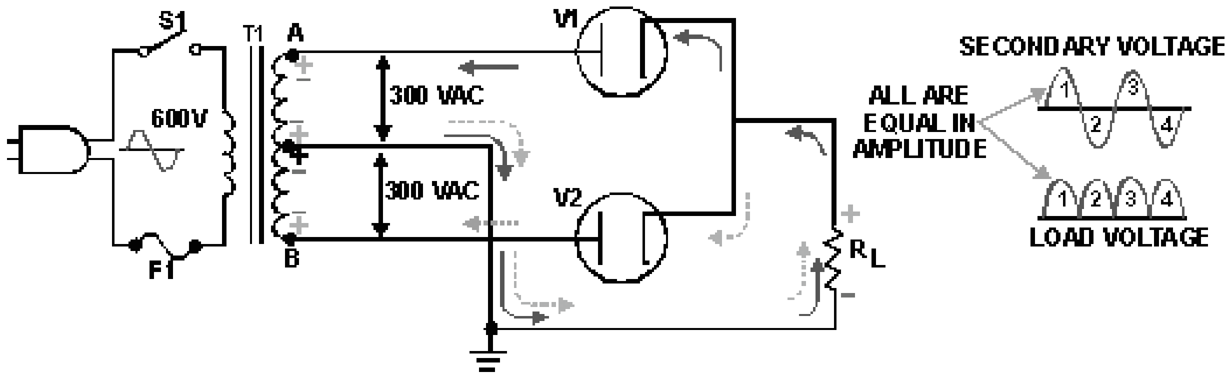


Figure 3-10.—Complete full-wave rectifier.

During the first half-cycle (as indicated by solid arrows) the plate of V1 is positive with respect to ground and the plate of V2 is negative. As shown, current flows from ground (center tap), up through the load resistor (R_L), through diode V1 to point A. In the transformer, current flows from point A, through the upper winding and back to ground (center tap). When V1 conducts, it acts like a closed switch so that the positive half-cycle is felt across the load.

During the second half-cycle (broken lines), the polarity of the applied voltage has reversed Now the plate of V2 is positive with respect to ground and the plate of V1 is negative. Only V2 can conduct. Current now flows, as shown, from ground (center tap), up through the load resistor (R_L), through diode V2 to point B of T1. In the transformer, current flows from point B up through the lower windings and back to ground (center tap). Notice that the current flows across the load resistor (R_L) in the **SAME DIRECTION** for both halves of the input cycles.

The output waveform from the full-wave rectifier consists of two pulses of current (or voltage) for each cycle of input voltage. The ripple frequency at the output of the full-wave rectifier is therefore **TWICE THE LINE FREQUENCY**.

The higher ripple frequency at the output of a full-wave rectifier has a distinct advantage: Because of the higher pulse frequency, the output is closely approximate to pure dc. This higher frequency also makes filtering much easier than the output of the half-wave rectifier.

In terms of peak value, the average value of current and voltage at the output of the full-wave rectifier is twice as great as the average current or voltage at the output of the half-wave rectifier. The relationship between peak and average values is illustrated in figure 3-11.

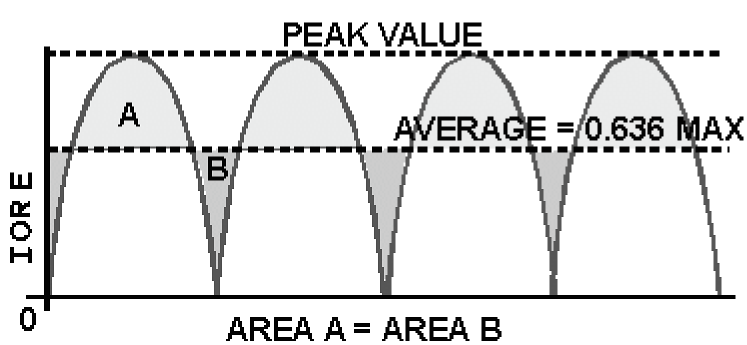


Figure 3-11.—Peak and average values for a full-wave rectifier.

Since the output waveform is essentially a sine wave with both alternations at the same polarity, the average current or voltage is 63.7 percent (or .637) of the peak current or voltage.

As an equation:

E_{avg} (the average load voltage) = $.637 \times E_{max}$

Where

E_{max} = The peak value of the load voltage pulse

And

I_{avg} (the average load current) = $.637 \times I_{max}$

Where:

I_{max} = The peak value of the load current pulse

Example: The total voltage across the high-voltage secondary of a transformer used to supply a full-wave rectifier is 600 volts. Find the average load voltage. (Ignore the drop across the rectifier tube.)

Solution: Since the total secondary voltage is 600 volts, each diode is supplied one-half of this value, or 300 volts. As the secondary voltage is an rms value, the peak load voltage is:

$E_{max} = 1.414 \times E_s$

$E_{max} = 1.414 \times 300$

$E_{max} = 424$ volts

The average load voltage is:

$E_{avg} = .637 \times E_{max}$

$E_{avg} = .637 \times 424$

$E_{avg} = 270$ volts

NOTE: If you have problems with this equation, review *NEETS*, module 2, pertaining to this area.

As you may recall from your past studies in electricity, there are advantages and disadvantages in every circuit. The full-wave rectifier is no exception. In studying the full-wave rectifier, you have found that when the output frequency is doubled, the average voltage is also doubled, and the resulting signal is much easier to filter because of the high-ripple frequency. The only disadvantage is that the peak voltage in a full-wave rectifier is only half the peak voltage in a half-wave rectifier. This is because the secondary of the power transformer in a full-wave rectifier is center tapped; therefore only half the source voltage goes to each diode.

Fortunately, there is a rectifier that produces the same peak voltage as a half-wave rectifier and the same ripple frequency as a full-wave rectifier. This circuit, called the **BRIDGE RECTIFIER**, will be the chapter of our next discussion.

Q12. What is the ripple frequency of a full-wave rectifier with an input frequency of 60 Hz?

Q13. What is the average voltage (E_{avg}) output of a full-wave rectifier that has an output of 100 volts peak?

The Bridge Rectifier

When four diodes are connected as shown in figure 3-12, the circuit is called a **BRIDGE RECTIFIER**. The input to the circuit is applied to the diagonally opposite corners of the network, and the output is taken from the remaining two corners.

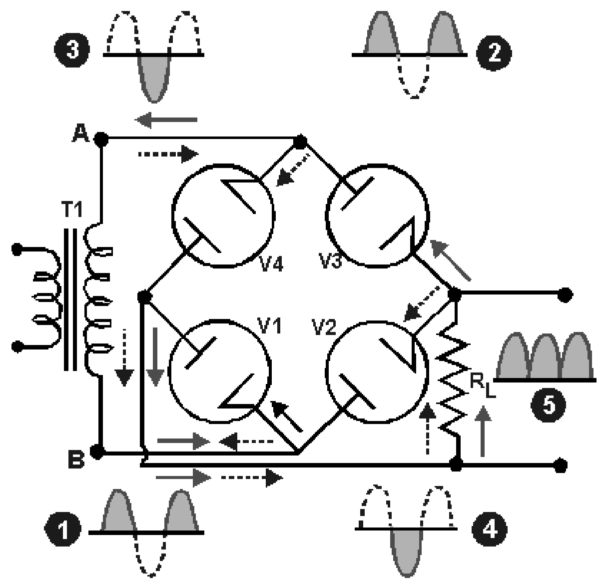


Figure 3-12.—Bridge rectifier circuit.

During one half-cycle of the applied voltage, point A becomes positive with respect to point B by the amount of voltage induced into the secondary of the transformer. During this time, the voltage between points A and B may be considered to be impressed across V1, the load resistor R_L, and V3, in series. The voltage applied across these tubes makes their plates more positive than their cathodes, and current flows from point B through tube V1 in an upward direction across the load resistor, through tube V3, to point A. This path is indicated by the solid arrows. The waveform is shown as numbers (1) and (2).

One half-cycle later, the polarity across the secondary reverses, making the plates of V1 and V3 negative with respect to their cathodes. At the same time, the plates of V2 and V4 become positive with respect to their cathodes, and current flows in the direction indicated by the dashed arrows. The current through R_L is always in the same direction. This current, in flowing through R_L, develops a voltage corresponding to that shown in waveform (5) of the figure. The bridge rectifier is a full-wave rectifier since current flows through the load during both half cycles of the applied alternating voltage.

One advantage of a bridge rectifier over a conventional full-wave rectifier is that with a given transformer, the bridge rectifier produces a voltage output that is nearly twice that of the conventional full-wave circuit. We can show this by assigning values to some of the components as shown in figure 3-13, views (A) and (B). Assume that the same transformer is used in both circuits. The peak voltage developed between points X and Y is 1,000 volts in both circuits. In the conventional full-wave circuit, view (A), the peak voltage from the center tap to either X or Y is 500 volts. Since only one diode can conduct at any instant, the maximum voltage that can be rectified at any instant is 500 volts. Therefore, the maximum voltage that appears across the load resistor is nearly, but never exceeds, 500 volts (because of the small voltage drop across the tube). In the bridge rectifier of view (B), the maximum voltage that can be rectified is the full secondary voltage, 1,000 volts. Therefore, the peak output voltage across the load resistor is nearly 1,000 volts. Thus, with both circuits using the same transformer, the full-wave bridge circuit produces a higher output voltage than the conventional full-wave rectifier.

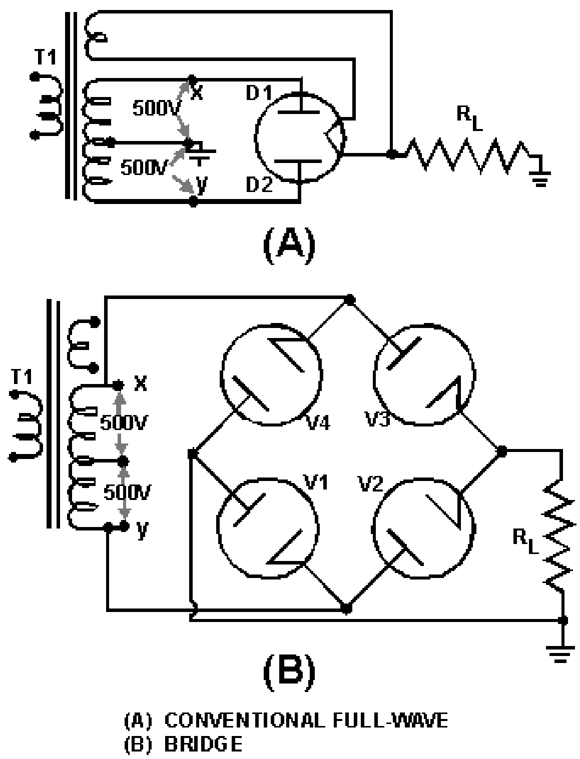


Figure 3-13.—Comparison of conventional full-wave and bridge rectifiers: A. Conventional full-wave circuit

A second advantage of the bridge rectifier is the low ratio of peak inverse voltage to average output voltage. For this reason bridge rectifiers that use vacuum tubes are widely used in high-voltage power supply applications.

If directly heated diodes are used in a bridge rectifier, three separate filament transformers are required. This is due to the different potentials existing at the filaments of the diodes. The filaments of V2 and V3 in figure 3-14 are at the same potential, but the filament of V1 is at a different potential from either V2 or V4. The three filament transformers must be well insulated from each other, and from ground, because of the high potentials to which they are subjected. The use of indirectly heated diodes would solve the filament transformer problem, but the high potential difference between cathode and heater would be likely to result in arcing.

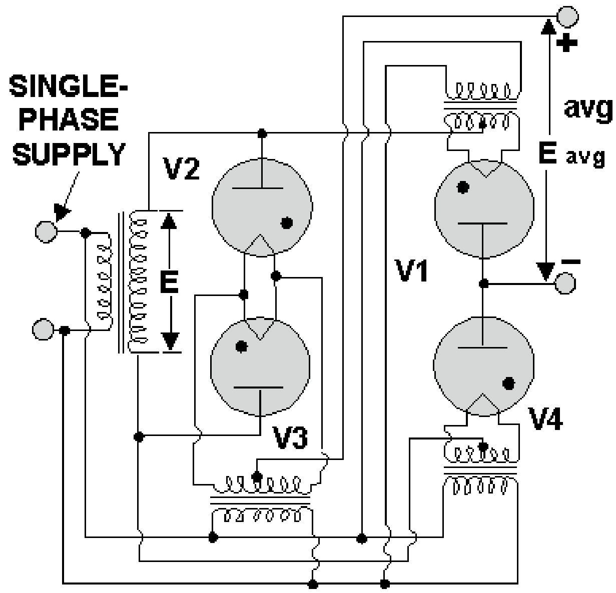


Figure 3-14.—Bridge rectifier with filament transformers.

Q14. What is the main disadvantage of the conventional full-wave rectifier?

Q15. What main advantage does a bridge rectifier have over a conventional full-wave rectifier?

FILTERS

While the output of a rectifier is a pulsating dc, most electronic circuits require a substantially pure dc for proper operation. This type of output may be provided by placing single or multisection filter circuits between the output of the rectifier and the load.

There are four basic types of filter circuits:

- Simple capacitor filter
- LC choke-input filter
- LC capacitor-input filter (pi-type)
- RC capacitor-input filter (pi-type)

We will cover the function of each of these filters in detail later in this chapter.

Filtering is done by using various combinations of capacitors, inductors, and resistors. Inductors are used as series impedances to oppose the change in flow of alternating (pulsating dc) current. Capacitors are used as shunt elements to bypass the alternating components of the signal around the load (to ground). Resistors are used in place of inductors in low current applications.

Let's briefly review the properties of a capacitor. First, a capacitor opposes any change in voltage. The opposition to a change in voltage is called capacitive reactance (X_C) and is measured in ohms. The capacitive reactance is determined by the frequency (f) of the applied voltage and capacitance (C) of the capacitor.

$$X_C = \frac{1}{2\pi fC} \text{ or } \frac{.159}{fC}$$

From the formula, you can see that if frequency or capacitance is increased, the X_C will decrease. Since filter capacitors are placed in parallel with the load, a low X_C will provide better filtering than a high X_C. This is done by providing a better shunting effect of the ac around the load, as shown in figure 3-15.

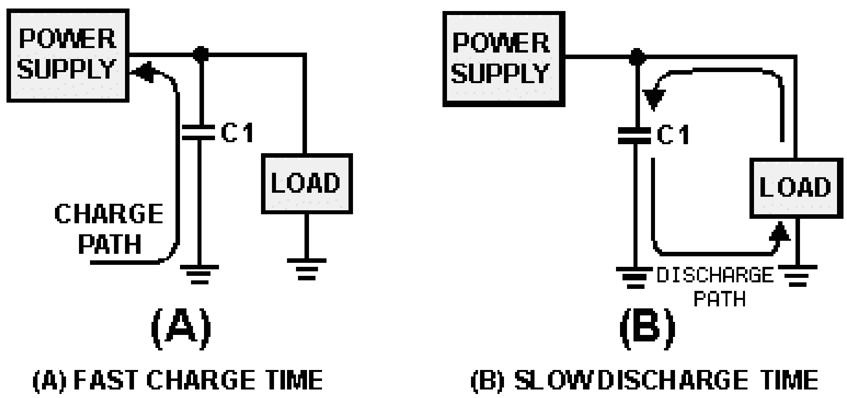


Figure 3-15.—Capacitor filter. Fast charge time

To obtain a steady dc output, the capacitor must charge almost instantaneously to the value of applied voltage. Once charged, the capacitor must retain the charge as long as possible. The capacitor must have a short charge time constant (view A) and a long discharge time constant (view B). This can be done by keeping the internal resistance of the power supply as small as possible (fast charge time) and the resistance of the load as large as possible (slow discharge time).

From your earlier studies in basic electricity, you may remember that one capacitor time constant is defined as the time it takes a capacitor to charge to 63.2 percent of the applied voltage or to discharge to 36.8 percent of its total charge. This can be expressed by the following equation:

$$t = R \times C$$

Where: R represents the resistance of the charge or discharge path

And: C represents the capacitance of the capacitor

You should also recall that a capacitor is considered fully charged after five RC time constants. Referring to figure 3-16, you should see that to obtain a steady dc output voltage, the capacitor should charge rapidly and discharge as slowly as possible.

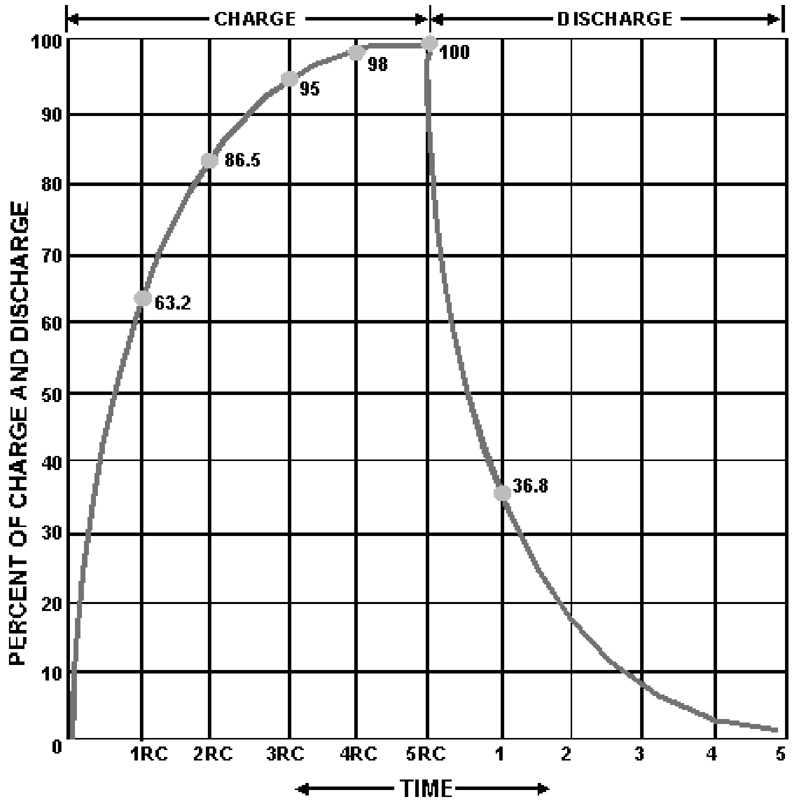


Figure 3-16.—RC time constant chart.

In filter circuits the capacitor is the common element to both the charge and discharge paths. Therefore, to obtain the longest possible discharge time, you want the capacitor to be as large as possible. Another way to look at this is: The capacitor acts as a short circuit around the load (as far as the ac component is concerned), and since

$$X_C = \frac{1}{2\pi fC}$$

the larger the value of the capacitor (C), the smaller the opposition (X_C) or resistance to ac.

Now let's look at inductors and their application in filter circuits. Remember, **AN INDUCTOR OPPOSES ANY CHANGE IN CURRENT**. In case you have forgotten, a change in current through an inductor produces a changing electromagnetic field. The changing field, in turn, cuts the windings of the wire in the inductor and thereby produces a counterelectromotive force (cemf). It is the cemf that opposes the change in circuit current. Opposition to a change in current at a given frequency is called inductive reactance (X_L) and is measured in ohms. The inductive reactance (X_L) of an inductor is determined by the applied frequency and the inductance of the inductor. Mathematically,

$$X_L = 2\pi fL$$

From the preceding formula, you know that if either frequency or inductance is increased, the X_L will increase. Since inductors are placed in series with the load (fig. 3-17), the larger the X_L, the larger the ac voltage developed across the inductor and the smaller the ac voltage developed across the load.

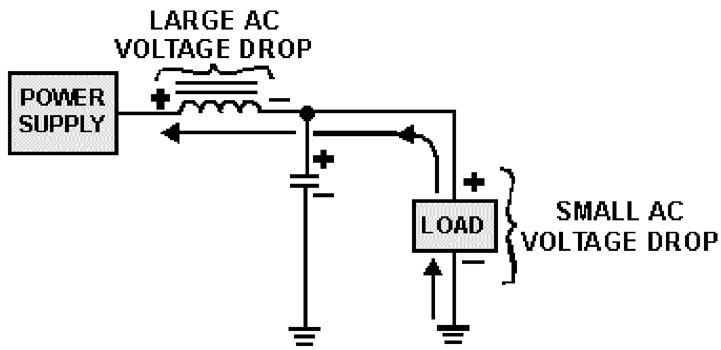


Figure 3-17.—Voltage drops in an inductive filter.

Now back to our circuit. As illustrated in figure 3-18, when the current starts to flow through the coil, an expanding magnetic field builds up around the inductor. This magnetic field around the coil develops the cemf that opposes the change in current. When the rectifier current decreases as shown in figure 3-19, the magnetic field collapses and again cuts the turns (windings) of wire, thus inducing current into the coil. This additional current adds to the rectifier current and attempts to keep it at its original level.

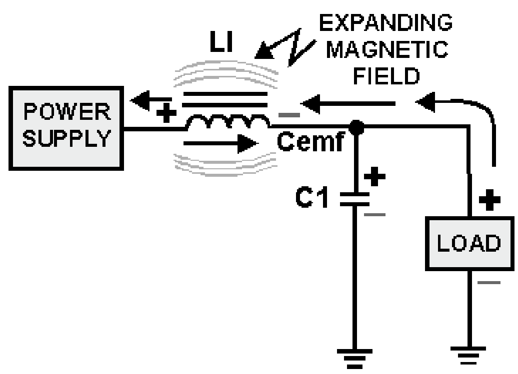


Figure 3-18.—Inductive filter (expanding field).

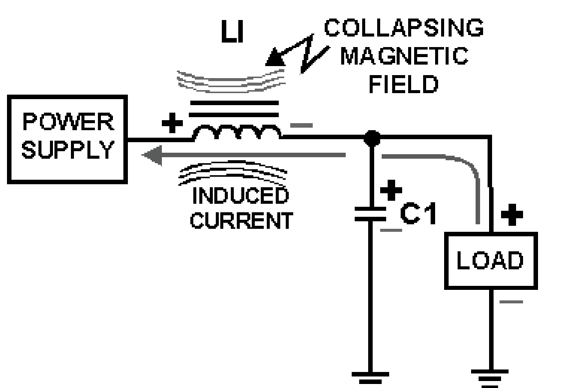


Figure 3-19.—Inductive filter (collapsing field).

Now that you have learned how the components in the filter circuits react to current flow from the rectifier, let's discuss the different types of filter circuits in use today.

Q16. If you increase the value of the capacitor will the X_C increase or decrease? Why?

The Capacitor Filter

The simple capacitor filter is the most basic type of power supply filter. The use of this filter is very limited. It is sometimes used on extremely high-voltage, low-current power supplies for cathode-ray and similar electron tubes that require very little load current from the supply. This filter is also used in circuits where the power-supply ripple frequency is not critical and can be relatively high.

The simple capacitor filter shown in figure 3-20 consists of a single-filter element. This capacitor (C1) is connected across the output of the rectifier in parallel with the load. The RC charge time of the filter capacitor (C1) must be short and the RC discharge time must be long to eliminate ripple action when using this filter. In other words, the capacitor must charge up fast with preferably no discharge at all. Better filtering also results when the frequency is high; therefore, the full-wave rectifier output is easier to filter than the half-wave rectifier because of its higher frequency.

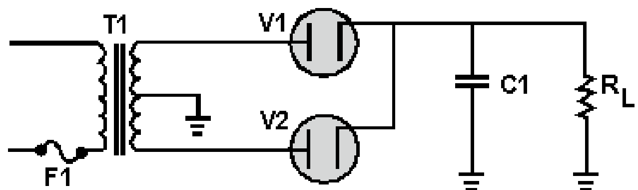


Figure 3-20.—Full-wave rectifier with a capacitor filter.

To understand better the effect that filtering has on E_{avg}, compare the rectifier circuits without filters in figure 3-21 to those with filters in figure 3-22. The output waveforms in figure 3-21 represent the unfiltered outputs of the half-wave and full-wave rectifier circuits. Current pulses flow through the load resistance (R_L) each time a diode conducts. The dashed line indicates the average value of output voltage. For the half-wave rectifier, E_{avg} is less than half the peak output voltage (or approximately 0.318 of the peak output voltage). For the full-wave rectifier, E_{avg} is approximately 0.637. This value is still much less than the applied voltage. With no capacitor connected across the output of the rectifier circuit, the waveform has a large pulsating component (ripple) compared with the average or dc component. Now refer to figure 3-22. When a capacitor is connected across the output (in parallel with R_L), the average value of output voltage (E_{avg}) is increased due to the filtering action of capacitor C1.

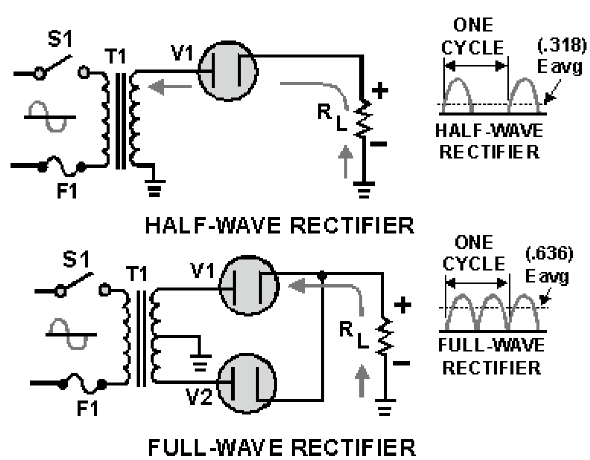


Figure 3-21.—Half-wave/full-wave rectifiers (without filters).

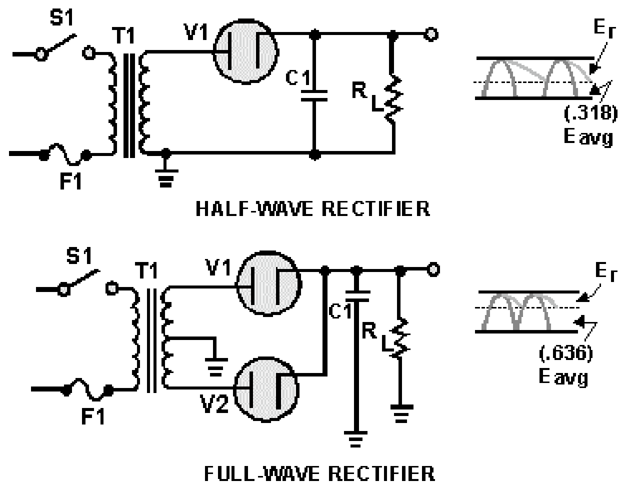


Figure 3-22.—Half-wave/full-wave rectifiers (with capacitor filters).

The value of the capacitor is fairly large (several microfarads); it thus presents a relatively low reactance to the pulsating current and stores a substantial charge. The rate of charge for the capacitor is limited only by the relatively low resistance of the conducting diode. The RC charge time of the circuit is, therefore, relatively short. As a result, when the pulsating voltage is first applied to the circuit, the capacitor charges rapidly and almost reaches the peak value of the rectified voltage within the first few cycles. The capacitor attempts to charge to the peak value of the rectified voltage anytime a diode is conducting, and tends to retain its charge when the rectifier output falls to zero. (The capacitor cannot discharge immediately). The capacitor slowly discharges through the load resistance (R_L) during the time the rectifier is nonconducting.

The rate of discharge of the capacitor is determined by the value of capacitance and the value of the load resistance. If the capacitance and load resistance values are large, the RC discharge time for the circuit is relatively long.

From the waveforms shown in figure 3-22, you should see that the addition of C1 to the circuit results in an increase in the average value of output voltage (E_{avg}) and a reduction in the amplitude of the ripple component (E_r) present across the load resistance.

Now, let's consider a complete cycle of operation using a half-wave rectifier, a capacitive filter (C1), and a load resistor (R_L).

As shown in figure 3-23, C1 is assumed to be large enough to ensure a small reactance to the pulsating rectified current. The resistance of R_L is assumed to be much greater than the reactance of C1 at the input frequency.

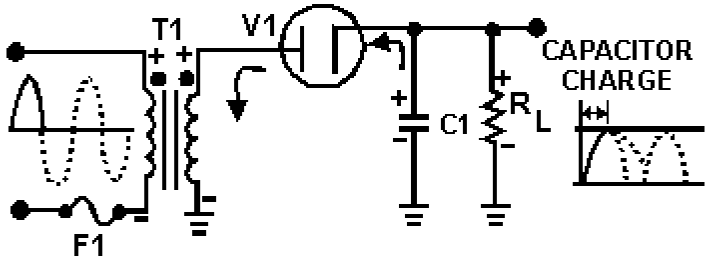


Figure 3-23.—Half-wave rectifier capacitor filter (positive input cycle).

When the circuit is energized, the diode conducts on the positive half cycle and current flows through the circuit allowing C1 to charge. C1 will charge to approximately the peak value of the input voltage. The charge is less than the peak value because of the voltage drop across diode V1. The charge on C1 is indicated by the heavy, solid line on the waveform.

As illustrated in figure 3-24, the diode (V1) cannot conduct on the negative half cycle because the plate of V1 is negative in respect to the cathode. During this interval, C1 discharges through load resistance R_L. The discharge of C1 produces the downward slope indicated by the solid line on the waveform in the figure.

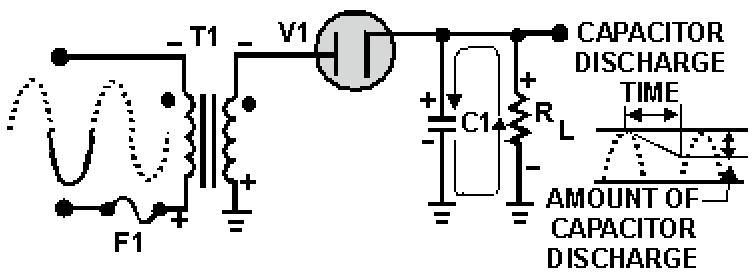


Figure 3-24.—Half-wave rectifier capacitor filter (negative input cycle).

During the discharge period, in contrast to the abrupt fall of the applied ac voltage from peak value to zero, the voltage across C1 (and thus across R_L) gradually decreases until the time of the next half cycle of rectifier operation. Keep in mind that for good filtering, the filter capacitor should <u>charge</u> as <u>fast</u> as possible and <u>discharge</u> as <u>little</u> as possible.

Since practical values of C1 and R_L ensure a more or less gradual decrease of the discharge voltage, a substantial charge remains on the capacitor at the time of the next half cycle of operation. As a result, no current can flow through the diode until the rising ac input voltage at the plate of the diode exceeds the voltage of the charge remaining on C1. The charge on C1 is the cathode potential of the diode. When the potential on the plate exceeds the potential on the cathode (the charge on C1), the diode again conducts, and C1 commences to charge to approximately the peak value of the applied voltage.

After the capacitor has charged to its peak value, it begins to discharge. Since the fall of the ac input voltage on the plate is considerably more rapid than the decrease on the capacitor voltage, the cathode quickly becomes more positive than the plate, and the diode ceases to conduct.

The operation of the simple capacitor filter using a full-wave rectifier is basically the same as the operation we discussed for the half-wave rectifier. Notice in figure 3-25 that because one of the diodes is always conducting on either alternation, the filter capacitor charges and discharges during each half cycle. (Note that each diode conducts only for that portion of time when the peak secondary voltage is greater than the charge across the capacitor.)

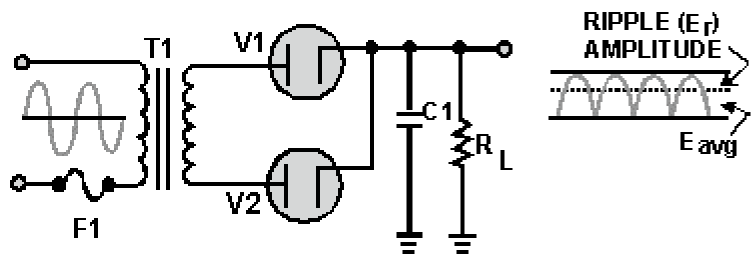


Figure 3-25.—Full-wave rectifier (with capacitor filter).

We stated before that a major advantage of full-wave and bridge rectifiers over half-wave rectifiers is the ease of filtering their output voltages. You can now see the reason for this. The ripple frequency is doubled; therefore, the time period the capacitor is allowed to discharge is cut in half. This means that the capacitor discharges less. Thus, ripple amplitude is less, and a smoother output voltage occurs.

Another thing to keep in mind is that the ripple component (E_r) of the output voltage is an ac voltage and the average output voltage (E_{avg}) is the dc component of the output. Since the filter capacitor offers a relatively low impedance to ac, the majority of the ac component flows through the filter capacitor. The ac component is therefore bypassed (shunted) around the load resistance and the entire dc component (or E_{avg}) flows through the load resistance. To clarify this statement, let's take a look at the formula for X_C in a half-wave and full-wave rectifier. First, you must establish some values for the circuit.

HALFWAVE RECTIFIER

FREQUENCY AT
RECTIFIER OUTPUT: 60 Hz

VALUE OF FILTER
CAPACITOR: 30µF

LOAD RESISTANCE: 10kΩ

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{.159}{fC}$$

$$X_C = \frac{.159}{60 \times .000030}$$

$$X_C = \frac{.159}{.0018}$$

$$X_C = 88.3\Omega$$

FULLWAVE RECTIFIER

FREQUENCY AT
RECTIFIER OUTPUT: 120z

VALUE OF FILTER
CAPACITOR: 30µF

LOAD RESISTANCE: 10k

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{.159}{fC}$$

$$X_C = \frac{.159}{120 \times .000030}$$

$$X_C = \frac{.159}{.0036}$$

$$X_C = 44.16\Omega$$

As you can see from the calculations, when the output frequency of the rectifier is doubled, the impedance of the capacitor is reduced by one-half. Therefore, when the simple capacitor filter is used in conjunction with a full-wave or bridge rectifier, improved filtering is provided because the increased ripple frequency decreases the capacitive reactance of the filter capacitor. This allows the ac component to be passed through the capacitor more easily. Therefore, the output of a full-wave rectifier is much easier to filter than that of a half-wave rectifier.

It should be obvious that the smaller the X_C of the filter capacitor in respect to the load resistance, the better the filtering action. By using the largest possible capacitor, we achieve the best filtering. The load resistance is also an important consideration. If load resistance is made small, the load current increases, and the average value of output voltage (E_{avg}) decreases. The RC discharge time constant is a direct function of the value of the load resistance; therefore, the rate of capacitor voltage discharge is a direct function of the current through the load. The greater the load current, the more rapid the discharge of the capacitor, and the lower the average value of output voltage. For this reason, the simple capacitor filter is seldom used with rectifier circuits that must supply a relatively large load current.

Q17. What is the most basic type of filter?

Q18. In a capacitor filter, is the capacitor in series or parallel with the load?

Q19. Is better filtering achieved at a high frequency or at a low frequency at the input of the filter?

Q20. Does a filter circuit increase or decrease the average output voltage?

Q21. What determines the rate of discharge of the capacitor in a filter circuit?

Q22. Does low ripple voltage indicate good or bad filtering?

Q23. Is a full-wave rectifier output easier to filter than that of a half-wave rectifier?

In general, with the supply voltage removed from the input to the filter circuit, one terminal of the filter capacitor can be disconnected from the circuit.

CAUTION

REMEMBER-AN UNDISCHARGED CAPACITOR RETAINS ITS POLARITY AND HOLDS ITS CHARGE FOR LONG PERIODS OF TIME. TO BE SAFE, USE A PROPER SAFETY SHORTING PROBE TO DISCHARGE THE CAPACITOR TO BE TESTED WITH THE POWER OFF BEFORE CONNECTING TEST EQUIPMENT OR DISCONNECTING THE CAPACITOR.

You can check the capacitor by using a capacitance analyzer to determine its effective capacitance and leakage resistance. During these checks it is very important that you observe correct polarity if the capacitor is an electrolytic. A decrease in capacitance or losses within the capacitor can cause the output to be below normal and also cause excessive ripple amplitude.

If a suitable capacitance analyzer is not available, you can get an indication of leakage resistance by using an ohmmeter. You can make resistance measurements across the terminals of the capacitor to determine whether it is shorted, leaky, or open. When you test electrolytic capacitors, set the ohmmeter to the high range, and connect the test probes across the capacitor. Be careful to observe polarity. This is important because current flows through an electrolytic capacitor with less opposition in one direction than in the other. If you do not observe the correct polarity, you will get an incorrect reading and you may damage the meter. When you first connect the test probes, a large deflection of the meter should take place, and then the pointer should return slowly toward the infinite-ohms position as the capacitor charges. For a good capacitor with a rated working voltage of 450 volts dc, the final reading on the ohmmeter should be over 500,000 ohms. (A rough rule of thumb for high-voltage capacitors is at least 1000 ohms per volt.) Low-voltage electrolytic capacitors (below 100 volts rating) should indicate approximately 100,000 ohms.

If the ohmmeter does not deflect when you make the resistance check explained above, you have found an open-circuit capacitor.

A steady full-scale deflection of the pointer at zero ohms indicates that the capacitor being tested is shorted.

An indication of a leaky capacitor is a steady reading on the scale somewhere between zero and the minimum acceptable value. (Be certain this reading is not caused by an in-circuit shunting part.) To be valid, these capacitor checks should be made with the capacitor completely disconnected from the circuit in which it operates.

In high-voltage filter capacitor applications, paper and oil-filled capacitors are used in addition to mica and ceramic capacitors (for low-capacitance values). In this case, polarity is of no importance unless the capacitor terminals are marked plus or minus. It is, however, good maintenance practice to use the output polarity of the circuit as a guide, connecting positive to positive, and negative to negative. Thus, any adverse effects of polarity on circuit tests are minimized and the possibility of damage to components or to test equipment is eliminated.

The LC Choke-Input Filter

The LC choke-input filter is used primarily in power supplies where good voltage regulation is important and where the output current is relatively high and subject to varying load conditions. This filter is used in high-power applications such as those found in radar and communication transmitter power supplies.

In figure 3-26 you can see that this filter consists of an input inductor or filter-choke (L1) and an output filter capacitor (C1).

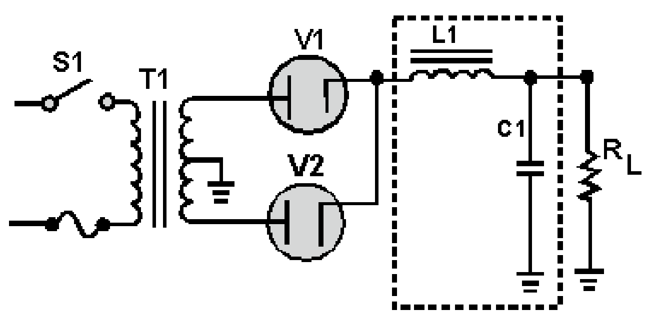


Figure 3-26.—Full-wave rectifier LC choke-input filter.

Inductor L1 is placed at the input to the filter and is in series with the output of the rectifier circuit. Since the action of an inductor is to oppose any change in current flow, the inductor tends to keep a constant current flowing to the load throughout the complete cycle of the applied voltage. As a result, the output voltage never reaches the peak value of the applied voltage; instead, the output voltage approximates the average value of the rectified input to the filter, as shown in figure 3-27.

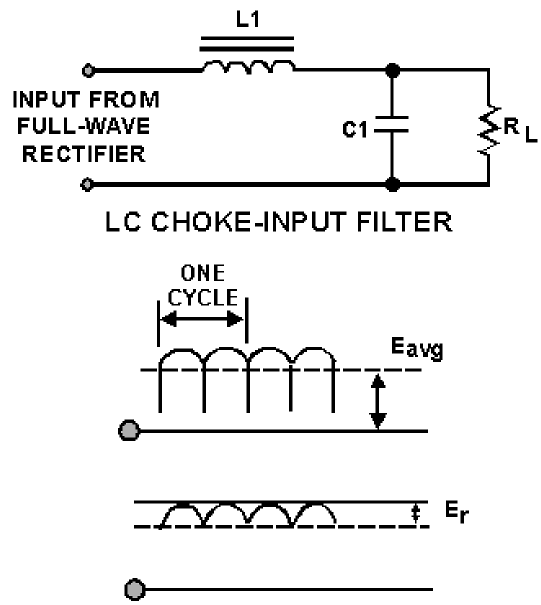


Figure 3-27.—Waveforms for a LC choke-input filter.

The reactance of the inductor (X_L) reduces the amplitude of ripple voltage without reducing the dc output voltage by an appreciable amount. (The dc resistance of the inductor is just a few ohms.)

The shunt capacitor (C1) charges and discharges at the ripple frequency rate, but the amplitude of the ripple voltage (E_r) is relatively small because the inductor (L1) tends to keep a constant current flowing from the rectifier circuit to the load. In addition, the reactance of the shunt capacitor (X_C) presents a low impedance to the ripple component existing at the output of the filter, and thus shunts the ripple component around the load. The capacitor attempts to hold the output voltage relatively constant at the average value of the voltage.

The value of the filter capacitor (C1) must be relatively large to present a low opposition (X_C) to the pulsating current and to store a substantial charge. The rate of the charge for the capacitor is limited by the low impedance of the ac source (transformer), the small resistance of the diode, and the counter emf developed by the coil. Therefore, the RC charge time constant (fig. 3-28) is short compared to its discharge time.

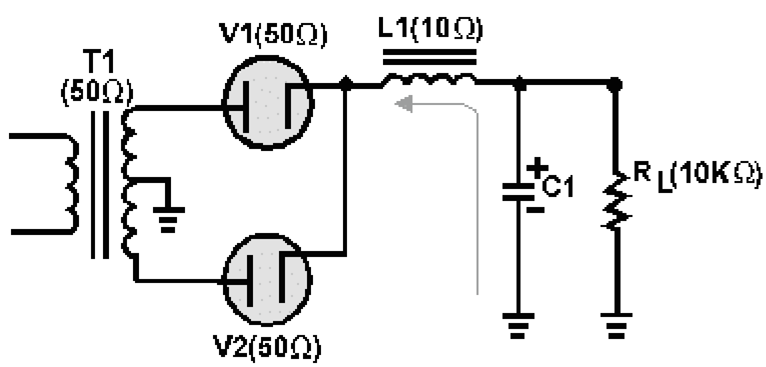


Figure 3-28.—LC choke-input filter (circuit resistance).

As a result, when the pulsating voltage is first applied to the LC choke-input filter, the inductor or filter choke (L1) produces a counter emf that opposes the constantly increasing input voltage. The net result is to effectively prevent the rapid charging of the filter capacitor (C1). Thus, instead of reaching the peak value of the input voltage, C1 only charges to the average value of the input voltage. After the input voltage reaches its peak and decreases sufficiently, the capacitor (C1) attempts to discharge through the load resistance (R_L). C1 will attempt to discharge as indicated in figure 3-29. Because of its relatively long discharge time constant, C1 can only partially discharge.

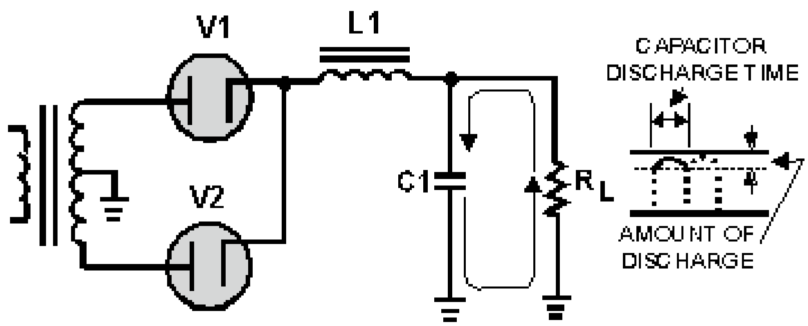


Figure 3-29.—LC choke-input filter (discharge path).

The larger the value of the filter capacitor, the better the filtering action. However, due to the physical size, there is a practical limitation to the maximum value of the capacitor.

The inductor or filter choke (L1) maintains the current flow to the filter output (capacitor C1 and load resistance R_L) at a nearly constant level during the charge and discharge periods of the filter capacitor.

The series inductor (L1) and the capacitor (C1) form a voltage divider for the ac component (ripple) of the applied input voltage. This is shown in figure 3-30. As far as the ripple component is concerned, the inductor offers a high impedance (Z) and the capacitor offers a low impedance. As a result, the ripple component (E_r) appearing across the load resistance is greatly attenuated (reduced). Since the inductance of the filter choke opposes changes in the value of the current flowing through it, the average value of the voltage produced across the capacitor contains a much smaller value of ripple component (E_r), as compared with the value of ripple produced across the coil.

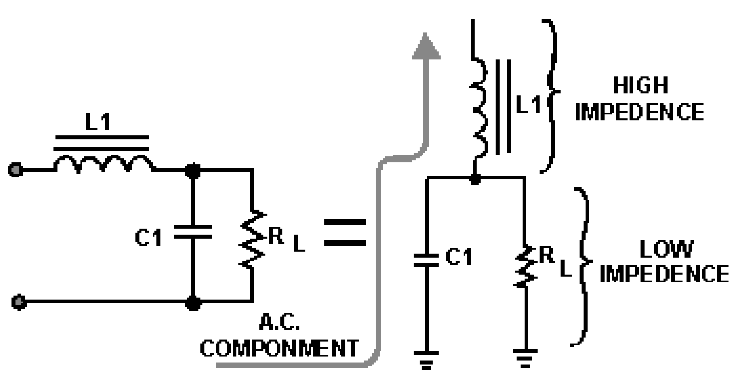


Figure 3-30.—LC choke-input filter (as voltage divider).

Now look at figure 3-31, which illustrates a complete cycle of operation where a full-wave rectifier circuit is used to supply the input voltage to the filter. The rectifier voltage is developed across capacitor C1. The ripple voltage in the output of the filter is the alternating component of the input voltage reduced in amplitude by the filter section.

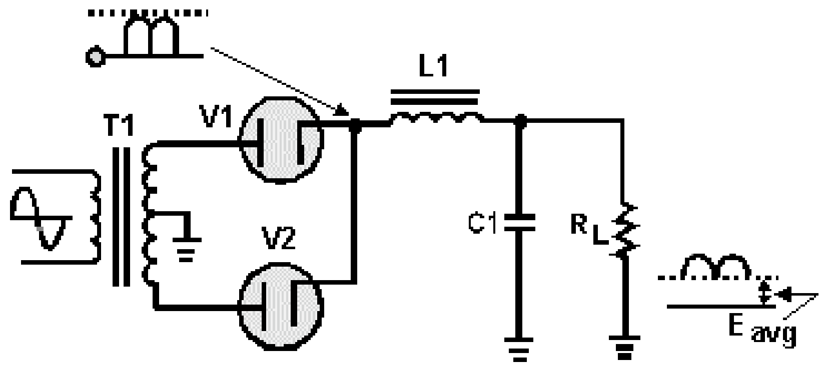


Figure 3-31.—Filtering action of an LC choke-input filter.

Each time the plate of a diode goes positive with respect to the cathode, the diode conducts and C1 charges. Conduction occurs twice during each cycle for a full-wave rectifier. For a 60-hertz supply, this produces a ripple frequency of 120 hertz. Although the diodes alternate (one conducts while the other is nonconducting), the filter input voltage is not steady. As the plate voltage of the conducting diode increases (on the positive half of the cycle), capacitor C1 charges-the charge being limited by the impedance of the secondary transformer winding, the diode's forward (cathode-to-plate) resistance, and the counter emf developed by the choke. During the nonconducting interval, (when the plate voltage drops below the capacitor charge voltage), C1 discharges through the load resistance R_L. The components in the discharge path cause a long time constant; thus C1 discharges slower than it charges.

The choke (L1) is usually of a large value, on the order of 1 to 20 henries, and offers a large inductive reactance to the 120-hertz ripple component produced by the rectifier. Therefore, the effect that L1 has on the charging of the capacitor (C1) must be considered. Since L1 is connected in series with the parallel branch consisting of C1 and R_L, a division of the ripple ac voltage and the output dc voltage occurs. The greater the impedance of the choke, the less the ripple voltage that appears across C1 and the output. The dc output voltage is fixed mainly by the dc resistance of the choke.

Now that you have read how the LC choke-input filter functions, let's take a look at it using actual component values. For simplicity, the input frequency at the primary of the transformer will be 117 volts 60 hertz. We will use both half-wave and full-wave rectifier circuits to provide the input to the filter.

Starting with the half-wave configuration as shown in figure 3-32, the basic parameters are: with 117 volts ac rms applied to the T1 primary, 165 volts ac peak-to-peak is available at the secondary [(117 V) × (1.414) = 165 V]. You should recall that the ripple frequency of this half-wave rectifier is 60 hertz. Therefore, the capacitive reactance of C1 is:

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{(2)(3.14)(60)(10)(10^{-6})}$$

$$X_C = \frac{(1)(10^{6})}{3768}$$

$$X_C = 265\Omega$$

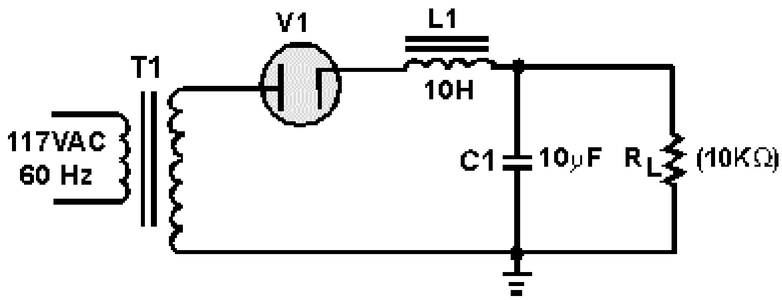


Figure 3-32.—Half-wave rectifier with an LC choke-input filter.

This means that the capacitor (C1) offers 265 ohms of opposition to the ripple current. Note, however, that the capacitor offers an infinite impedance to direct current. The inductive reactance of L1 is:

$$X_L = 2\pi fL$$

$$X_L = (2)\ (3.14)\ (60)\ (10)$$

$$X_L = 3.8\ \textit{Kilohms}$$

This shows that L1 offers a relatively high opposition (3.8 kilohms) to the ripple in comparison to the opposition offered by C1 (265 ohms). Thus, more ripple voltage will be dropped across L1 than across C1. In addition, the impedance of C1 (265 ohms) is relatively low in respect to the resistance of the load (10 kilohms). Therefore, more ripple current flows through C1 than the load. In other words, C1 shunts most of the ac component around the load.

Let's go a step further and redraw the filter circuit so that you can see the voltage divider action. (Refer to figure 3-33.) Remember, the 165 volts peak-to-peak 60 hertz provided by the rectifier consist of both an ac and a dc component. The first discussion will be about the ac component. Looking at figure 3-33, you see that the capacitor (C1) offers the least opposition (265 ohms) to the ac component; therefore, the greatest amount of ac will flow through C1. (The heavy line indicates current flow through the capacitor.) Thus the capacitor bypasses, or shunts, most of the ac around the load.

By combining the X_C of C1 and the resistance of R_L into an equivalent circuit, you will have an equivalent impedance of 265 ohms.

$$Z = \frac{R \times X_C}{\sqrt{(R^2 + X_C^2)}}$$

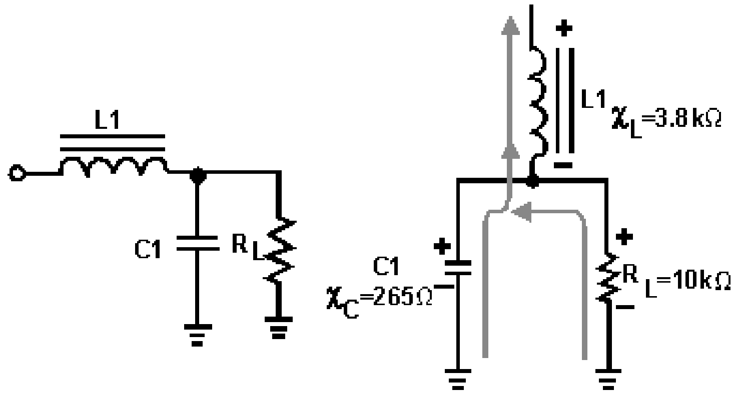


Figure 3-33.—AC component in an LC choke-input filter.

You now have a voltage divider as illustrated in figure 3-34. You should see that because of the impedance ratios, a large amount of ripple voltage is dropped across L1, and a substantially smaller amount is dropped across C1 and R_L. You can further increase the ripple voltage across L1 by increasing the inductance:

$$X_L = 2\pi fL$$

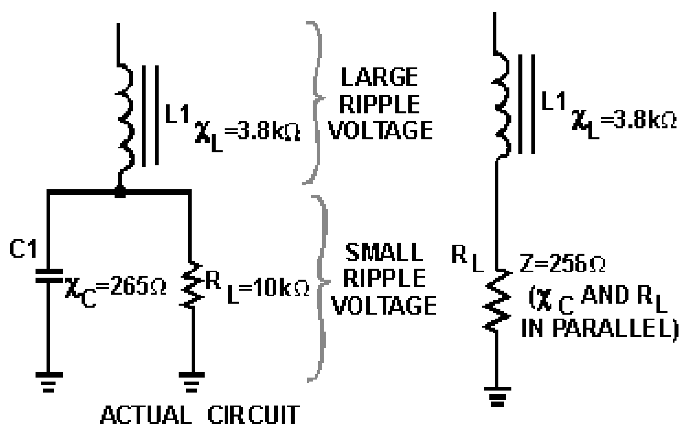


Figure 3-34.—Actual and equivalent circuits.

Now let's discuss the dc component of the applied voltage. Remember, a capacitor offers an infinite (∞) impedance to the flow of direct current. The dc component, therefore, must flow through R_L and L1. As far as the dc is concerned, the capacitor does not exist. The coil and the load are, therefore, in series with each other. The dc resistance of a filter choke is very low (50 ohms average). Therefore, most of the dc component is developed across the load and a very small amount of the dc voltage is dropped across the coil, as shown in figure 3-35.

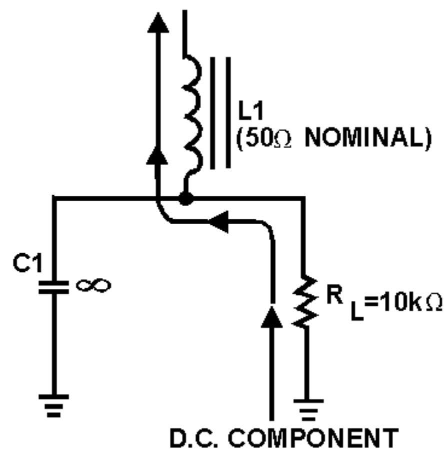


Figure 3-35.—DC component in an LC choke-input filter.

As you may have noticed, both the ac and the dc components flow through L1, and because the coil is frequency sensitive, it provides a large resistance to ac and a small resistance to dc. In other words, the coil opposes any change in current. This property makes the coil a highly desirable filter component. Note that the filtering action of the LC capacitor input filter is improved when the filter is used in conjunction with a full-wave rectifier as shown in figure 3-36. This is due to the decrease in the X_C of the filter capacitor and the increase in the X_L of the choke. Remember, the ripple frequency of a full-wave rectifier is twice that of a half-wave rectifier. For a 60-hertz input, the ripple will be 120 Hertz. Let's briefly calculate the X_C of C1 and the X_L of L1:

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{(2)(3.14)(120)(10)(10^{-6})}$$

$$X_C = \frac{(1)(10^6)}{7536}$$

$$X_C = 132.5\Omega$$

$$X_L = 2\pi fL$$

$$X_L = (2)(3.14)(120)(10)$$

$$X_L = 7.5 \text{ kilohms}$$

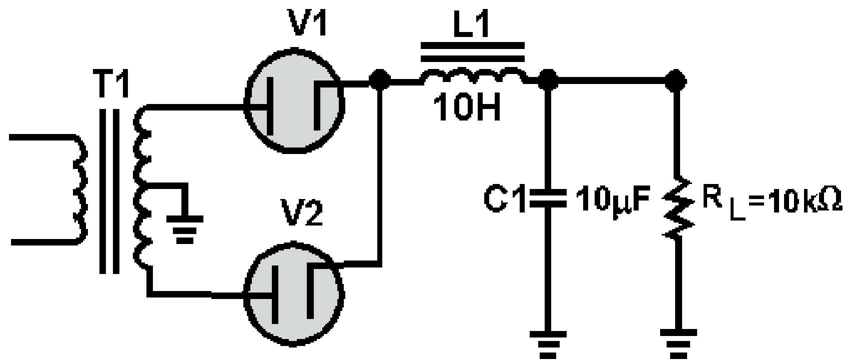


Figure 3-36.—Full-wave rectifier with an LC choke-input filter.

It should be apparent that when the X_C of a filter capacitor is decreased, it provides less opposition to the flow of ac. The greater the ac flow through the capacitor, the lower the flow through the load. Conversely, the larger the X_L of the choke, the greater the amount of ac ripple developed across the choke; consequently, less ripple is developed across the load. This condition provides better filtering.

Q24. In an LC choke-input filter, what prevents the rapid charging of the capacitor?

Q25. What is the value usually chosen for a filter choke?

Q26. If the inductance of a choke-input filter is increased, will the output ripple voltage amplitude (E_r) increase or decrease?

An LC choke-input filter is subject to several problems that can cause it to fail. The filter capacitors are subject to open circuits, short circuits, and excessive leakage. The series inductor is subject to open windings and, occasionally, shorted turns or a short circuit to the core.

The filter capacitor in the choke-input filter circuit is not subject to extreme voltage surges because of the protection offered by the inductor; however, the capacitor can become open, leaky, or shorted.

Shorted turns in the choke may reduce the value of inductance below the critical value. This will result in excessive peak-rectifier current, accompanied by an abnormally high output voltage, excessive ripple amplitude, and poor voltage regulation. A choke winding that is open, or a choke winding that is shorted to the core will result in a no-output condition. A choke winding that is shorted to the core may cause overheating of the rectifier element(s), blown fuses, and so forth.

To check the capacitor, first remove the supply voltage from the input to the filter circuit. Then disconnect one terminal of the capacitor from the circuit. Check the capacitor with a capacitance analyzer to determine its capacitance and leakage resistance. When the capacitor is electrolytic, be sure to use the correct polarity at all times. A decrease in capacitance or losses within the capacitor can decrease the efficiency of the filter and produce excessive ripple amplitude. If a suitable capacitance analyzer is not available, you can use an ohmmeter to check for leakage resistance. The test procedure is the same as that described for the input capacitor filter.

So far, this section has discussed in detail the operation and troubleshooting of the basic inductive and capacitive filter circuits. For the two remaining types of filters, we will discuss only the differences between them and the other basic filters.

Resistor-Capacitor (RC) Filters

The RC capacitor-input filter is limited to applications in which the load current is small. This type of filter is used in power supplies where the load current is constant and voltage regulation is not necessary. For example, RC filters are used in high-voltage power supplies for cathode-ray tubes and as part of decoupling networks for multistage amplifiers.

Figure 3-37 shows an RC capacitor-input filter and its associated waveforms. Both half-wave and full-wave rectifiers are used to provide the inputs.

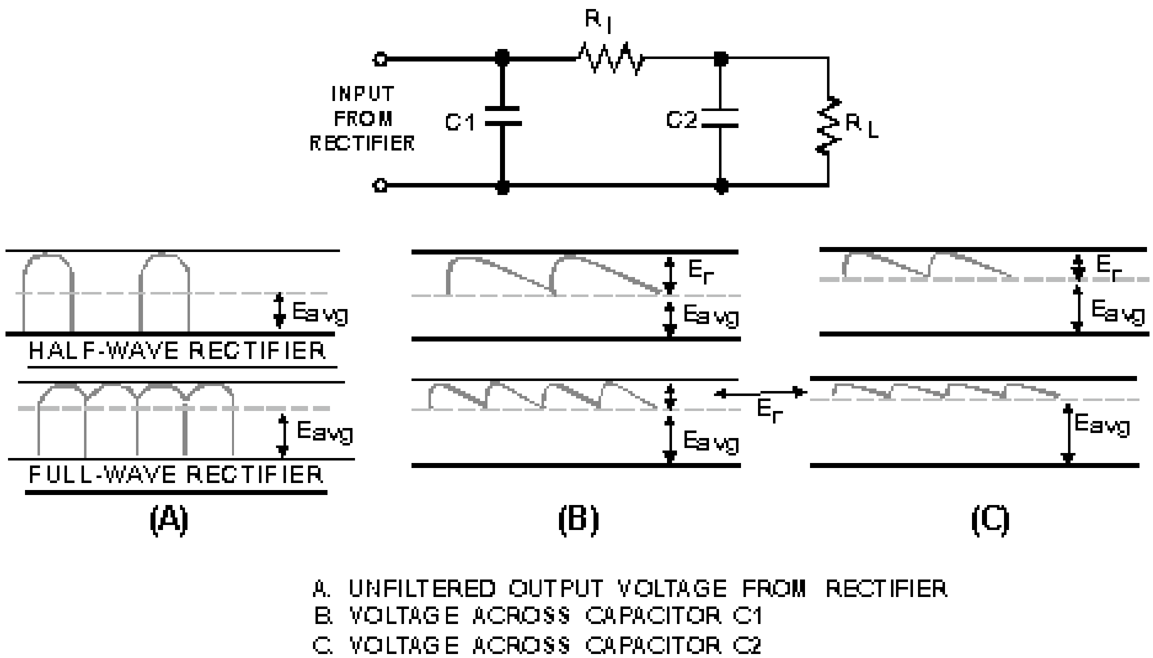


Figure 3-37.—RC filter and waveforms.

The RC filter in figure 3-37 consists of an input filter capacitor (C1), a series resistor (R1), and an output filter capacitor (C2). Although not part of the RC filter, R_L is shown to help explain the circuit. This filter is sometimes referred to as an RC *pi-section filter* because its schematic symbol resembles the Greek letter π.

Although the single capacitor filter is suitable for many noncritical, low-current applications, when the load resistance is very low or when the percent of ripple must be held to an absolute minimum, the capacitor must have an extremely large value. While electrolytic capacitors are available in sizes up to 10,000 µF or greater, the larger sizes are quite expensive. A more practical approach is to use a more sophisticated filter that can do the same job but that has lower capacitor values, such as the RC filter.

The waveforms shown in the figure represent the unfiltered output from a typical rectifier circuit. Note that the dashed line, which indicates the average value of output voltage (E_{avg}) for the half-wave rectifier, is less than half the amplitude of the voltage peaks (approximately 0.318). The average value of output voltage (E_{avg}) for the full-wave rectifier is greater than half (approximately 0.637), but is still much less than, the peak amplitude of the rectifier-output waveform. With no filter circuit connected across the output of the rectifier circuit (unfiltered), the waveform has a large value of pulsating component (ripple) as compared to the average (or dc) component.

An RC filter, such as a pi-section filter, does a much better job than a single capacitor filter.

Figure 3-37 illustrates an RC filter connected across the output of a rectifier. C1 performs the same function that it did in the single capacitor filter. It is used to reduce the percentage of ripple to a relatively low value. Thus, the voltage across C1 might consist of an average dc value of +100 volts with a ripple voltage of 10 volts. This voltage is passed on to the R1-C2 network, which reduces the ripple even further (view C).

C2 offers an infinite impedance (resistance) to the dc component of the output voltage. Thus, the dc voltage is passed to the load, but reduced in value by the amount of the voltage drop across R1. However, R1 is generally small compared to the load resistance. Therefore, the drop in the dc voltage by R1 is not a drawback.

Component values are designed so that the resistance of R1 is much greater than the reactance of C2 at the ripple frequency. C2 offers a very low impedance to the ac ripple frequency. Thus, the ac ripple senses a voltage divider consisting of R1 and C2 between the output of the rectifier and ground. Therefore, most of the ripple voltage is dropped across R1. Only a trace of the ripple voltage can be seen across C2 and the load.

In extreme cases where the ripple must be held to an absolute minimum, a second stage of RC filtering can be added. In practice, the second stage is rarely required. The RC filter is extremely popular because smaller capacitors can be used with good results.

The RC filter has some disadvantages, however. First, the voltage drop across R1 takes voltage away from the load. Second, power is wasted in R1 and is dissipated in the form of unwanted heat.

Finally, if the load resistance changes, the voltage across the load will change. Even so, the advantages of the RC filter overshadow these disadvantages in many cases.

Q27. Is an RC filter used when a large current or a small current demand is required?

Q28. Why is the use of large value capacitors in filter circuits discouraged?

Q29. When is a second RC filter stage used?

The resistor-capacitor (RC) filter is also subject to problems that can cause it to fail. The shunt capacitors (C1 and C2) are subject to an open circuit, a short circuit, or excessive leakage. The series filter resistor (R1) is subject to changes in value and occasionally opens. Any of these troubles can be easily detected.

The input capacitor (C1) has the greatest pulsating voltage applied to it and is the most susceptible to voltage surges. As a result, it is frequently subject to voltage breakdown and shorting. The remaining shunt capacitor (C2) in the filter circuit is not subject to voltage surges because of the protection offered by the series filter resistor (R1). However, a shunt capacitor <u>can</u> become open, leaky, or shorted.

A shorted capacitor or an open filter resistor results in a no-output indication. An open filter resistor results in an abnormally high dc voltage at the input to the filter and no voltage at the output of the filter. Leaky capacitors or filter resistors that have lost their effectiveness, or filter resistors that have decreased in value, result in an excessive ripple amplitude in the output of the supply.

LC Capacitor-Input Filter

The LC input filter is one of the most commonly used filters. This type of filter is used primarily in radio receivers, small audio amplifier power supplies, and in any type of power supply where the output current is low and the load current is relatively constant.

Figure 3-38 shows an LC capacitor-input filter and its associated waveforms. Both half-wave and full-wave rectifier circuits are used to provide the inputs.

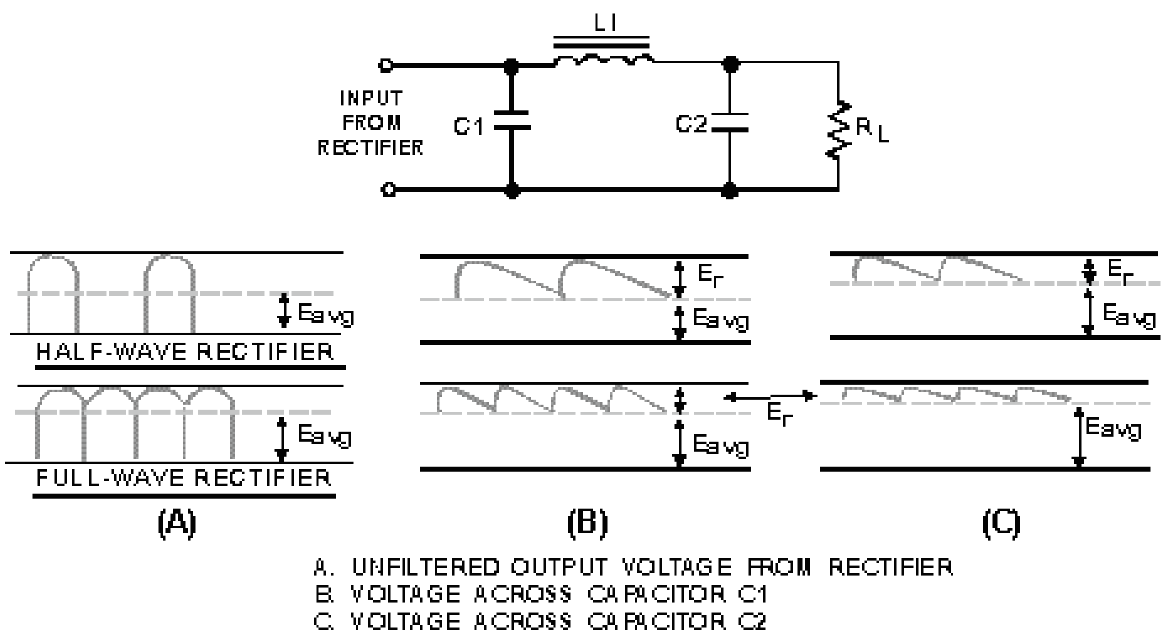


Figure 3-38.—LC capacitor-input filter and waveforms.

The waveforms shown in the figure represent the unfiltered output from a typical rectifier circuit. Note again, that the average value of output voltage (E_{avg}) for the half-wave rectifier is less than half the amplitude of the voltage peaks. This is indicated by the dashed line. The average value of output voltage (E_{avg}) for the full-wave rectifier is greater than half, but is still much less than the peak amplitude of the rectifier-output waveform. With no filter circuit connected across the output of the rectifier circuit (unfiltered), the waveform has a large value of pulsating component (ripple) as compared to the average (or dc) component.

A common type of LC filter is shown in figure 3-38. C1 performs the same functions as discussed earlier by reducing the ripple to a relatively low level. L1 and C2 form the LC filter, which reduces the ripple even further (view C).

L1 is a large value iron-core inductor (choke.) It has a high value of inductance and, therefore, a high value of X_L, which offers a high reactance to the ripple frequency. At the same time, C2 offers a very low reactance to the ac ripple. L1 and C2 form an ac voltage divider and, because the reactance of LI is much higher than that of C2, most of the ripple voltage is dropped across L1. Only a slight trace of the ripple appears across C2 and the load.

While the L1-C2 network greatly reduces the ac ripple, it has little effect on the dc. You should recall that an inductor offers no reactance to dc. The only opposition to current flow is the resistance of the wire in the choke. Generally, this resistance is very low and the dc voltage drop across the coil is minimal. Thus, the LC filter overcomes the disadvantages of the RC filter.

Aside from the voltage divider effect, the inductor improves filtering in another way. You should recall that an inductor resists changes in the magnitude of the current flowing through it. Consequently, when the inductor is placed in series with the load, the inductor tends to hold the current steady. This, in turn, helps to hold the voltage across the load constant.

The LC filter provides good filtering action over a wide range of currents. The capacitor filters best when the load is drawing little current. Thus, the capacitor discharges very slowly and the output voltage remains almost constant. On the other hand, the inductor filters best when the current is highest. The complementary nature of these components ensures good filtering over a wide range of current when size of components is a factor.

The LC filter has two disadvantages. The first is cost. The LC filter is more expensive than the RC filter because its iron-core choke costs more than the resistor of the RC filter. The second disadvantage is size, since the iron-core choke is bulky and heavy. Thus, the LC filter may be unsuitable for some applications but is still one of the most widely used.

Q30. What is the most commonly used filter in use today?

Q31. What are the two main disadvantages of an LC capacitor filter?

Several problems may cause the LC capacitor filter to fail. Shunt capacitors are subject to open circuits, short circuits, and excessive leakage; series inductors are subject to open windings and occasionally shorted turns or a short circuit to the core.

The input capacitor (C1) has the greatest pulsating voltage applied to it, is the most susceptible to voltage surges, and has a generally higher average voltage applied. As a result, the input capacitor is frequently subject to voltage breakdown and shorting. The output capacitor (C2) is not as susceptible to voltage surges because of the series protection offered by the series inductor (L1), but the capacitor can become open, leaky, or shorted.

A shorted capacitor, an open filter choke, or a choke winding that is shorted to the core, results in a no-output indication. A shorted capacitor, depending on the magnitude of the short, may cause a shorted rectifier, transformer, or filter choke and result in a blown fuse in the primary of the transformer. An open filter choke results in an abnormally high dc voltage at the input to the filter and no voltage at the output of the filter. A leaky or open capacitor in the filter circuit results in a low dc output voltage. This condition is generally accompanied by an excessive ripple amplitude. Shorted turns in the winding of a filter choke reduce the effective inductance of the choke and decrease its filtering efficiency. As a result, the ripple amplitude increases.

VOLTAGE REGULATION

Ideally, the output of most power supplies should be a constant voltage. Unfortunately, this is difficult to achieve. There are two factors that can cause the output voltage to change. First, the ac line voltage is not constant. The so-called *115 volts ac* can vary from about 105 volts ac to 125 volts ac. This means that the peak ac voltage to which the rectifier responds can vary from about 148 volts to 177 volts. The ac line voltage alone can be responsible for nearly a 20 percent change in the dc output voltage.

The second factor that can change the dc output voltage is a change in the load resistance. In complex electronic equipment, the load can change as circuits are switched in and out. In a television receiver, the load on a particular power supply may depend on the brightness of the screen, the control settings, or even the channel selected.

These variations in load resistance tend to change the applied dc voltage because the power supply has a fixed internal resistance. If the load resistance decreases, the internal resistance of the power supply drops more voltage. This causes a decrease in the voltage across the load.

Many circuits are designed to operate with a particular supply voltage. When the supply voltage changes, the operation of the circuit may be adversely affected. Consequently, some types of equipment must have power supplies that produce the same output voltage regardless of changes in the load resistance or changes in the ac line voltage. This constant supply of power may be achieved by adding a circuit called the **VOLTAGE REGULATOR** at the output of the filter.

LOAD REGULATION

A commonly used **FIGURE OF MERIT** for a power supply is its **PERCENT OF REGULATION**. The figure of merit gives us an indication of how much the output voltage changes over a range of load resistance values. The percent of regulation aids us in determining of the type of load regulation needed. Percent of regulation is determined by the equation:

$$\text{Percent of regulation} = \frac{(E_{nL} - E_{fL})}{E_{fL}} \times 100$$

This equation compares the change in output voltage at the two loading extremes to the voltage produced at full loading. For example, assume that a power supply produces 12 volts when the load current is zero. If the output voltage drops to 10 volts when full load current flows, then the percent of regulation is:

$$\text{Percent of regulation} = \frac{(E_{nL} - E_{fL})}{E_{fL}} \times 100$$

$$\text{Percent of regulation} = \frac{(12 - 10V)}{10V} \times 100$$

$$\text{Percent of regulation} = \frac{2V}{10V} \times 100$$

$$\text{Percent of regulation} = 20\%$$

Ideally, the output voltage should not change over the full range of operation. That is, a 12-volt power supply should produce 12 volts at no load, at full load, and at all points in between. In this case, the percent of regulation would be:

$$\text{Percent of regulation} = \frac{(E_{nL} - E_{fL})}{E_{fL}} \times 100$$

$$\text{Percent of regulation} = \frac{(12 - 12V)}{12V} \times 100$$

$$\text{Percent of regulation} = \frac{0V}{12V} \times 100$$

$$\text{Percent of regulation} = 0\%$$

Thus, zero-percent load regulation is the ideal situation. It means that the output voltage is constant under all load conditions. While you should strive for zero percent load regulation, in practical circuits you must settle for something less. Even so, by using a voltage regulator, you can hold the percent of regulation to a very low value.

If you are interested in reading more on this subject, refer to the *Electronic Installation and Maintenance Book (EIMB)* series or other similar books from your technical library.

Q32. What two factors can cause output dc voltage to change?

Q33. What is the commonly used figure of merit for a power supply?

Q34. If a power supply produces 20 volts with no load and 15 volts under full load, what is the percent of regulation?

Q35. What percent of regulation would be ideal?

REGULATORS

You know that the output of a power supply varies with changes in input voltage and circuit load current requirements. Because many military electronic equipments require operating voltages and currents that must remain constant, some form of regulation is necessary. The circuits that maintain power supply voltage or current outputs within specified limits, or tolerances, are called regulators. They are designated as dc voltage or dc current regulators, depending on their specific application.

Voltage regulator circuits are additions to basic power supply circuits and are made up of rectifier and filter sections. The purpose of the voltage regulator is to provide an output voltage with little or no variation. Regulator circuits sense changes in output voltages and compensate for the changes. Regulators that maintain voltages within plus or minus (±) 0.1 percent are quite common. The diagram in figure 3-39 clearly illustrates the purpose of the voltage regulator.

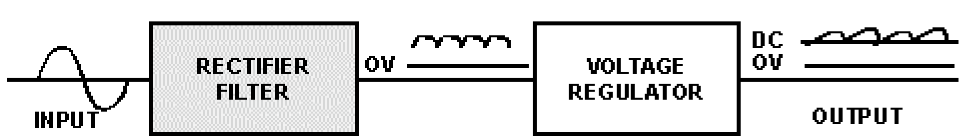


Figure 3-39.—Block diagram of a power supply and regulators

There are two basic types of voltage regulators, series and shunt. Whether a voltage regulator is classified as series or shunt depends on the location or position of the regulating element(s) in relation to the circuit load resistance.

Figure 3-40 illustrates the two basic types of voltage regulators. In actual practice the circuitry of regulating devices may be quite complex. We use the simplified drawings in the figure to emphasize that there are two basic types of voltage regulators. Broken lines highlight the differences between the series and shunt regulators.

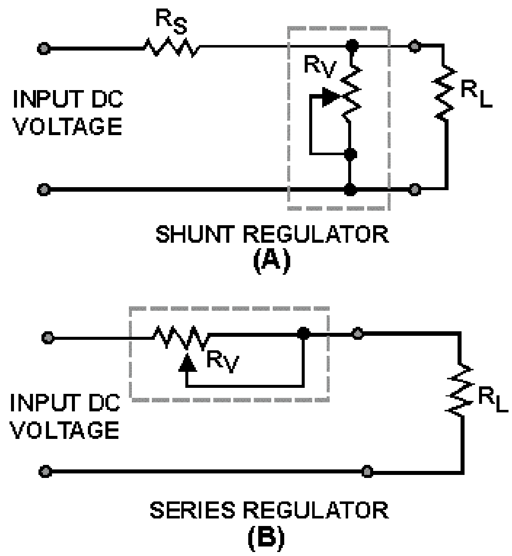


Figure 3-40.—Series and shunt regulators.

The schematic in view (A) is that of a shunt-type regulator. It is called a shunt-type regulator because the regulating device is connected in parallel with the load resistance. This is a characteristic of all shunt-type regulators. The schematic in view (B) is that of a series regulator. It is called a series regulator because the regulating device is connected in series with the load resistance.

Series Voltage Regulator

Figure 3-41 illustrates the principle of series voltage regulation. As you study the figure, notice that the regulator is in series with the load resistance and that all current passes through the regulator. In this example, variable resistor R_v is used for regulation. Examine the circuit to determine how the regulator functions. When the input voltage increases, the output voltage also increases. However, since the voltage regulator device (R_v) senses this change, the resistance of the regulating device increases and results in a greater voltage drop through R_v. This causes the output voltage to decrease to normal or, for all practical purposes, to remain constant.

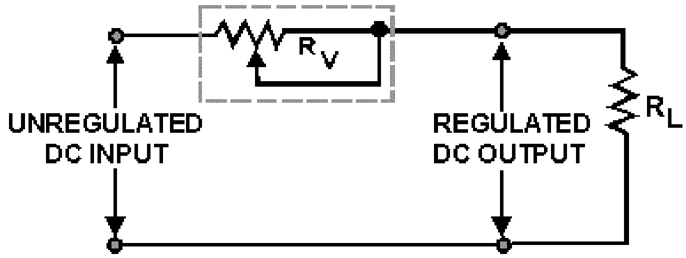


Figure 3-41.—Series voltage regulator.

You should be able to see that as the input voltage decreases, the resistance of the variable resistor R_v decreases almost simultaneously, thereby compensating for the voltage drop. Since there is a smaller voltage drop across R_v, the output voltage remains almost constant. Voltage fluctuations within the circuit occur in microseconds.

Shunt Voltage Regulator

The diagram in figure 3-42 represents a shunter voltage regulator. Notice that variable resistor R_v is in parallel with the load resistance R_L and that fixed resistor R_S is in series with the load resistance. You already know the voltage drop across a fixed resistor remains constant unless there is a variation (increase or decrease) in the current through it.

In a shunt regulator as shown in figure 3-42, output voltage regulation is determined by the current through the parallel resistances of the regulating device (R_v), the load resistance (R_L), and the series resistor (R_S). For now, assume that the circuit in figure 3-42 is operating under normal conditions, that the input is 120 volts dc, and that the desired regulated output is 100 volts dc. For a 100-volt output to be maintained, 20 volts must be dropped across the series resistor (R_S). If you assume that the value of R_S is 2 ohms, then you must have 10 amperes of current through R_v and R_L. (Remember: $E = IR$.) If the values of the resistance of R_v and R_L are equal, then 5 amperes of current will flow through each resistance (R_v and R_L).

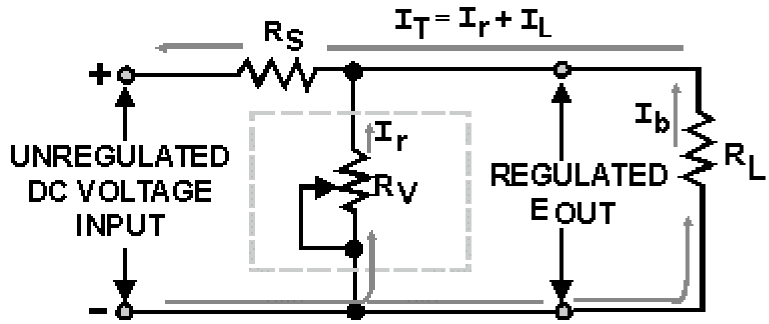


Figure 3-42.—Shunt voltage regulator.

Now, if the load resistance (R_L) increases, the current through R_L will decrease. For example, assume that the current through R_L is now 4 amperes and that the total current through R_S is 9 amperes. With this drop in current, the voltage drop across R_S is 18 volts; consequently, the output of the regulator has increased to 102 volts. At this time, the regulating device (R_v) decreases in resistance, and 6 amperes of current flows through this resistance (R_v). Thus, the total current through R_S is once again 10 amperes (6 amperes across R_v, 4 amperes through R_L); therefore, 20 volts will be dropped across R_S causing the output to decrease back to 100 volts.

You should know by now that if the load resistance (R_L) <u>increases</u>, the regulating device (R_v) <u>decreases</u> its resistance to compensate for the change. If R_L decreases, the opposite effect will occur and R_v will increase. Now take a look at the circuit when a decrease in load resistance takes place.

When R_L decreases, the current through R_L subsequently increases to 6 amperes. This action causes a total of 11 amperes to flow through R_S which now drops 22 volts. As a result, the output is now 98 volts. However, the regulating device (R_v) senses this change and increases its resistance so that less current (4 amperes) flows through R_v. The total current again becomes 10 amperes, and the output is again 100 volts.

From these examples, you should now understand that the shunt regulator maintains the desired output voltage by sensing the current change that occurs in the parallel resistance of the circuit.

Again refer to the schematic shown in figure 3-42 and consider how the voltage regulator operates to compensate for changes in input voltages. You know, of course, that the input voltage may vary and that any variation must be compensated for by the regulating device. Consider an <u>increase in input voltage</u>. When this happens the resistance of R_v automatically decreases to maintain the correct voltage division between R_v and R_S. You should see, therefore, that the regulator operates in the opposite way to compensate for a <u>decrease in input voltage.</u>

So far we have explained the operation of voltage regulators that use variable resistors; however, this type of regulation has limitations. Obviously, the variable resistor cannot be adjusted rapidly enough to compensate for frequent fluctuations in voltage. Since input voltages fluctuate frequently and rapidly, the variable resistor is <u>not</u> a practical method for voltage regulation. A voltage regulator that operates continuously and automatically to regulate the output voltage without external manipulation is required.

Q36. The purpose of a voltage regulator is to provide an output voltage with little or no ____.

Q37. The two basic types of voltage regulators are _______ and _______.

Q38. When a series voltage regulator is used to control output voltages, any increase in the input voltage results in <u>an increase/a decrease</u> in the resistance of the regulating device.

Q39. A shunt type voltage regulator is connected in <u>series/parallel</u> with the load resistance.

Basic VR Tube Regulator Circuit

Although we covered the electrical characteristics of the VR tube in chapter 2 of this module, we now need to cover the capabilities and limitations of the VR tube itself.

Figure 3-43 shows a basic VR tube regulating circuit. The voltage produced by the source is 150 volts. The VR 90 will provide a constant 90 volts across the load resistance (R_L) if the tube is operated in the normal glow discharge region. This means that 60 volts is dropped across R_S, which is the series limiting resistance used to limit the current through the VR tube.

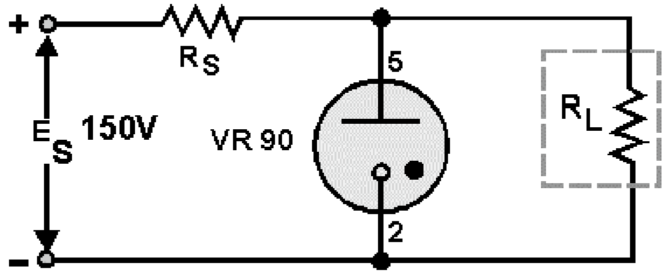


Figure 3-43.—Basic VR tube regulator.

Since the operating limits of a VR tube are determined by its maximum and minimum currents, circuits using such tubes should be designed to allow maximum variations in current above and below the normal point of operation. The normal point of operation, which allows maximum variation in current, must be midway between the current limits of the tube. This median current is called I_{mean}. and can be calculated by the use of the following equation:

$$I_{mean} = \frac{I_{max} + I_{min}}{2}$$

We can determine the mean current for the VR90-40 as shown in figure 3-44 by using the following values:

$$I_{mean} = \frac{40 + 5}{2} = \frac{45}{2} = 22.5\text{mA}$$

To calculate the value of series dropping resistance R_S, we use the following equation:

$$R_S = \frac{\text{Source Voltage - Regulated Voltage}}{I_{mean} + I_{load}(\text{average})}$$

If the average current flowing through the load of figure 3-44 is 100 milliamperes, we can find the series dropping resistance in the following manner:

$$R_S = \frac{150 - 90}{22.5 + 100} = \frac{60 \text{ volts}}{22.5\text{mA}} = 490 \text{ ohms}$$

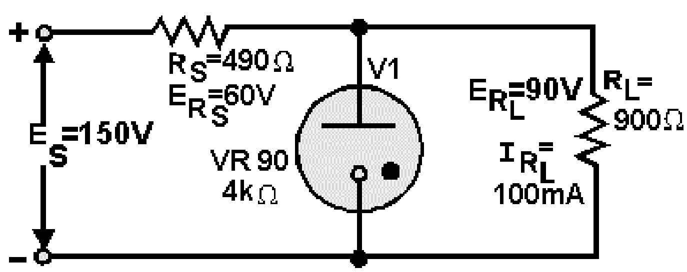


Figure 3-44.—Simplified VR tube regulator.

According to Ohm's law, the value of the load resistance for this circuit figure will be 900 ohms if a current of 100 milliamperes flows through R_L. The internal resistance of the VR tube can be calculated in a similar manner. With 22.5 milliamperes flowing and 90 volts dropped across the VR tube, its resistance is 4 kilohms.

To determine the voltage regulation in the circuit for figure 3-44, assume a constant supply voltage of 150 volts and a variable load resistance. If the value of R_L were to decrease to 857 ohms, the load current would increase to approximately 105 milliamperes to maintain 90 volts across the load resistance. R_S must drop 60 volts. To do so requires a current of 122.5 milliamperes flowing through the series resistance. Since 105 milliamperes is now flowing through the load, the current through the VR tube must decrease from 22.5 milliamperes to 17.5 milliamperes. We will discuss the sequence of events in more detail to help you better understand how the tube current is made to vary.

The original load resistance was 900 ohms. Changes in this resistance will not occur instantaneously, but will require some time to vary from 900 ohms to a new value. As resistance of the load begins to decrease, load current begins to increase. The minute increase in load current will flow through the series resistance R_S causing a slight increase in E_{RS}. This slight increase in voltage across R_S will result in the VR tube voltage dropping slightly. This slight drop in tube voltage will cause a decrease in the ionization of the tube gas, which in turn increases the resistance of the tube. As a result, less current flows through the tube.

Note that tube current can decrease only to a value of 5 milliamperes before deionization occurs. Therefore, the load current cannot exceed 117.5 milliamperes, for beyond this value, tube current becomes less than 5 milliamperes and regulation ceases.

If load resistance were to increase, load current would decrease. This would result in the VR tube current increasing to maintain a current of 122.5 milliamperes. The VR tube current can only increase to 40 milliamperes. Beyond this value of current, the tube enters the abnormal glow region and tube voltage increases.

The upper limit of the VR tube current will occur when load current decreases to a value of 82.5 milliamperes. When load current drops below this value, the VR tube ceases to regulate the load voltage. Therefore, with a constant source voltage but variable load resistance, the limits of regulation will be reached when current in the load exceeds 117.5 milliamperes or drops below 82.5 milliamperes.

The VR tube regulator can also compensate for changes in power supply voltage. Under these conditions, the load resistance will remain constant while the power supply voltage will be variable. Refer to figure 3-44 for the following discussion.

Assume that the source voltage begins to increase from an original value of 150 volts toward 155 volts. As this voltage increases, current through R_S increases from its original value of 122.5 milliamperes. Initially, this additional current is drawn from the load, causing a slight increase in load voltage. This increase in load voltage is felt across the VR tube and causes an increase in tube ionization. This decreases the internal resistance of the VR tube with a resultant increase in tube current. When source voltage reaches 155 volts, current through R_S is approximately 133 milliamperes (R_S = 490 ohms). Most of the additional current through R_S flows through the VR tube. As a result, approximately 33 milliamperes flows through the VR tube, maintaining the load voltage at 90 volts.

Since VR tube current decreases as source voltage decreases, tube current will drop below its lower limit of 5 milliamperes at some point. When source voltage drops below 141.4 volts, tube current will be less than 5 milliamperes and regulation will cease. The upper and lower limits of the supply voltage variations that can be allowed and still provide regulation in the circuit are 158.6 volts and 141.4 volts, respectively. Remember that tube voltage varies slightly through its operating range, but this voltage change is less than that which would exist without the use of a VR tube.

As the source voltage increases, the current through the VR tube increases. Since the upper limit of tube current is 40 milliamperes, there is a limit in the ability of the tube to regulate increasing voltage. When the supply voltage exceeds 158.6 volts, tube current will be greater than 40 milliamperes and regulation will cease.

If the source voltage decreases from 150 volts to 145 volts, only 55 volts must be dropped across the 490-ohm series resistance (R_S) to maintain the load voltage at 90 volts. Current through R_S for a 55 volt drop is 112 milliamperes. Since load current is 100 milliamperes, the remaining 12 milliamperes must flow through the VR tube. This represents a decrease in the ionization level of the VR tube, with a resultant increase in tube resistance. Under these conditions, 90 volts will be maintained across the load resistance.

VR Tubes Connected in Series

In applications where a regulated voltage in excess of the maximum rating of one tube is required, two or more tubes may be placed in series as shown in figure 3-45.

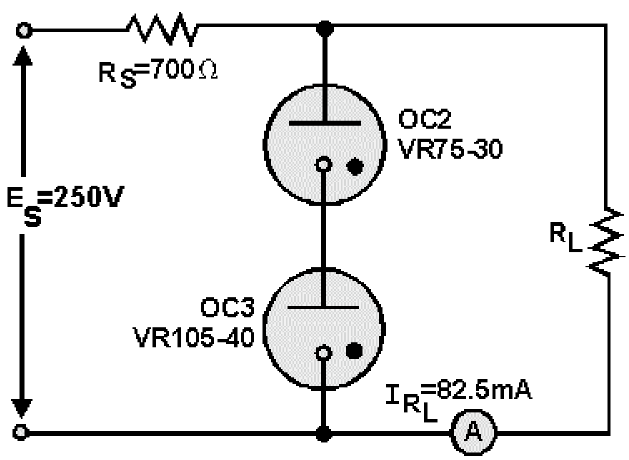


Figure 3-45.—VR tubes as voltage dividers.

In the figure, a VR75-30 and a VR105-40 are shown connected in series. The source voltage is 250 volts, and 82.5 milliamperes flows through the load resistance. Since current through the two VR tubes is common, the limits of regulation are determined by the tube having the smaller current limitations. (In this case, the VR 75-30). In computing I_{mean} for this circuit, I_{max} and I_{min} will be 30 milliamperes and 5 milliamperes, respectively. Therefore, the mean current will be 17.5 milliamperes.

The value of R_S in the figure can be computed using the source voltage of 250 volts and the total current through R_S (load current + I_{mean}). Using these values, R_S = 700 ohms. Note that the regulated voltage to the load is 180 volts. This provides a regulated voltage greater than would be possible using either VR tube by itself.

Another advantage of using VR tubes in series is illustrated in figure 3-46. In this circuit, several values of regulated voltages are obtained from a single power supply.

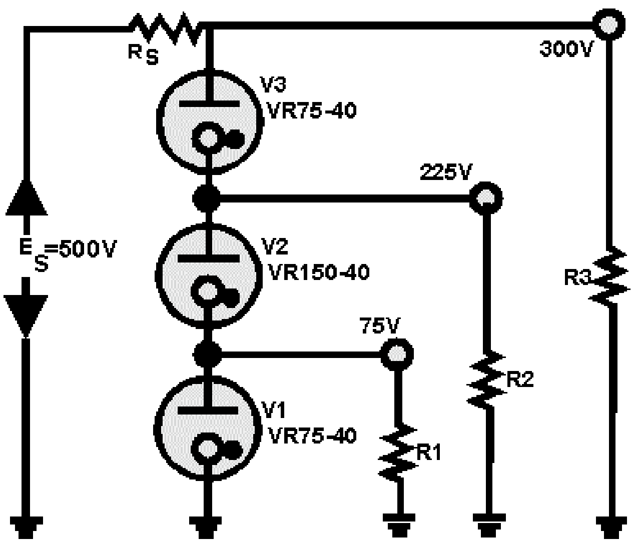


Figure 3-46.—VR tubes as voltage dividers

The current flowing through V2 in the figure is a combination of the current through R1 and the current through V1. The current through V3, on the other hand, is the sum of the currents through V2 and R2. Since V3 has more current flowing through it than any of the other VR tubes, it places or determines the limit on the maximum current in the VR tube circuit. Since the maximum rating of V3 is 40 milliamperes, the currents through R1 and R2 must be limited to only a few milliamperes, or the rating of V3 will be exceeded and regulation will cease.

The obvious advantage in using VR tubes in series is to provide several regulated voltages from a single power supply. The primary disadvantage is in the current limitations. Since it is impossible to have all VR tubes operating about their mean current values, this limits the ability of the circuit to regulate over wide ranges of variations in load resistance or source voltage.

VR Tubes Connected in Parallel

One might expect that connecting VR tubes in parallel as shown in figure 3-47 would increase the current handling capacity of the network. Although this is true for some gas-filled tubes, it is not true for VR tubes. In figure 3-47, two VR tubes are constructed in exactly the same way. The only difference will be a slight variation in their ionization potential. For the purpose of this discussion, VR tube VR1 will have a lower ionization potential than VR2. The potential that must be reached before a VR tube ionizes is considerably higher than its normal operating voltage.

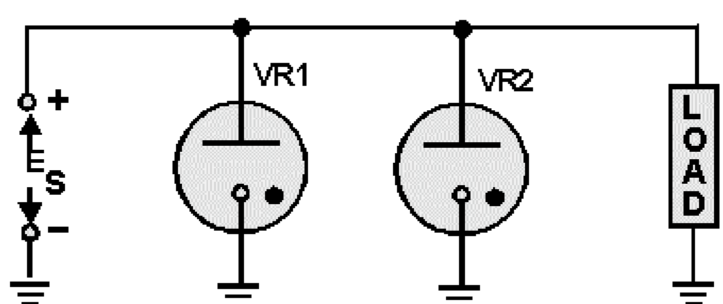


Figure 3-47.—VR tubes connected in parallel.

When voltage is applied to the circuit of figure 3-47, as soon as the correct potential is reached, VR1 begins to conduct and the potential across it decreases to its operating voltage. The potential across VR2 never becomes high enough to cause it to ionize. Therefore, placing the VR tubes in parallel accomplishes no useful purpose. When greater current handling capacity and better regulation are desired, electronic (vacuum tube) regulator circuits are used.

Several conditions may either indicate or cause problems with a VR tube regulator. Initially, you can get some indication of the trouble associated with a gas-tube regulator circuit by visually inspecting it to determine the presence of the characteristic glow from the ionized gas within the tube. When current through the tube is near its maximum rating, the tube is highly ionized. When the current is near its minimum rating, the tube is lightly ionized, Therefore, the intensity of the gaseous discharge within the tube is an indication of tube conduction. If the tube is not ionized, however, this does not necessarily mean that the tube is defective. The same indication (lack of characteristic glow) may also result from the following conditions: the series resistor (R_S) has increased in value, the dc input voltage (E_S) is below normal, the load current is below normal, or the load current is excessive. You therefore need to make dc voltage measurements at the input and output terminals of the voltage regulator circuit to determine whether the problem is inside the regulator circuit or outside of it.

You can check value of the series resistor (R_S) by using ohmmeter measurements to determine whether any change in resistance has occurred. If the maximum current rating of the regulator tube is exceeded for a considerable length of time, the tube may be damaged and lose its regulation characteristics; therefore, you can suspect the regulator tube itself as a possible source of trouble.

Although VR tubes are used extensively in electronic equipment, there are circuits that require a greater degree of regulation than a VR tube can provide. For these circuits, an electron tube voltage regulator is used.

Electron Tube Voltage Regulator

An electron tube may be considered a variable resistance. When the tube is passing a direct current, this resistance is simply the plate-to-cathode voltage divided by the current through the tube and is called the dc plate resistance (R_p). For a given plate voltage, the value of R_p depends upon the tube current, and the tube current depends upon the grid bias.

Refer to figure 3-48, view (A). The resistance of V1 is established initially by the bias on the tube. Assume that the voltage across the load is at the desired value. Then the cathode is positive with respect to ground by some voltage (E_L). The grid can be made positive relative to ground by a voltage (E2) that is less than E1. The potentiometer R2 is adjusted until the bias (grid-to-cathode voltage), which is E2 - E1, is sufficient to allow V1 to pass a current equal to the desired load current. With this bias, the resistance of V1 is established at the proper value to reduce the rectifier output voltage to the desired load voltage.

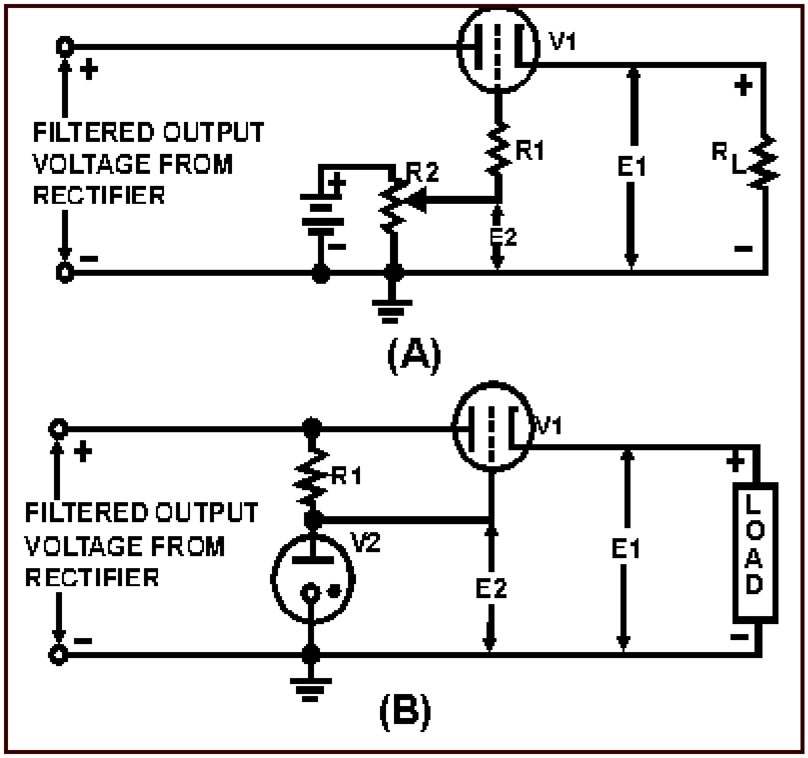


Figure 3-48.—Electron tube voltage regulator using a battery for the fixed bias

If the rectifier output voltage increases, the voltage at the cathode of V1 tends to increase. As E1 increases, the negative bias on the tube increases and the plate resistance of the tube becomes greater. Consequently, the voltage drop across V1 increases with the rise in input voltage. If the circuit is designed property, the increased voltage drop across V1 is approximately equal to the increase in voltage at the input. Thus the load voltage remains essentially constant.

The resistor (R1) is used to limit the grid current. This is necessary in this particular circuit because the battery is not disconnected when the power is turned off. However, the battery can be eliminated from the circuit by the use of a glow tube (V2), as shown in view (B) of the figure, to supply a fixed bias for the grid of the tube. The action of the circuit in view (B) is the same as the action of the circuit in view (A). The output voltage of the simple voltage regulators shown in the figure cannot remain absolutely constant. As the rectifier output voltage increases, the voltages on the cathode of V1 must rise slightly if the regulator is to function.

The voltage regulators shown in the figure compensate not only for changes in the output voltage from the rectifier, but also for changes in the load. For example, in view (B) if the load resistance decreases, the load current will increase. The load voltage will tend to fall because of the increased drop across V1. The decrease in load voltage is accompanied by a decrease in bias voltage on V1. The bias voltage on V1 is equal to E1 - E2. Thus the effective resistance of V1 is reduced at the same time the load current is increased. The IR drop across V1 increases only a slight amount because R decreases about as much as I increases. Therefore, the tendency for the load voltage to drop when the load is increased is checked by the decrease in resistance of the series triode.

Q40. In an electron tube regulator, the electron tube replaces what component?

CURRENT REGULATION

Before we go to the next section, there is one type of regulation that we should discuss-current regulation. In most power supplies, current is not regulated directly. Fuses and other circuit protection devices are used to set an upper limit on the amount of current that can flow in a power supply. Once this limit is exceeded, the fuse simply opens and the power supply is deenergized. Beyond this, current is usually left unregulated because the load will draw from the power supply only the amount of current that it needs. Decreases and increases in the power supply voltage caused by the variations in load current are usually controlled by the voltage regulator.

The Amperite Regulator

There are some cases in which current must be regulated or kept at a relatively constant value. The best example of this is the filament supply of a power transformer located in a power supply that is designed to supply filament power to many tubes. You can see this in view (A) of figure 3-49, which is a portion of a power supply designed to supply 50 vacuum tubes with both plate and filament voltages. Under normal conditions, circuit current will not exceed 2.5 amperes. For this reason, the power supply has been fused at 3 amperes. Because you are only interested in current regulation at this time, only the portion of the power supply that deals with current regulation is shown; namely, the power transformer and four of the 50 parallel connected vacuum-tube filaments. At operating temperatures, the resistance of each filament is 1 kilohm. Because the filaments are connected in parallel, the total filament resistance at operating temperature is 20 ohms. Ohm's law,

$$I = \frac{E}{R}$$

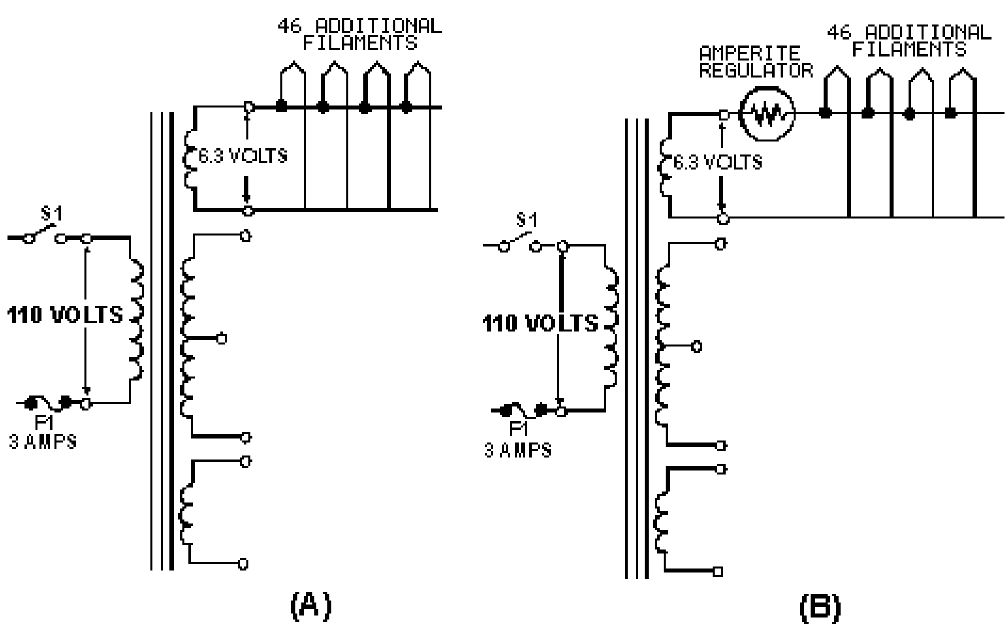


Figure 3-49.—Amperite regulator.

tells you that the filaments draw .315 amperes. You should know from your previous study that as conductors are heated, their resistance increases. Therefore, the cold resistance of the filaments is considerably lower than the hot resistance. In this case, assume 100 ohms per filament. The total resistance of the 50 parallel filaments is then 2 ohms when the power supply is first energized, and the filaments draw 3.15 amperes of current. If the current for the rest of the power supply is added to the filament current, the surge current will cause the power supply to draw 5 amperes when it is first energized. Unfortunately, the power supply is fused at 3 amperes. Under these conditions, it would be impossible to keep the power supply on the line long enough to get the filaments up to operating temperature.

There are three possible solutions to this problem. The first is simply to fuse the power supply at 5 amperes, but this could allow excessive current to flow in the power supply. Another solution is to use a slow-blow fuse. Unfortunately, the duration of the current surge may exceed the time limit that a slow-blow fuse can handle. Therefore, current regulation is the best solution to this problem.

Because of its quick-heating ability, the amperite tube is ideal as a current regulator. The amperite regulator is nothing more than an iron wire enclosed in a hydrogen-filled envelope. Because of its construction, the iron filament will heat quickly when current is applied to it.

View (B) of figure 3-49 shows the amperite regulator connected in series with the filaments of the load. When the power supply is first energized, the iron wire of the amperite gets hot quickly and presents a large resistance connected in series with the 2 ohms of filament resistance. As a result, most of the voltage is dropped across the amperite. Because of the large resistance of the amperite regulator, current in the circuit is held to an acceptable level in accordance with Ohm's law:

$$I = \frac{E}{R}$$

As the filaments warm up, their resistance increases, which causes circuit current to decrease. The decreasing circuit current allows the iron wire of the amperite to cool. As it cools, its resistance decreases until it reaches the approximate resistance of the circuit wiring. You might think that decreasing the resistance of the amperite would allow circuit current to increase again, but this does not happen. As the iron wire of the amperite cools and its resistance decreases, the resistance of the warming tube filaments increases. Throughout the entire heating cycle of the filaments, the total resistance of the series circuit, consisting of the amperite and tube filaments, remains fairly constant. When power is first applied, most of the resistance is in the amperite. Therefore, most of the voltage is dropped across the resistance of the amperite. Halfway through the cycle, the resistance of the amperite and the resistance of the filaments are approximately equal, and the voltage drops across the two series elements are equal. Finally, when the filaments have reached their operating temperature, most of the resistance is in the filaments of the tube. Therefore, most of the voltage is dropped across the tube filaments.

The important thing to note is that the total circuit resistance remains approximately the same throughout the heating cycle. As the cycle progresses, the resistance of the amperite decreases as the resistance of the tube filaments increases. Because resistance and voltage (6.3 volts) remain constant, current remains constant, except for the slight surge in the beginning of the heating cycle, which is necessary to heat up the iron wire of the amperite.

Now that we have discussed the different types of regulators, you should be able to see that there are many variables that affect good regulation.

Although you may not be required to design regulators, you will be required to maintain them because your electronic equipment depends upon good regulation to operate property.

Up to this point we have discussed only the individual sections of the electron tube power supply. In the next section, we will discuss the techniques of troubleshooting these individual sections and the total power supply.

Q41. What is the purpose of the amperite regulator?

Q42. As the tube filaments in the load heat up, will the circuit current increase or decrease?

TROUBLESHOOTING POWER SUPPLIES

Whenever you work with electricity, you **must** follow all the appropriate safety precautions. In the front of all electronic technical manuals, you will always find a section on safety precautions. You should also find posted on each piece of equipment a sign listing the specific precautions for that equipment. One hazardous area that is sometimes overlooked, especially on board ship, is grounding of equipment. By grounding the return side of the power transformer to the metal chassis, manufacturers can wire the cathodes of the tubes in both the power supply and the load being supplied by the power supply directly to the metal chassis. This eliminates the necessity of wiring each tube directly to the return side of the transformer, saving wire, and reducing the cost of building the equipment. While this solves one of the problems of the manufacturer, it creates a problem for you, the technician. Unless the chassis is physically grounded to the ship's ground (the hull), the chassis can be charged (or can float) several hundred volts above ship's ground. If you come in contact with the metal chassis at the same time you are in contact with the ship's hull, the current from the chassis can use your body as a low resistance path back to the ship's ac generators. At best this can be an unpleasant experience; at worst it can be fatal. For this reason Navy electronic equipment is always grounded to the ship's hull, and approved rubber mats are required in all spaces where electronic equipment is present. Therefore, before you start to work on any electronic or electrical equipment ALWAYS ENSURE THAT THE EQUIPMENT AND ANY TEST EQUIPMENT YOU ARE USING IS PROPERLY GROUNDED AND THAT THE RUBBER MAT YOU ARE STANDING ON IS IN GOOD CONDITION. As long as you follow these simple rules, you should be able to avoid the possibility of becoming an electrical conductor.

TESTING

There are two widely used checks in testing electronic equipment. The first is the **VISUAL CHECK**. Do not underestimate the importance of this check. Many technicians find defects right away simply by looking for them. A visual check does not take long; in fact you should be able to see the problem in about 2 minutes if it is the kind of problem that can be seen. You should learn the following procedure. You will find yourself using it quite often, as it is good not only for power supplies but also for any other type of electronic equipment you may be troubleshooting.

1. **BEFORE YOU PLUG IN THE EQUIPMENT, LOOK FOR:**

   a. LOOSE TUBES—A tube that is not properly seated in its socket may not be making proper contact with the rest of the circuit. It may very well be the source of your problem. Push the tube completely into place.

   b. SHORTS—Examine any terminal or connection that is close to the chassis or to any other terminal for the possibility of a short. A short in any part of the power supply can cause considerable damage. Look for and remove any stray drops of solder, bits of wire, nuts, or screws. It sometimes helps to shake the chassis and listen for any tell-tale rattles. Remember to correct any problem that may cause a short circuit. If it is not causing trouble now, it may cause problems in the future.

c. DISCOLORED OR LEAKING TRANSFORMER—This is a sure sign that there is a short somewhere. Locate it. If the equipment has a fuse, find out why the fuse did not blow; too large a size may have been installed, or there may be a short across the fuse holder.

d. LOOSE, BROKEN, OR CORRODED CONNECTIONS—Any connection that is not in good condition is a trouble spot If it is not causing problems now, it probably will in the future. Fix it.

e. DAMAGED RESISTORS OR CAPACITORS—A resistor that is discolored or charred has been subjected to an overload. An electrolytic capacitor will show a whitish deposit at the seal around the terminals. Check for a short whenever you notice a damaged resistor or capacitor. If there is no short, the trouble may be that the power supply has been overloaded in some way. Make a note to replace the part after signal tracing. There is no sense in risking a new part until you have located the trouble.

2. **PLUG IN THE POWER SUPPLY AND LOOK FOR:**

   a. SMOKING PARTS—If any part smokes or if you hear any boiling or sputtering sounds, pull the plug immediately. There is a short circuit somewhere that you have missed in your first inspection. Use an ohmmeter to check the part again; begin in neighborhood of the smoking part.

   b. COLD TUBES—After allowing the equipment about two minutes for warm-up, touch all the tubes. If a tube is cold, it is either burned out or there is a break in the heater connections and the tube is not receiving proper heater voltage. Remove the tube and connect an ohmmeter across the heater terminals to see if the filament is open (reads almost infinite resistance). If the filament reads open, it is burned out. Replace the bad tube with a good one. If the filament reads a low resistance, this indicates that the filament is all right. Use an ac voltmeter to find the break between the filament and the output of the transformer.

   c. SPARKING—Tap or shake the chassis. If you see or hear sparking, you have located a loose connection or a short. Check and repair the problem.

If you locate and repair any of the defects listed under the visual check, make a note of what you find and what you do to correct it. It is quite probable you have found the trouble. However, a good technician takes nothing for granted. You must prove to yourself that the equipment is operating properly and that no other troubles exist.

If you find none of the defects listed under the visual check, go ahead with the signal tracing procedure. The trouble is probably of such a nature that you cannot see it directly with your eye-you must see it through the eye of the oscilloscope.

The second type of testing is signal tracing. Tracing the ac signal through the equipment is the most rapid method of locating a trouble that you cannot find by a visual check. It also serves as a check on any repairs you may have made. The idea is to trace the ac voltage from the transformer, to see it change to pulsating dc at the rectifier tube filament, and then to see the pulsations smoothed out by the filter. The point where the signal stops or becomes distorted is the place to look for the trouble.

Before you begin signal tracing, it is a good idea to measure the dc voltage. The dc output voltage should be in the neighborhood of 340 volts. If you have no dc output voltage, you should look for an open or a short in your signal tracing. If you have a low dc voltage, you should look for a defective part and keep your eyes open for the place where the signal becomes distorted.

Signal tracing is done by observing the waveform at the input and output of each part of a circuit. It is the method used to localize trouble in a circuit.

Let's review what each part of a good power supply does to the signal, as shown in figure 3-50. The ac voltage is brought in from the power line through the line cord. This voltage is connected to the primary of the transformer through the **ON-OFF** switch (S1). At the secondary winding of the transformer (points 1 and 2), the scope shows you a picture of the stepped-up voltage developed across each half of the secondary winding—the picture is that of a complete sine wave. Each of the two stepped-up voltages is connected between ground and one of the two plates of the rectifier tube. At the two rectifier plates (points 4 and 5) there is still no change in the shape of the stepped-up voltage the scope picture still shows a complete sine wave.

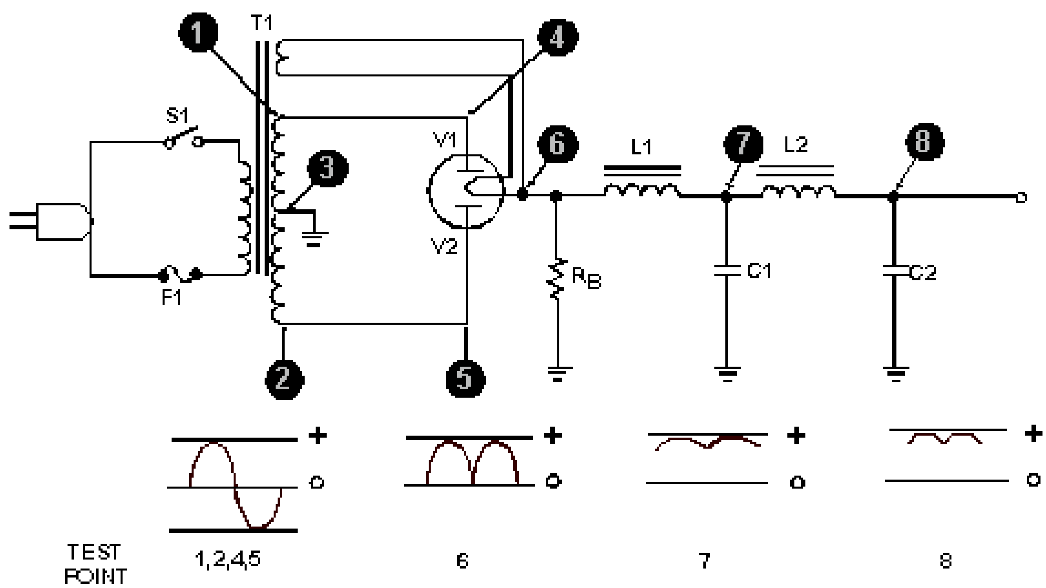


Figure 3-50.—Complete power supply (without regulator).

However, when you look at the scope pattern for point 6 (the voltage at the rectifier heater), you see the wave shape for pulsating direct current. This pulsating dc is fed through the first choke (L1) and filter capacitor (C1), which remove a large part of the ripple or "hum," as shown by the waveform for point 7. Finally, the dc voltage is fed through the second choke (L2) and filter capacitor (C2), which remove nearly all of the remaining ripple. See the waveform for point 8, which shows almost no visible ripple. You now have almost pure dc.

No matter what power supplies you may encounter in the future, they all do the same thing—they change ac voltage into dc voltage.

COMPONENT PROBLEMS

The following paragraphs will give you an indication of troubles that occur with many different electronic circuit components.

Tube Troubles

The symptoms of tube trouble will vary with every type of circuit and each type of tube. However, the problems that can develop with a tube are common to every tube. Here are the five possible tube troubles that you should keep in mind. The meaning of each trouble will be clear by the time you end your study of vacuum tubes, even though you may not quite understand them now.

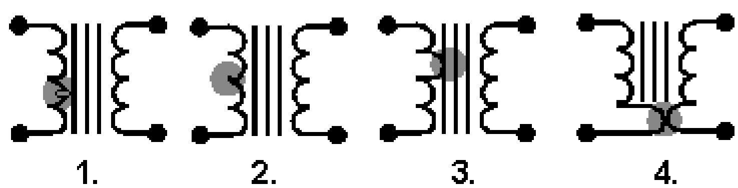
1. The filament, after long service, may be unable to emit as many electrons as are required for proper operation.
2. The filament may burn out.
3. A tube element-the plate, for instance-may break its connection with the tube base pin.
4. Two elements, such as filament and plate, may short together.
5. The tube may become gassy.

The symptoms you will come across in signal tracing will be many and varied. You will need to combine your "know-how" of the circuit and your knowledge of these five possible tube troubles to determine if the tube could in some way be causing the symptoms. If you suspect the tube of causing trouble, either try another tube in its place or check it on a tube tester. But remember, the final check of whether or not the old tube was bad is whether or not the equipment works properly when a good tube is put in its place. Therefore, putting in a good tube and then trying out the equipment is the best check.

Transformer and Choke Troubles

As you should know by now, the transformer and choke are quite similar in construction. Therefore, it is no coincidence that the basic troubles they can develop are the same.

1. A winding can open.
2. Two or more turns of one winding can short together.
3. A winding can short to the casing, which is usually grounded.
4. Two windings can short together. This trouble is possible, of course, only in transformers.



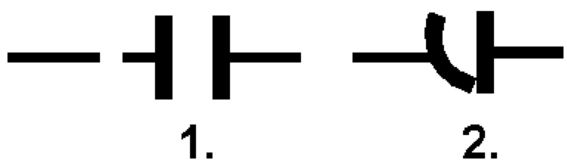
As with the tube, the symptoms of these troubles will vary with the type of circuit. However, when you have decided that one of these four possible troubles could be causing the symptoms, there are definite steps to take. If you surmise that there is an open winding or windings shorted together or to ground, an ohmmeter continuity check will locate the trouble. If the turns of a winding are shorted together, you may not be able to detect a difference in winding resistance. Therefore, you need to connect a good transformer in the place of the old one and see if the symptoms are eliminated; but keep in mind that transformers are difficult to replace. Make absolutely sure that the trouble is not elsewhere in the circuit before you change the transformer.

Occasionally, shorts will appear only when operating voltages are applied to the transformer. In this case you might find the trouble with a megger—an instrument that applies a high voltage as it reads resistance.

Capacitor and Resistor Troubles

Only two things can happen to a capacitor:

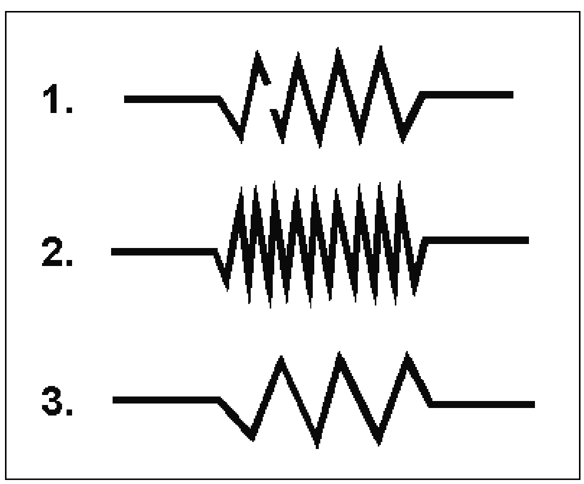
1. It may open up, removing itself completely from the circuit.
2. It may develop an internal short circuit. This means that it begins to pass current as though it were a resistor or a direct short.



You can check a capacitor you suspect of being open by disconnecting it from the circuit and checking it with a capacitor analyzer. You can check a capacitor you suspect of being leaky with an ohmmeter; if it reads less than 500 kilohms, it is more than likely bad. However, capacitor troubles are difficult to find since they may appear intermittently or only under operating voltages. Therefore, the best check for a faulty capacitor is to replace it with one you know to be good. If this restores proper operation, the fault was in the capacitor.

Resistor troubles are the simplest; but like the rest, you must keep them in mind.

1. A resistor can open up.
2. A resistor can increase in value.
3. A resistor can decrease in value.



You already know how to check possible resistor troubles. Just use an ohmmeter after making sure no parallel circuit is connected across the resistor you wish to measure. When you know a parallel circuit is connected across the resistor or when you are in doubt, disconnect one end of the resistor before measuring it. The ohmmeter check will usually be adequate. However, never forget that intermittent troubles may develop in resistors as well as in any other electronic parts. Also remember that the final proof that a resistor is bad is when you replace it with another resistor and the equipment operates satisfactorily.

Although you may observe problems that we have not covered specifically in this chapter, you should have gained enough knowledge to localize and repair any problem that may occur.

Q43. What is the most important thing to remember when troubleshooting?

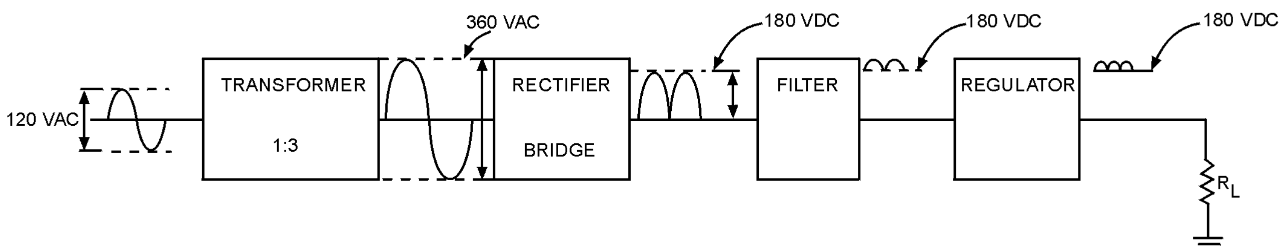
Q44. What is the main reason for grounding the return side of the transformer to the chassis?

Q45. What are two types of checks used in troubleshooting power supplies?

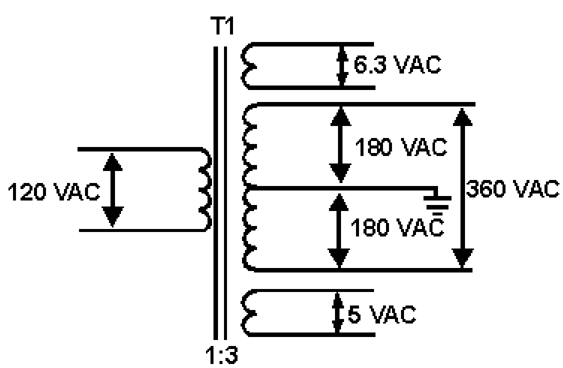
SUMMARY

In this chapter, we have presented you a basic description of the theory and operation of a basic power supply and its components. The following summary should enhance your understanding of power supplies.

POWER SUPPLIES are electronic circuits designed to convert ac to dc at any desired level. Almost all power supplies are composed of four sections: transformer, rectifier, filter, and regulator.

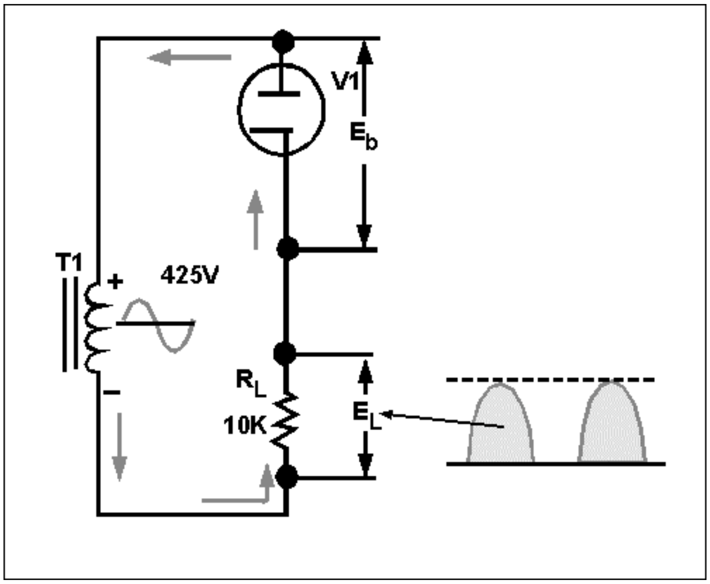


The **POWER TRANSFORMER** is the input transformer for the power supply. In addition to the high voltage, the power transformer also supplies filament voltage.

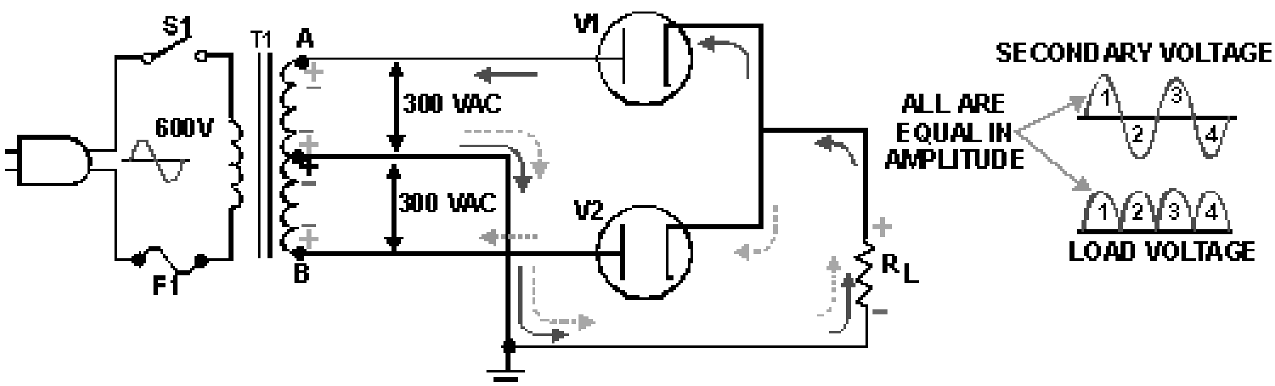


The **RECTIFIER** is the section of the power supply that contains the secondary windings of the power transformer and the rectifier circuit. The rectifier uses the ability of a diode to conduct during one half cycle of ac to convert ac to dc.

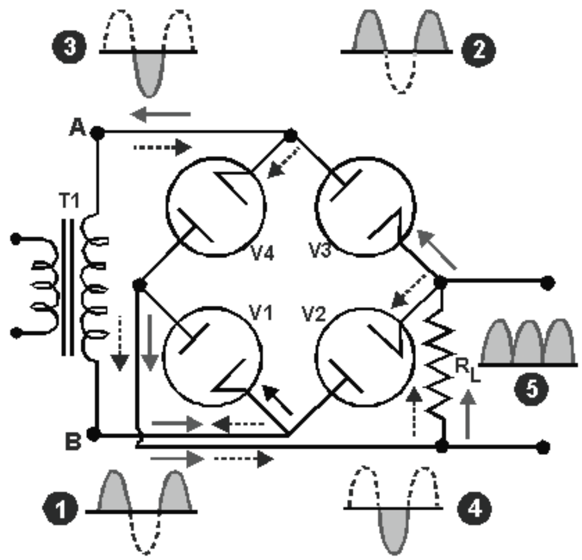
HALF-WAVE RECTIFIERS give an output on only one half cycle of the input ac. For this reason, the pulses of dc are separated by a period of one half cycle of zero potential voltage.



FULL-WAVE RECTIFIERS conduct on both halves of the input ac cycles. As a result, the dc pulses are not separated from each other. A characteristic of full-wave rectifiers is the use of a center-tapped, high-voltage secondary. Because of the center tap, the output of the rectifier is limited to one-half of the input voltage of the high-voltage secondary.

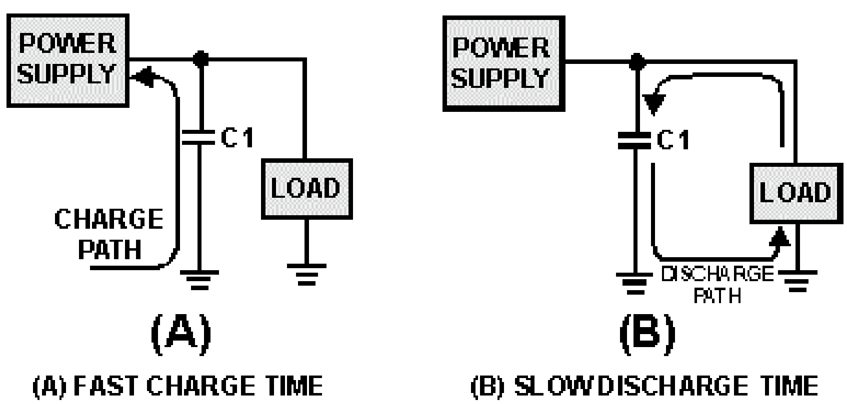


BRIDGE RECTIFIERS are full-wave rectifiers that do not use a center-tapped, high-voltage secondary. Because of this their dc output voltage is equal to the input voltage from the high-voltage secondary of the power transformer. Bridge rectifiers use four diodes connected in a bridge network. Tubes conduct in diagonal pairs to give a full-wave pulsating dc output.

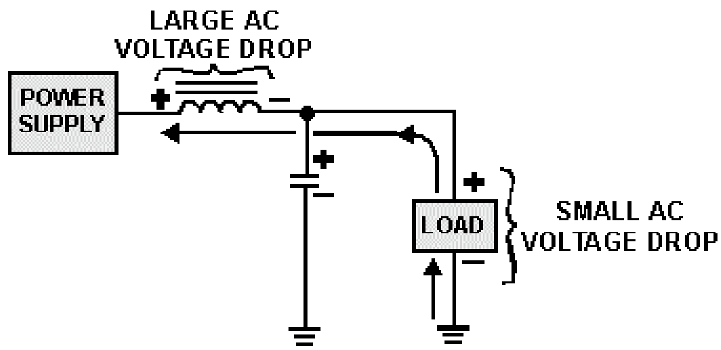


FILTER CIRCUITS are designed to smooth, or filter, the ripple voltage present on the pulsating dc output of the rectifier. This is done by an electrical device that has the ability to store energy and to release the stored energy.

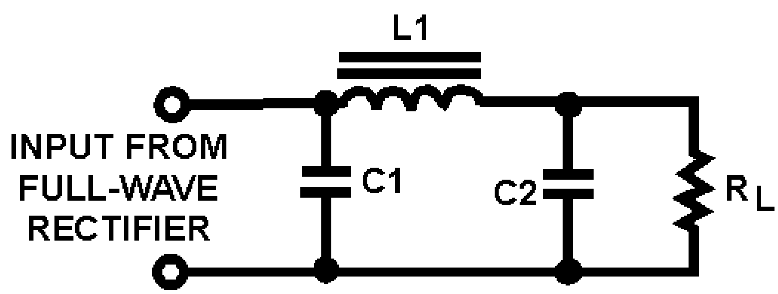
CAPACITANCE FILTERS are nothing more than large capacitors placed across the output of the rectifier section. Because of the large size of the capacitors, fast charge paths, and slow discharge paths, the capacitor will charge to average value, which will keep the pulsating dc output from reaching zero volts.



INDUCTOR FILTERS use an inductor called a choke to filter the pulsating dc input. Because of the impedance offered to circuit current, the output of the filter is at a lower amplitude than the input.

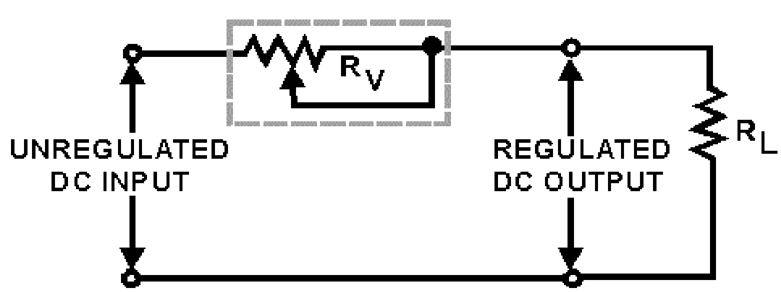


PI-TYPE FILTERS use both capacitive and inductive filters connected in a pi-type configuration. Because of the combination of filtering devices, the ability of the pi filter to remove ripple voltage is superior to that of either the capacitance or inductance filter.

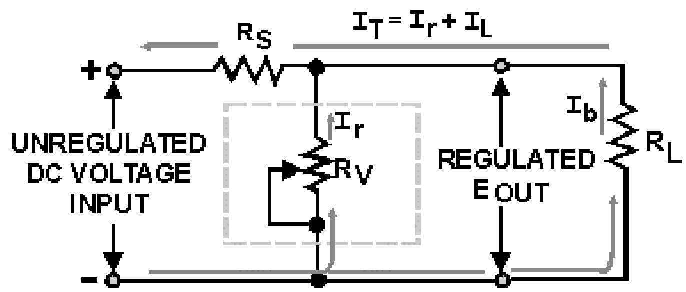


VOLTAGE REGULATORS are circuits designed to maintain the output of power supplies at a constant amplitude despite variations of the ac source voltage or changes of the resistance of the load. This is done by creating a voltage divider of a resistive element in the regulator and the resistance of the load. Regulation is achieved by varying the resistance of the resistive element in the regulator.

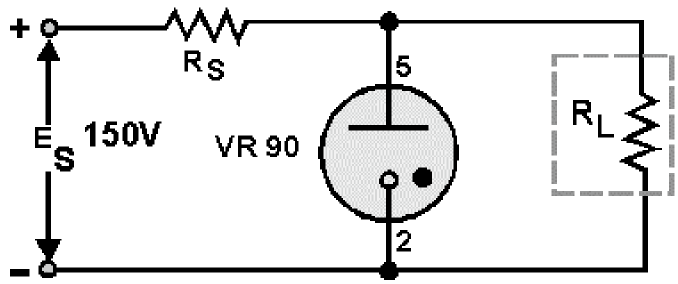
A **SERIES REGULATOR** uses a variable resistance in series with the load. Regulation is achieved by varying this resistance either to increase or decrease the voltage drop across the resistive element of the regulator. Characteristically, the resistance of the variable resistance moves in the same direction as the load. When the resistance of the load increases, the variable resistance of the regulator increases; when load resistance decreases, the variable resistance of the regulator decreases.



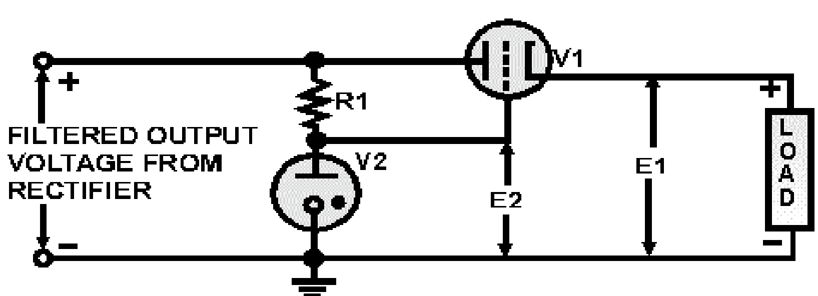
SHUNT REGULATORS use a variable resistance placed in parallel with the load. Regulation is achieved by keeping the resistance of the load constant. Characteristically, the resistance of the shunt moves in the opposite direction of the resistance of the load.



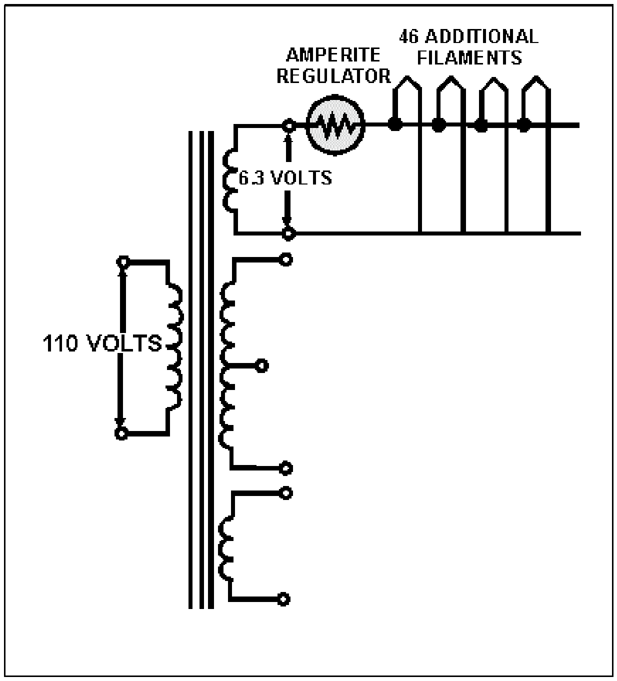
VR-TUBE REGULATORS are shunt regulators that use a cold cathode as a variable resistance in parallel with the load. Because of their ability to maintain a constant voltage potential between their plates and cathode, glow tubes can be connected in series to regulate any voltage. Additionally, glow tubes can be used to deliver different voltages to different loads.



SIMPLE ELECTRON TUBE REGULATORS use the dc plate resistance of a triode as a variable resistance in series with the load. The resistance of the vacuum tube is varied by changing the amount of conduction of the tube. This is done by holding the control grid voltage at a constant level and allowing the cathode voltage to vary with the output voltage.



The **AMPERITE VOLTAGE REGULATOR** or **BALLAST TUBE** is normally used to control current surges. This is done by heating an iron wire in a hydrogen-filled envelope. The hot iron will present a large resistance to current flow.



ANSWERS TO QUESTIONS Q1. THROUGH Q45.

A1. *Transformer, rectifier, filter, regulator.*

A2. *To maintain a constant voltage to the load.*

A3. *It couples the power supply to the ac line voltage, isolates the ac line voltage from the load, and steps this voltage up or down to the desired level.*

A4. *Filament voltage to the electron tubes.*

A5. *Provides capability of developing two high-voltage outputs.*

A6. *Positive.*

A7. *Cutoff.*

A8. *Pulsating dc.*

A9. *Series.*

A10. *60 hertz.*

A11. $E_{avg} = 0.318 \times E_{max}$.

A12. *120 hertz.*

A13. *63.6 volts.*

A14. *The peak voltage is half that of a half-wave rectifier.*

A15. *The bridge rectifier can produce double the voltage with the same size transformer.*

A16. Decrease-Capacitance is inversely proportional to X_C.

A17. Capacitor.

A18. Parallel.

A19. High.

A20. Increase.

A21. Value of capacitance and load resistance.

A22. Good.

A23. Yes.

A24. Counter electro-motive force of the inductor.

A25. 1 to 20 henries.

A26. Decrease.

A27. Small.

A28. Expense.

A29. When ripple must be held at an absolute minimum.

A30. LC capacitor-input filter.

A31. Cost of the inductor and size of the inductor.

A32. Ac line voltage and a change in load resistance.

A33. Percent of regulator.

A34. 33.33%

A35. 0%.

A36. Variation.

A37. Series and shunt.

A38. Increase.

A39. Parallel.

A40. Variable resistor.

A41. Current regulation.

A42. Decrease.

A43. Safety precautions.

A44. Reduce the cost of manufacturing equipment.

A45. Visual and signal tracing.

APPENDIX I

GLOSSARY

ACCELERATING ANODE—An electrode charged several thousand volts positive and used to accelerate electrons toward the front of a cathode-ray tube.

ACORN TUBE—A very small tube with closely spaced electrodes and no base. The tube is connected to its circuits by short wire pins that are sealed in a glass or ceramic envelope. The acorn tube is used in low-power uhf circuits.

AMPLIFICATION—The ratio of output magnitude to input magnitude in a device intended to produce an output that is an enlarged reproduction of its input.

AMPLIFICATION FACTOR—The voltage gain of an amplifier with no load on the output.

AMPLITUDE DISTORTION—Distortion that is present in an amplifier when the amplitude of the output signal fails to follow exactly any increase or decrease in the amplitude of the input signal.

AMPERITE (BALLAST) TUBE—A current-controlling resistance device designed to maintain substantially constant current over a specified range of variation in applied voltage or resistance of a series circuit.

ANODE—A positive electrode of an electrochemical device (such as a primary or secondary electric cell) toward which the negative ions are drawn.

AQUADAG COATING—A special coating of a conductive material, such as graphite, which is applied to the inside of a CRT. This coating eliminates the effects of secondary emission and aids in the acceleration of electrons.

BEAM-POWER TUBE—An electron tube in which the grids are aligned with the control grid. Special beam-forming plates are used to concentrate the electron stream into a beam. Because of this action, the beam-power tube has high power-handling capabilities.

BRIGHTNESS CONTROL—The name given to the potentiometer used to vary the potential applied to the control grid of a CRT.

CATHODE—The general name for any negative electrode.

CATHODE BIAS—The method of biasing a vacuum tube by placing the biasing resistor in the common-cathode return circuit, thereby making the cathode more positive with respect to ground.

CATHODE-RAY TUBE (CRT)—An electron tube that has an electron gun, a deflection system, and a screen. This tube is used to display visual electronic signals.

CHOKE—An inductor used to impede the flow of pulsating dc or ac by means of self-inductance.

COLD-CATHODE TUBE—A gas-filled electron tube that conducts without the use of filaments. Cold-cathode tubes are used as voltage regulators.

CONTROL GRID—The electrode of a vacuum tube, other than a diode, upon which a signal voltage is impressed to regulate the plate current.

DEFLECTION PLATES—Two pairs of parallel electrodes, one pair set forward of the other and at right angles to each other, parallel to the axis of the electron stream within an electrostatic cathode-ray tube.

DEIONIZATION POTENTIAL—The potential at which ionization of the gas within a gas-filled tube ceases and conduction stops, also referred to as extinction potential.

DIFFERENCE OF POTENTIAL—The voltage existing between two points. It will result in the flow of electrons whenever a circuit is established between the two points.

DIODE—An electron tube containing two electrode, a cathode, and a plate.

DIRECTLY HEATED CATHODE—A wire, or filament, designed to emit electrons that flow from cathode to plate. This is done by passing a current through the filament; the current heats the filament to the point where electrons are emitted.

DISTORTION—An undesired change in the waveform of the original signal, resulting in an unfaithful reproduction of audio or video signals.

DOORKNOB TUBE—An electron tube that is similar to the acorn tube but larger. The doorknob tube is designed to operate (at high power) in the uhf frequencies.

E_p-I_p CURV—The characteristic curve of an electron tube used to graphically depict the relationship between plate voltage (E_p) and the plate current (I_p).

EDISON EFFECT—Also called Richardson Effect. The phenomenon wherein electrons emitted from a heated element within a vacuum tube will flow to a second element that is connected to a positive potential.

ELECTRON GUN—An electrode of a CRT that is equivalent to the cathode and control grid of conventional tubes. The electron gun produces a highly concentrated stream of electrons.

ELECTROSTATIC DEFLECTION—The method of deflecting an electron beam by passing it between parallel charged plates mounted inside a cathode-ray tube.

FILAMENT—The cathode of a thermionic tube, usually a wire or ribbon, which is heated by passing current through it.

FILTER—A selective network of resistors, capacitors, and inductors that offer comparatively little opposition to certain frequencies or to direct current, while blocking or attenuating other frequencies.

FIXED BIAS—A constant value of bias voltage.

FLEMING VALVE—An earlier name for a diode, or a two-electrode vacuum tube used as a detector.

FOCUSING ANODE—An electrode of a CRT that is used to focus the electrons into a tight beam.

FULL-WAVE RECTIFIER—A circuit that uses both positive and negative alternations in an alternating current to produce direct current.

GETTER—An alkali metal introduced into a vacuum tube during manufacture. It is fired after the tube has been evacuated to react chemically with (and eliminate) any remaining gases.

GRID BIAS—A constant potential applied between the grid and cathode of a vacuum tube to establish an operating point.

GRID CURRENT—The current that flows in the grid-to-cathode circuit of a vacuum tube.

GRID-LEAK BIAS—A high resistance connected across the grid capacitor or between the grid and cathode. It provides a dc path to limit the accumulation of a charge on the grid.

HALF-WAVE RECTIFIER—A rectifier using only one-half of each cycle to change ac to pulsating dc.

HEATER—Same as a filament.

HORIZONTAL-DEFLECTION PLATES—A pair of parallel electrodes in a CRT that moves the electron beam from side to side.

IMPLOSION—The inward bursting of a CRT due to high vacuum. The opposite of explosion.

INDIRECTLY HEATED CATHODE—Same as a directly heated cathode with one exception: The hot filament raises the temperature of the sleeve around the filament; the sleeve then becomes the electron emitter.

INTERELECTRODE CAPACITANCE—The capacitance between one electron-tube electrode and the next electrode toward the anode.

IONIZATION—The electrically charged particles produced by high energy radiation, such as light or ultraviolet rays, or by the collision of particles during thermal agitation.

IONIZATION POINT—The potential required to ionize the gas of a gas-filled tube. Sometimes called firing point.

LIGHTHOUSE TUBE—An electron tube shaped like a lighthouse, is designed to handle large amounts of power at uhf frequencies.

LINEAR—Having an output that varies in direct proportion to the input.

MULTI-ELECTRODE TUBE—An electron tube that is normally classified according to the number of grids. (The multi-electrode tube contains more than three grids).

MULTI-UNIT TUBE—An electron tube containing two or more units within the same envelope. The multi-unit tube is capable of operating as a single-unit tube or as separate tubes.

NONLINEAR—Having an output that does not rise of fall directly with the input.

OILCAN TUBE—A type of planar tube, similar to the lighthouse tube, which has cooling fins. The oilcan tube is designed to handle large amounts of power at uhf frequencies.

PEAK CURRENT—The maximum current that flows during a complete cycle.

PEAK-REVERSE VOLTAGE—The peak ac voltage which a rectifier tube will withstand in the reverse direction.

PEAK VOLTAGE—The maximum value present in a varying or alternating voltage. This may be positive or negative.

PENTODE TUBE—A five-electrode electron tube containing a plate, a cathode, a control grid, and two grids.

PERSISTENCE—The duration of time a display remains on the face of a CRT.

PHOSPHOR—The material used to convert the energy of electrons into visible light.

PLANAR TUBE—An electron tube, constructed with parallel electrodes and a ceramic envelope, which is used at uhf frequencies. It is commonly referred to as lighthouse tube.

PLATE DISSIPATION—The amount of power lost as heat in the plate of a vacuum tube.

PLATE RESISTANCE—The plate voltage change divided by the resultant plate current change in a vacuum tube, all other conditions being fixed.

POWER PENTODE—A special-purpose tube used to provide high-current gain or power amplification. Each grid wire is directly in line with the one before and after it, a fact that allows more electrons to reach the plate.

POWER SUPPLY—A unit that supplies electrical power to another unit. It changes ac to dc and maintains a constant voltage output within limits.

QUIESCENCE—The operating condition of a circuit when no input signal is being applied to the circuit.

RECTIFIER—A device which, by its construction characteristics, converts alternating current to a pulsating direct current.

REGULATOR—The section in a basic power supply that maintains the output of the power supply at a constant level in spite of large changes in load current or in input line voltage.

REMOTE-CUTOFF TUBE—An electron tube in which the control grid wires are farther apart at the centers than at the ends. This arrangement allows the tube to amplify large signals without being driven into cutoff. This tube is also called a **VARIABLE**-mu tube.

rgk—The symbol used to express the resistance between the grid and the cathode of an electron tube.

rpk—The symbol used to represent the variable resistance between the cathode and the plate of a tube.

RIPPLE FREQUENCY—The frequency of the ripple current. In a full-wave rectifier, it is twice the input line frequency.

RIPPLE VOLTAGE—The alternating component of unidirectional voltage. (This component is small compared to the direct component).

SATURATION—The point in a tube where a further increase in plate voltage no longer produces an increase in plate current. At this point the upper limit of the conduction capabilities of the tube has been reached.

SECONDARY EMISSION—The liberation of electrons from an element, other than the cathode, as a result of being struck by other high-velocity electrons.

SCREEN GRIN—A grid placed between a control grid and the plate and usually maintained at a positive potential.

SELF-BIAS—The voltage developed by the flow of vacuum-tube current through a resistor in a grid or cathode lead.

SHARP-CUTOFF TUBE—The opposite of a remote-cutoff tube. An electron tube which has evenly spaced grid wires. The amplification of the sharp-cutoff tube is limited by the bias voltage and characteristics.

SPACE CHARGE—An electrical charge distributed throughout a volume or space.

TETRODE TUBE—A four-electrode electron tube containing a plate, a cathode, a control grid, and a screen grid.

THERMIONIC EMISSION—Emission of electrons from a solid body as a result of elevated temperature.

THYRATRON TUBE—A gas-filled triode in which a sufficiently large positive pulse applied to the control grid ionizes the gas and causes the tube to conduct, after which the control grid has no effect in conduction.

TRANSCONDUCTANCE—A measure of the change in plate current to a change in grid voltage with the plate voltage held constant. Transconductance (gm) is usually expressed in micromhos or millimhos. Mathematically,

$$gm = \frac{I_p}{E_g}$$

TRANSIT TIME—The time an electron takes to cross the distance between the cathode and the plate.

TRIODE TUBE—A three-electrode electron tube containing a plate, a cathode, and a control grid.

VARIABLE-mu-TUBE—Same as remote-cut off tube.

VERTICAL DEFLECTION PLATES—A pair of parallel electrodes in a CRT that moves the electron beam up and down.

VOLTAGE GAIN—Ratio of voltage across a specified load.

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, “Introduction to Electron Tubes,” pages 1-1 through 1-56.

1-1. The electrons emitted by a heated conductor come from what source?

1. An external battery
2. An external ac source
3. Both 1 and 2 above
4. The conductor itself

1-2. What is another name for thermionic emission?

1. The gap effect
2. The heat effect
3. The Edison effect
4. The Fleming effect

1-3. Electrons emitted by a hot filament are able to cross the gap between the filament and the plate. What force enables them to do this?

1. Magnetic repulsion
2. Inductive reactance
3. Thermionic emission
4. Electrostatic attraction

1-4. Name the two series circuits that are associated with a diode electron tube.

1. The plate and anode circuits
2. The plate and filament circuits
3. The battery and filament circuits
4. The filament and cathode circuits

1-5. When an ac voltage is applied across the plate and filament of a diode, the current measured will represent what type of waveform?

1. Pulsating dc
2. Dc
3. Pulsating ac
4. Ac

1-6. A filament that uses a one-molecule-thick layer of barium and strontium is classified as what type of filament?

1. Tungsten
2. Oxide-coated
3. Tungsten-strontium
4. Thoriated-tungsten

1-7. Which of the following ac filament voltages is most likely to be considered a common voltage?

1. 1.5 volts
2. 3.0 volts
3. 6.3 volts
4. 9.0 volts

1-8. An ac directly heated filament has which of the following advantages?

1. Even spacing relative to the plate
2. Even emission across the filament
3. Constant emission throughout the ac cycle
4. Rapid heating effect

1-9. An indirectly heated cathode always uses what material for its emitting surface?

1. An oxide coating
2. A thorium coating
3. A tungsten coating
4. A graphite coating

1-10. What is the principal advantage of an indirectly heated cathode over a directly heated cathode?

1. It is larger
2. It is immune to ac heater current variations
3. It reaches an operating temperature more quickly
4. It has a lower operating temperature

1-11. When you view an electron tube and its socket connection from the bottom, in what direction are (a) the pins of the tube and (b) the pins of the socket numbered?

1. (a) Counterclockwise
   (b) Clockwise
2. (a) Counterclockwise
   (b) Counterclockwise
3. (a) Clockwise
   (b) Counterclockwise
4. (a) Clockwise
   (b) Clockwise

1-12. Electron tubes are identified by a number preceded by which of the following letter designations?

1. T
2. V
3. ET
4. VT

1-13. The getter in an electron tube serves what purpose?

1. It protects the plate from overheating
2. It allows the cathode to emit more electrons
3. It helps to produce a better vacuum
4. It anchors the tube elements in the base

IN ANSWERING QUESTIONS 1-14 THROUGH 1-16, MATCH EACH TERM LISTED IN COLUMN A WITH ITS ASSOCIATED ELECTRONIC SYMBOL LISTED IN COLUMN B.

	A. TERMS	B. SYMBOLS
1-14.	Dc plate resistance	1. E_p
1-15.	Dc plate current	2. e_p
1-16.	Dc plate voltage	3. I_p
		4. R_p

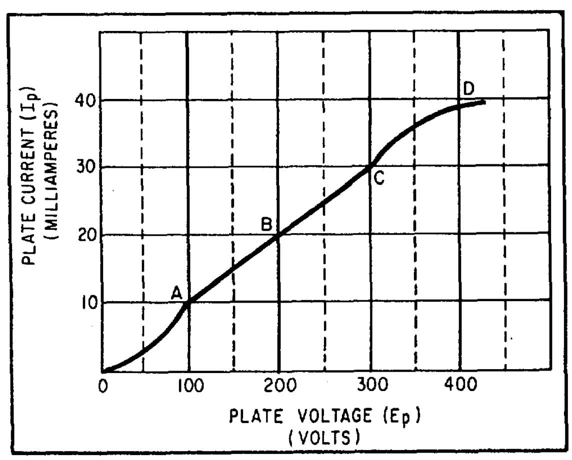


Figure 1A. — E_p — I_p characteristic curve.

IN ANSWERING QUESTIONS 1-17 THROUGH 1-23, REFER TO FIGURE 1A.

1-17. The area of the graph that lies between points C and D is referred to as

1. nonlinear
2. straight
3. linear
4. curved

1-18. In most applications, a designer would try to ensure that an electron tube operates at which of the following points on the curve?

1. A
2. B
3. C
4. D

1-19. An electron tube operating at point A on the curve would have what plate resistance?

1. 7 k Ω
2. 10 k Ω
3. 30 k Ω
4. 100 k Ω

1-20. An electron tube operating at point D can be said to be in what condition?

1. Plate saturation
2. Cathode saturation
3. Both 1 and 2 above
4. Normal operation

MATCH EACH ELECTRON- TUBE OPERATING CHARACTERISTIC IN COLUMN A WITH ITS CORRESPONDING CHARACTERISTIC- CURVE POINT IN COLUMN B.

	A. CHARACTERISTICS	B. POINTS
1-21.	Conduction occurs only at the outer fringe of the space charge	1. A
1-22.	All the electrons of the space charge are attracted to the plate	2. B
1-23.	The point at which the tube can be operated most efficiently	3. C
		4. D

1-24. An electron tube is operated at 300 volts and a plate current of 60 milliamperes. To avoid being damaged, the tube must have what minimum plate dissipation rating?

1. 5000 watts
2. 18 watts
3. 5 watts
4. 0.18 watt

1-25. Under which of the following conditions can a tube be considered operating beyond its peak inverse voltage rating?

1. When the plates glow cherry red
2. When current flows from the plate to the cathode
3. When current flows from the cathode in the form of an arc
4. When current flows from the cathode to the plate and damage occurs

1-26. Why does control grid voltage of a triode exercise greater control than plate voltage over conduction of the tube?

1. The grid is operated at a higher voltage than the plate
2. The grid adds electrons to the electron stream
3. The grid is closer to the plate than the cathode
4. The grid is closer to the cathode than the plate

1-27. The plate load resistor in an electron-tube circuit performs what function?

1. It converts variations in plate voltage to current variations
2. It limits the amount of plate voltage that can be applied to the tube
3. It converts variations in plate current to variations in plate voltage
4. It limits the amount of plate current that can flow through the tube

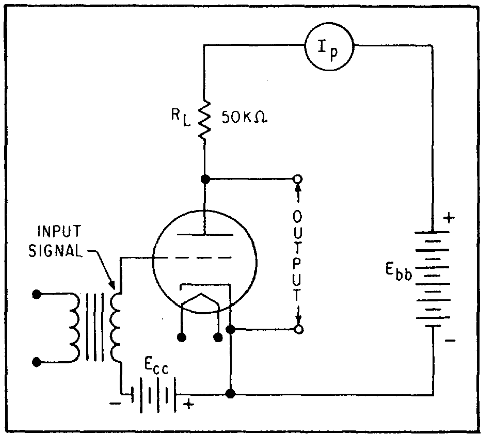


Figure 1B.—Triode operation.

1-28. The triode circuit depicted in figure 1B above contains a 50 k Ω load resistor. When a 10-volt peak-to-peak ac signal is applied to the grid, current flow in the tube varies between 5 milliamperes and 12 milliamperes. What is the peak-to-peak amplitude of the output?

1. 500 volts
2. 350 volts
3. 250 volts
4. 150 volts

1-29. Most amplifier circuits are designed to operate with the grid negative in relation to the cathode. This is done to avoid which of the following problems?

1. Excessive grid current
2. Excessive plate current
3. Distortion on small signals
4. Distortion on large negative signals

1-30. A triode amplifier has 350 volts applied to its plate across a 25 k Ω load resistor. With no input signal applied and a bias voltage of –9 volts, 4 milliamperes conducts across the tube. What is the quiescent plate voltage?

1. 0 V
2. 100 V
3. 250 V
4. 350 V

1-31. A triode electron tube is designed to conduct at 15 milliamperes of current when its grid is at 0 volts relative to its cathode. For every volt below this, conduction will decrease by 1.5 milliamperes. If the tube is biased at –3 volts and has a 6-volt peak-to-peak input signal, what is the minimum amount of current that will conduct through the tube?

1. 11.5 milliamperes
2. 6.0 milliamperes
3. 1.5 milliamperes
4. 0 milliamperes

1-32. Overdriving can be considered a form of distortion for which of the following reasons?

1. The output is not in phase with the input
2. The output does not have the same polarity as the input
3. The output is not a faithful reproduction of the input
4. The output does not have the same amplitude as the input

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 1-33 THROUGH 1-35, MATCH EACH CONDITION AFFECTING TRIODE AMPLIFIER OPERATION IN COLUMN A WITH ITS CORRESPONDING ELECTRONIC TERM IN COLUMN B.

	A. CONDITIONS	B. TERMS
1-33.	Condition that exists when the positive and negative excursions of the output are "flattopped"	1. Cutoff
1-34.	A form of distortion that can occur only during the positive excursion of the ac input of a triode amplifier	2. Saturation
1-35.	A form of distortion that can only occur in a triode amplifier during the negative excursion of the input	3. Overdriving
		4. Current limiting

1-36. Electronic equipment that uses fixed bias for its tube circuit receives its grid-bias voltage from what source?

1. A portion of the plate voltage
2. A power source internal to the circuit
3. Both 1 and 2 above
4. A power source external to the circuit

1-37. The effect of both cathode and grid biasing is to make the cathode (a) what polarity, relative to (b) what other tube element?

1. (a) Positive (b) the plate
2. (a) Negative (b) the plate
3. (a) Positive (b) the grid
4. (a) Negative (b) the grid

1-38. Which of the following types of biasing is most likely to use a battery supply?

1. Self
2. Grid
3. Fixed
4. Cathode

1-39. In an electron tube circuit using cathode biasing, the cathode is made positive in relation to the grid. This is done by a voltage dropped across what circuit element?

1. R_L
2. R_k
3. C_c
4. C_k

1-40. The cathode bias voltage level applied to the cathode is maintained at a constant level by what circuit component?

1. R_L
2. R_k
3. C_c
4. C_k

1-41. Which of the following undesirable characteristics is associated with cathode biasing?

1. Plate voltage is increased by the voltage amount of biasing
2. The cathode is forced to operate at a positive potential
3. The maximum negative output is limited
4. Current must flow in the circuit continuously

1-42. Grid-leak biasing develops a biasing voltage from (a) what portion of the input signal and (b) by what type of action?

1. (a) Negative (b) resistive
2. (a) Negative (b) capacitive
3. (a) Positive (b) capacitive
4. (a) Positive (b) resistive

1-43. During the charge cycle in grid-leak biasing, C_c, draws current through what circuit element?

1. R_g
2. rgk
3. R_L
4. R_k

1-44. During the discharge cycle in grid-leak biasing, C_c discharges across what circuit element?

1. R_g
2. rgk
3. R_L
4. R_k

1-45. The effect of grid-leak biasing is to rectify the input signal. Because of this, the amplitude of the biasing voltage depends upon which of the following factors?

1. Amplitude of the input
2. Frequency of the input
3. Size of R_g and C_c
4. All of the above

1-46. During the charging cycle in grid-leak biasing, the effective size of rgk is decreased. This is caused by what electronic principle?

1. Electrostatic repulsion between the grid and the plate
2. Electrostatic repulsion between the grid and the cathode
3. Electrostatic attraction between the cathode and the grid
4. Electrostatic attraction between the plate and the cathode

1-47. The charge and discharge of capacitor C_c, used in grid-leak circuits, will be equal when what condition occurs?

1. When Rgk becomes the same value as Rg
2. When C_c reaches its maximum charge-holding capacity
3. When the charge on C_c cuts the tube off
4. When R_g becomes larger than rgk

IN ANSWERING QUESTIONS 1-48 THROUGH 1-50, MATCH EACH CHARACTERISTIC OF AMPLIFIER OPERATION IN COLUMN A WITH ITS ASSOCIATED CLASS OF AMPLIFIER IN COLUMN B.

	A. CHARACTERISTICS	B. CLASSES
1-48.	Conduction occurs in the tube during only 50% of the entire input cycle	1. A 2. AB 3. B 4. C
1-49.	Conduction occurs in the tube throughout the entire input cycle	
1-50.	Conduction occurs in the tube for more than 50%, but less than 100% of the entire input cycle	

1-51. A triode amplifier has a load resistor rated at 150 kΩ. A +3-volt signal will cause 4 milliamperes of current to conduct through the tube. What is the voltage gain of the amplifier?

1. 450
2. 200
3. 100
4. 50

1-52. The amplification factor for an electron tube is identified by what electronic symbol?

1. A_r
2. V_g
3. g_m
4. μ

1-53. The grid voltage on an electron tube is increased from 2 volts to 4 volts. This causes plate current to increase from 2 milliamperes to 5.5 milliamperes. This same increase in plate current can be achieved by keeping the grid at +2 volts and raising the plate voltage from 200 volts to 400 volts. What is the mu of the tube?

1. 400
2. 200
3. 100
4. 50

1-54. What is the transconductance for the tube described in question 1-53?

1. 175 μmhos
2. 645 μmhos
3. 700 μmhos
4. 1750 μmhos

1-55. Transconductance is identified by what electronic symbol?

1. μ
2. g_m
3. rgk
4. t_c

1-56. In a triode, what interelectrode capacitance has the greatest effect on tube operation?

1. C_{pg}
2. C_{gk}
3. C_{pk}
4. C_{sg}

1-57. Interelectrode capacitance (C_{pg}) affects the gain of a triode stage because of what electronic feature?

1. Blocking
2. Feedback
3. Transit time
4. Phase inversion

1-58. The action of the screen grid in reducing interelectrode capacitance can be expressed mathematically as

1. $C_T = C1 + C2$
2. $C_T = C1 \times C2$
3. $C_T = \frac{C1 + C2}{C1 \times C2}$
4. $C_T = \frac{C1 \times C2}{C1 + C2}$

1-59. For normal operation, the screen grid of a tetrode is operated at a positive voltage in relation to (a) what tube element, and negative in relation to (b) what other tube element?

1. (a) Grid (b) plate
2. (a) Grid (b) cathode
3. (a) Plate (b) grid
4. (a) Cathod (b) grid

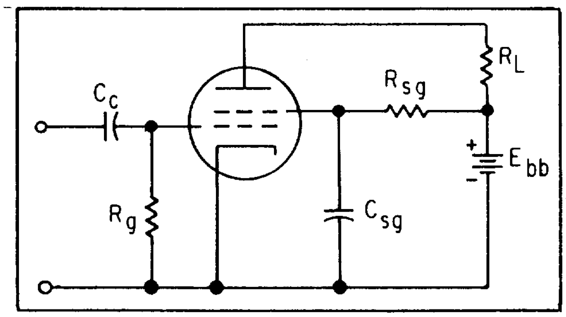


Figure 1C.—Basic tetrode circuit.

1-60. What is the function of C_{sg} in figure 1C above?

1. It serves as a feedback capacitor
2. It bypasses ac signals from the screen grid to ground
3. It keeps dc voltages from being applied to the screen grid from ground
4. It couples ac signals from the cathode to the screen grid

1-61. Which of the following undesirable characteristics is/are associated with tetrode operation?

1. The plate is isolated from the electron stream
2. The plate emits secondary emission electrons
3. The output is noisy
4. Both 2 and 3 above

1-62. Generally, tetrodes have a lower transconductance than triodes. This is caused by what feature of a tetrode?

1. The plate is isolated from the electron stream
2. The screen grid draws current from the electron steam
3. Secondary emission limits the amount of current the plate can draw from the electron stream
4. The screen grid is operated at a negative potential relative to the plate and electrons are repelled from the plate

1-63. The suppressor grid of a pentode is operated at what potential relative to (a) the cathode and (b) the plate?

1. (a) Positive
   (b) The same potential
2. (a) Negative
   (b) The same potential
3. (a) The same potentia
   (b) Negative
4. (a) The same potential
   (b) Positive

1-64. Voltage is supplied to the suppressor grid in a pentode from what source?

1. Through a resistor from the plate source voltage
2. Through a resistor from ground
3. By a separate voltage source
4. By a physical connection from the cathode

1-65. The suppressor grid is able to control the effects of secondary emission by using which of the following electronic actions?

1. By attracting electrons emitted by the plate through electromagnetic attraction
2. By repelling electrons emitted by the plate through electromagnetic repulsion
3. By attracting electrons emitted from the plate through electrostatic attraction
4. By repelling electrons emitted from the plate through electrostatic repulsion

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Special-Purpose Tubes," pages 2-1 through 2-39.

2-1. Which of the following types of tubes would be used as a voltage amplifier in an electronic circuit?

1. Diode
2. Triode
3. Duo-diode
4. Tetrahedral

2-2. Multielectrode tubes are normally classified according to the number of

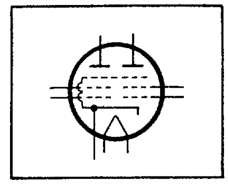
1. units contained in the tube
2. grids contained in the tube
3. elements contained in the tube
4. filaments contained in the tube

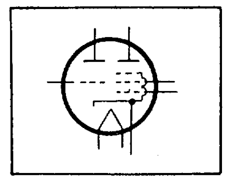
2-3. How many grids are there in a pentagrid tube?

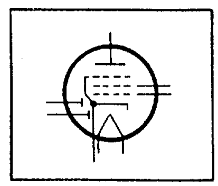
1. Five
2. Six
3. Seven
4. Eight

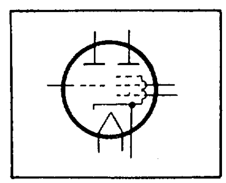
THIS SPACE LEFT BLANK INTENTIONALLY.

2-4. Which of the following diagrams represents a twin pentode?

1. 

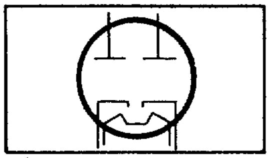
2. 

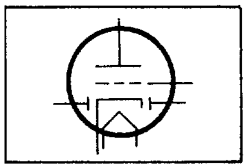
3. 

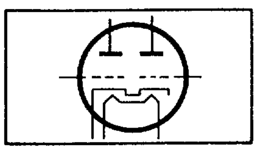
4. 

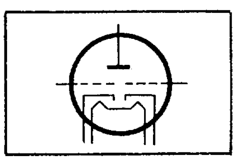
THIS SPACE LEFT BLANK INTENTIONALLY.

2-5. Which of the following diagrams represents a twin-input triode?

1. 

2. 

3. 

4. 

2-6. What advantage(s) does the in-line grid arrangement of the power pentode have over the staggered grid arrangement of the conventional pentode?

1. Higher efficiency
2. Higher power output
3. Both 1 and 2 above
4. Smaller current requirement

2-7. Which of the following is an advantage that a power pentode has over a conventional pentode?

1. Greater opposition to electron flow
2. Higher gain because of staggered grids
3. Greater sensitivity to small signals
4. Smaller plate current obtained from large signals

2-8. What is the primary purpose of the beam-forming plates in the beam-power tube?

1. To concentrate the electrons into a beam
2. To catch any stray electrons in the tube
3. To act as an extension of the cathode
4. To give the tube the appearance of a pentode

2-9. Which of the following is a name given to the variable-mu tube?

1. Sharp-cutoff tube
2. Remote-cutoff tube
3. Variable-spaced tube
4. Reversible-bias tube

2-10. What is the symbol for "mu"?

1. μ
2. β
3. I
4. L

2-11. Which of the following is an advantage of the variable-mu tube over conventional tubes?

1. It can be driven into cut-off by remote signals
2. It can be saturated quickly with a small input signal
3. It can amplify small input signals without distortion
4. It can amplify large input signals without distortion

2-12. What is the only difference between a remote-cutoff tube and a sharp-cutoff tube?

1. The spacing of the grid wires
2. The number of grids in each tube
3. The bias voltage used for conduction
4. The potential on the elements of each tube

2-13. Which of the following is the BEST method for reducing transit time in uhf tubes?

1. Placing the elements very close together
2. Increasing the voltage on the electrodes
3. Increasing the velocity of electrons
4. Placing the elements far apart

2-14. Which of the following is a disadvantage of uhf tubes?

1. Interelectrode capacitance is reduced
2. They are manufactured without socket bases
3. All the physical dimensions are scaled small
4. They have reduced power-handling capabilities

2-15. What is the only physical difference between the doorknob tube and the acorn tube?

1. Size
2. Base design
3. Filament material
4. Power-handling capability

2-16. How does the construction of a planar tube differ from that of a concentric tube?

1. Concentric tubes use filaments while planar tubes do not
2. Planar tubes use filaments while concentric tubes do not
3. The electrodes of the planar tubes are parallel to each other while those in concentric tubes are not

2-17. Why is the metallic ring of the planar tube grounded?

1. To eliminate shock hazards
2. To eliminate unwanted rf signals
3. To make removing the tube easier
4. To shunt the cathode current to ground

2-18. The metallic shell capacitive ground of a planar tube serves as what kind of capacitor in a cathode-bias circuit?

1. Grid
2. Bypass
3. Coupling
4. Plate-to-cathode

2-19. What is the major difference between the oilcan tube and the lighthouse tube?

1. The oilcan tube has cooling fins; the lighthouse tube does not
2. The oilcan tube functions as a triode
3. The lighthouse tube can handle more power
4. The lighthouse tube is a diode-type tube

2-20. Which of the following is an advantage that an oilcan tube has over a lighthouse tube?

1. The oilcan tube is smaller
2. The oilcan tube has no filaments
3. The oilcan tube can handle more power
4. The oilcan tube can operate at hf and uhf frequencies

2-21. The plate potential at which ionization occurs is known as the ionization point. Which of the following is also a name for this process?

1. Firing potential
2. Saturation potential
3. Extinction potential
4. Deionization potential

2-22. What name is given to the value of plate voltage at which ionization stops?

1. Firing potential
2. Saturation potential
3. Extinction potential
4. High plate potential

2-23. When a gas-filled triode ionizes, the grid loses control and the tube then functions as what type of tube?

1. Diode
2. Triode
3. Duo-diode
4. Trigatron

2-24. After the gas-filled triode ionizes and the grid loses control, which of the following methods is used to stop the conduction of the tube?

1. Increasing the plate potential
2. Increasing the grid potential
3. Removing the plate potential
4. Removing the grid potential

2-25. What is the name given to the gas-filled triode?

1. Variable triode
2. Trigatron
3. Thyristor
4. Thyratron

2-26. For what minimum amount of time must the filaments of a mercury-vapor tube have voltage applied before the plate voltage is applied to the tube?

1. 0.5 minute
2. 1.5 minutes
3. 2.5 minutes
4. 3.0 minutes

2-27. Which of the following conditions is/are responsible for the soft, blue glow of the gas-filled triode?

1. The tube is operating normally
2. The tube is gassy
3. The tube is saturated
4. The tube is ionized

2-28. Which of the following types of tubes is normally used as a voltage regulator?

1. Gas-filled triode
2. Gas-filled diode
3. Cold cathode
4. Cold plate

2-29. For a cold-cathode tube, how does the voltage regulator maintain a constant voltage drop across the tube?

1. By changing the current flow of the tube
2. By changing the resistance of the tube as current flow varies
3. By changing the plate potential of the tube as current varies
4. By changing the source voltage of the tube

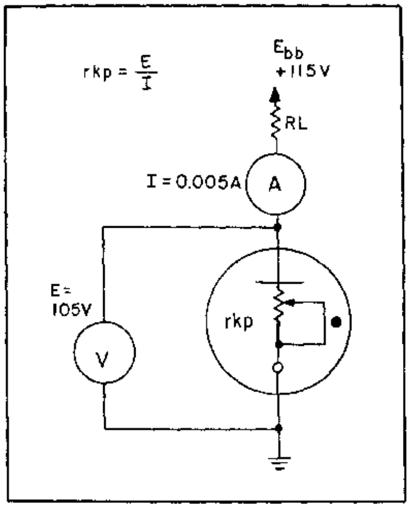


Figure 2A.—Cold-cathode tube operation.

2-30. If the source voltage in figure 2A is increased to 150 volts and the ammeter reads 20 milliamperes, what is the resistance (rkp) of the tube?

1. 2.65 kΩ
2. 5.25 kΩ
3. 6.25 kΩ
4. 8.15 kΩ

2-31. The electron gun of the CRT serves which of the following functions?

1. Deflects electrons into the plate
2. Concentrates electrons into a beam
3. Emits electrons
4. Both 2 and 3 above

2-32. Which of the following is a description of the grid in a CRT?

1. A metal cap with a hole in the center
2. A metal cap at ground potential
3. A metal cap with a positive potential
4. A metal cap with a wire screen in the center

2-33. What element of a television CRT is adjusted by the brightness control?

1. Cathode
2. Aquadag
3. Control grid
4. Focusing anode

2-34. Which of the following elements of the CRT helps prevent the beam of electrons from diverging?

1. Cathode
2. Aquadag
3. Focusing anode
4. Decelerating anode

2-35. Which of the following elements of a CRT has the highest positive potential?

1. The focusing anode
2. The electronic lens
3. The accelerating anode
4. The decelerating anode

2-36. What is the name of the florescent material that coats the inside face of a CRT?

1. Posporus
2. Phosphor
3. Flourine
4. Flourese

2-37. What is the purpose of the aquadag coating in the CRT?

1. It is used as a plate
2. It is used to focus the beam
3. It eliminates the space charge
4. It eliminates the effects of secondary emission

2-38. In which of the following equipment would you most likely find a cathode-ray tube?

1. Oscillator
2. Oscilloscope
3. Television set
4. Both 2 and 3

2-39. If deflection were not used in the CRT, what would be viewed on the screen of the tube?

1. A solid black screen
2. A solid white screen
3. A large spot on the left
4. A bright spot in the center

2-40. Which of the following types of deflection is used by much of the test equipment in the Navy?

1. Electromagnetic
2. Electrostatic
3. Magnetic
4. Static

2-41. Which of the following elements cause(s) the electron beam to move from left to right on a CRT?

1. Vertical deflection plates
2. Horizontal deflection plates
3. Suppressor grid
4. Control grid

THIS SPACE LEFT BLANK INTENTIONALLY.

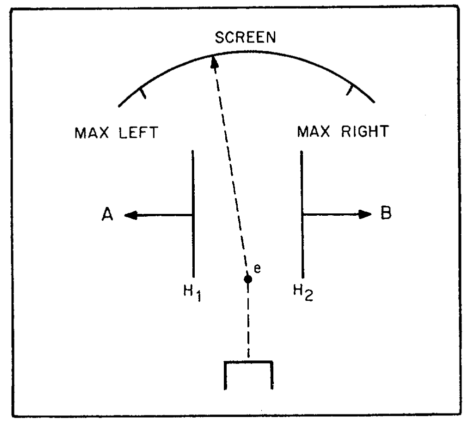


Figure 2B.—Deflection in a CRT.

2-42. In figure 2B, what potentials should be applied to points A and B to make electron (e) deflect in the direction as shown?

1. Point A zero and point B max positive
2. Point A max negative and point B max positive
3. Point A slightly positive and point B slightly negative
4. Point A slightly positive and point B zero

THIS SPACE LEFT BLANK INTENTIONALLY.

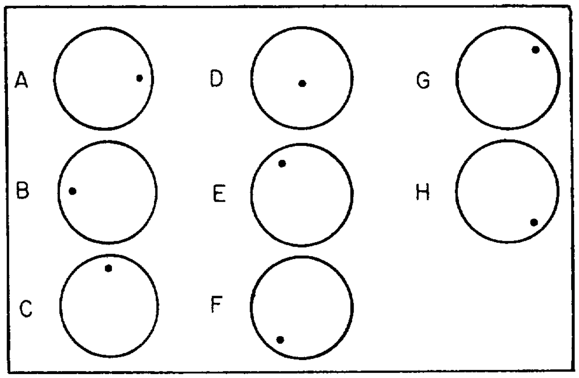
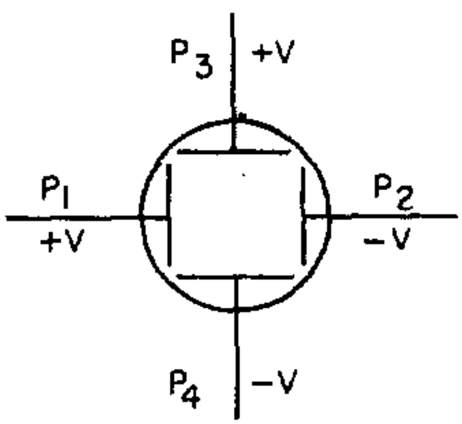


Figure 2C.—Deflection in a CRT (front view).

IN ANSWERING QUESTIONS 2-43 THROUGH 2-45, REFER TO FIGURE 2C. SELECT FROM CHOICES A THROUGH H THE DOT DISPLAY THAT MOST ACCURATELY REPRESENTS THE CONDITIONS OF THE PLATES SHOWN FOR EACH QUESTION.

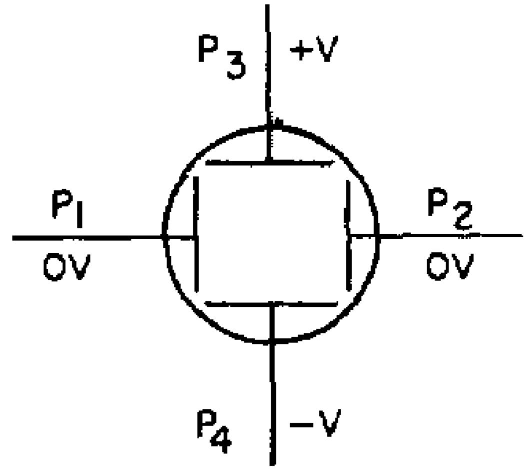
2-43.

1. E
2. F
3. G
4. H



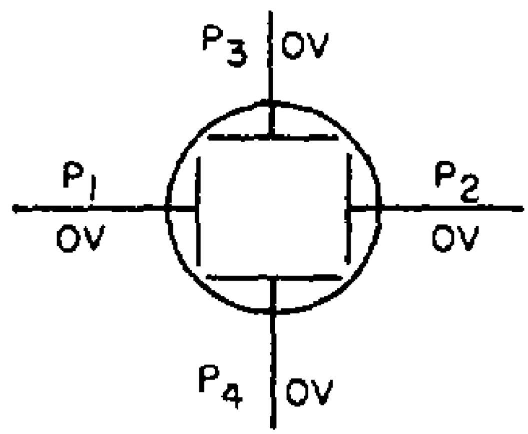
2-44.

1. A
2. B
3. C
4. D



2-45.

1. A
2. B
3. C
4. D



2-46. If a signal is to be viewed on a CRT, the signal should be applied to which of the following elements of the CRT?

1. Control grid
2. Vertical plates
3. Suppressor grid
4. Horizontal plates

IN ANSWERING QUESTIONS 2-47 THROUGH 2-56, SELECT FROM COLUMN B THE ELEMENT OF THE CRT WHICH PROVIDES THE FUNCTION DESCRIBED IN COLUMN A. YOU MAY USE THE ELEMENTS IN COLUMN B MORE THAN ONE TIME.

A. FUNCTIONS	B. ELEMENTS
2-47. Reduces interelectrode capacitance	1. Cathode
2-48. Controls electrons	2. Control grid
2-49. Source of electrons	3. Focusing anode
2-50. Accelerates electrons	4. Accelerating anode
2-51. Focuses electrons into a beam	

A. FUNCTIONS	B. ELEMENTS
2-52. Acts as suppressor grid	1. Vertical deflection plates
2-53. Displays electron beam	2. Horizontal deflection plates
2-54. Deflects beam to right	3. Aquadag coating
2-55. Deflects beam to down position	4. Screen
2-56. Eliminates secondary emission	

2-57. Which of the following actions must you take first before disposing of a cathode-ray tube?

1. Place the CRT carefully in a dumpster
2. Throw the CRT into deep water
3. Return the CRT to supply
4. Render the CRT harmless

2-58. What is the purpose of adding radioactive material to electron tubes?

1. The material reduces secondary emissions
2. The material aids ionization in the tube
3. The material increases therminonic emission
4. The material causes the tube to glow in the dark

2-59. Safety precautions and procedures for working with radioactive electron tubes can be found in which of the following publications?

1. Radiation, Health, and Protection Manual
2. Decontamination of Radioactivity Manual
3. Technical Manual for RF Radiation Hazards
4. Radiation Hazards of Shipboard Equipment Manual

ASSIGNMENT 3

Textbook assignment: Chapter 3, "Power Supplies," pages 3-1 through 3-59.

3-1. The electronic power supply was developed to fulfill which of the following needs?

1. Reliability
2. Convenience
3. Cost effectiveness
4. All of the above

3-2. Which of the following is NOT one of the four sections of a basic power supply?

1. Transformer
2. Oscillator
3. Rectifier
4. Filter

3-3. The primary purpose of the transformer in an electronic power supply is to isolate the power supply from ground.

1. True
2. False

3-4. What is the primary function of the rectifier section?

1. To convert dc to ac
2. To convert ac to pulsating dc
3. To increase average voltage output
4. To decrease average voltage output

3-5. What is/are the function(s) of the filter section?

1. To eliminate dc voltage
2. To increase the amplitude of the ac
3. To convert pulsating dc to steady dc
4. All of the above

3-6. The separate step-down windings in a transformer provide which of the following functions?

1. Filament voltage for power supply tubes
2. Filament voltage for the electronic load
3. Both 1 and 2 above
4. High voltage for the rectifier

3-7. The purpose of a center tap in a transformer is to provide

1. two separate filament voltages to the rectifier
2. a step-down voltage to the rectifier
3. pulsating dc to the rectifier
4. two outputs from one transformer

3-8. A diode vacuum tube is an ideal rectifier for which, if any, of the following reasons?

1. Current flows through the diode vacuum tube in one direction only
2. Current flows through the diode vacuum tube in both directions
3. The diode vacuum tube conducts only on the negative alternation of the input voltage
4. None of the above

3-9. When the plate of a diode tube is negative with respect to the cathode, the tube is said to be in what state?

1. Cutoff
2. Remission
3. Saturation
4. Conduction

3-10. In a simple half-wave rectifier, the diode tube will conduct for a maximum of how many degrees of the 360-degree input signal?

1. 45
2. 90
3. 180
4. 270

3-11. What term is used to describe current pulses that flow in the same direction?

1. Average current
2. Secondary current
3. Pure direct current
4. Pulsating direct current

3-12. For a diode to act as a rectifier, how should it be connected in a circuit?

1. In parallel with the input
2. In parallel with the load
3. In series with the input
4. In series with the load

3-13. What is the ripple frequency of a half-wave rectifier with an input line frequency of 60 Hz?

1. 30 Hz
2. 60 Hz
3. 90 Hz
4. 120 Hz

3-14. In a half-wave rectifier, what is the average voltage output when the peak voltage is 300 volts?

1. 190.8 volts
2. 95.4 volts
3. 19.08 volts
4. 9.4 volts

3-15. The full-wave rectifier was developed for which of the following reasons?

1. To obtain the highest average voltage and current
2. To increase the number of components
3. To increase the value of the voltage
4. To obtain better regulation

3-16. What is the ripple frequency of a full-wave rectifier with an input line frequency of 60 Hz?

1. 30 Hz
2. 60 Hz
3. 90 Hz
4. 120 Hz

3-17. What is the average voltage output of a full-wave rectifier that has an output of 10 volts peak?

1. 3.18 volts
2. 6.37 volts
3. 31.8 volts
4. 63.7 volts

3-18. The primary disadvantage of the conventional full-wave rectifier is that the peak output voltage is only half that of the half-wave rectifier.

1. True
2. False

THIS SPACE LEFT BLANK INTENTIONALLY.

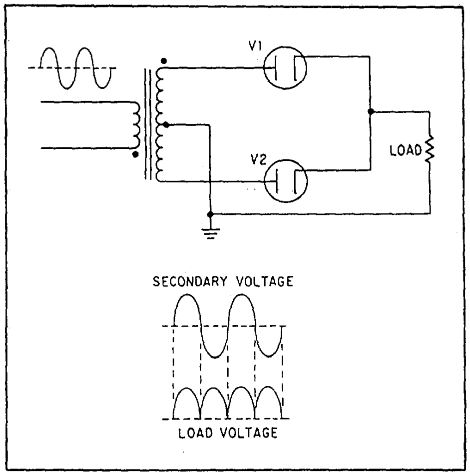


Figure 3A.—Complete full-wave rectifier.

IN ANSWERING QUESTIONS 3-19 AND 3-20, REFER TO FIGURE 3A. ASSUME THAT THE VOLTAGE ACROSS THE TRANSFORMER SECONDARY HAS AN RMS VALUE OF 240 VOLTS AC.

3-19. What is the peak value of the voltage pulses across the load?

1. 339.4 volts
2. 239.8 volts
3. 169.7 volts
4. 76.3 volts

3-20. What is the average output voltage?

1. 237.5 volts
2. 169.7 volts
3. 152.3 volts
4. 108.1 volts

THIS SPACE LEFT BLANK INTENTIONALLY.

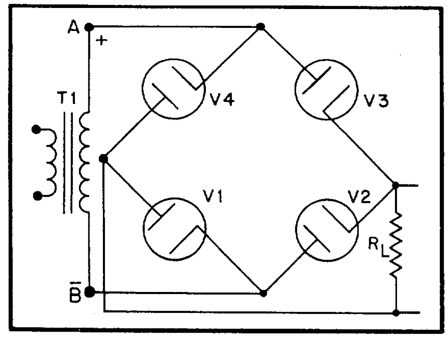


Figure 3B.—Bridge rectifier.

IN ANSWERING QUESTIONS 3-21 AND 3-22, REFER TO FIGURE 3B.

3-21. When the voltage across the secondary of the transformer has the polarity shown, which of the diodes will conduct?

1. V1 and V3
2. V2 and V4
3. V1 and V2
4. V3 and V4

3-22. When the polarity reverses, which of the diodes will conduct?

1. V3 and V1
2. V4 and V2
3. V2 and V1
4. V4 and V3

3-23. In filter circuits, inductors are used as what kind of impedances?

1. Shunt impedances to oppose changes in current
2. Shunt impedances to oppose changes in voltage
3. Series impedances to oppose changes in current
4. Series impedances to oppose changes in voltage

3-24. To retain its charge, the capacitor in a simple capacitor filter must have a long charge time constant and a short discharge time constant.

1. True
2. False

3-25. If the capacitance in a circuit increases, X_C, will increase.

1. True
2. False

3-26. To provide a steady dc output in a simple capacitor circuit, the capacitor must charge almost instantaneously to the value of the applied voltage.

1. True
2. False

3-27. What is the most basic type of filter?

1. Capacitor
2. LC choke input
3. LC capacitor input
4. RC capacitor input

3-28. In a circuit with a capacitor filter, how is the capacitor connected?

1. In series with the load
2. In parallel with the load
3. In parallel with the output
4. Both 2 and 3 above

3-29. Which, if any, of the following factors determines the rate of discharge of the capacitor in a filter circuit?

1. The value of the load resistance
2. The amount of voltage
3. The type of capacitor
4. None of the above

3-30. A half-wave rectifier has an output frequency of 60 hertz, a filter capacitor value of 40 microfarads, and a load resistance of 10 kilohms. What is the value of X_C?

1. 132.51 ohms
2. 66.25 ohms
3. 33.13 ohms
4. 16.57 ohms

3-31. A full-wave rectifier has an output frequency of 120 hertz, a filter capacitor value of 25 microfarads, and a load resistance of 10 kilohms. What is the value of X_C?

1. 5.3 ohms
2. 53 ohms
3. 106 ohms
4. 1060 ohms

3-32. The LC choke-input filter is used primarily where which of the following types of regulation is/are important?

1. Frequency
2. Current only
3. Voltage only
4. Voltage and current

3-33. In an LC choke-input filter circuit, the capacitor charges only to the average value of the input voltage. Which of the following components inhibits the capacitor from reaching the peak value of the input voltage?

1. Diode
2. Capacitor
3. Filter choke
4. Load resistor

3-34. In an LC choke-input filter, the larger the value of the filter capacitor, the better the filtering action. Which of the following factors represents the major limitation in obtaining the maximum value of the capacitor used?

1. Cost
2. Reliability
3. Availability
4. Physical size

3-35. What is the most common range of values selected for a power supply choke?

1. 1 to 20 henries
2. 5 to 25 henries
3. 25 to 30 henries
4. 10 to 200 henries

3-36. If the impedance of the choke in an LC choke-input filter is increased, the ripple voltage amplitude will

1. decrease
2. increase
3. oscillate
4. remain the same

3-37. A full-wave rectifier has an output frequency of 120 hertz, a filter choke with a value of 10 henries, and a load resistance of 10 kilohms. What is the value of X_L?

1. 75.0 ohms
2. 7.5 ohms
3. 75.0 kilohms
4. 7.5 kilohms

3-38. The filter capacitor in the LC choke-input filter is NOT subject to extreme voltage surges because of the protection provided by which of the following components?

1. Inductor
2. Load resistor
3. Series resistor
4. Shunt capacitor

3-39. Shorted turns in the choke of an LC choke-input filter may reduce the value of inductance below the critical value. When this happens, which of the following problems may occur?

1. Poor voltage regulation
2. Excessive ripple amplitude
3. Abnormally high output voltage
4. Each of the above

3-40. The use of the RC capacitor-input filter is limited to which of the following situations?

1. When the load current is large
2. When the load current is small
3. When the load voltage is large
4. When the load voltage is small

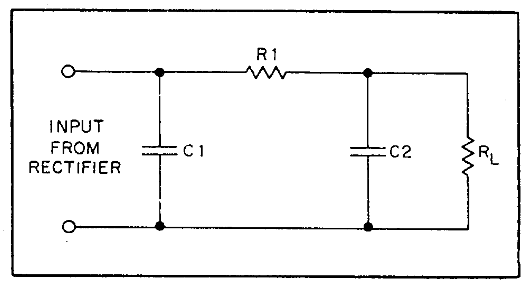


Figure 3C.—RC Capacitor-input filter.

IN ANSWERING QUESTIONS 3-41 AND 3-42, REFER TO FIGURE 3C.

3-41. Which of the following components will have the highest failure rate?

1. C1
2. C2
3. R1
4. R_L

3-42. Which of the following components provides protection against voltage surges in the circuit?

1. C1
2. C2
3. R1
4. R_L

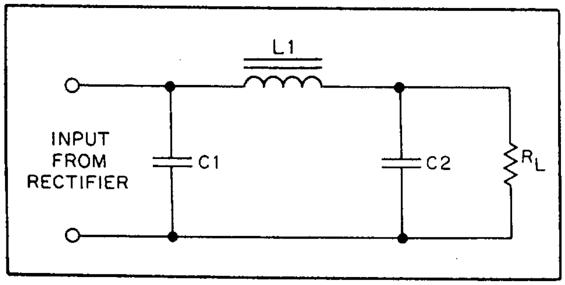


Figure 3D.—LC capacitor-input filter.

IN ANSWERING QUESTIONS 3-43 AND 3-44, REFER TO FIGURE 3D.

3-43. Components L1 and C2 form what type of circuit?

1. Ac voltage doubler
2. Dc voltage doubler
3. Ac voltage divider
4. Dc voltage divider

3-44. If component L1 shorts to the core, which of the following conditions will result?

1. No output
2. Excessively high output
3. Excessive ripple frequency
4. Low output ripple frequency

3-45. In a voltage regulator, what percent of regulation would be ideal?

1. 1 percent
2. 0 percent
3. 3 percent
4. 5 percent

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 3-46 THROUGH 3-48, REFER TO THE FOLLOWING FORMULA:

$$\text{Percent of Regulation} = \frac{E_{\text{no load}} - E_{\text{full load}}}{E_{\text{full load}}} \times 100$$

3-46. If a power supply produces 30 volts with no load and 25 volts under full load, what is the percent of regulation?

1. 5 percent
2. 10 percent
3. 20 percent
4. 30 percent

3-47. If a power supply produces 10 volts with no load and 9 volts under full load, what is the percent of regulation?

1. 8 percent
2. 9 percent
3. 10 percent
4. 11 percent

3-48. If a power supply produces 20 volts with no load and 20 volts under full load, what is the percent of regulation?

1. 1 percent
2. 2 percent
3. 3 percent
4. 0 percent

3-49. The simple series voltage regulator was designed to function as what type of resistance?

1. Fixed resistance in series with the load
2. Fixed resistance in parallel with the load
3. Variable resistance in series with the load
4. Variable resistance in parallel with the load

3-50. In a simple shunt voltage regulator, what is the purpose of the shunt element?

1. To regulate voltage through series resistance R_S
2. To regulate voltage through parallel resistance V1
3. To regulate current through series resistance R_S
4. To regulate current through parallel resistance V1

3-51. In an electron tube voltage regulator, the electron tube replaces which of the following components?

1. Variable resistor R_V
2. Parallel resistor R_P
3. Series resistor R_S
4. Load resistor R_L

3-52. The primary purpose of the amperite regulator is to regulate

1. power
2. voltage
3. current
4. resistance

3-53. What method is used by a manufacturer of electronic equipment to reduce the cost of extensive wiring?

1. Grounding the output of the power supply to the chassis
2. Grounding the return side of the power transformer to the chassis
3. Connecting all components in parallel
4. Connecting all components in series

3-54. When working on electronic equipment, the technician should observe which of the following safety precautions?

1. Make certain that the electronic equipment is properly grounded
2. Make certain that the test equipment is properly grounded
3. Make certain that the rubber mats are in good condition
4. All of the above

3-55. Which of the following is/are the most widely used check(s) for testing electronic equipment?

1. Smoke
2. Visual
3. Signal tracing
4. Both 2 and 3 above

3-56. Any connection that is located close to the chassis or any other terminal should be examined for the possibility of which of the following problems?

1. A short
2. An open
3. A low resistance
4. A high resistance

3-57. What is the condition of a transformer that is discolored or leaking?

1. Operational
2. Shorted
3. Cracked
4. Open

3-58. As a technician, you notice that a resistor is discolored and charred. The resistor has most likely been subjected to which, if any, of the following conditions?

1. Overload
2. Open circuit
3. Ambient temperature
4. None of the above

3-59. You are in the process of energizing a power supply. You hear a boiling or sputtering noise and notice smoke coming from a section of the power supply. Which, if any, of the following actions should you take first?

1. Secure power immediately
2. Examine the problem area
3. Remove the defective component
4. None of the above

3-60. Which, if any, of the following is the most rapid and accurate method for testing electronic circuits?

1. Smoke test
2. Visual test
3. Signal tracing
4. None of the above

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 7—Introduction to Solid-State Devices and Power Supplies

NAVEDTRA 14179

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its references to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 7 of a series.

History of the course:

Sep 1998: Original edition released.
Jun 2003: Administrative update released. Entered administrative updates. Technical content was *not revised.*

NAVSUP Logistics Tracking Number
0504-LP-026-8320

TABLE OF CONTENTS

ASSIGNMENT QUESTIONS follow Index.

CHAPTER 1

SEMICONDUCTOR DIODES

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the NRTC, you indicate that you have met the objectives and have learned the information. The learning objective are listed below.

Upon completion of this chapter, you should be able to do the following:

1. State, in terms of energy bands, the differences between a conductor, an insulator, and a semiconductor.
2. Explain the electron and the hole flow theory in semiconductors and how the semiconductor is affected by doping.
3. Define the term "diode" and give a brief description of its construction and operation.
4. Explain how the diode can be used as a half-wave rectifier and as a switch.
5. Identify the diode by its symbology, alphanumerical designation, and color code.
6. List the precautions that must be taken when working with diodes and describe the different ways to test them.

INTRODUCTION TO SOLID-STATE DEVICES

As you recall from previous studies in this series, semiconductors have electrical properties somewhere between those of insulators and conductors. The use of semiconductor materials in electronic components is not new; some devices are as old as the electron tube. Two of the most widely known semiconductors in use today are the JUNCTION DIODE and TRANSISTOR. These semiconductors fall under a more general heading called solid-state devices. A SOLID-STATE DEVICE is nothing more than an electronic device, which operates by virtue of the movement of electrons within a solid piece of semiconductor material.

Since the invention of the transistor, solid-state devices have been developed and improved at an unbelievable rate. Great strides have been made in the manufacturing techniques, and there is no foreseeable limit to the future of these devices. Solid-state devices made from semiconductor materials offer compactness, efficiency, ruggedness, and versatility. Consequently, these devices have invaded virtually every field of science and industry. In addition to the junction diode and transistor, a whole new family of related devices has been developed: the ZENER DIODE, LIGHT-EMITTING DIODE, FIELD EFFECT TRANSISTOR, etc. One development that has dominated solid-state technology, and probably has had a greater impact on the electronics industry than either the electron tube or transistor, is the INTEGRATED CIRCUIT. The integrated circuit is a minute piece of semiconductor material that can produce complete electronic circuit functions.

As the applications of solid-state devices mount, the need for knowledge of these devices becomes increasingly important. Personnel in the Navy today will have to understand solid-state devices if they are to become proficient in the repair and maintenance of electronic equipment. Therefore, our objective in this module is to provide a broad coverage of solid-state devices and, as a broad application, power supplies. We will begin our discussion with some background information on the development of the semiconductor. We will then proceed to the semiconductor diode, the transistor, special devices and, finally, solid-state power supplies.

SEMICONDUCTOR DEVELOPMENT

Although the semiconductor was late in reaching its present development, its story began long before the electron tube. Historically, we can go as far back as 1883 when Michael Faraday discovered that silver sulfide, a semiconductor, has a negative temperature coefficient. The term *negative temperature coefficient* is just another way of saying its resistance to electrical current flow decreases as temperature increases. The opposite is true of the conductor. It has a positive temperature coefficient. Because of this particular characteristic, semiconductors are used extensively in power-measuring equipment.

Only 2 years later, another valuable characteristic was reported by Munk A. Rosenshold. He found that certain materials have rectifying properties. Strange as it may seem, his finding was given such little notice that it had to be rediscovered 39 years later by F. Braun.

Toward the close of the 19th century, experimenters began to notice the peculiar characteristics of the chemical element SELENIUM. They discovered that in addition to its rectifying properties (the ability to convert ac into dc), selenium was also light sensitive-its resistance decreased with an increase in light intensity. This discovery eventually led to the invention of the photophone by Alexander Graham Bell. The photophone, which converted variations of light into sound, was a predecessor of the radio receiver; however, it wasn't until the actual birth of radio that selenium was used to any extent. Today, selenium is an important and widely used semiconductor.

Many other materials were tried and tested for use in communications. SILICON was found to be the most stable of the materials tested while GALENA, a crystalline form of lead sulfide, was found the most sensitive for use in early radio receivers. By 1915, Carl Beredicks discovered that GERMANIUM, another metallic element, also had rectifying capabilities. Later, it became widely used in electronics for low-power, low-frequency applications.

Although the semiconductor was known long before the electron tube was invented, the semiconductor devices of that time could not match the performance of the tube. Radio needed a device that could not only handle power and amplify but rectify and detect a signal as well. Since tubes could do all these things, whereas semiconductor devices of that day could not, the semiconductor soon lost out.

It wasn't until the beginning of World War II that interest was renewed in the semiconductor. There was a dire need for a device that could work within the ultra-high frequencies of radar. Electron tubes had interelectrode capacitances that were too high to do the job. The point-contact semiconductor diode, on the other hand, had a very low internal capacitance. Consequently, it filled the bill; it could be designed to work within the ultra-high frequencies used in radar, whereas the electron tube could not.

As radar took on greater importance and communication-electronic equipment became more sophisticated, the demands for better solid-state devices mounted. The limitations of the electron tube made necessary a quest for something new and different. An amplifying device was needed that was smaller, lighter, more efficient, and capable of handling extremely high frequencies. This was asking a lot, but if progress was to be made, these requirements had to be met. A serious study of semiconductor materials began in the early 1940's and has continued since.

In June 1948, a significant breakthrough took place in semiconductor development. This was the discovery of POINT-CONTACT TRANSISTOR. Here at last was a semiconductor that could amplify. This discovery brought the semiconductor back into competition with the electron tube. A year later, JUNCTION DIODES and TRANSISTORS were developed. The junction transistor was found superior to the point-contact type in many respects. By comparison, the junction transistor was more reliable, generated less noise, and had higher power-handling ability than its point-contact brother. The junction transistor became a rival of the electron tube in many uses previously uncontested.

Semiconductor diodes were not to be slighted. The initial work of Dr. Carl Zener led to the development of ZENER DIODE, which is frequently used today to regulate power supply voltages at precise levels. Considerably more interest in the solid-state diode was generated when Dr. Leo Esaki, a Japanese scientist, fabricated a diode that could amplify. The device, named the TUNNEL DIODE, has amazing gain and fast switching capabilities. Although it is used in the conventional amplifying and oscillating circuits, its primary use is in computer logic circuits.

Another breakthrough came in the late 1950's when it was discovered that semiconductor materials could be combined and treated so that they functioned as an entire circuit or subassembly rather than as a circuit component. Many names have been given to this solid-circuit concept, such as INTEGRATED CIRCUITS, MICROELECTRONICS, and MICROCIRCUITRY.

So as we see, in looking back, that the semiconductor is not something new, but it has come a long way in a short time.

Q1. What is a solid-state device?

Q2. Define the term negative temperature coefficient.

SEMICONDUCTOR APPLICATIONS

In the previous paragraphs, we mentioned just a few of the many different applications of semiconductor devices. The use of these devices has become so widespread that it would be impossible to list all their different applications. Instead, a broad coverage of their specific application is presented.

Semiconductor devices are all around us. They can be found in just about every commercial product we touch, from the family car to the pocket calculator. Semiconductor devices are contained in television sets, portable radios, stereo equipment, and much more.

Science and industry also rely heavily on semiconductor devices. Research laboratories use these devices in all sorts of electronic instruments to perform tests, measurements, and numerous other experimental tasks. Industrial control systems (such as those used to manufacture automobiles) and automatic telephone exchanges also use semiconductors. Even today heavy-duty versions of the solid-state rectifier diode are being use to convert large amounts of power for electric railroads. Of the many different applications for solid-state devices, space systems, computers, and data processing equipment are some of the largest consumers.

The various types of modem military equipment are literally loaded with semiconductor devices. Many radars, communication, and airborne equipment are transistorized. Data display systems, data processing units, computers, and aircraft guidance-control assemblies are also good examples of electronic equipments that use semiconductor devices. All of the specific applications of semiconductor devices would make a long impressive list. The fact is, semiconductors are being used extensively in commercial products, industry, and the military.

SEMICONDUCTOR COMPETITION

It should not be difficult to conclude, from what you already know, that semiconductor devices can and do perform all the conventional functions of rectification, amplification, oscillation, timing, switching, and sensing. Simply stated, these devices perform the same basic functions as the electron tube; but they perform more efficiently, economically, and for a longer period of time. Therefore, it should be no surprise to you to see these devices used in place of electron tubes. Keeping this in mind, we see that it is only natural and logical to compare semiconductor devices with electron tubes.

Physically, semiconductor devices are much smaller than tubes. You can see in figure 1-1 that the difference is quite evident. This illustration shows some commonly used tube sizes alongside semiconductor devices of similar capabilities. The reduction in size can be as great as 100:1 by weight and 1000:1 by volume. It is easy to see that size reduction favors the semiconductor device. Therefore, whenever miniaturization is required or is convenient, transistors are favored over tubes. Bear in mind, however, that the extent of practical size reduction is a big factor; many things must be considered. Miniature electron tubes, for example, may be preferred in certain applications to transistors, thus keeping size reduction to a competitive area.

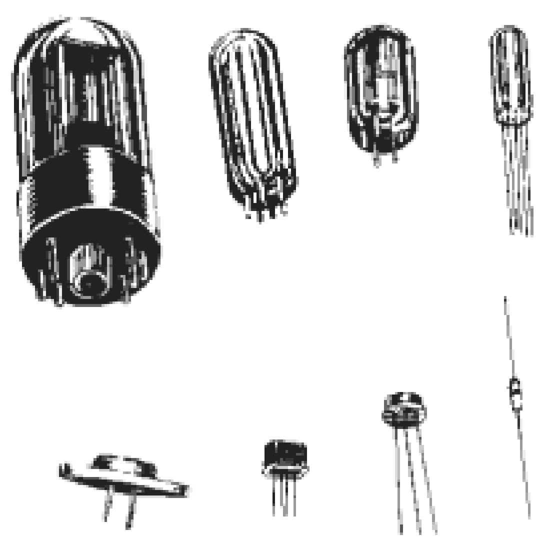


Figure 1-1.—Size comparisons of electron tubes and semiconductors.

Power is also a two-sided story. For low-power applications, where efficiency is a significant factor, semiconductors have a decided advantage. This is true mainly because semiconductor devices perform very well with an extremely small amount of power; in addition, they require no filaments or heaters as in the case of the electron tube. For example, a computer operating with over 4000 solid-state devices may require no more than 20 watts of power. However, the same number of tubes would require several kilowatts of power.

For high-power applications, it is a different story — tubes have the upper hand. The high-power tube has no equivalent in any semiconductor device. This is because a tube can be designed to operate with over a thousand volts applied to its plate whereas the maximum allowable voltage for a transistor is limited to about 200 volts (usually 50 volts or less). A tube can also handle thousands of watts of power. The maximum power output for transistor generally ranges from 30 milliwatts to slightly over 100 watts.

When it comes to ruggedness and life expectancy, the tube is still in competition. Design and functional requirements usually dictate the choice of device. However, semiconductor devices are rugged and long-lived. They can be constructed to withstand extreme vibration and mechanical shock. They have been known to withstand impacts that would completely shatter an ordinary electron tube. Although some specially designed tubes render extensive service, the life expectancy of transistors is better than three to four times that of ordinary electronic tubes. There is no known failure mechanism (such as an open filament in a tube) to limit the semiconductor's life. However, semiconductor devices do have some limitations. They are usually affected more by temperature, humidity, and radiation than tubes are.

Q3. Name three of the largest users of semiconductor devices.

Q4. State one requirement of an electron tube, which does not exist for semiconductors, that makes the tube less efficient than the semiconductor.

SEMICONDUCTOR THEORY

To understand why solid-state devices function as they do, we will have to examine closely the composition and nature of semiconductors. This entails theory that is fundamental to the study of solid-state devices.

Rather than beginning with theory, let's first become reacquainted with some of the basic information you studied earlier concerning matter and energy (NEETS, Module 1).

ATOMIC STRUCTURE

The universe, as we know it today, is divided into two parts: matter and energy. Matter, which is our main concern at this time, is anything that occupies space and has weight. Rocks, water, air, automobiles, clothing, and even our own bodies are good examples of matter. From this, we can conclude that matter may be found in any one of three states: SOLIDS, LIQUIDS, and GASES. All matter is composed of either an element or combination of elements. As you know, an element is a substance that cannot be reduced to a simpler form by chemical means. Examples of elements with which you are in contact everyday are iron, gold, silver, copper, and oxygen. At present, there are over 100 known elements of which all matter is composed.

As we work our way down the size scale, we come to the atom, the smallest particle into which an element can be broken down and still retain all its original properties. The atoms of one element, however, differ from the atoms of all other elements. Since there are over 100 known elements, there must be over 100 different atoms, or a different atom for each element.

Now let us consider more than one element at a time. This brings us to the term "compound." A compound is a chemical combination of two or more elements. Water, table salt, ethyl alcohol, and ammonia are all examples of compounds. The smallest part of a compound, which has all the characteristics of the compound, is the molecule. Each molecule contains some of the atoms of each of the elements forming the compound.

Consider sugar, for example. Sugar in general terms is matter, since it occupies space and has weight. It is also a compound because it consists of two or more elements. Take a lump of sugar and crush it into small particles; each of the particles still retains its original identifying properties of sugar. The only thing that changed was the physical size of the sugar. If we continue this subdividing process by grinding the sugar into a fine power, the results are the same. Even dissolving sugar in water does not change its identifying properties, in spite of the fact that the particles of sugar are now too small to be seen even with a microscope. Eventually, we end up with a quantity of sugar that cannot be further divided without its ceasing to be sugar. This quantity is known as a molecule of sugar. If the molecule is further divided, it is found to consist of three simpler kinds of matter: carbon, hydrogen, and oxygen. These simpler forms are called elements. Therefore, since elements consist of atoms, then a molecule of sugar is made up of atoms of carbon, hydrogen, and oxygen.

As we investigate the atom, we find that it is basically composed of electrons, protons, and neutrons. Furthermore, the electrons, protons, and neutrons of one element are identical to those of any other element. There are different kinds of elements because the number and the arrangement of electrons and protons are different for each element.

The electron carries a small <u>negative charge</u> of electricity. The proton carries a <u>positive charge</u> of electricity equal and opposite to the charge of the electron. Scientists have measured the mass and size of the electron and proton, and they know how much charge each possesses. Both the electron and proton have the same quantity of charge, although the mass of the proton is approximately 1,827 times that of the electron. In some atoms there exists a neutral particle called a neutron. The neutron has a mass approximately equal to that of a proton, but it has <u>no electrical charge</u>.

According to theory, the electrons, protons, and neutrons of the atoms are thought to be arranged in a manner similar to a miniature solar system. Notice the helium atom in figure 1-2. Two protons and two neutrons form the heavy nucleus with a positive charge around which two very light electrons revolve. The path each electron takes around the nucleus is called an orbit. The electrons are continuously being acted upon in their orbits by the force of attraction of the nucleus. To maintain an orbit around the nucleus, the electrons travel at a speed that produces a counterforce equal to the attraction force of the nucleus. Just as energy is required to move a space vehicle away from the earth, energy is also required to move an electron away from the nucleus. Like a space vehicle, the electron is said to be at a higher energy level when it travels a larger orbit. Scientific experiments have shown that the electron requires a certain amount of energy to stay in orbit. This quantity is called the electron's energy level. By virtue of just its motion alone, the electron contains kinetic energy. Because of its position, it also contains potential energy. The total energy contained by an electron (kinetic energy plus potential energy) is the main factor that determines the radius of the electron's orbit. For an electron to remain in this orbit, it must neither gain nor lose energy.

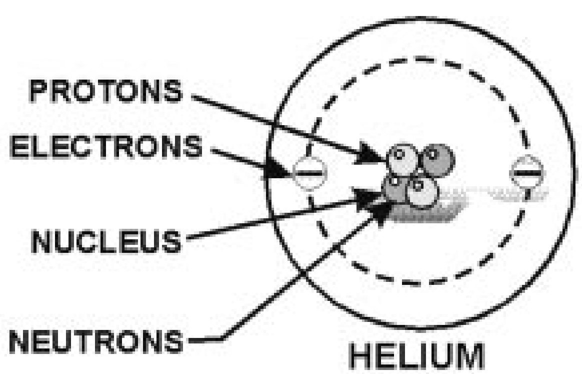


Figure 1-2.—The composition of a simple helium atom.

The orbiting electrons do not follow random paths, instead they are confined to definite energy levels. Visualize these levels as shells with each successive shell being spaced a greater distance from the nucleus. The shells, and the number of electrons required to fill them, may be predicted by using Pauli's exclusion principle. Simply stated, this principle specifies that each shell will contain a maximum of $2n^2$ electrons, where n corresponds to the shell number starting with the one closest to the nucleus. By this principle, the second shell, for example, would contain $2(2)^2$ or 8 electrons when full.

In addition to being numbered, the shells are also given letter designations starting with the shell closest to the nucleus and progressing outward as shown in figure 1-3. The shells are considered to be full, or complete, when they contain the following quantities of electrons: 2 in the K(1st) shell, 8 in the L(2nd) shell, 18 in the M(3rd) shell, and so on, in accordance with the exclusion principle. Each of these shells is a major shell and can be divided into subshells, of which there are four, labeled s, p, d, and f. Like the major shells, the subshells are also limited as to the number of electrons they contain. Thus, the "s" subshell is complete when it contains 2 electrons, the "p" subshell when it contains 6, the "d" subshell when it contains 10, and the "f" subshell when it contains 14 electrons.

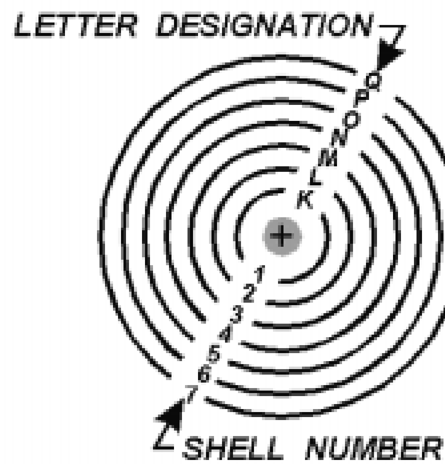


Figure 1-3.—Shell designation.

Inasmuch as the K shell can contain no more than 2 electrons, it must have only one subshell, the s subshell. The M shell is composed of three subshells: s, p, and d. If the electrons in the s, p, and d subshells are added together, their total is found to be 18, the exact number required to fill the M shell. Notice the electron configuration of copper illustrated in figure 1-4. The copper atom contains 29 electrons, which completely fill the first three shells and subshells, leaving one electron in the "s" subshell of the N shell. A list of all the other known elements, with the number of electrons in each atom, is contained in the PERIODIC TABLE OF ELEMENTS. The periodic table of elements is included in appendix 2.

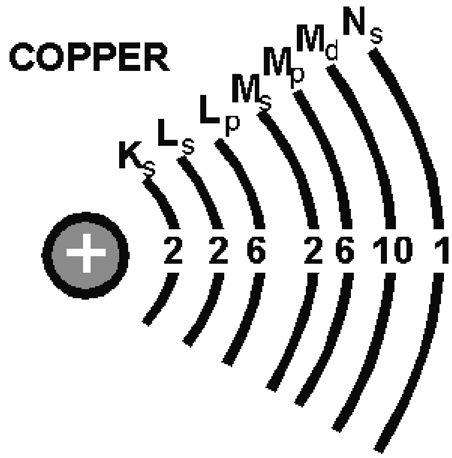


Figure 1-4.—Copper atom.

Valence is an atom's ability to combine with other atoms. The number of electrons in the outermost shell of an atom determines its valence. For this reason, the outer shell of an atom is called VALENCE SHELL, and the electrons contained in this shell are called VALENCE ELECTRONS. The valence of an atom determines its ability to gain or lose an electron, which in turn determines the chemical and electrical properties of the atom. An atom that is lacking only one or two electrons from its outer shell will easily gain electrons to complete its shell, but a large amount of energy is required to free any of its electrons. An atom having a relatively small number of electrons in its outer shell in comparison to the number of electrons required to fill the shell will easily lose these valence electrons. The valence shell always refers to the outermost shell.

Q5. Define matter and list its three different states.

Q6. What is the smallest particle into which an element can be broken down and still retain all its original properties?

Q7. What are the three particles that comprise an atom and state the type of charge they hold?

Q8. What is the outer shell of an atom called?

ENERGY BANDS

Now that you have become reacquainted with matter and energy, we will continue our discussion with electron behavior.

As stated earlier, orbiting electrons contain energy and are confined to definite energy levels. The various shells in an atom represent these levels. Therefore, to move an electron from a lower shell to a higher shell a certain amount of energy is required. This energy can be in the form of electric fields, heat, light, and even bombardment by other particles. Failure to provide enough energy to the electron, even if the energy supplied is just short of the required amount, will cause it to remain at its present energy level. Supplying more energy than is needed will only cause the electron to move to the next higher shell and the remaining energy will be wasted. In simple terms, energy is required in definite units to move electrons from one shell to the next higher shell. These units are called QUANTA (for example 1, 2, or 3 quanta).

Electrons can also lose energy as well as receive it. When an electron loses energy, it moves to a lower shell. The lost energy, in some cases, appears as heat.

If a sufficient amount of energy is absorbed by an electron, it is possible for that electron to be completely removed from the influence of the atom. This is called IONIZATION. When an atom loses electrons or gains electrons in this process of electron exchange, it is said to be ionized. For ionization to take place, there must be a transfer of energy that results in a change in the internal energy of the atom. An atom having more than its normal amount of electrons acquires a negative charge, and is called a NEGATIVE ION. The atom that gives up some of its normal electrons is left with fewer negative charges than positive charges and is called a POSITIVE ION. Thus, we can define ionization as the process by which an atom loses or gains electrons.

Up to this point in our discussion, we have spoken only of isolated atoms. When atoms are spaced far enough apart, as in a gas, they have very little influence upon each other, and are very much like lone atoms. But atoms within a solid have a marked effect upon each other. The forces that bind these atoms together greatly modify the behavior of the other electrons. One consequence of this close proximity of atoms is to cause the individual energy levels of an atom to break up and form bands of energy. Discrete (separate and complete) energy levels still exist within these energy bands, but there are many more energy levels than there were with the isolated atom. In some cases, energy levels will have disappeared. Figure 1-5 shows the difference in the energy arrangement between an isolated atom and the atom in a solid. Notice that the isolated atom (such as in gas) has energy levels, whereas the atom in a solid has energy levels grouped into ENERGY BANDS.

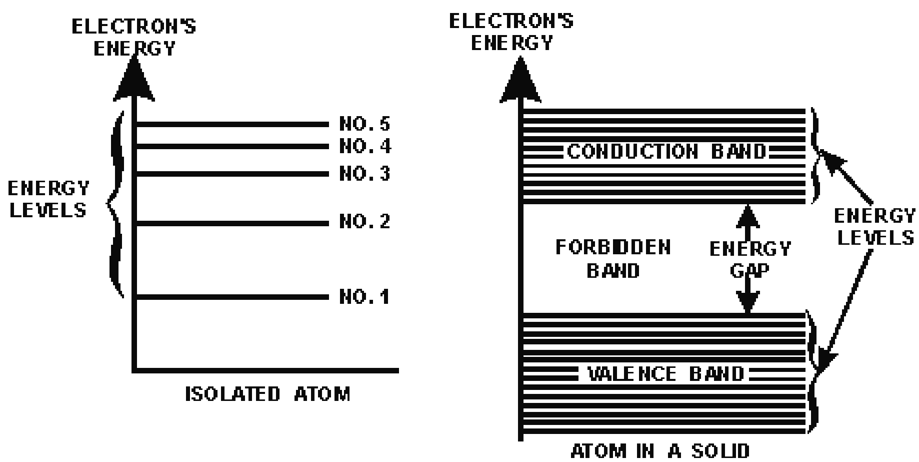


Figure 1-5.—The energy arrangement in atoms.

The upper band in the solid lines in figure 1-5 is called the CONDUCTION BAND because electrons in this band are easily removed by the application of external electric fields. Materials that have a large number of electrons in the conduction band act as good conductors of electricity.

Below the conduction band is the FORBIDDEN BAND or energy gap. Electrons are never found in this band, but may travel back and forth through it, provided they do not come to rest in the band.

The last band or VALENCE BAND is composed of a series of energy levels containing valence electrons. Electrons in this band are more tightly bound to the individual atom than the electrons in the conduction band. However, the electrons in the valence band can still be moved to the conduction band with the application of energy, usually thermal energy. There are more bands below the valence band, but they are not important to the understanding of semiconductor theory and will not be discussed.

The concept of energy bands is particularly important in classifying materials as conductors, semiconductors, and insulators. An electron can exist in either of two energy bands, the conduction band or the valence band. All that is necessary to move an electron from the valence band to the conduction band so it can be used for electric current, is enough energy to carry the electron through the forbidden band. The width of the forbidden band or the separation between the conduction and valence bands determines whether a substance is an insulator, semiconductor, or conductor. Figure 1-6 uses energy level diagrams to show the difference between insulators, semiconductors, and conductors.

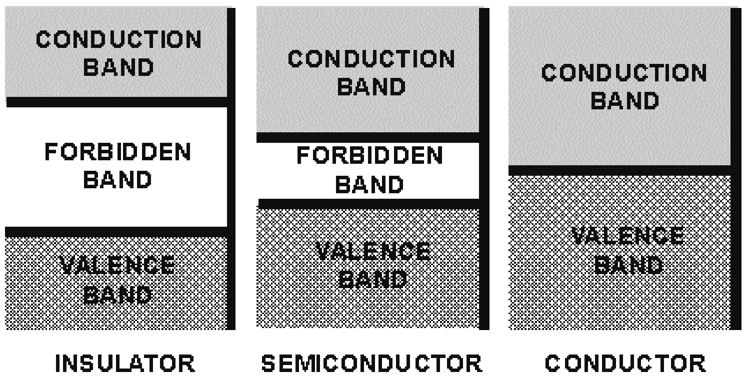


Figure 1-6.—Energy level diagram.

The energy diagram for the insulator shows the insulator with a very wide energy gap. The wider this gap, the greater the amount of energy required to move the electron from the valence band to the conduction band. Therefore, an insulator requires a large amount of energy to obtain a small amount of current. The insulator "insulates" because of the wide forbidden band or energy gap.

The semiconductor, on the other hand, has a smaller forbidden band and requires less energy to move an electron from the valence band to the conduction band. Therefore, for a certain amount of applied voltage, more current will flow in the semiconductor than in the insulator.

The last energy level diagram in figure 1-6 is that of a conductor. Notice, there is no forbidden band or energy gap and the valence and conduction bands overlap. With no energy gap, it takes a small amount of energy to move electrons into the conduction band; consequently, conductors pass electrons very easily.

Q9. *What term is used to describe the definite discrete amounts of energy required to move an electron from a low shell to a higher shell?*

Q10. *What is a negative ion?*

Q11. *What is the main difference in the energy arrangement between an isolated atom and the atom in a solid?*

Q12. *What determines, in terms of energy bands, whether a substance is a good insulator, semiconductor, or conductor?*

COVALENT BONDING

The chemical activity of an atom is determined by the number of electrons in its valence shell. When the valence shell is complete, the atom is stable and shows little tendency to combine with other atoms to form solids. Only atoms that possess eight valence electrons have a complete outer shell. These atoms are referred to as inert or inactive atoms. However, if the valence shell of an atom lacks the required number of electrons to complete the shell, then the activity of the atom increases.

Silicon and germanium, for example, are the most frequently used semiconductors. Both are quite similar in their structure and chemical behavior. Each has four electrons in the valence shell. Consider just silicon. Since it has fewer than the required number of eight electrons needed in the outer shell, its atoms will unite with other atoms until eight electrons are shared. This gives each atom a total of eight electrons in its valence shell; four of its own and four that it borrowed from the surrounding atoms. The sharing of valence electrons between two or more atoms produces a COVALENT BOND between the atoms. It is this bond that holds the atoms together in an orderly structure called a CRYSTAL. A crystal is just another name for a solid whose atoms or molecules are arranged in a three-dimensional geometrical pattern commonly referred to as a lattice. Figure 1-7 shows a typical crystal structure. Each sphere in the figure represents the nucleus of an atom, and the arms that join the atoms and support the structure are the covalent bonds.

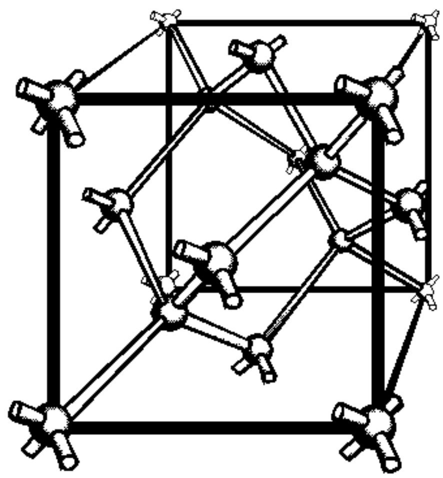


Figure 1-7.—A typical crystal structure.

As a result of this sharing process, the valence electrons are held tightly together. This can best be illustrated by the two-dimensional view of the silicon lattice in figure 1-8. The circles in the figure represent the nuclei of the atoms. The +4 in the circles is the net charge of the nucleus plus the inner shells (minus the valence shell). The short lines indicate valence electrons. Because every atom in this pattern is bonded to four other atoms, the electrons are not free to move within the crystal. As a result of this bonding, pure silicon and germanium are poor conductors of electricity. The reason they are not insulators but semiconductors is that with the proper application of heat or electrical pressure, electrons can be caused to break free of their bonds and move into the conduction band. Once in this band, they wander aimlessly through the crystal.

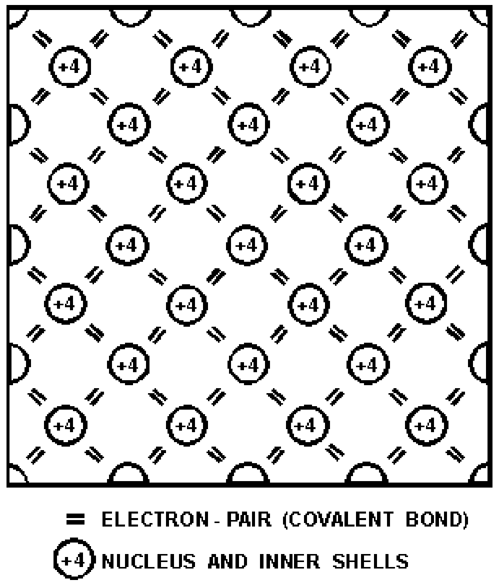


Figure 1-8.—A two-dimensional view of a silicon cubic lattice.

Q13. What determines the chemical activity of an atom?

Q14. What is the term used to describe the sharing of valence electrons between two or more atoms?

CONDUCTION PROCESS

As stated earlier, energy can be added to electrons by applying heat. When enough energy is absorbed by the valence electrons, it is possible for them to break some of their covalent bonds. Once the bonds are broken, the electrons move to the conduction band where they are capable of supporting electric current. When a voltage is applied to a crystal containing these conduction band electrons, the electrons move through the crystal toward the applied voltage. This movement of electrons in a semiconductor is referred to as electron current flow.

There is still another type of current in a pure semiconductor. This current occurs when a covalent bond is broken and a vacancy is left in the atom by the missing valence electron. This vacancy is commonly referred to as a "hole." The hole is considered to have a positive charge because its atom is deficient by one electron, which causes the protons to outnumber the electrons. As a result of this hole, a chain reaction begins when a nearby electron breaks its own covalent bond to fill the hole, leaving another hole. Then another electron breaks its bond to fill the previous hole, leaving still another hole. Each time an electron in this process fills a hole, it enters into a covalent bond. Even though an electron has moved from one covalent bond to another, the most important thing to remember is that the hole is also moving. Therefore, since this process of conduction resembles the movement of holes rather than electrons, it is termed hole flow (short for hole current flow or conduction by holes). Hole flow is very similar to electron flow except that the holes move toward a negative potential and in an opposite direction to that of the electron. Since hole flow results from the breaking of covalent bonds, which are at the valence band level, the electrons associated with this type of conduction contain only valence band energy and must remain in the valence band. However, the electrons associated with electron flow have conduction band energy and can, therefore, move throughout the crystal. A good analogy of hole flow is the movement of a hole through a tube filled with balls (figure 1-9).

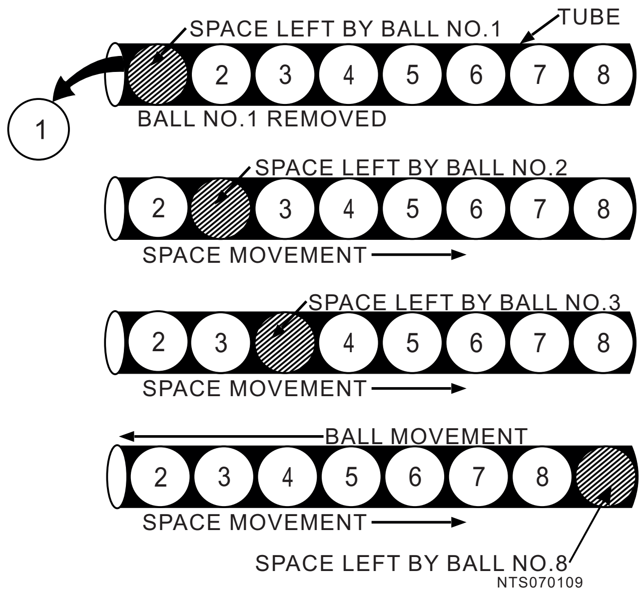


Figure 1-9.—Analogy of hole flow.

When ball number 1 is removed from the tube, a hole is left. This hole is then filled by ball number 2, which leaves still another hole. Ball number 3 then moves into the hole left by ball number 2. This causes still another hole to appear where ball 3 was. Notice the holes are moving to the right side of the tube. This action continues until all the balls have moved one space to the left in which time the hole moved eight spaces to the right and came to rest at the right-hand end of the tube.

In the theory just described, two current carriers were created by the breaking of covalent bonds: the negative electron and the positive hole. These carriers are referred to as electron-hole pairs. Since the semiconductor we have been discussing contains no impurities, the number of holes in the electron-hole pairs is always equal to the number of conduction electrons. Another way of describing this condition where no impurities exist is by saying the semiconductor is INTRINSIC. The term *intrinsic* is also used to distinguish the pure semiconductor that we have been working with from one containing impurities.

Q15. Name the two types of current flow in a semiconductor.

Q16. What is the name given to a piece of pure semiconductor material that has an equal number of electrons and holes?

DOPING PROCESS

The pure semiconductor mentioned earlier is basically neutral. It contains no free electrons in its conduction bands. Even with the application of thermal energy, only a few covalent bonds are broken, yielding a relatively small current flow. A much more efficient method of increasing current flow in semiconductors is by adding very small amounts of selected additives to them, generally no more than a few parts per million. These additives are called impurities and the process of adding them to crystals is referred to as DOPING. The purpose of semiconductor doping is to increase the number of free charges that can be moved by an external applied voltage. When an impurity increases the number of free electrons, the doped semiconductor is NEGATIVE or N TYPE, and the impurity that is added is known as an N-type impurity. However, an impurity that reduces the number of free electrons, causing more holes, creates a POSITIVE or P-TYPE semiconductor, and the impurity that was added to it is known as a P-type impurity. Semiconductors which are doped in this manner — either with N- or P-type impurities — are referred to as EXTRINSIC semiconductors.

N-Type Semiconductor

The N-type impurity loses its extra valence electron easily when added to a semiconductor material, and in so doing, increases the conductivity of the material by contributing a free electron. This type of impurity has 5 valence electrons and is called a PENTAVALENT impurity. Arsenic, antimony, bismuth, and phosphorous are pentavalent impurities. Because these materials give or donate one electron to the doped material, they are also called DONOR impurities.

When a pentavalent (donor) impurity, like arsenic, is added to germanium, it will form covalent bonds with the germanium atoms. Figure 1-10 illustrates this by showing an arsenic atom (AS) in a germanium (GE) lattice structure. Notice the arsenic atom in the center of the lattice. It has 5 valence electrons in its outer shell but uses only 4 of them to form covalent bonds with the germanium atoms, leaving 1 electron relatively free in the crystal structure. Pure germanium may be converted into an N-type semiconductor by "doping" it with any donor impurity having 5 valence electrons in its outer shell. Since this type of semiconductor (N-type) has a surplus of electrons, the electrons are considered MAJORITY carriers, while the holes, being few in number, are the MINORITY carriers.

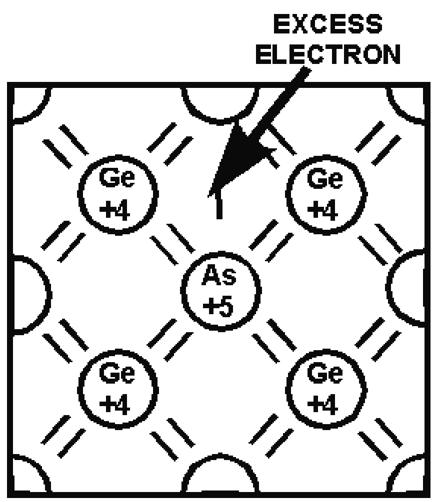


Figure 1-10.—Germanium crystal doped with arsenic.

P-Type Semiconductor

The second type of impurity, when added to a semiconductor material, tends to compensate for its deficiency of 1 valence electron by acquiring an electron from its neighbor. Impurities of this type have only 3 valence electrons and are called TRIVALENT impurities. Aluminum, indium, gallium, and boron are trivalent impurities. Because these materials accept 1 electron from the doped material, they are also called ACCEPTOR impurities.

A trivalent (acceptor) impurity element can also be used to dope germanium. In this case, the impurity is 1 electron short of the required amount of electrons needed to establish covalent bonds with 4 neighboring atoms. Thus, in a single covalent bond, there will be only 1 electron instead of 2. This arrangement leaves a hole in that covalent bond. Figure 1-11 illustrates this theory by showing what happens when germanium is doped with an indium (In) atom. Notice, the indium atom in the figure is 1 electron short of the required amount of electrons needed to form covalent bonds with 4 neighboring atoms and, therefore, creates a hole in the structure. Gallium and boron, which are also trivalent impurities, exhibit these same characteristics when added to germanium. The holes can only be present in this type semiconductor when a trivalent impurity is used. Note that a hole carrier is not created by the removal of an electron from a neutral atom, but is created when a trivalent impurity enters into covalent bonds with a tetravalent (4 valence electrons) crystal structure. The holes in this type of semiconductor (P-type) are considered the MAJORITY carriers since they are present in the material in the greatest quantity. The electrons, on the other hand, are the MINORITY carriers.

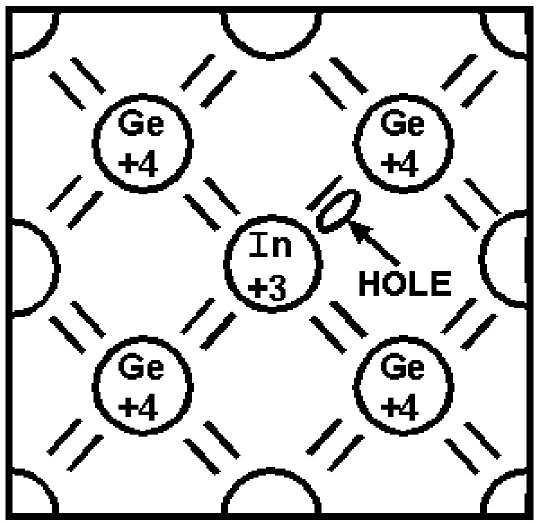


Figure 1-11.—Germanium crystal doped with indium.

Q17. What is the name given to a doped germanium crystal with an excess of free holes?

Q18. What are the majority carriers in an N-type semiconductor?

SEMICONDUCTOR DIODE

If we join a section of N-type semiconductor material with a similar section of P-type semiconductor material, we obtain a device known as a PN JUNCTION. (The area where the N and P regions meet is appropriately called the junction.) The usual characteristics of this device make it extremely useful in electronics as a diode rectifier. The diode rectifier or PN junction diode performs the same function as its counterpart in electron tubes but in a different way. The diode is nothing more than a two-element semiconductor device that makes use of the rectifying properties of a PN junction to convert alternating current into direct current by permitting current flow in only one direction. The schematic symbol of a PN junction diode is shown in figure 1-12. The vertical bar represents the cathode (N-type material) since it is the source of electrons and the arrow represents the anode. (P-type material) since it is the destination of the electrons. The label "CR1" is an alphanumerical code used to identify the diode. In this figure, we have only one diode so it is labeled CR1 (crystal rectifier number one). If there were four diodes shown in the diagram, the last diode would be labeled CR4. The heavy dark line shows electron flow. Notice it is against the arrow. For further clarification, a pictorial diagram of a PN junction and an actual semiconductor (one of many types) are also illustrated.

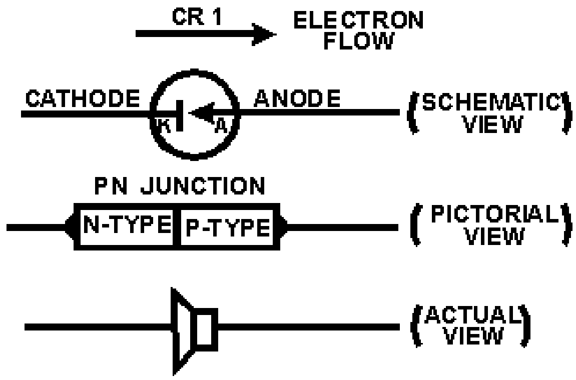


Figure 1-12.—The PN junction diode.

CONSTRUCTION

Merely pressing together a section of P material and a section of N material, however, is not sufficient to produce a rectifying junction. The semiconductor should be in one piece to form a proper PN junction, but divided into a P-type impurity region and an N-type impurity region. This can be done in various ways. One way is to mix P-type and N-type impurities into a single crystal during the manufacturing process. By so doing, a P-region is grown over part of a semiconductor's length and N-region is grown over the other part. This is called a GROWN junction and is illustrated in view A of figure 1-13. Another way to produce a PN junction is to melt one type of impurity into a semiconductor of the opposite type impurity. For example, a pellet of acceptor impurity is placed on a wafer of N-type germanium and heated. Under controlled temperature conditions, the acceptor impurity fuses into the wafer to form a P-region within it, as shown in view B of figure 1-13. This type of junction is known as an ALLOY or FUSED-ALLOY junction, and is one of the most commonly used junctions. In figure 1-14, a POINT-CONTACT type of construction is shown. It consists of a fine metal wire, called a cat whisker, that makes contact with a small area on the surface of an N-type semiconductor as shown in view A of the figure. The PN union is formed in this process by momentarily applying a high-surge current to the wire and the N-type semiconductor. The heat generated by this current converts the material nearest the point of contact to a P-type material (view B).

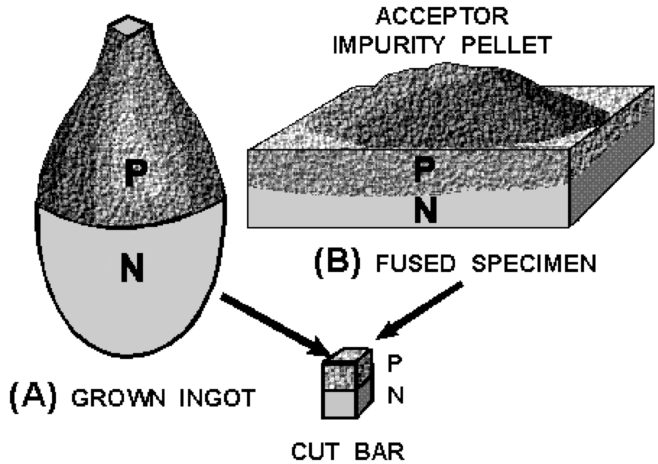


Figure 1-13.—Grown and fused PN junctions from which bars are cut.

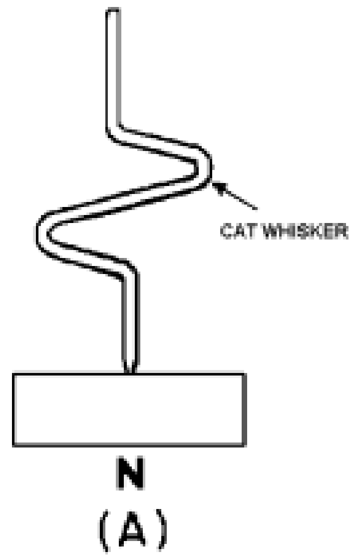


Figure 1-14A.—The point-contact type of diode construction.

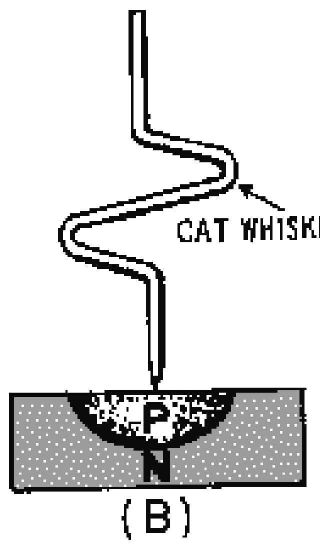


Figure 1-14B.—The point-contact type of diode construction.

Still another process is to heat a section of semiconductor material to near melting and then diffuse impurity atoms into a surface layer. Regardless of the process, the objective is to have a perfect bond everywhere along the union (interface) between P and N materials. Proper contact along the union is important because, as we will see later, the union (junction or interface) is the rectifying agent in the diode.

Q19. What is the purpose of a PN junction diode?

Q20. In reference to the schematic symbol for a diode, do electrons flow toward or away from the arrow?

Q21. What type of PN diode is formed by using a fine metal wire and a section of N-type semiconductor material?

PN JUNCTION OPERATION

Now that you are familiar with P- and N-type materials, how these materials are joined together to form a diode, and the function of the diode, let us continue our discussion with the operation of the PN junction. But before we can understand how the PN junction works, we must first consider current flow in the materials that make up the junction and what happens initially within the junction when these two materials are joined together.

Current Flow in the N-Type Material

Conduction in the N-type semiconductor, or crystal, is similar to conduction in a copper wire. That is, with voltage applied across the material, electrons will move through the crystal just as current would flow in a copper wire. This is shown in figure 1-15. The positive potential of the battery will attract the free electrons in the crystal. These electrons will leave the crystal and flow into the positive terminal of the battery. As an electron leaves the crystal, an electron from the negative terminal of the battery will enter the crystal, thus completing the current path. Therefore, the majority current carriers in the N-type material (electrons) are repelled by the negative side of the battery and move through the crystal toward the positive side of the battery.

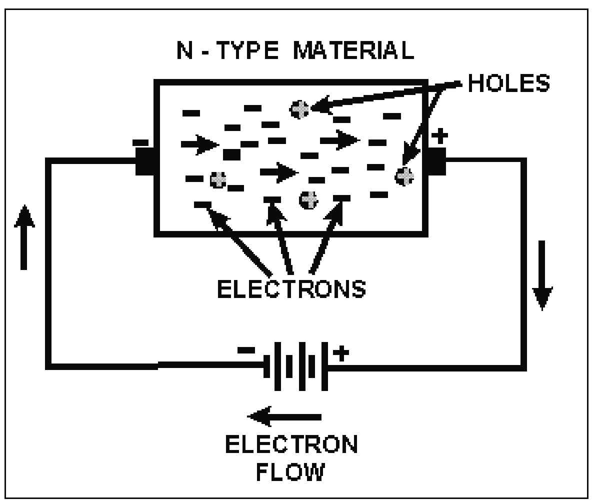


Figure 1-15.—Current flow In the N-type material.

Current Flow in the P-Type Material

Current flow through the P-type material is illustrated in figure 1-16. Conduction in the P material is by positive holes, instead of negative electrons. A hole moves from the positive terminal of the P material to the negative terminal. Electrons from the external circuit enter the negative terminal of the material and fill holes in the vicinity of this terminal. At the positive terminal, electrons are removed from the covalent bonds, thus creating new holes. This process continues as the steady stream of holes (hole current) moves toward the negative terminal.

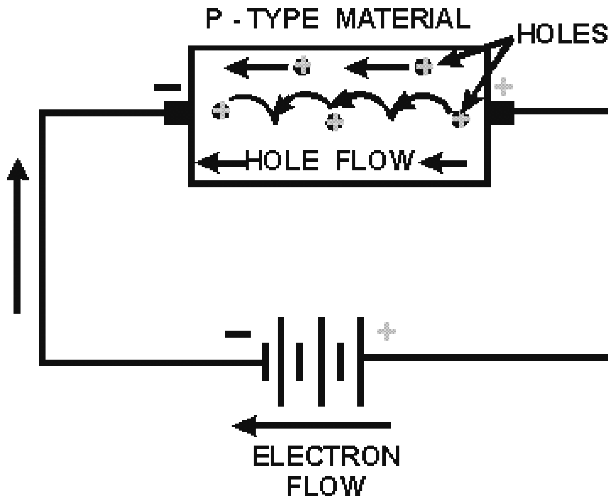


Figure 1-16.—Current flow In the P-type material.

Notice in both N-type and P-type materials, current flow in the external circuit consists of electrons moving out of the negative terminal of the battery and into the positive terminal of the battery. Hole flow, on the other hand, only exists within the material itself.

Q22. What are the majority carriers in a P-type semiconductor?

Q23. Conduction in which type of semiconductor material is similar to conduction in a copper wire?

Junction Barrier

Although the N-type material has an excess of free electrons, it is still electrically neutral. This is because the donor atoms in the N material were left with positive charges after free electrons became available by covalent bonding (the protons outnumbered the electrons). Therefore, for every free electron in the N material, there is a corresponding positively charge atom to balance it. The end result is that the N material has an overall charge of zero.

By the same reasoning, the P-type material is also electrically neutral because the excess of holes in this material is exactly balanced by the number of electrons. Keep in mind that the holes and electrons are still free to move in the material because they are only loosely bound to their parent atoms.

It would seem that if we joined the N and P materials together by one of the processes mentioned earlier, all the holes and electrons would pair up. On the contrary, this does not happen. Instead the electrons in the N material diffuse (move or spread out) across the junction into the P material and fill some of the holes. At the same time, the holes in the P material diffuse across the junction into the N material and are filled by N material electrons. This process, called JUNCTION RECOMBINATION, reduces the number of free electrons and holes in the vicinity of the junction. Because there is a depletion, or lack of free electrons and holes in this area, it is known as the DEPLETION REGION.

The loss of an electron from the N-type material created a positive ion in the N material, while the loss of a hole from the P material created a negative ion in that material. These ions are fixed in place in the crystal lattice structure and cannot move. Thus, they make up a layer of fixed charges on the two sides of the junction as shown in figure 1-17. On the N side of the junction, there is a layer of positively charged ions; on the P side of the junction, there is a layer of negatively charged ions. An electrostatic field, represented by a small battery in the figure, is established across the junction between the oppositely charged ions. The diffusion of electrons and holes across the junction will continue until the magnitude of the electrostatic field is increased to the point where the electrons and holes no longer have enough energy to overcome it, and are repelled by the negative and positive ions respectively. At this point equilibrium is established and, for all practical purposes, the movement of carriers across the junction ceases. For this reason, the electrostatic field created by the positive and negative ions in the depletion region is called a barrier.

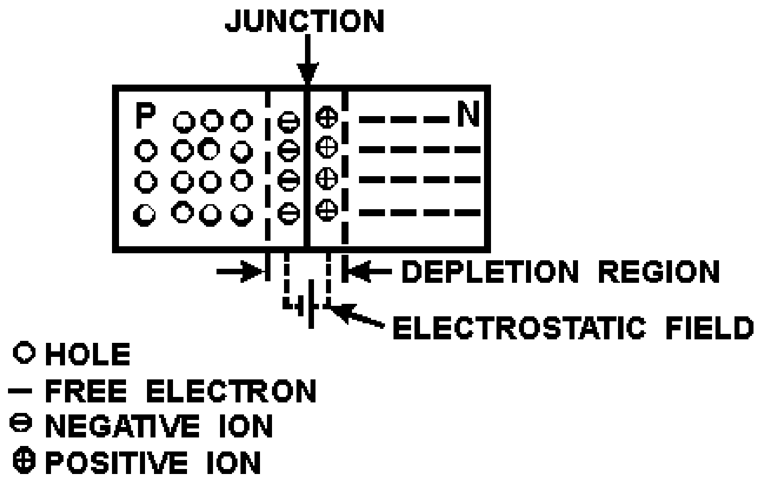


Figure 1-17.—The PN junction barrier formation.

The action just described occurs almost instantly when the junction is formed. Only the carriers in the immediate vicinity of the junction are affected. The carriers throughout the remainder of the N and P material are relatively undisturbed and remain in a balanced condition.

FORWARD BIAS.—An external voltage applied to a PN junction is call BIAS. If, for example, a battery is used to supply bias to a PN junction and is connected so that its voltage opposes the junction field, it will reduce the junction barrier and, therefore, aid current flow through the junction. This type of bias is known as forward bias, and it causes the junction to offer only minimum resistance to the flow of current.

Forward bias is illustrated in figure 1-18. Notice the positive terminal of the bias battery is connected to the P-type material and the negative terminal of the battery is connected to the N-type material. The positive potential repels holes toward the junction where they neutralize some of the negative ions. At the same time the negative potential repels electrons toward the junction where they neutralize some of the positive ions. Since ions on both sides of the barrier are being neutralized, the width of the barrier decreases. Thus, the effect of the battery voltage in the forward-bias direction is to reduce the barrier potential across the junction and to allow majority carriers to cross the junction. Current flow in the forward-biased PN junction is relatively simple. An electron leaves the negative terminal of the battery and moves to the terminal of the N-type material. It enters the N material, where it is the majority carrier and moves to the edge of the junction barrier. Because of forward bias, the barrier offers less opposition to the electron and it will pass through the depletion region into the P-type material. The electron loses energy in overcoming the opposition of the junction barrier, and upon entering the P material, combines with a hole. The hole was produced when an electron was extracted from the P material by the positive potential of the battery. The created hole moves through the P material toward the junction where it combines with an electron.

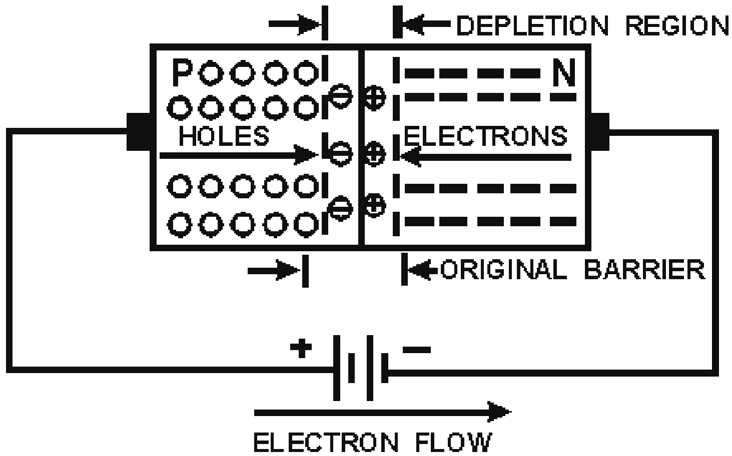


Figure 1-18.—Forward-biased PN junction.

It is important to remember that in the forward biased condition, conduction is by MAJORITY current carriers (holes in the P-type material and electrons in the N-type material). Increasing the battery voltage will increase the number of majority carriers arriving at the junction and will therefore increase the current flow. If the battery voltage is increased to the point where the barrier is greatly reduced, a heavy current will flow and the junction may be damaged from the resulting heat.

REVERSE BIAS.—If the battery mentioned earlier is connected across the junction so that its voltage aids the junction, it will increase the junction barrier and thereby offer a high resistance to the current flow through the junction. This type of bias is known as reverse bias.

To reverse bias a junction diode, the negative battery terminal is connected to the P-type material, and the positive battery terminal to the N-type material as shown in figure 1-19. The negative potential attracts the holes away from the edge of the junction barrier on the P side, while the positive potential attracts the electrons away from the edge of the barrier on the N side. This action increases the barrier width because there are more negative ions on the P side of the junction, and more positive ions on the N side of the junction. Notice in the figure the width of the barrier has increased. This increase in the number of ions prevents current flow across the junction by majority carriers. However, the current flow across the barrier is not quite zero because of the minority carriers crossing the junction. As you recall, when the crystal is subjected to an external source of energy (light, heat, etc.), electron-hole pairs are generated. The electron-hole pairs produce minority current carriers. There are minority current carriers in both regions: holes in the N material and electrons in the P material. With reverse bias, the electrons in the P-type material are repelled toward the junction by the negative terminal of the battery. As the electron moves across the junction, it will neutralize a positive ion in the N-type material. Similarly, the holes in the N-type material will be repelled by the positive terminal of the battery toward the junction. As the hole crosses the junction, it will neutralize a negative ion in the P-type material. This movement of minority carriers is called MINORITY CURRENT FLOW, because the holes and electrons involved come from the electron-hole pairs that are generated in the crystal lattice structure, and not from the addition of impurity atoms.

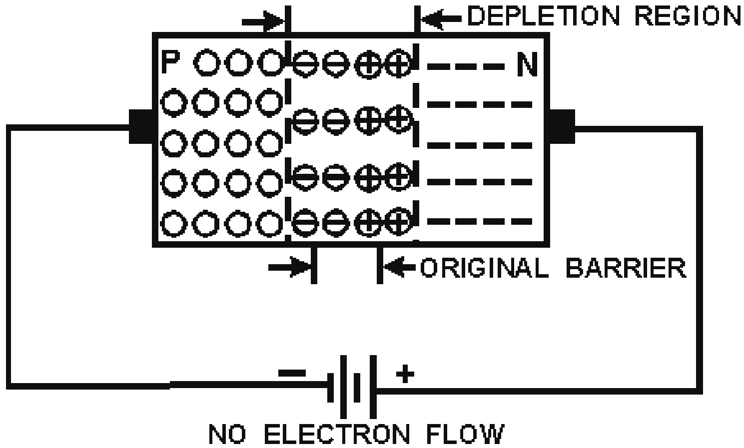


Figure 1-19.—Reverse-biased PN junction.

Therefore, when a PN junction is reverse biased, there will be no current flow because of majority carriers but a very small amount of current because of minority carriers crossing the junction. However, at normal operating temperatures, this small current may be neglected.

In summary, the most important point to remember about the PN junction diode is its ability to offer very little resistance to current flow in the forward-bias direction but maximum resistance to current flow when reverse biased. A good way of illustrating this point is by plotting a graph of the applied voltage versus the measured current. Figure 1-20 shows a plot of this voltage-current relationship (characteristic curve) for a typical PN junction diode.

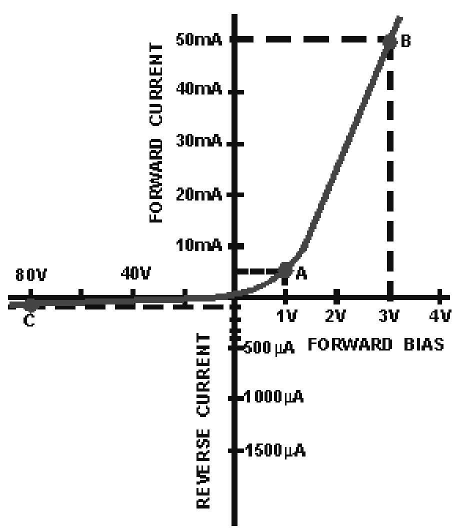


Figure 1-20.—PN junction diode characteristic curve.

To determine the resistance from the curve in this figure we can use Ohm's law:

$$R = \frac{E}{I}$$

For example at point A the forward-bias voltage is 1 volt and the forward-bias current is 5 milliamperes. This represents 200 ohms of resistance (1 volt/5mA = 200 ohms). However, at point B the voltage is 3 volts and the current is 50 milliamperes. This results in 60 ohms of resistance for the diode. Notice that when the forward-bias voltage was tripled (1 volt to 3 volts), the current increased 10 times (5mA to 50 mA). At the same time the forward-bias voltage increased, the resistance decreased from 200 ohms to 60 ohms. In other words, when forward bias increases, the junction barrier gets smaller and its resistance to current flow decreases.

On the other hand, the diode conducts very little when reverse biased. Notice at point C the reverse bias voltage is 80 volts and the current is only 100 microamperes. This results in 800 k ohms of resistance, which is considerably larger than the resistance of the junction with forward bias. Because of these unusual features, the PN junction diode is often used to convert alternating current into direct current (rectification).

Q24. What is the name of the area in a PN junction that has a shortage of electrons and holes?

Q25. In order to reverse bias in a PN junction, what terminal of a battery is connected to the P material?

Q26. What type of bias opposes the PN junction barrier?

PN JUNCTION APPLICATION

Until now, we have mentioned only one application for the diode-rectification, but there are many more applications that we have not yet discussed. Variations in doping agents, semiconductor materials, and manufacturing techniques have made it possible to produce diodes that can be used in many different applications. Examples of these types of diodes are signal diodes, rectifying diodes, Zener diodes (voltage protection diodes for power supplies), varactors (amplifying and switching diodes), and many more. Only applications for two of the most commonly used diodes, the signal diode and rectifier diode, will be presented in this chapter. The other diodes will be explained later on in this module.

Half-Wave Rectifier

One of the most important uses of a diode is rectification. The normal PN junction diode is well-suited for this purpose as it conducts very heavily when forward biased (low-resistance direction) and only slightly when reverse biased (high-resistance direction). If we place this diode in series with a source of ac power, the diode will be forward and reverse biased every cycle. Since in this situation current flows more easily in one direction than the other, rectification is accomplished. The simplest rectifier circuit is a half-wave rectifier (fig. 1-21 view A and view B) which consists of a diode, an ac power source, and a load resister.

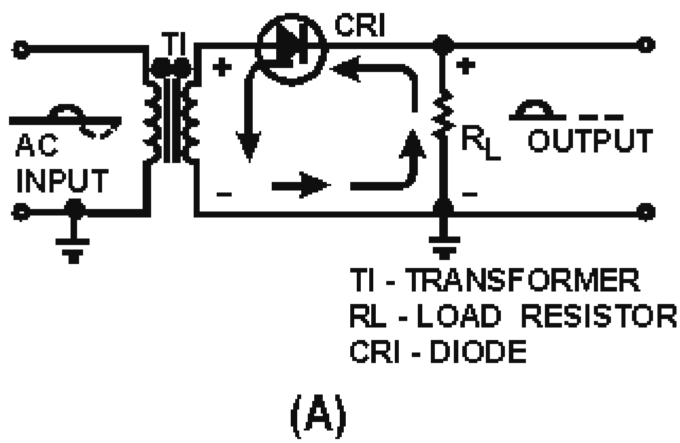


Figure 1-21A.—Simple half-wave rectifier.

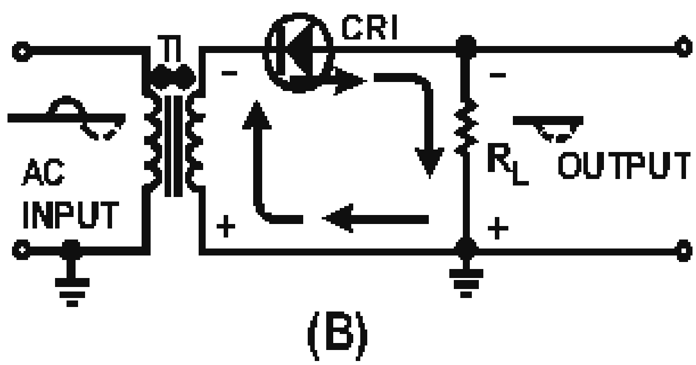


Figure 1-21B.—Simple half-wave rectifier.

The transformer (T1) in the figure provides the ac input to the circuit; the diode (CR1) provides the rectification; and the load resistor (R_L) serves two purposes: it limits the amount of current flow in the circuit to a safe level, and it also develops the output signal because of the current flow through it.

Before describing how this circuit operates, the definition of the word "load" as it applies to power supplies must be understood. Load is defined as <u>any device</u> that <u>draws current</u>. A device that draws little current is considered a light load, whereas a device that draws a large amount of current is a heavy load. Remember that when we speak of "load," we are speaking about the device that draws current from the power source. This device may be a simple resistor, or one or more complicated electronic circuits.

During the positive half-cycle of the input signal (solid line) in figure 1-21 view A, the top of the transformer is positive with respect to ground. The dots on the transformer indicate points of the same polarity. With this condition the diode is forward biased, the depletion region is narrow, the resistance of the diode is low, and current flows through the circuit in the direction of the solid lines. When this current flows through the load resistor, it develops a negative to positive voltage drop across it, which appears as a positive voltage at the output terminal.

When the ac input goes in a negative direction (fig. 1-21 view A), the top of the transformer becomes negative and the diode becomes reverse biased. With reverse bias applied to the diode, the depletion region increases, the resistance of the diode is high, and minimum current flows through the diode. For all practical purposes, there is no output developed across the load resistor during the negative alternation of the input signal as indicated by the broken lines in the figure. Although only one cycle of input is shown, it should be realized that the action described above continually repeats itself, as long as there is an input. Therefore, since only the positive half-cycles appear at the output this circuit converted the ac input into a positive pulsating dc voltage. The frequency of the output voltage is equal to the frequency of the applied ac signal since there is one pulse out for each cycle of the ac input. For example, if the input frequency is 60 hertz (60 cycles per second), the output frequency is 60 pulses per second (pps).

However, if the diode is reversed as shown in view B of figure 1-21, a negative output voltage would be obtained. This is because the current would be flowing from the top of R_L toward the bottom, making the output at the top of R_L negative with respect to the bottom or ground. Because current flows in this circuit only during half of the input cycle, it is called a half-wave rectifier.

The semiconductor diode shown in the figure can be replaced by a metallic rectifier and still achieve the same results. The metallic rectifier, sometimes referred to as a dry-disc rectifier, is a metal-to-semiconductor, large-area contact device. Its construction is distinctive; a semiconductor is sandwiched between two metal plates, or electrodes, as shown in figure 1-22. Note in the figure that a barrier, with a resistance many times greater than that of the semiconductor material, is constructed on one of the metal electrodes. The contact having the barrier is a rectifying contact; the other contact is nonrectifying. Metallic rectifiers act just like the diodes previously discussed in that they permit current to flow more readily in one direction than the other. However, the metallic rectifier is fairly large compared to the crystal diode as can be seen in figure 1-23. The reason for this is: metallic rectifier units are stacked (to prevent inverse voltage breakdown), have large area plates (to handle high currents), and usually have cooling fins (to prevent overheating).

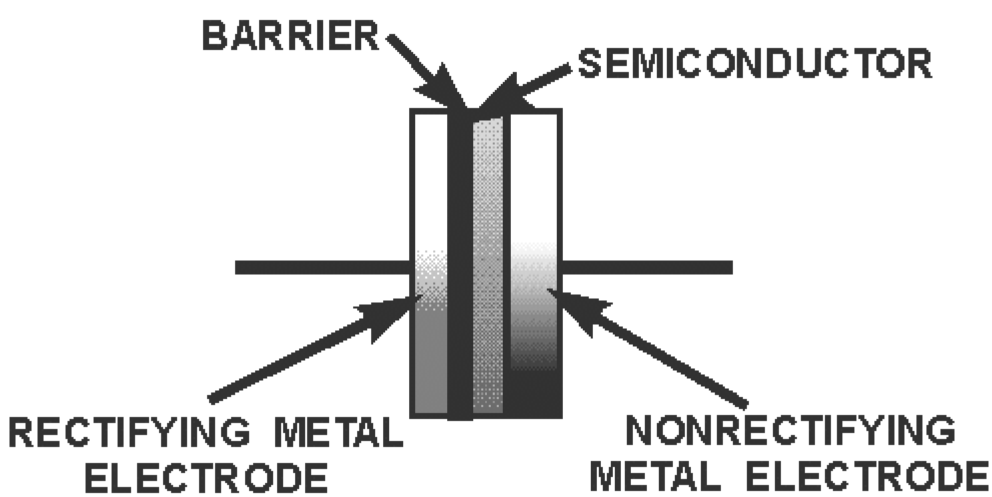


Figure 1-22.—A metallic rectifier.

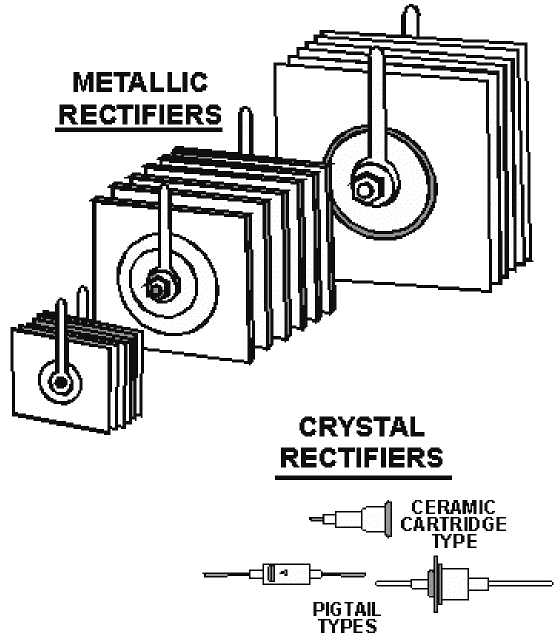


Figure 1-23.—Different types of crystal and metallic rectifiers.

There are many known metal-semiconductor combinations that can be used for contact rectification. Copper oxide and selenium devices are by far the most popular. Copper oxide and selenium are frequently used over other types of metallic rectifiers because they have a large forward current per unit contact area, low forward voltage drop, good stability, and a lower aging rate. In practical application, the selenium rectifier is used where a relatively large amount of power is required. On the other hand, copper-oxide rectifiers are generally used in small-current applications such as ac meter movements or for delivering direct current to circuits requiring not more than 10 amperes.

Since metallic rectifiers are affected by temperature, atmospheric conditions, and aging (in the case of copper oxide and selenium), they are being replaced by the improved silicon crystal rectifier. The silicon rectifier replaces the bulky selenium rectifier as to current and voltage rating, and can operate at higher ambient (surrounding) temperatures.

Diode Switch

In addition to their use as simple rectifiers, diodes are also used in circuits that mix signals together (mixers), detect the presence of a signal (detector), and act as a switch "to open or close a circuit." Diodes used in these applications are commonly referred to as "signal diodes." The simplest application of a signal diode is the basic diode switch shown in figure 1-24.

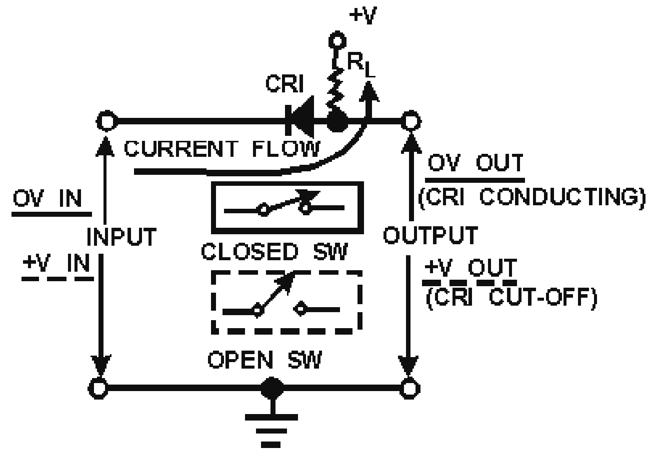


Figure 1-24.—Basic diode switch.

When the input to this circuit is at zero potential, the diode is forward biased because of the zero potential on the cathode and the positive voltage on the anode. In this condition, the diode conducts and acts as a straight piece of wire because of its very low forward resistance. In effect, the input is directly coupled to the output resulting in zero volts across the output terminals. Therefore, the diode, acts as a closed switch when its anode is positive with respect to its cathode.

If we apply a positive input voltage (equal to or greater than the positive voltage supplied to the anode) to the diode's cathode, the diode will be reverse biased. In this situation, the diode is cut off and acts as an open switch between the input and output terminals. Consequently, with no current flow in the circuit, the positive voltage on the diode's anode will be felt at the output terminal. Therefore, the diode acts as an open switch when it is reverse biased.

Q27. What is a load?

Q28. What is the output of a half-wave rectifier?

Q29. What type of rectifier is constructed by sandwiching a section of semiconductor material between two metal plates?

Q30. What type of bias makes a diode act as a closed switch?

DIODE CHARACTERISTICS

Semiconductor diodes have properties that enable them to perform many different electronic functions. To do their jobs, engineers and technicians must be supplied with data on these different types of diodes. The information presented for this purpose is called DIODE CHARACTERISTICS. These characteristics are supplied by manufacturers either in their manuals or on specification sheets (data sheets). Because of the scores of manufacturers and numerous diode types, it is not practical to put before you a specification sheet and call it typical. Aside from the difference between manufacturers, a single manufacturer may even supply specification sheets that differ both in format and content. Despite these differences, certain performance and design information is normally required. We will discuss this information in the next few paragraphs.

A standard specification sheet usually has a brief description of the diode. Included in this description is the type of diode, the major area of application, and any special features. Of particular interest is the specific application for which the diode is suited. The manufacturer also provides a drawing of the diode which gives dimension, weight, and, if appropriate, any identification marks. In addition to the above data, the following information is also provided: a static operating table (giving spot values of parameters under fixed conditions), sometimes a characteristic curve similar to the one in figure 1-20 (showing how parameters vary over the full operating range), and diode ratings (which are the limiting values of operating conditions outside which could cause diode damage).

Manufacturers specify these various diode operating parameters and characteristics with "letter symbols" in accordance with fixed definitions. The following is a list, by letter symbol, of the major electrical characteristics for the rectifier and signal diodes.

RECTIFIER DIODES

DC BLOCKING VOLTAGE [V_R]—the maximum reverse dc voltage that will not cause breakdown.

AVERAGE FORWARD VOLTAGE DROP [$V_{F(AV)}$]—the average forward voltage drop across the rectifier given at a specified forward current and temperature.

AVERAGE RECTIFIER FORWARD CURRENT [$I_{F(AV)}$]—the average rectified forward current at a specified temperature, usually at 60 Hz with a resistive load.

AVERAGE REVERSE CURRENT [$I_{R(AV)}$]—the average reverse current at a specified temperature, usually at 60 Hz.

PEAK SURGE CURRENT [I_{SURGE}]—the peak current specified for a given number of cycles or portion of a cycle.

SIGNAL DIODES

PEAK REVERSE VOLTAGE [PRV]—the maximum reverse voltage that can be applied before reaching the breakdown point. (PRV also applies to the rectifier diode.)

REVERSE CURRENT [I_R]—the small value of direct current that flows when a semiconductor diode has reverse bias.

MAXIMUM FORWARD VOLTAGE DROP AT INDICATED FORWARD CURRENT [$V_F@I_F$]—the maximum forward voltage drop across the diode at the indicated forward current.

REVERSE RECOVERY TIME [t_{rr}]—the maximum time taken for the forward-bias diode to recover its reverse bias.

The ratings of a diode (as stated earlier) are the limiting values of operating conditions, which if exceeded could cause damage to a diode by either voltage breakdown or overheating. The PN junction diodes are generally rated for: MAXIMUM AVERAGE FORWARD CURRENT, PEAK RECURRENT FORWARD CURRENT, MAXIMUM SURGE CURRENT, and PEAK REVERSE VOLTAGE.

Maximum average forward current is usually given at a special temperature, usually 25º C, (77º F) and refers to the maximum amount of average current that can be permitted to flow in the forward direction. If this rating is exceeded, structure breakdown can occur.

Peak recurrent forward current is the maximum peak current that can be permitted to flow in the forward direction in the form of recurring pulses.

Maximum surge current is the maximum current permitted to flow in the forward direction in the form of nonrecurring pulses. Current should not equal this value for more than a few milliseconds.

Peak reverse voltage (PRV) is one of the most important ratings. PRV indicates the maximum reverse-bias voltage that may be applied to a diode without causing junction breakdown.

All of the above ratings are subject to change with temperature variations. If, for example, the operating temperature is above that stated for the ratings, the ratings must be decreased.

Q31. What is used to show how diode parameters vary over a full operating range?

Q32. What is meant by diode ratings?

DIODE IDENTIFICATION

There are many types of diodes varying in size from the size of a pinhead (used in subminiature circuitry) to large 250-ampere diodes (used in high-power circuits). Because there are so many different types of diodes, some system of identification is needed to distinguish one diode from another. This is accomplished with the semiconductor identification system shown in figure 1-25. This system is not only used for diodes but transistors and many other special semiconductor devices as well. As illustrated in this figure, the system uses numbers and letters to identify different types of semiconductor devices. The first number in the system indicates the number of junctions in the semiconductor device and is a number, one less than the number of active elements. Thus 1 designates a diode; 2 designates a transistor (which may be considered as made up of two diodes); and 3 designates a tetrode (a four-element transistor). The letter "N" following the first number indicates a semiconductor. The 2- or 3-digit number following the letter "N" is a serialized identification number. If needed, this number may contain a suffix letter after the last digit. For example, the suffix letter "M" may be used to describe matching pairs of separate semiconductor devices or the letter "R" may be used to indicate reverse polarity. Other letters are used to indicate modified versions of the device which can be substituted for the basic numbered unit. For example, a semiconductor diode designated as type 1N345A signifies a two-element diode (1) of semiconductor material (N) that is an improved version (A) of type 345.

XNYYY

XN YYY

COMPONENT IDENTIFICATION NUMBER

X - NUMBER OF SEMICONDUCTOR JUNCTIONS
N - A SEMICONDUCTOR
YYY - IDENTIFICATION NUMBER (ORDER OR REGISTRATION NUMBER) ALSO INCLUDES SUFFIX LETTER (IF APPLICABLE) TO INDICATE

1. MATCHING DEVICES
2. REVERSE POLARITY
3. MODIFICATION

EXAMPLE - 1N345A (AN IMPROVED VERSION OF THE SEMICONDUCTOR DIODE TYPE 345)

Figure 1-25.—Semiconductor identification codes.

When working with these different types of diodes, it is also necessary to distinguish one end of the diode from the other (anode from cathode). For this reason, manufacturers generally code the cathode end of the diode with a "k," "+," "cath," a color dot or band, or by an unusual shape (raised edge or taper) as shown in figure 1-26. In some cases, standard color code bands are placed on the cathode end of the diode. This serves two purposes: (1) it identifies the cathode end of the diode, and (2) it also serves to identify the diode by number.

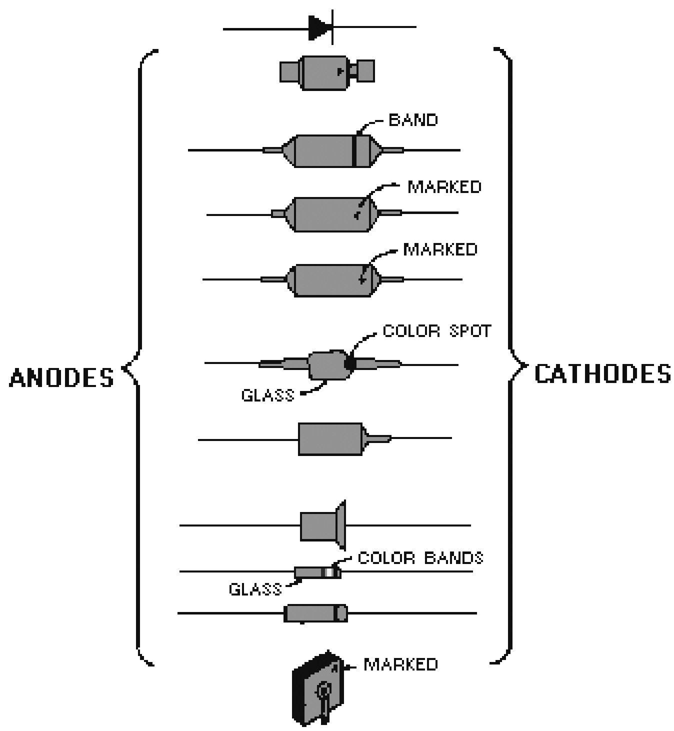
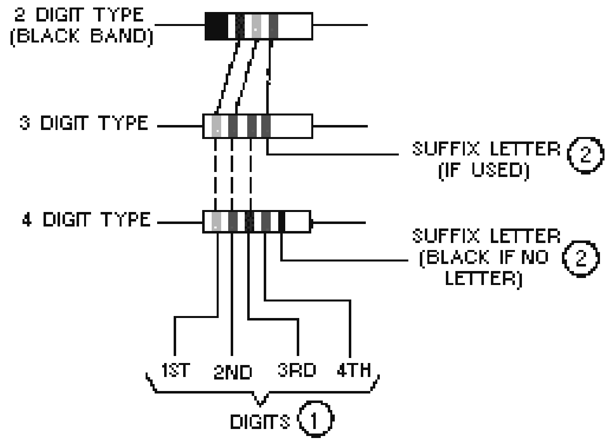


Figure 1-26.—Semiconductor diode markings.

The standard diode color code system is shown in figure 1-27. Take, for example, a diode with brown, orange, and white bands at one terminal and figure out its identification number. With brown being a "1," orange a "3," and white "9," the device would be identified as a type 139 semiconductor diode, or specifically 1N139.



COLOR	1 DIGIT	2 DIODE SUFFIX LETTER
BLACK	0	-
BROWN	1	A
RED	2	B
ORANGE	3	C
YELLOW	4	D
GREEN	5	E
BLUE	6	F
VIOLET	7	G
GRAY	8	H
WHITE	9	J
SILVER	-	-
GOLD	-	-
NONE	-	-

Figure 1-27.—Semiconductor diode color code system.

Keep in mind, whether the diode is a small crystal type or a large power rectifier type, both are still represented schematically, as explained earlier, by the schematic symbol shown in figure 1-12.

Q33. What does the letter "N" indicate in the semiconductor identification system?

Q34. What type of diode has orange, blue, and gray bands?

DIODE MAINTENANCE

Diodes are rugged and efficient. They are also expected to be relatively trouble free. Protective encapsulation processes and special coating techniques have even further increased their life expectancies. In theory, a diode should last indefinitely. However, if diodes are subjected to current overloads, their junctions will be damaged or destroyed. In addition, the application of excessively high operating voltages can damage or destroy junctions through arc-over, or excessive reverse currents. One of the greatest dangers to the diode is heat. Heat causes more electron-hole pairs to be generated, which in turn increases current flow. This increase in current generates more heat and the cycle repeats itself until the diode draws excessive current. This action is referred to as THERMAL RUNAWAY and eventually causes diode destruction. Extreme caution should be used when working with equipment containing diodes to ensure that these problems do not occur and cause irreparable diode damage.

The following is a list of some of the special safety precautions that should be observed when working with diodes:

- Never remove or insert a diode into a circuit with voltage applied.
- Never pry diodes to loosen them from their circuits.
- Always be careful when soldering to ensure that excessive heat is not applied to the diode.
- When testing a diode, ensure that the test voltage does not exceed the diode's maximum allowable voltage.
- Never put your fingers across a signal diode because the static charge from your body could short it out.
- Always replace a diode with a direct replacement, or with one of the same type.
- Ensure a replacement diode is put into a circuit in the correct direction.

If a diode has been subjected to excessive voltage or temperature and is suspected of being defective, it can be checked in various ways. The most convenient and quickest way of testing a diode is with an ohmmeter (fig. 1-28). To make the check, simply disconnect one of the diode leads from the circuit wiring, and make resistance measurements across the leads of the diode. The resistance measurements obtained depend upon the test-lead polarity of the ohmmeter; therefore, two measurements must be taken. The first measurement is taken with the test leads connected to either end of the diode and the second measurement is taken with the test leads reversed on the diode. The larger resistance value is assumed to be the reverse (back) resistance of the diode, and the smaller resistance (front) value is assumed to be the forward resistance. Measurement can be made for comparison purposes using another identical-type diode (known to be good) as a standard. Two high-value resistance measurements indicate that the diode is open or has a high forward resistance. Two low-value resistance measurements indicate that the diode is shorted or has a low reverse resistance. A normal set of measurements will show a high resistance in the reverse direction and a low resistance in the forward direction. The diode's efficiency is determined by how low the forward resistance is compared with the reverse resistance. That is, it is desirable to have as great a ratio (often known as the front-to-back ratio or the back-to-front ratio) as possible between the reverse and forward resistance measurements. However, as a rule of thumb, a small signal diode will have a ratio of several hundred to one, while a power rectifier can operate satisfactorily with a ratio of 10 to 1.

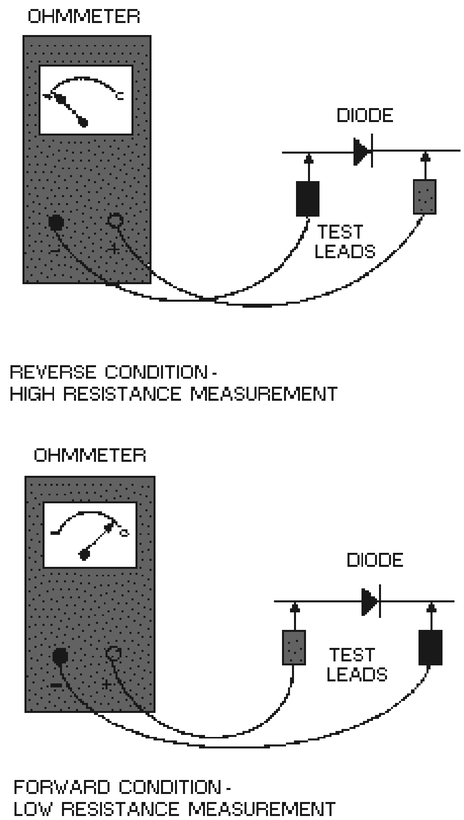


Figure 1-28.—Checking a diode with an ohmmeter.

One thing you should keep in mind about the ohmmeter check-it is not conclusive. It is still possible for a diode to check good under this test, but break down when placed back in the circuit. The problem is that the meter used to check the diode uses a lower voltage than the diode usually operates at in the circuit.

Another important point to remember is that a diode should not be condemned because two ohmmeters give different readings on the diode. This occurs because of the different internal resistances of the ohmmeters and the different states of charge on the ohmmeter batteries. Because each ohmmeter sends a different current through the diode, the two resistance values read on the meters will not be the same.

Another way of checking a diode is with the substitution method. In this method, a good diode is substituted for a questionable diode. This technique should be used only after you have made voltage and resistance measurements to make certain that there is no circuit defect that might damage the substitution diode. If more than one defective diode is present in the equipment section where trouble has been localized, this method becomes cumbersome, since several diodes may have to be replaced before the trouble is corrected. To determine which stages failed and which diodes are not defective, all of the removed diodes must be tested. This can be accomplished by observing whether the equipment operates correctly as each of the removed diodes is reinserted into the equipment.

In conclusion, the only valid check of a diode is a dynamic electrical test that determines the diode's forward current (resistance) and reverse current (resistance) parameters. This test can be accomplished using various crystal diode test sets that are readily available from many manufacturers.

Q35. What is the greatest threat to a diode?

Q36. When checking a diode with an ohmmeter, what is indicated by two high resistance measurements?

SUMMARY

Now that we have completed this chapter, a short review of the more important points covered in the chapter will follow. You should be thoroughly familiar with these points before continuing on to chapter 2.

The **UNIVERSE** consists of two main parts-matter and energy.

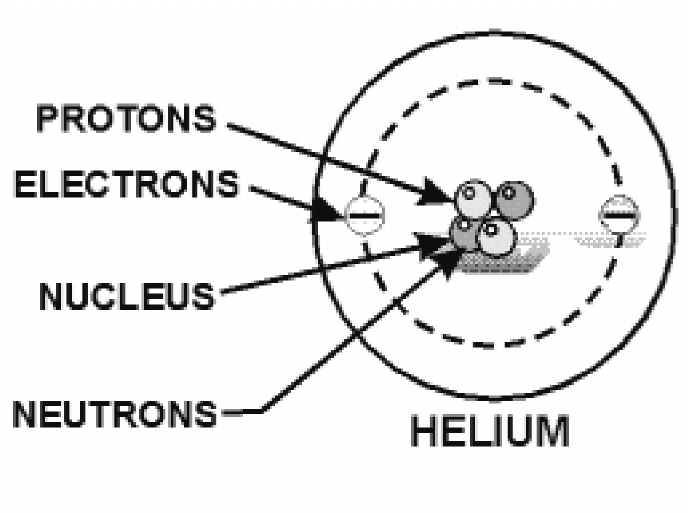
MATTER is anything that occupies space and has weight. Rocks, water, and air are examples of matter. Matter may be found in any one of three states: solid, liquid and gaseous. It can also be composed of either an element or a combination of elements.

An **ELEMENT** is a substance that cannot be reduced to a simpler form by chemical means. Iron, gold, silver, copper, and oxygen are all good examples of elements.

A **COMPOUND** is a chemical combination of two or more elements. Water, table salt, ethyl alcohol, and ammonia are all examples of compounds.

A **MOLECULE** is the smallest part of a compound that has all the characteristics of the compound. Each molecule contains some of the atoms of each of the elements forming the compound.

The **ATOM** is the smallest particle into which an element can be broken down and still retain all its original properties. An atom is made up of electrons, protons, and neutrons. The number and arrangement of these particles determine the kind of element.



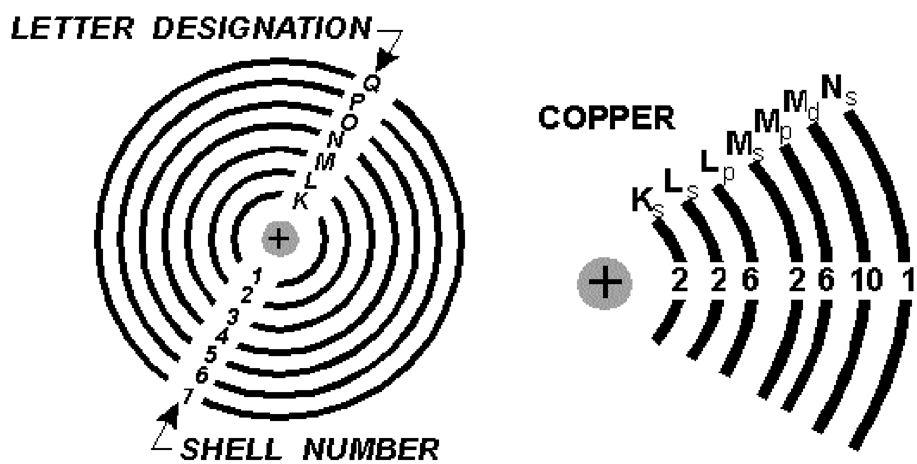
An **ELECTRON** carries a small negative charge of electricity.

The **PROTON** carries a positive charge of electricity that is equal and opposite to the charge of the electron. However, the mass of the proton is approximately 1,837 times that of the electron.

The **NEUTRON** is a neutral particle in that it has no electrical charge. The mass of the neutron is approximately equal to that of the proton.

An **ELECTRON'S ENERGY LEVEL** is the amount of energy required by an electron to stay in orbit. Just by the electron's motion alone, it has kinetic energy. The electron's position in reference to the nucleus gives it potential energy. An energy balance keeps the electron in orbit and as it gains or loses energy, it assumes an orbit further from or closer to the center of the atom.

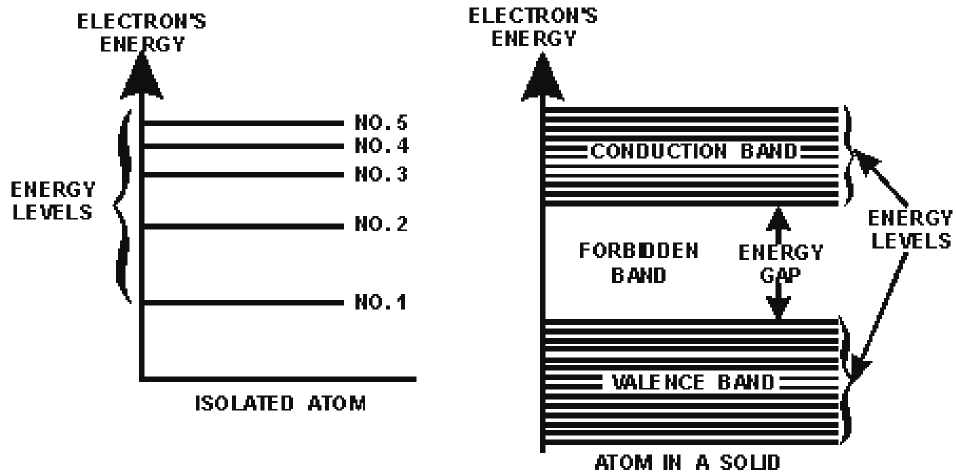
SHELLS and **SUBSHELLS** are the orbits of the electrons in an atom. Each shell can contain a maximum number of electrons, which can be determined by the formula $2n^2$. Shells are lettered K through Q, starting with K, which is the closest to the nucleus. The shell can also be split into four subshells labeled s, p, d, and f, which can contain 2, 6, 10, and 14 electrons, respectively.



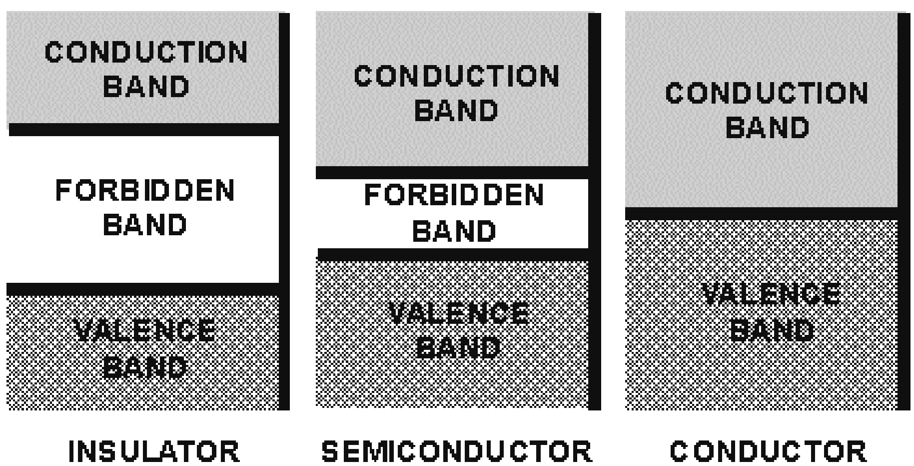
VALENCE is the ability of an atom to combine with other atoms. The valence of an atom is determined by the number of electrons in the atom's outermost shell. This shell is referred to as the VALENCE SHELL. The electrons in the outermost shell are called VALENCE ELECTRONS.

IONIZATION is the process by which an atom loses or gains electrons. An atom that loses some of its electrons in the process becomes positively charged and is called a POSITIVE ION. An atom that has an excess number of electrons is negatively charged and is called a **NEGATIVE ION**.

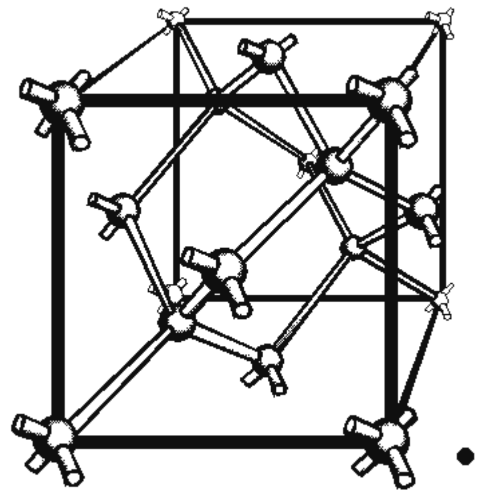
ENERGY BANDS are groups of energy levels that result from the close proximity of atoms in a solid. The three most important energy bands are the CONDUCTION BAND, FORBIDDEN BAND, and VALENCE BAND.



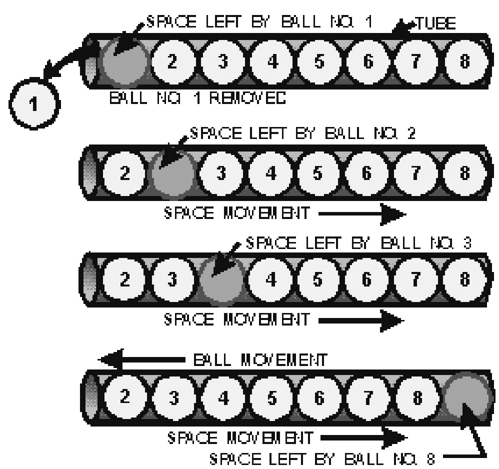
CONDUCTORS, **SEMICONDUCTORS**, and **INSULATORS** are categorized as such by using the energy band concept. It is the width of the forbidden band that determines whether a material is an insulator, a semiconductor, or a conductor. A CONDUCTOR has a very narrow forbidden band or none at all. A SEMICONDUCTOR has a medium width forbidden band. An INSULATOR has a wide forbidden band.



COVALENT BONDING is the sharing of valence electrons between two or more atoms. It is this bonding that holds the atoms together in an orderly structure called a CRYSTAL.

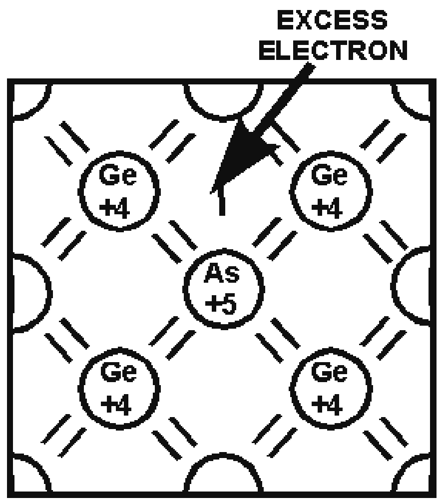


The **CONDUCTION PROCESS** in a **SEMICONDUCTOR** is accomplished by two different types of current flow: **HOLE FLOW** and **ELECTRON FLOW**. Hole flow is very similar to electron flow except that holes (positive charges) move toward a negative potential and in an opposite direction to that of the electrons. In an **INTRINSIC** semiconductor (one which does not contain any impurities), the number of holes always equals the number of conducting electrons.

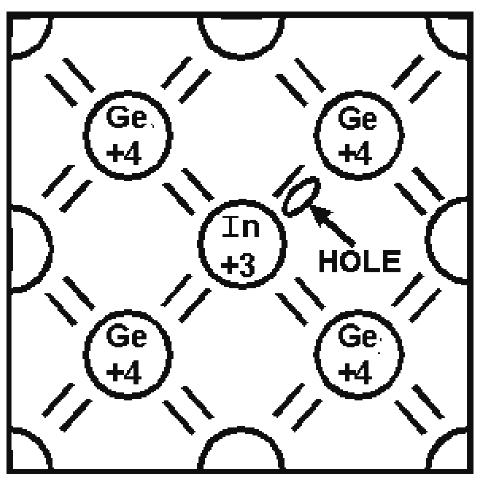


DOPING is the process by which small amounts of selected additives, called impurities, are added to semiconductors to increase their current flow. Semiconductors that undergo this treatment are referred to as **EXTRINSIC SEMICONDUCTORS**.

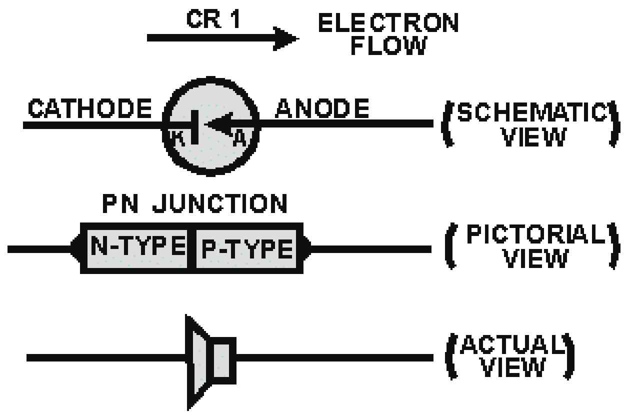
An **N-TYPE SEMICONDUCTOR** is one that is doped with an **N-TYPE** or donor impurity (an impurity that easily loses its extra electron to the semiconductor causing it to have an excess number of free electrons). Since this type of semiconductor has a surplus of electrons, the electrons are considered the majority current carriers, while the holes are the minority current carriers.



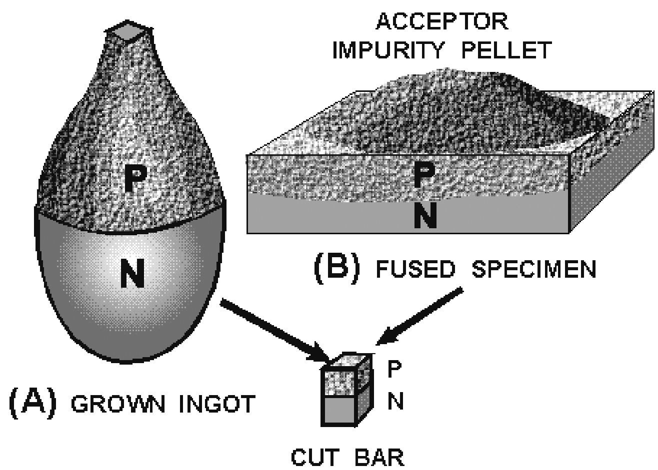
A **P-TYPE SEMICONDUCTOR** is one which is doped with a P-TYPE or acceptor impurity (an impurity that reduces the number of free electrons causing more holes). The holes in this type semiconductor are the majority current carriers since they are present in the greatest quantity while the electrons are the minority current carriers.



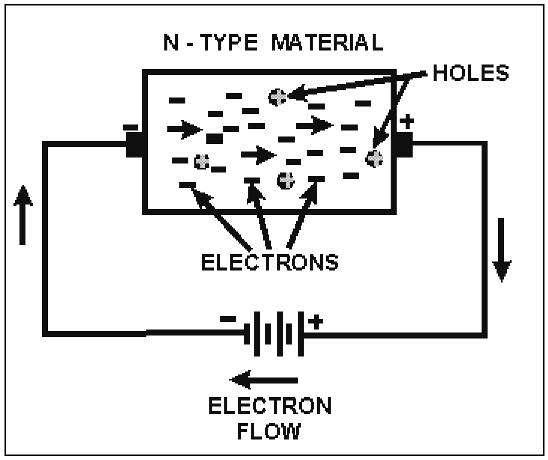
The **SEMICONDUCTOR DIODE**, also known as **a PN JUNCTION DIODE**, is a two-element semiconductor device that makes use of the rectifying properties of a PN junction to convert alternating current into direct current by permitting current flow in only one direction.



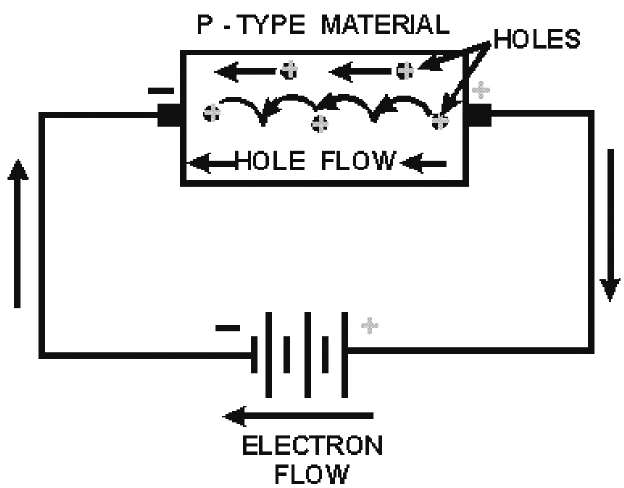
A **PN JUNCTION CONSTRUCTION** varies from one manufacturer to the next. Some of the more commonly used manufacturing techniques are: **GROWN**, **ALLOY** or **FUSED-ALLOY**, **DIFFUSED**, and **POINT-CONTACT**.



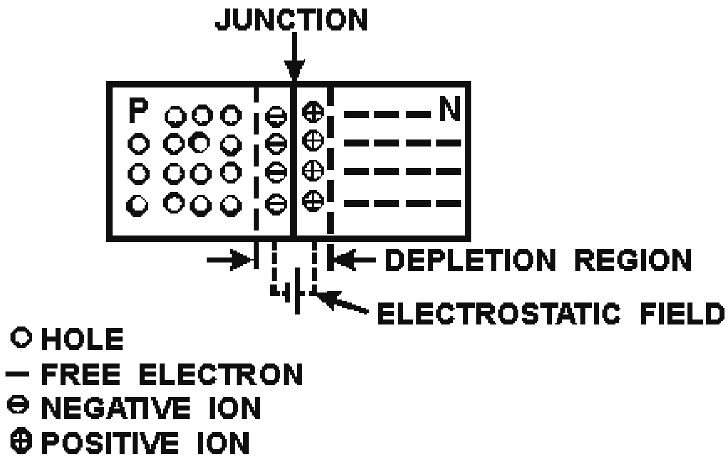
CURRENT FLOW in an **N-TYPE MATERIAL** is similar to conduction in a copper wire. That is, with voltage applied across the material, electrons will move through the crystal toward the positive terminal just like current flows in a copper wire.



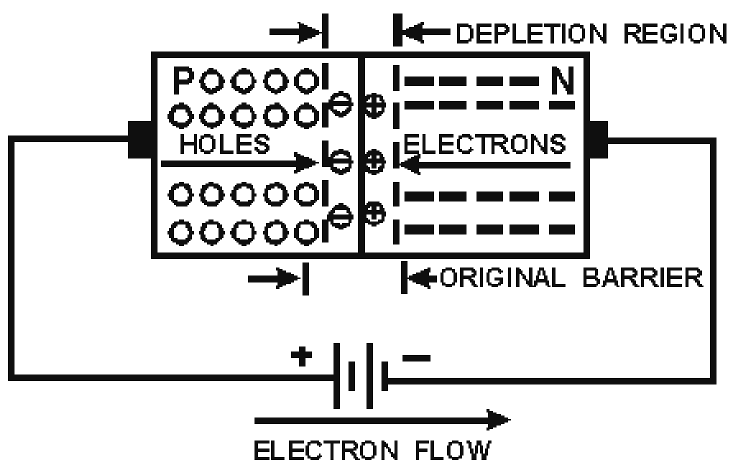
CURRENT FLOW in a **P-TYPE MATERIAL** is by positive holes, instead of negative electrons. Unlike the electron, the hole moves from the positive terminal of the P material to the negative terminal.



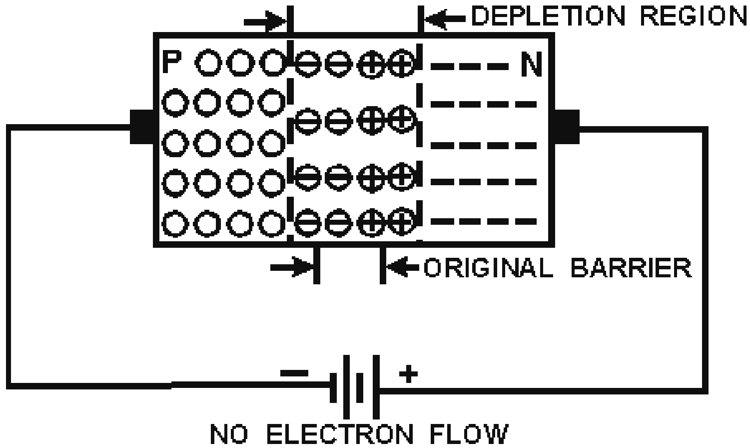
JUNCTION BARRIER is an electrostatic field that has been created by the joining of a section of N material with a section of P material. Since holes and electrons must overcome this field to cross the junction, the electrostatic field is commonly called a BARRIER. Because there is a lack or depletion of free electrons and holes in the area around the barrier, this area has become known as the DEPLETION REGION.



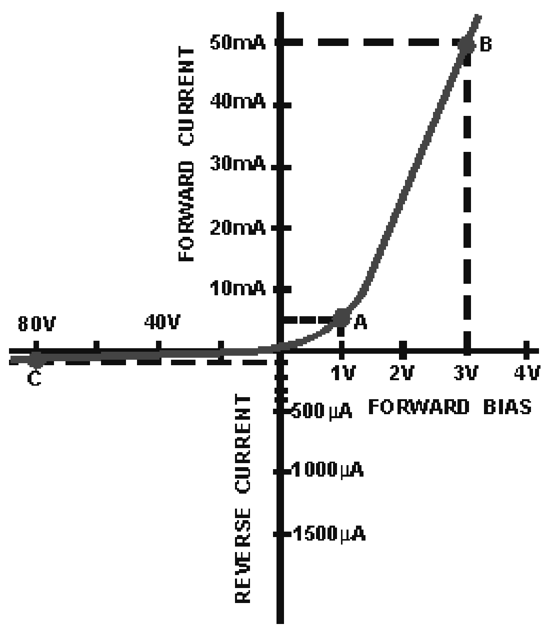
FORWARD BIAS is an external voltage that is applied to a PN junction to reduce its barrier and, therefore, aid current flow through the junction. To accomplish this function, the external voltage is connected so that it opposes the electrostatic field of the junction.



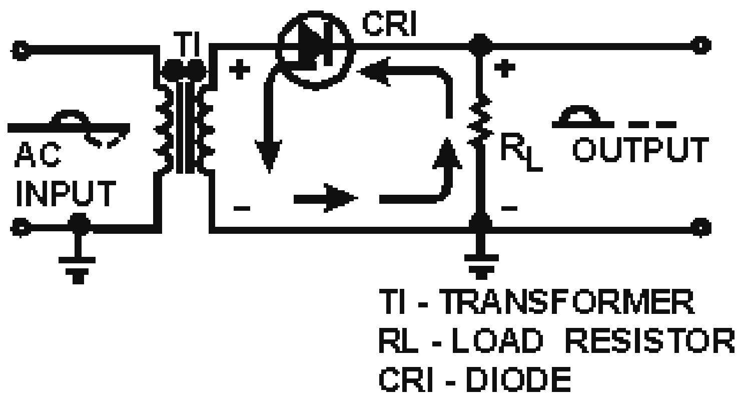
REVERSE BIAS is an external voltage that is connected across a PN junction so that its voltage aids the junction and, thereby, offers a high resistance to the current flow through the junction.



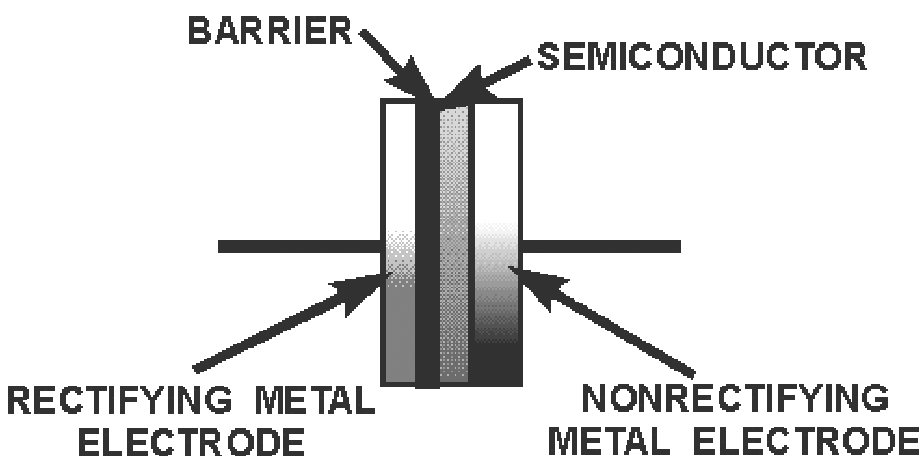
The **PN JUNCTION** has a unique ability to offer very little resistance to current flow in the forward-bias direction, but maximum resistance to current flow when reverse biased. For this reason, the PN junction is commonly used as a diode to convert ac to dc.



The **PN JUNCTION'S APPLICATION** expands many different areas-from a simple voltage protection device to an amplifying diode. Two of the most commonly used applications for the PN junction are the **SIGNAL DIODE** (mixing, detecting, and switching signals) and the **RECTIFYING DIODE** (converting ac to dc).



The **METALLIC RECTIFIER** or dry-disc rectifier is a metal-to-semiconductor device that acts just like a diode in that it permits current to flow more readily in one direction than the other. Metallic rectifiers are used in many applications where a relatively large amount of power is required.



DIODE CHARACTERISTICS is the information supplied by manufacturers on different types of diodes, either in their manuals or on specification sheets.

DIODE RATINGS are the limiting value of operating conditions of a diode. Operation of the diode outside of its operating limits could damage the diode. Diodes are generally rated for: **MAXIMUM AVERAGE FORWARD CURRENT**, **PEAK RECURRENT FORWARD CURRENT**, **MAXIMUM SURGE CURRENT**, and **PEAK REVERSE VOLTAGE**.

The **SEMICONDUCTOR IDENTIFICATION SYSTEM** is an alphanumerical code used to distinguish one semiconductor from another. It is used for diodes, transistors, and many other special semiconductor devices.

XNYYY

XN YYY

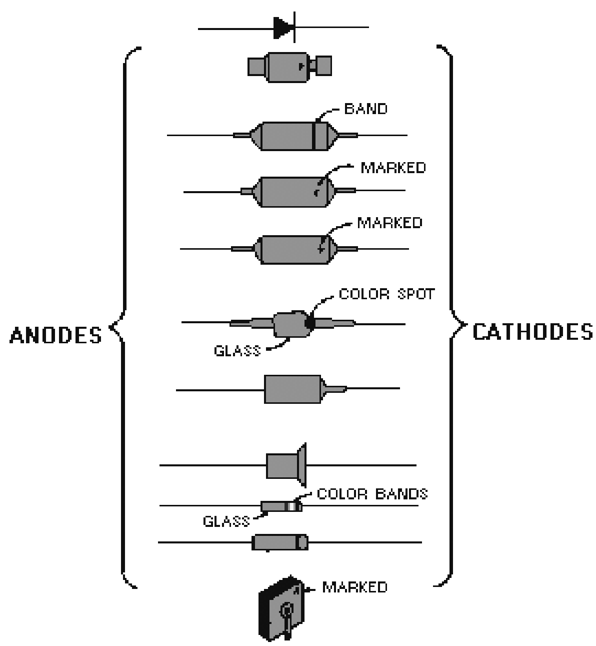
COMPONENT IDENTIFICATION NUMBER

X - NUMBER OF SEMICONDUCTOR JUNCTIONS
N - A SEMICONDUCTOR
YYY - IDENTIFICATION NUMBER (ORDER OR REGISTRATION NUMBER) ALSO INCLUDES SUFFIX LETTER (IF APPLICABLE) TO INDICATE

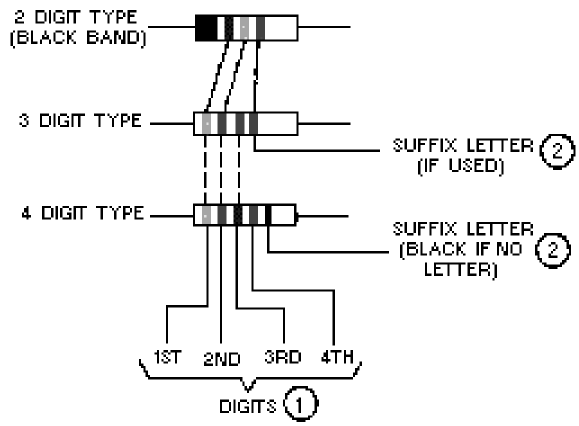
1. MATCHING DEVICES
2. REVERSE POLARITY
3. MODIFICATION

EXAMPLE - 1N345A (AN IMPROVED VERSION OF THE SEMICONDUCTOR DIODE TYPE 345)

DIODE MARKINGS are letters and symbols placed on the diode by manufacturers to distinguish one end of the diode from the other. In some cases, an unusual shape or the addition of color code bands is used to distinguish the cathode from the anode.



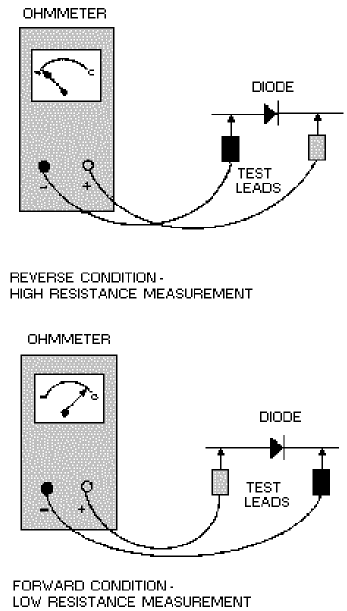
The **STANDARD DIODE COLOR CODE SYSTEM** serves two purposes when it is used: (1) it identifies the cathode end of the diode, and (2) it also serves to identify the diode by number.



COLOR	1 DIGIT	2 DIODE SUFFIX LETTER
BLACK	0	-
BROWN	1	A
RED	2	B
ORANGE	3	C
YELLOW	4	D
GREEN	5	E
BLUE	6	F
VIOLET	7	G
GRAY	8	H
WHITE	9	J
SILVER	-	-
GOLD	-	-
NONE	-	-

DIODE MAINTENANCE is the procedures or methods used to keep a diode in good operating condition. To prevent diode damage, you should observe standard diode safety precautions and ensure that diodes are not subjected to heat, current overloads, and excessively high operating voltages.

TESTING A DIODE can be accomplished by using an ohmmeter, the substitution method, or a dynamic diode tester. The most convenient and quickest way of testing a diode is with an ohmmeter.



ANSWERS TO QUESTIONS Q1. THROUGH Q36.

A1. *An electronic device that operates by virtue of the movement of electrons within a solid piece of semiconductor material.*

A2. *It is the decrease in a semiconductor's resistance as temperature rises.*

A3. *Space systems, computers, and data processing equipment.*

A4. *The electron tube requires filament or heater voltage, whereas the semiconductor device does not; consequently, no power input is spent by the semiconductor for conduction.*

A5. *Anything that occupies space and has weight. Solid, liquid, and gas.*

A6. *The atom.*

A7. *Electrons-negative, protons-positive, and neutrons-neutral.*

A8. *The valence shell.*

A9. *Quanta.*

A10. *A negatively charged atom having more than its normal amount of electrons.*

A11. *The energy levels of an atom in a solid group together to form energy bands, whereas the isolated atom does not.*

A12. *The width of the forbidden band.*

A13. *The number of electrons in the valence shell.*

A14. *Covalent bonding.*

A15. Electron flow and hole flow.

A16. Intrinsic.

A17. P-type crystal.

A18. Electrons.

A19. To convert alternating current into direct current.

A20. Toward the arrow.

A21. Point-contact.

A22. Holes.

A23. N-type material.

A24. Depletion region.

A25. Negative.

A26. Forward.

A27. Any device that draws current.

A28. A pulsating dc voltage.

A29. Metallic rectifier.

A30. Forward bias.

A31. A characteristic curve.

A32. They are the limiting values of operating conditions outside which operations could cause diode damage.

A33. A semiconductor.

A34. 1N368.

A35. Heat.

A36. The diode is open or has a high-forward resistance.

CHAPTER 2

TRANSISTORS

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to do the following:

1. Define the term *transistor* and give a brief description of its construction and operation.
2. Explain how the transistor can be used to amplify a signal.
3. Name the four classes of amplifiers and give an explanation for each.
4. List the three different transistor circuit configurations and explain their operation.
5. Identify the different types of transistors by their symbology and alphanumerical designations.
6. List the precautions to be taken when working with transistors and describe ways to test them.
7. Explain the meaning of the expression "integrated circuits."
8. Give a brief description on how integrated circuits are constructed and the advantages they offer over conventional transistor circuits.
9. Name the two types of circuit boards.
10. State the purpose and function of modular circuitry.

INTRODUCTION TO TRANSISTORS

The discovery of the first transistor in 1948 by a team of physicists at the Bell Telephone Laboratories sparked an interest in solid-state research that spread rapidly. The transistor, which began as a simple laboratory oddity, was rapidly developed into a semiconductor device of major importance. The transistor demonstrated for the first time in history that amplification in solids was possible. Before the transistor, amplification was achieved only with electron tubes. Transistors now perform numerous electronic tasks with new and improved transistor designs being continually put on the market. In many cases, transistors are more desirable than tubes because they are small, rugged, require no filament power, and operate at low voltages with comparatively high efficiency. The development of a family of transistors has even made possible the miniaturization of electronic circuits. Figure 2-1 shows a sample of the many different types of transistors you may encounter when working with electronic equipment.

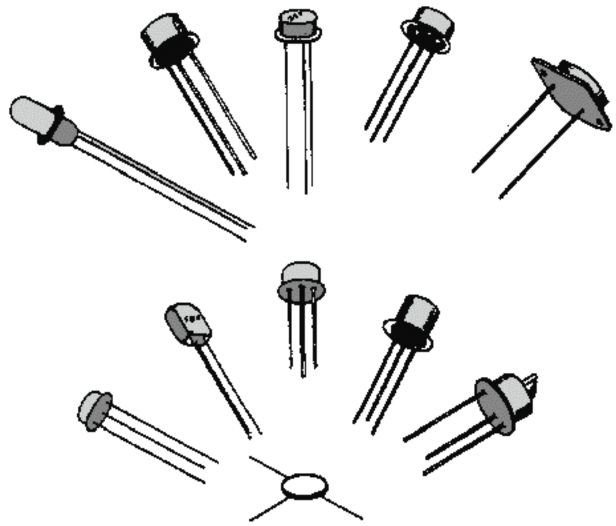


Figure 2-1.—An assortment of different types of transistors.

Transistors have infiltrated virtually every area of science and industry, from the family car to satellites. Even the military depends heavily on transistors. The ever increasing uses for transistors have created an urgent need for sound and basic information regarding their operation.

From your study of the PN-junction diode in the preceding chapter, you now have the basic knowledge to grasp the principles of transistor operation. In this chapter you will first become acquainted with the basic types of transistors, their construction, and their theory of operation. You will also find out just how and why transistors amplify. Once this basic information is understood, transistor terminology, capabilities, limitations, and identification will be discussed. Last, we will talk about transistor maintenance, integrated circuits, circuit boards, and modular circuitry.

TRANSISTOR FUNDAMENTALS

The first solid-state device discussed was the two-element semiconductor diode. The next device on our list is even more unique. It not only has one more element than the diode but it can <u>amplify</u> as well. Semiconductor devices that have-three or more elements are called TRANSISTORS. The term transistor was derived from the words <u>TRANS</u>fer and res<u>ISTOR</u>. This term was adopted because it best describes the operation of the transistor - the transfer of an input signal current from a low-resistance circuit to a high-resistance circuit. Basically, the transistor is a solid-state device that amplifies by controlling the flow of current carriers through its semiconductor materials.

There are many different types of transistors, but their basic theory of operation is all the same. As a matter of fact, the theory we will be using to explain the operation of a transistor is the same theory used earlier with the PN-junction diode except that now two such junctions are required to form the three elements of a transistor. The three elements of the two-junction transistor are (1) the EMITTER, which gives off, or emits," current carriers (electrons or holes); (2) the BASE, which controls the flow of current carriers; and (3) the COLLECTOR, which collects the current carriers.

CLASSIFICATION

Transistors are classified as either NPN or PNP according to the arrangement of their N and P materials. Their basic construction and chemical treatment is implied by their names, "NPN" or "PNP." That is, an NPN transistor is formed by introducing a thin region of P-type material between two regions of N-type material. On the other hand, a PNP transistor is formed by introducing a thin region of N-type material between two regions of P-type material. Transistors constructed in this manner have two PN junctions, as shown in figure 2-2. One PN junction is between the emitter and the base; the other PN junction is between the collector and the base. The two junctions share one section of semiconductor material so that the transistor actually consists of three elements.

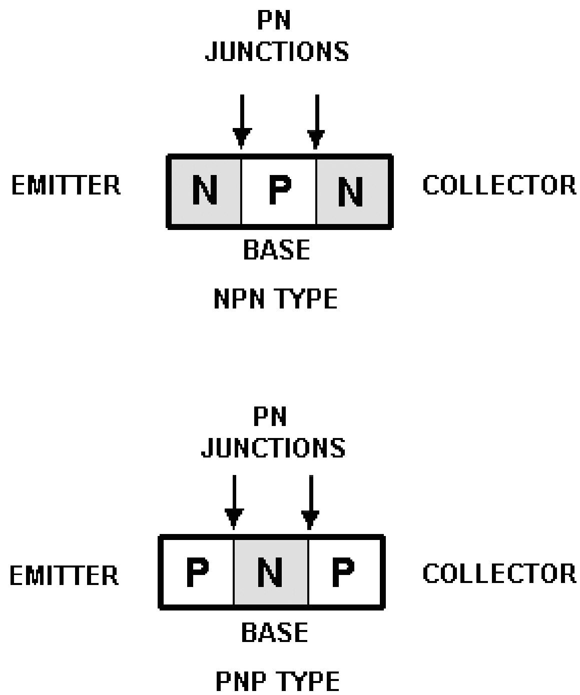


Figure 2-2.—Transistor block diagrams.

Since the majority and minority current carriers are different for N and P materials, it stands to reason that the internal operation of the NPN and PNP transistors will also be different. The theory of operation of the NPN and PNP transistors will be discussed separately in the next few paragraphs. Any additional information about the PN junction will be given as the theory of transistor operation is developed.

To prepare you for the forthcoming information, the two basic types of transistors along with their circuit symbols are shown in figure 2-3. It should be noted that the two symbols are different. The horizontal line represents the base, the angular line with the arrow on it represents the emitter, and the other angular line represents the collector. The direction of the arrow on the emitter distinguishes the NPN from the PNP transistor. If the arrow points in, (Points iN) the transistor is a PNP. On the other hand if the arrow points out, the transistor is an NPN (Not Pointing iN).

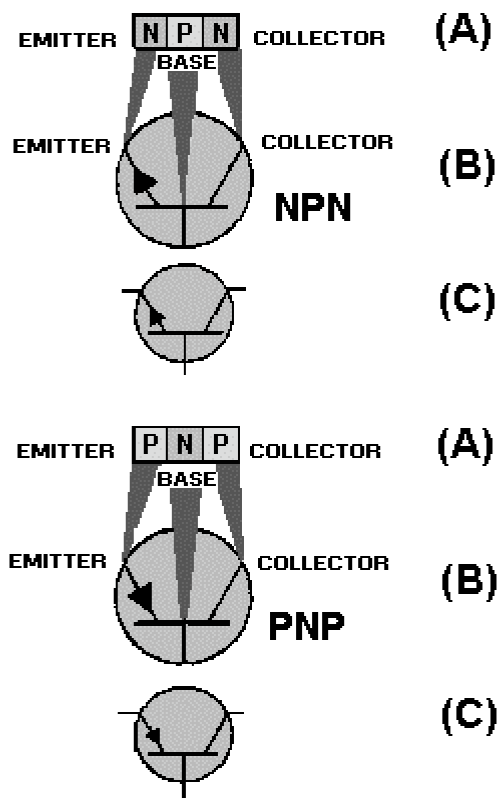


Figure 2-3.—Transistor representations.

Another point you should keep in mind is that the arrow always points in the direction of hole flow, or from the P to N sections, no matter whether the P section is the emitter or base. On the other hand, electron flow is always toward or against the arrow, just like in the junction diode.

CONSTRUCTION

The very first transistors were known as point-contact transistors. Their construction is similar to the construction of the point-contact diode covered in chapter 1. The difference, of course, is that the point-contact transistor has two P or N regions formed instead of one. Each of the two regions constitutes an electrode (element) of the transistor. One is named the emitter and the other is named the collector, as shown in figure 2-4, view A.

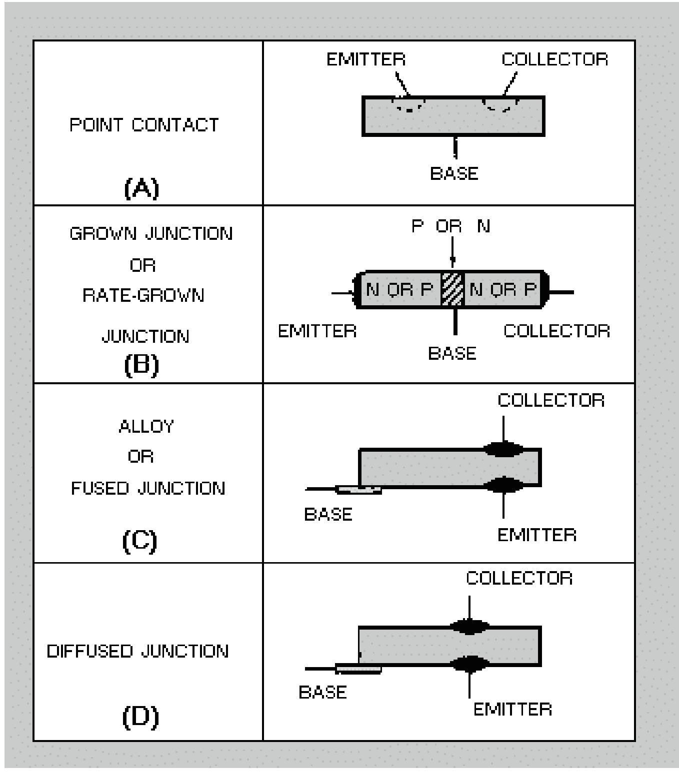


Figure 2-4.—Transistor constructions.

Point-contact transistors are now practically obsolete. They have been replaced by junction transistors, which are superior to point-contact transistors in nearly all respects. The junction transistor generates less noise, handles more power, provides higher current and voltage gains, and can be mass-produced more cheaply than the point-contact transistor. Junction transistors are manufactured in much the same manner as the PN junction diode discussed earlier. However, when the PNP or NPN material is grown (view B), the impurity mixing process must be reversed twice to obtain the two junctions required in a transistor. Likewise, when the alloy-junction (view C) or the diffused-junction (view D) process is used, two junctions must also be created within the crystal.

Although there are numerous ways to manufacture transistors, one of the most important parts of any manufacturing process is quality control. Without good quality control, many transistors would prove unreliable because the construction and processing of a transistor govern its thermal ratings, stability, and electrical characteristics. Even though there are many variations in the transistor manufacturing processes, certain structural techniques, which yield good reliability and long life, are common to all processes: (1) Wire leads are connected to each semiconductor electrode; (2) the crystal is specially mounted to protect it against mechanical damage; and (3) the unit is sealed to prevent harmful contamination of the crystal.

Q1. What is the name given to the semiconductor device that has three or more elements?

Q2. What electronic function made the transistor famous?

Q3. In which direction does the arrow point on an NPN transistor?

Q4. What was the name of the very first transistor?

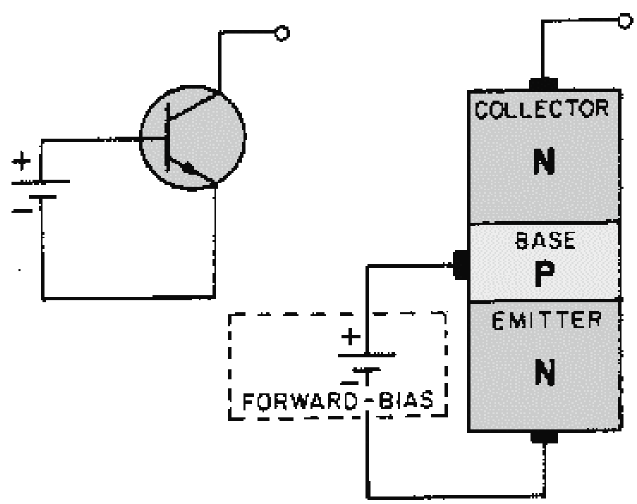
Q5. What is one of the most important parts of any transistor manufacturing process?

TRANSISTOR THEORY

You should recall from an earlier discussion that a forward-biased PN junction is comparable to a low-resistance circuit element because it passes a high current for a given voltage. In turn, a reverse-biased PN junction is comparable to a high-resistance circuit element. By using the Ohm's law formula for power ($P = I^2R$) and assuming current is held constant, you can conclude that the power developed across a high resistance is greater than that developed across a low resistance. Thus, if a crystal were to contain two PN junctions (one forward-biased and the other reverse-biased), a low-power signal could be injected into the forward-biased junction and produce a high-power signal at the reverse-biased junction. In this manner, a power gain would be obtained across the crystal. This concept, which is merely an extension of the material covered in chapter 1, is the basic theory behind how the transistor amplifies. With this information fresh in your mind, let's proceed directly to the NPN transistor.

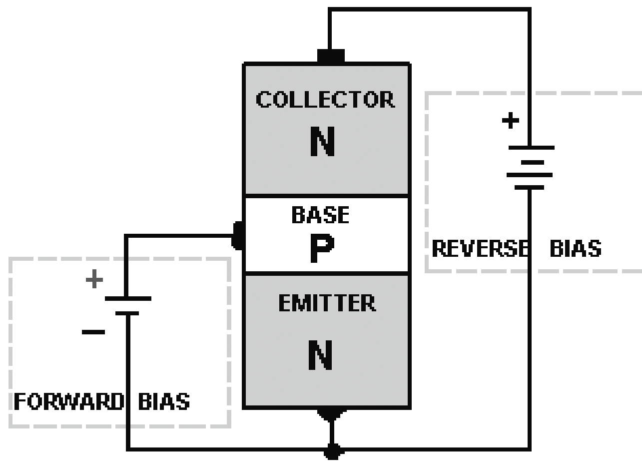
NPN Transistor Operation

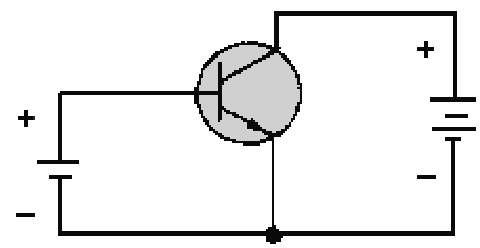
Just as in the case of the PN junction diode, the N material comprising the two end sections of the NP N transistor contains a number of free electrons, while the center P section contains an excess number of holes. The action at each junction between these sections is the same as that previously described for the diode; that is, depletion regions develop and the junction barrier appears. To use the transistor as an amplifier, each of these junctions must be modified by some external bias voltage. For the transistor to function in this capacity, the first PN junction (emitter-base junction) is biased in the forward, or low-resistance, direction. At the same time the second PN junction (base-collector junction) is biased in the reverse, or high-resistance, direction. A simple way to remember how to properly bias a transistor is to observe the NPN or PNP elements that make up the transistor. The letters of these elements indicate what polarity voltage to use for correct bias. For instance, notice the NPN transistor below:



1. The emitter, which is the first letter in the NPN sequence, is connected to the negative side of the battery while the base, which is the second letter (NPN), is connected to the positive side.

2. However, since the second PN junction is required to be reverse biased for proper transistor operation, the collector must be connected to an opposite polarity voltage (positive) than that indicated by its letter designation(NPN). The voltage on the collector must also be more positive than the base, as shown below:





We now have a properly biased NPN transistor.

In summary, the base of the NPN transistor must be positive with respect to the emitter, and the collector must be more positive than the base.

NPN FORWARD-BIASED JUNCTION.—An important point to bring out at this time, which was not necessarily mentioned during the explanation of the diode, is the fact that the N material on one side of the forward-biased junction is more heavily doped than the P material. This results in more current being carried across the junction by the majority carrier electrons from the N material than the majority carrier holes from the P material. Therefore, conduction through the forward-biased junction, as shown in figure 2-5, is mainly by majority carrier electrons from the N material (emitter).

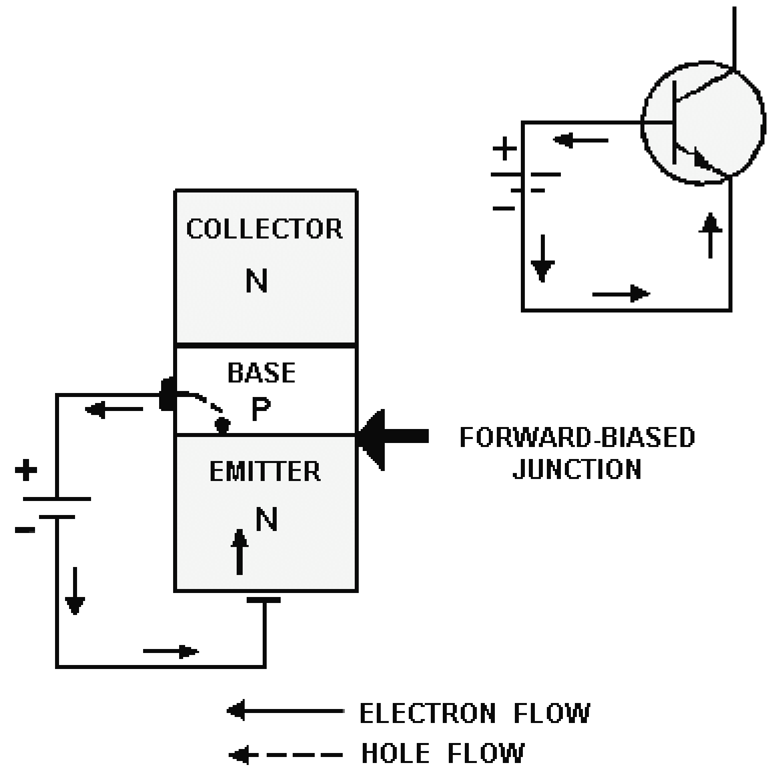


Figure 2-5.—The forward-biased junction in an NPN transistor.

With the emitter-to-base junction in the figure biased in the forward direction, electrons leave the negative terminal of the battery and enter the N material (emitter). Since electrons are majority current carriers in the N material, they pass easily through the emitter, cross over the junction, and combine with holes in the P material (base). For each electron that fills a hole in the P material, another electron will leave the P material (creating a new hole) and enter the positive terminal of the battery.

NPN REVERSE-BIASED JUNCTION.—The second PN junction (base-to-collector), or reverse-biased junction as it is called (fig. 2-6), blocks the majority current carriers from crossing the junction. However, there is a very small current, mentioned earlier, that does pass through this junction. This current is called minority current, or reverse current. As you recall, this current was produced by the electron-hole pairs. The minority carriers for the reverse-biased PN junction are the electrons in the P material and the holes in the N material. These minority carriers actually conduct the current for the reverse-biased junction when electrons from the P material enter the N material, and the holes from the N material enter the P material. However, the minority current electrons (as you will see later) play the most important part in the operation of the NPN transistor.

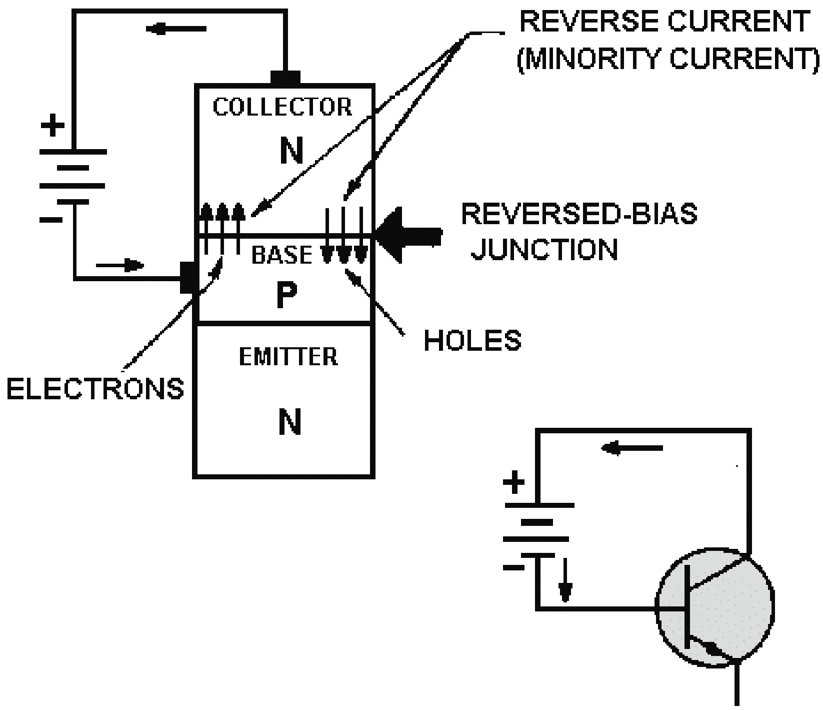


Figure 2-6.—The reverse-biased junction in an NPN transistor.

At this point you may wonder why the second PN junction (base-to-collector) is not forward biased like the first PN junction (emitter-to-base). If both junctions were forward biased, the electrons would have a tendency to flow from each end section of the N P N transistor (emitter and collector) to the center P section (base). In essence, we would have two junction diodes possessing a common base, thus eliminating any amplification and defeating the purpose of the transistor. A word of caution is in order at this time. If you should mistakenly bias the second PN junction in the forward direction, the excessive current could develop enough heat to destroy the junctions, making the transistor useless. Therefore, be sure your bias voltage polarities are correct before making any electrical connections.

NPN JUNCTION INTERACTION.—We are now ready to see what happens when we place the two junctions of the NPN transistor in operation at the same time. For a better understanding of just how the two junctions work together, refer to figure 2-7 during the discussion.

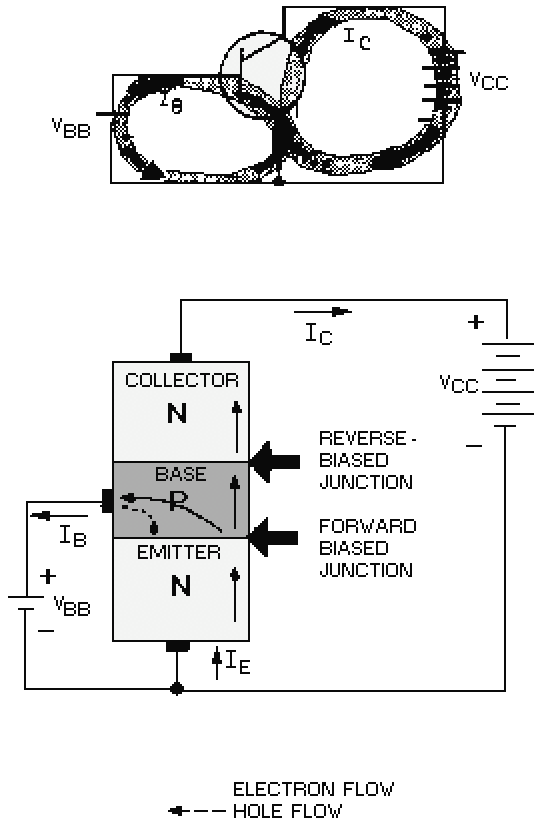


Figure 2-7.—NPN transistor operation.

The bias batteries in this figure have been labeled V_{CC} for the collector voltage supply, and V_{BB} for the base voltage supply. Also notice the base supply battery is quite small, as indicated by the number of cells in the battery, usually 1 volt or less. However, the collector supply is generally much higher than the base supply, normally around 6 volts. As you will see later, this difference in supply voltages is necessary to have current flow from the emitter to the collector.

As stated earlier, the current flow in the external circuit is always due to the movement of free electrons. Therefore, electrons flow from the negative terminals of the supply batteries to the N-type emitter. This combined movement of electrons is known as emitter current (I_E). Since electrons are the majority carriers in the N material, they will move through the N material emitter to the emitter-base junction. With this junction forward biased, electrons continue on into the base region. Once the electrons are in the base, which is a P-type material, they become minority carriers. Some of the electrons that move into the base recombine with available holes. For each electron that recombines, another electron moves out through the base lead as base current I_B (creating a new hole for eventual combination) and returns to the base supply battery V_{BB}. The electrons that recombine are lost as far as the collector is concerned. Therefore, to make the transistor more efficient, the base region is made very thin and lightly doped. This reduces the opportunity for an electron to recombine with a hole and be lost. Thus, most of the electrons that move into the base region come under the influence of the large collector reverse bias. This bias acts as forward bias for the minority carriers (electrons) in the base and, as such, accelerates them through the base-collector junction and on into the collector region. Since the collector is made of an N-type material, the electrons that reach the collector again become majority current carriers. Once in the collector, the electrons move easily through the N material and return to the positive terminal of the collector supply battery V_{CC} as collector current (I_C).

To further improve on the efficiency of the transistor, the collector is made physically larger than the base for two reasons: (1) to increase the chance of collecting carriers that diffuse to the side as well as directly across the base region, and (2) to enable the collector to handle more heat without damage.

In summary, total current flow in the NPN transistor is through the emitter lead. Therefore, in terms of percentage, I_E is 100 percent. On the other hand, since the base is very thin and lightly doped, a smaller percentage of the total current (emitter current) will flow in the base circuit than in the collector circuit. Usually no more than 2 to 5 percent of the total current is base current (I_B) while the remaining 95 to 98 percent is collector current (I_C). A very basic relationship exists between these two currents:

$$I_E = I_B + I_C$$

In simple terms this means that the emitter current is separated into base and collector current. Since the amount of current leaving the emitter is solely a function of the emitter-base bias, and because the collector receives most of this current, a small change in emitter-base bias will have a far greater effect on the magnitude of collector current than it will have on base current. In conclusion, the relatively small emitter-base bias controls the relatively large emitter-to-collector current.

Q6. To properly bias an NPN transistor, what polarity voltage is applied to the collector, and what is its relationship to the base voltage?

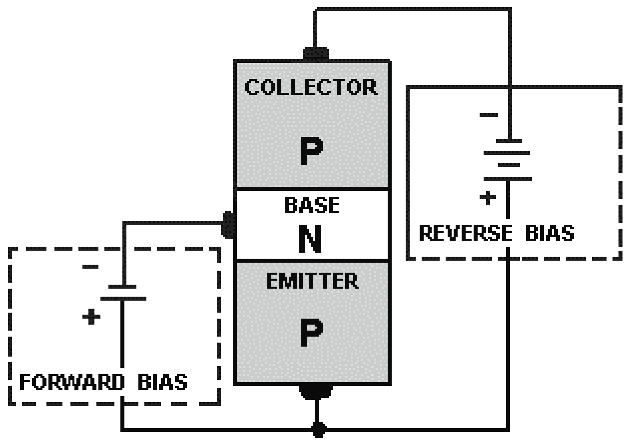
Q7. Why is conduction through the forward-biased junction of an NPN transistor primarily in one direction, namely from the emitter to base?

Q8. In the NPN transistor, what section is made very thin compared with the other two sections?

Q9. What percentage of current in an NPN transistor reaches the collector?

PNP Transistor Operation

The PNP transistor works essentially the same as the NPN transistor. However, since the emitter, base, and collector in the PNP transistor are made of materials that are different from those used in the NPN transistor, different current carriers flow in the PNP unit. The majority current carriers in the PNP transistor are holes. This is in contrast to the NPN transistor where the majority current carriers are electrons. To support this different type of current (hole flow), the bias batteries are reversed for the PNP transistor. A typical bias setup for the PNP transistor is shown in figure 2-8. Notice that the procedure used earlier to properly bias the NPN transistor also applies here to the PNP transistor. The first letter (P) in the PNP sequence indicates the polarity of the voltage required for the emitter (positive), and the second letter (N) indicates the polarity of the base voltage (negative). Since the base-collector junction is always reverse biased, then the opposite polarity voltage (negative) must be used for the collector. Thus, the base of the PNP transistor must be negative with respect to the emitter, and the collector must be more negative than the base. Remember, just as in the case of the NPN transistor, this difference in supply voltage is necessary to have current flow (hole flow in the case of the PNP transistor) from the emitter to the collector. Although hole flow is the predominant type of current flow in the PNP transistor, hole flow only takes place within the transistor itself, while electrons flow in the external circuit. However, it is the internal hole flow that leads to electron flow in the external wires connected to the transistor.



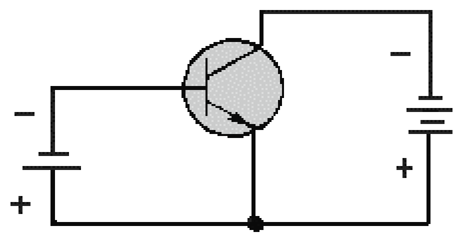


Figure 2-8.—A properly biased PNP transistor.

PNP FORWARD-BIASED JUNCTION.—Now let us consider what happens when the emitter-base junction in figure 2-9 is forward biased. With the bias setup shown, the positive terminal of the battery repels the emitter holes toward the base, while the negative terminal drives the base electrons toward the emitter. When an emitter hole and a base electron meet, they combine. For each electron that combines with a hole, another electron leaves the negative terminal of the battery, and enters the base. At the same time, an electron leaves the emitter, creating a new hole, and enters the positive terminal of the battery. This movement of electrons into the base and out of the emitter constitutes base current flow (I_B), and the path these electrons take is referred to as the emitter-base circuit.

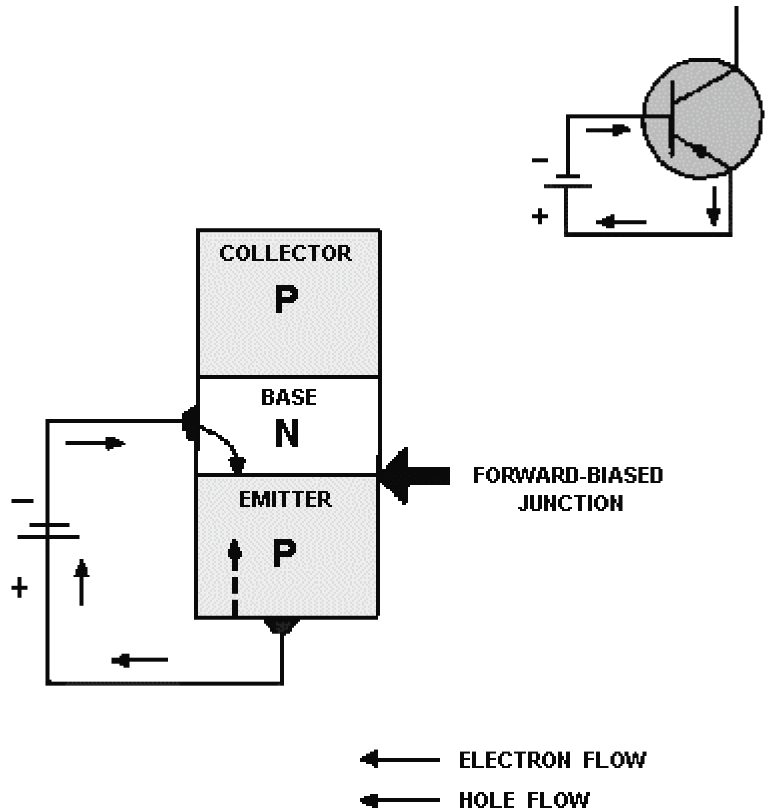


Figure 2-9.—The forward-biased junction in a PNP transistor.

PNP REVERSE-BIASED JUNCTION.—In the reverse-biased junction (fig. 2-10), the negative voltage on the collector and the positive voltage on the base block the majority current carriers from crossing the junction. However, this same negative collector voltage acts as forward bias for the minority current holes in the base, which cross the junction and enter the collector. The minority current electrons in the collector also sense forward bias-the positive base voltage-and move into the base. The holes in the collector are filled by electrons that flow from the negative terminal of the battery. At the same time the electrons leave the negative terminal of the battery, other electrons in the base break their covalent bonds and enter the positive terminal of the battery. Although there is only minority current flow in the reverse-biased junction, it is still very small because of the limited number of minority current carriers.

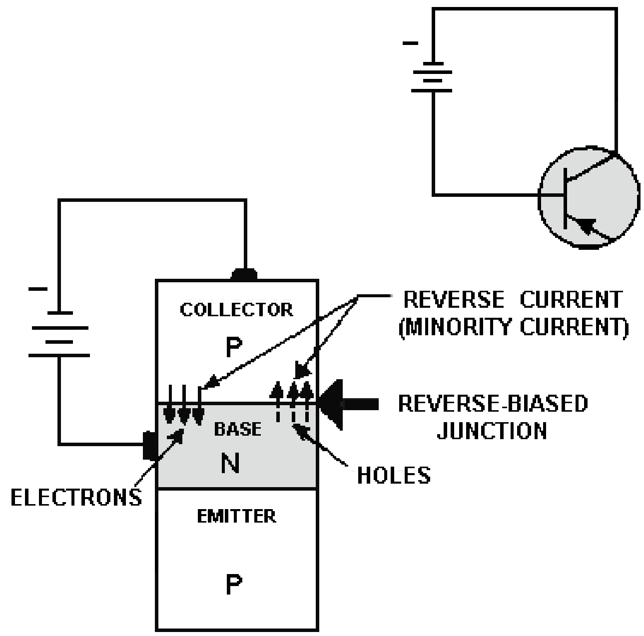


Figure 2-10.—The reverse-biased junction in a PNP transistor.

PNP JUNCTION INTERACTION.—The interaction between the forward- and reverse-biased junctions in a PNP transistor is very similar to that in an NPN transistor, except that in the PNP transistor, the majority current carriers are holes. In the PNP transistor shown in figure 2-11, the positive voltage on the emitter repels the holes toward the base. Once in the base, the holes combine with base electrons. But again, remember that the base region is made very thin to prevent the recombination of holes with electrons. Therefore, well over 90 percent of the holes that enter the base become attracted to the large negative collector voltage and pass right through the base. However, for each electron and hole that combine in the base region, another electron leaves the negative terminal of the base battery (V_{BB}) and enters the base as base current (I_B). At the same time an electron leaves the negative terminal of the battery, another electron leaves the emitter as IE (creating a new hole) and enters the positive terminal of V_{BB}. Meanwhile, in the collector circuit, electrons from the collector battery (V_{CC}) enter the collector as Ic and combine with the excess holes from the base. For each hole that is neutralized in the collector by an electron, another electron leaves the emitter and starts its way back to the positive terminal of V_{CC}.

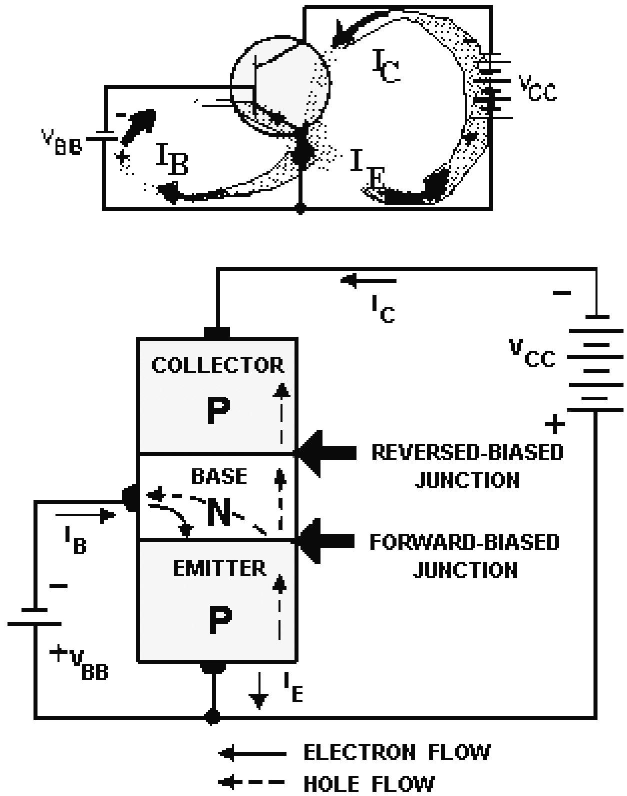


Figure 2-11.—PNP transistor operation.

Although current flow in the external circuit of the PNP transistor is opposite in direction to that of the NPN transistor, the majority carriers always flow from the emitter to the collector. This flow of majority carriers also results in the formation of two individual current loops within each transistor. One loop is the base-current path, and the other loop is the collector-current path. The combination of the current in both of these loops ($I_B + I_C$) results in total transistor current (I_E). The most important thing to remember about the two different types of transistors is that the emitter-base voltage of the PNP transistor has the same controlling effect on collector current as that of the NPN transistor. In simple terms, increasing the forward-bias voltage of a transistor reduces the emitter-base junction barrier. This action allows more carriers to reach the collector, causing an increase in current flow from the emitter to the collector and through the external circuit. Conversely, a decrease in the forward-bias voltage reduces collector current.

Q10. What are the majority current carriers in a PNP transistor?

Q11. What is the relationship between the polarity of the voltage applied to the PNP transistor and that applied to the NPN transistor?

Q12. What is the letter designation for base current?

Q13. Name the two current loops in a transistor.

THE BASIC TRANSISTOR AMPLIFIER

In the preceding pages we explained the internal workings of the transistor and introduced new terms, such as emitter, base, and collector. Since you should be familiar by now with all of the new terms mentioned earlier and with the internal operation of the transistor, we will move on to the basic transistor amplifier.

To understand the overall operation of the transistor amplifier, you must only consider the current in and out of the transistor and through the various components in the circuit. Therefore, from this point on, only the schematic symbol for the transistor will be used in the illustrations, and rather than thinking about majority and minority carriers, we will now start thinking in terms of emitter, base, and collector current.

Before going into the basic transistor amplifier, there are two terms you should be familiar with: AMPLIFICATION and AMPLIFIER. Amplification is the process of increasing the strength of a SIGNAL. A signal is just a general term used to refer to any particular current, voltage, or power in a circuit. An amplifier is the device that provides amplification (the increase in current, voltage, or power of a signal) without appreciably altering the original signal.

Transistors are frequently used as amplifiers. Some transistor circuits are CURRENT amplifiers, with a small load resistance; other circuits are designed for VOLTAGE amplification and have a high load resistance; others amplify POWER.

Now take a look at the NPN version of the basic transistor amplifier in figure 2-12 and let's see just how it works.

So far in this discussion, a separate battery has been used to provide the necessary forward-bias voltage. Although a separate battery has been used in the past for convenience, it is not practical to use a battery for emitter-base bias. For instance, it would take a battery slightly over .2 volts to properly forward bias a germanium transistor, while a similar silicon transistor would require a voltage slightly over .6 volts. However, common batteries do not have such voltage values. Also, since bias voltages are quite critical and must be held within a few tenths of one volt, it is easier to work with bias currents flowing through resistors of high ohmic values than with batteries.

By inserting one or more resistors in a circuit, different methods of biasing may be achieved and the emitter-base battery eliminated. In addition to eliminating the battery, some of these biasing methods compensate for slight variations in transistor characteristics and changes in transistor conduction resulting from temperature irregularities. Notice in figure 2-12 that the emitter-base battery has been eliminated and the bias resistor R_B has been inserted between the collector and the base. Resistor R_B provides the necessary forward bias for the emitter-base junction. Current flows in the emitter-base bias circuit from ground to the emitter, out the base lead, and through R_B to V_{CC}. Since the current in the base circuit is very small (a few hundred microamperes) and the forward resistance of the transistor is low, only a few tenths of a volt of positive bias will be felt on the base of the transistor. However, this is enough voltage on the base, along with ground on the emitter and the large positive voltage on the collector, to properly bias the transistor.

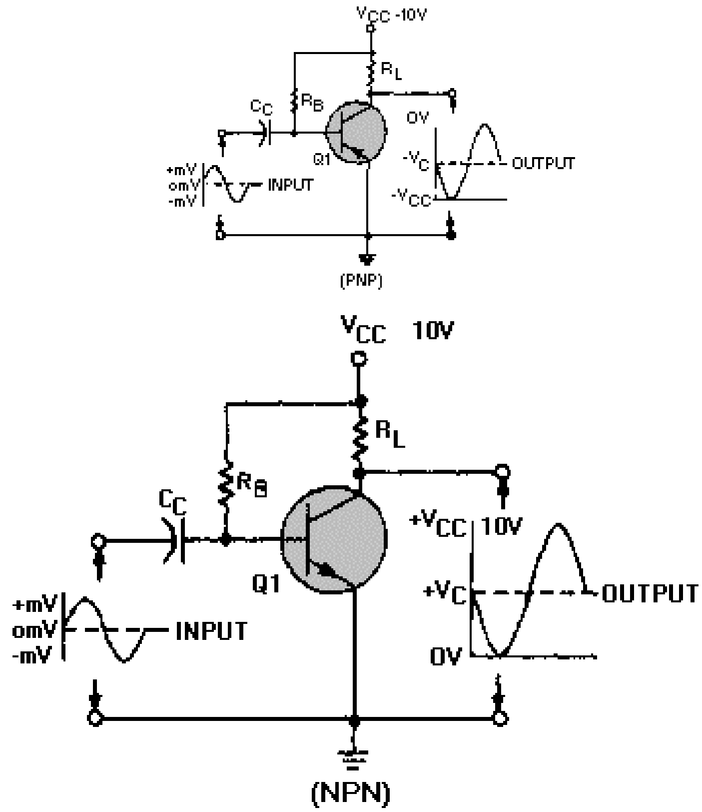


Figure 2-12.—The basic transistor amplifier.

With Q1 properly biased, direct current flows continuously, with or without an input signal, throughout the entire circuit. The direct current flowing through the circuit develops more than just base bias; it also develops the collector voltage (V_C) as it flows through Q1 and R_L. Notice the collector voltage on the output graph. Since it is present in the circuit without an input signal, the output signal starts at the V_C level and either increases or decreases. These dc voltages and currents that exist in the circuit before the application of a signal are known as QUIESCENT voltages and currents (the quiescent state of the circuit).

Resistor R_L, the collector load resistor, is placed in the circuit to keep the full effect of the collector supply voltage off the collector. This permits the collector voltage (V_C) to change with an input signal, which in turn allows the transistor to amplify voltage. Without R_L in the circuit, the voltage on the collector would always be equal to V_{CC}.

The coupling capacitor (C_C) is another new addition to the transistor circuit. It is used to pass the ac input signal and block the dc voltage from the preceding circuit. This prevents dc in the circuitry on the left of the coupling capacitor from affecting the bias on Q1. The coupling capacitor also blocks the bias of Q1 from reaching the input signal source.

The input to the amplifier is a sine wave that varies a few millivolts above and below zero. It is introduced into the circuit by the coupling capacitor and is applied between the base and emitter. As the input signal goes positive, the voltage across the emitter-base junction becomes more positive. This in effect increases forward bias, which causes base current to increase at the same rate as that of the input sine wave. Emitter and collector currents also increase but much more than the base current. With an increase in collector current, more voltage is developed across R_L. Since the voltage across R_L and the voltage across Q1 (collector to emitter) must add up to V_{CC}, an increase in voltage across R_L results in an equal decrease in voltage across Q1. Therefore, the output voltage from the amplifier, taken at the collector of Q1 with respect to the emitter, is a negative alternation of voltage that is larger than the input, but has the same sine wave characteristics.

During the negative alternation of the input, the input signal opposes the forward bias. This action decreases base current, which results in a decrease in both emitter and collector currents. The decrease in current through R_L decreases its voltage drop and causes the voltage across the transistor to rise along with the output voltage. Therefore, the output for the negative alternation of the input is a positive alternation of voltage that is larger than the input but has the same sine wave characteristics.

By examining both input and output signals for one complete alternation of the input, we can see that the output of the amplifier is an exact reproduction of the input except for the reversal in polarity and the increased amplitude (a few millivolts as compared to a few volts).

The PNP version of this amplifier is shown in the upper part of the figure. The primary difference between the NPN and PNP amplifier is the polarity of the source voltage. With a negative V_{CC}, the PNP base voltage is slightly negative with respect to ground, which provides the necessary forward bias condition between the emitter and base.

When the PNP input signal goes positive, it opposes the forward bias of the transistor. This action cancels some of the negative voltage across the emitter-base junction, which reduces the current through the transistor. Therefore, the voltage across the load resistor decreases, and the voltage across the transistor increases. Since V_{CC} is negative, the voltage on the collector (V_C) goes in a negative direction (as shown on the output graph) toward $-V_{CC}$ (for example, from -5 volts to -7 volts). Thus, the output is a negative alternation of voltage that varies at the same rate as the sine wave input, but it is opposite in polarity and has a much larger amplitude.

During the negative alternation of the input signal, the transistor current increases because the input voltage aids the forward bias. Therefore, the voltage across R_L increases, and consequently, the voltage across the transistor decreases or goes in a positive direction (for example: from −5 volts to −3 volts). This action results in a positive output voltage, which has the same characteristics as the input except that it has been amplified and the polarity is reversed.

In summary, the input signals in the preceding circuits were amplified because the small change in base current caused a large change in collector current. And, by placing resistor R_L in series with the collector, voltage amplification was achieved.

Q14. What is the name of the device that provides an increase in current, voltage, or power of a signal without appreciably altering the original signal?

Q15. Besides eliminating the emitter-base battery, what other advantages can different biasing methods offer?

Q16. In the basic transistor amplifier discussed earlier, what is the relationship between the polarity of the input and output signals?

Q17. What is the primary difference between the NPN and PNP amplifiers?

TYPES OF BIAS

One of the basic problems with transistor amplifiers is establishing and maintaining the proper values of quiescent current and voltage in the circuit. This is accomplished by selecting the proper circuit-biasing conditions and ensuring these conditions are maintained despite variations in ambient (surrounding) temperature, which cause changes in amplification and even distortion (an unwanted change in a signal). Thus a need arises for a method to properly bias the transistor amplifier and at the same time stabilize its dc operating point (the no signal values of collector voltage and collector current). As mentioned earlier, various biasing methods can be used to accomplish both of these functions. Although there are numerous biasing methods, only three basic types will be considered.

Base-Current Bias (Fixed Bias)

The first biasing method, called BASE CURRENT BIAS or sometimes FIXED BIAS, was used in figure 2-12. As you recall, it consisted basically of a resistor (R_B) connected between the collector supply voltage and the base. Unfortunately, this simple arrangement is quite thermally unstable. If the temperature of the transistor rises for any reason (due to a rise in ambient temperature or due to current flow through it), collector current will increase. This increase in current also causes the dc operating point, sometimes called the quiescent or static point, to move away from its desired position (level). This reaction to temperature is undesirable because it affects amplifier gain (the number of times of amplification) and could result in distortion, as you will see later in this discussion.

Self-Bias

A better method of biasing is obtained by inserting the bias resistor directly between the base and collector, as shown in figure 2-13. By tying the collector to the base in this manner, feedback voltage can be fed from the collector to the base to develop forward bias. This arrangement is called SELF-BIAS. Now, if an increase of temperature causes an increase in collector current, the collector voltage (V_C) will fall because of the increase of voltage produced across the load resistor (R_L). This drop in V_C will be fed back to the base and will result in a decrease in the base current. The decrease in base current will oppose the original increase in collector current and tend to stabilize it. The exact opposite effect is produced when the collector current decreases.

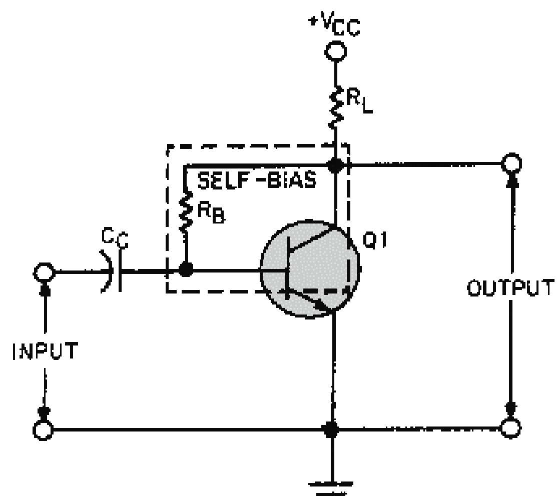


Figure 2-13.—A basic transistor amplifier with self-bias.

Self-bias has two small drawbacks: (1) It is only partially effective and, therefore, is only used where moderate changes in ambient temperature are expected; (2) it reduces amplification since the signal on the collector also affects the base voltage. This is because the collector and base signals for this particular amplifier configuration are 180 degrees out of phase (opposite in polarity) and the part of the collector signal that is fed back to the base cancels some of the input signal. This process of returning a part of the output back to its input is known as DEGENERATION or NEGATIVE FEEDBACK. Sometimes degeneration is desired to prevent amplitude distortion (an output signal that fails to follow the input exactly) and self-bias may be used for this purpose.

Combination Bias

A combination of fixed and self-bias can be used to improve stability and at the same time overcome some of the disadvantages of the other two biasing methods. One of the most widely used combination-bias systems is the voltage-divider type shown in figure 2-14. Fixed bias is provided in this circuit by the voltage-divider network consisting of R1, R2, and the collector supply voltage (V_{CC}). The dc current flowing through the voltage-divider network biases the base positive with respect to the emitter. Resistor R3, which is connected in series with the emitter, provides the emitter with self-bias. Should I_E increase, the voltage drop across R3 would also increase, reducing V_C. This reaction to an increase in I_E by R3 is another form of degeneration, which results in less output from the amplifier. However, to provide long-term or dc thermal stability, and at the same time, allow minimal ac signal degeneration, the bypass capacitor (C_{bp}) is placed across R3. If C_{bp} is large enough, rapid signal variations will not change its charge materially and no degeneration of the signal will occur.

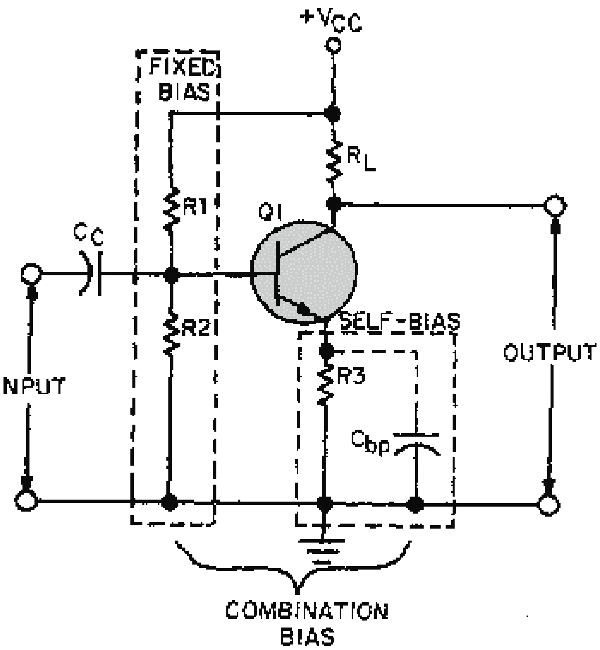


Figure 2-14.—A basic transistor amplifier with combination bias.

In summary, the fixed-bias resistors, R1 and R2, tend to keep the base bias constant while the emitter bias changes with emitter conduction. This action greatly improves thermal stability and at the same time maintains the correct operating point for the transistor.

Q18. Which biasing method is the most unstable?

Q19. What type of bias is used where only moderate changes in ambient temperature are expected?

Q20. When is degeneration tolerable in an amplifier?

Q21. What is the most widely used combination-bias system?

AMPLIFIER CLASSES OF OPERATION

In the previous discussions, we assumed that for every portion of the input signal there was an output from the amplifier. This is not always the case with amplifiers. It may be desirable to have the transistor conducting for only a portion of the input signal. The portion of the input for which there is an output determines the class of operation of the amplifier. There are four classes of amplifier operations. They are class A, class AB, class B, and class C.

Class A Amplifier Operation

Class A amplifiers are biased so that variations in input signal polarities occur within the limits of CUTOFF and SATURATION. In a PNP transistor, for example, if the base becomes positive with respect to the emitter, holes will be repelled at the PN junction and no current can flow in the collector circuit. This condition is known as cutoff. Saturation occurs when the base becomes so negative with respect to the emitter that changes in the signal are not reflected in collector-current flow.

Biasing an amplifier in this manner places the dc operating point between cutoff and saturation and allows collector current to flow during the complete cycle (360 degrees) of the input signal, thus providing an output which is a replica of the input. Figure 2-12 is an example of a class A amplifier. Although the output from this amplifier is 180 degrees out of phase with the input, the output current still flows for the complete duration of the input.

The class A operated amplifier is used as an audio- and radio-frequency amplifier in radio, radar, and sound systems, just to mention a few examples.

For a comparison of output signals for the different amplifier classes of operation, refer to figure 2-15 during the following discussion.

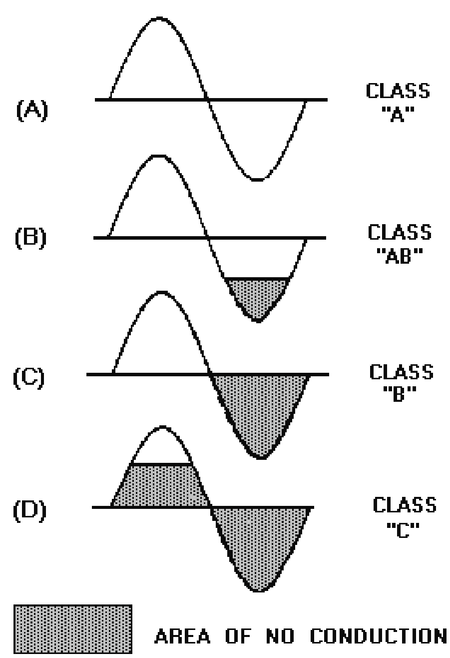


Figure 2-15.—A comparison of output signals for the different amplifier classes of operation.

Class AB Amplifier Operation

Amplifiers designed for class AB operation are biased so that collector current is zero (cutoff) for a portion of one alternation of the input signal. This is accomplished by making the forward-bias voltage less than the peak value of the input signal. By doing this, the base-emitter junction will be reverse biased during one alternation for the amount of time that the input signal voltage opposes and exceeds the value of forward-bias voltage. Therefore, collector current will flow for more than 180 degrees but less than 360 degrees of the input signal, as shown in figure 2-15 view B. As compared to the class A amplifier, the dc operating point for the class AB amplifier is closer to cutoff.

The class AB operated amplifier is commonly used as a push-pull amplifier to overcome a side effect of class B operation called crossover distortion.

Class B Amplifier Operation

Amplifiers biased so that collector current is cut off during one-half of the input signal are classified class B. The dc operating point for this class of amplifier is set up so that base current is zero with no input signal. When a signal is applied, one half cycle will forward bias the base-emitter junction and I_C will flow. The other half cycle will reverse bias the base-emitter junction and I_C will be cut off. Thus, for class B operation, collector current will flow for approximately 180 degrees (half) of the input signal, as shown in figure 2-15 view C.

The class B operated amplifier is used extensively for audio amplifiers that require high-power outputs. It is also used as the driver- and power-amplifier stages of transmitters.

Class C Amplifier Operation

In class C operation, collector current flows for less than one half cycle of the input signal, as shown in figure 2-15 view D. The class C operation is achieved by reverse biasing the emitter-base junction, which sets the dc operating point below cutoff and allows only the portion of the input signal that overcomes the reverse bias to cause collector current flow.

The class C operated amplifier is used as a radio-frequency amplifier in transmitters.

From the previous discussion, you can conclude that two primary items determine the class of operation of an amplifier — (1) the amount of bias and (2) the amplitude of the input signal. With a given input signal and bias level, you can change the operation of an amplifier from class A to class B just by removing forward bias. Also, a class A amplifier can be changed to class AB by increasing the input signal amplitude. However, if an input signal amplitude is increased to the point that the transistor goes into saturation and cutoff, it is then called an OVERDRIVEN amplifier.

You should be familiar with two terms used in conjunction with amplifiers — FIDELITY and EFFICIENCY. Fidelity is the faithful reproduction of a signal. In other words, if the output of an amplifier is just like the input except in amplitude, the amplifier has a high degree of fidelity. The opposite of fidelity is a term we mentioned earlier — distortion. Therefore, a circuit that has high fidelity has low distortion. In conclusion, a class A amplifier has a high degree of fidelity. A class AB amplifier has less fidelity, and class B and class C amplifiers have low or "poor" fidelity.

The efficiency of an amplifier refers to the ratio of output-signal power compared to the total input power. An amplifier has two input power sources: one from the signal, and one from the power supply. Since every device takes power to operate, an amplifier that operates for 360 degrees of the input signal uses more power than if operated for 180 degrees of the input signal. By using more power, an amplifier has less power available for the output signal; thus the efficiency of the amplifier is low. This is the case with the class A amplifier. It operates for 360 degrees of the input signal and requires a relatively large input from the power supply. Even with no input signal, the class A amplifier still uses power from the power supply. Therefore, the output from the class A amplifier is relatively small compared to the total input power. This results in low efficiency, which is acceptable in class A amplifiers because they are used where efficiency is not as important as fidelity.

Class AB amplifiers are biased so that collector current is cut off for a portion of one alternation of the input, which results in less total input power than the class A amplifier. This leads to better efficiency.

Class B amplifiers are biased with little or no collector current at the dc operating point. With no input signal, there is little wasted power. Therefore, the efficiency of class B amplifiers is higher still.

The efficiency of class C is the highest of the four classes of amplifier operations.

Q22. What amplifier class of operation allows collector current to flow during the complete cycle of the input?

Q23. What is the name of the term used to describe the condition in a transistor when the emitter-base junction has zero bias or is reverse biased and there is no collector current?

Q24. What two primary items determine the class of operation of an amplifier?

Q25. What amplifier class of operation is the most inefficient but has the least distortion?

TRANSISTOR CONFIGURATIONS

A transistor may be connected in any one of three basic configurations (fig. 2-16): common emitter (CE), common base (CB), and common collector (CC). The term common is used to denote the element that is common to both input and output circuits. Because the common element is often grounded, these configurations are frequently referred to as grounded emitter, grounded base, and grounded collector.

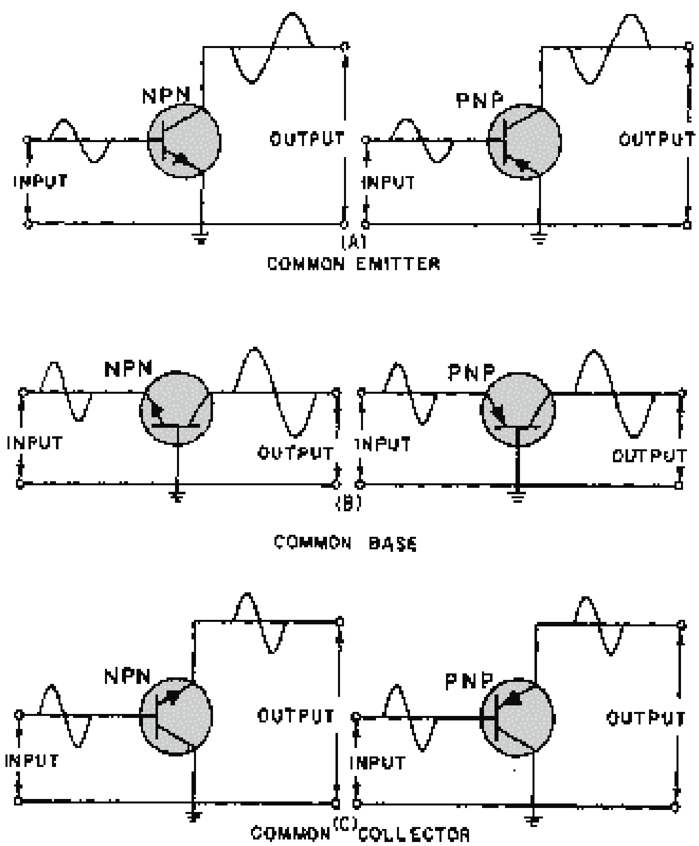


Figure 2-16.—Transistor configurations.

Each configuration, as you will see later, has particular characteristics that make it suitable for specific applications. An easy way to identify a specific transistor configuration is to follow three simple steps:

1. Identify the element (emitter, base, or collector) to which the input signal is applied.

2. Identify the element (emitter, base, or collector) from which the output signal is taken.

3. The remaining element is the common element, and gives the configuration its name.

Therefore, by applying these three simple steps to the circuit in figure 2-12, we can conclude that this circuit is more than just a basic transistor amplifier. It is a common-emitter amplifier.

Common Emitter

The common-emitter configuration (CE) shown in figure 2-16 view A is the arrangement most frequently used in practical amplifier circuits, since it provides good voltage, current, and power gain. The common emitter also has a somewhat low input resistance (500 ohms-1500 ohms), because the input is applied to the forward-biased junction, and a moderately high output resistance (30 kilohms-50 kilohms or more), because the output is taken off the reverse-biased junction. Since the input signal is applied to the base-emitter circuit and the output is taken from the collector-emitter circuit, the emitter is the element common to both input and output.

Since you have already covered what you now know to be a common-emitter amplifier (fig. 2-12), let's take a few minutes and review its operation, using the PNP common-emitter configuration shown in figure 2-16 view A.

When a transistor is connected in a common-emitter configuration, the input signal is injected between the base and emitter, which is a low resistance, low-current circuit. As the input signal swings positive, it also causes the base to swing positive with respect to the emitter. This action decreases forward bias which reduces collector current (I_C) and increases collector voltage (making V_C more negative). During the negative alternation of the input signal, the base is driven more negative with respect to the emitter. This increases forward bias and allows more current carriers to be released from the emitter, which results in an increase in collector current and a decrease in collector voltage (making V_C less negative or swing in a positive direction). The collector current that flows through the high resistance reverse-biased junction also flows through a high resistance load (not shown), resulting in a high level of amplification.

Since the input signal to the common emitter goes positive when the output goes negative, the two signals (input and output) are 180 degrees out of phase. The common-emitter circuit is the only configuration that provides a phase reversal.

The common-emitter is the most popular of the three transistor configurations because it has the best combination of current and voltage gain. The term *GAIN* is used to describe the amplification capabilities of the amplifier. It is basically a ratio of output versus input. Each transistor configuration gives a different value of gain even though the same transistor is used. The transistor configuration used is a matter of design consideration. However, as a technician you will become interested in this output versus input ratio (gain) to determine whether or not the transistor is working properly in the circuit.

The current gain in the common-emitter circuit is called BETA (β). Beta is the relationship of collector current (output current) to base current (input current). To calculate beta, use the following formula:

$$\beta = \frac{\Delta I_C}{\Delta I_B}$$

(Δ is the Greek letter delta, it is used to indicate a small change)

For example, if the input current (I_B) in a common emitter changes from 75 uA to 100 uA and the output current (I_C) changes from 1.5 mA to 2.6 mA, the current gain (β) will be 44.

$$\beta = \frac{\Delta I_C}{\Delta I_B} = \frac{11 \times 10^{-3}}{25 \times 10^{-6}} = 44$$

This simply means that a change in base current produces a change in collector current which is 44 times as large.

You may also see the term h_{fe} used in place of β. The terms h_{fe} and β are equivalent and may be used interchangeably. This is because "h_{fe}" means:

h = hybrid (meaning mixture)

f = forward current transfer ratio

e = common emitter configuration

The resistance gain of the common emitter can be found in a method similar to the one used for finding beta:

$$R = \frac{R_{out}}{R_{in}}$$

Once the resistance gain is known, the voltage gain is easy to calculate since it is equal to the current gain (β) multiplied by the resistance gain ($E = \beta R$). And, the power gain is equal to the voltage gain multiplied by the current gain β ($P = \beta E$).

Common Base

The common-base configuration (CB) shown in figure 2-16, view B is mainly used for impedance matching, since it has a low input resistance (30 ohms-160 ohms) and a high output resistance (250 kilohms-550 kilohms). However, two factors limit its usefulness in some circuit applications: (1) its low input resistance and (2) its current gain of less than 1. Since the CB configuration will give voltage amplification, there are some additional applications, which require both a low-input resistance and voltage amplification, that could use a circuit configuration of this type; for example, some microphone amplifiers.

In the common-base configuration, the input signal is applied to the emitter, the output is taken from the collector, and the base is the element common to both input and output. Since the input is applied to the emitter, it causes the emitter-base junction to react in the same manner as it did in the common-emitter circuit. For example, an input that aids the bias will increase transistor current, and one that opposes the bias will decrease transistor current.

Unlike the common-emitter circuit, the input and output signals in the common-base circuit are in phase. To illustrate this point, assume the input to the PNP version of the common-base circuit in figure 2-16 view B is positive. The signal adds to the forward bias, since it is applied to the emitter, causing the collector current to increase. This increase in Ic results in a greater voltage drop across the load resistor R_L (not shown), thus lowering the collector voltage V_C. The collector voltage, in becoming less negative, is swinging in a positive direction, and is therefore in phase with the incoming positive signal.

The current gain in the common-base circuit is calculated in a method similar to that of the common emitter except that the input current is I_E not I_B and the term ALPHA (α) is used in place of beta for gain. Alpha is the relationship of collector current (output current) to emitter current (input current). Alpha is calculated using the formula:

$$\alpha = \frac{\Delta I_C}{\Delta I_E}$$

For example, if the input current (I_E) in a common base changes from 1 mA to 3 mA and the output current (I_C) changes from 1 mA to 2.8 mA, the current gain (α) will be 0.90 or:

$$\alpha = \frac{\Delta I_C}{\Delta I_E} = \frac{18 \times 10^{-3}}{2 \times 10^{-3}} = 0.90$$

This is a current gain of less than 1.

Since part of the emitter current flows into the base and does not appear as collector current, collector current will always be less than the emitter current that causes it. (Remember, $I_E = I_B + I_C$) Therefore, ALPHA is ALWAYS LESS THAN ONE FOR A COMMON-BASE CONFIGURATION.

Another term for "α" is h_f. These terms (and h_f) are equivalent and may be used interchangeably. The meaning for the term h_f is derived in the same manner as the term h_{fe} mentioned earlier, except that the last letter "e" has been replaced with "b" to stand for common- base configuration.

Many transistor manuals and data sheets only list transistor current gain characteristics in terms of β or h_{fe}. To find alpha (α) when given beta (β), use the following formula to convert β to α for use with the common-base configuration:

$$\alpha = \frac{\beta}{\beta + 1}$$

To calculate the other gains (voltage and power) in the common-base configuration when the current gain (α) is known, follow the procedures described earlier under the common-emitter section.

Common Collector

The common-collector configuration (CC) shown in figure 2-16 view C is used mostly for impedance matching. It is also used as a current driver, because of its substantial current gain. It is particularly useful in switching circuitry, since it has the ability to pass signals in either direction (bilateral operation).

In the common-collector circuit, the input signal is applied to the base, the output is taken from the emitter, and the collector is the element common to both input and output. The common collector is equivalent to our old friend the electron-tube cathode follower. Both have high input and low output resistance. The input resistance for the common collector ranges from 2 kilohms to 500 kilohms, and the output resistance varies from 50 ohms to 1500 ohms. The current gain is higher than that in the common emitter, but it has a lower power gain than either the common base or common emitter. Like the common base, the output signal from the common collector is in phase with the input signal. The common collector is also referred to as an emitter-follower because the output developed on the emitter follows the input signal applied to the base.

Transistor action in the common collector is similar to the operation explained for the common base, except that the current gain is not based on the emitter-to-collector current ratio, alpha (α). Instead, it is based on the emitter-to-base current ratio called GAMMA (γ), because the output is taken off the emitter. Since a small change in base current controls a large change in emitter current, it is still possible to obtain high current gain in the common collector. However, since the emitter current gain is offset by the low output resistance, the voltage gain is always less than 1 (unity), exactly as in the electron-tube cathode follower

The common-collector current gain, gamma (γ), is defined as

$$\gamma = \frac{\Delta I_E}{\Delta I_B}$$

and is related to collector-to-base current gain, beta (β), of the common-emitter circuit by the formula:

$$\gamma = \beta + 1$$

Since a given transistor may be connected in any of three basic configurations, there is a definite relationship, as pointed out earlier, between alpha (α), beta (β), and gamma (γ). These relationships are listed again for your convenience:

$$\alpha = \frac{\beta}{\beta + 1} \quad \beta = \frac{\alpha}{1 - \alpha} \quad \gamma = \beta + 1$$

Take, for example, a transistor that is listed on a manufacturer's data sheet as having an alpha of 0.90. We wish to use it in a common emitter configuration. This means we must find beta. The calculations are:

$$\beta = \frac{\alpha}{1 - \alpha} = \frac{0.90}{1 - 0.90} = \frac{0.90}{0.1} = 9$$

Therefore, a change in base current in this transistor will produce a change in collector current that will be 9 times as large.

If we wish to use this same transistor in a common collector, we can find gamma (γ) by:

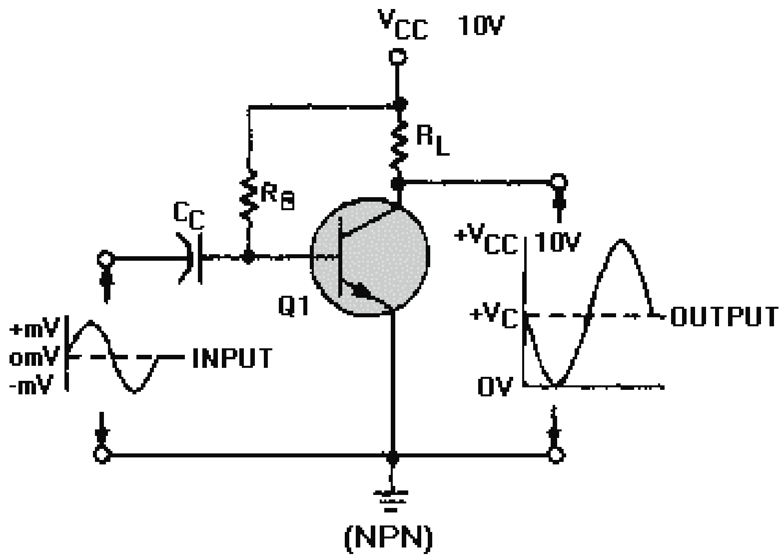
$$\gamma = \beta + 1 = 9 + 1 = 10$$

To summarize the properties of the three transistor configurations, a comparison chart is provided in table 2-1 for your convenience.

Table 2-1.—Transistor Configuration Comparison Chart

AMPLIFIER TYPE	COMMON BASE	COMMON EMITTER	COMMON COLLECTOR
INPUT/OUTPUT PHASE RELATIONSHIP	0º	180º	0º
VOLTAGE GAIN	HIGH	MEDIUM	LOW
CURRENT GAIN	LOW(α)	MEDIUM(β)	HIGH(γ)
POWER GAIN	LOW	HIGH	MEDIUM
INPUT RESISTANCE	LOW	MEDIUM	HIGH
OUTPUT RESISTANCE	HIGH	MEDIUM	LOW

Now that we have analyzed the basic transistor amplifier in terms of bias, class of operation, and circuit configuration, let's apply what has been covered to figure 2-12. A reproduction of figure 2-12 is shown below for your convenience.



This illustration is not just the basic transistor amplifier shown earlier in figure 2-12 but a class A amplifier configured as a common emitter using fixed bias. From this, you should be able to conclude the following:

- Because of its fixed bias, the amplifier is thermally unstable.
- Because of its class A operation, the amplifier has low efficiency but good fidelity.
- Because it is configured as a common emitter, the amplifier has good voltage, current, and power gain.

In conclusion, the type of bias, class of operation, and circuit configuration are all clues to the function and possible application of the amplifier.

Q26. What are the three transistor configurations?

Q27. Which transistor configuration provides a phase reversal between the input and output signals?

Q28. What is the input current in the common-emitter circuit?

Q29. What is the current gain in a common-base circuit called?

Q30. Which transistor configuration has a current gain of less than 1?

Q31. What is the output current in the common-collector circuit?

Q32. Which transistor configuration has the highest input resistance?

Q33. What is the formula for GAMMA (γ)?

TRANSISTOR SPECIFICATIONS

Transistors are available in a large variety of shapes and sizes, each with its own unique characteristics. The characteristics for each of these transistors are usually presented on SPECIFICATION SHEETS or they may be included in transistor manuals. Although many properties of a transistor could be specified on these sheets, manufacturers list only some of them. The specifications listed vary with different manufacturers, the type of transistor, and the application of the transistor. The specifications usually cover the following items.

1. A general description of the transistor that includes the following information:

    a. The kind of transistor. This covers the material used, such as germanium or silicon; the type of transistor (NPN or PNP); and the construction of the transistor(whether alloy-junction, grown, or diffused junction, etc.).

    b. Some of the common applications for the transistor, such as audio amplifier, oscillator, rf amplifier, etc.

    c. General sales features, such as size and packaging mechanical data).

2. The "Absolute Maximum Ratings" of the transistor are the direct voltage and current values that if exceeded in operation may result in transistor failure. Maximum ratings usually include collector-to-base voltage, emitter-to-base voltage, collector current, emitter current, and collector power dissipation.

3. The typical operating values of the transistor. These values are presented only as a guide. The values vary widely, are dependent upon operating voltages, and also upon which element is common in the circuit. The values listed may include collector-emitter voltage, collector current, input resistance, load resistance, current-transfer ratio (another name for alpha or beta), and collector cutoff current, which is leakage current from collector to base when no emitter current is applied. Transistor characteristic curves may also be included in this section. A transistor characteristic curve is a graph plotting the relationship between currents and voltages in a circuit. More than one curve on a graph is called a "family of curves."

4. Additional information for engineering-design purposes.

So far, many letter symbols, abbreviations, and terms have been introduced, some frequently used and others only rarely used. For a complete list of all semiconductor letter symbols and terms, refer to EIMB series 000-0140, Section III.

TRANSISTOR IDENTIFICATION

Transistors can be identified by a Joint Army-Navy (JAN) designation printed directly on the case of the transistor. The marking scheme explained earlier for diodes is also used for transistor identification. The first number indicates the number of junctions. The letter "N" following the first number tells us that the component is a semiconductor. And, the 2- or 3-digit number following the N is the manufacturer's identification number. If the last number is followed by a letter, it indicates a later, improved version of the device. For example, a semiconductor designated as type 2N130A signifies a three-element transistor of semiconductor material that is an improved version of type 130:

2	N	130	A
NUMBER OF JUNCTIONS (TRANSISTOR)	SEMI-CONDUCTOR	IDENTIFICATION NUMBER	FIRST MODIFICATION

You may also find other markings on transistors that do not relate to the JAN marking system. These markings are manufacturers' identifications and may not conform to a standardized system. If in doubt, always replace a transistor with one having identical markings. To ensure that an identical replacement or a correct substitute is used, consult an equipment or transistor manual for specifications on the transistor.

TRANSISTOR MAINTENANCE

Transistors are very rugged and are expected to be relatively trouble free. Encapsulation and conformal coating techniques now in use promise extremely long life expectancies. In theory, a transistor should last indefinitely. However, if transistors are subjected to current overloads, the junctions will be damaged or even destroyed. In addition, the application of excessively high operating voltages can damage or destroy the junctions through arc-over or excessive reverse currents. One of the greatest dangers to the transistor is heat, which will cause excessive current flow and eventual destruction of the transistor.

To determine if a transistor is good or bad, you can check it with an ohmmeter or a transistor tester. In many cases, you can substitute a transistor known to be good for one that is questionable and thus determine the condition of a suspected transistor. This method of testing is highly accurate and sometimes the quickest, but it should be used only after you make certain that there are no circuit defects that might damage the replacement transistor. If more than one defective transistor is present in the equipment where the trouble has been localized, this testing method becomes cumbersome, as several transistors may have to be replaced before the trouble is corrected. To determine which stages failed and which transistors are not defective, all the removed transistors must be tested. This test can be made by using a standard Navy ohmmeter, transistor tester, or by observing whether the equipment operates correctly as each of the removed transistors is reinserted into the equipment. A word of caution-indiscriminate substitution of transistors in critical circuits should be avoided.

When transistors are soldered into equipment, substitution is not practicable; it is generally desirable to test these transistors in their circuits.

Q34. List three items of information normally included in the general description section of a specification sheet for a transistor.

Q35. What does the number "2" (before the letter "N") indicate in the JAN marking scheme?

Q36. What is the greatest danger to a transistor?

Q37. What method for checking transistors is cumbersome when more than one transistor is bad in a circuit?

PRECAUTIONS

Transistors, although generally more rugged mechanically than electron tubes, are susceptible to damage by electrical overloads, heat, humidity, and radiation. Damage of this nature often occurs during transistor servicing by applying the incorrect polarity voltage to the collector circuit or excessive voltage to the input circuit. Careless soldering techniques that overheat the transistor have also been known to cause considerable damage. One of the most frequent causes of damage to a transistor is the electrostatic discharge from the human body when the device is handled. You may avoid such damage before starting repairs by discharging the static electricity from your body to the chassis containing the transistor. You can do this by simply touching the chassis. Thus, the electricity will be transferred from your body to the chassis before you handle the transistor.

To prevent transistor damage and avoid electrical shock, you should observe the following precautions when you are working with transistorized equipment:

1. Test equipment and soldering irons should be checked to make certain there is no leakage current from the power source. If leakage current is detected, isolation transformers should be used.

2. Always connect a ground between test equipment and circuit before attempting to inject or monitor a signal.

3. Ensure test voltages do not exceed maximum allowable voltage for circuit components and transistors. Also, never connect test equipment outputs directly to a transistor circuit.

4. Ohmmeter ranges that require a current of more than one milliampere in the test circuit should not be used for testing transistors.

5. Battery eliminators should not be used to furnish power for transistor equipment because they have poor voltage regulation and, possibly, high-ripple voltage.

6. The heat applied to a transistor, when soldered connections are required, should be kept to a minimum by using a low-wattage soldering iron and heat shunts, such as long-nose pliers, on the transistor leads.

7. When it becomes necessary to replace transistors, never pry transistors to loosen them from printed circuit boards.

8. All circuits should be checked for defects before replacing a transistor.

9. The power must be removed from the equipment before replacing a transistor.

10. Using conventional test probes on equipment with closely spaced parts often causes accidental shorts between adjacent terminals. These shorts rarely cause damage to an electron tube but may ruin a transistor. To prevent these shorts, the probes can be covered with insulation, except for a very short length of the tips.

LEAD IDENTIFICATION

Transistor lead identification plays an important part in transistor maintenance; because, before a transistor can be tested or replaced, its leads or terminals must be identified. Since there is <u>no</u> standard method of identifying transistor leads, it is quite possible to mistake one lead for another. Therefore, when you are replacing a transistor, you should pay close attention to how the transistor is mounted, particularly to those transistors that are soldered in, so that you do not make a mistake when you are installing the new transistor. When you are testing or replacing a transistor, if you have any doubts about which lead is which, consult the equipment manual or a transistor manual that shows the specifications for the transistor being used.

There are, however, some typical lead identification schemes that will be very helpful in transistor troubleshooting. These schemes are shown in figure 2-17. In the case of the oval-shaped transistor shown in view A, the collector lead is identified by a wide space between it and the base lead. The lead farthest from the collector, in line, is the emitter lead. When the leads are evenly spaced and in line, as shown in view B, a colored dot, usually red, indicates the collector. If the transistor is round, as in view C, a red line indicates the collector, and the emitter lead is the shortest lead. In view D the leads are in a triangular arrangement that is offset from the center of the transistor. The lead opposite the blank quadrant in this scheme is the base lead. When viewed from the bottom, the collector is the first lead clockwise from the base. The leads in view E are arranged in the same manner as those is view D except that a tap is used to identify the leads. When viewed from the bottom in a clockwise direction, the first lead following the tab is the emitter, followed by the base and collector.

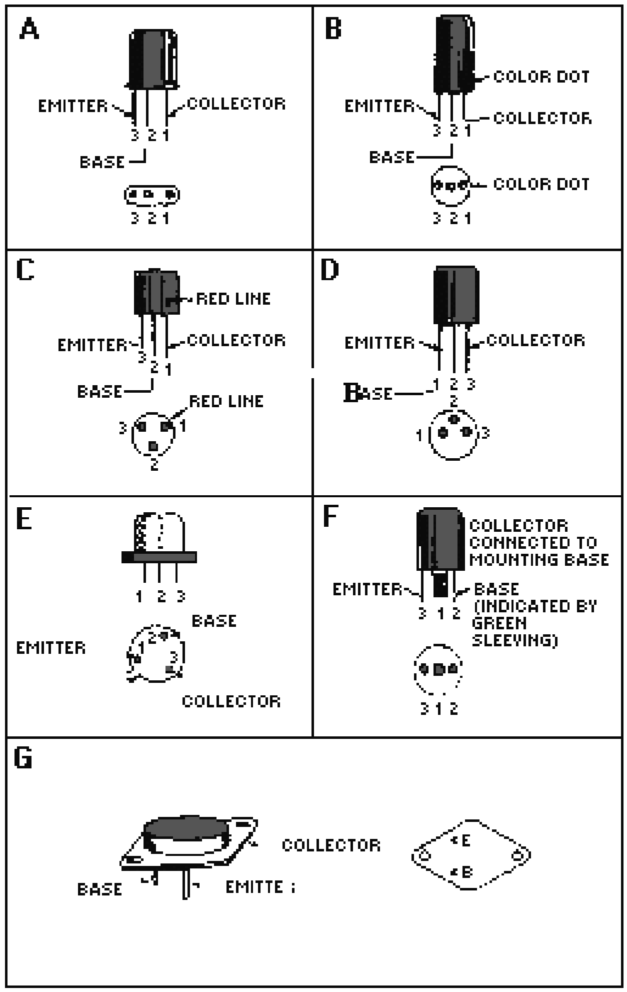


Figure 2-17.—Transistor lead identification.

In a conventional power transistor as shown in views F and G, the collector lead is usually connected to the mounting base. For further identification, the base lead in view F is covered with green sleeving. While the leads in view G are identified by viewing the transistor from the bottom in a clockwise direction (with mounting holes occupying 3 o'clock and 9 o'clock positions), the emitter lead will be either at the 5 o'clock or 11 o'clock position. The other lead is the base lead.

TRANSISTOR TESTING

There are several different ways of testing transistors. They can be tested while in the circuit, by the substitution method mentioned, or with a transistor tester or ohmmeter.

Transistor testers are nothing more than the solid-state equivalent of electron-tube testers (although they do not operate on the same principle). With most transistor testers, it is possible to test the transistor in or out of the circuit.

There are four basic tests required for transistors in practical troubleshooting: gain, leakage, breakdown, and switching time. For maintenance and repair, however, a check of two or three parameters is usually sufficient to determine whether a transistor needs to be replaced.

Since it is impractical to cover all the different types of transistor testers and since each tester comes with its own operator's manual, we will move on to something you will use more frequently for testing transistors-the ohmmeter.

Testing Transistors with an Ohmmeter

Two tests that can be done with an ohmmeter are gain, and junction resistance. Tests of a transistor's junction resistance will reveal leakage, shorts, and opens.

TRANSISTOR GAIN TEST.—A basic transistor gain test can be made using an ohmmeter and a simple test circuit. The test circuit can be made with just a couple of resistors and a switch, as shown in figure 2-18. The principle behind the test lies in the fact that little or no current will flow in a transistor between emitter and collector until the emitter-base junction is forward biased. The only precaution you should observe is with the ohmmeter. Any internal battery may be used in the meter provided that it does not exceed the maximum collector-emitter breakdown voltage.

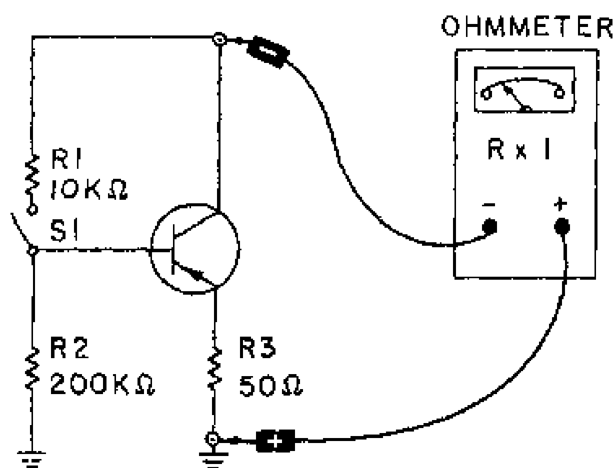


Figure 2-18.—Testing a transistor's gain with an ohmmeter.

With the switch in figure 2-18 in the open position as shown, no voltage is applied to the PNP transistor's base, and the emitter-base junction is not forward biased. Therefore, the ohmmeter should read a high resistance, as indicated on the meter. When the switch is closed, the emitter-base circuit is forward biased by the voltage across R1 and R2. Current now flows in the emitter-collector circuit, which causes a lower resistance reading on the ohmmeter. A 10-to-1 resistance ratio in this test between meter readings indicates a normal gain for an audio-frequency transistor.

To test an NPN transistor using this circuit, simply reverse the ohmmeter leads and carry out the procedure described earlier.

TRANSISTOR JUNCTION RESISTANCE TEST.—An ohmmeter can be used to test a transistor for leakage (an undesirable flow of current) by measuring the base-emitter, base-collector, and collector-emitter forward and reverse resistances.

For simplicity, consider the transistor under test in each view of figure 2-19 (view A, view B and view C) as two diodes connected back to back. Therefore, each diode will have a low forward resistance and a high reverse resistance. By measuring these resistances with an ohmmeter as shown in the figure, you can determine if the transistor is leaking current through its junctions. When making these measurements, avoid using the R1 scale on the meter or a meter with a high internal battery voltage. Either of these conditions can damage a low-power transistor.

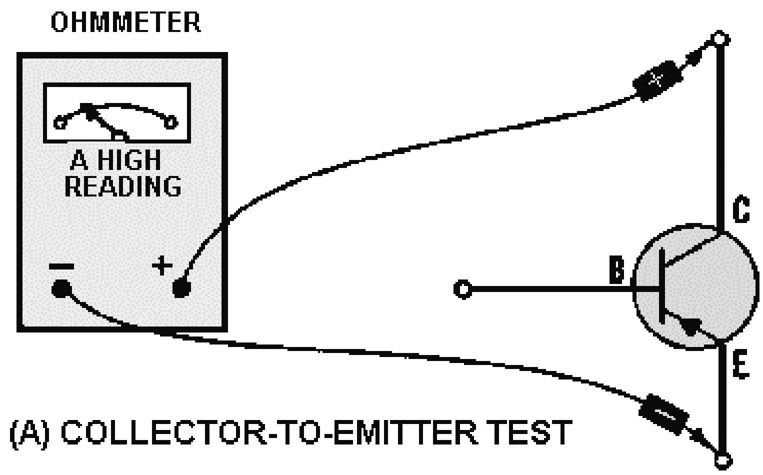


Figure 2-19A.—Testing a transistor's leakage with an ohmmeter. COLLECTOR-TO-EMITTER TEST

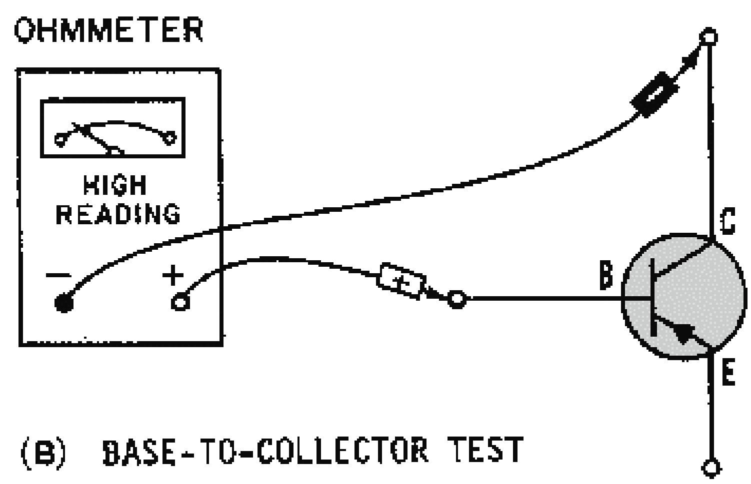


Figure 2-19B.—Testing a transistor's leakage with an ohmmeter. BASE-TO-COLLECTOR TEST

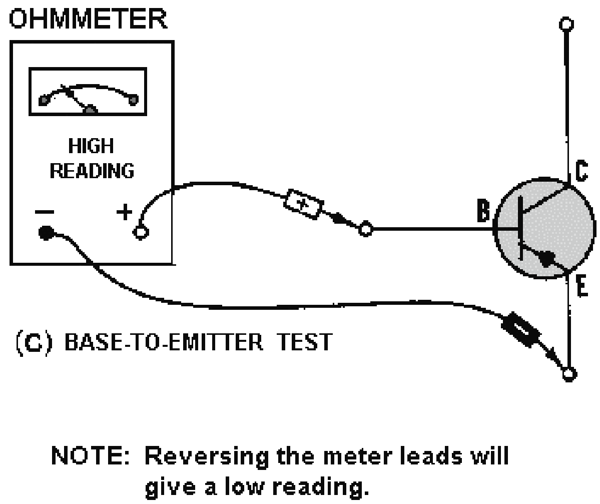


Figure 2-19C.—Testing a transistor's leakage with an ohmmeter. BASE-TO-EMITTER TEST

Now consider the possible transistor problems that could exist if the indicated readings in figure 2-19 are not obtained. A list of these problems is provided in table 2-2.

Table 2-2.—Possible Transistor Problems from Ohmmeter Readings

RESISTANCE READINGS		PROBLEMS
FORWARD	REVERSE	The transistor is:
LOW (NOT SHORTED)	LOW (NOT SHORTED)	LEAKING
LOW (SHORTED)	LOW (SHORTED)	SHORTED
HIGH	HIGH	OPEN*
*Except collector-to-emitter test.		

By now, you should recognize that the transistor used in figure 2-19 (view A, view B and view C) is a PNP transistor. If you wish to test an NPN transistor for leakage, the procedure is identical to that used for testing the PNP except the readings obtained are reversed.

When testing transistors (PNP or NPN), you should remember that the actual resistance values depend on the ohmmeter scale and the battery voltage. Typical forward and reverse resistances are insignificant. The best indicator for showing whether a transistor is good or bad is the <u>ratio of forward-to-reverse resistance</u>. If the transistor you are testing shows a ratio of at least 30 to 1, it is probably good. Many transistors show ratios of 100 to 1 or greater.

Q38. What safety precaution must be taken before replacing a transistor?

Q39. How is the collector lead identified on an oval-shaped transistor?

Q40. What are two transistor tests that can be done with an ohmmeter?

Q41. When you are testing the gain of an audio-frequency transistor with an ohmmeter, what is indicated by a 10-to-1 resistance ratio?

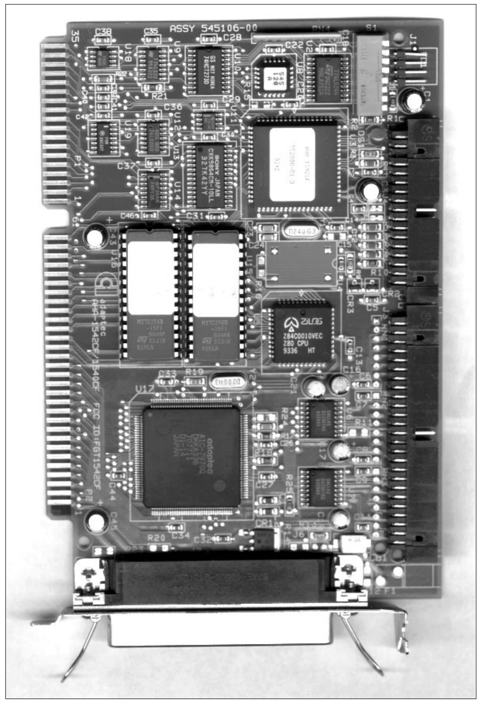
Q42. When you are using an ohmmeter to test a transistor for leakage, what is indicated by a low, but not shorted, reverse resistance reading?

MICROELECTRONICS

Up to now the various semiconductors, resistors, capacitors, etc., in our discussions have been considered as separately packaged components, called DISCRETE COMPONENTS. In this section we will introduce some of the more complex devices that contain complete circuits packaged as a single component. These devices are referred to as INTEGRATED CIRCUITS and the broad term used to describe the use of these devices to miniaturize electronic equipment is called MICROELECTRONICS.

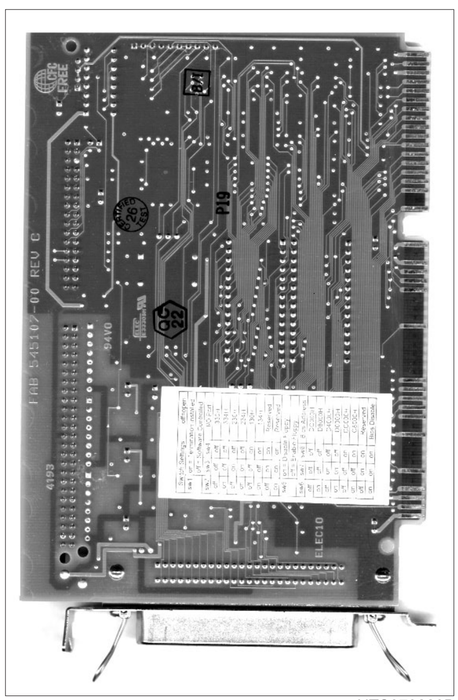
With the advent of the transistor and the demand by the military for smaller equipment, design engineers set out to miniaturize electronic equipment. In the beginning, their efforts were frustrated because most of the other components in a circuit such as resistors, capacitors, and coils were larger than the transistor. Soon these other circuit components were miniaturized, thereby pushing ahead the development of smaller electronic equipment. Along with miniature resistors, capacitors, and other circuit elements, the production of components that were actually smaller than the space required for the interconnecting wiring and cabling became possible. The next step in the research process was to eliminate these bulky wiring components. This was accomplished with the PRINTED CIRCUIT BOARD (PCB).

A printed circuit board is a flat insulating surface upon which printed wiring and miniaturized components are connected in a predetermined design, and attached to a common base. Figure 2-20 (view A and view B) shows a typical printed circuit board. Notice that various components are connected to the board and the printed wiring is on the reverse side. With this technique, all interconnecting wiring in a piece of equipment, except for the highest power leads and cabling, is reduced to lines of conducting material (copper, silver, gold, etc.) deposited directly on the surface of an insulating "circuit board." Since printed circuit boards are readily adapted as plug-in units, the elimination of terminal boards, fittings and tie points, not to mention wires, results in a substantial reduction in the overall size of electronic equipment.



A. FRONT SIDE NTS070220A

Figure 2-20A.—A typical printed circuit board (PCB). FRONT SIDE



NTS070220B
B. REVERSE SIDE

Figure 2-20B.—A typical printed circuit board (PCB). REVERSE SIDE

After the printed circuit boards were perfected, efforts to miniaturize electronic equipment were then shifted to assembly techniques, which led to MODULAR CIRCUITRY. In this technique, printed circuit boards are stacked and connected together to form a module. This increases the packaging density of circuit components and results in a considerable reduction in the size of electronic equipment. Since the module can be designed to perform any electronic function, it is also a very versatile unit.

However, the drawback to this approach was that the modules required a considerable number of connections that took up too much space and increased costs. In addition, tests showed the reliability was adversely affected by the increase in the number of connections.

A new technique was required to improve reliability and further increase packaging density. The solution was INTEGRATED CIRCUITS.

An integrated circuit is a device that integrates (combines) both active components (transistors, diodes, etc.) and passive components (resistors, capacitors, etc.) of a complete electronic circuit in a single chip (a tiny slice or wafer of semiconductor crystal or insulator).

Integrated circuits (ICs) have almost eliminated the use of individual electronic components (resistors, capacitors, transistors, etc.) as the building blocks of electronic circuits. Instead, tiny CHIPS have been developed whose functions are not that of a single part, but of dozens of transistors, resistors, capacitors, and other electronic elements, all interconnected to perform the task of a complex circuit. Often these comprise a number of complete conventional circuit stages, such as a multistage amplifier (in one extremely small component). These chips are frequently mounted on a printed circuit board, as shown in figure 2-21, which plugs into an electronic unit.

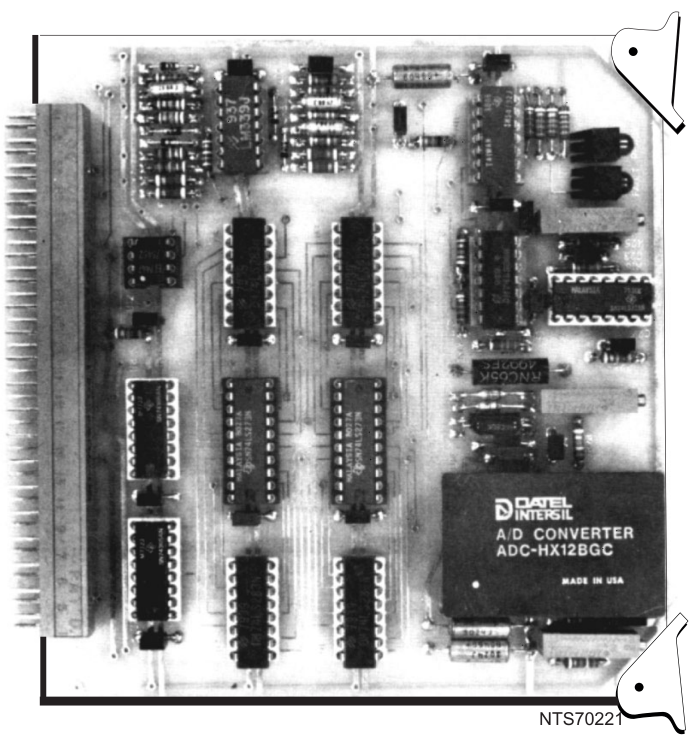


Figure 2-21.—ICs on a printed circuit board.

Integrated circuits have several advantages over conventional wired circuits of discrete components. These advantages include (1) a drastic reduction in size and weight, (2) a large increase in reliability, (3) lower cost, and (4) possible improvement in circuit performance. However, integrated circuits are composed of parts so closely associated with one another that repair becomes almost impossible. In case of trouble, the entire circuit is replaced as a single component.

Basically, there are two general classifications of integrated circuits: HYBRID and MONOLITHIC. In the monolithic integrated circuit, all elements (resistors, transistors, etc.) associated with the circuit are fabricated inseparably within a continuous piece of material (called the SUBSTRATE), usually silicon. The monolithic integrated circuit is made very much like a single transistor. While one part of the crystal is being doped to form a transistor, other parts of the crystal are being acted upon to form the associated resistors and capacitors. Thus, all the elements of the complete circuit are created in the crystal by the same processes and in the same time required to make a single transistor. This produces a considerable cost savings over the same circuit made with discrete components by lowering assembly costs.

Hybrid integrated circuits are constructed somewhat differently from the monolithic devices. The PASSIVE components (resistors, capacitors) are deposited onto a substrate (foundation) made of glass, ceramic, or other insulating material. Then the ACTIVE components (diodes, transistors) are attached to the substrate and connected to the passive circuit components on the substrate using very fine (.001 inch) wire. The term hybrid refers to the fact that different processes are used to form the passive and active components of the device.

Hybrid circuits are of two general types: (1) thin film and (2) thick film. "Thin" and "thick" film refer to the relative thickness of the deposited material used to form the resistors and other passive components. Thick film devices are capable of dissipating more power, but are somewhat more bulky.

Integrated circuits are being used in an ever increasing variety of applications. Small size and weight and high reliability make them ideally suited for use in airborne equipment, missile systems, computers, spacecraft, and portable equipment. They are often easily recognized because of the unusual packages that contain the integrated circuit. A typical packaging sequence is shown in figure 2-22. These tiny packages protect and help dissipate heat generated in the device. One of these packages may contain one or several stages, often having several hundred components. Some of the most common package styles are shown in figure 2-23.

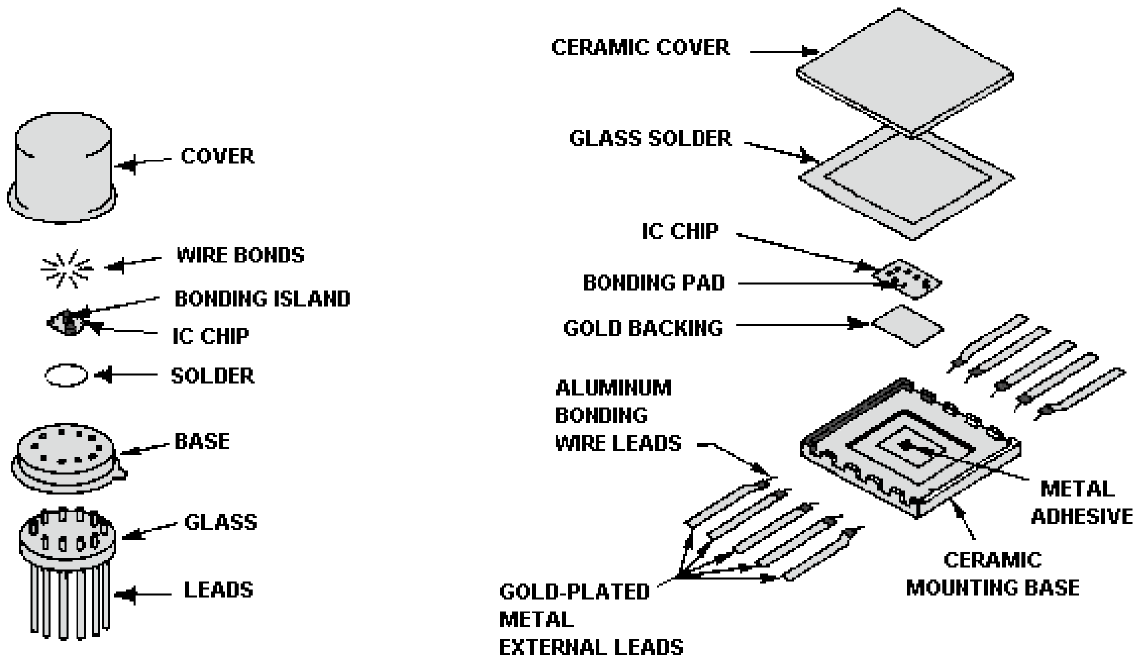


Figure 2-22.—A typical integrated circuit packaging sequence.

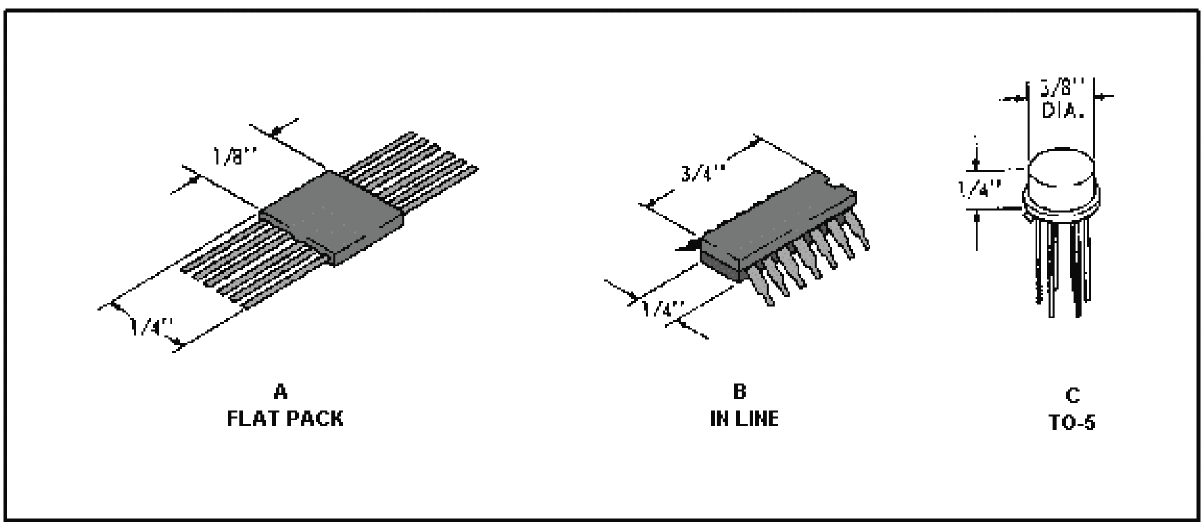


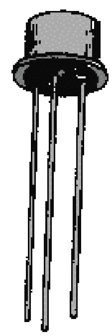
Figure 2-23.—Common IC packaging styles.

The preceding information was presented to give you a brief introduction into integrated circuits. If you wish to pursue this subject further, additional information is available in your ship's or station's library.

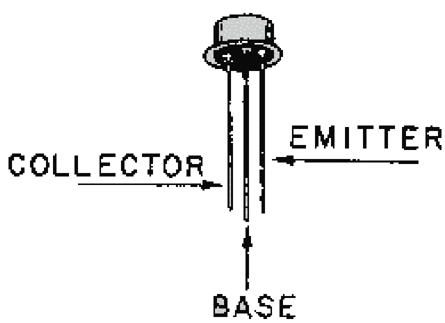
SUMMARY

Now that you have completed this chapter, a short review of the more important points covered in the chapter will follow. This review should refresh your memory of transistors, their theory of operation, and how they are tested with an ohmmeter.

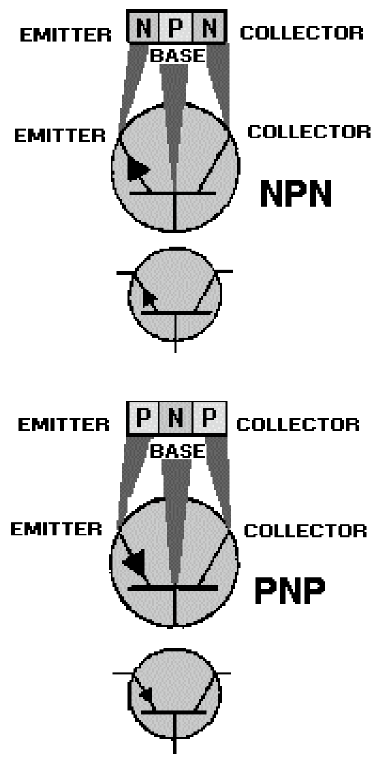
A **TRANSISTOR** is a three or more element solid-state device that amplifies by controlling the flow of current carriers through its semiconductor materials.



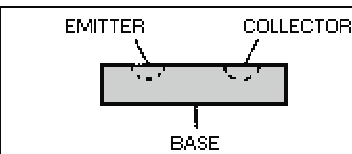
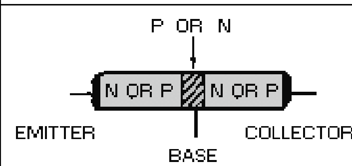
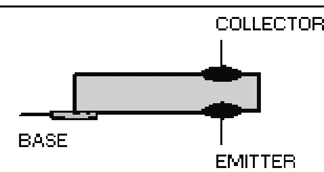
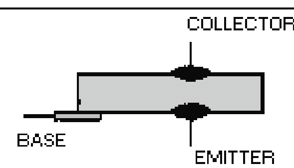
The **THREE ELEMENTS OF A TRANSISTOR** are (1) the EMITTER, which gives off current carriers, (2) the BASE, which controls the carriers, and (3) the COLLECTOR, which collects the carriers.



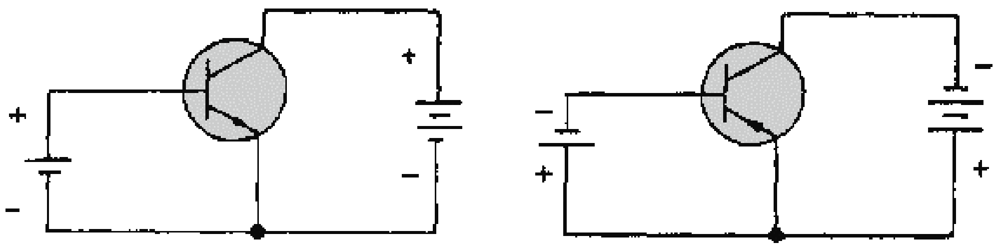
The two **BASIC TYPES** OF **TRANSISTORS** are the NPN and PNP. The only difference in symbology between the two transistors is the direction of the arrow on the emitter. If the arrow points in, it is a PNP transistor and if it points outward, it is an NPN transistor.



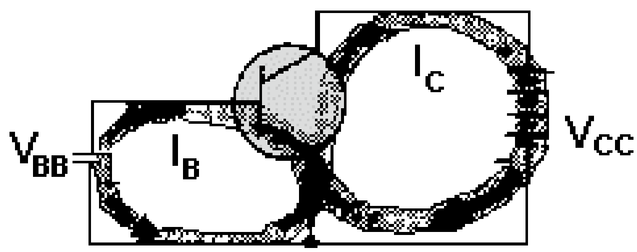
The four **TRANSISTOR MANUFACTURING PROCESSES** are the (1) point contact, (2) grown or rate-grown junction, (3) alloy or fused junction, and (4) diffused junction.

POINT CONTACT (A)	
GROWN JUNCTION OR RATE-GROWN JUNCTION (B)	
ALLOY OR FUSED JUNCTION (C)	
DIFFUSED JUNCTION (D)	

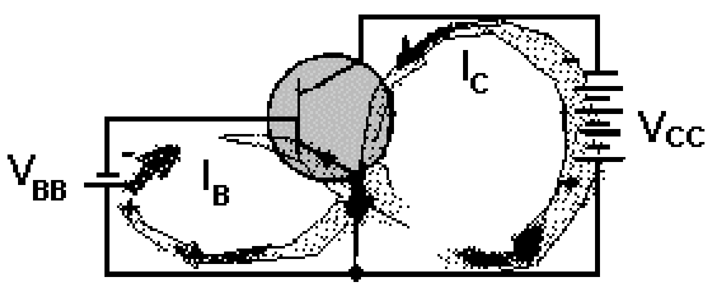
The **PROPER BIASING OF A TRANSISTOR** enables the transistor to be used as an amplifier. To function in this capacity, the emitter-to-base junction of the transistor is forward biased, while the base-to-collector junction is reverse biased.



NPN TRANSISTOR OPERATION is basically the action of a relatively small emitter-base bias voltage controlling a relatively large emitter-to-collector current.



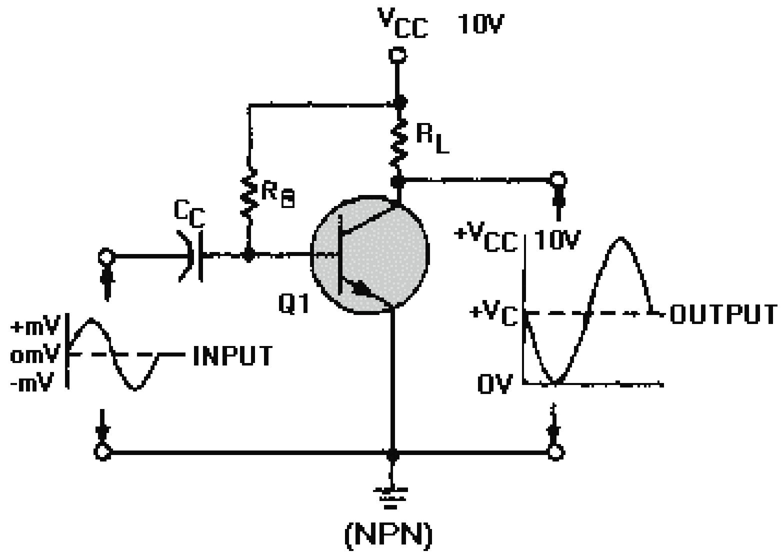
PNP TRANSISTOR OPERATION is essentially the same as the NPN operation except the majority current carriers are holes and the bias batteries are reversed.



AMPLIFICATION is the process of increasing the strength of a signal.

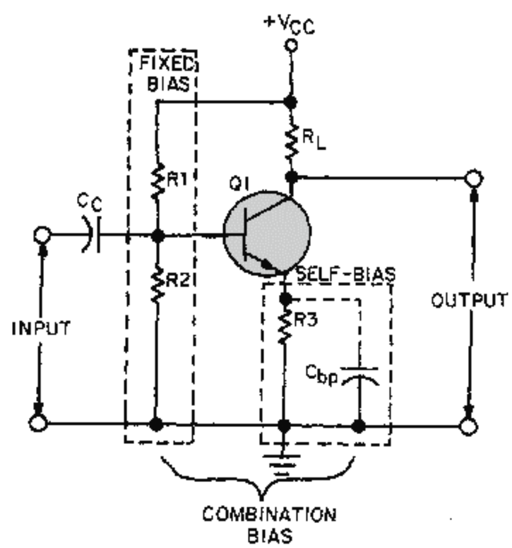
An **AMPLIFIER** is the device that provides amplification without appreciably altering the original signal.

The **BASIC TRANSISTOR AMPLIFIER** amplifies by producing a large change in collector current for a small change in base current. This action results in voltage amplification because the load resistor placed in series with the collector reacts to these large changes in collector current which, in turn, results in large variations in the output voltage.



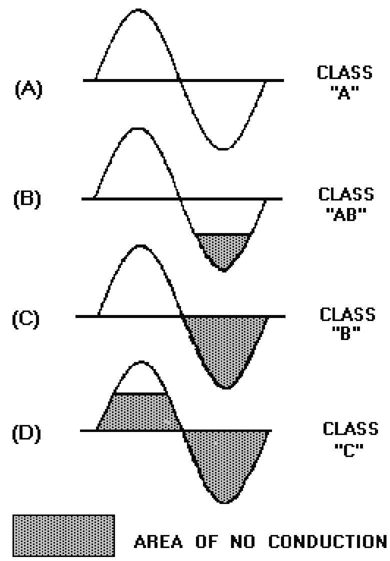
The three types of BIAS used to properly bias a transistor are base-current bias (fixed bias), self-bias, and combination bias.

Combination bias is the one most widely used because it improves circuit stability and at the same time overcomes some of the disadvantages of base-current bias and self-bias.



THE CLASS OF AMPLIFIER OPERATION is determined by the portion of the input signal for which there is an output.

There are four classes of amplifier operations: class A, class AB, class B, and class C.

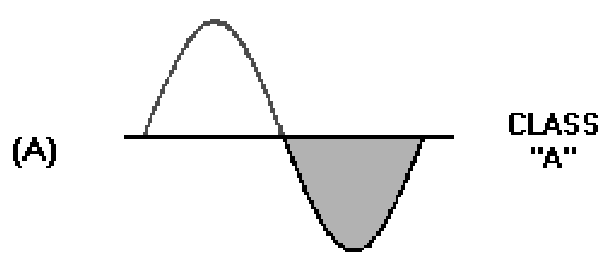


CUTOFF occurs when the base-to-emitter bias prevents current from flowing in the emitter circuit. For example, in the PNP transistor, if the base becomes positive with respect to the emitter, holes are repelled at the emitter-base junction. This prevents current from flowing in the collector circuit.

SATURATION occurs in a PNP transistor when the base becomes so negative, with respect to the emitter, that changes in the signal are not reflected in collector-current flow.

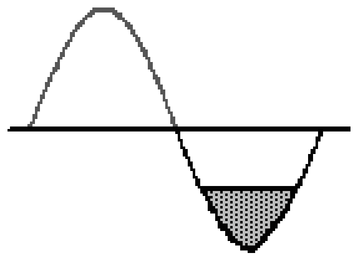
CLASS A AMPLIFIERS are biased so that variations in input signal polarities occur within the limits of cutoff and saturation. Biasing an amplifier in this manner allows collector current to flow during the complete cycle (360 degrees) of the input signal, thus providing an output which is a replica of the input but 180 degrees out of phase.

Class A operated amplifiers are used as audio- and radio-frequency amplifiers in radio, radar, and sound systems.



CLASS AB AMPLIFIERS are biased so that collector current is zero (cutoff) for a portion of one alternation of the input signal. Therefore, collector current will flow for more than 180 degrees but less than 360 degrees of the input signal. The class AB amplifier is commonly used as a push-pull amplifier to overcome a side effect of class B operations.

(B)

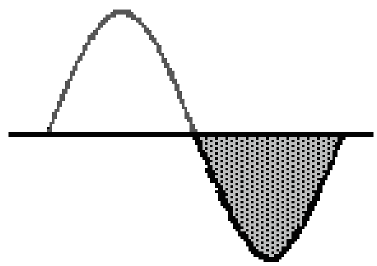


CLASS "AB"

CLASS B AMPLIFIERS are biased so that collector current is cut off during one-half of the input signal. Thus, for a class B operation, collector current will flow for approximately 180 degrees (half) of the input signal.

The class B operated amplifier is used as an audio amplifier and sometimes as the driver- and power-amplifier stage of transmitters.

(C)

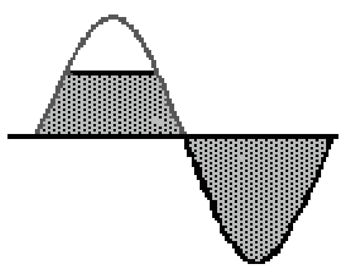


CLASS "B"

CLASS C AMPLIFIERS are biased so that collector current flows for less than one-half cycle of the input signal.

The class C operated amplifier is used as a radio-frequency amplifier in transmitters.

(D)

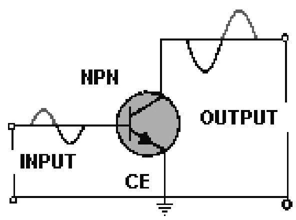


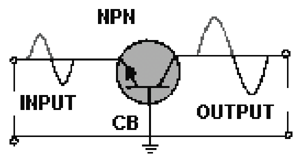
CLASS "C"

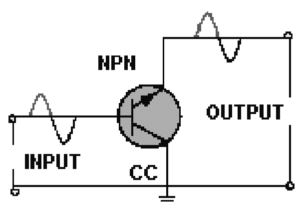
FIDELITY and **EFFICIENCY** are two terms used in conjunction with amplifiers. Fidelity is the faithful reproduction of a signal, while efficiency is the ratio of output signal power compared to the total input power.

The class A amplifier has the highest degree of fidelity, but the class C amplifier has the highest efficiency.

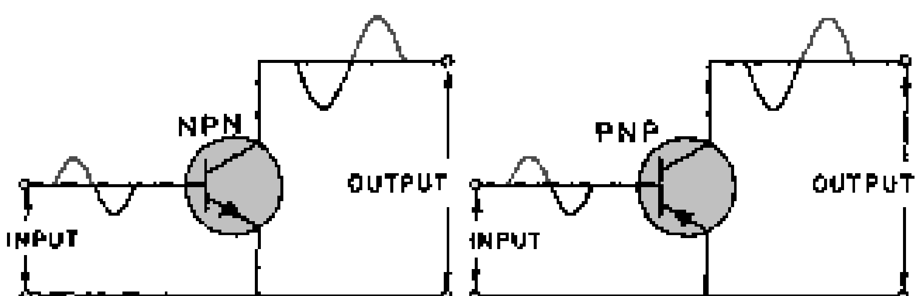
A **TRANSISTOR CONFIGURATION** is the particular way a transistor is connected in a circuit. A transistor may be connected in any one of three different configurations: common emitter (CE), common base (CB), and common collector (CC).







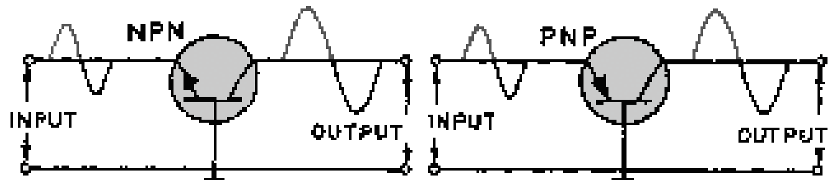
The **COMMON-EMITTER CONFIGURATION (CE)** is the most frequently used configuration in practical amplifier circuits, since it provides good voltage, current, and power gain. The input to the CE is applied to the base-emitter circuit and the output is taken from the collector-emitter circuit, making the emitter the element "common" to both input and output. The CE is set apart from the other configurations, because it is the only configuration that provides a phase reversal between input and output signals.



(A)
COMMON EMITTER

The **COMMON-BASE CONFIGURATION (CB)** is mainly used for impedance matching, since it has a low input resistance and a high output resistance. It also has a current gain of less than 1.

In the CB, the input is applied to the emitter, the output is taken from the collector, and the base is the element common to both input and output.

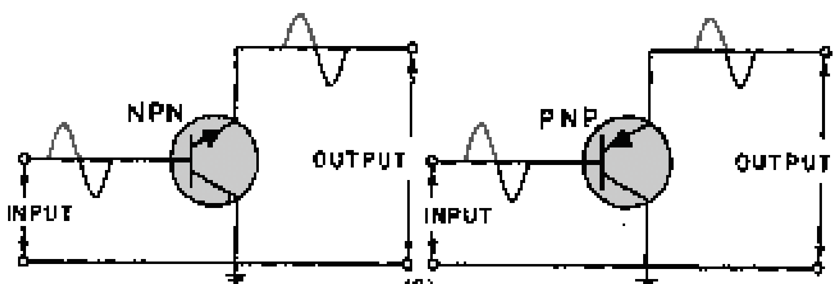


(B)

COMMON BASE

The **COMMON-COLLECTOR CONFIGURATION (CC)** is used as a current driver for impedance matching and is particularly useful in switching circuits. The CC is also referred to as an emitter-follower and is equivalent to the electron-tube cathode follower. Both have high input impedance and low output impedance.

In the CC, the input is applied to the base, the output is taken from the emitter, and the collector is the element common to both input and output.



COMMON (C) COLLECTOR

GAIN is a term used to describe the amplification capabilities of an amplifier. It is basically a ratio of output to input. The current gain for the three transistor configurations (CB, CE, and CC) are ALPHA (α), BETA (β), and GAMMA (γ), respectively.

$$\alpha = \frac{\Delta I_C}{\Delta I_E}$$

$$\beta = \frac{\Delta I_C}{\Delta I_B}$$

$$\gamma = \frac{\Delta I_E}{\Delta I_B}$$

The **TRANSISTOR CONFIGURATION COMPARISON CHART** gives a rundown of the different properties of the three configurations.

AMPLIFIER TYPE	COMMON BASE	COMMON EMITTER	COMMON COLLECTOR
INPUT/OUTPUT PHASE RELATIONSHIP	0º	180º	0º
VOLTAGE GAIN	HIGH	MEDIUM	LOW
CURRENT GAIN	LOW(α)	MEDIUM(β)	HIGH(γ)
POWER GAIN	LOW	HIGH	MEDIUM
INPUT RESISTANCE	LOW	MEDIUM	HIGH
OUTPUT RESISTANCE	HIGH	MEDIUM	LOW

TRANSISTOR CHARACTERISTICS are usually presented on specification sheets. These sheets usually cover the following items:

1. The kind of transistor.
2. The absolute maximum ratings of the transistor.
3. The typical operating values of the transistor.
4. Additional engineering/design information.

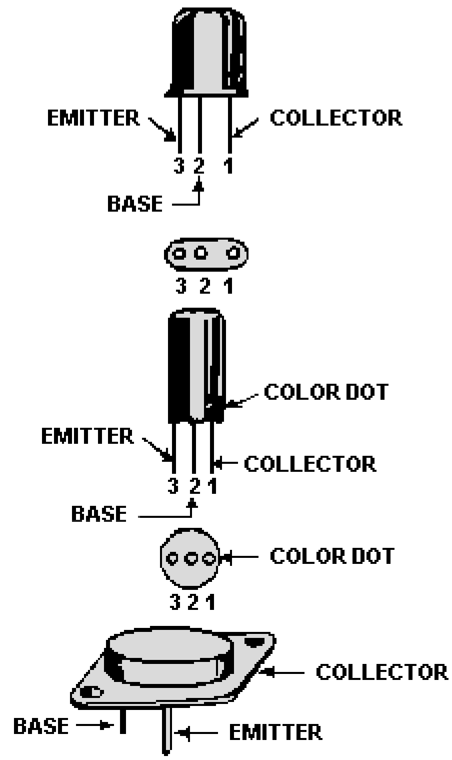
TRANSISTORS ARE IDENTIFIED by a Joint Army-Navy (JAN) designation printed directly on the case of the transistor. If in doubt about a transistor's markings, always replace a transistor with one having identical markings, or consult an equipment or transistor manual to ensure that an identical replacement or substitute is used.

2	N	130	A
NUMBER OF JUNCTIONS (TRANSISTOR)	SEMI-CONDUCTOR	IDENTIFICATION NUMBER	FIRST MODIFICATION

TESTING A TRANSISTOR to determine if it is good or bad can be done with an ohmmeter or transistor tester or by the substitution method.

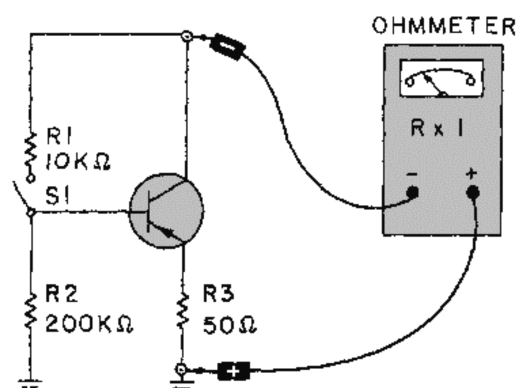
PRECAUTIONS should be taken when working with transistors since they are susceptible to damage by electrical overloads, heat, humidity, and radiation.

TRANSISTOR LEAD IDENTIFICATION plays an important part in transistor maintenance because before a transistor can be tested or replaced, its leads must be identified. Since there is NO standard method of identifying transistor leads, check some typical lead identification schemes or a transistor manual before attempting to replace a transistor.

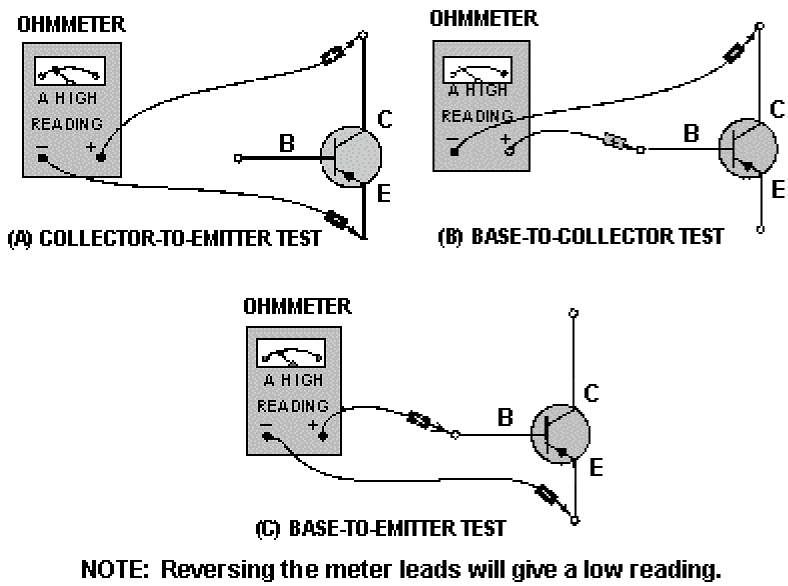


A **TRANSISTOR GAIN TEST** can be made using an ohmmeter and a simple test circuit. The principle behind this test lies in the fact that little or no current will flow in a transistor between emitter and collector until the emitter-base junction is forward biased.

A 10-to-1 resistance ratio in the test between meter readings indicates normal gain.



TRANSISTOR JUNCTION RESISTANCE TEST can also be made using an ohmmeter by measuring the base-emitter, base-collector, and collector-emitter forward and reverse resistances.

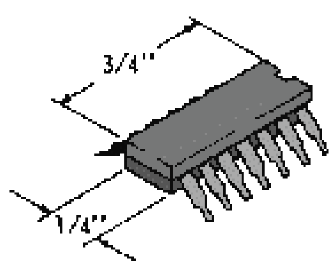


MICROELECTRONICS is a broad term used to describe the use of integrated circuits to miniaturize electronic equipment.

A PRINTED CIRCUIT BOARD (PCB) is a flat, insulating surface upon which printed wiring and miniaturized components are connected in a predetermined design and attached to a common base.

MODULAR CIRCUITRY is an assembly technique in which printed circuit boards are stacked and connected together to form a module. This technique increases the packaging density of circuit components and results in a considerable reduction in the size of electronic equipment.

An **INTEGRATED CIRCUI**T is a device that integrates (combines) both active components (transistors, diodes, etc.) and passive components (resistors, capacitors, etc.) of a complete electronic circuit in a single chip.



The two basic types of ICs are the **HYBRID** and the **MONOLITHIC**.

In the **MONOLITHIC IC**, all elements (resistors, transistors, etc.) associated with the circuit are fabricated inseparably with a continuous piece of material (called the substrate).

In the **HYBRID IC**, the passive components (resistors, capacitors) are deposited onto a substrate (foundation) made of glass, ceramic, or other insulating material. Then the active components (diodes, transistors) are attached to the substrate and connected to the passive components using fine wire.

ANSWERS TO QUESTIONS Q1. THROUGH Q42.

A1. *Transistor*

A2. *Amplification.*

A3. *Outward.*

A4. *Point-contact.*

A5. *Quality control.*

A6. *Positive, more positive.*

A7. *Because the N material on one side of the forward-biased junction is more heavily doped than the P-material.*

A8. *The P or base section.*

A9. *98 percent.*

A10. *Holes.*

A11. *The polarity of voltage applied to the PNP transistor is opposite of that applied to the NPN transistor*

A12. I_B.

A13. *The base current loop and the collector current loop.*

A14. *Amplifier.*

A15. *Compensation for slight variations in transistor characteristics and changes in transistor conduction because of temperature variations.*

A16. *The signals are opposite in polarity or 180 degrees out of phase with each other.*

A17. *The polarity of the source voltage.*

A18. *Base current bias or fixed bias.*

A19. *Self-bias.*

A20. *When it is necessary to prevent amplitude distortion.*

A21. *The voltage-divider type.*

A22. *Class A.*

A23. *Cutoff.*

A24. The amount of bias and the amplitude of the input signal.

A25. Class A.

A26. Common emitter (CE), common base (CB), and common collector (CC).

A27. Common emitter.

A28. Base current (I_B).

A29. Alpha (α).

A30. Common base.

A31. I_E.

A32. Common collector.

A33.

$$\gamma = \frac{I_E}{I_B}$$

A34. The kind of transistor, the transistor's common applications, and mechanical data.

A35. The number of junctions in the device, which in this case indicates a transistor.

A36. Heat.

A37. The substitution method.

A38. The power must be removed from the circuit.

A39. By the wide space between the collector lead and the other two leads (emitter and base).

A40. Gain and junction resistance.

A41. Normal gain.

A42. A leaking transistor

CHAPTER 3

SPECIAL DEVICES

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to:

1. Explain the basic operation and the major applications of the Zener diode.
2. Describe the basic operation of the tunnel diode and the varactor.
3. Explain the basic operation of the silicon controlled rectifier and the TRIAC, and compare the advantages and disadvantages of each.
4. List the five most commonly used optoelectronic devices and explain the uses of each.
5. Describe the basic operation, applications, and major advantages of the unijunction transistor.
6. Describe the basic operation, applications, and major advantages of the field effect transistor and the metal oxide semiconductor field effect transistor.
7. Explain the basic operation and the major applications of the Zener diode.
8. Describe. the basic operation of the tunnel diode and the varactor.
9. Explain the basic operation of the silicon controlled rectifier and the TRIAC, and compare the advantages and disadvantages of each.
10. List the five most commonly used optoelectronic devices and explain the uses of each.
11. Describe the basic operation, applications, and major advantages of the unijunction transistor.
12. Describe the basic operation, applications, and major advantages of the field-effect transistor and the metal-oxide semiconductor field-effect transistor.

INTRODUCTION TO SPECIAL DEVICES

If you consider the sensitive nature and the various interacting properties of semiconductors, it should not be surprising to you that solid state devices can be designed for many different purposes. In fact, devices with special features are so numerous and new designs are so frequently introduced that it would be beyond the scope of this chapter to describe all of the devices in use today. Therefore, this chapter will include a variety of representative devices that are used extensively in Navy equipment to give you an idea of the diversity and versatility that have been made possible. These devices have been grouped into three categories: diodes, optoelectronic devices, and transistors. In this chapter each device will be described and the basic operation of each one will be discussed.

DIODES

Diodes are two terminal semiconductors of various types that are used in seemingly endless applications. The operation of normal PN-junction diodes has already been discussed, but there are a number of diodes with special properties with which you should be familiar. A discussion of all of the developments in the diode field would be impossible so some of the more commonly used special diodes have been selected for explanation. These include Zener diodes, tunnel diodes, varactors, silicon controlled rectifiers (SCR), and TRIACs.

Zener Diodes

When a PN-junction diode is reverse biased, the majority carriers (holes in the P-material and electrons in the N-material) move away from the junction. The barrier or depletion region becomes wider, as illustrated in figure 3-1, (view A, view B, view C) and majority carrier current flow becomes very difficult across the high resistance of the wide depletion region. The presence of minority carriers causes a small leakage current that remains nearly constant for all reverse voltages up to a certain value. Once this value has been exceeded, there is a sudden increase in the reverse current. The voltage at which the sudden increase in current occurs is called the BREAKDOWN VOLTAGE. At breakdown, the reverse current increases very rapidly with a slight increase in the reverse voltage. Any diode can be reverse biased to the point of breakdown, but not every diode can safely dissipate the power associated with breakdown. A Zener diode is a PN junction designed to operate in the reverse-bias breakdown region.

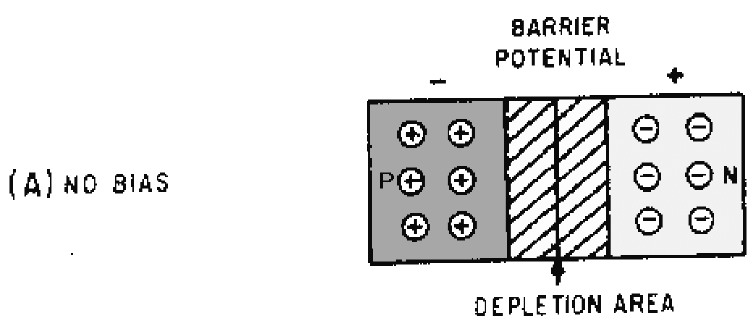


Figure 3-1A.—Effects of bias on the depletion region of a PN junction.

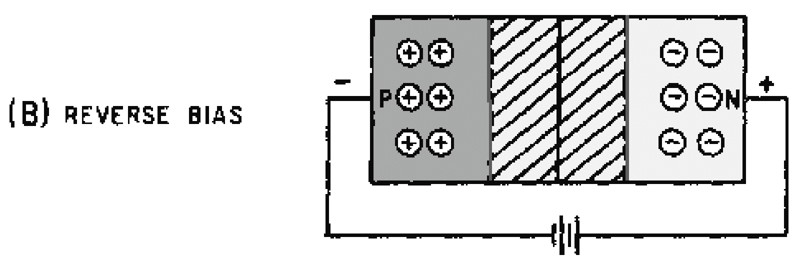


Figure 3-1B.—Effects of bias on the depletion region of a PN junction.

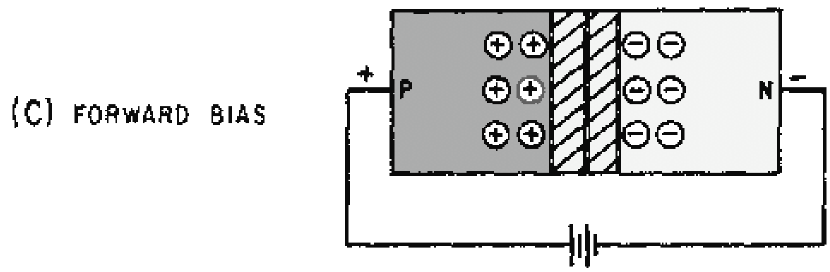


Figure 3-1C.—Effects of bias on the depletion region of a PN junction.

There are two distinct theories used to explain the behavior of PN junctions during breakdown: one is the ZENER EFFECT and the other is the AVALANCHE EFFECT.

The ZENER EFFECT was first proposed by Dr. Carl Zener in 1934. According to Dr. Zener's theory, electrical breakdown in solid dielectrics occurs by a process called QUANTUM-MECHANICAL TUNNELING. The Zener effect accounts for the breakdown below 5 volts; whereas, above 5 volts the breakdown is caused by the avalanche effect. Although the avalanche effect is now accepted as an explanation of diode breakdown, the term *Zener diode* is used to cover both types.

The true Zener effect in semiconductors can be described in terms of energy bands; however, only the two upper energy bands are of interest. The two upper bands, illustrated in figure 3-2, view A, are called the conduction band and the valence band.

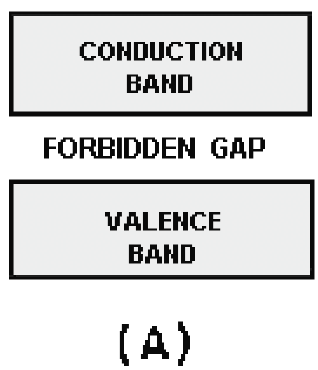


Figure 3-2A.-Energy diagram for Zener diode.

The CONDUCTION BAND is a band in which the energy level of the electrons is high enough that the electrons will move easily under the influence of an external field. Since current flow is the movement of electrons, the readily mobile electrons in the conduction band are capable of maintaining a current flow when an external field in the form of a voltage is applied. Therefore, solid materials that have many electrons in the conduction band are called conductors.

The VALENCE BAND is a band in which the energy level is the same as the valence electrons of the atoms. Since the electrons in these levels are attached to the atoms, the electrons are not free to move around as are the conduction band electrons. With the proper amount of energy added, however, the electrons in the valence band may be elevated to the conduction band energy level. To do this, the electrons must cross a gap that exists between the valence band energy level and the conduction band energy level. This gap is known as the FORBIDDEN ENERGY BAND or FORBIDDEN GAP. The energy difference across this gap determines whether a solid material will act as a conductor, a semiconductor, or an insulator.

A conductor is a material in which the forbidden gap is so narrow that it can be considered nonexistent. A semiconductor is a solid that contains a forbidden gap, as shown in figure 3-2, view A. Normally, a semiconductor has no electrons at the conduction band energy level. The energy provided by room temperature heat, however, is enough energy to overcome the binding force of a few valence electrons and to elevate them to the conduction band energy level. The addition of impurities to the semiconductor material increases both the number of free electrons in the conduction band and the number of electrons in the valence band that can be elevated to the conduction band. Insulators are materials in which the forbidden gap is so large that practically no electrons can be given enough energy to cross the gap. Therefore, unless extremely large amounts of heat energy are available, these materials will not conduct electricity.

View B of figure 3-2 is an energy diagram of a reverse-biased Zener diode. The energy bands of the P and N materials are naturally at different levels, but reverse bias causes the valence band of the P material to overlap the energy level of the conduction band in the N material. Under this condition, the valence electrons of the P material can cross the extremely thin junction region at the overlap point without acquiring any additional energy. This action is called tunneling. When the breakdown point of the PN junction is reached, large numbers of minority carriers "tunnel" across the junction to form the current that occurs at breakdown. The tunneling phenomenon only takes place in heavily doped diodes such as Zener diodes.

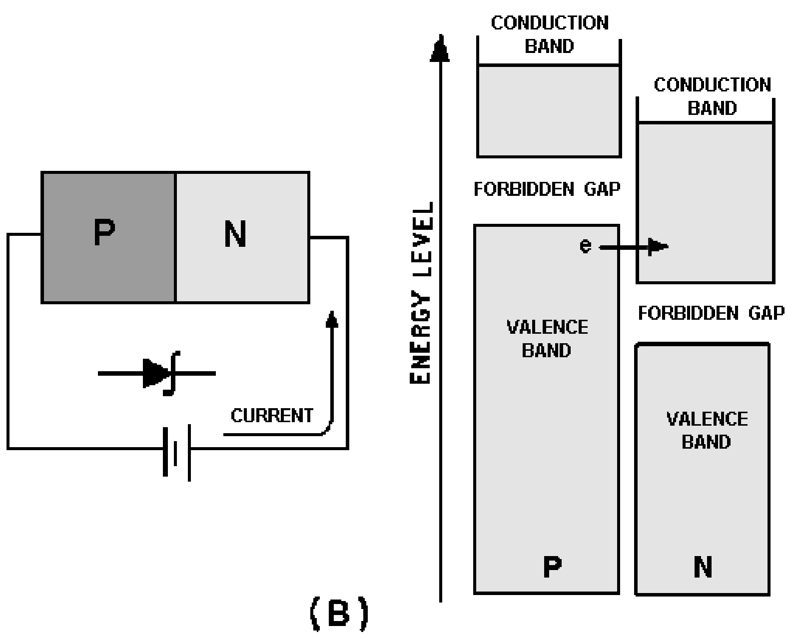


Figure 3-2B.-Energy diagram for Zener diode.

The second theory of reverse breakdown effect in diodes is known as AVALANCHE breakdown and occurs at reverse voltages beyond 5 volts. This type of breakdown diode has a depletion region that is deliberately made narrower than the depletion region in the normal PN-junction diode, but thicker than that in the Zener-effect diode. The thicker depletion region is achieved by decreasing the doping level from the level used in Zener-effect diodes. The breakdown is at a higher voltage because of the higher resistivity of the material. Controlling the doping level of the material during the manufacturing process can produce breakdown voltages ranging between about 2 and 200 volts.

The mechanism of avalanche breakdown is different from that of the Zener effect. In the depletion region of a PN junction, thermal energy is responsible for the formation of electron-hole pairs. The leakage current is caused by the movement of minority electrons, which is accelerated in the electric field across the barrier region. As the reverse voltage across the depletion region is increased, the reverse voltage eventually reaches a critical value. Once the critical or breakdown voltage has been reached, sufficient energy is gained by the thermally released minority electrons to enable the electrons to rupture covalent bonds as they collide with lattice atoms. The released electrons are also accelerated by the electric field, resulting in the release of further electrons, and so on, in a chain or avalanche effect. This process is illustrated in figure 3-3.

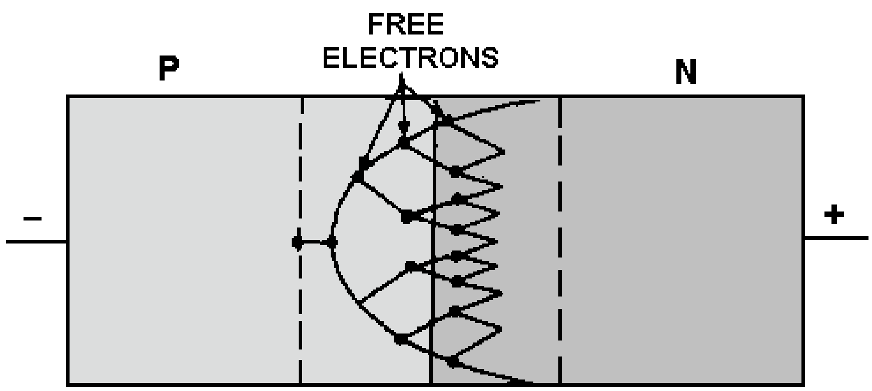


Figure 3-3.—Avalanche multiplication.

For reverse voltage slightly higher than breakdown, the avalanche effect releases an almost unlimited number of carriers so that the diode essentially becomes a short circuit. The current flow in this region is limited only by an external series current-limiting resistor. Operating a diode in the breakdown region does not damage it, as long as the maximum power dissipation rating of the diode is not exceeded. Removing the reverse voltage permits all carriers to return to their normal energy values and velocities.

Some of the symbols used to represent Zener diodes are illustrated in figure 3-4 (view A, view B, view C, view D, and view E). Note that the polarity markings indicate electron flow is with the arrow symbol instead of against it as in a normal PN-junction diode. This is because breakdown diodes are operated in the reverse-bias mode, which means the current flow is by minority current carriers.

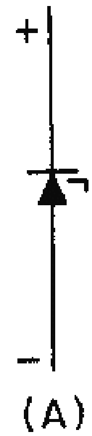


Figure 3-4A.—Schematic symbols for Zener diodes.

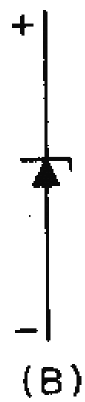


Figure 3-4B.—Schematic symbols for Zener diodes.

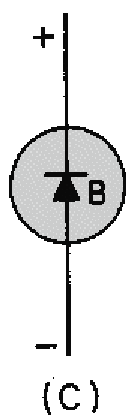


Figure 3-4C.—Schematic symbols for Zener diodes.

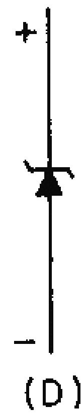


Figure 3-4D.—Schematic symbols for Zener diodes.

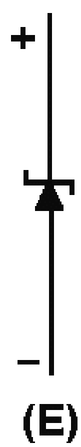


Figure 3-4E.—Schematic symbols for Zener diodes.

Zener diodes of various sorts are used for many purposes, but their most widespread use is as voltage regulators. Once the breakdown voltage of a Zener diode is reached, the voltage across the diode remains almost constant regardless of the supply voltage. Therefore they hold the voltage across the load at a constant level. This characteristic makes Zener diodes ideal voltage regulators, and they are found in almost all solid-state circuits in this capacity.

Q1. In a reverse biased PN-junction, which current carriers cause leakage current?

Q2. The action of a PN-junction during breakdown can be explained by what two theories?

Q3. Which breakdown theory explains the action that takes place in a heavily doped PN-junction with a reverse bias of less than 5 volts?

Q4. What is the doping level of an avalanche effect diode when compared to the doping level of a Zener-effect diode?

Q5. During avalanche effect breakdown, what limits current flow through the diode?

Q6. Why is electron flow with the arrow in the symbol of a Zener diode instead of against the arrow as it is in a normal diode?

The Tunnel Diode

In 1958, Leo Esaki, a Japanese scientist, discovered that if a semiconductor junction diode is heavily doped with impurities, it will have a region of negative resistance. The normal junction diode uses semiconductor materials that are lightly doped with one impurity atom for ten-million semiconductor atoms. This low doping level results in a relatively wide depletion region. Conduction occurs in the normal junction diode only if the voltage applied to it is large enough to overcome the potential barrier of the junction.

In the TUNNEL DIODE, the semiconductor materials used in forming a junction are doped to the extent of one-thousand impurity atoms for ten-million semiconductor atoms. This heavy doping produces an extremely narrow depletion zone similar to that in the Zener diode. Also because of the heavy doping, a tunnel diode exhibits an unusual current-voltage characteristic curve as compared with that of an ordinary junction diode. The characteristic curve for a tunnel diode is illustrated in figure 3-5.

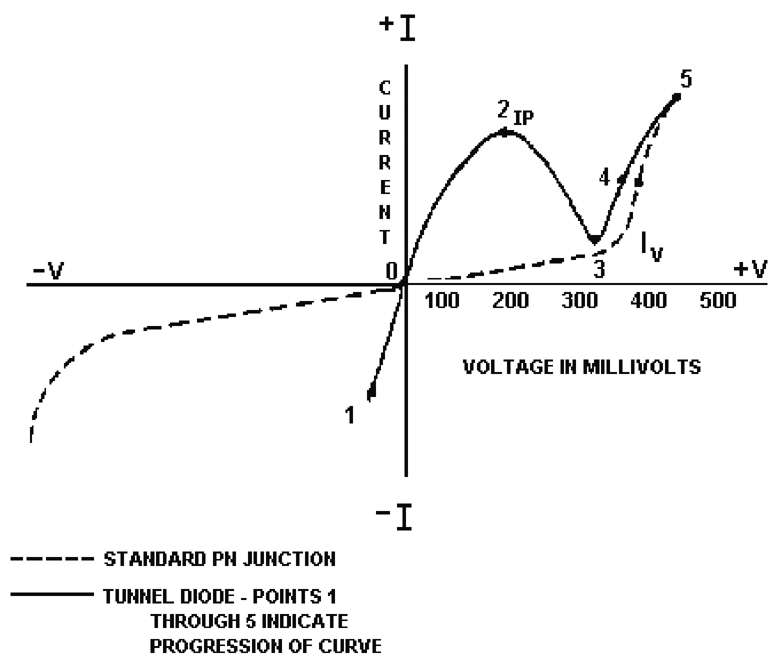


Figure 3-5.—Characteristic curve of a tunnel diode compared to that of a standard PN junction.

The three most important aspects of this characteristic curve are (1) the forward current increase to a peak (I_P) with a small applied forward bias, (2) the decreasing forward current with an increasing forward bias to a minimum valley current (I_V), and (3) the normal increasing forward current with further increases in the bias voltage. The portion of the characteristic curve between I_P and I_V is the region of negative resistance. An explanation of why a tunnel diode has a region of negative resistance is best understood by using energy levels as in the previous explanation of the Zener effect.

Simply stated the theory known as quantum-mechanical tunneling is an electron crossing a PN-junction without having sufficient energy to do so otherwise. Because of the heavy doping the width of the depletion region is only one-millionth of an inch. You might think of the process simply as an arc-over between the N- and the P-side across the depletion region.

Figure 3-6 shows the equilibrium energy level diagram of a tunnel diode with no bias applied. Note in view A that the valence band of the P-material overlaps the conduction band of the N-material. The majority electrons and holes are at the same energy level in the equilibrium state. If there is any movement of current carriers across the depletion region due to thermal energy, the net current flow will be zero because equal numbers of current carriers flow in opposite directions. The zero net current flow is marked by a "0" on the current-voltage curve illustrated in view B.

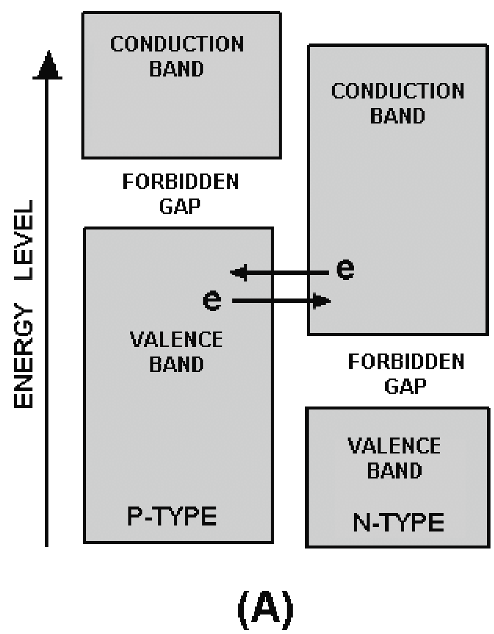


Figure 3-6A.—Tunnel diode energy diagram with no bias.

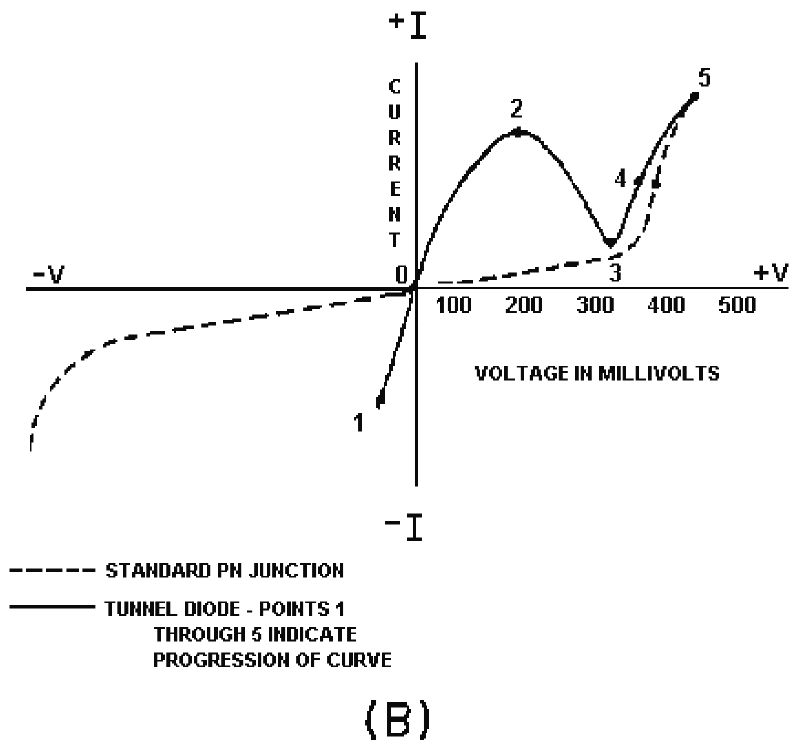


Figure 3-6B.—Tunnel diode energy diagram with no bias.

Figure 3-7, view A, shows the energy diagram of a tunnel diode with a small forward bias (50 millivolts) applied. The bias causes unequal energy levels between some of the majority carriers at the energy band overlap point, but not enough of a potential difference to cause the carriers to cross the forbidden gap in the normal manner. Since the valence band of the P-material and the conduction band of the N-material still overlap, current carriers tunnel across at the overlap and cause a substantial current flow. The amount of current flow is marked by point 2 on the curve in view B. Note in view A that the amount of overlap between the valence band and the conduction band decreased when forward bias was applied.

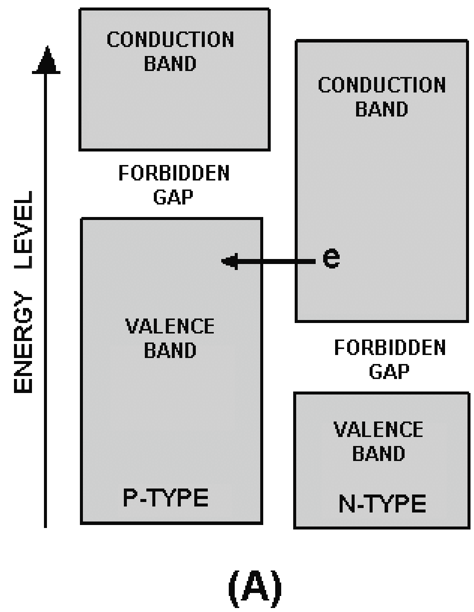


Figure 3-7A.—Tunnel diode energy diagram with 50 millivolts bias.

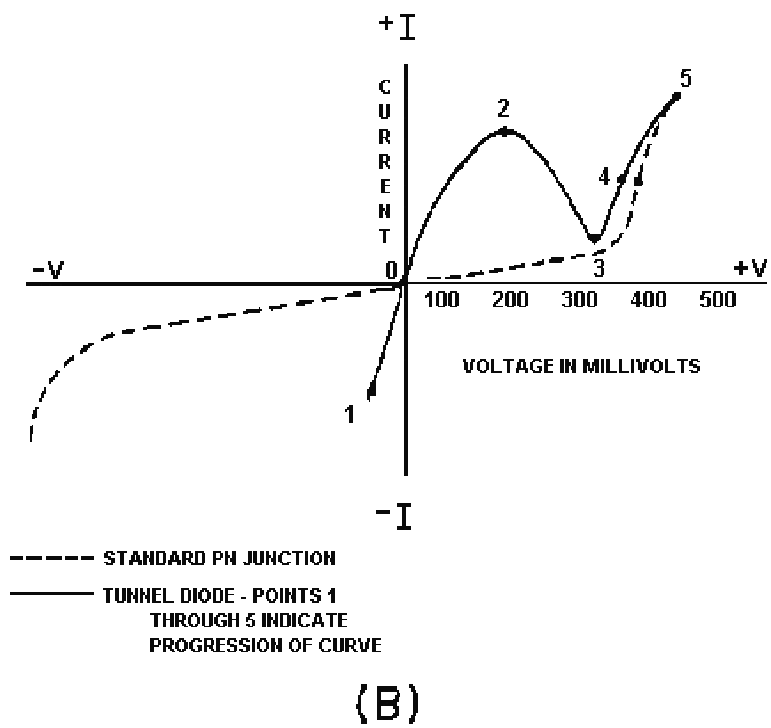


Figure 3-7B.—Tunnel diode energy diagram with 50 millivolts bias.

Figure 3-8, view A, is the energy diagram of a tunnel diode in which the forward bias has been increased to 450 millivolts. As you can see, the valence band and the conduction band no longer overlap at this point, and tunneling can no longer occur. The portion of the curve in view B from point 2 to point 3 shows the decreasing current that occurs as the bias is increased, and the area of overlap becomes smaller. As the overlap between the two energy bands becomes smaller, fewer and fewer electrons can tunnel across the junction. The portion of the curve between point 2 and point 3 in which current decreases as the voltage increases is the negative resistance region of the tunnel diode.

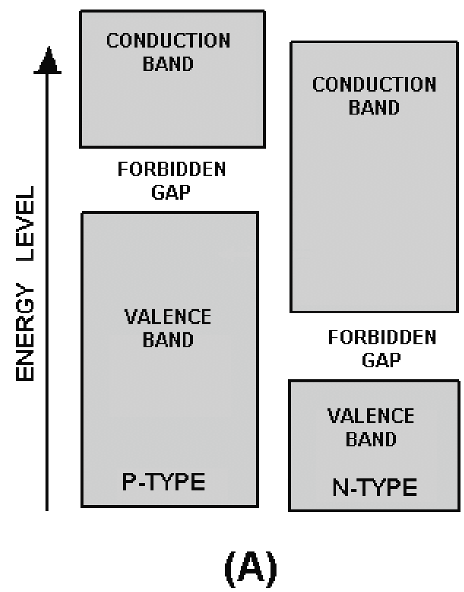


Figure 3-8A.—Tunnel diode energy diagram with 450 millivolts bias.

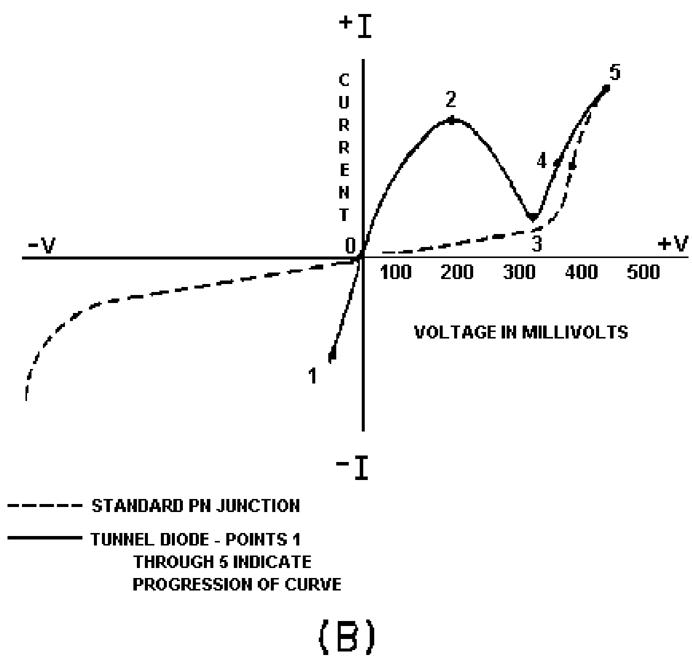


Figure 3-8B.—Tunnel diode energy diagram with 450 millivolts bias.

Figure 3-9, view A, is the energy diagram of a tunnel diode in which the forward bias has been increased even further. The energy bands no longer overlap and the diode operates in the same manner as a normal PN junction, as shown by the portion of the curve in view (B) from point 3 to point 4.

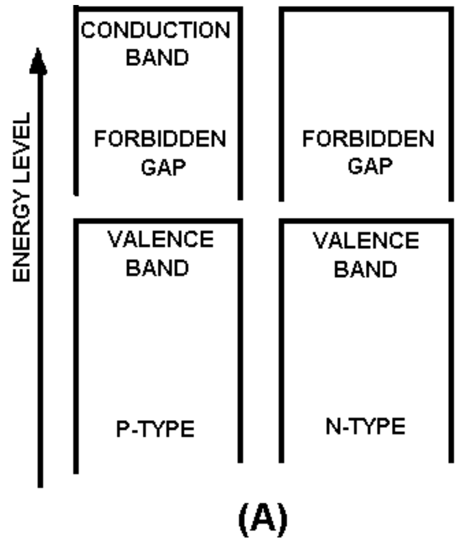


Figure 3-9A.—Tunnel diode energy diagram with 600 millivolts bias.

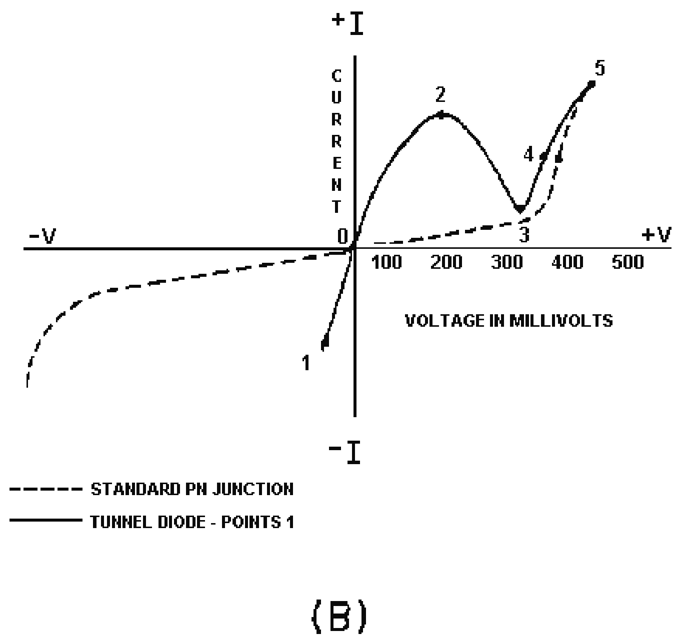


Figure 3-9B.—Tunnel diode energy diagram with 600 millivolts bias.

The negative resistance region is the most important and most widely used characteristic of the tunnel diode. A tunnel diode biased to operate in the negative resistance region can be used as either an oscillator or an amplifier in a wide range of frequencies and applications. Very high frequency applications using the tunnel diode are possible because the tunneling action occurs so rapidly that there is no transit time effect and therefore no signal distortion. Tunnel diodes are also used extensively in high-speed switching circuits because of the speed of the tunneling action.

Several schematic symbols are used to indicate a tunnel diode. These symbols are illustrated in figure 3-10 (view A, view B, view C, and view D).

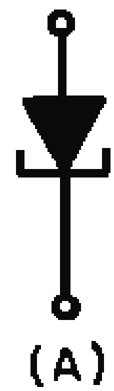


Figure 3-10A.—Tunnel diode schematic symbols.

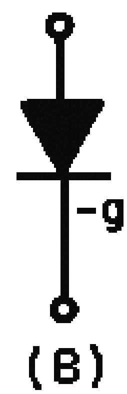


Figure 3-10B.—Tunnel diode schematic symbols.

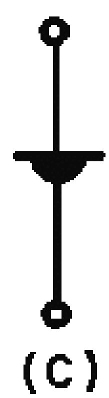


Figure 3-10C.—Tunnel diode schematic symbols.

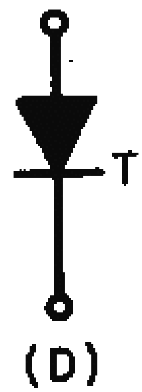


Figure 3-10D.—Tunnel diode schematic symbols.

Varactor

The VARACTOR, or varicap, as the schematic drawing in figure 3-11 suggests, is a diode that behaves like a variable capacitor, with the PN junction functioning like the dielectric and plates of a common capacitor. Understanding how the varactor operates is an important prerequisite to understanding field-effect transistors, which will be covered later in this topic.

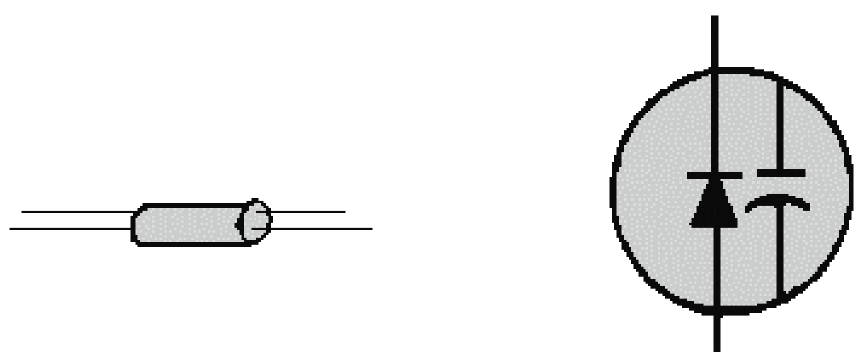


Figure 3-11.—Varactor diode.

Figure 3-12 shows a PN junction. Surrounding the junction of the P and N materials is a narrow region void of both positively and negatively charged current carriers. This area is called the depletion region.

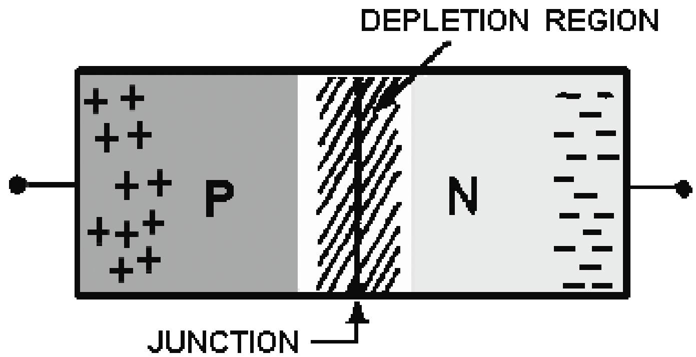


Figure 3-12.—PN junction.

The size of the depletion region in a varactor diode is directly related to the bias. Forward biasing makes the region smaller by repelling the current carriers toward the PN junction. If the applied voltage is large enough (about .5 volt for silicon material), the negative particles will cross the junction and join with the positive particles, as shown in figure 3-13. This forward biasing causes the depletion region to decrease, producing a low resistance at the PN junction and a large current flow across it. This is the condition for a forward-biased diode. On the other hand, if reverse-bias voltage is applied to the PN junction, the size of its depletion region increases as the charged particles on both sides move away from the junction. This condition, shown in figure 3-14, produces a high resistance between the terminals and allows little current flow (only in the microampere range). This is the operating condition for the varactor diode, which is nothing more than a special PN junction.

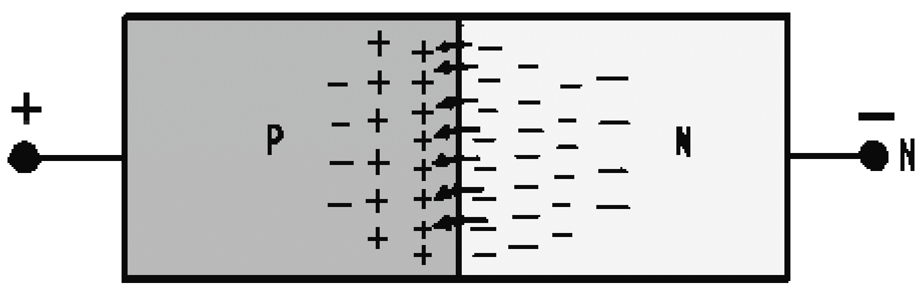


Figure 3-13.—Forward-biased PN junction.

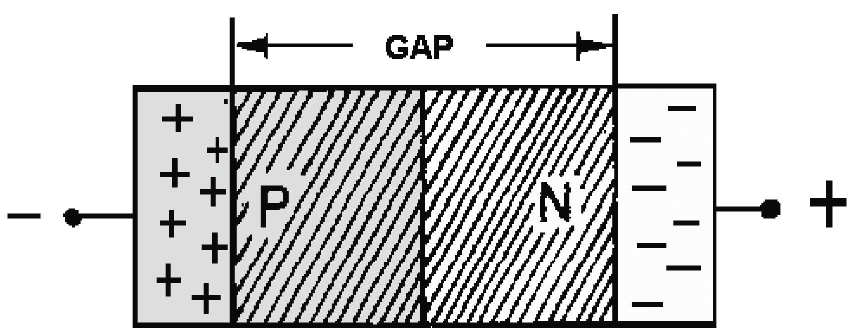


Figure 3-14.—Reverse-biased PN junction.

As the figure shows, the insulation gap formed by reverse biasing of the varactor is comparable to the layer of dielectric material between the plates of a common capacitor. Furthermore, the formula used to calculate capacitance

$$C = \frac{AK}{d}$$

Where

A = plate area

K = a constant value

d = distance between plates

can be applied to both the varactor and the capacitor. In this case, the size of the insulation gap of the varactor, or depletion region, is substituted for the distance between the plates of the capacitor. By varying the reverse-bias voltage applied to the varactor, the width of the "gap" may be varied. An increase in reverse bias increases the width of the gap (d) which reduces the capacitance (C) of the PN junction. Therefore, the capacitance of the varactor is inversely proportional to the applied reverse bias.

The ratio of varactor capacitance to reverse-bias voltage change may be as high as 10 to 1. Figure 3-15 shows one example of the voltage-to-capacitance ratio. View A shows that a reverse bias of 3 volts produces a capacitance of 20 picofarads in the varactor. If the reverse bias is increased to 6 volts, as shown in view B, the depletion region widens and capacitance drops to 5 picofarads. Each 1-volt increase in bias voltage causes a 5-picofarad decrease in the capacitance of the varactor; the ratio of change is therefore 5 to 1. Of course any decrease in applied bias voltage would cause a proportionate increase in capacitance, as the depletion region narrows. Notice that the value of the capacitance is small in the picofarad range.

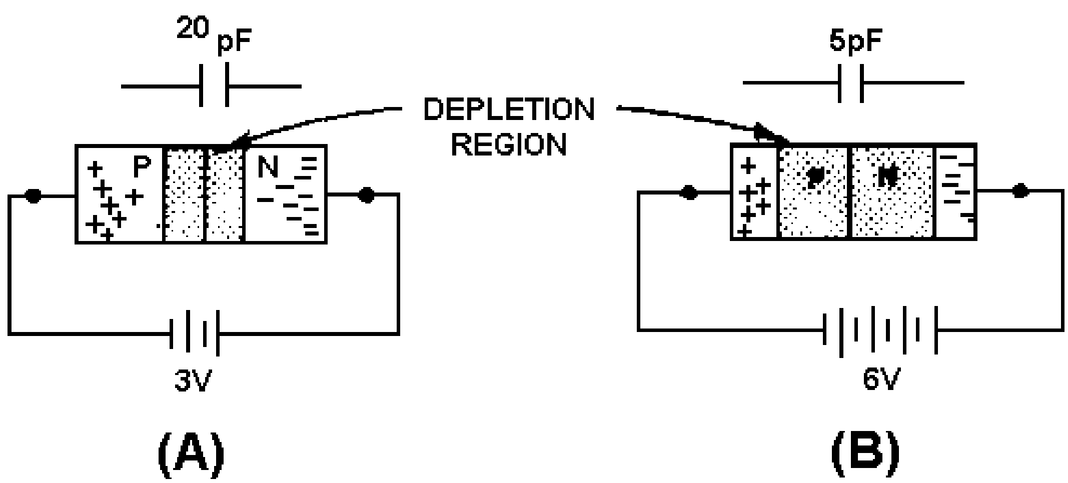


Figure 3-15.—Varactor capacitance versus bias voltage.

In general, varactors are used to replace the old style variable capacitor tuning. They are used in tuning circuits of more sophisticated communication equipment and in other circuits where variable capacitance is required. One advantage of the varactor is that it allows a dc voltage to be used to tune a circuit for simple remote control or automatic tuning functions. One such application of the varactor is as a variable tuning capacitor in a receiver or transmitter tank circuit like that shown in figure 3-16.

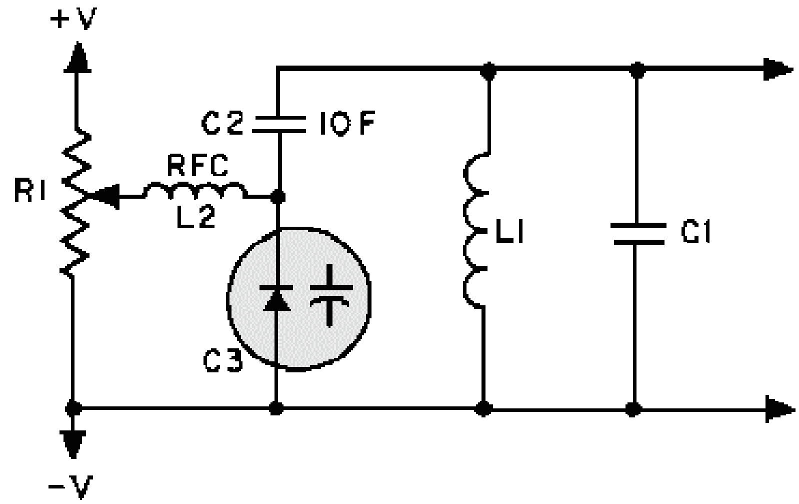


Figure 3-16.—Varactor tuned resonant circuit.

Figure 3-16 shows a dc voltage felt at the wiper of potentiometer R1 which can be adjusted between +V and –V. The dc voltage, passed through the low resistance of radio frequency choke L2, acts to reverse bias varactor diode C3. The capacitance of C3 is in series with C2, and the equivalent capacitance of C2 and C3 is in parallel with tank circuit L1-C1. Therefore, any variation in the dc voltage at R1 will vary both the capacitance of C3 and the resonant frequency of the tank circuit. The radio-frequency choke provides high inductive reactance at the tank frequency to prevent tank loading by R1. C2 acts to block dc from the tank as well as to fix the tuning range of C3.

An ohmmeter can be used to check a varactor diode in a circuit. A high reverse-bias resistance and a low forward-bias resistance with a 10 to 1 ratio in reverse-bias to forward-bias resistance is considered normal.

Q7. What is the main difference in construction between normal PN junction diodes and tunnel diodes?

Q8. What resistance property is found in tunnel diodes but not in normal diodes?

Q9. When compared to the ordinary diode, the tunnel diode has what type of depletion region?

Q10. In the tunnel diode, the tunneling current is at what level when the forbidden gap of the N-type material is at the same energy level as the empty states of the P-type material?

Q11. The varactor displays what useful electrical property?

Q12. When a PN junction is forward biased, what happens to the depletion region?

Q13. When the reverse bias on a varactor is increased, what happens to the effective capacitance?

Silicon Controlled Rectifier (SCR)

The SILICON CONTROLLED RECTIFIER, usually referred to as an SCR, is one of the family of semiconductors that includes transistors and diodes. A drawing of an SCR and its schematic representation is shown in views A and B of figure 3-17. Not all SCRs use the casing shown, but this is typical of most of the high-power units.

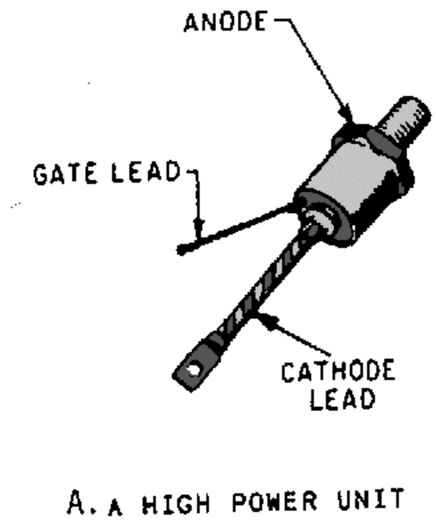


Figure 3-17A.—Silicon controlled rectifier.

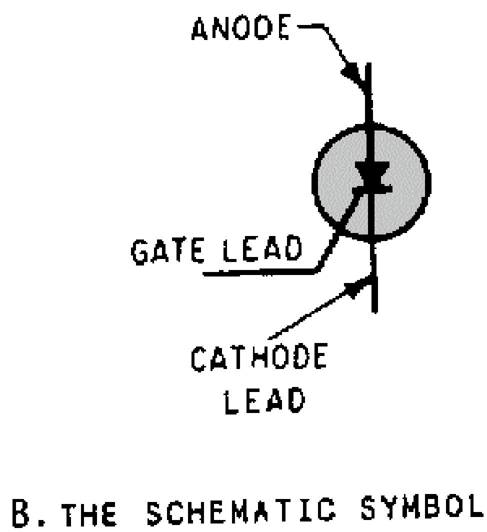


Figure 3-17B.—Silicon controlled rectifier.

Although it is not the same as either a diode or a transistor, the SCR combines features of both. Circuits using transistors or rectifier diodes may be greatly improved in some instances through the use of SCRs.

The basic purpose of the SCR is to function as a switch that can turn on or off small or large amounts of power. It performs this function with no moving parts that wear out and no points that require replacing. There can be a tremendous power gain in the SCR; in some units a very small triggering current is able to switch several hundred amperes without exceeding its rated abilities. The SCR can often replace much slower and larger mechanical switches. It even has many advantages over its more complex and larger electron tube equivalent, the thyratron.

The SCR is an extremely fast switch. It is difficult to cycle a mechanical switch several hundred times a minute; yet, some SCRs can be switched 25,000 times a second. It takes just microseconds (millionths of a second) to turn on or off these units. Varying the time that a switch is on as compared to the time that it is off regulates the amount of power flowing through the switch. Since most devices can operate on pulses of power (alternating current is a special form of alternating positive and negative pulse), the SCR can be used readily in control applications. Motor-speed controllers, inverters, remote switching units, controlled rectifiers, circuit overload protectors, latching relays, and computer logic circuits all use the SCR.

The SCR is made up of four layers of semiconductor material arranged PNPN. The construction is shown in view A of figure 3-18. In function, the SCR has much in common with a diode, but the theory of operation of the SCR is best explained in terms of transistors.

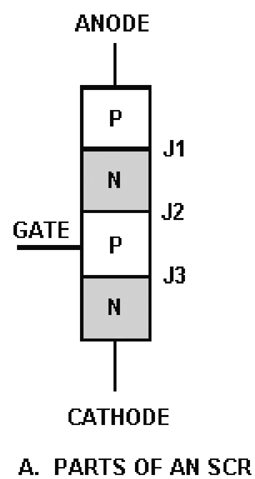


Figure 3-18A.—SCR structure.

Consider the SCR as a transistor pair, one PNP and the other NPN, connected as shown in views B and C. The anode is attached to the upper P-layer; the cathode, C, is part of the lower N-layer; and the gate terminal, G, goes to the P-layer of the NPN triode.

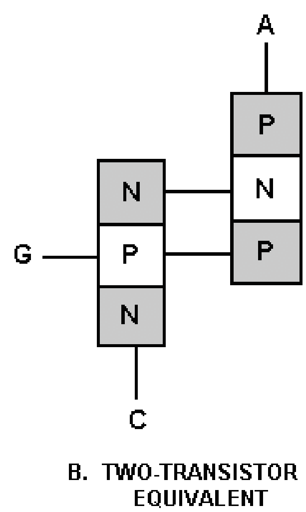
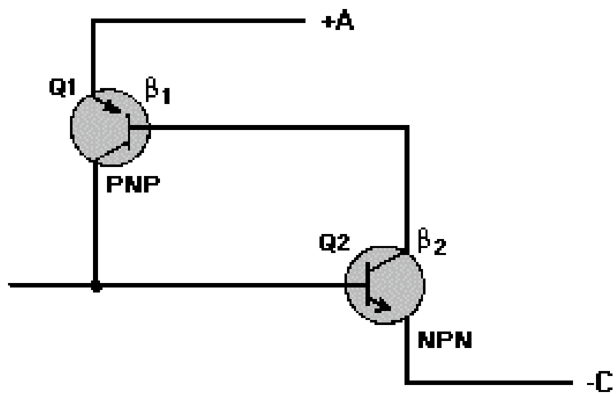


Figure 3-18B.—SCR structure.



C. TWO-TRANSISTOR SCHEMATIC

Figure 3-18C.—SCR structure.

In operation the collector of Q2 drives the base of Q1, while the collector of Q1 feeds back to the base of Q2. (Beta) 1 is the current gain of Q1, and (Beta) 2 is the current gain of Q2. The gain of this positive feedback loop is their product, 1 times 2. When the product is less than one, the circuit is stable; if the product is greater than unity, the circuit is regenerative. A small negative current applied to terminal G will bias the NPN transistor into cutoff, and the loop gain is less than unity. Under these conditions, the only current that can exist between output terminals A and C is the very small cutoff collector current of the two transistors. For this reason the impedance between A and C is very high.

When a positive current is applied to terminal G, transistor Q2 is biased into conduction, causing its collector current to rise. Since the current gain of Q2 increases with increased collector current, a point (called the breakover point) is reached where the loop gain equals unity and the circuit becomes regenerative. At this point, collector current of the two transistors rapidly increases to a value limited only by the external circuit. Both transistors are driven into saturation, and the impedance between A and C is very low. The positive current applied to terminal G, which served to trigger the self-regenerative action, is no longer required since the collector of PNP transistor Q1 now supplies more than enough current to drive Q2. The circuit will remain on until it is turned off by a reduction in the collector current to a value below that necessary to maintain conduction.

The characteristic curve for the SCR is shown in figure 3-19. With no gate current, the leakage current remains very small as the forward voltage from cathode to anode is increased until the breakdown point is reached. Here the center junction breaks down, the SCR begins to conduct heavily, and the drop across the SCR becomes very low.

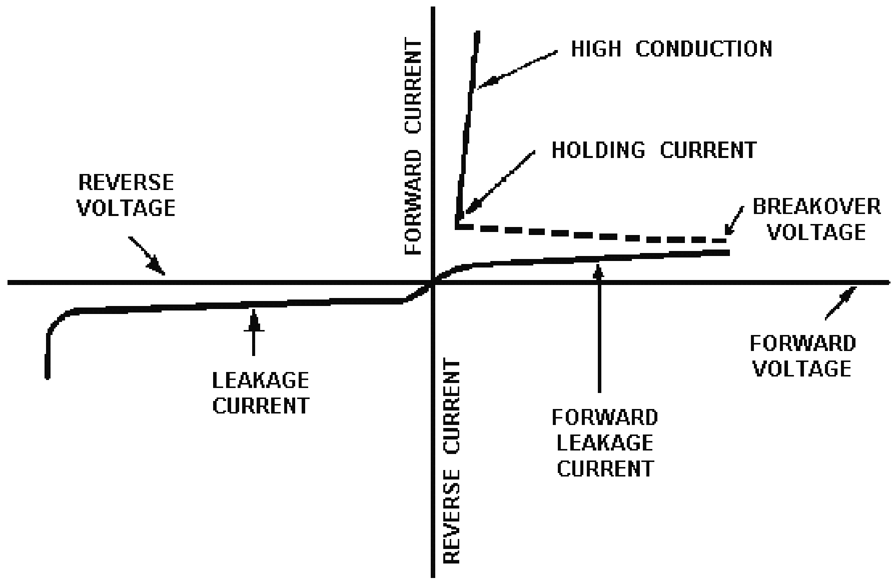


Figure 3-19.—Characteristic curve for an SCR.

The effect of a gate signal on the firing of an SCR is shown in figure 3-20. Breakdown of the center junction can be achieved at speeds approaching a microsecond by applying an appropriate signal to the gate lead, while holding the anode voltage constant. After breakdown, the voltage across the device is so low that the current through it from cathode to anode is essentially determined by the load it is feeding.

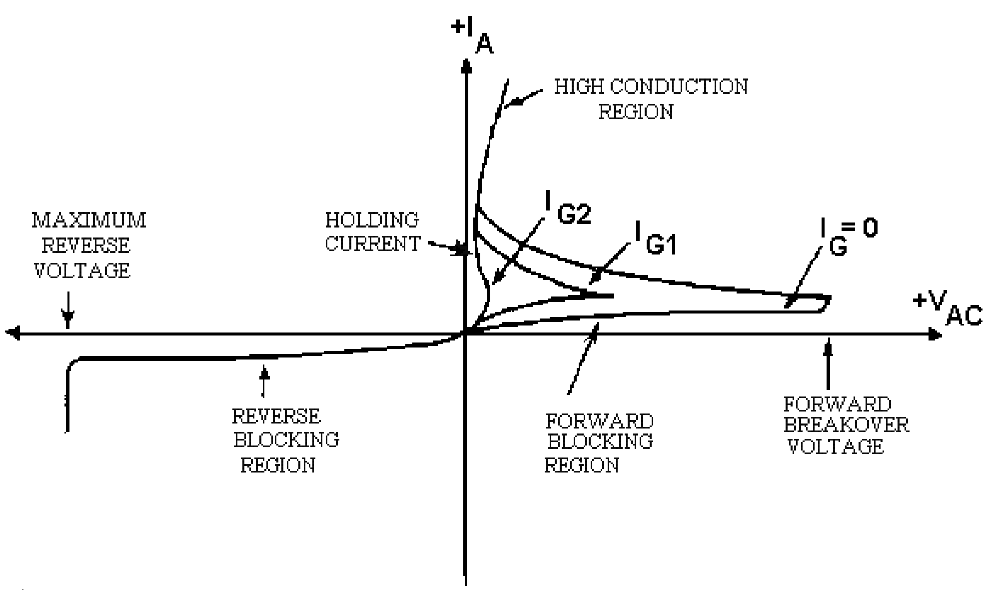


Figure 3-20.—SCR characteristic curve with various gate signals.

The important thing to remember is that a small current from gate to cathode can fire or trigger the SCR, changing it from practically an open circuit to a short circuit. The only way to change it back again (to commutate it) is to reduce the load current to a value less than the minimum forward-bias current. Gate current is required only until the anode current has completely built up to a point sufficient to sustain conduction (about 5 microseconds in resistive-load circuits). After conduction from cathode to anode begins, removing the gate current has no effect.

The basic operation of the SCR can be compared to that of the thyratron. The thyratron is an electron tube, normally gas filled, that uses a filament or a heater. The SCR and the thyratron function in a very similar manner. Figure 3-21 shows the schematic of each with the corresponding elements labeled. In both types of devices, control by the input signal is lost after they are triggered. The control grid (thyratron) and the gate (SCR) have no further effect on the magnitude of the load current after conduction begins. The load current can be interrupted by one or more of three methods: (1) the load circuit must be opened by a switch, (2) the plate (anode) voltage must be reduced below the ionizing potential of the gas (thyratron), (3) the forward-bias current must be reduced below a minimum value required to sustain conduction (SCR). The input resistance of the SCR is relatively low (approximately 100 ohms) and requires a current for triggering; the input resistance of the thyratron is exceptionally high, and requires a voltage input to the grid for triggering action.

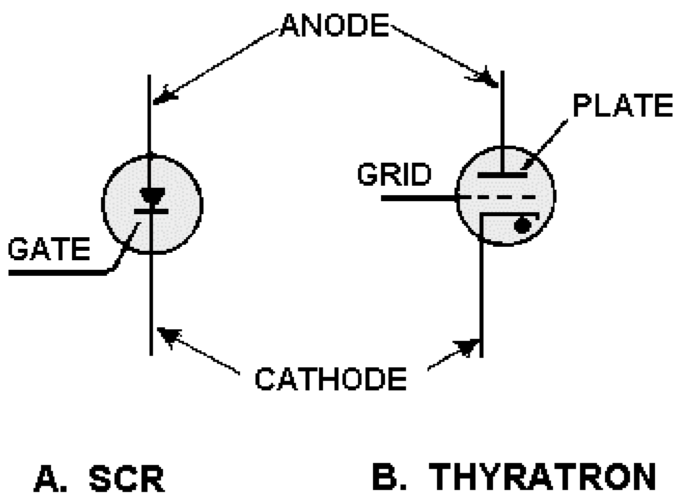


Figure 3-21.—Comparison of an SCR and a thyratron.

The applications of the SCR as a rectifier are many. In fact, its many applications as a rectifier give this semiconductor device its name. When alternating current is applied to a rectifier, only the positive or negative halves of the sine wave flow through. All of each positive or negative half cycle appears in the output. When an SCR is used, however, the controlled rectifier may be turned on at any time during the half cycle, thus controlling the amount of dc power available from zero to maximum, as shown in figure 3-22. Since the output is actually dc pulses, suitable filtering can be added if continuous direct current is needed. Thus any dc operated device can have controlled amounts of power applied to it. Notice that the SCR must be turned on at the desired time for each cycle.

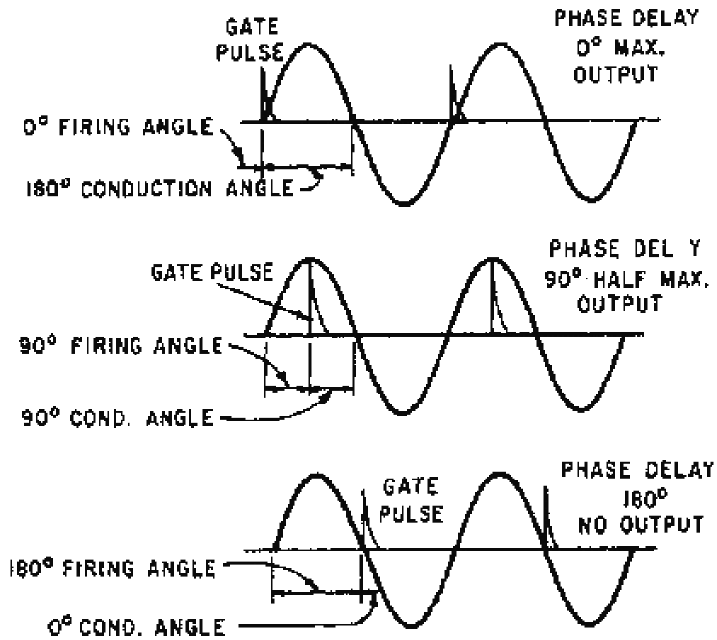


Figure 3-22.—SCR gate control signals.

When an ac power source is used, the SCR is turned off automatically, since current and voltage drop to zero every half cycle. By using one SCR on positive alternations and one on negative, full-wave rectification can be accomplished, and control is obtained over the entire sine wave. The SCR serves in this application just as its name implies—as a controlled rectifier of ac voltage.

Q14. The SCR is primarily used for what function?

Q15. When an SCR is forward biased, what is needed to cause it to conduct?

Q16. What is the only way to cause an SCR to stop conducting?

TRIAC

The TRIAC is a three-terminal device similar in construction and operation to the SCR. The TRIAC controls and conducts current flow during both alternations of an ac cycle, instead of only one. The schematic symbols for the SCR and the TRIAC are compared in figure 3-23. Both the SCR and the TRIAC have a gate lead. However, in the TRIAC the lead on the same side as the gate is "main terminal 1," and the lead opposite the gate is "main terminal 2." This method of lead labeling is necessary because the TRIAC is essentially two SCRs back to back, with a common gate and common terminals. Each terminal is, in effect, the anode of one SCR and the cathode of another, and either terminal can receive an input. In fact, the functions of a TRIAC can be duplicated by connecting two actual SCRs as shown in figure 3-24. The result is a three-terminal device identical to the TRIAC. The common anode-cathode connections form main terminals 1 and 2, and the common gate forms terminal 3.

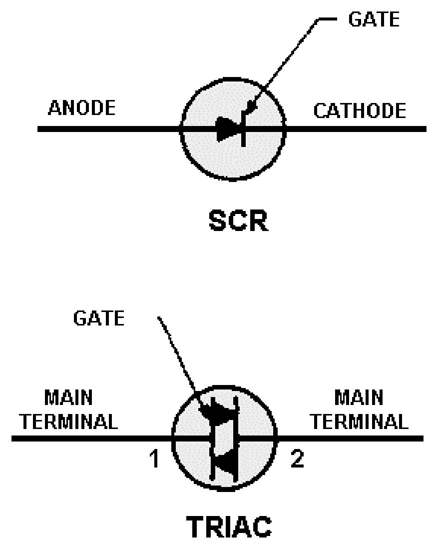


Figure 3-23.—Comparison of SCR and TRIAC symbols.

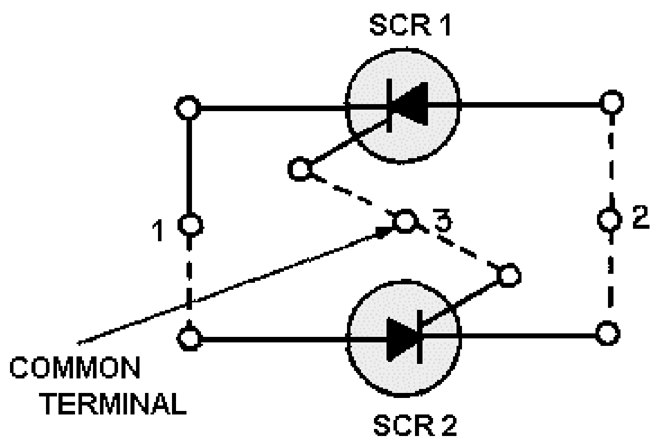


Figure 3-24.—Back to back SCR equivalent circuit.

The difference in current control between the SCR and the TRIAC can be seen by comparing their operation in the basic circuit shown in figure 3-25.

In the circuit shown in view A, the SCR is connected in the familiar half-wave arrangement. Current will flow through the load resistor (R_L) for one alternation of each input cycle. Diode CR1 is necessary to ensure a positive trigger voltage.

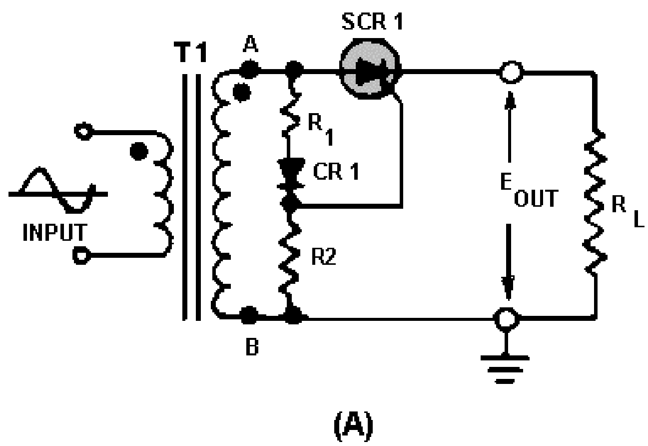


Figure 3-25A.—Comparison of SCR and TRIAC circuits.

In the circuit shown in view B, with the TRIAC inserted in the place of the SCR, current flows through the load resistor during both alternations of the input cycle. Because either alternation will trigger the gate of the TRIAC, CR1 is not required in the circuit. Current flowing through the load will reverse direction for half of each input cycle. To clarify this difference, a comparison of the waveforms seen at the input, gate, and output points of the two devices is shown in figure 3-26.

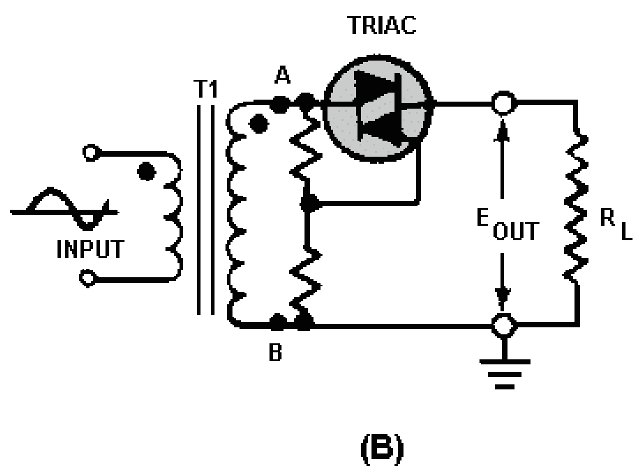


Figure 3-25B.—Comparison of SCR and TRIAC circuits.

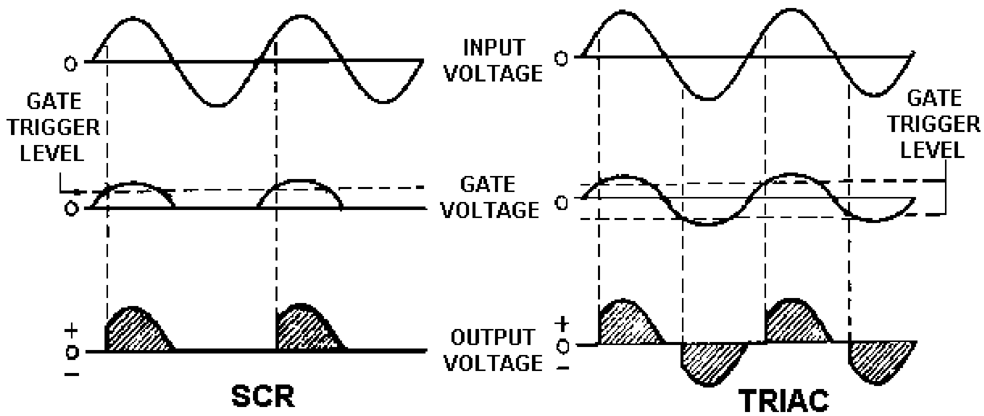


Figure 3-26.—Comparison of SCR and TRIAC waveforms.

Q17. The TRIAC is similar in operation to what device?

Q18. When used for ac current control, during which alternation of the ac cycle does the TRIAC control current flow?

Optoelectronic Devices

OPTOELECTRONIC devices either produce light or use light in their operation. The first of these, the light-emitting diode (LED), was developed to replace the fragile, short-life incandescent light bulbs used to indicate on/off conditions on panels. A LIGHT-EMITTING DIODE is a diode which, when forward biased, produces visible light. The light may be red, green, or amber, depending upon the material used to make the diode.

Figure 3-27 shows an LED and its schematic symbol. The LED is designated by a standard diode symbol with two arrows pointing away from the cathode. The arrows indicate light leaving the diode. The circuit symbols for all optoelectronic devices have arrows pointing either toward them, if they use light, or away from them, if they produce light. The LED operating voltage is small, about 1.6 volts forward bias and generally about 10 milliamperes. The life expectancy of the LED is very long, over 100,000 hours of operation.

Figure 3-27.—LED.

LEDs are used widely as "power on" indicators of current and as displays for pocket calculators, digital voltmeters, frequency counters, etc. For use in calculators and similar devices, LEDs are typically placed together in seven-segment displays, as shown in figure 3-28 (view A and view B). This display uses seven LED segments, or bars (labeled A through G in the figure), which can be lit in different combinations to form any number from "0" through "9." The schematic, view A, shows a common-anode display. All anodes in a display are internally connected. When a negative voltage is applied to the proper cathodes, a number is formed. For example, if negative voltage is applied to all cathodes except that of LED "E," the number "9" is produced, as shown in view A of figure 3-29. If the negative voltage is changed and applied to all cathodes except LED "B," the number "9" changes to "6" as shown in view B.

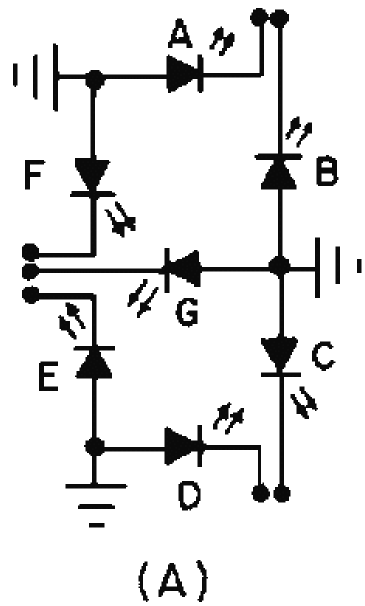


Figure 3-28A.—Seven-segment LED display.

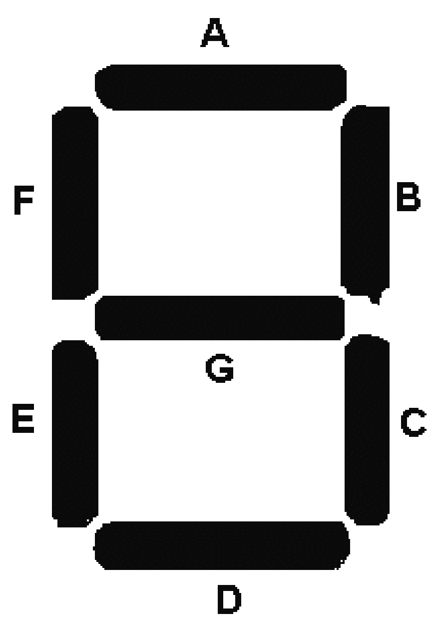


Figure 3-28B.—Seven-segment LED display.

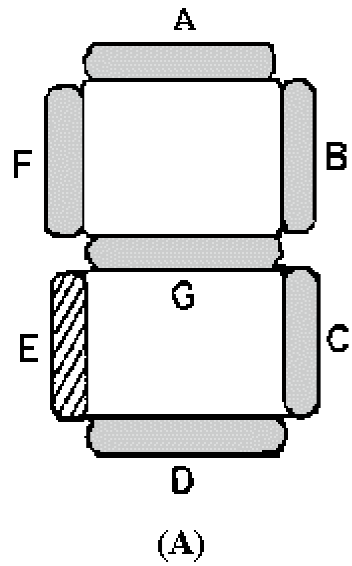


Figure 3-29A.—Seven-segment LED display examples.

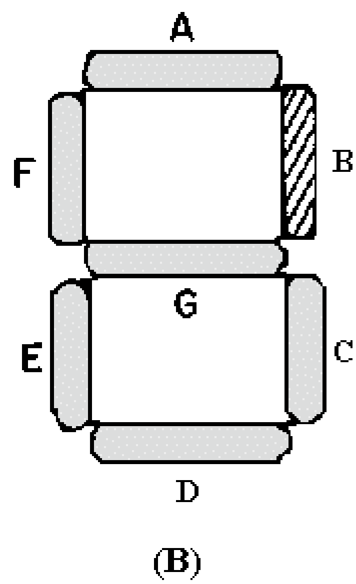


Figure 3-29B.—Seven-segment LED display examples.

Seven-segment displays are also available in common-cathode form, in which all cathodes are at the same potential. When replacing LED displays, you must ensure the replacement display is the same type as the faulty display. Since both types look alike, you should always check the manufacturer's number.

LED seven-segment displays range from the very small, often not much larger than standard typewritten numbers, to about an inch. Several displays may be combined in a package to show a series of numbers, such as the one shown in figure 3-30.

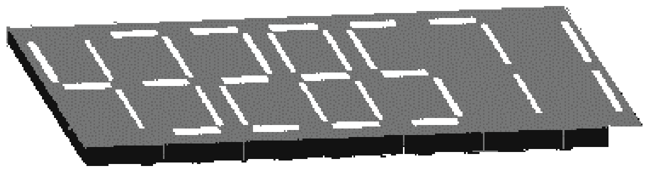


Figure 3-30.—Stacked seven-segment display.

Another special optoelectronic device in common use today is the photodiode. Unlike the LED, which produces light, the photodiode uses light to accomplish special circuit functions. Basically, the PHOTODIODE is a light-controlled variable resistor. In total darkness, it has a relatively high resistance and therefore conducts little current. However, when the PN junction is exposed to an external light source, internal resistance decreases and current flow increases. The photodiode is operated with reverse-bias and conducts current in direct proportion to the intensity of the light source.

Figure 3-31 shows a photodiode with its schematic symbol. The arrows pointing toward the symbol indicate that light is required for operation of the device. A light source is aimed at the photodiode through a transparent "window" placed over the semiconductor chip. Switching the light source on or off changes the conduction level of the photodiode. Varying the light intensity controls the amount of conduction. Because photodiodes respond quickly to changes in light intensity, they are extremely useful in digital applications such as computer card readers, paper tape readers, and photographic light meters. They are also used in some types of optical scanning equipment.

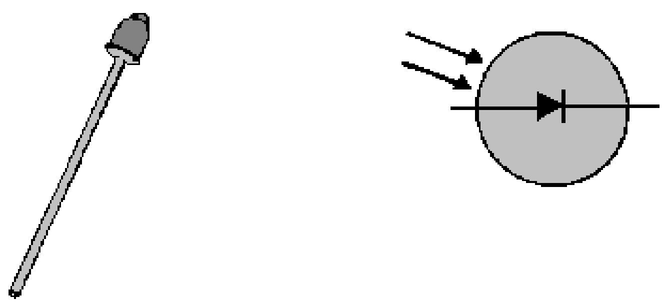


Figure 3-31.—Photodiode.

A second optoelectronic device that conducts current when exposed to light is the PHOTOTRANSISTOR. A phototransistor, however, is much more sensitive to light and produces more output current for a given light intensity that does a photodiode. Figure 3-32 shows one type of phototransistor, which is made by placing a photodiode in the base circuit of an NPN transistor. Light falling on the photodiode changes the base current of the transistor, causing the collector current to be amplified. Phototransistors may also be of the PNP type, with the photodiode placed in the base-collector circuit.

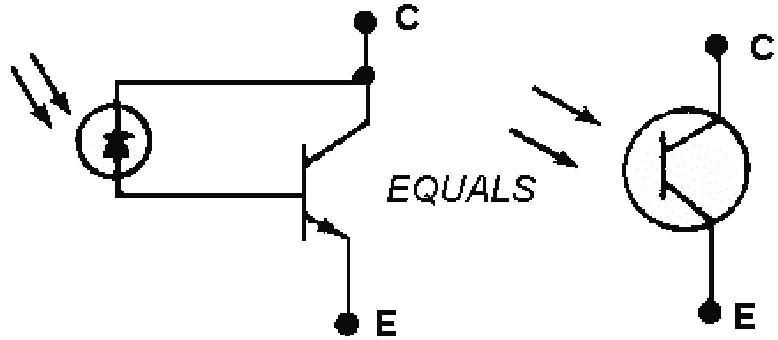


Figure 3-32.—Phototransistor.

Figure 3-33 illustrates the schematic symbols for the various types of phototransistors. Phototransistors may be of the two-terminal type, in which the light intensity on the photodiode alone determines the amount of conduction. They may also be of the three-terminal type, which have an added base lead that allows an electrical bias to be applied to the base. The bias allows an optimum transistor conduction level, and thus compensates for ambient (normal room) light intensity.

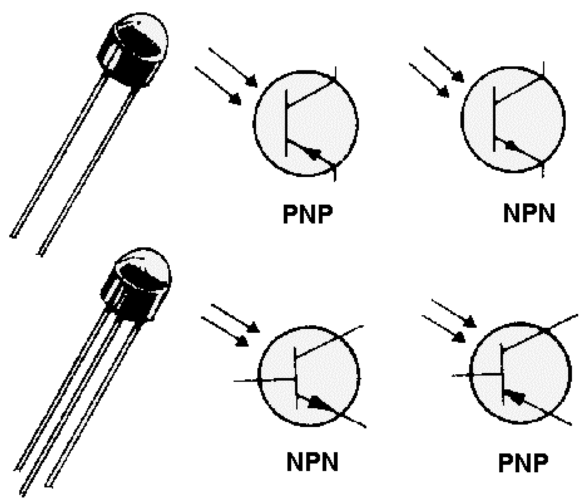


Figure 3-33.—2-terminal and 3-terminal phototransistors.

An older device that uses light in a way similar to the photodiode is the photoconductive cell, or PHOTOCELL, shown with its schematic symbol in figure 3-34. Like the photodiode, the photocell is a light-controlled variable resistor. However, a typical light-to-dark resistance ratio for a photocell is 1:1000. This means that its resistance could range from 1000 ohms in the light to 1000 kilohms in the dark, or from 2000 ohms in the light to 2000 kilohms in the dark, and so forth. Of course, other ratios are also available. Photocells are used in various types of control and timing circuits as, for example, the automatic street light controllers in most cities.

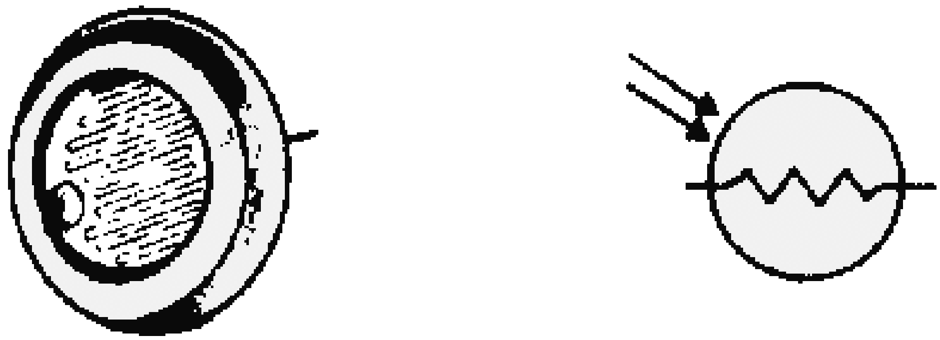


Figure 3-34.—Photocell.

The photovoltaic cell, or solar cell, is a device which converts light energy into electrical energy. An example of a solar cell and its schematic symbol are shown in figure 3-35. The symbol is similar to that of a battery. The device itself acts much like a battery when exposed to light and produces about .45 volt across its terminals, with current capacity determined by its size. As with batteries, solar cells may be connected in series or parallel to produce higher voltages and currents. The device is finding widespread application in communications satellites and solar-powered homes.

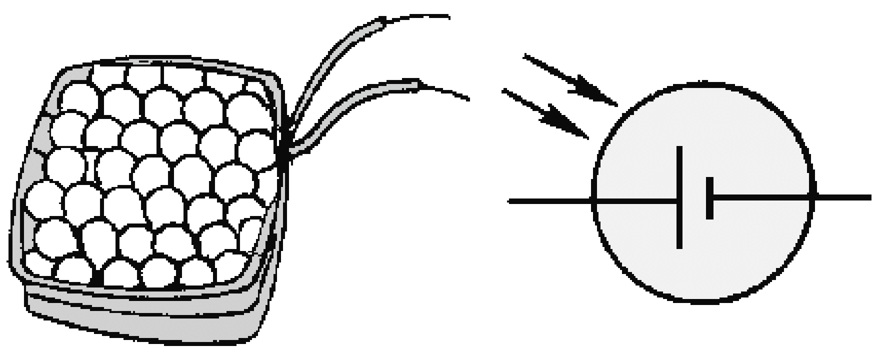


Figure 3-35.—Solar cell.

When it is necessary to block the voltage between one electronic circuit and another, and transfer the signal at the same time, an amplifier coupling capacitor is often used as shown in figure 3-36. Although this method of coupling does block dc between the circuits, voltage isolation is not complete. A newer method, making use of optoelectronic devices to achieve electrical isolation, is the optical coupler, shown in figure 3-37. The coupler is composed of an LED and a photodiode contained in a light-conducting medium. As the polarity signs in figure 3-37 show, the LED is forward biased, while the photodiode is reverse biased. When the input signal causes current through the LED to increase, the light produce by the LED increases. This increased light intensity causes current flow through the photodiode to increase. In this way, changes in input current produce proportional changes in the output, even though the two circuits are electrically isolated.

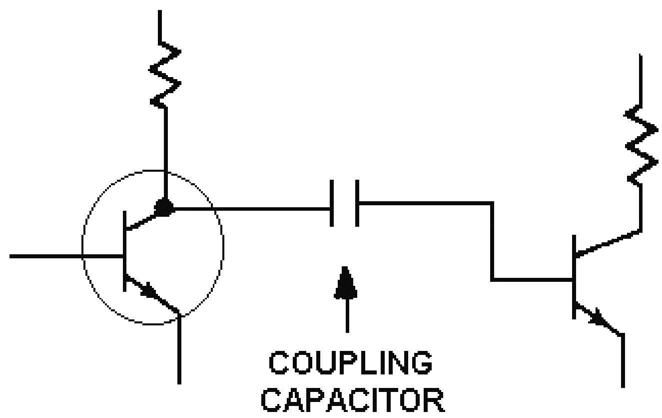


Figure 3-36.—Dc blocking with a coupling capacitor.

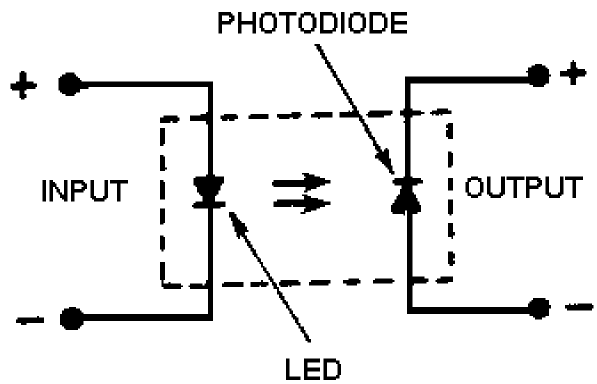


Figure 3-37.—Optical coupler.

The optical coupler is suitable for frequencies in the low megahertz range. The photodiode type shown above can handle only small currents; however, other types of couplers, combining phototransistors with the SCR, can be used where more output is required. Optical couplers are replacing transformers in low-voltage and low-current applications. Sensitive digital circuits can use the coupler to control large current and voltages with low-voltage logic levels.

Q19. What type of bias is required to cause an LED to produce light?

Q20. When compared to incandescent lamps, what is the power requirement of an LED?

Q21. In a common anode, seven-segment LED display, an individual LED will light if a negative voltage is applied to what element?

Q22. What is the resistance level of a photodiode in total darkness?

Q23. What type of bias is required for proper operation of a photodiode?

Q24. What is a typical light-to-dark resistance ratio for a photocell?

Q25. What semiconductor device produces electrical energy when exposed to light?

TRANSISTORS

Transistors are semiconductor devices with three or more terminals. The operation of normal transistors has already been discussed, but there are several transistors with special properties that should be explained. As with diodes, a discussion of all the developments in the transistor field would be impossible. The unijunction transistor (UJT) and the field effect transistor (FET) will be discussed because of their widespread application in Navy equipment. Many other special transistors have been developed and will be discussed in later *NEETS* modules.

The Unijunction Transistor (UJT)

The UNIJUNCTION TRANSISTOR (UJT), originally called a double-based diode, is a three-terminal, solid-state device that has several advantages over conventional transistors. It is very stable over a wide range of temperatures and allows a reduction of components when used in place of conventional transistors. A comparison is shown in figure 3-38. View A is a circuit using conventional transistors, and view B is the same circuit using the UJT. As you can see, the UJT circuit has fewer components. Reducing the number of components reduces the cost, size, and probability of failure.

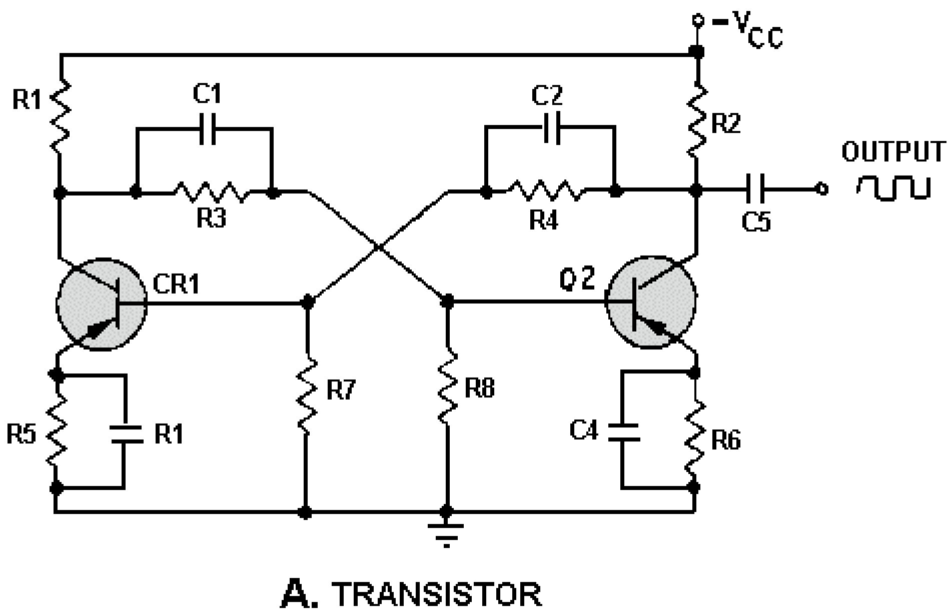


Figure 3-38A.—Comparison of conventional transistors and UJT circuits.

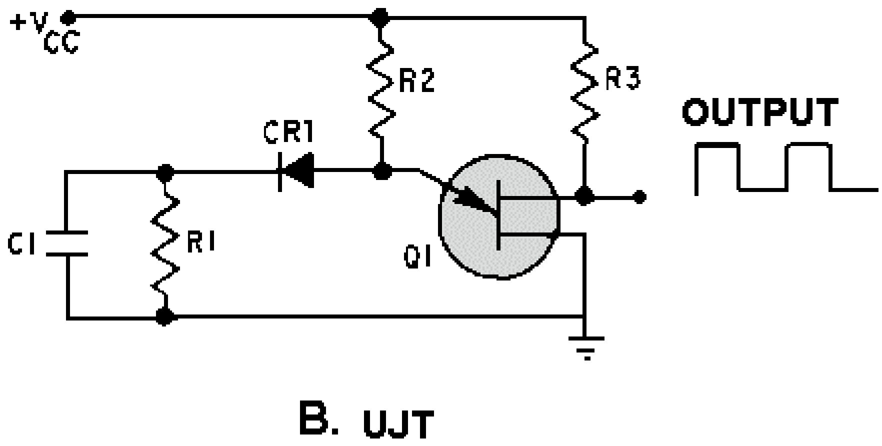


Figure 3-38B.—Comparison of conventional transistors and UJT circuits.

The physical appearance of the UJT is identical to that of the common transistor. As shown in figure 3-39, both have three leads and the same basic shape; the tab on the case indicates the emitter on both devices. The UJT, however, has a second base instead of a collector.

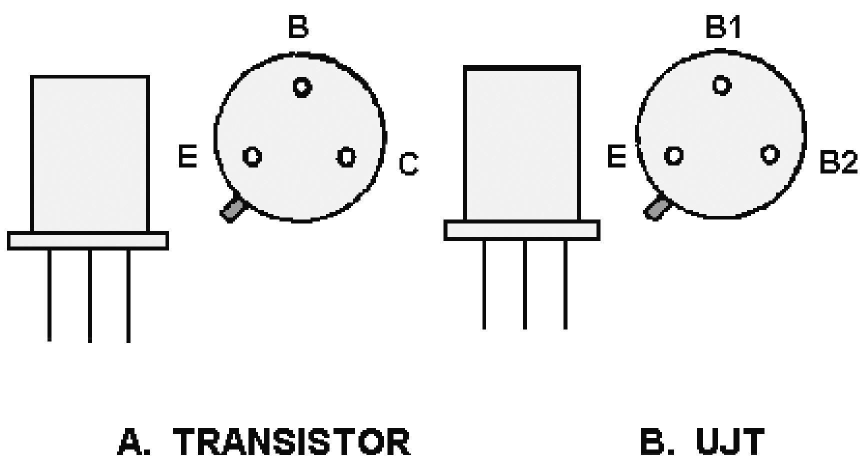


Figure 3-39.—Transistor and UJT.

As indicated in the block diagram shown in views A and B of figure 3-40, the lead differences are even more pronounced. Unlike the transistor, the UJT has only one PN junction. The area between base 1 and base 2 acts as a resistor when the UJT is properly biased. A conventional transistor needs a certain bias level between the emitter, base, and collector for proper conduction. The same principle is true for the UJT; it needs a certain bias level between the emitter and base 1 and also between base 1 and base 2 for proper conduction.

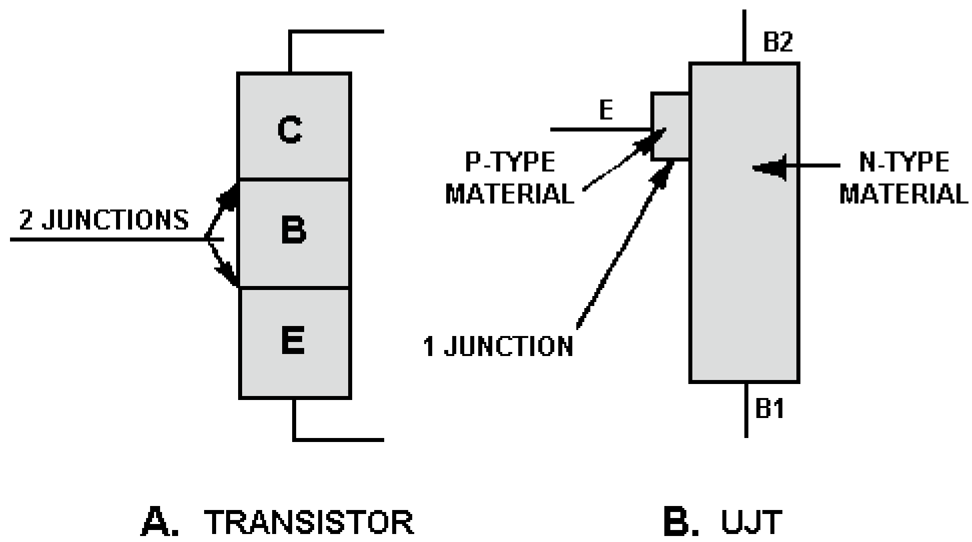


Figure 3-40.—Transistor and UJT structure.

The normal bias arrangement for the UJT is illustrated in figure 3-41, view A. A positive 10 volts is placed on base 2 and a ground on base 1. The area between base 1 and base 2 acts as a resistor. If a reading were taken between base 1 and base 2, the meter would indicate the full 10 volts as shown in view B. Theoretically, if one meter lead were connected to base 1 and the other lead to some point between base 1 and base 2, the meter would read some voltage less than 10 volts. This concept is illustrated in figure 3-42, view A. View B is an illustration of the voltage levels at different points between the two bases. The sequential rise in voltage is called a voltage gradient.

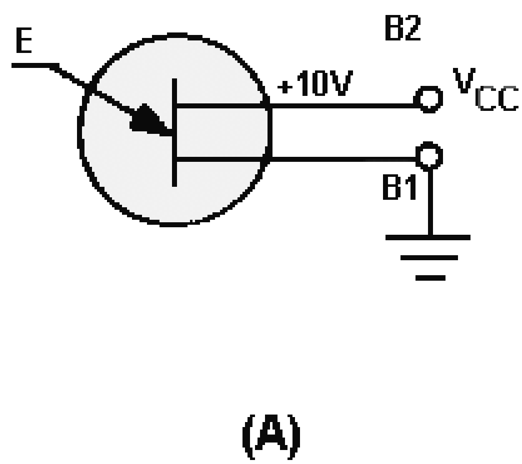


Figure 3-41A.—UJT biasing.

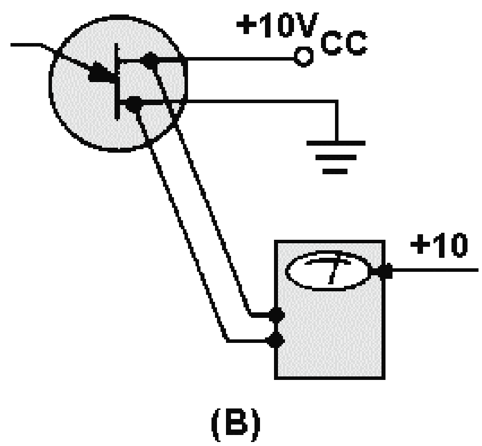


Figure 3-41B.—UJT biasing.

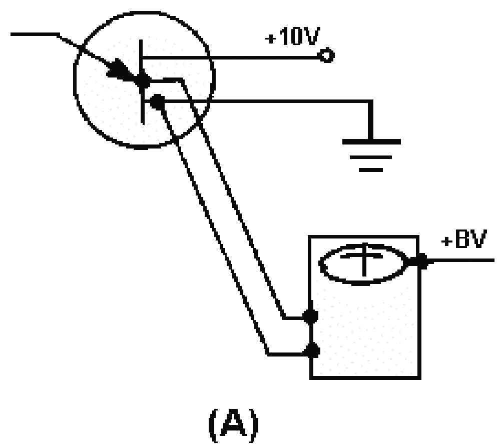


Figure 3-42A.—UJT voltage gradient.

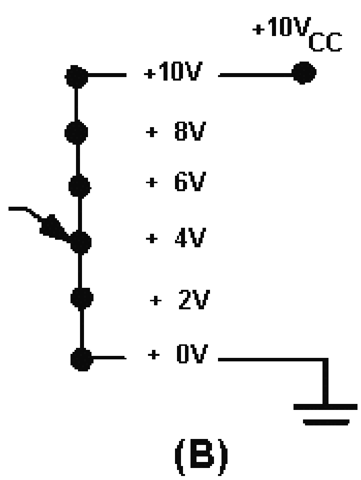


Figure 3-42B.—UJT voltage gradient.

The emitter of the UJT can be viewed as the wiper arm of a variable resistor. If the voltage level on the emitter is more positive than the voltage gradient level at the emitter-base material contact point, the UJT is forward biased. The UJT will conduct heavily (almost a short circuit) from base 1 to the emitter. The emitter is fixed in position by the manufacturer. The level of the voltage gradient therefore depends upon the amount of bias voltage, as shown in figure 3-43.

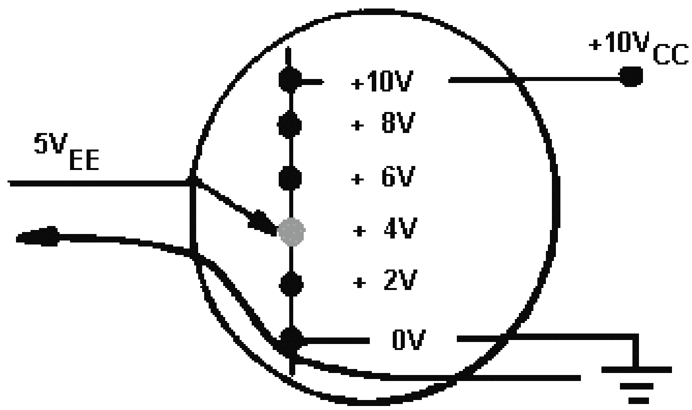


Figure 3-43.—Forward bias point on UJT voltage gradient.

If the voltage level on the emitter is less positive than the voltage gradient opposite the emitter, the UJT is reverse biased. No current will flow from base 1 to the emitter. However, a small current, called reverse current, will flow from the emitter to base 2. The reverse current is caused by the impurities used in the construction of the UJT and is in the form of minority carriers.

More than 40 distinct types of UJTs are presently in use. One of the most common applications is in switching circuits. They are also used extensively in oscillators and wave-shaping circuits.

Q26. The UJT has how many PN junctions?

Q27. The area between base 1 and base 2 in a UJT acts as what type of common circuit component?

Q28. The sequential rise in voltage between the two bases of the UJT is called what?

Q29. What is the normal current path for a UJT?

Field Effect Transistors

Although it has brought about a revolution in the design of electronic equipment, the bipolar (PNP/NPN) transistor still has one very undesirable characteristic. The low input impedance associated with its base-emitter junction causes problems in matching impedances between interstage amplifiers.

For years, scientists searched for a solution that would combine the high input impedance of the vacuum tube with the many other advantages of the transistor. The result of this research is the FIELD-EFFECT TRANSISTOR (FET). In contrast to the bipolar transistor, which uses bias current between base and emitter to control conductivity, the FET uses voltage to control an electrostatic field within the transistor. Because the FET is voltage-controlled, much like a vacuum tube, it is sometimes called the "solid-state vacuum tube."

The elements of one type of FET, the junction type (JFET), are compared with the bipolar transistor and the vacuum tube in figure 3-44. As the figure shows, the JFET is a three-element device comparable to the other two. The "gate" element of the JFET corresponds very closely in operation to the base of the transistor and the grid of the vacuum tube. The "source" and "drain" elements of the JFET correspond to the emitter and collector of the transistor and to the cathode and plate of the vacuum tube.

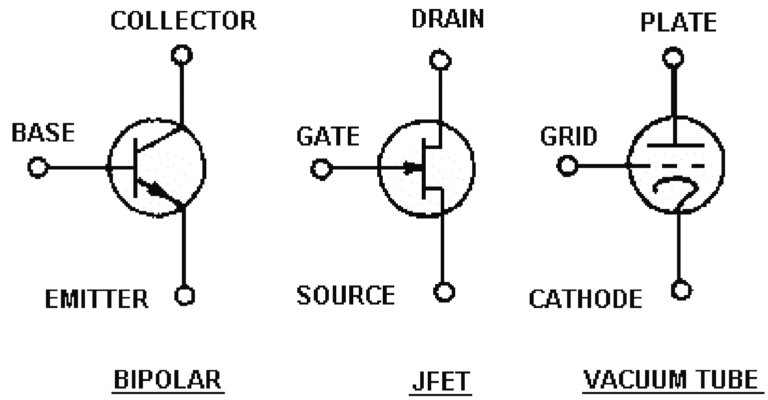


Figure 3-44.—Comparison of JFET, transistor, and vacuum tube symbols.

The construction of a JFET is shown in figure 3-45. A solid bar, made either of N-type or P-type material, forms the main body of the device. Diffused into each side of this bar are two deposits of material of the opposite type from the bar material, which form the "gate." The portion of the bar between the deposits of gate material is of a smaller cross section than the rest of the bar and forms a "channel" connecting the source and the drain. Figure 3-45 shows a bar of N-type material and a gate of P-type material. Because the material in the channel is N-type, the device is called an N-channel JFET.

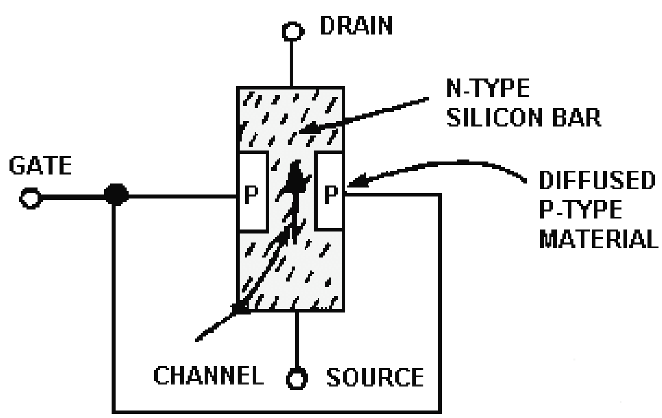


Figure 3-45.—JFET structure.

In a P-channel JFET, the channel is made of P-type material and the gate of N-type material. In figure 3-46, schematic symbols for the two types of JFET are compared with those of the NPN and PNP bipolar transistors. Like the bipolar transistor types, the two types of JFET differ only in the configuration of bias voltages required and in the direction of the arrow within the symbol. Just as it does in transistor symbols, the arrow in a JFET symbol always points towards the N-type material. Thus the symbol of the N-channel JFET shows the arrow pointing toward the drain/source channel, whereas the P-channel symbol shows the arrow pointing away from the drain/source channel toward the gate.

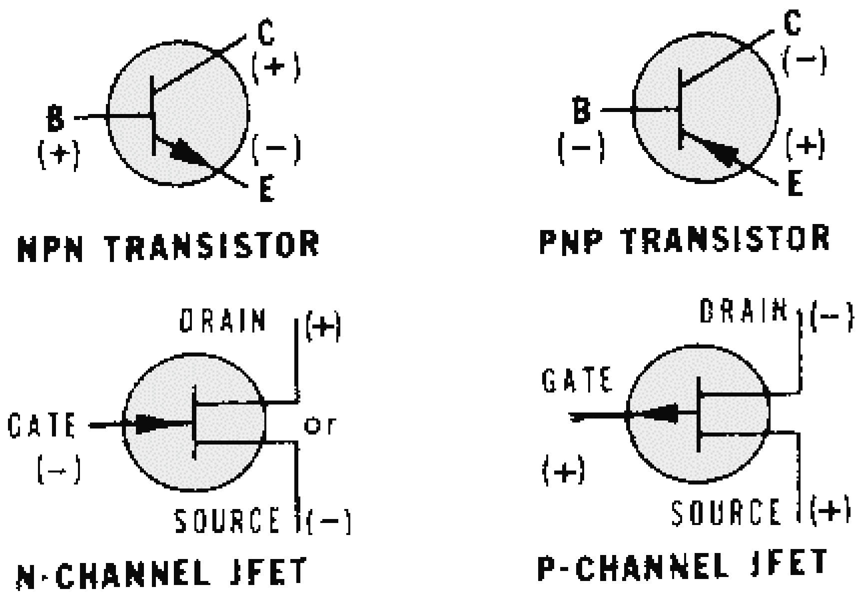


Figure 3-46.—Symbols and bias voltages for transistors and JFET.

The key to FET operation is the effective cross-sectional area of the channel, which can be controlled by variations in the voltage applied to the gate. This is demonstrated in the figures which follow.

Figure 3-47 shows how the JFET operates in a zero gate bias condition. Five volts are applied across the JFET so that current flows through the bar from source to drain, as indicated by the arrow. The gate terminal is tied to ground. This is a zero gate bias condition. In this condition, a typical bar represents a resistance of about 500 ohms. A milliammeter, connected in series with the drain lead and dc power, indicates the amount of current flow. With a drain supply (V_{DD}) of 5 volts, the milliammeter gives a drain current (I_D) reading of 10 milliamperes. The voltage and current subscript letters (V_{DD}, I_D) used for an FET correspond to the elements of the FET just as they do for the elements of transistors.

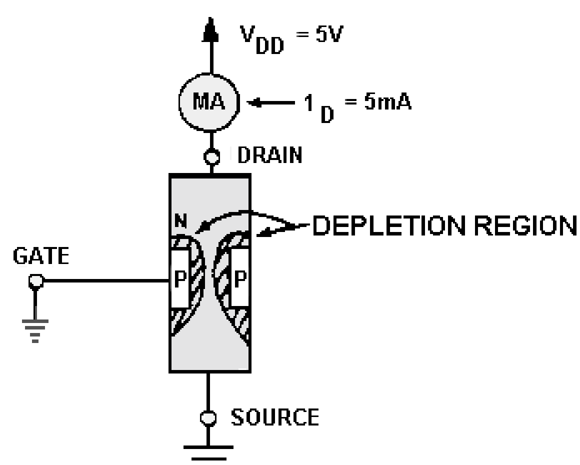


Figure 3-47.—JFET operation with zero gate bias.

In figure 3-48, a small reverse-bias voltage is applied to the gate of the JFET. A gate-source voltage (V_{GG}) of negative 1 volt applied to the P-type gate material causes the junction between the P- and N-type material to become reverse biased. Just as it did in the varactor diode, a reverse-bias condition causes a "depletion region" to form around the PN junction of the JFET. Because this region has a reduced number of current carriers, the effect of reverse biasing is to reduce the effective cross-sectional area of the "channel." This reduction in area increases the source-to-drain resistance of the device and decreases current flow.

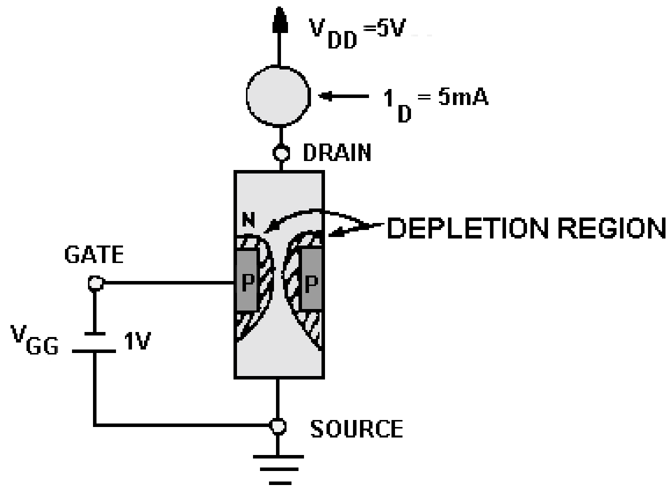


Figure 3-48.—JFET with reverse bias.

The application of a large enough negative voltage to the gate will cause the depletion region to become so large that conduction of current through the bar stops altogether. The voltage required to reduce drain current (I_D) to zero is called "pinch-off" voltage and is comparable to "cut-off" voltage in a vacuum tube. In figure 3-48, the negative 1 volt applied, although not large enough to completely stop conduction, has caused the drain current to decrease markedly (from 10 milliamperes under zero gate bias conditions to 5 milliamperes). Calculation shows that the 1-volt gate bias has also increased the resistance of the JFET (from 500 ohms to 1 kilohm). In other words, a 1-volt change in gate voltage has doubled the resistance of the device and cut current flow in half.

These measurements, however, show only that a JFET operates in a manner similar to a bipolar transistor, even though the two are constructed differently. As stated before, the main advantage of an FET is that its input impedance is significantly higher than that of a bipolar transistor. The higher input impedance of the JFET under reverse gate bias conditions can be seen by connecting a microammeter in series with the gate-source voltage (V_{GG}), as shown in figure 3-49.

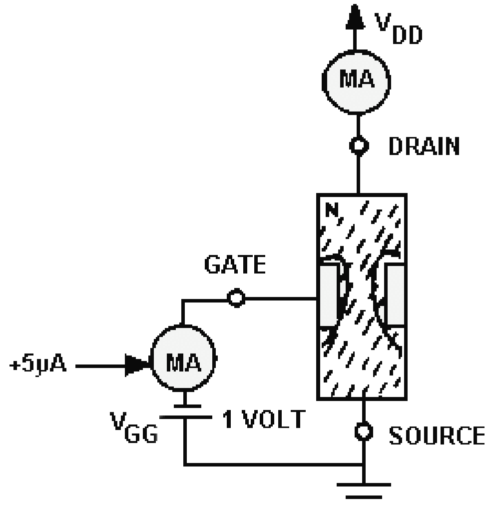


Figure 3-49.—JFET input impedance.

With a V_{GG} of 1 volt, the microammeter reads .5 microamps. Applying Ohm's law (1V ÷ .5 μA) illustrates that this very small amount of current flow results in a very high input impedance (about 2 megohms). By contrast, a bipolar transistor in similar circumstances would require higher current flow (e.g., .1 to −1 mA), resulting in a much lower input impedance (about 1000 ohms or less). The higher input impedance of the JFET is possible because of the way reverse-bias gate voltage affects the cross-sectional area of the channel.

The preceding example of JFET operation uses an N-channel JFET. However, a P-channel JFET operates on identical principles. The differences between the two types are shown in figure 3-50.

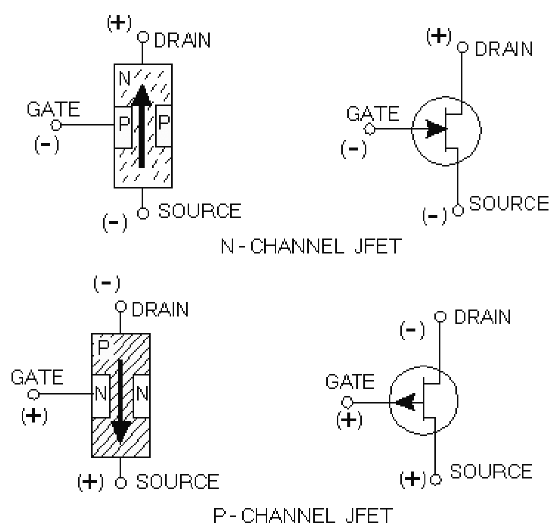


Figure 3-50.—JFET symbols and bias voltages.

Because the materials used to make the bar and the gate are reversed, source voltage potentials must also be reversed. The P-channel JFET therefore requires a positive gate voltage to be reverse biased, and current flows through it from drain to source.

Figure 3-51 shows a basic common-source amplifier circuit containing an N-channel JFET. The characteristics of this circuit include high input impedance and a high voltage gain. The function of the circuit components in this figure is very similar to those in a triode vacuum tube common-cathode amplifier circuit. C1 and C3 are the input and output coupling capacitors. R1 is the gate return resistor and functions much like the grid return resistor in a vacuum tube circuit. It prevents unwanted charge buildup on the gate by providing a discharge path for C1. R2 and C2 provide source self-bias for the JFET, which operates like cathode self-bias. R3 is the drain load resistor, which acts like the plate or collector load resistor.

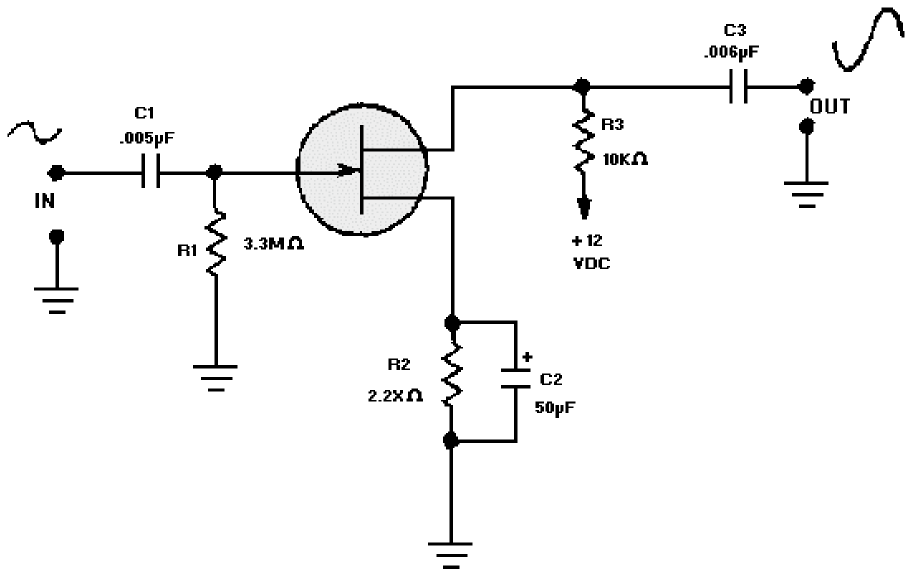


Figure 3-51.—JFET common source amplifier.

The phase shift of 180 degrees between input and output signals is the same as that of common-cathode vacuum tube circuits (and common-emitter transistor circuits). The reason for the phase shift can be seen easily by observing the operation of the N-channel JFET. On the positive alternation of the input signal, the amount of reverse bias on the P-type gate material is reduced, thus increasing the effective cross-sectional area of the channel and decreasing source-to-drain resistance. When resistance decreases, current flow through the JFET increases. This increase causes the voltage drop across R3 to increase, which in turn causes the drain voltage to decrease. On the negative alternation of the cycle, the amount of reverse bias on the gate of the JFET is increased and the action of the circuit is reversed. The result is an output signal, which is an amplified 180-degree-out-of-phase version of the input signal.

A second type of field-effect transistor has been introduced in recent years that has some advantages over the JFET. This device is the metal oxide semiconductor field effect transistor (MOSFET). The MOSFET has an even higher input impedance than the JFET (10 to 100 million megohms). Therefore, the MOSFET is even less of a load on preceding circuits. The extremely high input impedance, combined with a high gain factor, makes the MOSFET a highly efficient input device for RF/IF amplifiers and mixers and for many types of test equipment.

The MOSFET is normally constructed so that it operates in one of two basic modes: the depletion mode or the enhancement mode. The depletion mode MOSFET has a heavily doped channel and uses reverse bias on the gate to cause a depletion of current carriers in the channel. The JFET also operates in this manner. The enhancement mode MOSFET has a lightly doped channel and uses forward bias to enhance the current carriers in the channel. A MOSFET can be constructed that will operate in either mode depending upon what type of bias is applied, thus allowing a greater range of input signals.

In addition to the two basic modes of operation, the MOSFET, like the JFET, is of either the P-channel type or the N-channel type. Each type has four elements: gate, source, drain, and substrate. The schematic symbols for the four basic variations of the MOSFET are shown in views A, B, C, and D of figure 3-52.

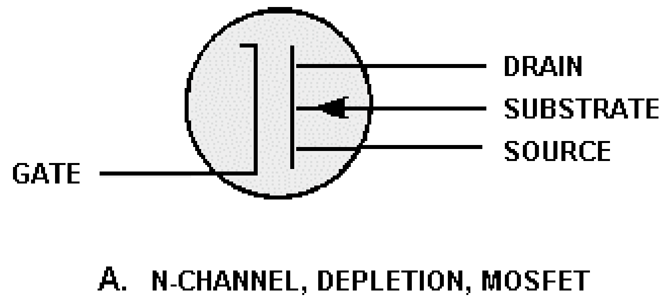


Figure 3-52A.—MOSFET symbols.

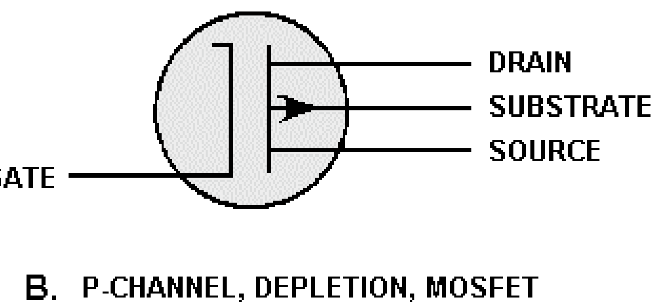


Figure 3-52B.—MOSFET symbols.

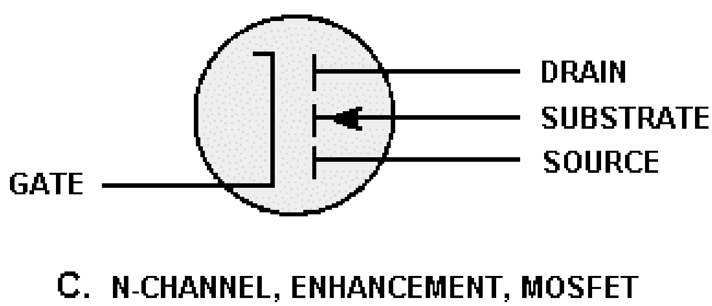


Figure 3-52C.—MOSFET symbols.

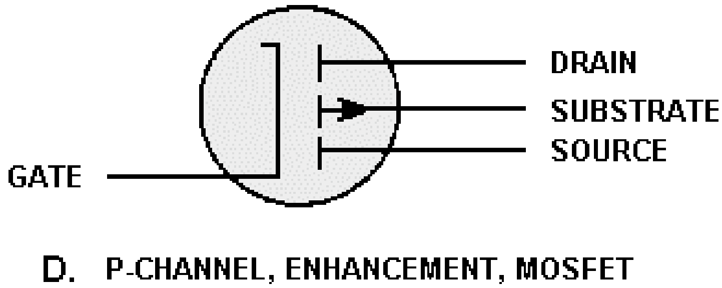


Figure 3-52D.—MOSFET symbols.

The construction of an N-channel MOSFET is shown in figure 3-53. Heavily doped N-type regions (indicated by the N+) are diffused into a P-type substrate or base. A channel of regular N-type material is diffused between the heavily doped N-type regions. A metal oxide insulating layer is then formed over the channel, and a metal gate layer is deposited over the insulating layer. There is no electrical connection between the gate and the rest of the device. This construction method results in the extremely high input impedance of the MOSFET. Another common name for the device, derived from the construction method, is the insulated gate field effect transistor (IGFET).

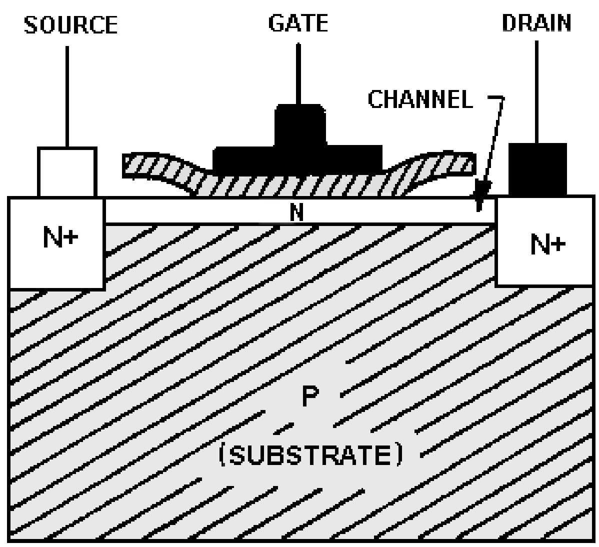


Figure 3-53.—MOSFET structure.

The operation of the MOSFET, or IGFET, is basically the same as the operation of the JFET. The current flow between the source and drain can be controlled by using either of two methods or by using a combination of the two methods. In one method the drain voltage controls the current when the gate potential is at zero volts. A voltage is applied to the gate in the second method. An electric field is formed by the gate voltage that affects the current flow in the channel by either depleting or enhancing the number of current carriers available. As previously stated, a reverse bias applied to the gate depletes the carriers, and a forward bias enhances the carriers. The polarity of the voltages required to forward or reverse bias a MOSFET depends upon whether it is of the P-channel type or the N-channel type. The effects of reverse-bias voltage on a MOSFET designed to operate in the depletion mode are illustrated in views A, B, and C of figure 3-54. The amount of reverse bias applied has a direct effect on the width of the current channel and, thus, the amount of drain current (I_D).

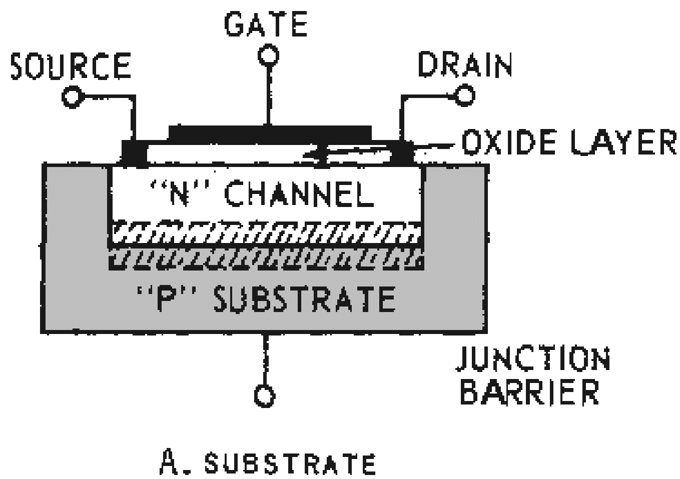


Figure 3-54A.—Effects of bias on N-channel depletion MOSFET.

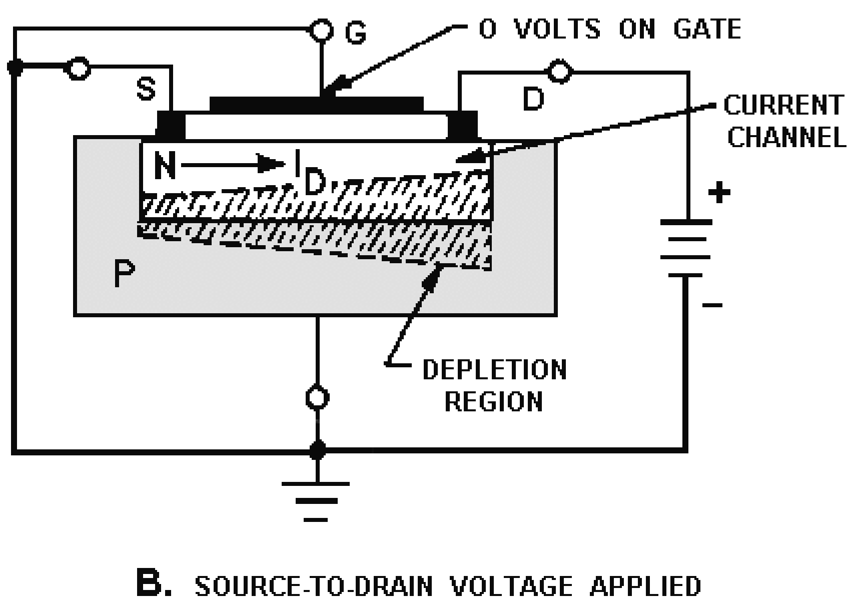


Figure 3-54B.—Effects of bias on N-channel depletion MOSFET.

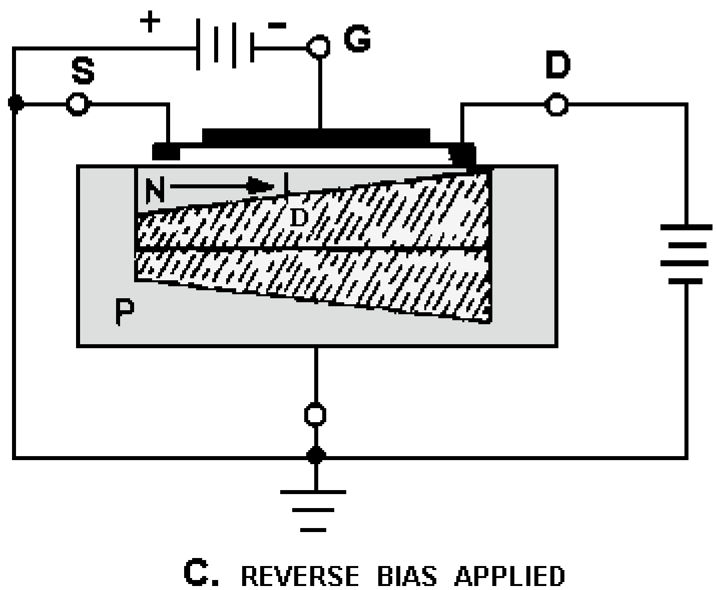


Figure 3-54C.—Effects of bias on N-channel depletion MOSFET.

Figure 3-55 (view A, view B, and view C) illustrates the effect of forward bias on an enhancement mode N-channel MOSFET. In this case, a positive voltage applied to the gate increases the width of the current channel and the amount of drain current (I_D).

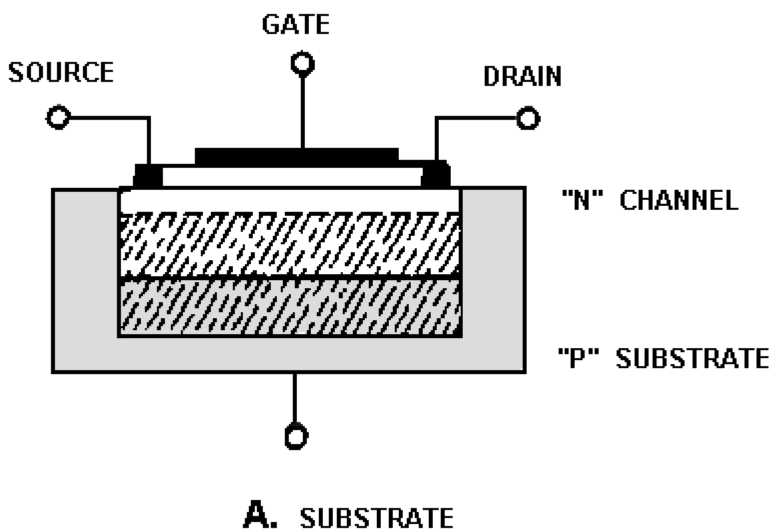


Figure 3-55A.—Effects of bias on N-channel enhancement MOSFET.

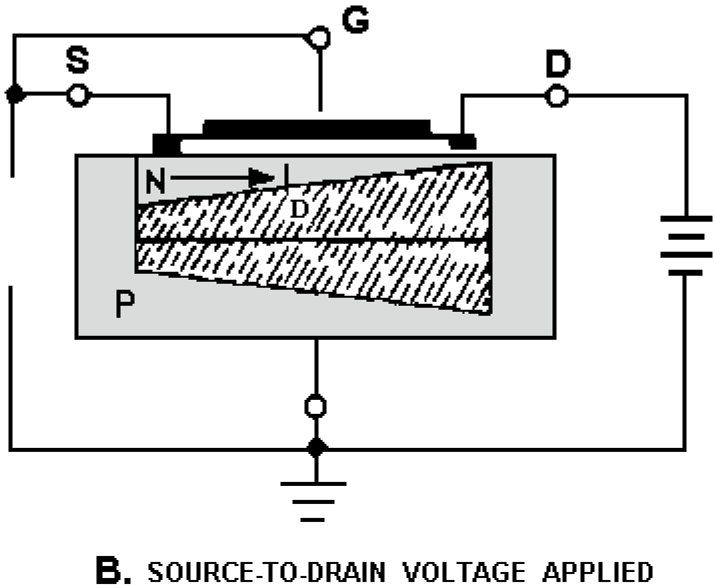


Figure 3-55B.—Effects of bias on N-channel enhancement MOSFET.

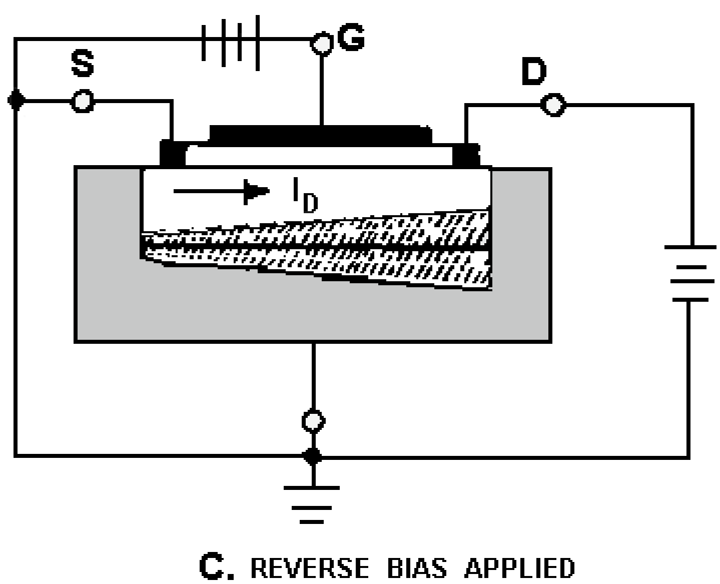


Figure 3-55C.—Effects of bias on N-channel enhancement MOSFET.

Another type of MOSFET is the induced-channel type MOSFET. Unlike the MOSFETs discussed so far, the induced-channel type has no actual channel between the source and the drain. The induced channel MOSFET is constructed by making the channel of the same type material as the substrate, or the opposite of the source and the drain material. As shown in figure 3-56, the source and the drain are of P-type material, and the channel and the substrate are of N-type material.

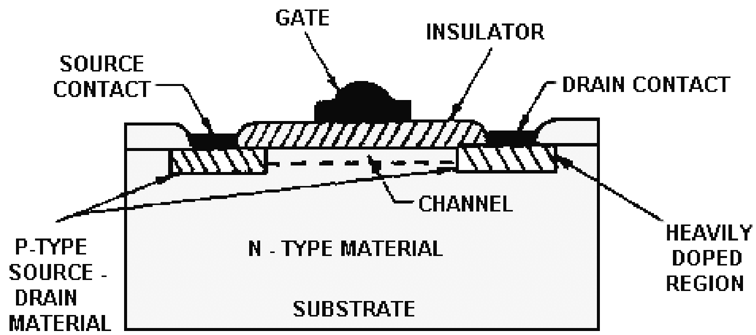


Figure 3-56.—Induced channel MOSFET construction.

The induced-channel MOSFET is caused to conduct from source to drain by the electric field that is created when a voltage is applied to the gate. For example, assume that a negative voltage is applied to the MOSFET in figure 3-56. The effect of the negative voltage modifies the conditions in the substrate material. As the gate builds a negative charge, free electrons are repelled, forming a depletion region. Once a certain level of depletion has occurred (determined by the composition of the substrate material), any additional gate bias attracts positive holes to the surface of the substrate. When enough holes have accumulated at the surface channel area, the channel changes from an N-type material to a P-type material, since it now has more positive carriers than negative carriers. At this point the channel is considered to be to inverted, and the two P-type regions at the source and the drain are now connected by a P-type inversion layer or channel. As with the MOSFET, the gate signal determines the amount of current flow through the channel as long as the source and drain voltages remain constant. When the gate voltage is at zero, essentially no current flows since a gate voltage is required to form a channel.

The MOSFETs discussed up to this point have been single-gate MOSFETs. Another type of MOSFET, the dual-gate type, is shown in figure 3-57. As the figure shows, the gates in a dual-gate MOSFET can be compared to the grids in a multi-grid vacuum tube. Because the substrate has been connected directly to the source terminal, the dual-gate MOSFET still has only four leads: one each for source and drain, and two for the gates. Either gate can control conduction independently, making this type of MOSFET a truly versatile device.

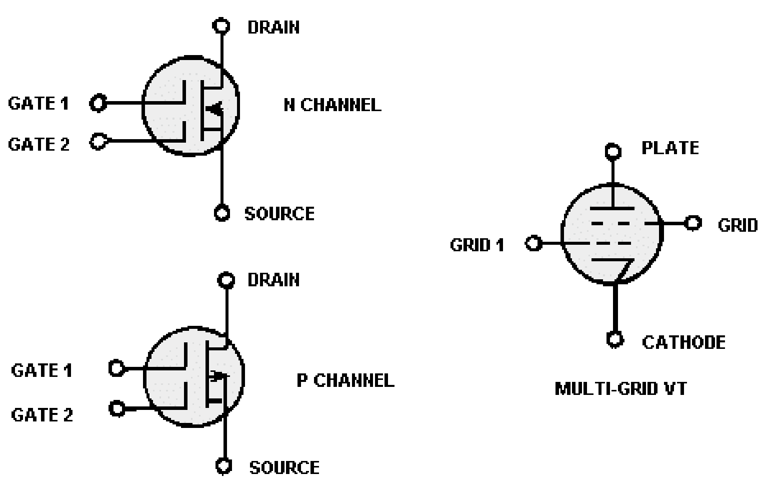


Figure 3-57.—Dual-gate MOSFET.

One problem with both the single- and dual-gate MOSFETs is that the oxide layer between gate and channel can be destroyed very easily by ordinary static electricity. Replacement MOSFETs come packaged with their leads shorted together by a special wire loop or spring to avoid accidental damage. The rule to remember with these shorting springs is that they must not be removed until after the MOSFET has been soldered or plugged into a circuit. One such spring is shown in figure 3-58.

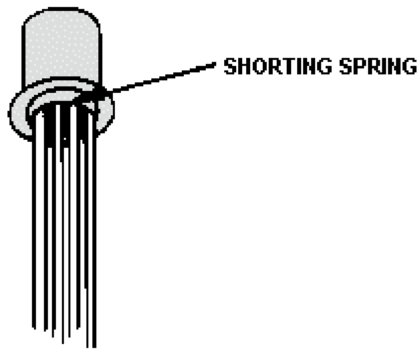


Figure 3-58.—MOSFET shorting spring.

Q30. What is one of the primary advantages of the FET when compared to the bipolar transistor?

Q31. The FET and the vacuum tube have what in common?

Q32. The base of a transistor serves a purpose similar to what element of the FET?

Q33. What are the two types of JFET?

Q34. The source and drain of an N-channel JFET are made of what type of material?

Q35. What is the key to FET operation?

Q36. What is the normal current path in an N-channel JFET?

Q37. Applying a reverse bias to the gate of an FET has what effect?

Q38. The input and output signals of a JFET amplifier have what phase relationship?

Q39. When compared to the JFET, what is the input impedance of the MOSFET?

Q40. What are the four elements of the MOSFET?

Q41. The substrate of an N-channel MOSFET is made of what material?

Q42. In a MOSFET, which element is insulated from the channel material?

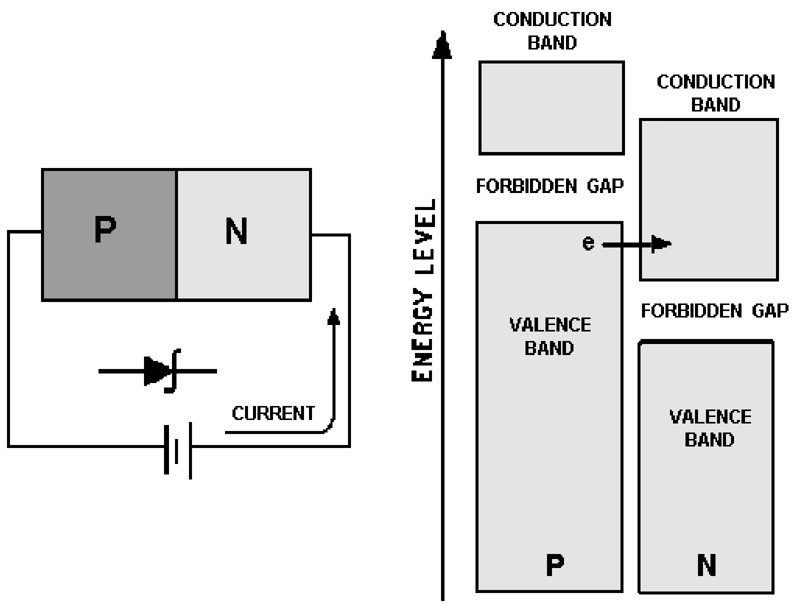
Q43. What type of MOSFET can be independently controlled by two separate signals?

Q44. What is the purpose of the spring or wire around the leads of a new MOSFET?

SUMMARY

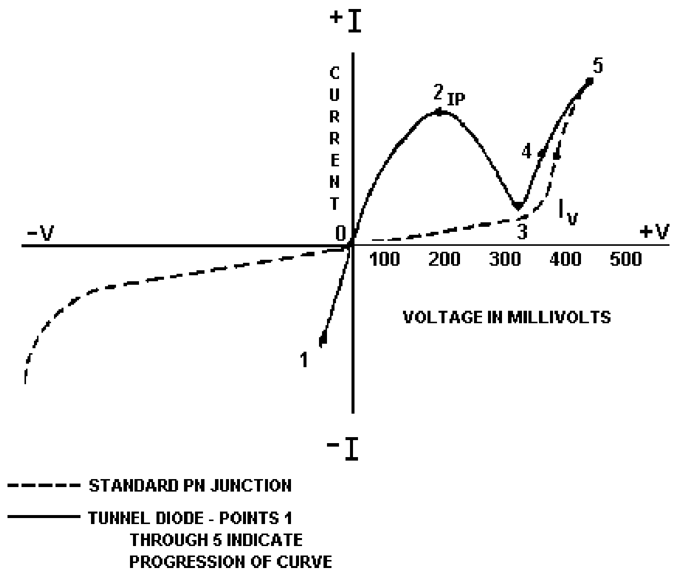
This chapter introduced you to a representative selection of solid-state devices that have special properties. The basic operating principles of the devices discussed in this chapter are summarized in the following paragraphs for you to use as a review and a future reference.

The **ZENER DIODE** is a PN junction that is designed to operate in the reverse-bias breakdown mode. When the applied voltage reaches the breakdown point, the Zener diode, for all practical purposes, becomes a short circuit. The reverse bias and breakdown mode of operation cause the Zener diode to conduct with (in the direction of) the arrow in the symbol as shown.

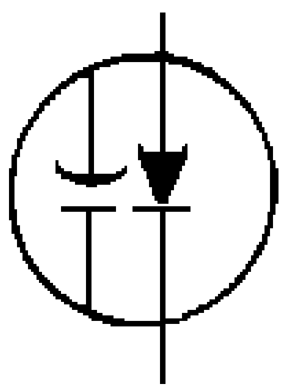


Two theories are used to explain the breakdown action of Zener diodes. The **ZENER EFFECT** explains the breakdown of diodes below 5 volts. The heavy doping used in these diodes allows the valence band of one material to overlap the energy level of the conduction band of the other material. This situation allows electrons to tunnel across the PN junction at the point where the two energy bands overlap. Zener diodes that operate above 5 volts are explained by the AVALANCHE EFFECT in which free electrons colliding with bound electrons cause an ever-increasing number of free current carriers in a multiplying action. The Zener diode is used primarily as a voltage regulator in electronic circuits.

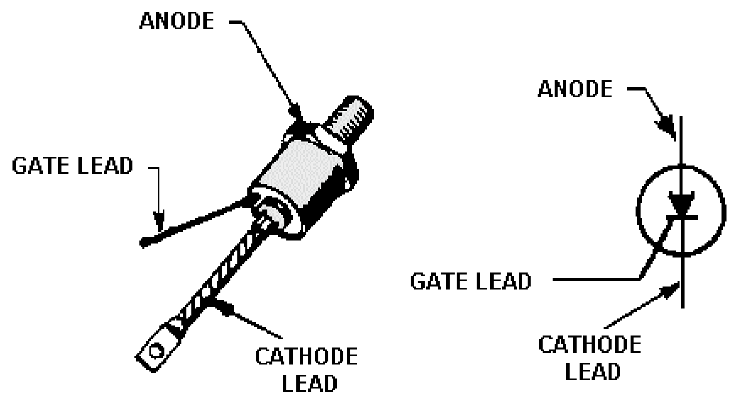
The **TUNNEL DIODE** is a heavily doped PN junction that exhibits negative resistance over part of its range of operation, as can be seen in the curve in the illustration. The heavy doping causes the tunnel diode to have a very narrow depletion region and also causes the valence band of one of the semiconductor materials to overlap the energy level of the conduction band of the other semiconductor material. At the energy overlap point, electrons can cross from the valence band of one material to the conduction band of the other material without acquiring any additional energy. This action is called tunneling. Tunnel diodes are used as amplifiers, oscillators, and high-speed switching devices.



The **VARACTOR** is a diode that exhibits the characteristics of a variable capacitor. The depletion region at the PN junction acts as the dielectric of a capacitor and is caused to expand and contract by the voltage applied to the diode. This action increases and decreases the capacitance. The schematic symbol for the varactor is shown below. Varactors are used in tuning circuits and can be used as high-frequency amplifiers.

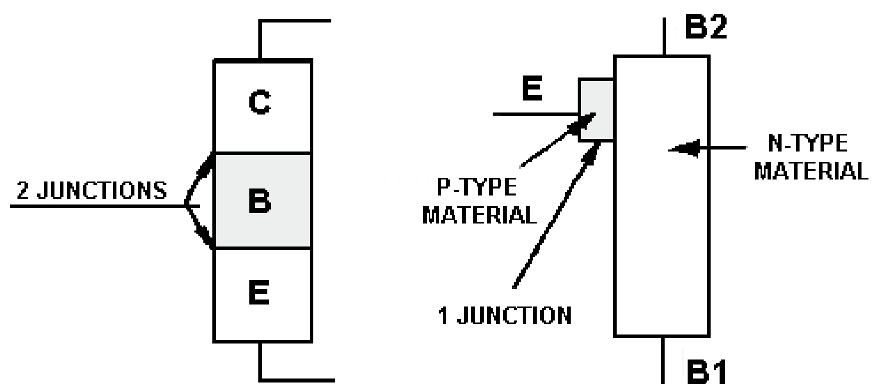


The **SILICON CONTROLLED RECTIFIER (SCR)** is a four-element, solid-state device that combines characteristics of both diodes and transistors. The symbol for the SCR is shown below. A signal must be applied to the gate to cause the SCR to conduct. When the proper gate signal is applied, the SCR conducts or "fires" until the bias potential across the device drops below the minimum required to sustain current flow. Removal of the gate signal does not shut off the SCR. In fact, the gate signal is often a very narrow voltage pulse or trigger. The SCR is ideal for use in situations where a small, low-power gate can be used to turn on larger currents, such as those found in rectifier and switching circuits. SCRs are used extensively in power supply circuits as rectifiers.

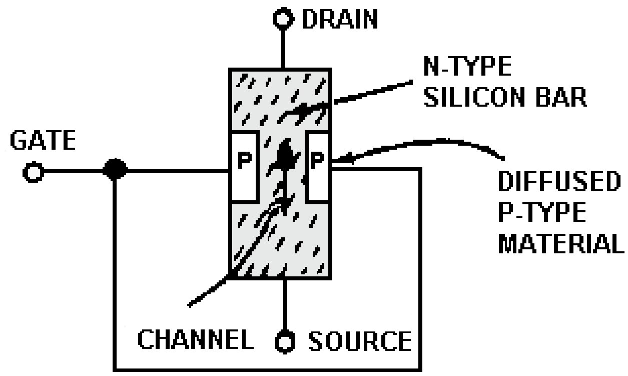


OPTOELECTRONIC DEVICES are of two basic types: light producers or light users. The LED is the most widely used light-producing device. When the LED is forward biased it emits energy in the form of light. LEDs are used in several configurations as digital equipment readout displays. The PHOTODIODE, the PHOTOTRANSISTOR, and the PHOTOCELL are all devices that use light to modify conduction through them. The SOLAR CELL uses light to produce voltage.

The **UNIJUNCTION TRANSISTOR (UJT)** is a three-terminal, solid-state device with only one PN junction. The block diagram below shows the difference in construction between normal transistors and the UJT. The area between base 1 and base 2 of the UJT acts as a variable resister. The emitter of the UJT acts as the wiper arm. The sequential rise in voltage between the bases is called a voltage gradient. The UJT conducts when the emitter is more positive than the voltage gradient at the emitter/base contact point. There are many variations of the UJT which are used in switching circuits, oscillators, and wave-shaping circuits.



The FIELD-EFFECT TRANSISTOR combines the high input impedance of the vacuum tube with all the other advantages of the transistor. The elements of the FET are the gate, source, and drain, which are comparable to the base, emitter, and collector of a standard transistor. The JFET or "junction FET" is made of a solid bar of either P- or N-semiconductor material, and the gate is made of the opposite type material, as illustrated below. The FET is called P-channel or N-channel depending upon the type of material used to make the bar between the source and drain. Voltage applied to the gate controls the width of the channel and consequently controls the current flow from the source to the drain. The JFET is normally operated with reverse bias that controls the channel width by increasing or decreasing the depletion region.



The **MOSFET** is an FET that has even higher input impedances than the JFET because the gate of the MOSFET is completely insulated from the rest of the device. The MOSFET operates in either the depletion mode or the forward-bias enhancement mode and can be either N-channel or P-channel. The induced-channel and the dual-gate MOSFETs are variations of the basic MOSFET.

ANSWERS TO QUESTIONS Q1. THROUGH Q44.

A1. *The minority carriers.*

A2. *Zener effect and avalanche effect.*

A3. *Zener effect.*

A4. *The doping level of an avalanche effect diode is lower.*

A5. *An external current-limiting resistor.*

A6. *Because Zener diodes are operated in the reverse bias mode.*

A7. *The amount of doping.*

A8. *Negative resistance.*

A9. *The tunnel diode has a very narrow depletion region.*

A10. *Minimum.*

A11. *Variable capacitance.*

A12. *The depletion region decreases.*

A13. *Capacitance decreases.*

A14. *The SCR is primarily used for switching power on or off.*

A15. *A gate signal.*

A16. The forward bias must be reduced below the minimum conduction level.

A17. SCR.

A18. During both alternations.

A19. Forward bias.

A20. Very low.

A21. The cathode.

A22. Very high.

A23. Reverse bias.

A24. 1:1000.

A25. Photovoltaic cell.

A26. One.

A27. Variable resistor.

A28. A voltage gradient.

A29. From base 1 to the emitter.

A30. High input impedance.

A31. Voltage controls conduction.

A32. Gate.

A33. N-channel and P-channel.

A34. N-type material.

A35. Effective cross-sectional area of the channel.

A36. From source to drain.

A37. Source-to-drain resistance increases.

A38. They are 180 degrees out of phase.

A39. The MOSFET has a higher input impedance.

A40. Gate, source, drain, and substrate.

A41. P-type material.

A42. The gate terminal.

A43. The dual-gate MOSFET.

A44. To prevent damage from static electricity.

CHAPTER 4

SOLID-STATE POWER SUPPLIES

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Identify the various sections of a power supply.
2. State the purpose of each section of a power supply.
3. Describe the operation of the power supply from both a whole unit standpoint and from a subunit standpoint.
4. Describe the purpose of the various types of rectifier circuits used in power supplies.
5. Describe the purpose of the various types of filter circuits used in power supplies.
6. Describe the operation of the various voltage and current regulators in a power supply.
7. Describe the operation of the various types of voltage multipliers.
8. Trace the flow of ac and dc in a power supply, from the ac input to the dc output on a schematic diagram.
9. Identify faulty components through visual checks.
10. Identify problems within specific areas of a power supply by using a logical isolation method of troubleshooting.
11. Apply safety precautions when working with electronic power supplies.

In today's Navy all electronic equipment, both ashore and on board ship, requires a power supply. The discovery of the silicon diode and other solid-state components made possible the reduction in size and the increase in reliability of electronic equipment. This is especially important on board ship where space and accessibility to spare parts are a major concern.

In this chapter, you will read about the individual sections of the power supply, their components, and the purpose of each within the power supply.

THE BASIC POWER SUPPLY

View A of figure 4-1 shows the block diagram of a basic power supply. Most power supplies are made up of four basic sections: a TRANSFORMER, a RECTIFIER, a FILTER, and a REGULATOR.

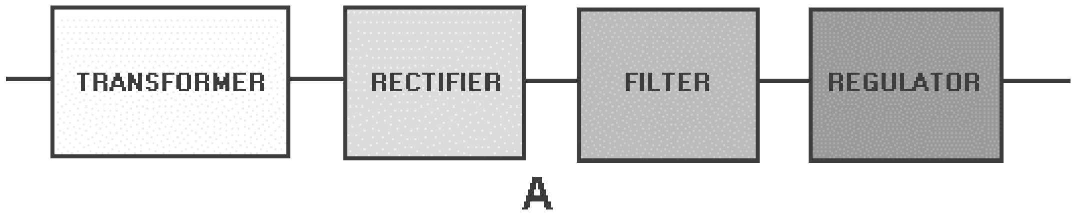


Figure 4-1A.—Block diagram of a basic power supply.

As illustrated in view B of figure 4-1, the first section is the TRANSFORMER. The transformer steps up or steps down the input line voltage and isolates the power supply from the power line. The RECTIFIER section converts the alternating current input signal to a pulsating direct current. However, as you proceed in this chapter you will learn that pulsating dc is not desirable. For this reason a FILTER section is used to convert pulsating dc to a purer, more desirable form of dc voltage.

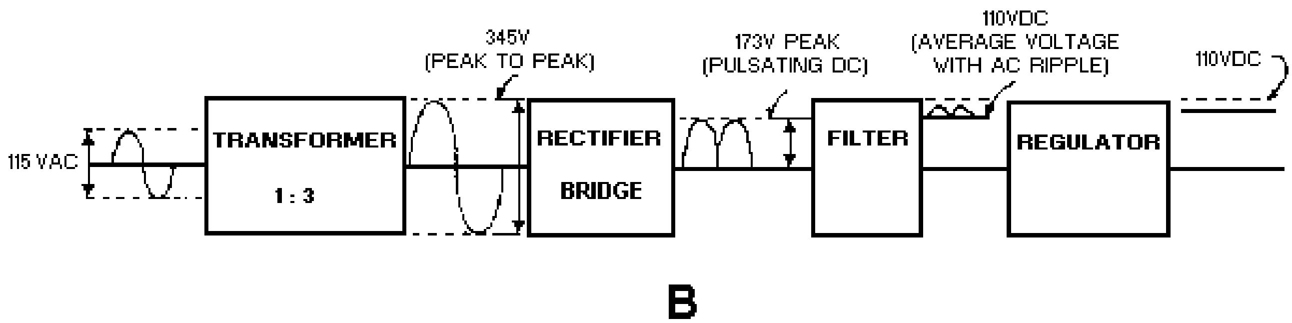


Figure 4-1B.—Block diagram of a basic power supply.

The final section, the REGULATOR, does just what the name implies. It maintains the output of the power supply at a constant level in spite of large changes in load current or input line voltages.

Now that you know what each section does, let's trace an ac signal through the power supply. At this point you need to see how this signal is altered within each section of the power supply. Later on in the chapter you will see how these changes take place. In view B of figure 4-1, an input signal of 115 volts ac is applied to the primary of the transformer. The transformer is a step-up transformer with a turns ratio of 1:3. You can calculate the output for this transformer by multiplying the input voltage by the ratio of turns in the primary to the ratio of turns in the secondary; therefore, 115 volts ac $\times$ 3 = 345 volts ac (peak-to-peak) at the output. Because each diode in the rectifier section conducts for 180 degrees of the 360-degree input, the output of the rectifier will be one-half, or approximately 173 volts of pulsating dc. The filter section, a network of resistors, capacitors, or inductors, controls the rise and fall time of the varying signal; consequently, the signal remains at a more constant dc level. You will see the filter process more clearly in the discussion of the actual filter circuits. The output of the filter is a signal of 110 volts dc, with ac ripple riding on the dc. The reason for the lower voltage (average voltage) will be explained later in this chapter. The regulator maintains its output at a constant 110-volt dc level, which is used by the electronic equipment (more commonly called the load).

Q1. What are the four basic sections of a power supply?

Q2. What is the purpose of the rectifier section?

Q3. What is the purpose of the filter section?

Q4. What is the purpose of the regulator section?

THE POWER TRANSFORMER

In some cases a power supply may not use a transformer; therefore, the power supply would be connected directly to the source line voltage. This type of connection is used primarily because it is economical. However, unless the power supply is completely insulated, it presents a dangerous shock hazard to anyone who comes in contact with it. When a transformer is not being used, the return side of the ac line is connected to the metal chassis. To remove this potential shock hazard and to have the option of stepping up or stepping down the input voltage to the rectifier, a transformer must be used.

View A of figure 4-2 shows the schematic diagram for a STEP-UP transformer; view B shows a STEP-DOWN transformer; and, view C shows a STEP-UP, CENTER-TAPPED transformer. The step-up and step-down transformers were discussed in earlier *NEETS* modules, so only the center-tapped transformer will be mentioned in this chapter. The primary purpose of the center-tapped transformer is to provide two equal voltages to the conventional full-wave rectifier.

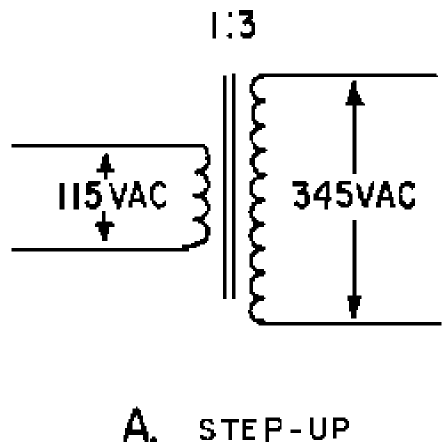


Figure 4-2A.—Common types of transformers. STEP-UP

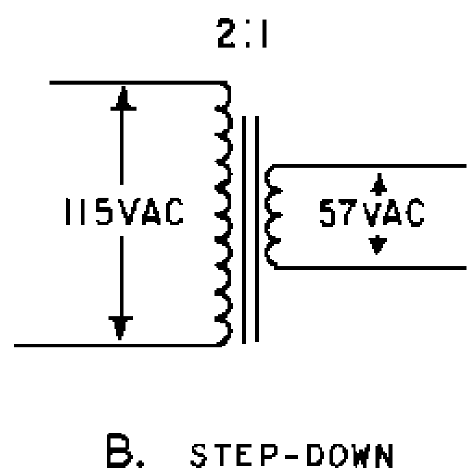


Figure 4-2B.—Common types of transformers. STEP-DOWN

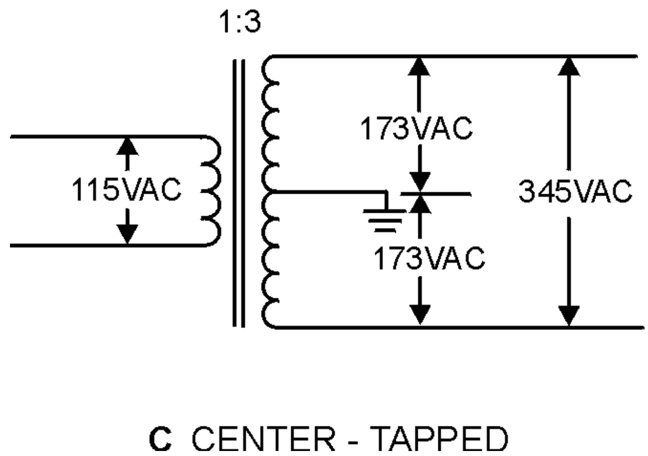


Figure 4-2C.—Common types of transformers. CENTER-TAPPED

THE RECTIFIER

From previous discussions, you should know that rectification is the conversion of an alternating current to a pulsating direct current. Now let's see how the process of RECTIFICATION occurs in both a half-wave and a full-wave rectifier.

The Half-Wave Rectifier

Since a silicon diode will pass current in only one direction, it is ideally suited for converting alternating current (ac) to direct current (dc). When ac voltage is applied to a diode, the diode conducts ONLY ON THE POSITIVE ALTERNATION OF VOLTAGE; that is, when the anode of the diode is positive with respect to the cathode. This simplest type of rectifier is the half-wave rectifier. As shown in view A of figure 4-3, the half-wave rectifier uses only one diode. During the positive alternation of input voltage, the sine wave applied to the diode makes the anode positive with respect to the cathode. The diode then conducts, and current (I) flows from the negative supply lead (the secondary of the transformer), through the milliammeter, through the diode, and to the positive supply lead. As indicated by the shaded area of the output waveform in view B, this current exists during the entire period of time that the anode is positive with respect to the cathode (in other words, for the first 180 degrees of the input sine wave).

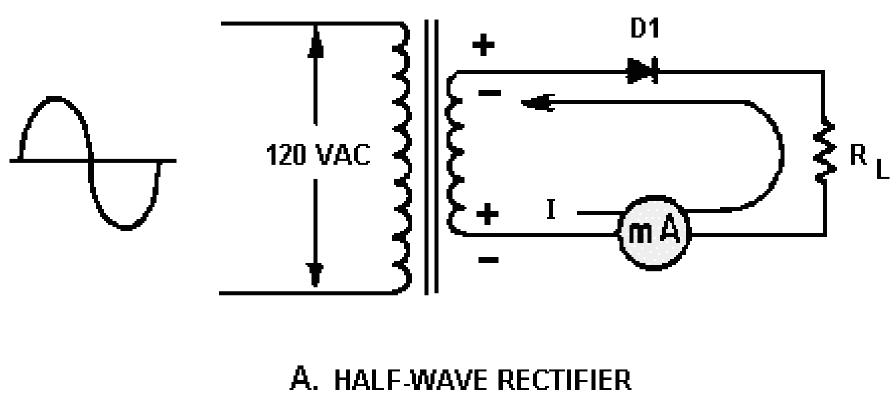


Figure 4-3A.—Simple half-wave rectifier. HALF-WAVE RECTIFIER

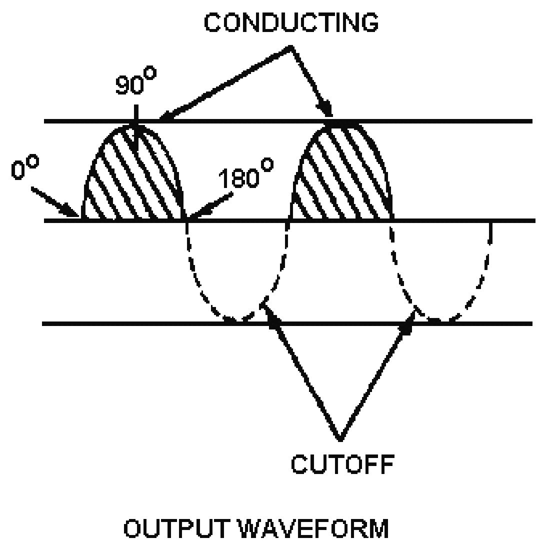


Figure 4-3B.—Simple half-wave rectifier. OUTPUT WAVEFORM

During the negative alternation of input voltage (dotted polarity signs), the anode is driven negative and the diode cannot conduct. When conditions such as these exist, the diode is in cutoff and remains in cutoff for 180 degrees, during which time no current flows in the circuit. The circuit current therefore has the appearance of a series of positive pulses, as illustrated by the shaded areas on the waveform in view B. Notice that although the current is in the form of pulses, the current always flows in the same direction. Current that flows in pulses in the same direction is called PULSATING DC. The diode has thus RECTIFIED the ac input voltage.

Rms, Peak, and Average Values

View A of figure 4-4 is a comparison of the rms, peak, and average values of the types of waveforms associated with the half-wave rectifier. Ac voltages are normally specified in terms of their rms values. Thus, when a 115-volt ac power source is mentioned in this chapter, it is specifying the rms value of 115 volts ac. In terms of peak values,

$E_{rms} = E_{peak} \times .707$

The peak value is always higher than the rms value. In fact,

$E_{peak} = E_{rms} \times 1.414$

therefore, if the rms value is 115 volts ac, then the peak value must be:

$E_{peak} = E_{rms} \times 1.414$

$E_{peak} = 115 \text{ volts ac} \times 1.414$

$E_{peak} = 162.6 \text{ volts}$

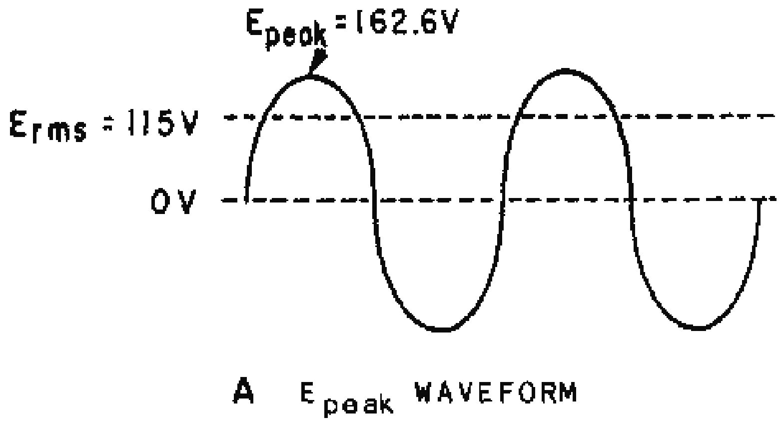


Figure 4-4A.—Comparison of E_{peak} to E_{avg} in a half-wave rectifier. E_{peak} WAVEFORM.

The average value of a sine wave is 0 volts. View B of figure 4-4 shows how the average voltage changes when the negative portion of the sine wave is clipped off. Since the wave form swings positive but never negative (past the "zero-volt" reference line), the average voltage is positive. The average voltage (E_{avg}) is determined by the equation:

Where: $E_{avg} = E_{peak} \times .318$

Thus: $E_{avg} = 162.6 \times .318$

$E_{avg} = 51.7$ volts

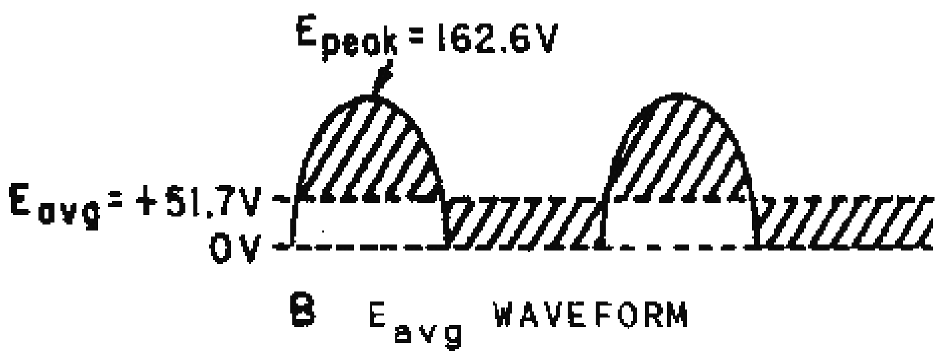


Figure 4-4B.—Comparison of E_{peak} to E_{avg} in a half-wave rectifier. E_{peak} WAVEFORM

Ripple Frequency

The half-wave rectifier gets its name from the fact that it conducts during only half the input cycle. Its output is a series of pulses with a frequency that is the same as the input frequency. Thus when operated from a 60-hertz line, the frequency of the pulses is 60 hertz. This is called RIPPLE FREQUENCY.

Q5. What is the name of the simplest type of rectifier which uses one diode?

Q6. If the output of a half-wave rectifier is 50-volts peak, what is the average voltage?

Q7. In addition to stepping up or stepping down the input line voltage, what additional purpose does the transformer serve?

The Conventional Full-Wave Rectifier

A full-wave rectifier is a device that has two or more diodes arranged so that load current flows in the same direction during each half cycle of the ac supply.

A diagram of a simple full-wave rectifier is shown in figure 4-5. The transformer supplies the source voltage for two diode rectifiers, D1 and D2. This power transformer has a center-tapped, high-voltage secondary winding that is divided into two equal parts (W1 and W2). W1 provides the source voltage for D1, and W2 provides the source voltage for D2. The connections to the diodes are arranged so that the diodes conduct on alternate half cycles.

During one alternation of the secondary voltage, the polarities are as shown in view A. The source for D2 is the voltage induced into the lower half of the secondary winding of the transformer (W2). At the specific instant of time shown in the figure, the anode voltage on D2 is negative, and D2 cannot conduct. Throughout the period of time during which the anode of D2 is negative, the anode of D1 is positive. Since the anode of D1 is positive, it conducts, causing current to flow through the load resistor in the direction shown by the arrow.

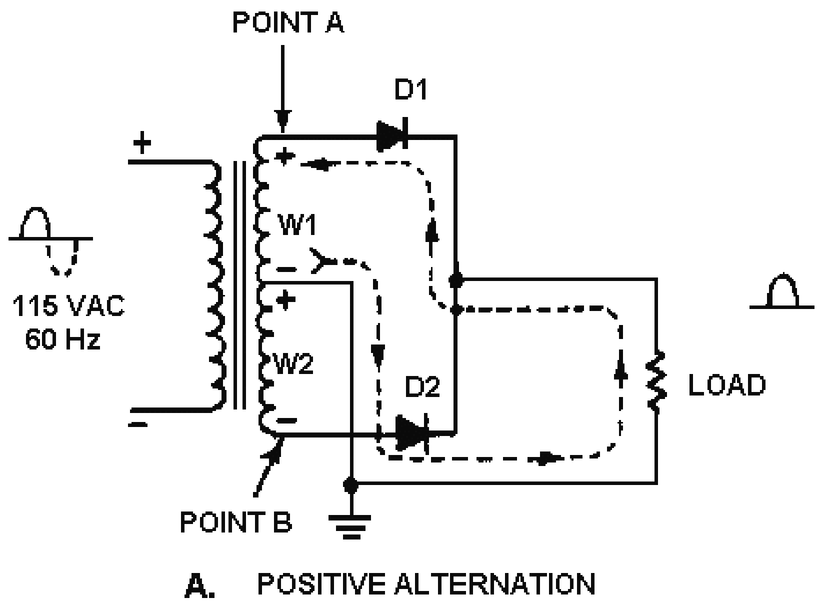


Figure 4-5A.—Full-wave rectifier. POSITIVE ALTERNATION

View B shows the next half cycle of secondary voltage. Now the polarities across W1 and W2 are reversed. During this alternation, the anode of D1 is driven negative and D1 cannot conduct. For the period of time that the anode of D1 is negative, the anode of D2 is positive, permitting D2 to conduct. Notice that the anode current of D2 passes through the load resistor in the same direction as the current of D1 did. In this circuit arrangement, a pulse of load current flows during each alternation of the input cycle. Since both alternations of the input voltage cycle are used, the circuit is called a FULL-WAVE RECTIFIER.

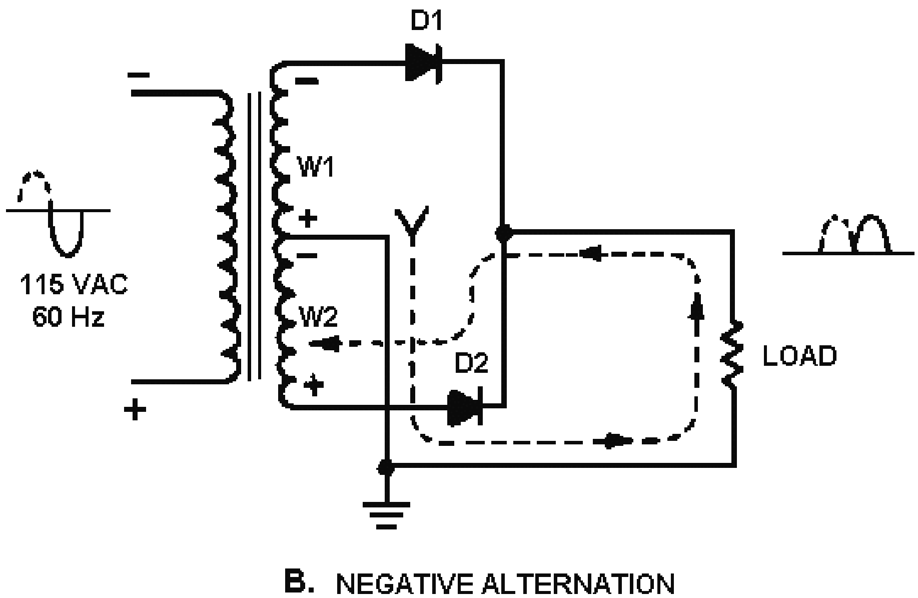


Figure 4-5B.—Full-wave rectifier. NEGATIVE ALTERNATION

Now that you have a basic understanding of how a full-wave rectifier works, let's cover in detail a practical full-wave rectifier and its waveforms.

A Practical Full-Wave Rectifier

A practical full-wave rectifier circuit is shown in view A of figure 4-6. It uses two diodes (D1 and D2) and a center-tapped transformer (T1). When the center tap is grounded, the voltages at the opposite ends of the secondary windings are 180 degrees out of phase with each other. Thus, when the voltage at point A is positive with respect to ground, the voltage at point B is negative with respect to ground. Let's examine the operation of the circuit during one complete cycle.

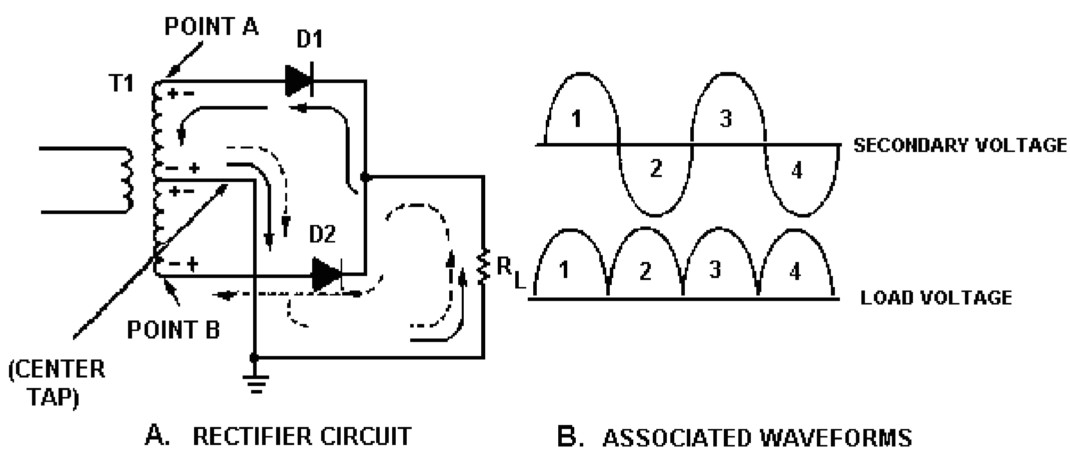


Figure 4-6.—Practical full-wave rectifier.

During the first half cycle (indicated by the solid arrows), the anode of D1 is positive with respect to ground and the anode of D2 is negative. As shown, current flows from ground (center tap), up through the load resistor (R_L), through diode D1 to point A. In the transformer, current flows from point A, through the upper winding, and back to ground (center tap). When D1 conducts, it acts like a closed switch so that the positive half cycle is felt across the load (R_L).

During the second half cycle (indicated by the dotted lines), the polarity of the applied voltage has reversed. Now the anode of D2 is positive with respect to ground and the anode of D1 is negative. Now only D2 can conduct. Current now flows, as shown, from ground (center tap), up through the load resistor (R_L), through diode D2 to point B of T1. In the transformer, current flows from point B up through the lower windings and back to ground (center tap). Notice that the current flows across the load resistor (R_L) in the same direction for both halves of the input cycle.

View B represents the output waveform from the full-wave rectifier. The waveform consists of two pulses of current (or voltage) for each cycle of input voltage. The ripple frequency at the output of the full-wave rectifier is therefore twice the line frequency.

The higher frequency at the output of a full-wave rectifier offers a distinct advantage: Because of the higher ripple frequency, the output is closely approximate to pure dc. The higher frequency also makes filtering much easier than it is for the output of the half-wave rectifier.

In terms of peak value, the average value of current and voltage at the output of the full-wave rectifier is twice as great as that at the output of the half-wave rectifier. The relationship between the peak value and the average value is illustrated in figure 4-7. Since the output waveform is essentially a sine wave with both alternations at the same polarity, the average current or voltage is 63.7 percent (or 0.637) of the peak current or voltage.

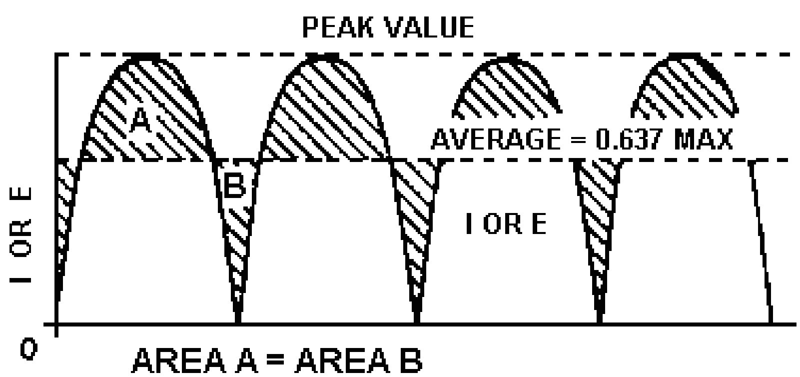


Figure 4-7.—Peak and average values for a full-wave rectifier.

As an equation:

Where:

E_{max} = The peak value of the load voltage pulse

$E_{avg} = 0.637 \times E_{max}$ (the average load voltage)

I_{max} = The peak value of the load current pulse

$I_{avg} = 0.637 \times I_{max}$ (the average load current)

Example: The total voltage across the high-voltage secondary of a transformer used to supply a full-wave rectifier is 300 volts. Find the average load voltage (ignore the drop across the diode).

Solution: Since the total secondary voltage (E_S) is 300 volts, each diode is supplied one-half of this value, or 150 volts. Because the secondary voltage is an rms value, the peak load voltage is:

$$E_{max} = 1.414 \times E_S$$

$$E_{max} = 1.414 \times 150$$

$$E_{max} = 212 \text{ volts}$$

The average load voltage is:

$$E_{avg} = 0.637 \times E_{max}$$

$$E_{avg} = 0.637 \times 212$$

$$E_{avg} = 135 \text{ volts}$$

NOTE: If you have problems with this equation, review the portion of *NEETS*, module 2, that pertain to this subject.

As you may recall from your past studies in electricity, every circuit has advantages and disadvantages. The full-wave rectifier is no exception. In studying the full-wave rectifier, you may have found that by doubling the output frequency, the average voltage has doubled, and the resulting signal is much easier to filter because of the high ripple frequency. The only disadvantage is that the peak voltage in the full-wave rectifier is only half the peak voltage in the half-wave rectifier. This is because the secondary of the power transformer in the full-wave rectifier is center tapped; therefore, only half the source voltage goes to each diode.

Fortunately, there is a rectifier which produces the same peak voltage as a half-wave rectifier and the same ripple frequency as a full-wave rectifier. This circuit, known as the BRIDGE RECTIFIER, will be the subject of our next discussion.

Q8. What was the major factor that led to the development of the full-wave rectifier?

Q9. What is the ripple frequency of a full-wave rectifier with an input frequency of 60 Hz?

Q10. What is the average voltage (E_{avg}) Output of a full-wave rectifier with an output of 100 volts peak?

The Bridge Rectifier

When four diodes are connected as shown in figure 4-8, the circuit is called a BRIDGE RECTIFIER. The input to the circuit is applied to the diagonally opposite corners of the network, and the output is taken from the remaining two corners.

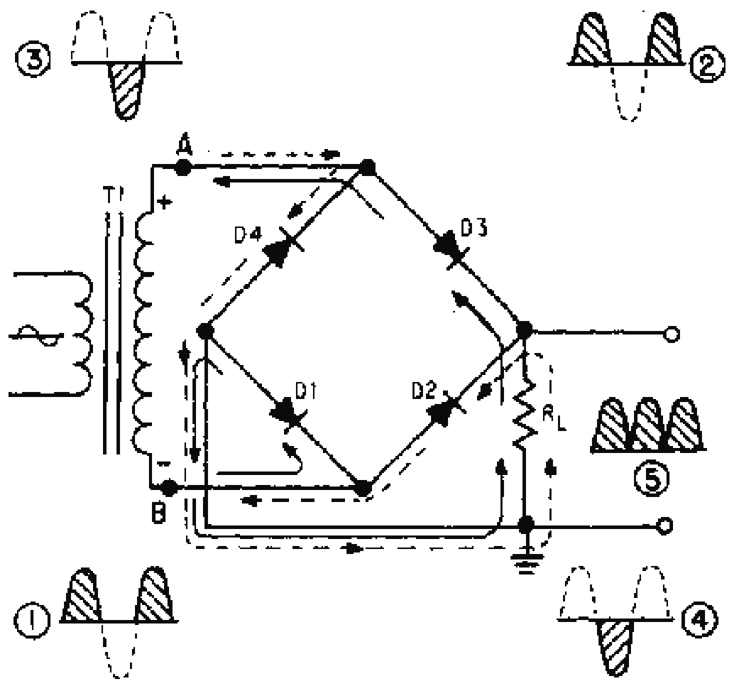


Figure 4-8.—Bridge rectifier.

One complete cycle of operation will be discussed to help you understand how this circuit works. We have discussed transformers in previous modules in the *NEETS* series and will not go into their characteristics at this time. Let us assume the transformer is working properly and there is a positive potential at point A and a negative potential at point B. The positive potential at point A will forward bias D3 and reverse bias D4. The negative potential at point B will forward bias D1 and reverse bias D2. At this time D3 and D1 are forward biased and will allow current flow to pass through them; D4 and D2 are reverse biased and will block current flow. The path for current flow is from point B through D1, up through R_L, through D3, through the secondary of the transformer back to point B. This path is indicated by the solid arrows. Waveforms (1) and (2) can be observed across D1 and D3.

One-half cycle later the polarity across the secondary of the transformer reverses, forward biasing D2 and D4 and reverse biasing D1 and D3. Current flow will now be from point A through D4, up through R_L, through D2, through the secondary of T1, and back to point A. This path is indicated by the broken arrows. Waveforms (3) and (4) can be observed across D2 and D4. You should have noted that the current flow through R_L is always in the same direction. In flowing through R_L this current develops a voltage corresponding to that shown in waveform (5). Since current flows through the load (R_L) during both half cycles of the applied voltage, this bridge rectifier is a full-wave rectifier.

One advantage of a bridge rectifier over a conventional full-wave rectifier is that with a given transformer the bridge rectifier produces a voltage output that is nearly twice that of the conventional full-wave circuit. This may be shown by assigning values to some of the components shown in views A and B of figure 4-9. Assume that the same transformer is used in both circuits. The peak voltage developed between points X and Y is 1000 volts in both circuits. In the conventional full-wave circuit shown in view A, the peak voltage from the center tap to either X or Y is 500 volts. Since only one diode can conduct at any instant, the maximum voltage that can be rectified at any instant is 500 volts. Therefore, the maximum voltage that appears across the load resistor is nearly — but never exceeds — 500 volts, as a result of the small voltage drop across the diode. In the bridge rectifier shown in view B, the maximum voltage that can be rectified is the full secondary voltage, which is 1000 volts. Therefore, the peak output voltage across the load resistor is nearly 1000 volts. With both circuits using the same transformer, the bridge rectifier circuit produces a higher output voltage than the conventional full-wave rectifier circuit.

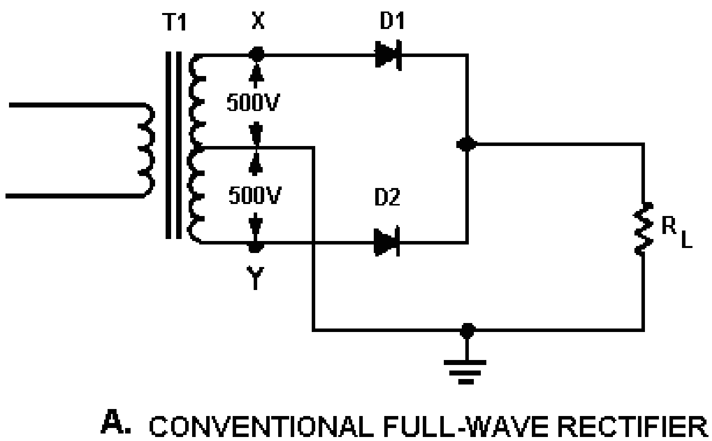


Figure 4-9A.—Comparison of a conventional and bridge full-wave rectifier. CONVENTIONAL FULL-WAVE RECTIFIER

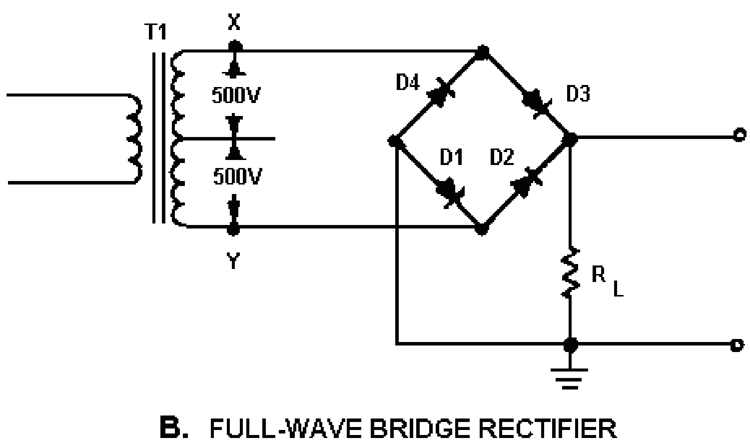


Figure 4-9B.—Comparison of a conventional and bridge full-wave rectifier. FULL-WAVE BRIDGE RECTIFIER

Q11. What is the main disadvantage of a conventional full-wave rectifier?

Q12. What main advantage does a bridge rectifier have over a conventional full-wave rectifier?

FILTERS

While the output of a rectifier is a pulsating dc, most electronic circuits require a substantially pure dc for proper operation. This type of output is provided by single or multisection filter circuits placed between the output of the rectifier and the load.

There are four basic types of filter circuits:

- Simple capacitor filter
- LC choke-input filter

- LC capacitor-input filter (pi-type)
- RC capacitor-input filter (pi-type)

The function of each of these filters will be covered in detail in this chapter.

Filtering is accomplished by the use of capacitors, inductors, and/or resistors in various combinations. Inductors are used as series impedances to oppose the flow of alternating (pulsating dc) current. Capacitors are used as shunt elements to bypass the alternating components of the signal around the load (to ground). Resistors are used in place of inductors in low current applications.

Let's briefly review the properties of a capacitor. First, a capacitor opposes any change in voltage. The opposition to a change in current is called capacitive reactance (X_C) and is measured in ohms. The capacitive reactance is determined by the frequency (f) of the applied voltage and the capacitance (C) of the capacitor.

$$X_C = \frac{1}{2\pi fC} \text{ or } \frac{.159}{fC}$$

From the formula, you can see that if frequency or capacitance is increased, the X_C decreases. Since filter capacitors are placed in parallel with the load, a low X_C will provide better filtering than a high X_C. For this to be accomplished, a better shunting effect of the ac around the load is provided, as shown in figure 4-10.

To obtain a steady dc output, the capacitor must charge almost instantaneously to the value of applied voltage. Once charged, the capacitor must retain the charge as long as possible. The capacitor must have a short charge time constant (view A). This can be accomplished by keeping the internal resistance of the power supply as small as possible (fast charge time) and the resistance of the load as large as possible (for a slow discharge time as illustrated in view B).

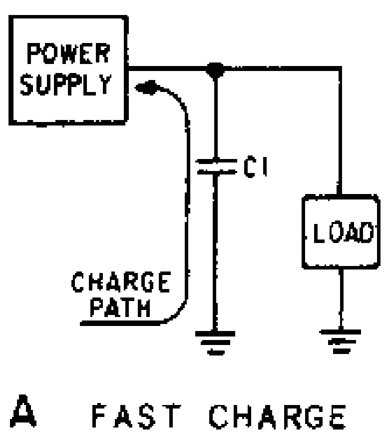


Figure 4-10A.—Capacitor filter. FAST CHARGE

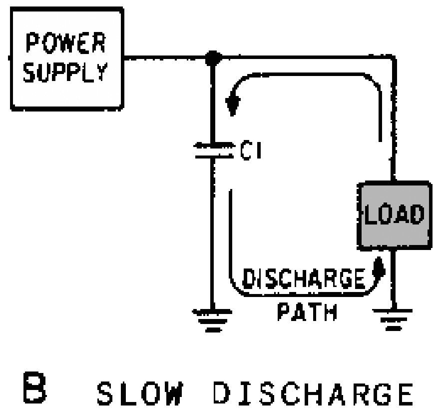


Figure 4-10B.—Capacitor filter. SLOW DISCHARGE

From your earlier studies in basic electricity, you may remember that one time constant is defined as the time it takes a capacitor to charge to 63.2 percent of the applied voltage or to discharge to 36.8 percent of its total charge. This action can be expressed by the following equation:

$$t = RC$$

Where: R represents the resistance of the charge or discharge path

And: C represents the capacitance of the capacitor.

You should also recall that a capacitor is considered fully charged after five RC time constants. Refer to figure 4-11. You can see that a steady dc output voltage is obtained when the capacitor charges rapidly and discharges as slowly as possible.

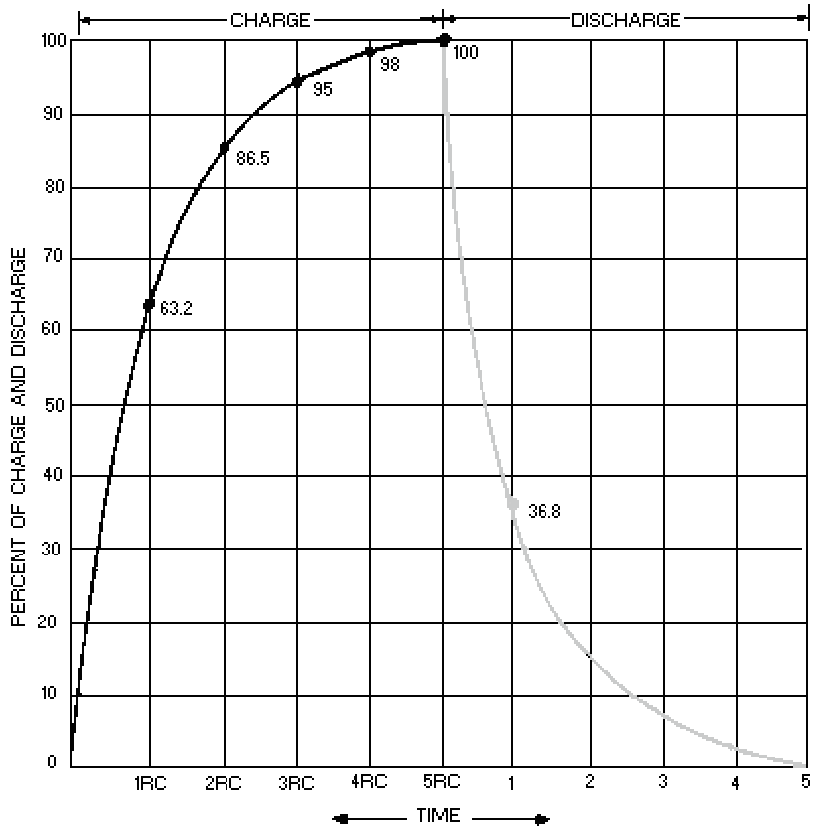


Figure 4-11.—RC time constant.

In filter circuits the capacitor is the common element to both the charge and the discharge paths. Therefore, to obtain the longest possible discharge time, you want the capacitor to be as large as possible. Another way to look at it is: The capacitor acts as a short circuit around the load (as far as the ac component is concerned), and since

$$X_C = \frac{1}{2\pi fC}$$

the larger the value of the capacitor (C), the smaller the opposition (X_C) or reactance to ac.

Now let's look at inductors and their application in filter circuits. Remember, AN INDUCTOR OPPOSES ANY CHANGE IN CURRENT. In case you have forgotten, a change in current through an inductor produces a changing electromagnetic field. The changing field, in turn, cuts the windings of the wire in the inductor and thereby produces a counter electromotive force (CEMF). It is the CEMF that opposes the change in circuit current. Opposition to a change in current at a given frequency is called inductive reactance (X_L) and is measured in ohms. The inductive reactance (X_L) of an inductor is determined by the applied frequency and the inductance of the inductor.

Mathematically,

$$X_L = 2\pi fL$$

If frequency or inductance is increased, the X_L increases. Since inductors are placed in series with the load (as shown in figure 4-12), the larger the X_L, the larger the ac voltage developed across the load.

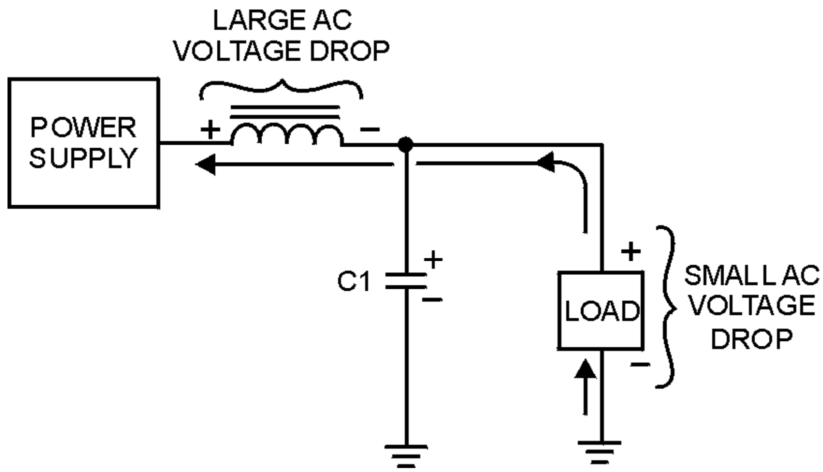


Figure 4-12.—Voltage drops in an inductive filter.

Now refer to figure 4-13. When the current starts to flow through the coil, an expanding magnetic field builds up around the inductor. This magnetic field around the coil develops the CEMF that opposes the change in current. When the rectifier current decreases, as shown in figure 4-14, the magnetic field collapses and again cuts the turns (windings) of wire, thus inducing current into the coil. This additional current merges with the rectifier current and attempts to keep it at its original level.

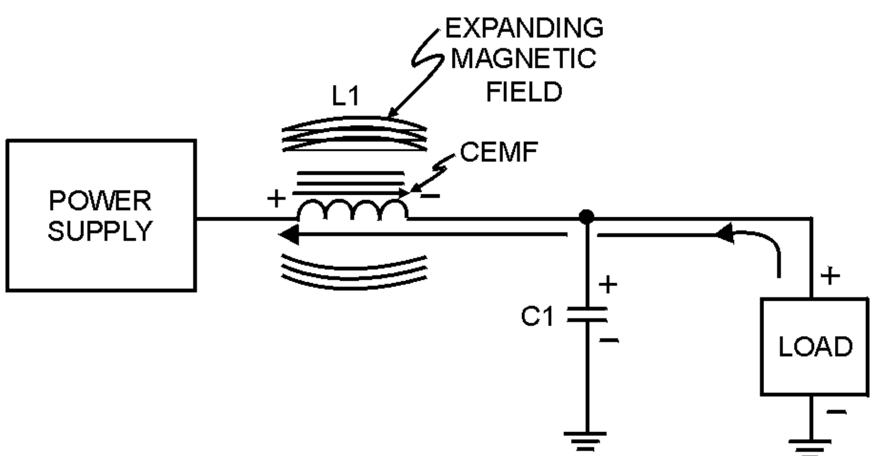


Figure 4-13.—Inductive filter (expanding field).

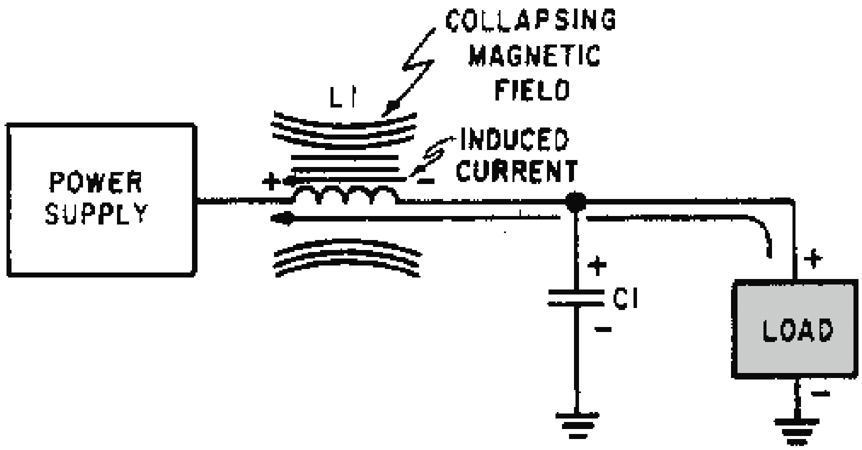


Figure 4-14.—Inductive filter (collapsing field).

Now that you have read how the components in a filter circuit react to current flow from the rectifier, the different types of filter circuits in use today will be discussed.

Q13. If you increase the value of the capacitor, will the X_C increase or decrease? Why?

The Capacitor Filter

The simple capacitor filter is the most basic type of power supply filter. The application of the simple capacitor filter is very limited. It is sometimes used on extremely high-voltage, low-current power supplies for cathode-ray and similar electron tubes, which require very little load current from the supply. The capacitor filter is also used where the power-supply ripple frequency is not critical; this frequency can be relatively high. The capacitor (C1) shown in figure 4-15 is a simple filter connected across the output of the rectifier in parallel with the load.

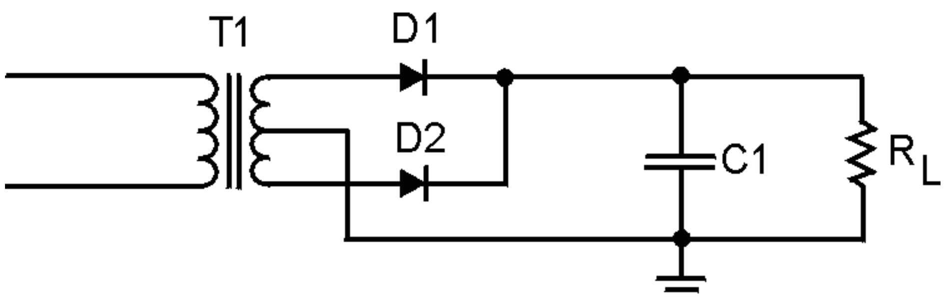


Figure 4-15.—Full-wave rectifier with a capacitor filter.

When this filter is used, the RC charge time of the filter capacitor (C1) must be short and the RC discharge time must be long to eliminate ripple action. In other words, the capacitor must charge up fast, preferably with no discharge at all. Better filtering also results when the input frequency is high; therefore, the full-wave rectifier output is easier to filter than that of the half-wave rectifier because of its higher frequency.

For you to have a better understanding of the effect that filtering has on E_{avg}, a comparison of a rectifier circuit with a filter and one without a filter is illustrated in views A and B of figure 4-16. The output waveforms in figure 4-16 represent the unfiltered and filtered outputs of the half-wave rectifier circuit. Current pulses flow through the load resistance (R_L) each time a diode conducts. The dashed line indicates the average value of output voltage. For the half-wave rectifier, E_{avg} is less than half (or approximately 0.318) of the peak output voltage. This value is still much less than that of the applied voltage. With no capacitor connected across the output of the rectifier circuit, the waveform in view A has a large pulsating component (ripple) compared with the average or dc component. When a capacitor is connected across the output (view B), the average value of output voltage (E_{avg}) is increased due to the filtering action of capacitor C1.

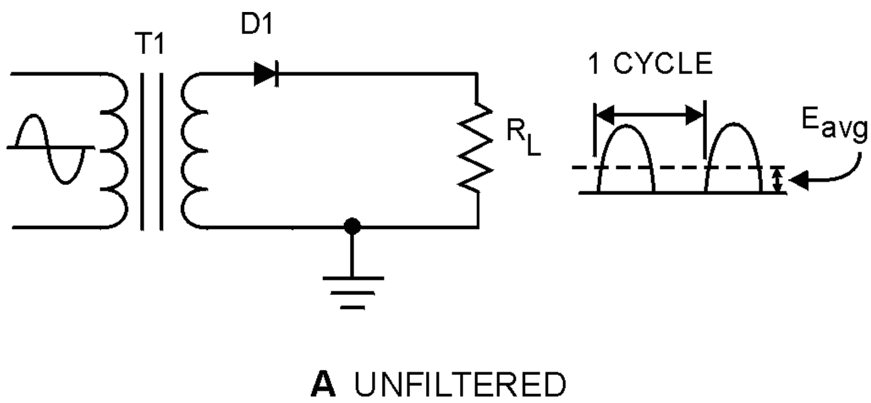


Figure 4-16A.—Half-wave rectifier with and without filtering. UNFILTERED

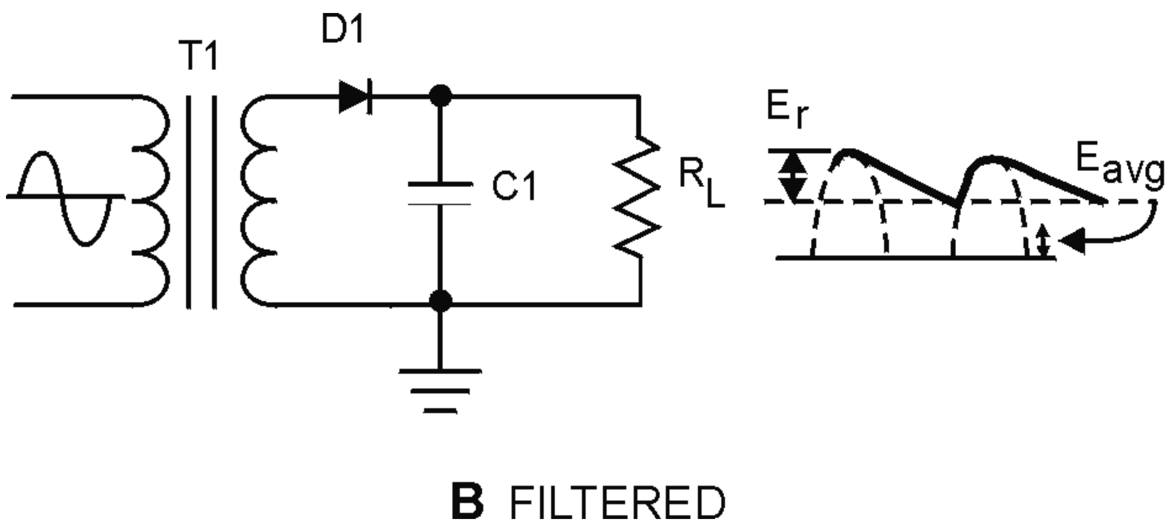


Figure 4-16B.—Half-wave rectifier with and without filtering. FILTERED

The value of the capacitor is fairly large (several microfarads), thus it presents a relatively low reactance to the pulsating current and it stores a substantial charge.

The rate of charge for the capacitor is limited only by the resistance of the conducting diode which is relatively low. Therefore, the RC charge time of the circuit is relatively short. As a result, when the pulsating voltage is first applied to the circuit, the capacitor charges rapidly and almost reaches the peak value of the rectified voltage within the first few cycles. The capacitor attempts to charge to the peak value of the rectified voltage anytime a diode is conducting, and tends to retain its charge when the rectifier output falls to zero. (The capacitor cannot discharge immediately.) The capacitor slowly discharges through the load resistance (R_L) during the time the rectifier is nonconducting.

The rate of discharge of the capacitor is determined by the value of capacitance and the value of the load resistance. If the capacitance and load-resistance values are large, the RC discharge time for the circuit is relatively long.

A comparison of the waveforms shown in figure 4-16 (view A and view B) illustrates that the addition of C1 to the circuit results in an increase in the average of the output voltage (E_{avg}) and a reduction in the amplitude of the ripple component (E_r) which is normally present across the load resistance.

Now, let's consider a complete cycle of operation using a half-wave rectifier, a capacitive filter (C1), and a load resistor (R_L). As shown in view A of figure 4-17, the capacitive filter (C1) is assumed to be large enough to ensure a small reactance to the pulsating rectified current. The resistance of R_L is assumed to be much greater than the reactance of C1 at the input frequency. When the circuit is energized, the diode conducts on the positive half cycle and current flows through the circuit, allowing C1 to charge. C1 will charge to approximately the peak value of the input voltage. (The charge is less than the peak value because of the voltage drop across the diode (D1)). In view A of the figure, the charge on C1 is indicated by the heavy solid line on the waveform. As illustrated in view B, the diode cannot conduct on the negative half cycle because the anode of D1 is negative with respect to the cathode. During this interval, C1 discharges through the load resistor (R_L). The discharge of C1 produces the downward slope as indicated by the solid line on the waveform in view B. In contrast to the abrupt fall of the applied ac voltage from peak value to zero, the voltage across C1 (and thus across R_L) during the discharge period gradually decreases until the time of the next half cycle of rectifier operation. Keep in mind that for good filtering, the filter capacitor should charge up as fast as possible and discharge as little as possible.

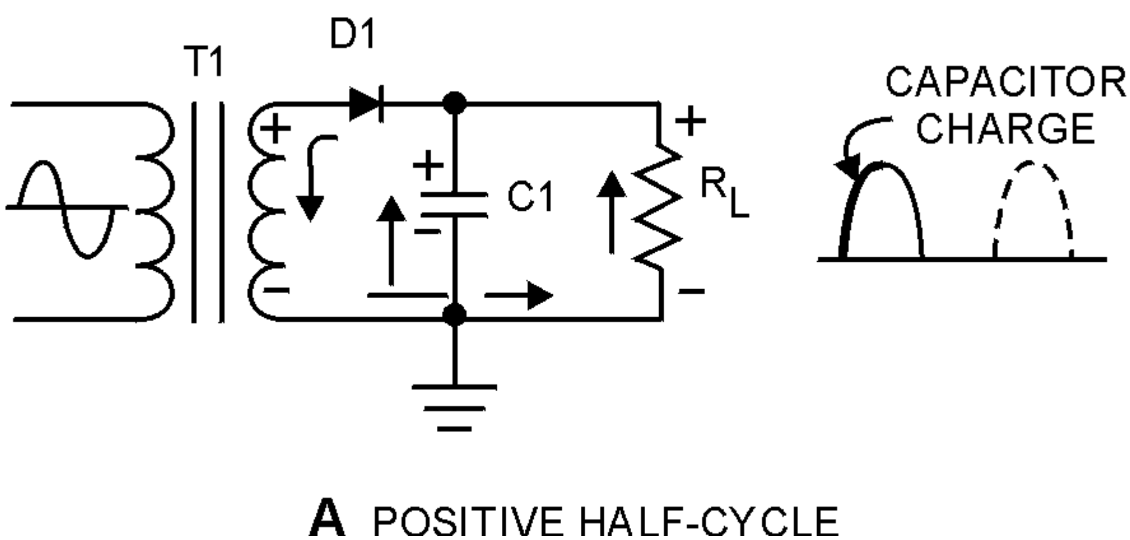


Figure 4-17A.—Capacitor filter circuit (positive and negative half cycles). POSITIVE HALF-CYCLE

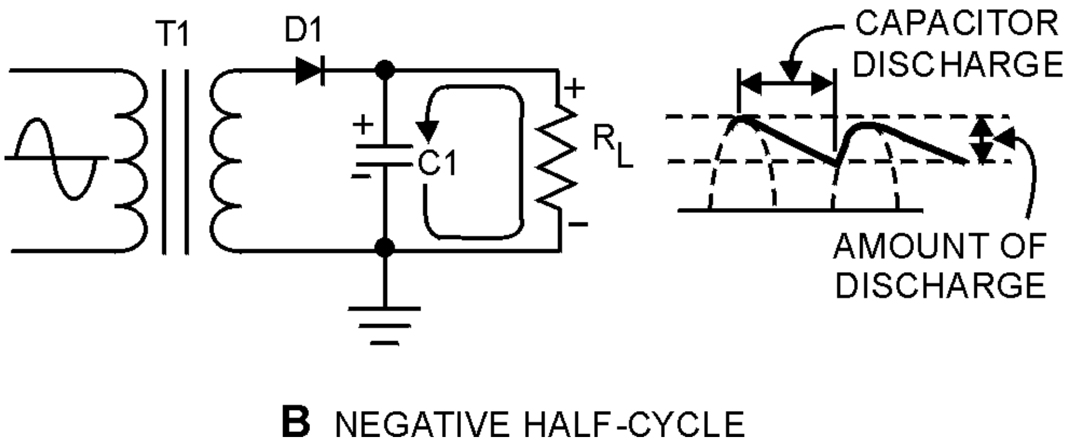


Figure 4-17B.—Capacitor filter circuit (positive and negative half cycles). NEGATIVE HALF-CYCLE

Since practical values of C1 and R_L ensure a more or less gradual decrease of the discharge voltage, a substantial charge remains on the capacitor at the time of the next half cycle of operation. As a result, no current can flow through the diode until the rising ac input voltage at the anode of the diode exceeds the voltage on the charge remaining on C1. The charge on C1 is the cathode potential of the diode. When the potential on the anode exceeds the potential on the cathode (the charge on C1), the diode again conducts, and C1 begins to charge to approximately the peak value of the applied voltage.

After the capacitor has charged to its peak value, the diode will cut off and the capacitor will start to discharge. Since the fall of the ac input voltage on the anode is considerably more rapid than the decrease on the capacitor voltage, the cathode quickly become more positive than the anode, and the diode ceases to conduct.

Operation of the simple capacitor filter using a full-wave rectifier is basically the same as that discussed for the half-wave rectifier. Referring to figure 4-18, you should notice that because one of the diodes is always conducting on. either alternation, the filter capacitor charges and discharges during each half cycle. (Note that each diode conducts only for that portion of time when the peak secondary voltage is greater than the charge across the capacitor.)

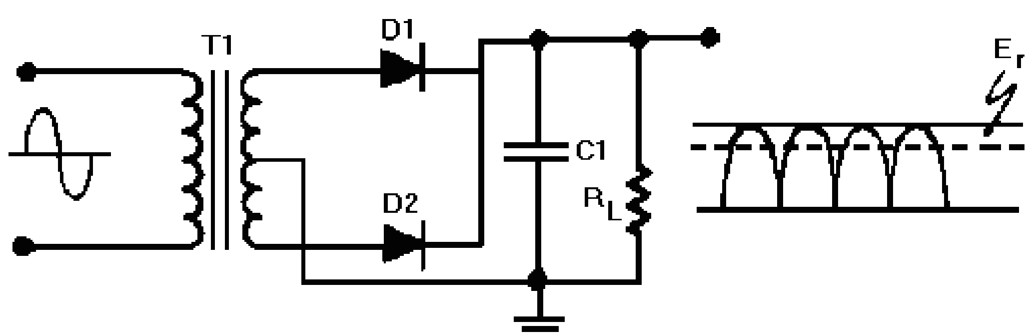


Figure 4-18.—Full-wave rectifier (with capacitor filter).

Another thing to keep in mind is that the ripple component (E) of the output voltage is an ac voltage and the average output voltage (E_{avg}) is the dc component of the output. Since the filter capacitor offers a relatively low impedance to ac, the majority of the ac component flows through the filter capacitor. The ac component is therefore bypassed (shunted) around the load resistance, and the entire dc component (or

E_{avg}) flows through the load resistance. This statement can be clarified by using the formula for X_C in a half-wave and full-wave rectifier. First, you must establish some values for the circuit.

HALFWAVE RECTIFIER

FREQUENCY AT
RECTIFIER OUTPUT: 60 Hz

VALUE OF FILTER
CAPACITOR: 30μF

LOAD RESISTANCE: 10kΩ

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{.159}{fC}$$

$$X_C = \frac{.159}{60 \times .000030}$$

$$X_C = \frac{.159}{.0018}$$

$$X_C = 88.3\Omega$$

FREQUENCY AT
RECTIFIER OUTPUT: 120Hz

VALUE OF FILTER
CAPACITOR: 30µF

LOAD RESISTANCE: 10kΩ

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{.159}{fC}$$

$$X_C = \frac{.159}{120 \times .000030}$$

$$X_C = \frac{.159}{.0036}$$

$$X_C = 44.16\Omega$$

As you can see from the calculations, by doubling the frequency of the rectifier, you reduce the impedance of the capacitor by one-half. This allows the ac component to pass through the capacitor more easily. As a result, a full-wave rectifier output is much easier to filter than that of a half-wave rectifier. Remember, the smaller the X_C of the filter capacitor with respect to the load resistance, the better the filtering action. Since

$$X_C = \frac{1}{2\pi fC}$$

the largest possible capacitor will provide the best filtering. Remember, also, that the load resistance is an important consideration. If load resistance is made small, the load current increases, and the average value of output voltage (E_{avg}) decreases. The RC discharge time constant is a direct function of the value of the load resistance; therefore, the rate of capacitor voltage discharge is a direct function of the current through the load. The greater the load current, the more rapid the discharge of the capacitor, and the lower the average value of output voltage. For this reason, the simple capacitive filter is seldom used with rectifier circuits that must supply a relatively large load current. Using the simple capacitive filter in conjunction with a full-wave or bridge rectifier provides improved filtering because the increased ripple frequency decreases the capacitive reactance of the filter capacitor.

Q14. What is the most basic type of filter?

Q15. In a capacitor filter, is the capacitor in series or in parallel with the load?

Q16. Is filtering better at a high frequency or at a low frequency?

Q17. Does a filter circuit increase or decrease the average output voltage?

Q18. What determines the rate of discharge of the capacitor in a filter circuit?

Q19. Does low ripple voltage indicate good or bad filtering?

Q20. Is a full-wave rectifier output easier to filter than that of a half-wave rectifier?

LC Choke-Input Filter

The LC choke-input filter is used primarily in power supplies where voltage regulation is important and where the output current is relatively high and subject to varying load conditions. This filter is used in high power applications such as those found in radars and communication transmitters.

Notice in figure 4-19 that this filter consists of an input inductor (L1), or filter choke, and an output filter capacitor (C1). Inductor L1 is placed at the input to the filter and is in series with the output of the rectifier circuit. Since the action of an inductor is to oppose any change in current flow, the inductor tends to keep a constant current flowing to the load throughout the complete cycle of the applied voltage. As a result, the output voltage never reaches the peak value of the applied voltage. Instead, the output voltage approximates the average value of the rectified input to the filter, as shown in the figure. The reactance of the inductor (X_L) reduces the amplitude of ripple voltage without reducing the dc output voltage by an appreciable amount. (The dc resistance of the inductor is just a few ohms.)

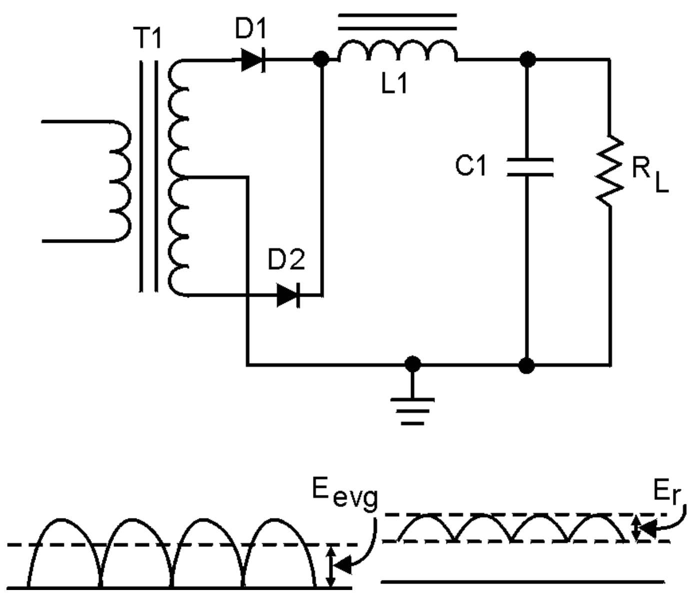


Figure 4-19.—LC choke-input filter.

The shunt capacitor (C1) charges and discharges at the ripple frequency rate, but the amplitude of the ripple voltage (E_r) is relatively small because the inductor (L1) tends to keep a constant current flowing from the rectifier circuit to the load. In addition, the reactance of the shunt capacitor (X_C) presents a low impedance to the ripple component existing at the output of the filter, and thus shunts the ripple component around the load. The capacitor attempts to hold the output voltage relatively constant at the average value of the voltage.

The value of the filter capacitor (C1) must be relatively large to present a low opposition (X_C) to the pulsating current and to store a substantial charge. The rate of the charge for the capacitor is limited by the low impedance of the ac source (the transformer), by the small resistance of the diode, and by the counter electromotive force (CEMF) developed by the coil. Therefore, the RC charge time constant is short compared to its discharge time. (This comparison in RC charge and discharge paths is illustrated in views A and B of figure 4-20.) Consequently, when the pulsating voltage is first applied to the LC choke-input filter, the inductor (L1) produces a CEMF which opposes the constantly increasing input voltage. The net result is to effectively prevent the rapid charging of the filter capacitor (C1). Thus, instead of reaching the peak value of the input voltage, C1 only charges to the average value of the input voltage. After the input voltage reaches its peak and decreases sufficiently, the capacitor C1) attempts to discharge through the load resistance R_L). C1 will only partially discharge, as indicated in view B of the figure, because of its relatively long discharge time constant. The larger the value of the filter capacitor, the better the filtering action. However, because of physical size, there is a practical limitation to the maximum value of the capacitor.

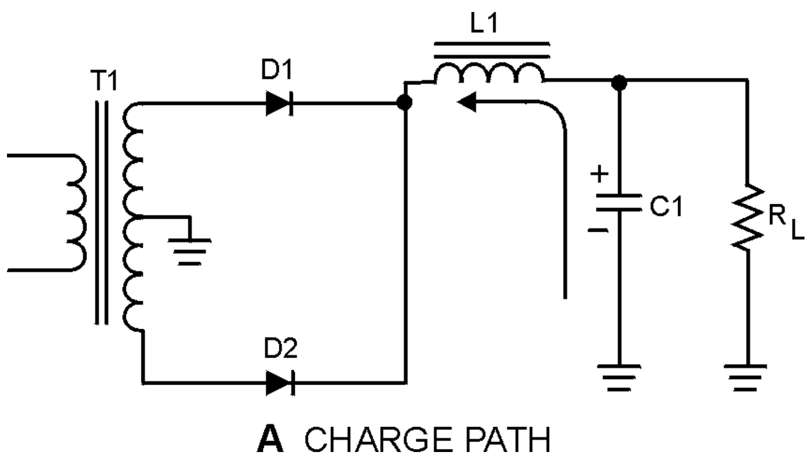


Figure 4-20A.—LC choke-input filter (charge and discharge paths). CHARGE PATH

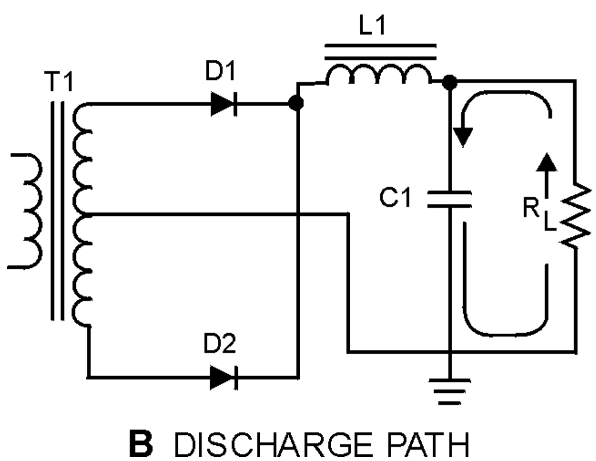


Figure 4-20B.—LC choke-input filter (charge and discharge paths). DISCHARGE PATH

The inductor (also referred to as the filter choke or coil) serves to maintain the current flow to the filter output (R_L) at a nearly constant level during the charge and discharge periods of the filter capacitor. The inductor (L1) and the capacitor (C1) form a voltage divider for the ac component (ripple) of the applied input voltage. This is shown in views A and B of figure 4-21. As far as the ripple component is concerned, the inductor offers a high impedance (Z) and the capacitor offers a low impedance (view B). As a result, the ripple component (E_r) appearing across the load resistance is greatly attenuated (reduced). The inductance of the filter choke opposes changes in the value of the current flowing through it;

therefore, the average value of the voltage produced across the capacitor contains a much smaller value of ripple component (E_r) than the value of ripple produced across the choke.

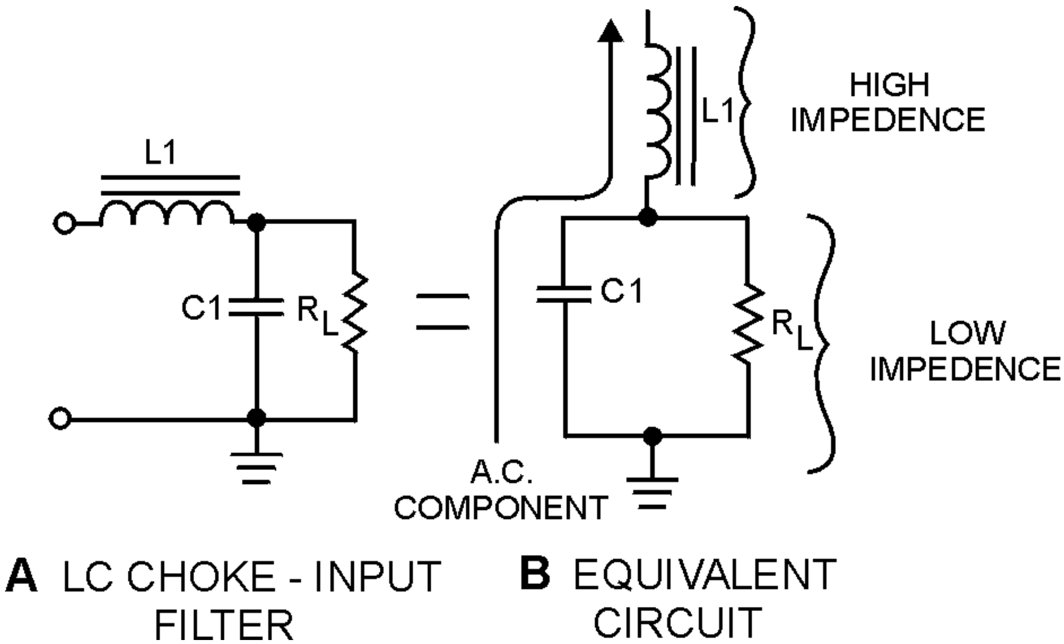


Figure 4-21.—LC choke-input filter.

Now look at figure 4-22 which illustrates a complete cycle of operation for a full-wave rectifier circuit used to supply the input voltage to the filter. The rectifier voltage is developed across the capacitor (C1). The ripple voltage at the output of the filter is the alternating component of the input voltage reduced in amplitude by the filter section. Each time the anode of a diode goes positive with respect to the cathode, the diode conducts and C1 charges. Conduction occurs twice during each cycle for a full-wave rectifier. For a 60-hertz supply, this produces a 120-hertz ripple voltage. Although the diodes alternate (one conducts while the other is nonconducting), the filter input voltage is not steady. As the anode voltage of the conducting diode increases (on the positive half of the cycle), capacitor C1 charges-the charge being limited by the impedance of the secondary transformer winding, the diode's forward (cathode-to-anode) resistance, and the counter electromotive force developed by the choke. During the nonconducting interval (when the anode voltage drops below the capacitor charge voltage), C1 discharges through the load resistor (R_L). The components in the discharge path have a long time constant; thus, C1 discharges more slowly than it charges.

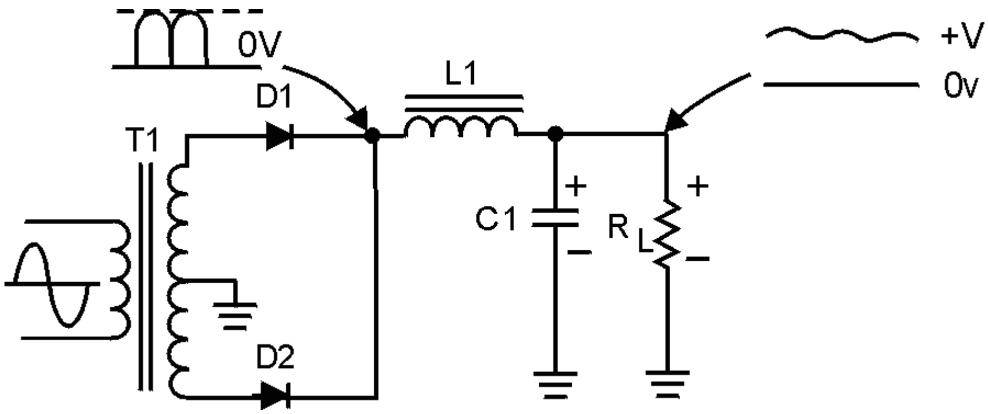


Figure 4-22.—Filtering action of the LC choke-input filter.

The choke (L1) is usually a large value, from 1 to 20 henries, and offers a large inductive reactance to the 120-hertz ripple component produced by the rectifier. Therefore, the effect that L1 has on the charging of the capacitor (C1) must be considered. Since L1 is connected in series with the parallel branch consisting of C1 and R_L, a division of the ripple (ac) voltage and the output (dc) voltage occurs. The greater the impedance of the choke, the less the ripple voltage that appears across C1 and the output. The dc output voltage is fixed mainly by the dc resistance of the choke.

Now that you have read how the LC choke-input filter functions, it will be discussed with actual component values applied. For simplicity, the input frequency at the primary of the transformer will be 117 volts 60 hertz. Both half-wave and full-wave rectifier circuits will be used to provide the input to the filter.

Starting with the half-wave configuration shown in figure 4-23, the basic parameters are: With 117 volts ac rms applied to the T1 primary, 165 volts ac peak is available at the secondary [(117 V) × (1.414) = 165 V]. You should recall that the ripple frequency of this half-wave rectifier is 60 hertz. Therefore, the capacitive reactance of C1 is:

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{(2)(3.14)(60)(10)(10^{-6})}$$

$$X_C = \frac{(1)(10^6)}{3768}$$

$$X_C = 265\Omega$$

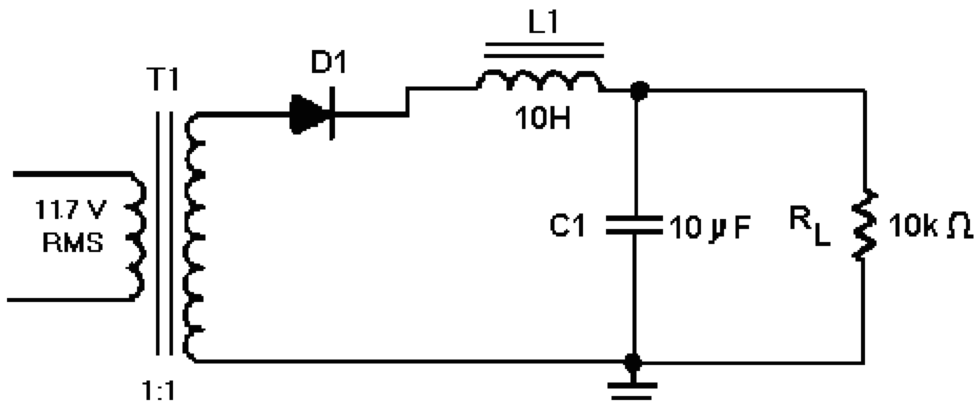


Figure 4-23.—Half-wave rectifier with an LC choke-input filter.

This means that the capacitor (C1) offers 265 ohms of opposition to the ripple current. Note, however, that the capacitor offers an infinite impedance to direct current. The inductive reactance of L1 is:

$$X_L = 2\pi fL$$
$$X_L = (2)(3.14)(60)(10)$$
$$X_L = 3.8 \text{ kilohms}$$

The above calculation shows that L1 offers a relatively high opposition (3.8 kilohms) to the ripple in comparison to the opposition offered by C1 (265 ohms). Thus, more ripple voltage will be dropped across L1 than across C1. In addition, the impedance of C1 (265 ohms) is relatively low with respect to the resistance of the load (10 kilohms). Therefore, more ripple current flows through C1 than the load. In other words, C1 shunts most of the ac component around the load.

Let's go a step further and redraw the filter circuit so that you can see the voltage divider action. Refer to view A of figure 4-24. Remember, the 165 volts peak 60 hertz provided by the rectifier consists of both an ac and a dc component. This first discussion will be about the ac component. From the figure, you see that the capacitor (C1) offers the least opposition (265 ohms) to the ac component. Therefore, the greater amount of ac will flow through C1. (The heavy line in view B indicates the ac current flow through the capacitor.) Thus the capacitor bypasses, or shunts, most of the ac around the load.

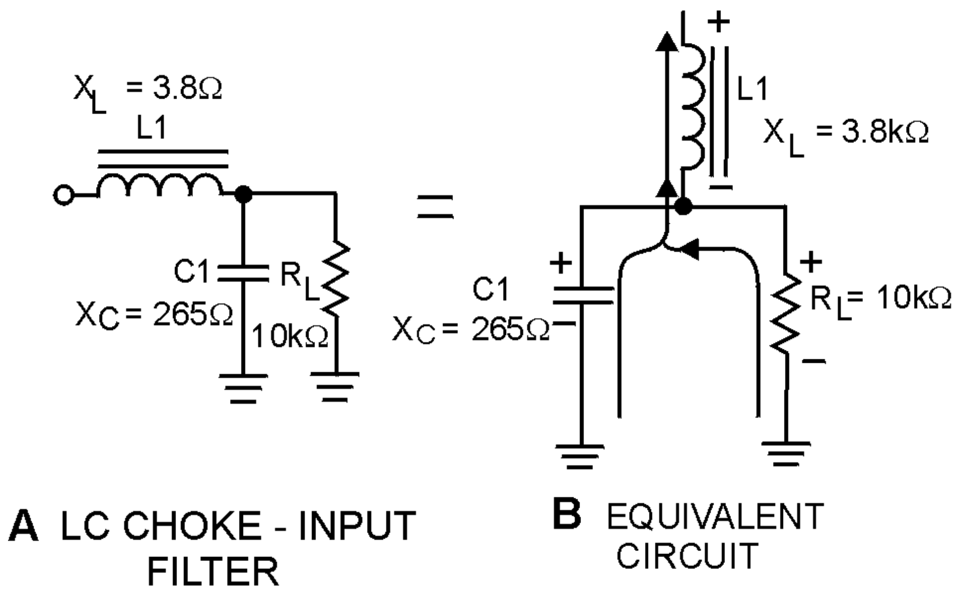


Figure 4-24.—Ac component in an LC choke-input filter.

By combining the X_C of C1 and the resistance of R_L into an equivalent circuit (view B), you will have an equivalent impedance of 265 ohms.

As a formula;

$$RT = \frac{(R1)(R2)}{R1 + R2}$$

You now have a voltage divider as illustrated in figure 4-25. You should see that because of the impedance ratios, a large amount of ripple voltage is dropped across L1, and a substantially smaller amount is dropped across C1 and R_L. You can further increase the ripple voltage across L1 by increasing the inductance ($X_L = 2\pi fL$).

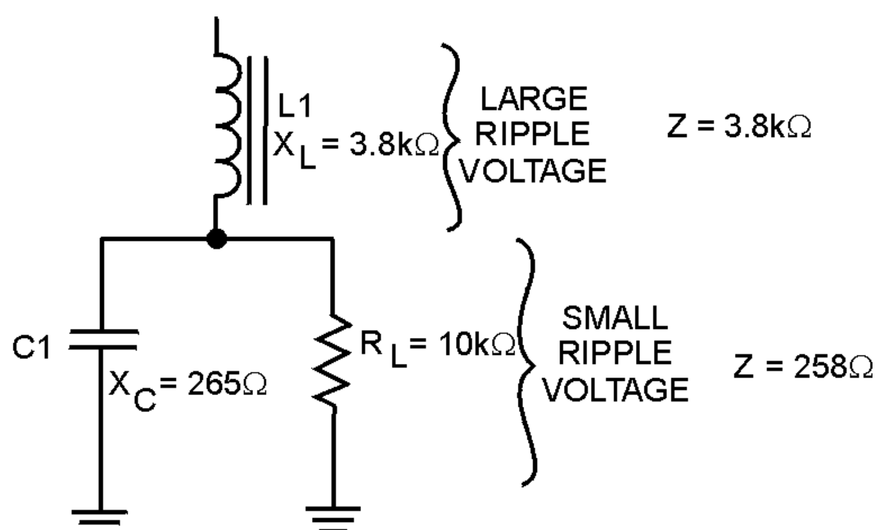


Figure 4-25.—Equivalent circuit of an LC choke-input filter.

Now let's discuss the dc component of the applied voltage. Remember, a capacitor offers an infinite (∞) impedance to the flow of direct current. The dc component, therefore, must flow through R_L and L1. As far as the dc is concerned, the capacitor does not exist. The coil and the load are therefore in series with each other. The dc resistance of a filter choke is very low (50 ohms average). Consequently, most of the dc component is developed across the load and a very small amount of the dc voltage is dropped across the coil, as shown in figure 4-26.

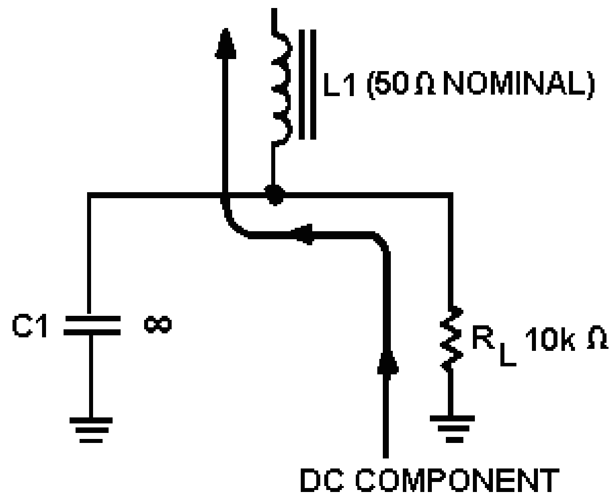


Figure 4-26.—Dc component in an LC choke-input filter.

As you may have noticed, both the ac and the dc components flow through L1. Because it is frequency sensitive, the coil provides a large resistance to ac and a small resistance to dc. In other words, the coil opposes any change in current. This property makes the coil a highly desirable filter component. Note that the filtering action of the LC choke-input filter is improved when the filter is used in conjunction with a full-wave rectifier, as shown in figure 4-27. This is due to the decrease in the X_C of the filter capacitor and the increase in the X_L of the choke. Remember, ripple frequency of a full-wave rectifier is twice that of a half-wave rectifier. For 60-hertz input, the ripple will be 120 hertz. The X_C of C1 and the X_L of L1 are calculated as follows:

$$X_C = \frac{1}{2\pi fC}$$

$$X_C = \frac{1}{(2)(3.14)(120)(10)(10^{-6})}$$

$$X_C = \frac{(1)(10^6)}{7536}$$

$$X_C = 132.5\Omega$$

$$X_L = 2\pi fL$$

$$X_L = (2)(3.14)(120)(10)$$

$$X_L = 7.5 \text{ kilohms}$$

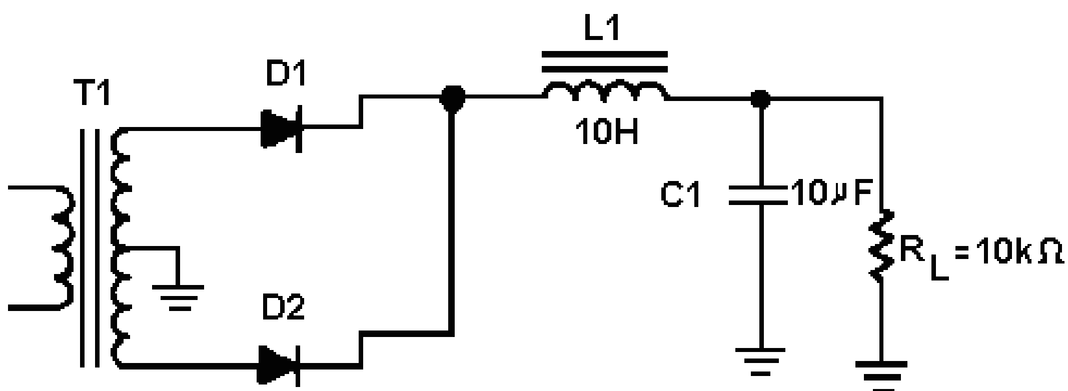


Figure 4-27.—Full-wave rectifier with an LC choke-input filter.

When the X_C of a filter capacitor is decreased, it provides less opposition to the flow of ac. The greater the ac flow through the capacitor, the lower the flow through the load. Conversely, the larger the X_L of the choke, the greater the amount of ac ripple developed across the choke; consequently, less ripple is developed across the load and better filtering is obtained.

Q21. In an LC choke-input filter, what prevents the rapid charging of the capacitor?

Q22. What is the range of values usually chosen for a choke?

Q23. If the impedance of the choke is increased, will the ripple amplitude increase or decrease?

FAILURE ANALYSIS OF AN LC CHOKE-INPUT FILTER.—The filter capacitors are subject to open circuits, short circuits, and excessive leakage; the series inductor is subject to open windings and, occasionally, shorted turns or a short circuit to the core.

The filter capacitor in the LC choke-input filter circuit is not subject to extreme voltage surges because of the protection offered by the inductor. However, the capacitor can become open, leaky, or shorted.

Shorted turns in the choke may reduce the value of inductance below the critical value. This will result in excessive peak-rectifier current, accompanied by an abnormally high output voltage, excessive ripple amplitude, and poor voltage regulation.

A choke winding that is open, or a choke winding which is shorted to the core will result in a no-output condition. A choke winding which is shorted to the core may cause overheating of the rectifier element(s) and blown fuses.

With the supply voltage removed from the input to the filter circuit, one terminal of the capacitor can be disconnected from the circuit. The capacitor should be checked with a capacitance analyzer to determine its capacitance and leakage resistance. When the capacitor is electrolytic, you must use the correct polarity at all times. A decrease in capacitance or losses within the capacitor can decrease the efficiency of the filter and can produce excessive ripple amplitude.

Resistor-Capacitor (RC) Filters

The RC capacitor-input filter is limited to applications in which the load current is small. This type of filter is used in power supplies where the load current is constant and voltage regulation is not necessary. For example, RC filters are used in high-voltage power supplies for cathode-ray tubes and in decoupling networks for multistage amplifiers.

Figure 4-28 shows an RC capacitor-input filter and associated waveforms. Both half-wave and full-wave rectifiers are used to provide the inputs. The waveform shown in view A of the figure represent the unfiltered output from a typical rectifier circuit. Note that the dashed lines in view A indicate the average value of output voltage (E_{avg}) for the half-wave rectifier. The average output voltage (E_{avg}) is less than half (approximately 0.318) the amplitude of the voltage peaks. The average value of output voltage (E_{avg}) for the full-wave rectifier is greater than half (approximately 0.637), but is still much less than, the peak amplitude of the rectifier-output waveform. With no filter circuit connected across the output of the rectifier circuit (unfiltered), the waveform has a large value of pulsating component (ripple) as compared to the average (or dc) component.

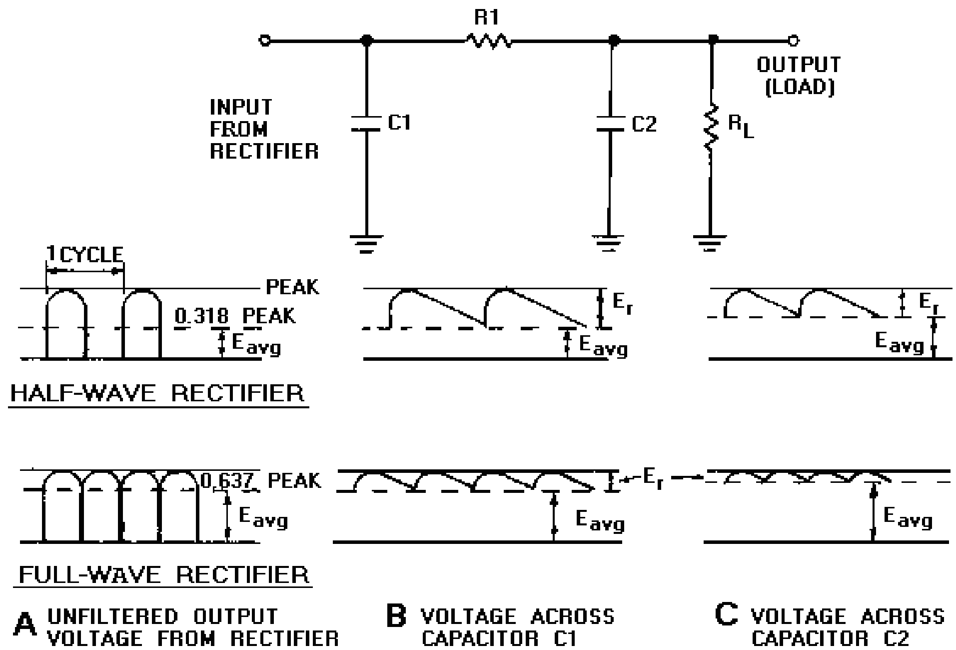


Figure 4-28.—RC filter and waveforms.

The RC filter in figure 4-28 consists of an input filter capacitor (C1), a series resistor (R1), and an output filter capacitor (C2). (This filter is sometimes referred to as an RC pi-section filter because its schematic symbol resembles the Greek letter π).

The single capacitor filter is suitable for many noncritical, low-current applications. However, when the load resistance is very low or when the percent of ripple must be held to an absolute minimum, the capacitor value required must be extremely large. While electrolytic capacitors are available in sizes up to 10,000 microfarads or greater, the large sizes are quite expensive. A more practical approach is to use a more sophisticated filter that can do the same job but that has lower capacitor values, such as the RC filter.

Views A, B, and C of figure 4-28 show the output waveforms of a half-wave and a full-wave rectifier. Each waveform is shown with an RC filter connected across the output. The following explanation of how a filter works will show you that an RC filter of this type does a much better job than the single capacitor filter.

C1 performs exactly the same function as it did in the single capacitor filter. It is used to reduce the percentage of ripple to a relatively low value. Thus, the voltage across C1 might consist of an average dc value of +100 volts with a ripple voltage of 10 volts peak-to-peak. This voltage is passed on to the R1-C2 network, which reduces the ripple even further.

C2 offers an infinite impedance (resistance) to the dc component of the output voltage. Thus, the dc voltage is passed to the load, but reduced in value by the amount of the voltage drop across R1. However, R1 is generally small compared to the load resistance. Therefore, the drop in the dc voltage by R1 is not a drawback.

Component values are designed so that the resistance of R1 is much greater than the reactance (X_C) of C2 at the ripple frequency. C2 offers a very low impedance to the ac ripple frequency. Thus, the ac ripple senses a voltage divider consisting of R1 and C2 between the output of the rectifier and ground. Therefore, most of the ripple voltage is dropped across R1. Only a trace of the ripple voltage can be seen across C2 and the load. In extreme cases where the ripple must be held to an absolute minimum, a second stage of RC filtering can be added. In practice, the second stage is rarely required. The RC filter is extremely popular because smaller capacitors can be used with good results.

The RC filter has some disadvantages. First, the voltage drop across R1 takes voltage away from the load. Second, power is wasted in R1 and is dissipated in the form of unwanted heat. Finally, if the load resistance changes, the voltage across the load will change. Even so, the advantages of the RC filter overshadow these disadvantages in many cases.

Q24. Why is the use of large value capacitors in filter circuits discouraged?

Q25. When is a second RC filter stage used?

FAILURE ANALYSIS OF THE RESISTOR-CAPACITOR (RC) FILTER.—The shunt capacitors (C1 and C2) are subject to an open circuit, a short circuit, or excessive leakage. The series filter resistor (R1) is subject to changes in value and occasionally opens. Any of these troubles can be easily detected.

The input capacitor (C1) has the greatest pulsating voltage applied to it and is the most susceptible to voltage surges. As a result, the input capacitor is frequently subject to voltage breakdown and shorting. The remaining shunt capacitor (C2) in the filter circuit is not subject to voltage surges because of the protection offered by the series filter resistor (R1). However, a shunt capacitor can become open, leaky, or shorted.

A shorted capacitor or an open filter resistor results in a no-output indication. An open filter resistor results in an abnormally high dc voltage at the input to the filter and no voltage at the output of the filter. Leaky capacitors or filter resistors that have lost their effectiveness, or filter resistors that have decreased in value, result in an excessive ripple amplitude in the output of the supply.

LC Capacitor-Input Filter

The LC capacitor-input filter is one of the most commonly used filters. This type of filter is used primarily in radio receivers, small audio amplifier power supplies, and in any type of power supply where the output current is low and the load current is relatively constant.

Figure 4-29 shows an LC capacitor-input filter and associated waveforms. Both half-wave and full-wave rectifier circuits are used to provide the input. The waveforms shown in view A of the figure represent the unfiltered output from a typical rectifier circuit. Note that the average value of output voltage (E_{avg}), indicated by the dashed lines, for the half-wave rectifier is less than half the amplitude of the voltage peaks. The average value of output voltage (E_{avg}) for the full-wave rectifier is greater than half, but is still much less than the peak amplitude of the rectifier-output waveform. With no filter connected across the output of the rectifier circuit (which results in unfiltered output voltage), the waveform has a large value of pulsating component (ripple) as compared to the average (or dc) component.

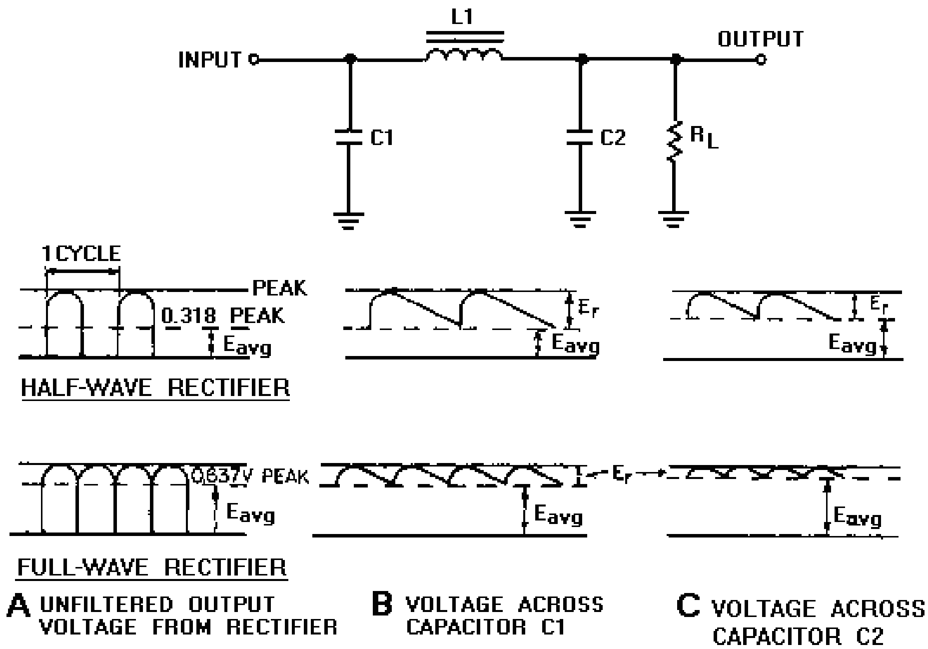


Figure 4-29.—LC filter and waveforms.

C1 reduces the ripple to a relatively low level (view B). L1 and C2 form the LC filter, which reduces the ripple even further. L1 is a large value iron-core induct (choke). L1 has a high value of inductance an therefore, a high value of X_L which offers a high reactance to the ripple frequency. At the same time, C2 offers a very low reactance to ac ripple. L1 and C2 for an ac voltage divider and, because the reactance of L1 much higher than that of C2, most of the ripple voltage is dropped across L1. Only a slight trace of ripple appears across C2 and the load (view C).

While the L1-C2 network greatly reduces ac ripple it has little effect on dc. You should recall that an inductor offers no reactance to dc. The only opposition to current flow is the resistance of the wire in the choke. Generally, this resistance is very low and the dc voltage drop across the coil is minimal. Thus, the LC filter overcomes the disadvantages of the RC filter.

Aside from the voltage divider effect, the inductor improves filtering in another way. You should recall that an inductor resists changes in the magnitude of the current flowing through it. Consequently, when the inductor is placed in series with the load, the inductor maintains steady current. In turn, this helps the voltage across the load remain constant when size of components is a factor.

The LC filter provides good filtering action over a wide range of currents. The capacitor filters best when the load is drawing little current. Thus, the capacitor discharges very slowly and the output voltage remains almost constant. On the other hand, the inductor filters best when the current is highest. The complementary nature of these two components ensures that good filtering will occur over a wide range of currents.

The LC filter has two disadvantages. First, it is more expensive than the RC filter because an iron-core choke costs more than a resistor. The second disadvantage is size. The iron-core choke is bulky and heavy, a fact which may render the LC filter unsuitable for many applications.

Q26. What is the most commonly used filter today?

Q27. What are the two main disadvantages of an LC capacitor filter?

FALURE ANALYSIS OF THE LC CAPACITOR-INPUT FILTER.—Shunt capacitors are subject to open circuits, short circuits, and excessive leakage; series inductors are subject to open windings and occasionally shorted turns or a short circuit to the core.

The input capacitor (C1) has the greatest pulsating voltage applied to it, is the most susceptible to voltage surges, and has a generally higher average voltage applied. As a result, the input capacitor is frequently subject to voltage breakdown and shorting. The output capacitor (C2) is not as susceptible to voltage surges because of the series protection offered by the series inductor (L1), but the capacitor can become open, leaky, or shorted.

A shorted capacitor, an open filter choke, or a choke winding which is shorted to the core, results in a no-output indication. A shorted capacitor, depending on the magnitude of the short, may cause a shorted rectifier, transformer, or filter choke, and may result in a blown fuse in the primary of the transformer. An open filter choke results in an abnormally high dc voltage at the input to the filter and no voltage at the output of the filter. A leaky or open capacitor in the filter circuit results in a low dc output voltage. This condition is generally accompanied by an excessive ripple amplitude. Shorted turns in the winding of a filter choke reduce the effective inductance of the choke and decrease its filtering efficiency. As a result, the ripple amplitude increases.

VOLTAGE REGULATION

Ideally, the output of most power supplies should be a constant voltage. Unfortunately, this is difficult to achieve. There are two factors that can cause the output voltage to change. First, the ac line voltage is not constant. The so-called 115 volts ac can vary from about 105 volts ac to 125 volts ac. This means that the peak ac voltage to which the rectifier responds can vary from about 148 volts to 177 volts. The ac line voltage alone can be responsible for nearly a 20 percent change in the dc output voltage. The second factor that can change the dc output voltage is a change in the load resistance. In complex electronic equipment, the load can change as circuits are switched in and out. In a television receiver, the load on a particular power supply may depend on the brightness of the screen, the control settings, or even the channel selected.

These variations in load resistance tend to change the applied dc voltage because the power supply has a fixed internal resistance. If the load resistance decreases, the internal resistance of the power supply drops more voltage. This causes a decrease in the voltage across the load.

Many circuits are designed to operate with a particular supply voltage. When the supply voltage changes, the operation of the circuit may be adversely affected. Consequently, some types of equipment must have power supplies that produce the same output voltage regardless of changes in the load resistance or changes in the ac line voltage. This constant output voltage may be achieved by adding a circuit called the VOLTAGE REGULATOR at the output of the filter. There are many different types of regulators in use today and to discuss all of them would be beyond the scope of this chapter.

LOAD REGULATION

A commonly used FIGURE OF MERIT for a power supply is its PERCENT OF REGULATION. The figure of merit gives us an indication of how much the output voltage changes over a range of load resistance values. The percent of regulation aids in the determination of the type of load regulation needed. Percent of regulation is determined by the equation:

$$\text{Percent of regulation} = \frac{(E_{nL} - E_{fL})}{E_{fL}} \times 100$$

This equation compares the change in output voltage at the two loading extremes to the voltage produced at full loading. For example, assume that a power supply produces 12 volts when the load current is zero. If the output voltage drops to 10 volts when full load current flows, then the percent of regulation is:

$$\text{Percent of regulation} = \frac{(E_{nL} - E_{fL})}{E_{fL}} \times 100$$

$$\text{Percent of regulation} = \frac{(12 - 10V)}{10V} \times 100$$

$$\text{Percent of regulation} = \frac{2V}{10V} \times 100$$

$$\text{Percent of regulation} = 20\%$$

Ideally, the output voltage should not change over the full range of operation. That is, a 12-volt power supply should produce 12 volts at no load, at full load, and at all points in between. In this case, the percent of regulation would be:

$$\text{Percent of regulation} = \frac{(E_{nL} - E_{fL})}{E_{fL}} \times 100$$

$$\text{Percent of regulation} = \frac{(12 - 12V)}{12V} \times 100$$

$$\text{Percent of regulation} = \frac{0V}{12V} \times 100$$

$$\text{Percent of regulation} = 0\%$$

Thus, zero-percent load regulation is the ideal situation. It means that the output voltage is constant under all load conditions. While you should strive for zero percent load regulation, in practical circuits you must settle for something less ideal. Even so, by using a voltage regulator, you can hold the percent of regulation to a very low value.

REGULATORS

You should know that the output of a power supply varies with changes in input voltage and circuit load current requirements. Because many electronic equipments require operating voltages and currents that must remain constant, some form of regulation is necessary. Circuits that maintain power supply voltages or current outputs within specified limits, or tolerances are called REGULATORS. They are designated as dc voltage or dc current regulators, depending on their specific application.

Voltage regulator circuits are additions to basic power supply circuits, which are made up of rectified and filter sections (figure 4-30). The purpose of the voltage regulator is to provide an output voltage with little or no variation. Regulator circuits sense changes in output voltages and compensate for the changes. Regulators that maintain voltages within plus or minus (±) 0.1 percent are quite common.

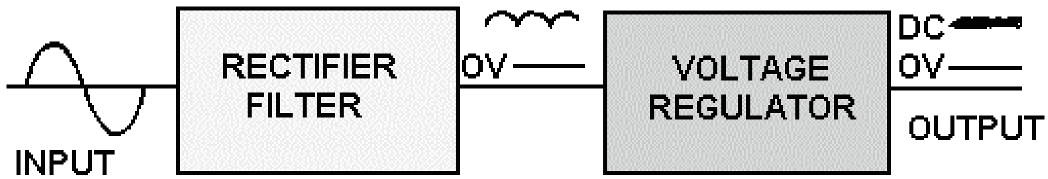


Figure 4-30.—Block diagram of a power supply and regulator.

Series and Shunt Voltage Regulators

There are two basic types of voltage regulators. Basic voltage regulators are classified as either SERIES or SHUNT, depending on the location or position of the regulating element(s) in relation to the circuit load resistance. Figure 4-31 (view A and view B) illustrates these two basic types of voltage regulators. In actual practice the circuitry of regulating devices may be quite complex. Broken lines have been used in the figure to highlight the differences between the series and shunt regulators.

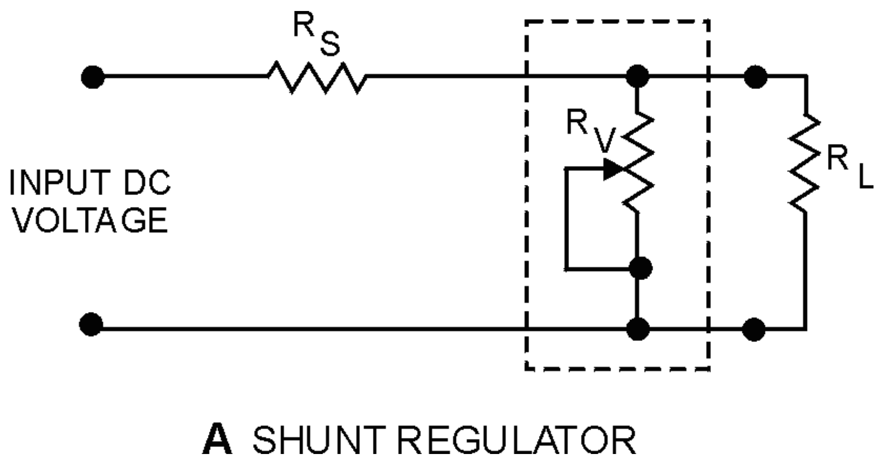


Figure 4-31A.—Simple series and shunt regulators. SHUNT REGULATOR.

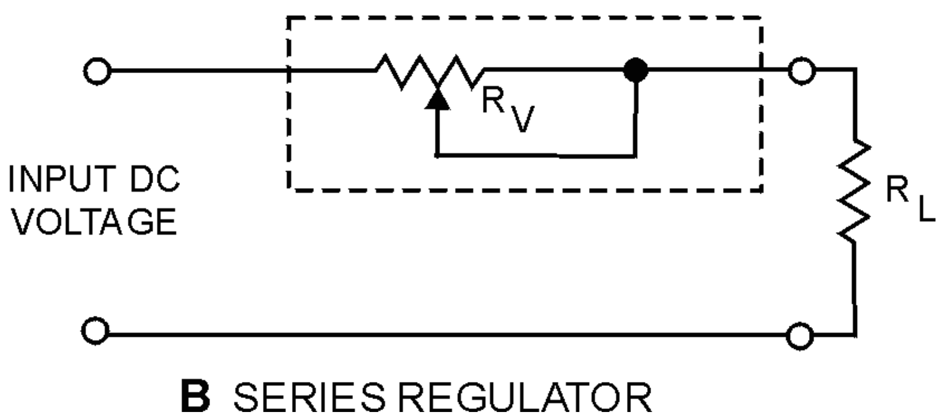


Figure 4-31B.—Simple series and shunt regulators. SERIES REGULATOR.

The schematic drawing in view A is that of a shunt-type regulator. It is called a shunt-type regulator because the regulating device is connected in parallel with the load resistance. The schematic drawing in view B is that of a series regulator. It is called a series regulator because the regulating device is connected in series with the load resistance. Figure 4-32 illustrates the principle of series voltage regulation. As you study the figure, notice that the regulator is in series with the load resistance (R_L) and that the fixed resistor (R_S) is in series with the load resistance.

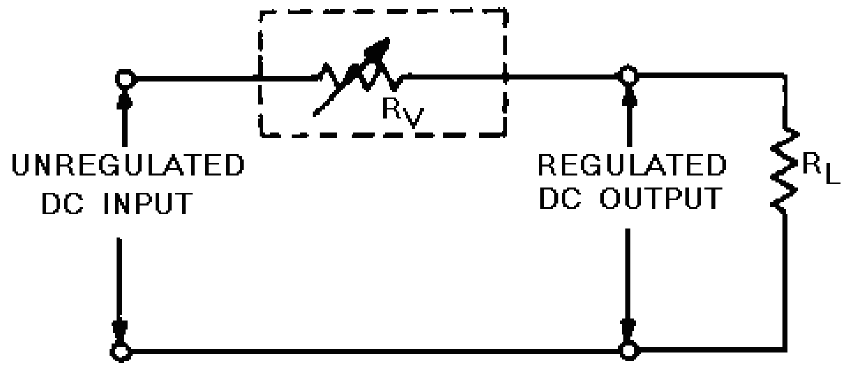


Figure 4-32.—Series voltage regulator.

You already know the voltage drop across a fixed resistor remains constant unless the current flowing through it varies (increases or decreases). In a shunt regulator, as shown in figure 4-33, output voltage regulation is determined by the current through the parallel resistance of the regulating device (R_V), the load resistance (R_L), and the series resistor (R_S). For now, assume that the circuit is operating under normal conditions, that the input is 120 volts dc, and that the desired regulated output is 100 volts dc. For a 100-volt output to be maintained, 20 volts must be dropped across the series resistor (R_S). If you assume that the value of R_S is 2 ohms, you must have 10 amperes of current through R_V and R_L. (Remember: E = IR.) If the values of the resistance of R_V and R_L are equal, 5 amperes of current will flow through each resistance (R_V and R_L).

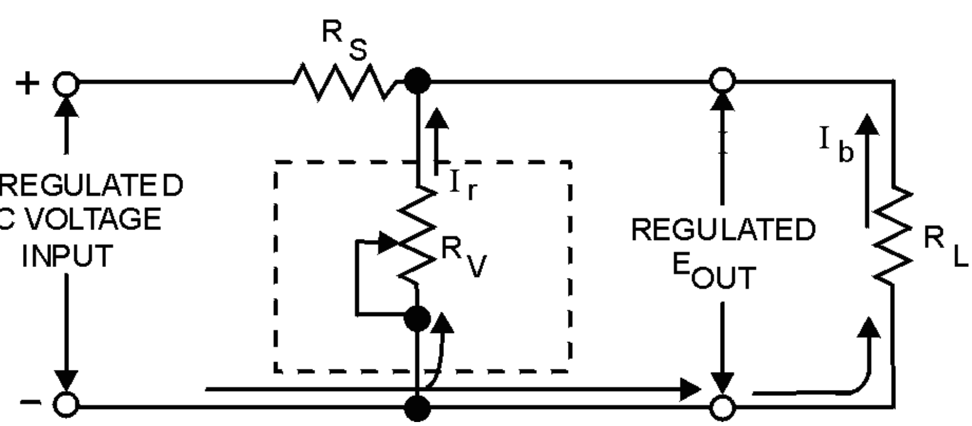


Figure 4-33.—Shunt voltage regulator.

Now, if the load resistance (R_L) increases, the current through R_L will decrease. For example, assume that the current through R_L is now 4 amperes and that the total current through R_S is 9 amperes. With this drop in current, the voltage drop across R_S is 18 volts; consequently, the output of the regulator has increased to 102 volts. At this time, the regulating device (R_V) decreases in resistance, and 6 amperes of current flows through this resistance (R_V). Thus, the total current R_S is once again 10 amperes (6 amperes through R_V; 4 amperes through R_L). Therefore, 20 volts is dropped across R_S causing the output to decrease back to 100 volts. You should know by now that if the load resistance (R_L) <u>increases</u>, the regulating device (R_V) <u>decreases</u> its resistance to compensate for the change. If R_L decreases, the opposite effect occurs and R_V increases.

Now consider the circuit when a decrease in load resistance takes place. When R_L decreases, the current through R_L subsequently increases to 6 amperes. This action causes a total of 11 amperes to flow through R_S which then drops 22 volts. As a result, the output is 98 volts. However, the regulating device (R_V) senses this change and increases its resistance so that less current (4 amperes) flows through R_V. The total current again becomes 10 amperes, and the output is again 100 volts.

From these examples, you should now understand that the shunt regulator maintains the desired output voltage first by sensing the current change in the parallel resistance of the circuit and then by compensating for the change.

Again refer to the schematic shown in figure 4-33 and consider how the voltage regulator operates to compensate for changes in input voltages. You know, of course, that the input voltage may vary and that any variation must be compensated for by the regulating device. If an <u>increase in input voltage occurs</u>, the resistance of R_V automatically decreases to maintain the correct voltage division between R_V and R_S. You should see, therefore, that the regulator operates in the opposite way to compensate for a <u>decrease in input voltage</u>.

So far only voltage regulators that use variable <u>resistors</u> have been explained. However, this type of regulation has limitations. Obviously, the variable resistor cannot be adjusted rapidly enough to compensate for frequent fluctuations in voltages. Since input voltages fluctuate frequently and rapidly, the variable resistor is <u>not</u> a practical method for voltage regulation. A voltage regulator that operates continuously and automatically to regulate the output voltage without external manipulation is required for practical regulation.

Q28. Circuits which maintain constant voltage or current outputs are called dc voltage or dc current ___.

Q29. The purpose of a voltage regulator is to provide an output voltage with little or no ___.

Q30. The two basic types of voltage regulators are ___ and ___.

Q31. When a series voltage regulator is used to control output voltages, any increase in the input voltage results in <u>an increase/a decrease</u> (which one) in the resistance of the regulating device.

Q32. A shunt-type voltage regulator is connected in <u>serial/parallel</u> (which one) with the load resistance.

The schematic for a typical series voltage regulator is shown in figure 4-34. Notice that this regulator has a transistor (Q1) in the place of the variable resistor found in figure 4-32. Because the total load current passes through this transistor, it is sometimes called a "pass transistor." Other components which make up the circuit are the current limiting resistor (R1) and the Zener diode (CR1).

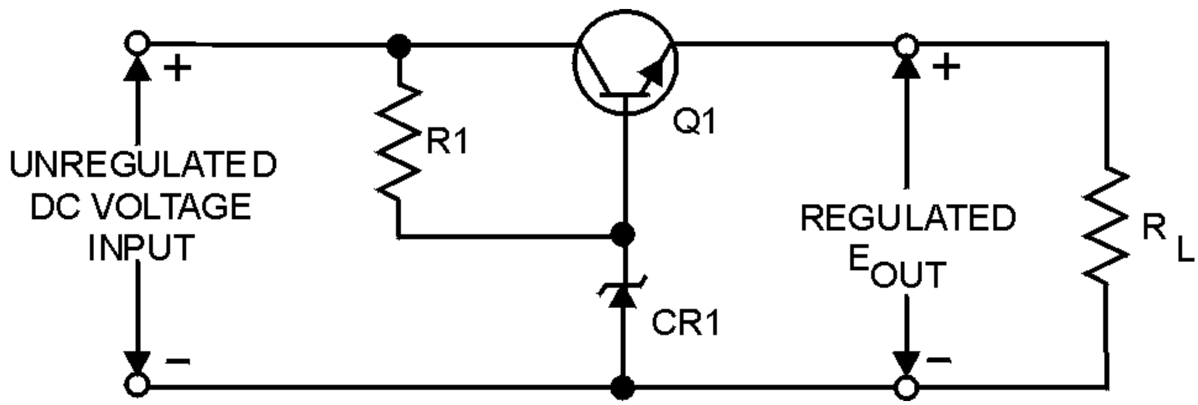


Figure 4-34.—Series voltage regulator.

Recall that a Zener diode is a diode that block current until a specified voltage is applied. Remember also that the applied voltage is called the breakdown, or Zener voltage. Zener diodes are available with different Zener voltages. When the Zener voltage is reached, the Zener diode conducts from its anode to its cathode (with the direction of the arrow).

In this voltage regulator, Q1 has a constant voltage applied to its base. This voltage is often called the reference voltage. As changes in the circuit output voltage occur, they are sensed at the emitter of Q1 producing a corresponding change in the forward bias of the transistor. In other words, Q1 compensates by increasing or decreasing its resistance in order to change the circuit voltage division.

Now, study figure 4-35. Voltages are shown to help you understand how the regulator operates. The Zener used in this regulator is a 15-volt Zener. In this instance the Zener or breakdown voltage is 15 volts. The Zener establishes the value of the base voltage for Q1. The output voltage will equal the Zener voltage minus a 0.7-volt drop across the forward biased base-emitter junction of Q1, or 14.3 volts. Because the output voltage is 14.3 volts, the voltage drop across Q1 must be 5.7 volts.

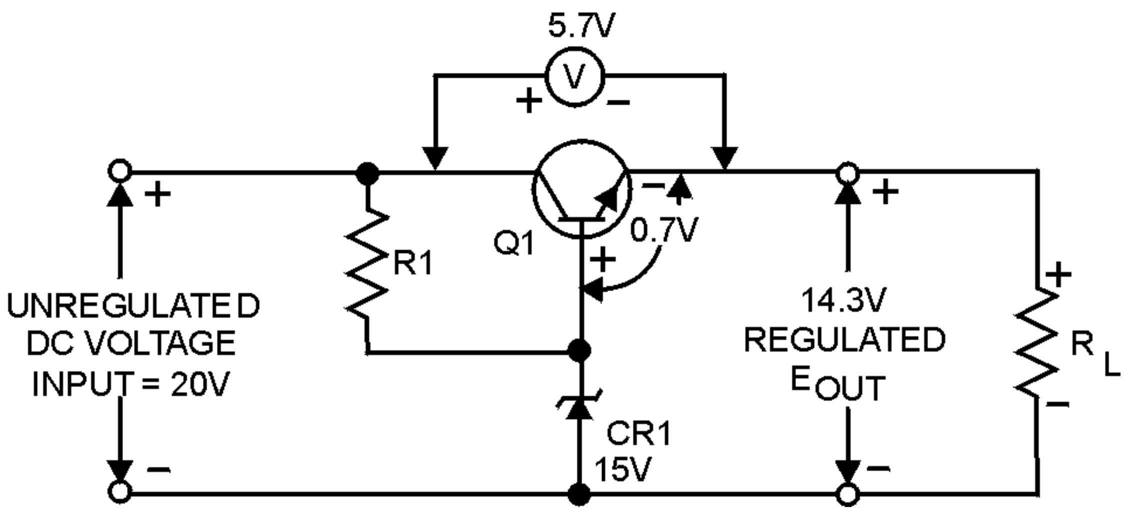


Figure 4-35.—Series voltage regulator (with voltages).

Study figure 4-36, view A, in order to understand what happens when the input voltage exceeds 20 volts. Notice the input and output voltages of 20.1 and 14.4 volts, respectively. The 14.4 output voltage is a momentary deviation, or variation, from the required regulated output voltage of 14.3 and is the result of a rise in the input voltage to 20.1 volts. Since the base voltage of Q1 is held at 15 volts by CR1, the forward bias of Q1 changes to 0.6 volt. Because this bias voltage is less than the normal 0.7 volt, the resistance of Q1 increases, thereby increasing the voltage drop across the transistor to 5.8 volts. This voltage drop restores the output voltage to 14.3 volts. The entire cycle takes only a fraction of a second and, therefore, the change is not visible on an oscilloscope or readily measurable with other standard test equipment.

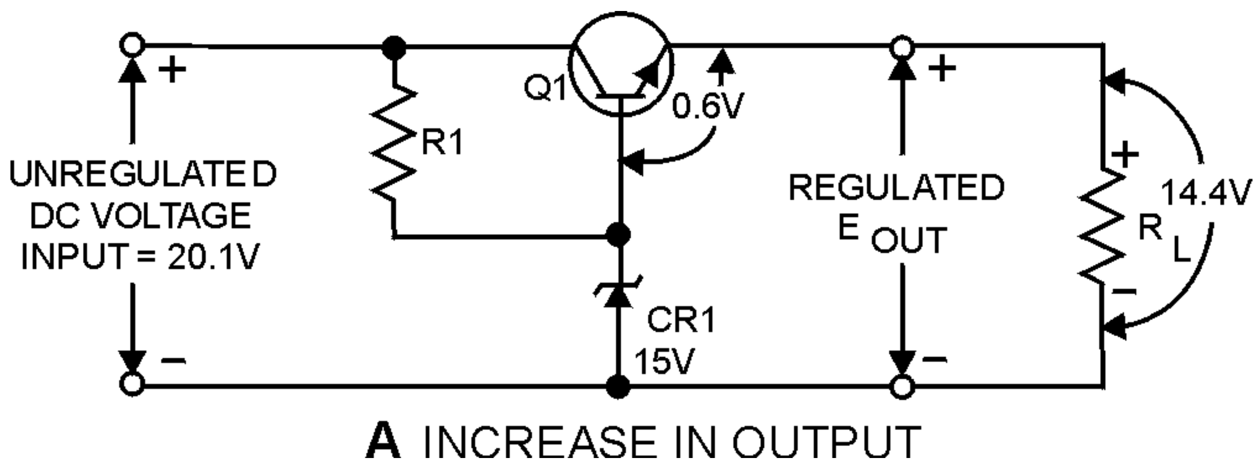


Figure 4-36A.—Series voltage regulator. INCREASE IN OUTPUT

View B is a schematic diagram for the same series voltage regulator with one significant difference. The output voltage is shown as 14.2 volts instead of the desired 14.3 volts. In this case, the load has increased causing a lowered voltage drop across R_L to 14.2 volts. When the output decreases, the forward bias of Q1 increases to 0.8 volt because Zener diode CR1 maintains the base voltage of Q1 at 15 volts. This 0.8 volt is the difference between the Zener reference voltage of 15 volts and the momentary output voltage. (15 V – 14.2 V = 0.8 V). At this point, the larger forward bias on Q1 causes the resistance of Q1 to decrease, thereby causing the voltage drop across Q1 to return to 5.7 volts. This then causes the output voltage to return to 14.3 volts.

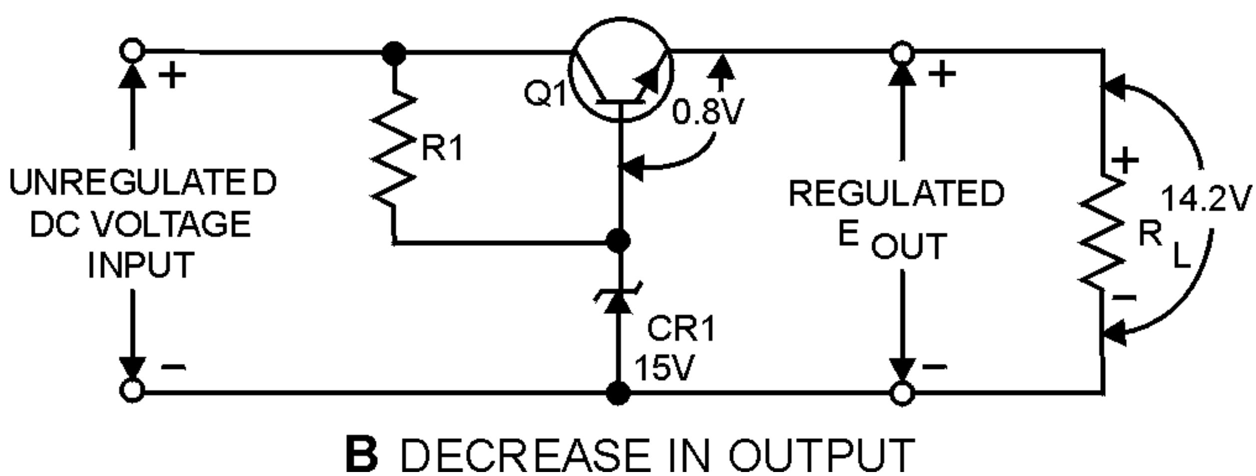


Figure 4-36B.—Series voltage regulator. DECREASE IN OUTPUT

The schematic shown in figure 4-37 is that of a shunt voltage regulator. Notice that Q1 is in parallel with the load. Components of this circuit are identical with those of the series voltage regulator except for the addition of fixed resistor R_S. As you study the schematic, you will see that this resistor is connected in series with the output load resistance. The current limiting resistor (R1) and Zener diode (CR1) provide a constant reference voltage for the base-collector junction of Q1. Notice that the bias of Q1 is determined by the voltage drop across R_S and R1. As you should know, the amount of forward bias across a transistor affects its total resistance. In this case, the voltage drop across R_S is the key to the total circuit operation.

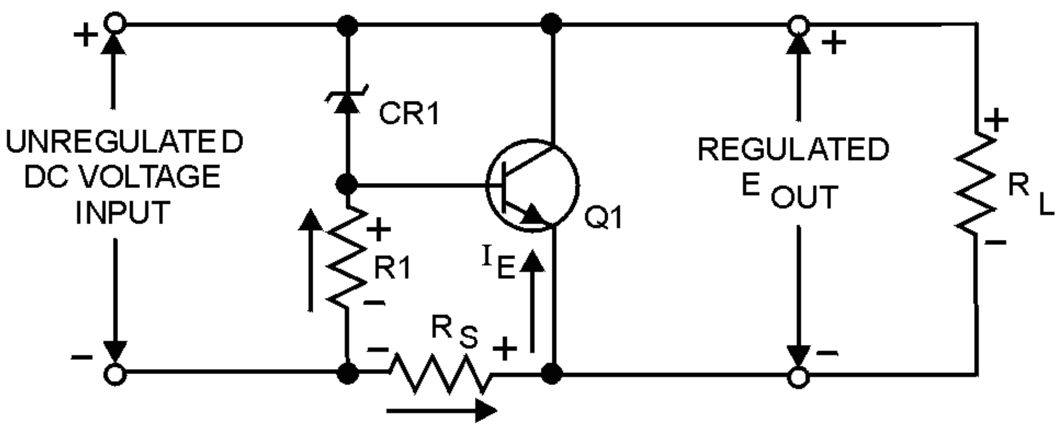


Figure 4-37.—Shunt voltage regulator.

Figure 4-38 is the schematic for a typical shunt-type regulator. Notice that the schematic is identical to the schematic shown in figure 4-37 except that voltages are shown to help you understand the functions of the various components. In the circuit shown, the voltage drop across the Zener diode (CR1) remains constant at 5.6 volts. This means that with a 20-volt input voltage, the voltage drop across R1 is 14.4 volts. With a base-emitter voltage of 0.7 volt, the output voltage is equal to the sum of the voltages across CR1 and the voltage at the base-emitter junction of Q1. In this example, with an output voltage of 6.3 volts and a 20-volt input voltage, the voltage drop across R_S equals 13.7 volts. Study the schematic to understand fully how these voltages are developed. Pay close attention to the voltages shown.

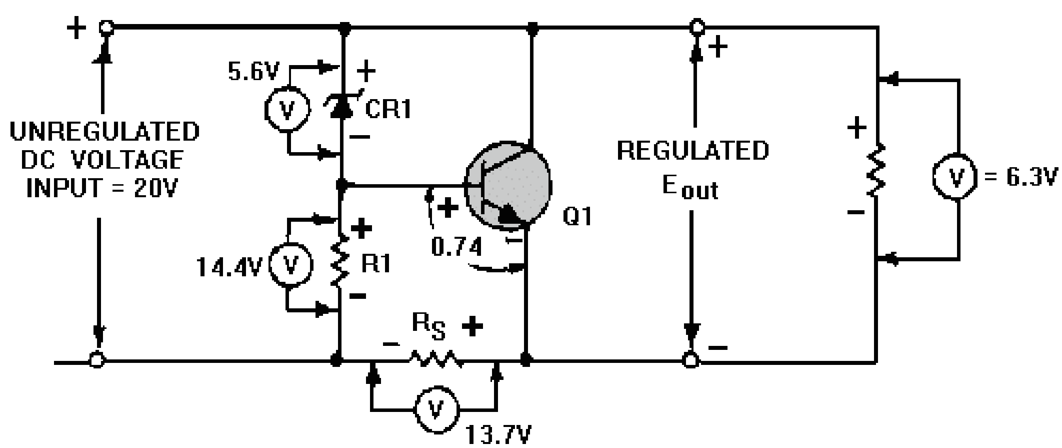


Figure 4-38.—Shunt voltage regulator (with voltages).

Now, refer to view A of figure 4-39. This figure shows the schematic diagram of the same shunt voltage regulator as that shown in figure 4-38 with an increased input voltage of 20.1 volts. This increases the forward bias on Q1 to 0.8 volt. Recall that the voltage drop across CR1 remains constant at 5.6 volts. Since the output voltage is composed of the Zener voltage and the base-emitter voltage, the output voltage momentarily increases to 6.4 volts. At this time, the increase in the forward bias of Q1 lowers the resistance of the transistor allowing more current to flow through it. Since this current must also pass through R_S, there is also an increase in the voltage drop across this resistor. The voltage drop across R_S is now 13.8 volts and therefore the output voltage is reduced to 6.3 volts. Remember, this change takes place in a fraction of a second.

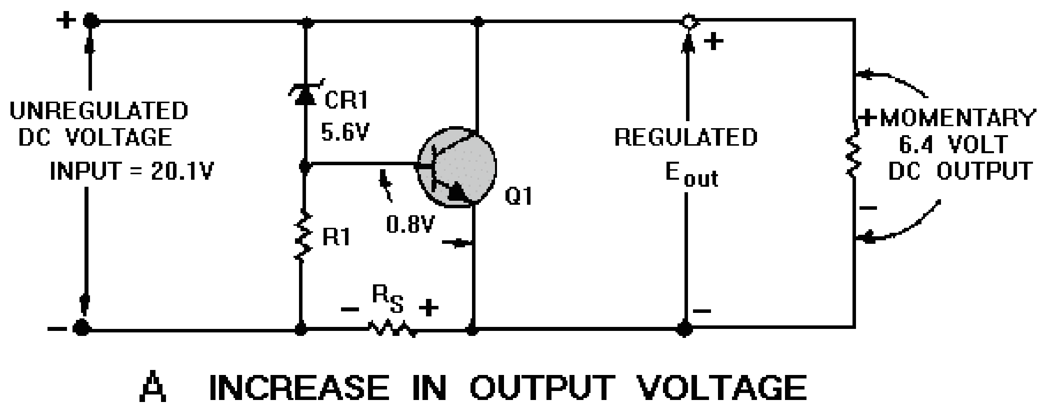


Figure 4-39A.—Shunt voltage regulator. INCREASE IN OUTPUT VOLTAGE

Study the schematic shown in view B. Although this schematic is identical to the other shunt voltage schematics previously illustrated and discussed, the output voltage is different. The load current has increased causing a <u>momentary</u> drop in voltage output to 6.2 volts. Recall that the circuit was designed to ensure a constant output voltage of 6.3 volts. Since the output voltage is less than that required, changes occur in the regulator to restore the output to 6.3 volts. Because of the 0.1 volt drop in the output voltage, the forward bias of Q1 is now 0.6 volt. This decrease in the forward bias increases the resistance of the transistor, thereby reducing the current flow through Q1 by the same amount that the load current increased. The current flow through R_S returns to its normal value and restores the output voltage to 6.3 volts.

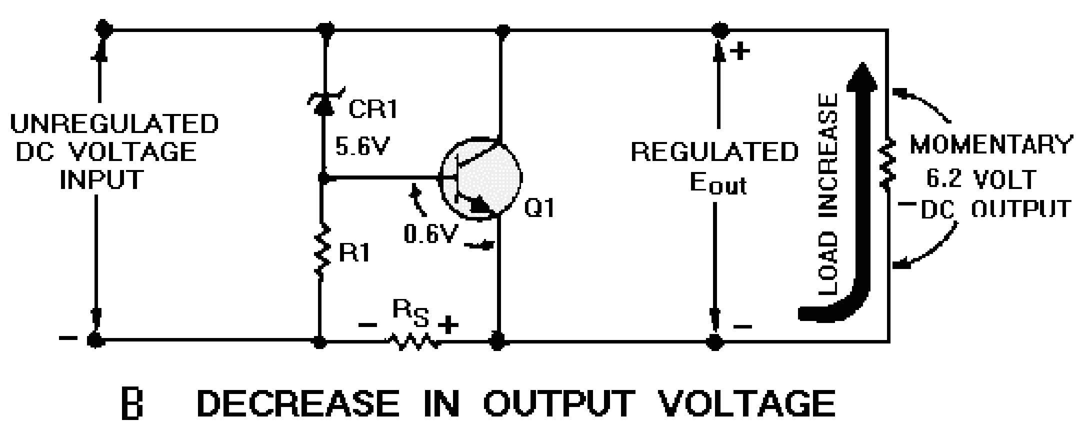


Figure 4-39B.—Shunt voltage regulator. DECREASE IN OUTPUT VOLTAGE

Q33. In figure 4-37, the voltage drop across R_S and R1 determines the amount of base-emitter ____ for Q1.

Q34. In figure 4-39, view A, when there is an increase in the input voltage, the forward bias of Q1 increases/decreases (which one).

Q35. In view B of figure 4-39, when the load current increases and the output voltage momentarily drops, the resistance of Q1 increase/decreases (which one) to compensate.

Current Regulators

You should now know how voltage regulators work to provide constant output voltages. In some circuits it may be necessary to regulate the current output. The circuitry which provides a constant current output is called a constant current regulator or just CURRENT REGULATOR. The schematic shown in figure 4-40 is a simplified schematic for a current regulator. The variable resistor shown on the schematic is used to illustrate the concept of current regulation. You should know from your study of voltage regulators that a variable resistor does not respond quickly enough to compensate for the changes. Notice that an ammeter has been included in this circuit to indicate that the circuit shown is that of a current regulator. When the circuit functions properly, the current reading of the ammeter remains constant. In this case the variable resistor (R_V) compensates for changes in the load or dc input voltage. Adequate current regulation results in the loss of voltage regulation. Studying the schematic shown, you should recall that any increase in load resistance causes a drop in current. To maintain a constant current flow, the resistance of R_V must be reduced whenever the load resistance increases. This causes the total resistance to remain constant. An increase in the input voltage must be compensated for by an increase in the resistance of R_V, thereby maintaining a constant current flow. The operation of a current regulator is similar to that of a voltage regulator. The basic difference is that one regulates current and the other regulates voltage.

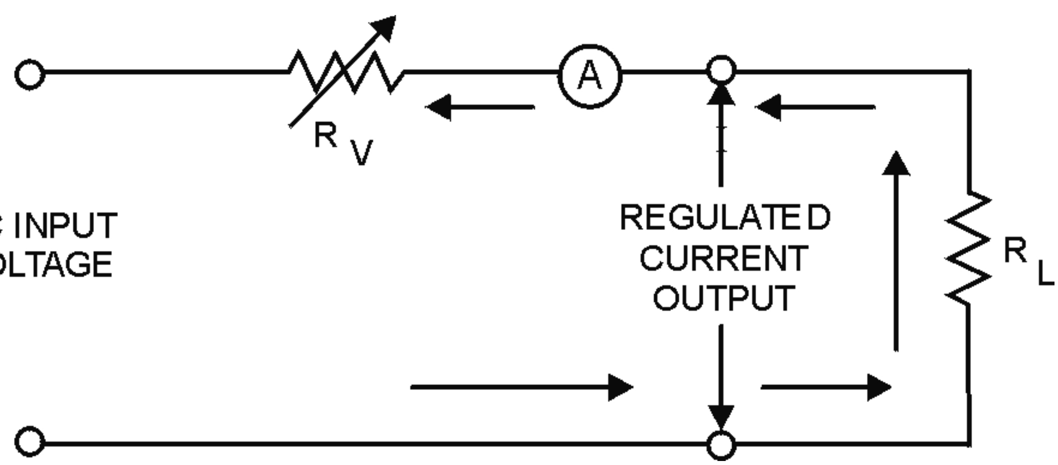


Figure 4-40.—Current regulator (simplified).

Since use of a variable resistor is not a practical way to control current fluctuation or variation, a transistor and a Zener diode, together with necessary resistors, are used. Recall that the Zener diode provides a constant reference voltage. The schematic shown in figure 4-41 is that of a current regulator circuit. Except for the addition of R1, the circuit shown in the figure is similar to that of a series voltage regulator. The resistor is connected in series with the load and senses any current changes in the load. Notice the voltage drop across R1 and the negative voltage polarity applied to the emitter of Q1. The voltage polarity is a result of current flowing through R1, and this negative voltage opposes the forward bias for Q1. However, since the regulated voltage across CR1 has an opposite polarity, the actual bias of the transistor is the difference between the two voltages. You should see that the purpose of R2 is to function as a current-limiting resistor for the Zener diode.

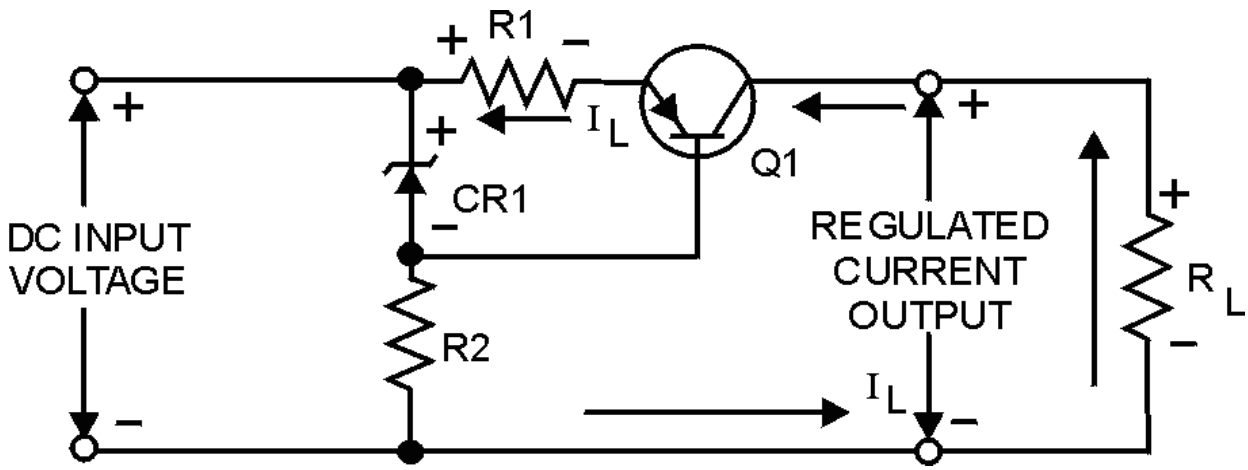


Figure 4-41.—Current regulator.

The purpose of a current regulator is to provide a constant current regardless of changes in the input voltage or load current. The schematic shown in figure 4-42 is that of a circuit designed to provide a constant current of 400 milliamperes. Voltmeters are shown in the schematic to emphasize the voltage drops across specific components. These voltages will help you understand how the current regulator operates. The voltage drop across the base-emitter junction of Q1 is 0.6 volt. This voltage is the difference between the Zener voltage and the voltage drop across R1. The 0.6-volt forward bias of Q1 permits proper operation of the transistor. The output voltage across R_L is 6 volts as shown by the voltmeter. With a regulated current output of 400 milliamperes, the transistor resistance (R_{Q1}) is 9 ohms. This can be proved by using Ohm's law and the values shown on the schematic. In this case, current (I) is equal to the voltage drop (E) divided by the resistance (R). Therefore: 12 volts divided by 30 ohms equals 0.4 ampere, or 400 milliamperes.

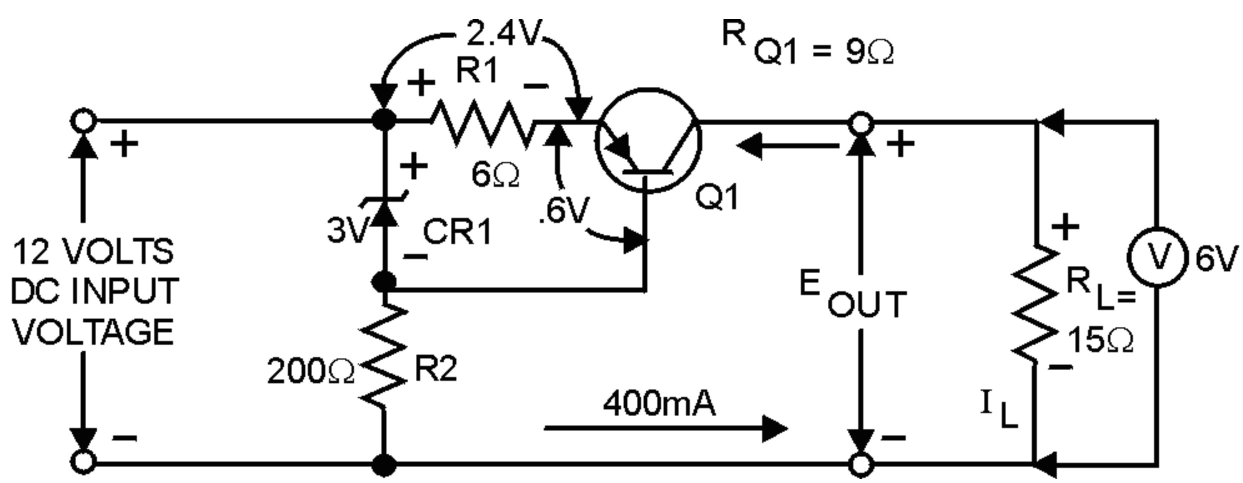


Figure 4-42.—Current regulator (with circuit values).

Since you are familiar with the basic current regulating circuitry, let's examine in detail how the various components work to maintain the constant 400-milliampere output. Refer to the schematic shown in figure 4-43. Remember a decrease in load resistance causes a corresponding increase in current flow. In the example shown, the load resistance R_L has dropped from 15 ohms to 10 ohms. This results in a larger voltage drop across R1 because of the increased current flow. The voltage drop has increased from 2.4 volts to 2.5 volts. Of course, the voltage drop across CR1 remains constant at 9 volts due to its regulating ability. Because of the increased voltage drop across R1, the forward bias on Q1 is now 0.5 volt. Since the forward bias of Q1 has decreased, the resistance of the transistor increases from 9 ohms to 14 ohms. Notice that the 5 ohm increase in resistance across the transistor corresponds to the 5 ohm decrease in the load resistance. Thus, the total resistance around the outside loop of the circuit remains constant. Since the circuit is a current regulator, you know that output voltages will vary as the regulator maintains a constant current output. In the figure, the voltage output is reduced to 4 volts, which is computed by multiplying current (I) times resistance (R) (400 mA × 10 ohms = 4 volts).

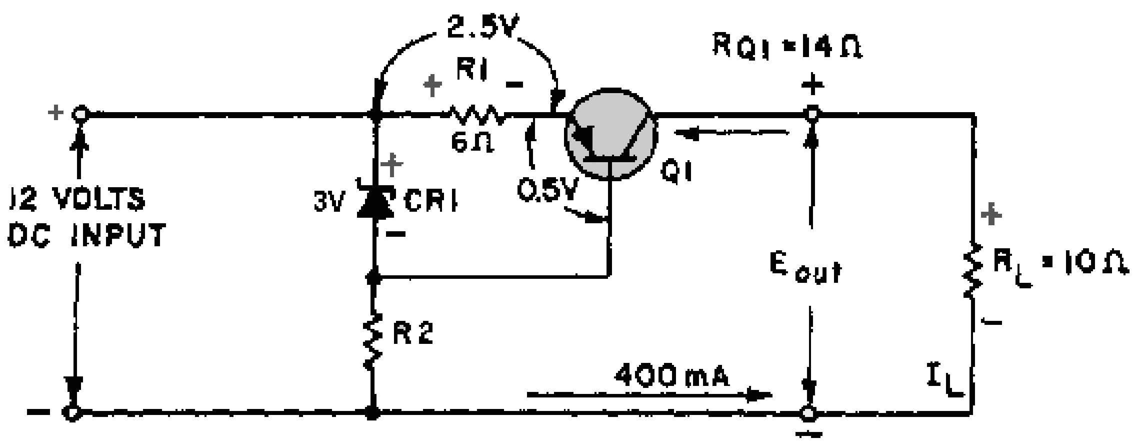


Figure 4-43.—Current regulator (with a decrease in R_L).

Q36. In figure 4-40, when there is an increase in the load resistance (R_L), the resistance of R_V <u>increases/decreases</u> (which one) to compensate for the change.

Q37. In figure 4-43 any decrease in the base-emitter forward bias across Q1 results in <u>an increase/a decrease</u> (which one) in the resistance of the transistor.

VOLTAGE MULTIPLIERS

You may already know how a transformer functions to increase or decrease voltages. You may also have learned that a transformer secondary may provide one or several ac voltage outputs which may be greater or less than the input voltage. When voltages are stepped up, current is decreased; when voltages are stepped down, current is increased.

Another method for increasing voltages is known as voltage multiplication. VOLTAGE MULTIPLIERS are used primarily to develop high voltages where low current is required. The most common application of the high voltage outputs of voltage multipliers is the anode of cathode-ray tubes (CRT), which are used for radar scope presentations, oscilloscope presentations, or TV picture tubes. The dc output of the voltage multiplier ranges from 1000 volts to 30,000 volts. The actual voltage depends upon the size of the CRT and its equipment application.

Voltage multipliers may also be used as primary power supplies where a 177 volt-ac input is rectified to pulsating dc. This dc output voltage may be increased (through use of a voltage multiplier) to as much as 1000 volts dc. This voltage is generally used as the plate or screen grid voltage for electron tubes.

If you have studied transformers, you may have learned that when voltage is stepped up, the output current decreases. This is also true of voltage multipliers. Although the measured output voltage of a voltage multiplier may be several times greater than the input voltage, once a load is connected the value of the output voltage decreases. Also any small fluctuation of load impedance causes a large fluctuation in the output voltage of the multiplier. For this reason, voltage multipliers are used only in special applications where the load is constant and has a high impedance or where input voltage stability is not critical.

Voltage multipliers may be classified as voltage doublers, triplers, or quadruplers. The classification depends on the ratio of the output voltage to the input voltage. For example, a voltage multiplier that increases the peak input voltage twice is called a voltage doubler. Voltage multipliers increase voltages through the use of series-aiding voltage sources. This can be compared to the connection of dry cells (batteries) in series.

The figures used in the explanation of voltage multipliers show a transformer input, even though for some applications a transformer is not necessary. The input could be directly from the power source or line voltage. This, of course, does not isolate the equipment from the line and creates a potentially hazardous condition. Most military equipments use transformers to minimize this hazard.

Figure 4-44 shows the schematic for a half-wave voltage doubler. Notice the similarities between this schematic and those of half-wave voltage rectifiers. In fact, the doubler shown is made up of two half-wave voltage rectifiers. C1 and CR1 make up one half-wave rectifier, and C2 and CR2 make up the other. The schematic of the first half-wave rectifier is indicated by the dark lines in view A of figure 4-45. The dotted lines and associated components represent the other half-wave rectifier and load resistor.

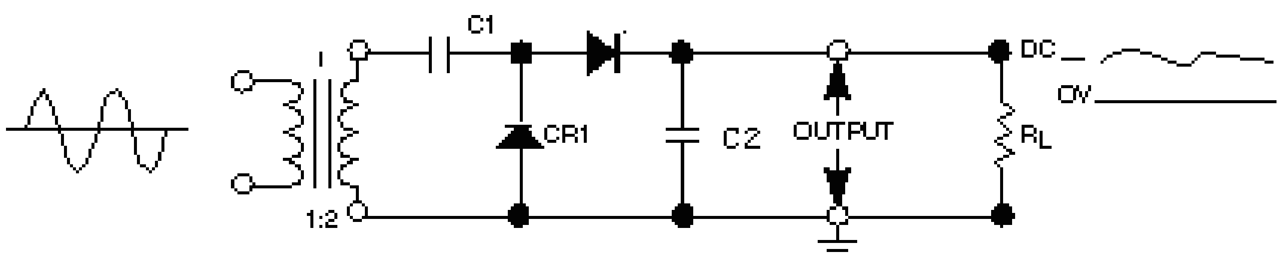


Figure 4-44.—Half-wave voltage doubler.

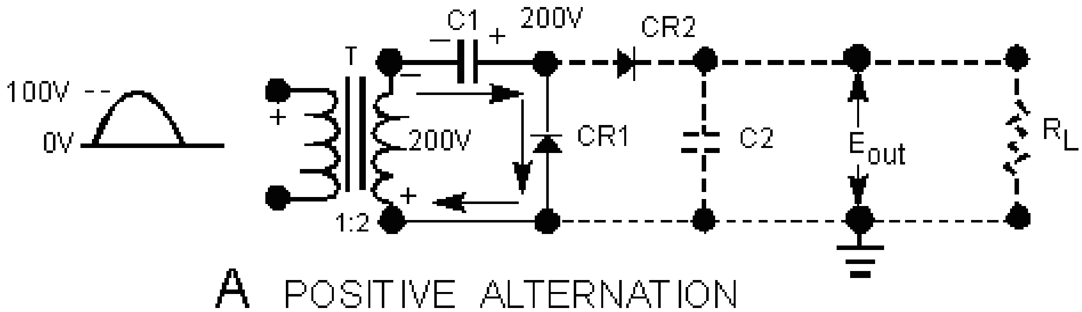


Figure 4-45A.—Rectifier action of CR1 and CR2. POSITIVE ALTERNATION

Notice that C1 and CR1 work exactly like a half-wave rectifier. During the positive alternation of the input cycle (view A), the polarity across the secondary winding of the transformer is as shown. Note that the top of the secondary is negative. At this time CR1 is forward biased (cathode negative in respect to the anode). This forward bias causes CR1 to function like a closed switch and allows current to follow the path indicated by the arrows. At this time, C1 charges to the peak value of the input voltage, or 200 volts, with the polarity shown.

During the period when the input cycle is negative, as shown in view B, the polarity across the secondary of the transformer is reversed. Note specifically that the top of the secondary winding is now positive. This condition now forward biases CR2 and reverse biases CR1. A series circuit now exists consisting of C1, CR2, C2, and the secondary of the transformer. The current flow is indicated by the arrows. The secondary voltage of the transformer now aids the voltage on C1. This results in a pulsating dc voltage of 400 volts, as shown by the waveform. The effect of series aiding is comparable to the connection of two 200-volt batteries in series. As shown in figure 4-46, C2 charges to the sum of these voltages, or 400 volts.

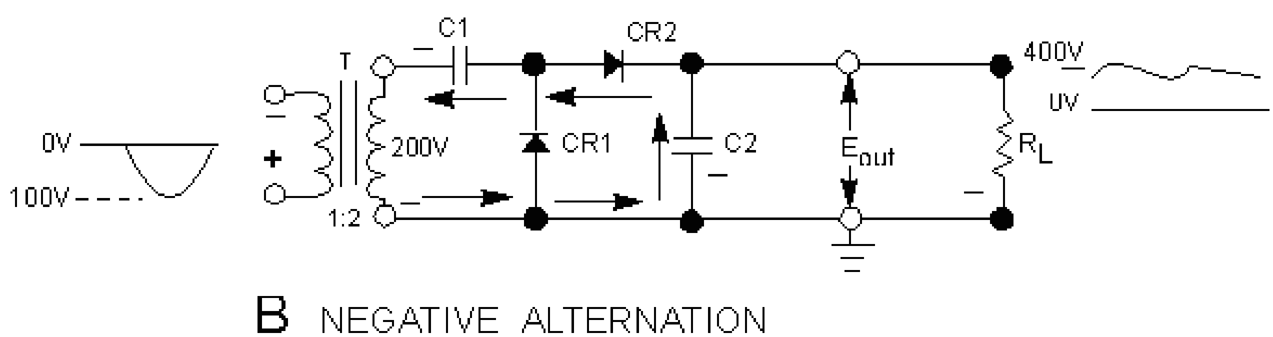


Figure 4-45B.—Rectifier action of CR1 and CR2. NEGATIVE ALTERNATION

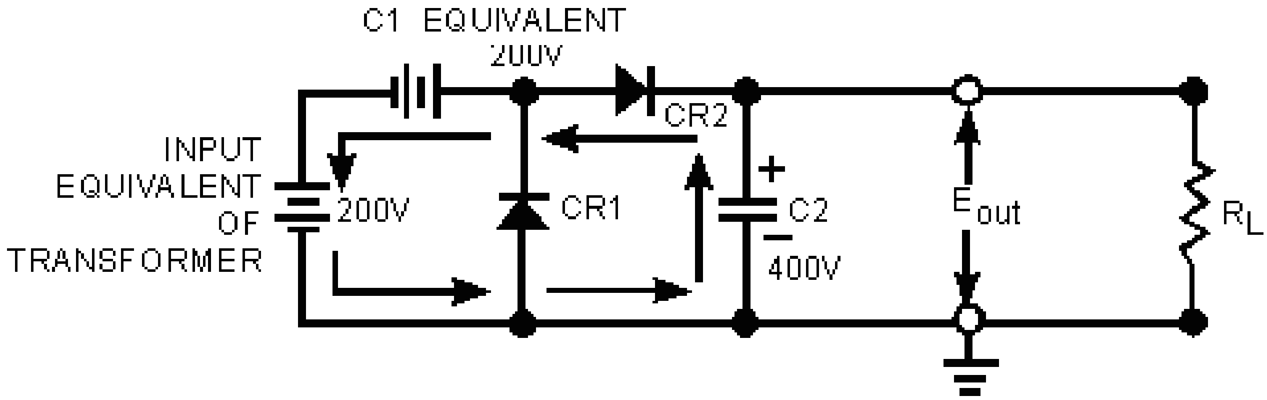


Figure 4-46.—Series-aiding sources.

The schematic shown in figure 4-47 is an illustration of a half-wave voltage tripler. When you compare figures 4-46 and 4-47, you should see that the circuitry is identical except for the additional parts, components, and circuitry shown by the dotted lines. (CR3, C3, and R2 make up the additional circuitry.) By themselves, CR3, C3, and R2 make up a half-wave rectifier. Of course, if you remove the added circuitry, you will once again have a half-wave voltage doubler.

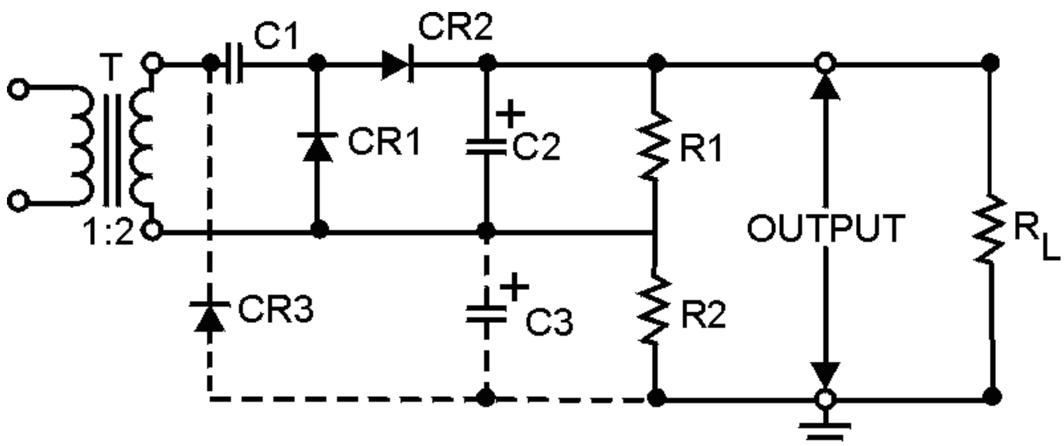


Figure 4-47.—Half-wave voltage tripler.

View A of figure 4-48 shows the schematic for the voltage tripler. Notice that CR3 is forward biased and functions like a closed switch. This allows C3 to charge to a peak voltage of 200 volts at the same time C1 is also charging to 200 volts.

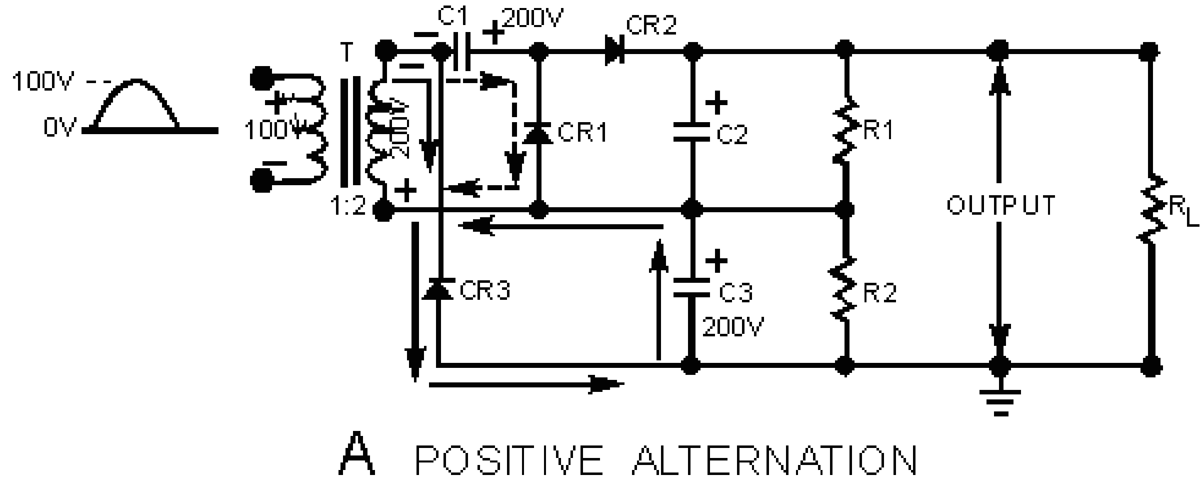


Figure 4-48A.—Voltage tripler. POSITIVE ALTERNATION

The other half of the input cycle is shown in view B. C2 is charged to twice the input voltage, or 400 volts, as a result of the voltage-doubling action of the transformer and C1. At this time, C2 and C3 are used as series-aiding devices, and the output voltage increases to the sum of their respective voltages, or 600 volts. R1 and R2 are proportional according to the voltages across C2 and C3. In this case, there is a 2 to 1 ratio.

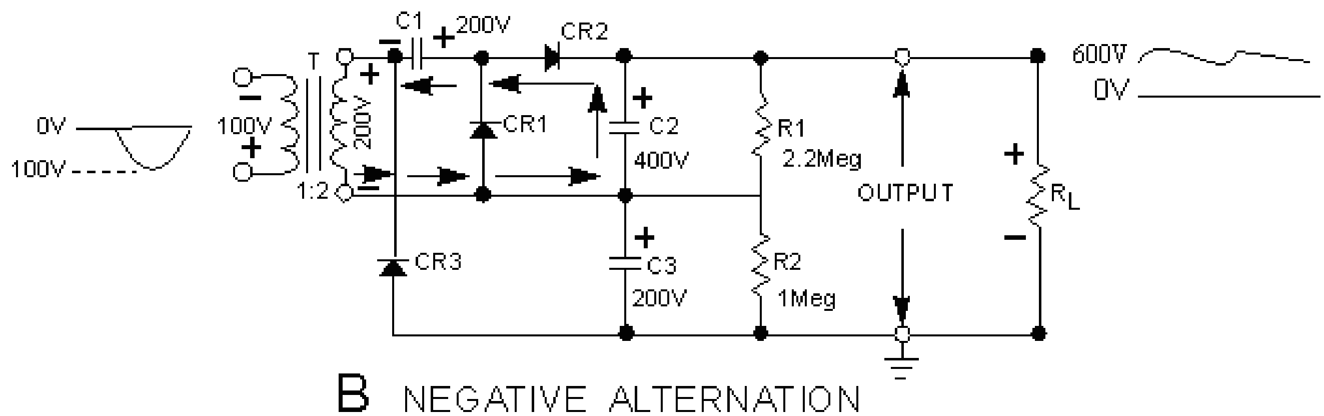


Figure 4-48B.—Voltage tripler. NEGATIVE ALTERNATION

The circuit shown in figure 4-49 is that of a full-wave voltage doubler. The main advantage of a full-wave doubler over a half-wave doubler is better voltage regulation, as a result of reduction in the output ripple amplitude and an increase in the ripple frequency. The circuit is, in fact, two half-wave rectifiers. These rectifiers function as series-aiding devices except in a slightly different way. During the alternation when the secondary of the transformer is positive at the top, C1 charges to 200 volts through CR1. Then, when the transformer secondary is negative at the top, C2 charges to 200 volts through CR2. R1 and R2 are equal value, balancing resistors that stabilize the charges of the two capacitors. Resistive load R_L is connected across C1 and C2 so that R_L receives the total charge of both capacitors. The output voltage is +400 volts when measured at the top of R_L, or point "A" with respect to point "B." If the output is measured at the bottom of R_L, it is –400 volts. Either way, the output is twice the peak value of the ac secondary voltage. As you can imagine, the possibilities for voltage multiplication are extensive.

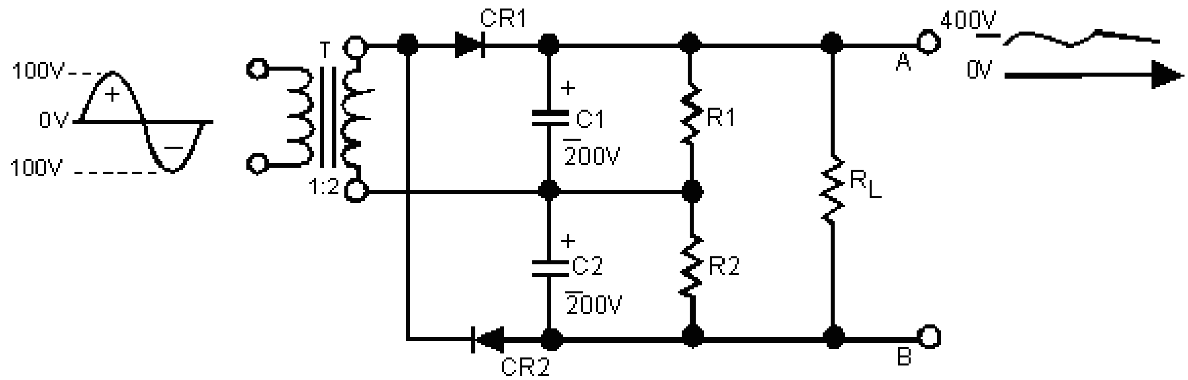


Figure 4-49.—Full-wave voltage doubler.

Q38. A half-wave voltage doubler is made up of how many half-wave rectifiers?

Q39. If a half-wave rectifier is added to a half-wave voltage doubler, the resulting circuit is a voltage ____.

Q40. In a full-wave voltage doubler, are the capacitors connected in series or in parallel with the output load?

Short Circuit Protection

The main disadvantage of a series regulator is that the pass transistor is in series with the load. If a short develops in the load, a large amount of current will flow in the regulator circuit. The pass transistor can be damaged by this excessive current flow. You could place a fuse in the circuit, but in many cases, the transistor will be damaged before the fuse blows. The best way to protect this circuit is to limit the current automatically to a safe value. A series regulator with a current-limiting circuit is shown in figure 4-50. You should recall that in order for a silicon NPN transistor to conduct, the base must be between 0.6 volt to 0.7 volt more positive than the emitter. Resistor R4 will develop a voltage drop of 0.6 volt when the load current reaches 600 milliamperes. This is illustrated using Ohm's law:

$$I = \frac{E}{R} = \frac{0.6 \ volt}{1 \ ohm} = .6 \text{ ampere or } 600 \text{ milliampere}$$

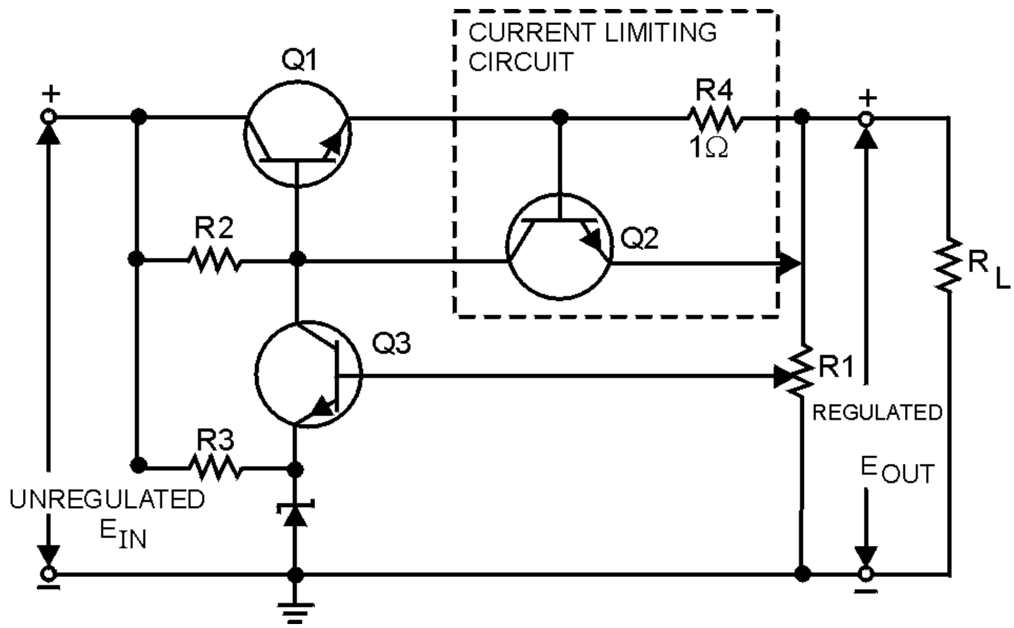


Figure 4-50.—Series regulator with current limiting.

When load current is below 600 milliamperes, the base-to-emitter voltage on Q2 is not high enough to allow Q2 to conduct. With Q2 cut off, the circuit acts like a series regulator.

When the load current increases above 600 milliamperes, the voltage drop across R4 increases to more than 0.6 volt. This causes Q2 to conduct through resistor R2, thereby decreasing the voltage on the base of pass transistor Q1. This action causes Q1 to conduct less. Therefore, the current cannot increase above 600 to 700 milliamperes.

By increasing the value of R4, you can limit the current to almost any value. For example, a 100-ohm resistor develops a voltage drop of 0.6 volt at 6 milliamperes of current. You may encounter current-limiting circuits that are more sophisticated, but the theory of operation is always the same. If you understand this circuit, you should have no problem with the others.

TROUBLESHOOTING POWER SUPPLIES

Whenever you are working with electricity, the proper use of safety precautions is of the utmost importance to remember. In the front of all electronic technical manuals, you will always find a section on safety precautions. Also posted on each piece of equipment should be a sign listing the specific precautions for that equipment. One area that is sometimes overlooked, and is a hazard especially on board ship, is the method in which equipment is grounded. By grounding the return side of the power transformer to the metal chassis, the load being supplied by the power supply can be wired directly to the metal chassis. Thereby the necessity of wiring directly to the return side of the transformer is eliminated. This method saves wire and reduces the cost of building the equipment, and while it solves one of the problems of the manufacturer, it creates a problem for you, the technician. Unless the chassis is physically grounded to the ship's ground (the hull), the chassis can be charged (or can float) several hundred volts above ship's ground. If you come in contact with the metal chassis at the same time you are in contact with the ship's hull, the current from the chassis can use your body as a low resistance path back to the ship's ac generators. At best this can be an unpleasant experience; at worst it can be fatal. For this reason Navy electronic equipment is always grounded to the ship's hull, and approved rubber mats are required in all spaces where electronic equipment is present. Therefore, before starting to work on any electronic or electrical equipment, ALWAYS ENSURE THAT THE EQUIPMENT AND ANY TEST EQUIPMENT YOU ARE USING IS PROPERLY GROUNDED AND THAT THE RUBBER MAT YOU ARE STANDING ON IS IN GOOD CONDITION. As long as you follow these simple rules, you should be able to avoid the possibility of becoming an electrical conductor.

TESTING

There are two widely used checks in testing electronic equipment, VISUAL and SIGNAL TRACING. The importance of the visual check should not be underestimated because many technicians find defects right away simply by looking for them. A visual check does not take long. In fact, you should be able to see the problem readily if it is the type of problem that can be seen. You should learn the following procedure. You could find yourself using it quite often. This procedure is not only for power supplies but also for any type of electronic equipment you may be troubleshooting. (Because diode and transistor testing was covered in chapter 1 and 2 of this module, it will not be discussed at this time. If you have problems in this area, refer to chapter 1 for diodes or chapter 2 for transistors.)

1. BEFORE YOU ENERGIZE THE EQUIPMENT, LOOK FOR:

   a. SHORTS—Any terminal or connection that is close to the chassis or to any other terminal should be examined for the possibility of a short. A short in any part of the power supply can cause considerable damage. Look for and remove any stray drops of solder, bits of wire, nuts, or screws. It sometimes helps to shake the chassis and listen for any tell-tale rattles. Remember to correct any problem that may cause a short circuit; if it is not causing trouble now, it may cause problems in the future.

   b. DISCOLORED OR LEAKING TRANSFORMER—This is a sure sign that there is a short somewhere. Locate it. If the equipment has a fuse, find out why the fuse did not blow; too large a size may have been installed, or there may be a short across the fuse holder.

   c. LOOSE, BROKEN, OR CORRODED CONNECTION—Any connection that is not in good condition is a trouble spot. If it is not causing trouble now, it will probably cause problems in the future. Fix it.

   d. DAMAGED RESISTORS OR CAPACITORS—A resistor that is discolored or charred has been subjected to an overload. An electrolytic capacitor will show a whitish deposit at the seal around the terminals. Check for a short whenever you notice a damaged resistor or a damaged capacitor. If there is no short, the trouble may be that the power supply has been overloaded in some way. Make a note to replace the part after signal tracing. There is no sense in risking a new part until the trouble has been located.

2. ENERGIZE THE EQUIPMENT AND LOOK FOR:

   a. SMOKING PARTS—If any part smokes or if you hear any boiling or sputtering sounds, remove the power immediately. There is a short circuit somewhere that you have missed in your first inspection. Use any ohmmeter to check the part once again. Start in the neighborhood of the smoking part.

   b. SPARKING—Tap or shake the chassis. If you see or hear sparking, you have located a loose connection or a short. Check and repair.

If you locate and repair any of the defects listed under the visual check, make a note of what you find and what you do to correct it. It is quite probable you have found the trouble. However, a good technician takes nothing for granted. You must prove to yourself that the equipment is operating properly and that no other troubles exist.

If you find none of the defects listed under the visual check, go ahead with the signal tracing procedure. The trouble is probably of such a nature that it cannot be seen directly-it may only be seen using an oscilloscope.

Tracing the ac signal through the equipment is the most rapid and accurate method of locating a trouble that cannot be found by a visual check, and it also serves as check on any repairs you may have made. The idea is to trace the ac voltage from the transformer, to see it change to pulsating dc at the rectifier output, and then see the pulsations smoothed out by the filter. The point where the signal stops or becomes distorted is the place look for the trouble. If you have no dc output voltage, you should look for an open or a short in your signal tracing. If you have a low dc voltage, you should look for a defective part and keep your eyes open for the place where the signal becomes distorted.

Signal tracing is one method used to localize trouble in a circuit. This is done by observing the waveform at the input and output of each part of a circuit.

Let's review what each part of a good power supply does to a signal, as shown in figure 4-51. The ac voltage is brought in from the power line by means of the line cord. This voltage is connected to the primary of the transformer through the ON-OFF switch (S1). At the secondary winding of the transformer (points 1 and 2), the scope shows you a picture of the stepped-up voltage developed across each half of the secondary winding-the picture is that of a complete sine wave. Each of the two stepped-up voltages is connected between ground and one of the two anodes of the rectifier diodes. At the two rectifier anodes (points 4 and 5), there is still no change in the shape of the stepped-up voltage-the scope picture still shows a complete sine wave.

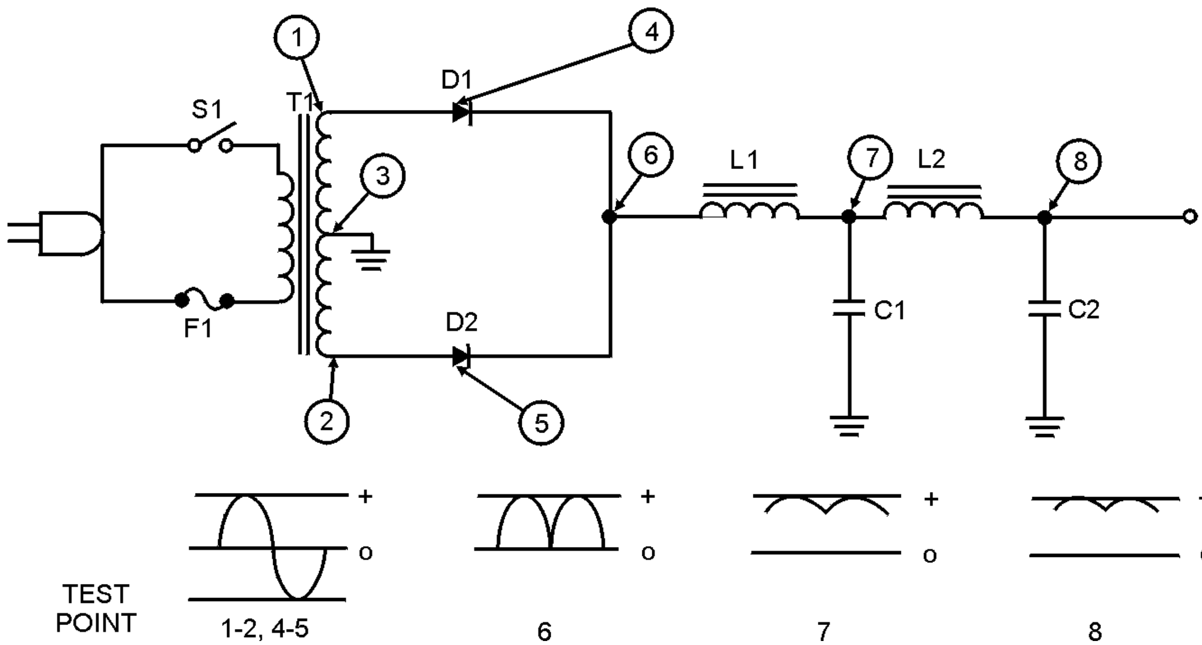


Figure 4-51.—Complete power supply (without regulator).

However, when you look at the scope pattern for point 6 (the voltage at the rectifier cathodes), you see the waveshape for pulsating direct current. This pulsating dc is fed through the first choke (L1) and filter capacitor (C1) which remove a large part of the ripple, or "hum," as shown by the waveform for point 7. Finally the dc voltage is fed through the second choke (L2) and filter capacitor (C2), which remove nearly all of the remaining ripple. (See the waveform for point 8, which shows almost no visible ripple.) You now have almost pure dc.

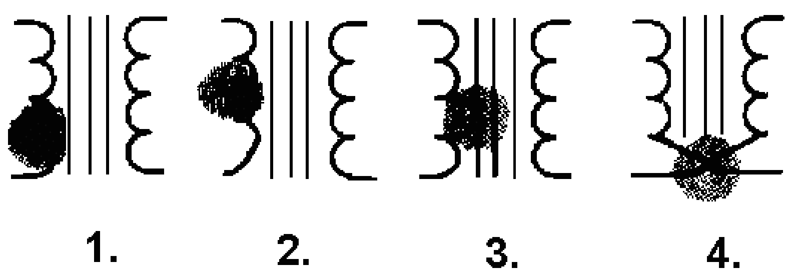
No matter what power supplies you use in the future, they all do the same thing—they change ac voltage into dc voltage.

Component Problems

The following paragraphs will give you an indication of troubles that occur with many different electronic circuit components.

TRANSFORMER AND CHOKE TROUBLES.—As you should know by now, the transformer and the choke are quite similar in construction. Likewise, the basic troubles that they may develop are comparable.

1. A winding can open.
2. Two or more turns of one winding can short together.
3. A winding can short to the casing, which is usually grounded.
4. Two windings (primary and secondary) can short together. This trouble is possible, of course, only in transformers.

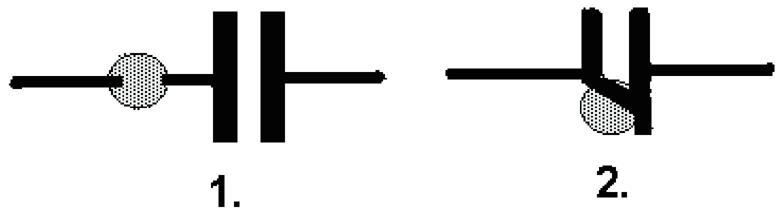


When you have decided which of these four possible troubles could be causing the symptoms, you have definite steps to take. If you surmise that there is an open winding, or windings shorted together or to ground, an ohmmeter continuity check will locate the trouble. If the turns of a winding are shorted together, you may not be able to detect a difference in winding resistance. Therefore, you need to connect a good transformer in the place of the old one and see if the symptoms are eliminated. Keep in mind that transformers are difficult to replace. Make absolutely sure that the trouble is not elsewhere in the circuit before you change the transformer.

Occasionally, the shorts will only appear when the operating voltages are applied to the transformer. In this case you might find the trouble with a megger-an instrument which applies a high voltage as it reads resistance.

CAPACITOR AND RESISTOR TROUBLES.—Just two things can happen to a capacitor:

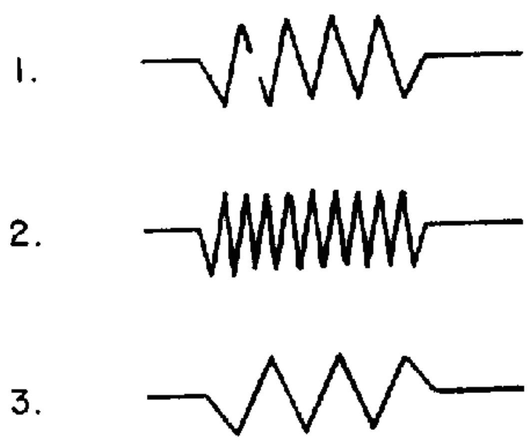
1. It may open up, removing the capacitor completely from the circuit.
2. It may develop an internal short circuit. This means that it begins to pass current as though it were a resistor or a direct short.



You may check a capacitor suspected of being open by disconnecting it from the circuit and checking it with a capacitor analyzer. You can check a capacitor suspected of being leaky with an ohmmeter; if it reads less than 500 kilohms, it is more than likely bad. However, capacitor troubles are difficult to find since they may appear intermittently or only under operating voltages. Therefore, the best check for a faulty capacitor is to replace it with one known to be good. If this restores proper operation, the fault was in the capacitor.

Resistor troubles are the simplest. However, like the others, they must be considered.

1. A resistor can open.
2. A resistor can increase in value.
3. A resistor can decrease in value.



You already know how to check possible resistor troubles. Just use an ohmmeter after making sure no parallel circuit is connected across the resistor you wish to measure. When you know a parallel circuit is connected across the resistor or when you are in doubt disconnect one end of the resistor before measuring it. The ohmmeter check will usually be adequate. However, never forget that occasionally intermittent troubles may develop in resistors as well as in any other electronic parts.

Although you may observe problems that have not been covered specifically in this chapter, you should have gained enough knowledge to localize and repair any problem that may occur.

Q41. What is the most important thing to remember when troubleshooting?

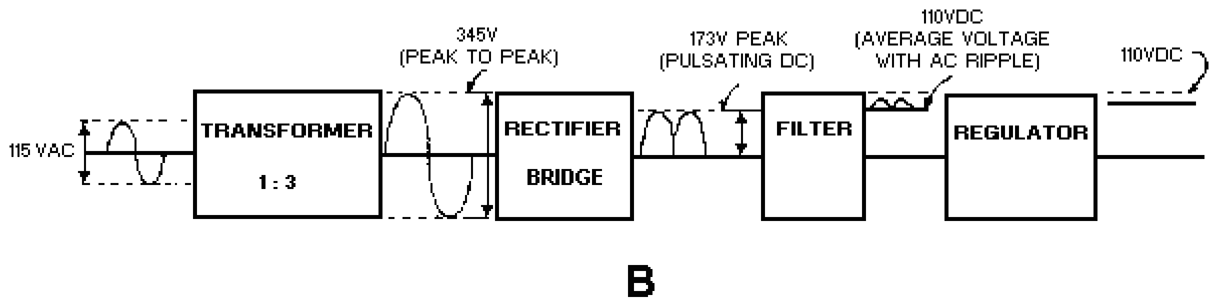
Q42. What is the main reason for grounding the return side of the transformer to the chassis?

Q43. What are two types of checks used in troubleshooting power supplies?

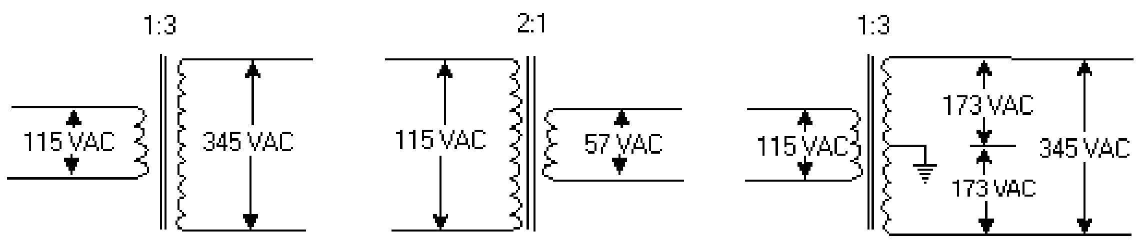
SUMMARY

This chapter has presented you with a basic description of the theory and operation of a basic power supply and its components. The following summary is provided to enhance your understanding of power supplies.

POWER SUPPLIES are electronic circuits designed to convert ac to dc at any desired level. Almost all power supplies are composed of four sections: transformer, rectifier, filter, and regulator.

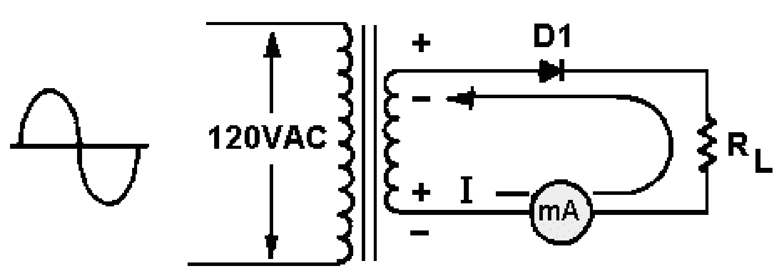


The **POWER TRANSFORMER** is the input transformer for the power supply.

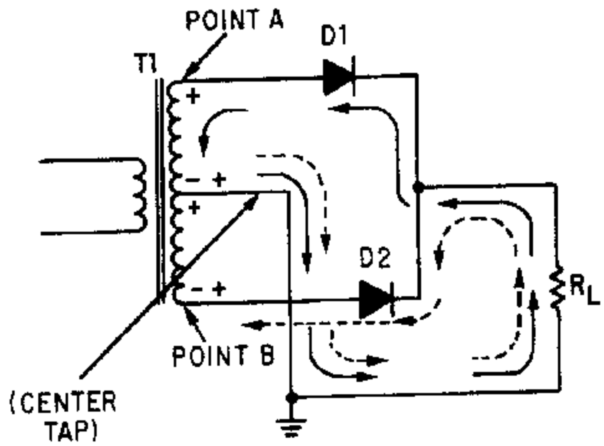


The **RECTIFIER** is the section of the power supply that contains the secondary windings of the power transformer and the rectifier circuit. The rectifier uses the ability of a diode to conduct during one half cycle of ac to convert ac to dc.

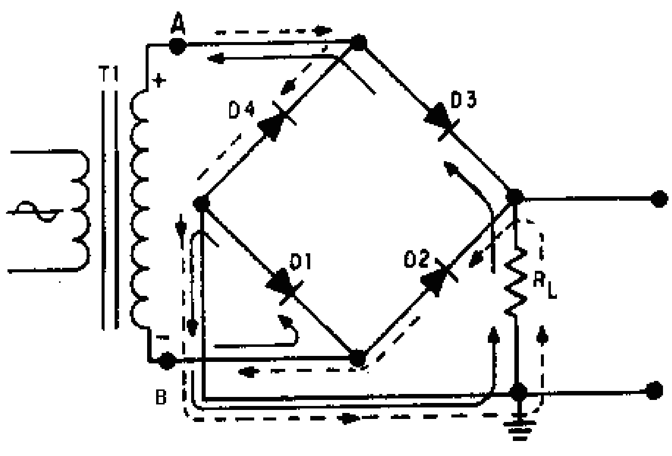
HALF-WAVE RECTIFIERS give an output on only one half cycle of the input ac. For this reason, the pulses of dc are separated by a period of one half cycle of zero potential voltage.



FULL-WAVE RECTIFIERS conduct on both halves of the input ac cycles. As a result, the dc pulses are not separated from each other. A characteristic of full-wave rectifiers is the use of a center-tapped, high-voltage secondary. Because of the center tap, the output of the rectifier is limited to one-half of the input voltage of the high-voltage secondary.

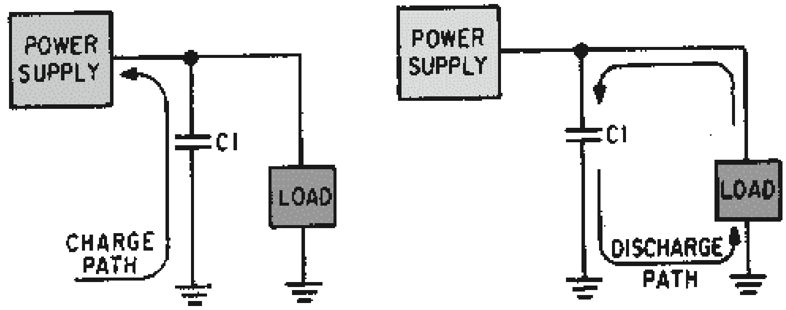


BRIDGE RECTIFIERS are full-wave rectifiers that do not use a center-tapped, high-voltage secondary. Because of this, their dc output voltage is equal to the input voltage from the high-voltage secondary of the power transformer. Bridge rectifiers use four diodes connected in a bridge network. Diodes conduct in diagonal pairs to give a full-wave pulsating dc output.

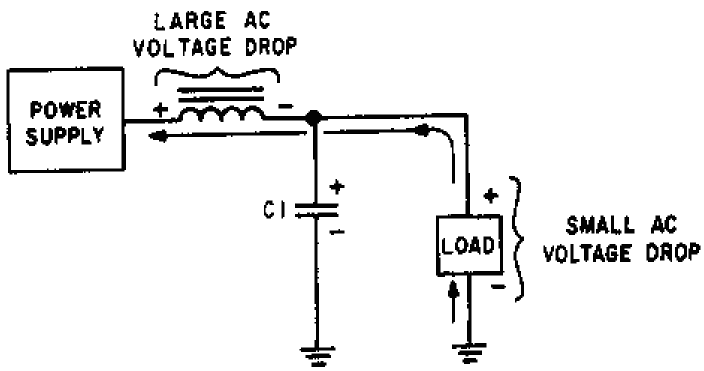


FILTER CIRCUITS are designed to smooth, or filter, the ripple voltage present on the pulsating dc output of the rectifier. This is done by an electrical device that has the ability to store energy and to release the stored energy.

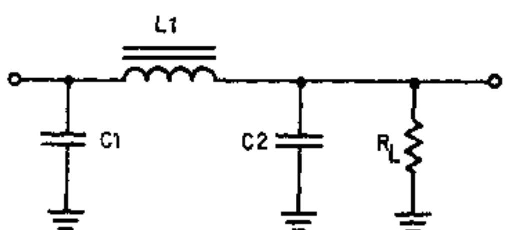
CAPACITANCE FILTERS are nothing more than large capacitors placed across the output of the rectifier section. Because of the large size of the capacitors, fast charge paths, and slow discharge paths, the capacitor will charge to average value, which will keep the pulsating dc output from reaching zero volts.

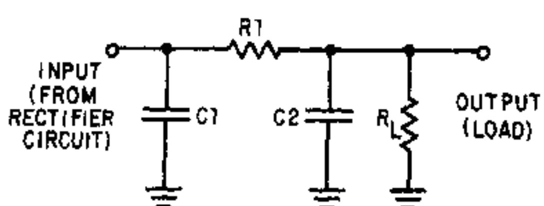


INDUCTOR FILTERS use an inductor called a choke to filter the pulsating dc input. Because of the impedance offered to circuit current, the output of the filter is at a lower amplitude than the input.



PI-TYPE FILTERS use both capacitive and inductive filters connected in a pi-type configuration. By combining filtering devices, the ability of the pi filter to remove ripple voltage is superior to that of either the capacitance or inductance filter.

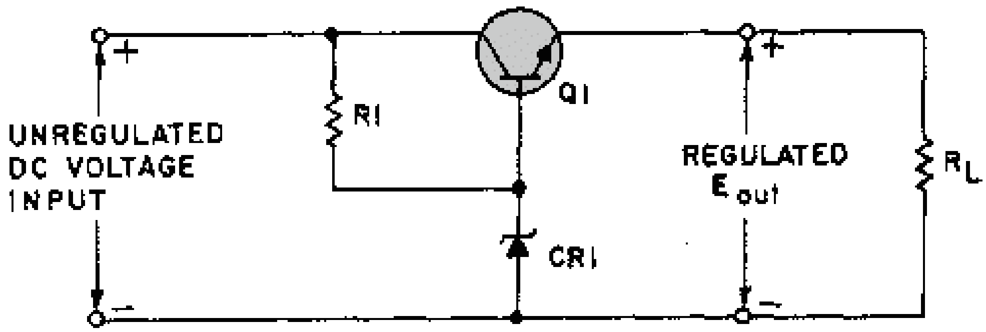




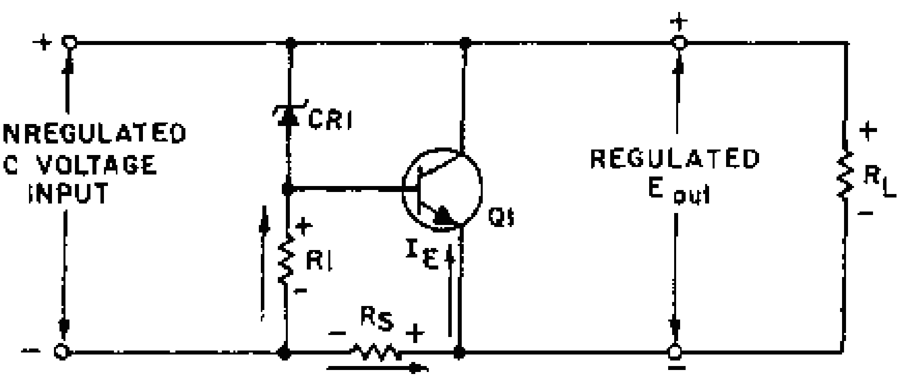
VOLTAGE REGULATORS are circuits designed to maintain the output of power supplies at a constant amplitude despite variations of the ac source voltage or changes of the resistance of the load.

This is done by creating a voltage divider of a resistive element in the regulator and the resistance of the load. Regulation is achieved by varying the resistance of the resistive element in the regulator.

A **SERIES REGULATOR** uses a variable resistance in series with the load. Regulation is achieved by varying this resistance either to increase or to decrease the voltage drop across the resistive element of the regulator. Characteristically, the resistance of the variable resistance moves in the same direction as the load. When the resistance of the load increases, the variable resistance of the regulator increases; when load resistance decreases, the variable resistance of the regulator decreases.



SHUNT REGULATORS use a variable resistance placed in parallel with the load. Regulation is achieved by keeping the resistance of the load constant. Characteristically the resistance of the shunt moves in the opposite direction of the resistance of the load.



The **CURRENT LIMITER** is a short-circuit protection device that automatically limits the current to a safe value. This is done when the current-limiting transistor senses an increase in load current. At this time the current-limiting transistor decreases the voltage on the base of the pass transistor in the regulator, causing a decrease in its conduction. Therefore, current cannot rise above a safe value.

TROUBLESHOOTING is a method of detecting and repairing problems in electronic equipment. Two methods commonly used are the **VISUAL CHECK** and **SIGNAL TRACING**. The visual check allows the technician to make a quick check of component problems, such as shorts, discolored or leaky transformers, loose or broken connections, damaged resistors or capacitors, smoking parts, or sparking. The signal tracing method is used when the technician cannot readily see the problem and needs to use test equipment. Component failure is also important in troubleshooting. In transformers and chokes, a winding can open, or two or more windings can short, either to themselves or to the case that is usually grounded. In a capacitor only two things can occur: either it can short and act as a resistor, or it can open, removing it from the circuit. A resistor can open, increase in value, or decrease in value.

ANSWERS TO QUESTIONS Q1. THROUGH Q43.

A1. *Transformer, rectifier, filter, regulator.*

A2. *To change ac to pulsating dc.*

A3. *To change pulsating dc to pure dc.*

A4. *To maintain a constant voltage to the load.*

A5. *The half-wave rectifier.*

A6. *15.9 volts.*

A7. *It isolates the chassis from the power line.*

A8. *The fact that the full-wave rectifier uses the full output, both half cycles, of the transformer.*

A9. *120 hertz.*

A10. *63.7 volts.*

A11. *Peak voltage is half that of the half-wave rectifier.*

A12. *The bridge rectifier can produce twice the voltage with the same size transformer.*

A13. *It will decrease. Capacitance is inversely proportional to:*

$$X_C \left(X_C = \frac{1}{2\pi fC}\right).$$

A14. *The capacitor filter.*

A15. *Parallel.*

A16. *At a high frequency.*

A17. *A filter circuit increases the average output voltage.*

A18. *Value of capacitance and load resistance.*

A19. *Good.*

A20. *Yes.*

A21. *The CEMF of the inductor.*

A22. *From 1 to 20 henries.*

A23. *Decrease.*

A24. *Expense.*

A25. *When ripple must be held at an absolute minimum.*

A26. *LC capacitor-input filter.*

A27. Cost and size of the inductor.

A28. Regulators.

A29. Variation.

A30. Series and shunt.

A31. An increase.

A32. In parallel.

A33. Bias.

A34. Increases.

A35. Increases.

A36. Decreases.

A37. An increase.

A38. Two.

A39. Trippler.

A40. In parallel.

A41. Safety precautions.

A42. To eliminate shock hazard.

A43. Visual and signal tracing.

APPENDIX I

GLOSSARY

ACCEPTOR IMPURITY—An impurity which, when added to a semiconductor, accepts one electron from a neighboring atom and creates a hole in the lattice structure of the crystal. Also called TRIVALENT impurities.

ALLOYED JUNCTION—A junction formed by recrystallization of a molten region of P-type material on an N-type substrate, or vice versa.

ALPHA—The emitter-to-collector current gain in a common-base circuit.

AMPLIFICATION—The ratio of output magnitude to input magnitude in a device that is intended to produce an output that is an enlarged reproduction of its input.

AMPLIFICATION FACTOR—The voltage of an amplifier with no load on the output.

AMPLIFIER—The device that provides amplification (the increase in current, voltage, or power of a signal) without appreciably altering the original signal.

AMPLITUDE DISTORTION—Distortion that is present in an amplifier when the amplitude of the output signal fails to follow exactly any increase or decrease in the amplitude of the input signal.

ANODE—A positive electrode of an electrochemical device (such as a primary or secondary electric cell) toward which the negative ions are drawn.

AVALANCHE EFFECT—A reverse breakdown effect in diodes that occurs at reverse voltage beyond 5 volts. The released electrons are accelerated by the electric field, which results in a release of more electrons in a chain or "avalanche" effect.

BASE—The element in a transistor that controls the flow of current carriers.

BETA—The ratio of a change in collector current to a corresponding change in base current, when the collector voltage is constant in a common-emitter circuit.

BREAKDOWN—The phenomenon occurring in a reverse-biased semiconductor diode. The start of the phenomenon is observed as a transition from a high dynamic resistance to one of substantially lower dynamic resistance. This is done to boost the reverse current.

CAPACITOR FILTER—This filter is used on extremely high-voltage, low current power supplies and also where the ripple frequency is not critical.

CATHODE—The negative terminal of a forward-biased semiconductor diode that is the source of the electrons.

CHOKE—An inductor used to impede the flow of pulsating dc or ac by means of self-inductance.

CLASS A AMPLIFIER OPERATION—The amplifier is biased so that variations in input signal polarities occur within the limits of cutoff and saturation.

CLASS AB AMPLIFIER OPERATION—The amplifier is biased so that collector current is cut off for a portion of the alternation of the input signal.

CLASS B AMPLIFIER OPERATION—The amplifier is biased so that collector current is cut off for one-half of the input signal.

CLASS C AMPLIFIER OPERATION—The amplifier is biased so that collector current is cut off for more than one-half of the input signal.

COLLECTOR—The element in a transistor which collects the current carriers.

COMMON BASE—A transistor circuit in which the base electrode is the common element to both input and output circuits.

COMMON COLLECTOR—A transistor circuit configuration in which the collector is the common element to the input circuit and to the output circuit.

COMMON EMITTER—Circuit configuration in which the emitter is the element common to both the input and the output circuit.

CONDUCTION BAND—A partially filled energy band in which electrons can move freely.

COVALENT BOND—A type of linkage between atoms.

CURRENT REGULATOR—A circuit that provides a constant current output.

DEGENERATION—The process whereby a part of the output signal of an amplifying device is returned to its input circuit in such a manner that it tends to cancel part of the input.

DEPLETION REGION—The region in a semiconductor where essentially all free electrons and holes have been swept out by the electrostatic field that exists there.

DIODE—A two element solid-state device made of either germanium or silicon. It is primarily used as a switching device.

DONOR—An impurity that can make a semiconductor material an N-type by donating extra "free" electrons to the conduction band.

DONOR IMPURITY—See PENTAVALENT IMPURITY.

DOPING—The process of adding impurities to semiconductor crystals to increase the number of free charges that can be moved by an external, applied voltage. Doping produces an N-type or P-type material.

DUAL-GATE MOSFET—A two-gate MOSFET in which either gate can control the conductor independently, a fact which makes this MOSFET very versatile.

EFFICIENCY—The ratio of output-signal power compared to the total input power.

EMITTER—The element in a transistor that emits current carriers (electrons or holes).

EXTRINSIC—A semiconductor in which impurities have been added to create certain charge carrier concentrations.

FIELD-EFFECT TRANSISTOR (FET)—A transistor consisting of a source, a gate, and a drain. Current flow is controlled by the transverse electric field under the gate.

FIDELITY-—The faithful reproduction of a signal. The accuracy with which a system reproduces a signal at its output that faithfully maintains the essential characteristics of the input signal.

FIXED BIAS—A constant value of bias voltage.

FORBIDDEN BAND—The energy band in an atom lying between the conduction band and the valence band. Electrons are never found in the forbidden band but may travel back and forth through it. The forbidden band determines whether a solid material will act as a conductor, a semiconductor, or an insulator.

FORWARD BIAS—An external voltage that is applied to a PN junction in the conducting direction so that the junction offers only minimum resistance to the flow of current. Conduction is by majority current carriers (holes in P-type material; electrons in N-type material).

FREE CHARGES—Those electrons that can be moved by an externally applied voltage.

FULL-WAVE RECTIFIER—A circuit that uses both positive and negative alternations in an alternating current to produce direct current.

FULL-WAVE VOLTAGE DOUBLER—Consists of two full-wave voltage rectifiers and is used to reduce the output ripple frequency.

FUSED-ALLOY JUNCTION—See ALLOYED-JUNCTION.

GALENA—A crystalline form of lead sulfide used in early radio receivers.

GAMMA—The emitter-to-base current ratio in a common-collector configuration.

GERMANIUM—A grayish-white metal having semiconductor properties.

GROWN JUNCTION—A method of mixing P-type and N-type impurities into a single crystal while the crystal is being grown.

HALF-WAVE RECTIFIER—A rectifier using only one-half of each cycle to change ac to pulsating dc.

HALF-WAVE VOLTAGE DOUBLER—Consists of two half-wave voltage rectifiers.

HOLE FLOW—In the valence band, a process of conduction in which electrons move into holes, thereby creating other holes that appear to move toward a negative potential. (The movement of holes is opposite the movement of electrons.)

HYBRID CIRCUIT—A circuit where passive components (resistors, capacitors) are deposited onto a substrate made of glass, ceramic, or other insulating material. Then the active components (diodes, transistors) are attached to the substrate and connected to the passive components on the substrate with a very fine wire.

IGFET—Any field-effect transistor that has an insulated gate.

INDUCED CHANNEL MOSFET—A MOSFET in which there is no actual channel between the source and the drain. This MOSFET is constructed by making the channel the same type of material as the substrate.

INDUCTANCE—The properties of a circuit that tend to oppose any change in current flow.

INTEGRATED CIRCUIT—A circuit in which many elements are fabricated and interconnected by a single process (into a single chip), as opposed to a "nonintegrated" circuit in which the transistors, diodes, resistors, and other components are fabricated separately and then assembled.

JUNCTION DIODE—A two-terminal device containing a single crystal of semiconducting material, which ranges from P-type at one terminal to N-type at the other.

JUNCTION TRANSISTOR—A bipolar transistor constructed from interacting PN junctions. The term is used to distinguish junction transistors from other types such as field-effect and point-contact.

LC CAPACITOR-INPUT FILTER—This is the most common type of filter. It is used in a power supply where output current is low and load current is relatively constant.

LC CHOKE-INPUT FILTER—This filter is used in power supplies where voltage regulation is important and where the output current is relatively high and subject to varying load conditions.

LIGHT-EMITTING DIODE (LED)—A PN junction diode that emits visible light when it is forward biased. Depending on the material used to make the diode, the light may be red, green or amber.

LINEAR—Having an output that varies in direct proportion to the input.

MAJORITY CARRIERS—The mobile charge carriers (holes or electrons) that are predominate in a semiconductor material; for example, electrons in an N-type region.

METAL-OXIDE SEMICONDUCTOR FIELD-EFFECT TRANSISTOR—See MOSFET.

METALLIC RECTIFIER—Also known as a DRY-DISC RECTIFIER. A metal to semiconductor large-area contact device in which a semiconductor is sandwiched between two metal plates. This asymmetrical construction permits current to flow more readily in one direction than the other.

MICROELECTRONICS—The solid-state concept of electronics in which compact semiconductor materials are designed to function as an entire circuit or subassembly rather than as circuit components.

MINORITY CARRIERS—Either electrons or holes, whichever is the less dominant carrier in a semiconductor device. In P-type semiconductors, electrons are the minority carriers; in N-type semiconductors, the holes are the minority carriers.

MINORITY CURRENT—A very small current that passes through the base-to-collector junction when this junction is reverse biased.

MODULAR CIRCUITRY—A technique where printed circuit boards are stacked and connected together to form a module.

MONOLITHIC CIRCUIT—A circuit where all elements (resistors, transistors, etc.) associated with the circuit are fabricated inseparably within a continuous piece of material (called the substrate), usually silicon.

METAL-OXIDE SEMICONDUCTOR FIELD-EFFECT TRANSISTOR—See MOSFET.

MOSFET—A semiconductor device that contains diffused source and drain regions on either side of a P- or N-channel area. Also contains a gate insulated from the channel area by silicon-oxide. Operates in either the depletion or the enhancement mode.

Mu—English spelling for the Greek letter μ.

NEGATIVE TEMPERATURE COEFFICIENT—A characteristic of a semiconductor material, such as silver sulfide, in which resistance to electrical current flow decreases as temperature increases.

NONLINEAR—Having an output that does not rise or fall directly with the input.

NPN—An NPN transistor is formed by introducing a thin region of P-type material between two regions of N-type material.

OPTICAL COUPLER—A coupler composed of an LED and a photodiode and contained in a light-conducting medium. Suitable for frequencies in the low-megahertz range.

OPTOELECTRONIC DEVICES—Devices that either produce or use light in their operation.

OVERDRIVEN—When the input signal amplitude is increased to the point that the transistor goes into saturation and cutoff.

PENTAVALENT IMPURITY—A type of impurity which contains five valence electrons and donates one electron to the doped material. Also called DONOR IMPURITY.

PHOTOCELL—A light-controlled variable resistor that has a light-to-dark resistance ratio of 1:1000. Used in various types of control and timing circuits.

PHOTODIODE—A light-controlled variable resistor. Current flow increases when the PN junction is exposed to an external light source

PHOTOTRANSISTOR—An optoelectronic device that conducts current when exposed to light. Produces more current and is much more sensitive to light than the photodiode.

PHOTOVOLTAIC CELL (SOLAR CELL)—A device that acts much like a battery when exposed to light and converts light energy into electrical energy.

POINT-CONTACT TRANSISTOR—A semiconductor diode that can work with and amplify the ultrahigh frequencies used in radar.

POSITIVE TEMPERATURE COEFFICIENT—The characteristic of a conductor in which the resistance increases as temperature increases.

POWER SUPPLY—A unit that supplies electrical power to another unit. It changes ac to dc and maintains a constant voltage output within limits.

PRINTED CIRCUIT BOARD—A flat insulating surface upon which printed wiring and miniaturized components are connected in a predetermined design and attached to a common base.

QUANTUM-MECHANICAL TUNNELING—When an electron is able to cross a PN junction because of tunnel effect.

QUIESCENCE—The operating condition of a circuit when no input signal is being applied to the circuit.

RC FILTER—This filter is used in applications where load current is low and constant, and voltage regulation is not necessary.

RECTIFIER—A device which, by its conduction characteristics, converts alternating current to a pulsating direct current.

REGULATOR—The section in a basic power supply that maintains the output of the power supply at a constant level in spite of large changes in load current or in input line voltage.

REVERSE BIAS—When an external voltage is applied to a PN junction and the junction offers a high resistance to current flow.

RIPPLE FREQUENCY—The frequency of the ripple current. In a full-wave rectifier, it is twice the input-line frequency.

RIPPLE VOLTAGE—The alternating component of unidirectional voltage. (This component is small compared to the direct component.)

SELENIUM—A chemical element which has rectification and light-sensitive properties that make it widely used as a semiconductor material.

SERIES VOLTAGE REGULATOR—A regulator with a regulating device that is in series with the load resistance.

SHUNT VOLTAGE REGULATOR—A regulator whose regulating device is in parallel with the load resistance.

SILICON—A metallic element which, in its pure state, is used as a semiconductor.

SILICON-CONTROLLED RECTIFIER (SCR)—A semiconductor device that functions as an electrically controlled switch.

SOLID-STATE DEVICE—An electronic device which operates by the movement of electrons within a solid piece of semiconductor material.

THERMAL RUNAWAY—A conduction that exists when heat causes more electron-hole pairs to be generated; which, in turn, causes more heat and may eventually cause diode destruction.

TRANSISTOR—A semiconductor device with three or more elements.

TRIAC—A three-terminal device that is similar to two SCRs back to back with a common gate and common terminals. Although similar in construction and operation to the SCR, the TRIAC controls and conducts current flow during both alternations of an ac cycle.

TRIVALENT IMPURITY—Acceptor impurities containing only three valence electrons.

TUNNEL DIODE—A heavily doped semiconductor device that has high gain and fast switching capabilities.

UNIJUNCTION TRANSISTOR (UJT)—A three-terminal, solid-state device that resembles a transistor but is stable over a wide range of temperatures and allows a reduction of components when used in place of a transistor. Used in switching circuits, oscillators, and wave-shaping circuits.

VARACTOR—A diode that behaves like a variable capacitor, with the PN junction functioning like the dielectric and the plates of a common capacitor.

VOLTAGE GAIN—Ratio of voltage across a specified load.

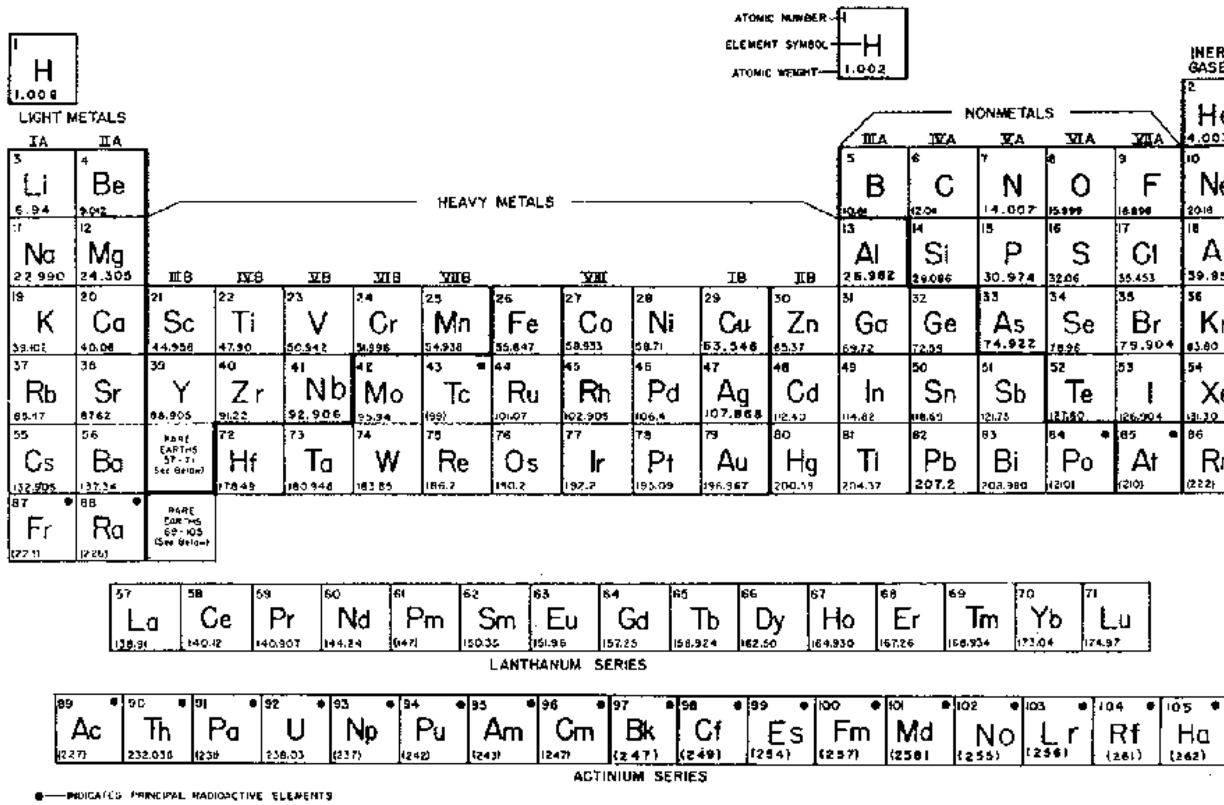
VOLTAGE MULTIPLIERS—Methods of increasing voltages used primarily where low current is required.

ZENER DIODE—A PN-junction diode designed to operate in the reverse-bias breakdown region.

ZENER EFFECT—A reverse breakdown effect in diodes in which breakdown occurs at reverse voltages below 5 volts. The presence of a high energy field at the junction of a semiconductor produces the breakdown.

APPENDIX II

PERIODIC TABLE OF THE ELEMENTS



SEE TABLE BELOW FOR INTERPRETATION OF SYMBOLS

Figure AII-1.—Periodic table of the elements.

Table AII-1.—Periodic Table of the Elements

Symbol	Name	Atomic Number	Atomic Weight
Ac	Actinium	89	1(227)
Ag	Silver	47	107.868
Al	Aluminum	13	6.982
Am	Americium	95	(243)
Ar	Argon	18	39.95
As	Arsenic	33	74.922
At	Astatine	85	(210)
Au	Gold	79	196.967
B	Boron	5	10.81
Ba	Barium	56	137.34
Be	Beryllium	4	9.012
Bi	Bismuth	83	208.980
Bk	Berkelium	97	(247)
Br	Bromine	35	79.904
C	Carbon	6	12.011
Ca	Calcium	20	40.08
Cd	Cadium	48	112.40
Ce	Cerium	58	140.12
Cf	Californium	98	(249)
Cl	Chlorine	17	35.453
Cm	Curium	96	(247)
Co	Cobalt	27	58.933
Cr	Chromium	24	51.996
Cs	Cesium	55	132.905
Cu	Copper	29	63.546
Dy	Dysprosium	66	162.50
Es	Einsteinium	99	(254)
Er	Erbium	68	167.26
Eu	Europium	63	151.96
F	Fluorine	9	18.998
Fe	Iron	26	55.847
Fm	Fermium	100	(257)
Fr	Francium	87	(223)
Ga	Gallium	31	69.72
Gd	Gadolinium	64	157.25
Ge	Germanium	32	72.59
H	Hydrogen	1	1.008
Ha	Hahnium	105	(262)
He	Helium	2	4.003
Hf	Hafnium	72	178.49
Hg	Mercury	80	200.59
Ho	Holmium	67	164.930
I	Iodine	53	126.904

Table AII-1.—Periodic Table of the Elements—Continued

Symbol	Name	Atomic Number	Atomic Weight
In	Indium	49	114.82
Ir	Iridium	77	192.2
K	Potassium	19	39.102
Kr	Krypton	36	83.80
La	Lanthanum	57	138.91
Li	Lithium	3	6.94
Lr	Lawrencium	103	(256)
Lu	Lutetium	71	174.97
Md	Mendelevium	101	(258)
Mg	Magnesium	12	24.305
Mn	Manganese	25	54.938
Mo	Molybdenum	42	95.94
N	Nitrogen	7	14.007
Na	Sodium	11	22,990
Nb	Niobium	41	92.906
Nd	Neodymium	60	144.24
Ne	Neon	10	20.18
Ni	Nickel	28	58.71
No	Nobelium	102	(255)
Np	Neptunium	93	(237)
O	Oxygen	8	15.999
Os	Osmium	76	190.2
P	Phosphorus	15	30.974
Pa	Protactinium	91	(231)
Pb	Lead	82	207.2
Pd	Palladium	46	106.4
Pm	Promethium	61	(147)
Po	Polonium	84	(210)
Pr	Praseodymium	59	140.907
Pt	Platinum	78	195.09
Pu	Plutonium	94	(242)
Ra	Radium	88	(226)
Rb	Rubidium	37	85.47
Re	Rhenium	75	186.2
Rf	Rutherfordium	104	(261)
Rh	Rhodium	45	102.905
Rn	Radon	86	(222)
Ru	Ruthenium	44	101.07
S	Sulfur	16	32.06
Sb	Antimony	51	121.75
Sc	Scandium	21	44.956
Se	Selenium	34	78.96

Table AII-1.—Periodic Table of the Elements—Continued

Symbol	Name	Atomic Number	Atomic Weight
Si	Silicon	14	28.086
Sm	Samarium	62	150.35
Sn	Tin	50	118.69
Sr	Strontium	38	87.62
Ta	Tantalum	73	180.948
Tb	Terbium	65	158.924
Tc	Technetium	43	(99)
Te	Tellurium	52	127.60
Th	Thorium	90	232.038
Ti	Titanium	22	47.90
Tl	Thallium	81	204.37
Tm	Thulium	69	158.934
U	Uranium	92	238.03
V	Vanadium	23	50.942
W	Tungsten	74	183.85
Xe	Xenon	54	131.30
Y	Yttrium	39	88.905
Yb	Ytterbium	70	173.04
Zn	Zinc	30	65.37
Zr	Zirconium	40	91.22

INDEX-1

U

V

Z

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Semiconductor Diodes," pages 1-1 through 1-47.

1-1. Which of the following electronic devices operates by virtue of the movement of electrons within a solid piece of semiconductor material?

1. Transistor
2. Junction diode
3. Solid-state device
4. Each of the above

1-2. Which of the following electronic devices is a minute piece of semiconductor material that can produce complete electronic circuit functions?

1. Zener diode
2. Light-emitting diode
3. Integrated circuit
4. Field effect transistor

1-3. Which of the following terms is used for the decrease in resistance as the temperature of the semiconductor increases?

1. Positive temperature coefficient
2. Negative temperature coefficient
3. Faraday temperature coefficient
4. Zero temperature coefficient

1-4. In addition to rectifying properties, selenium has the property of being light sensitive. How is selenium's resistance affected by light?

1. It decreases with an increase in light intensity
2. It increases with an increase in light intensity
3. It remains constant with variation in light intensity
4. It increases regardless of the variation in light intensity

1-5. One of the most sensitive elements of semiconductor materials is galena. Galena is a crystalline form of what material?

1. Krypton
2. Bismuth
3. Strontium
4. Lead sulfide

1-6. What significant discovery caused a breakthrough in the development of semiconductor devices?

1. The junction diode
2. The junction barrier
3. The extrinsic semiconductor
4. The point-contact transistor

1-7. Which of the following devices is frequently used to regulate power supply voltages at precise levels?

1. Junction diode
2. Tunnel diode
3. Esaki diode
4. Zener diode

1-8. Which of the following solid-state devices has both gain and fast-switching capabilities?

1. Zener diode
2. Tunnel diode
3. Junction diode
4. Point-contact diode

1-9. Which of the following advantages, if any, does a conventional electron tube have over a semiconductor device?

1. It is more efficient
2. It has a longer life
3. It is more economical
4. None of the above

1-10. When compared to an electron tube, the semiconductor device has which of the following limitations?

1. The semiconductor is more sensitive to temperature
2. The semiconductor is used only in radar equipment
3. The semiconductor is difficult to adapt to commercial products
4. Each of the above

1-11. Matter can be found in which of the following forms?

1. Solid
2. Liquid
3. Gas
4. Each of the above

1-12. A substance that cannot be reduced to a simpler form by chemical means is called a/an

1. element
2. mixture
3. compound
4. solution

1-13. An atom is the smallest possible particle that retains the characteristics of which of the following substances?

1. An element
2. A mixture
3. A compound
4. A solution

1-14. A molecule is the smallest possible particle that retains the characteristics of which of the following substances?

1. An element
2. A mixture
3. A compound
4. A solution

1-15. Which part of the atom has a negative charge and a small mass?

1. Proton
2. Electron
3. Positron
4. Neutron

1-16. Which part of the atom has a positive charge and a large mass?

1. Proton
2. Electron
3. Positron
4. Neutron

1-17. Which part of the atom has no electrical charge?

1. Proton
2. Electron
3. Positron
4. Neutron

1-18. What name is given to the outermost shell of an atom?

1. First shell
2. M shell
3. Valence shell
4. Subshell

1-19. What term is used to describe an atom which has more than its normal amount of electrons?

1. Ion
2. Ionization
3. Positive ion
4. Negative ion

1-20. Which of the following terms is defined as the process by which an atom gains or loses electrons?

1. Quanta
2. Ionization
3. Loss of energy
4. Remaining energy

1-21. Electrons are NEVER found in which of the following bands?

1. Energy band
2. Valence band
3. Forbidden band
4. Conduction band

1-22. What determines whether a substance is an insulator, semi-conductor, or conductor?

1. The separation between the valence and forbidden bands
2. The separation between the conduction and valence bands
3. The separation between the conduction and forbidden bands
4. The separation between the forbidden band and the energy gap

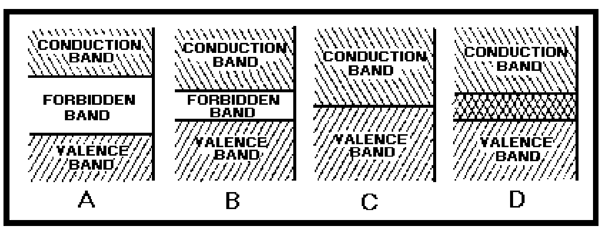


Figure 1A.—Energy level diagram.

IN ANSWERING QUESTIONS 1-23 AND 1-24, REFER TO FIGURE 1-A.

1-23. What energy level in figure 1-A is classified as the best insulator?

1. A
2. B
3. C
4. D

THIS SPACE LEFT BLANK INTENTIONALLY.

1-24. When the insulator is compared to the semiconductor, (a) which one requires the least energy to move an electron and (b) from which point in the energy level in figure 1-A does the electron leave in its travel to the conduction band?

1. (a) Semiconductor (b) Valence band
2. (a) Insulator (b) Valence band
3. (a) Semiconductor (b) Forbidden band
4. (a) Insulator (b) Forbidden band

1-25. Which of the following terms applies to the process that holds the atom together in a crystal?

1. Suhl effect
2. Superposition
3. Boundary defect
4. Covalent bonding

1-26. The movement of electrons in a semiconductor toward the applied voltage is termed

1. hole flow
2. positive conduction
3. negative conduction
4. electron current flow

1-27. When the theory of semiconductors is discussed, what term(s) is/are used to describe the current that flows in the semiconductor?

1. Hole flow
2. Electron flow
3. Both 1 and 2 above
4. Electromotive flow

1-28. What process takes place within the semiconductor to cause hole flow?

1. The breaking of covalent bonds
2. The combining of valence bands
3. The flexing of the material
4. The splitting of atoms

1-29. A material which has an equal number of electron-hole pairs and conducting electrons is known as what type of semiconductor material?

1. Extrinsic
2. Intrinsic
3. N-type
4. P-type

1-30. The process of adding impurities to crystals is known by which of the following terms?

1. Charging
2. Doping
3. Honing
4. Processing

1-31. When doping increases the number of free electrons in a semiconductor, what type of impurity has been added?

1. E-type
2. N-type
3. O-type
4. P-type

1-32. The semiconductor doping impurities—arsenic, antimony, and bismuth—are classified as what type of impurities?

1. Active
2. Neutral
3. Trivalent
4. Pentavalent

1-33. In the P-type semiconductor, what are the majority carriers?

1. The electrons
2. The holes
3. The inactive atoms
4. The inert atoms

1-34. What is/are the purpose(s) of the PN junction diode?

1. To rectify only
2. To amplify only
3. To rectify and amplify
4. To switch

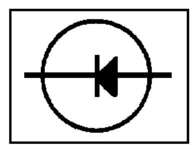
1-35. In a schematic diagram of a PN junction diode, which of the following symbols represents the cathode (N-type material)?

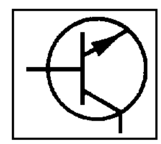
1. The arrow
2. The circle
3. The vertical bar
4. The horizontal line

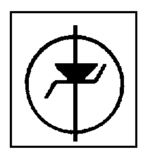
1-36. Which of the following alphanumerical codes correctly identifies the diode, crystal rectifier number 3, in a circuit?

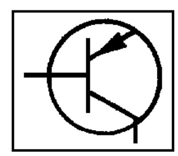
1. CR3
2. DR3
3. RD3
4. CRD3

1-37. Which of the following is the schematic symbol of the PN junction diode?

1. 

3. 

2. 

4 

1-38. The placing of an impurity on a semiconductor of the opposite impurity and fusing the two together produce what type of semiconductor junction?

1. Grown junction
2. Alloy Junction
3. Barrier junction
4. Intrinsic junction

1-39. A perfect bond at the junction of the two diode materials is important for which of the following reasons?

1. It is the point at which rectification takes place
2. It is the point at which amplification takes place
3. It is the main structural point from where the diode gets its strength
4. All of the above

1-40. Current flow in a copper wire can be compared to current flow in what type(s) of semiconductor material?

1. N-type only
2. P-type only
3. N- and P-type
4. All types

1-41. What is the overall electrical charge of the N-material in a semiconductor?

1. Zero
2. Some negative value
3. Some positive value
4. Depending upon the balance of electrons, it will be positive or negative

1-42. What is the overall electrical charge of the P-material in a semiconductor?

1. Zero
2. Some negative value
3. Some positive value
4. Depending upon the balance of electrons, it will be positive or negative

1-43. What causes the process called junction recombination to occur when N and P materials are joined together?

1. The diffusion of electrons and holes moving across the junction into the two materials
2. The generation of heat which causes the electrons to bombard the holes at the junction
3. The development of an electrostatic field on each side of the junction
4. The loss of electrons to the depletion region

1-44. After the junction recombination process has reached equilibrium, what is the area that surrounds the junction called?

1. The anode
2. The free ion space
3. The depletion region
4. The electrostatic field

1-45. A voltage applied to a PN junction so that it reduces the junction barrier and aids current flow is what type of bias?

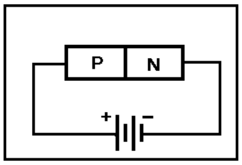
1. Indirect
2. Reverse
3. Forward
4. Direct

1-46. In a forward-biased PN junction, when an electron leaves the negative terminal of the battery and enters the N material, it becomes what type of carrier?

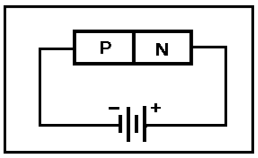
1. Loop
2. Signal
3. Majority
4. Minority

1-47. Which of the following illustrations depicts a properly forward-biased PN junction? (Note: The number of elements in the battery indicates the applied voltages.)

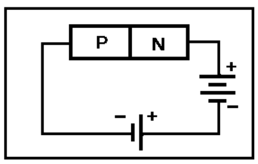
1.



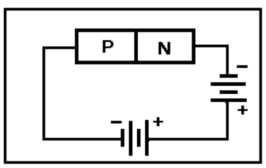
2.



3.



4.



1-48. In the PN junction, which of the following actions will increase the number of majority carriers and increase current flow in a forward-biased condition?

1. Increasing the size of the P material
2. Decreasing the size of the P material
3. Increasing battery voltage
4. Decreasing battery voltage

1-49. A voltage applied to a PN junction so that it will increase the junction barrier and offer a high resistance to current flow is called what type of bias?

1. Direct
2. Forward
3. Reverse
4. Indirect

1-50. When the negative terminal of a battery is connected to the P material, and the positive terminal is connected to the N material, what type of bias is being used?

1. Self
2. Forward
3. Reverse
4. Inverse

1-51. What provides you with information concerning the voltage-current relationship of a PN junction diode?

1. The body color of the diode
2. The color coded bands on the diode
3. The printed information on the diode
4. The characteristic curve graph of the diode

THIS SPACE LEFT BLANK INTENTIONALLY.

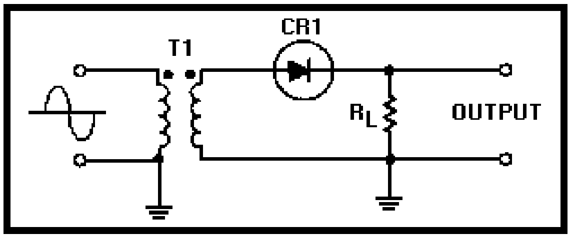


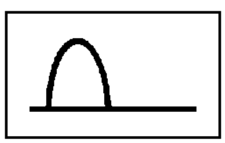
Figure 1B.—An electronic circuit.

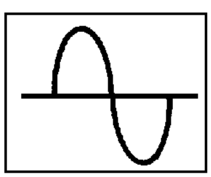
IN ANSWERING QUESTIONS 1-52 THROUGH 1-55, REFER TO FIGURE 1-B.

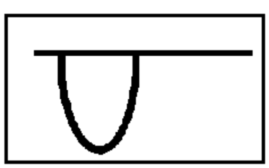
1-52. What type of circuit is shown in figure 1-B?

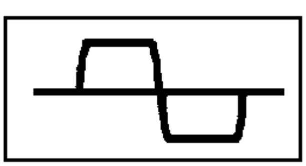
1. Full-wave rectifier
2. Half-wave rectifier
3. Clipper
4. Clamper

1-53. With the input shown, which of the following outputs would be correct?

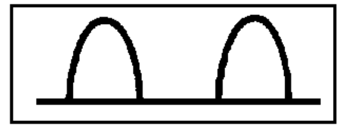
1. 

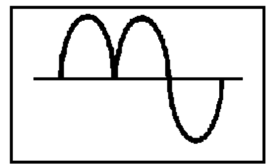
2. 

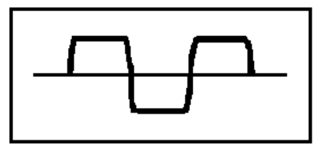
3. 

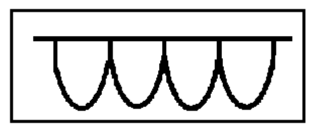
4. 

1-54. Which of the following outputs would be correct with two alternations of the ac input signal applied?

1. 

2. 

3. 

4. 

1-55. What is/are the purposes(s) of RL?

1. It limits the amount of current flow in the circuit
2. It develops the output signal
3. Both 1 and 2 above
4. It maintains the proper bias on the diode

THIS SPACE LEFT BLANK INTENTIONALLY.

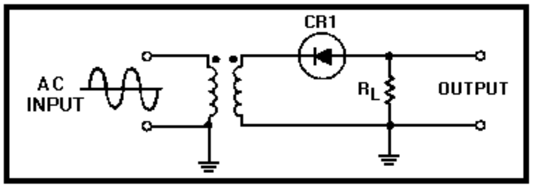


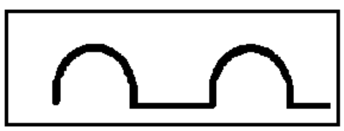
Figure 1C

IN ANSWERING QUESTIONS 1-56 AND 1-57, REFER TO FIGURE 1-C.

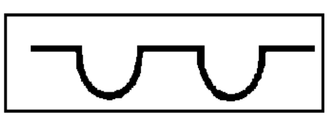
1-56. What type of circuit is shown in figure 1-C?

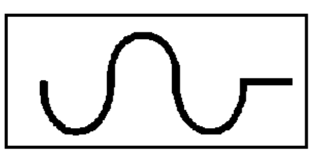
1. Positive half-wave rectifier
2. Negative half-wave rectifier
3. Full-wave rectifier
4. Clipper

1-57. With the input shown in figure 1-C, which of the following outputs would be correct?

1. 

2. 

3. 

4. 

1-58. If the input frequency to a half-wave rectifier is 120 hertz, what is the output frequency of the rectified dc?

1. 30 pps
2. 60 pps
3. 120 pps
4. 240 pps

1-59. Why are the units (plates) of the metallic rectifier stacked?

1. To dissipate heat
2. To be used in more than one circuit
3. To prevent inverse voltage breakdown
4. To handle high current applications

1-60. Which of the following types of rectifiers replaces the bulky selenium rectifier?

1. Copper-oxide rectifier
2. Half-wave rectifier
3. Metallic rectifier
4. Silicon rectifier

1-61. Signal diodes are used for which of the following purposes?

1. As mixers
2. As switches
3. As detectors
4. Each of the above

1-62. What type of bias makes a diode act as an open switch?

1. Direct
2. Reverse
3. Forward
4. Switching

1-63. A standard specification sheet for a diode contains which of the following information?

1. A brief description of the diode
2. Major application of the diode
3. Special features of the diode
4. All of the above

IN ANSWERING QUESTIONS 1-64 THROUGH 1-66, MATCH THE ELECTRICAL LETTER SYMBOL FOR RECTIFIER DIODES IN COLUMN B TO THE DEFINITION IN COLUMN A.

	A. DEFINITIONS		B. SYMBOLS
1-64.	The maximum reverse dc voltage that will not cause breakdown	1.	V_R
1-65.	The peak current specified for a given number of cycles or portion of a cycle	2.	I_{SURGE}
1-66.	The average reverse current at a specified temperature, usually 60 hertz	3.	$I_{F(AV)}$
		4.	$I_{R(AV)}$

1-67. A matching pair of diodes is indicated by which of the following numbers?

1. 2N325
2. 1N325C
3. 2N325M
4. 1N325M

1-68. The number 3N345 identifies which of the following semiconductors, if any?

1. Diode
2. Transistor
3. Tetrode transistor
4. None of the above

1-69. What type of diode has green, blue, and orange bands?

1. 1N463
2. 1N572
3. 1N663
4. 1N563

1-70. One of the prime dangers to the semiconductor diode is heat. Excessive current generated by heat which eventually destroys a diode is called

1. junction overload
2. thermal runaway
3. thermoplastic action
4. thermionic emission

1-71. When replacing a diode in a circuit, which of the following safety precautions should you observe in removing the diode from the circuit?

1. Do not pry the diode from the circuit
2. Do not use excessive heat to remove the diode
3. Do not remove the diode from the circuit while voltage is applied
4. All of the above

1-72. As you make a front-to-back ratio check of a diode with an ohmmeter, your first measurement (forward) is a low resistance reading, and your second measurement (reverse) is also a low reading. What should be your evaluation of the diode?

1. It is open
2. It is shorted
3. It is good
4. It is leaky

1-73. What are normally the front-to-back ratio of (a) a power rectifier and (b) a signal diode?

1. (a) 10:1 (b) 50:1
2. (a) 10:1 (b) 300:1
3. (a) 300:1 (b) 10:1
4. (a) 300:1 (b) 50:1

1-74. Of the following tests, which is the most valid for checking a diode?

1. A forward and reverse resistance check with an ohmmeter
2. The substitution of a new diode for the questionable one
3. A dynamic electrical check with a diode test set
4. A forward and reverse resistance check using two different ohmmeters

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Transistors," pages 2-1 through 2-53.

2-1. What term is used for a semiconductor that has three or more elements?

1. Diode
2. Transistor
3. Duo-diode
4. Point contact

2-2. The term transistor was derived from which of the following words?

1. Resistance and capacitance
2. Transformer and resistor
3. Resistor and transformer
4. Transfer and resistor

2-3. What are the three elements of a transistor?

1. Anode, base, and collector
2. Cathode, base, and collector
3. Emitter, collector, and base
4. Collector, emitter, and cathode

2-4. In a transistor, the flow of current carriers is controlled by which element(s)?

1. Emitter
2. Collector
3. Both 1 and 2 above
4. Base

2-5. In a transistor schematic, what is indicated by (a) the angular line with the arrowhead and (b) the direction of the arrow?

1. (a) Cathode
   (b) Direction of current flow
2. (a) Base
   (b) Direction of current flow
3. (a) Emitter
   (b) Type of transistor
4. (a) Collector
   (b) Type of transistor

2-6. Junction transistors have replaced point-contact transistors for which of the following reasons?

1. Junction transistors generate less noise
2. Junction transistors handle more power
3. Junction transistors provide higher current and voltage gains
4. All of the above

2-7. What is the total number of PN junctions in a transistor?

1. One
2. Two
3. Three
4. Four

2-8. What are the two junctions of a transistor?

1. Emitter-base and emitter-collector
2. Emitter-collector and base-collector
3. Emitter-base and collector-emitter
4. Emitter-base and base-collector

2-9. With proper bias applied to a transistor, what should be the relative resistance of (a) the emitter-base junction and (b) the base-collector junction?

1. (a) High (b) low
2. (a) High (b) high
3. (a) Low (b) low
4. (a) Low (b) high

2-10. For normal operation of a transistor, what is the bias of the (a) emitter-base junction and (b) base-collector junction?

1. (a) Forward (b) reverse
2. (a) Forward (b) forward
3. (a) Reverse (b) forward
4. (a) Reverse (b) reverse

A. I_C	E. V_{BB}
B. I_B	F. V_{CEO}
C. I_E	G. V_{EB}
D. V_{CC}	H. I_{CBO}

Figure 2A.—Transistor symbology.

IN ANSWERING QUESTIONS 2-11 THROUGH 2-15, REFER TO FIGURE 2-A. MATCH THE SYMBOL TO THE TERM GIVEN IN THE QUESTION.

2-11. The symbol for base current.

1. H
2. D
3. C
4. B

2-12. The symbol for collector current.

1. A
2. C
3. G
4. H

2-13. The symbol for emitter current.

1. B
2. C
3. D
4. F

2-14. The symbol for collector voltage supply.

1. C
2. D
3. E
4. F

2-15. The symbol for base voltage supply.

1. E
2. F
3. G
4. H

2-16. In a transistor, what percent of the total current flows through the emitter lead?

1. 100
2. 98
3. 60
4. 5

2-17. What are the majority current carriers in (a) the PNP transistor and (b) the NPN transistor?

1. (a) Holes (b) holes
2. (a) Holes (b) electrons
3. (a) Elements (b) holes
4. (a) Electrons (b) electrons

2-18. How will the transistor currents be affected if the forward bias provided by V_{BB} is increased?

1. I_B will decrease, I_E will decrease, and I_C will decrease
2. I_B will increase, I_E will decrease, and I_C will decrease
3. I_B will increase, I_E will decrease, and I_C will increase
4. I_B will increase, I_E will increase, and I_C will increase

2-19. What device provides an increase in current, voltage, or power of a signal without appreciably altering the original signal?

1. Diode
2. Amplifier
3. Oscillator
4. Power supply

2-20. The resistor that provides forward bias for the emitter-base junction of a transistor is indicated by which of the following symbols?

1. R_T
2. R_g
3. R_L
4. R_B

2-21. The collector load resistor is represented by which of the following symbols?

1. R_T
2. R_g
3. R_L
4. R_B

2-22. In the quiescent state of a transistor circuit, what does the symbol VC indicate?

1. Collector voltage supply
2. Collector voltage
3. Current gain
4. Capacitor voltage

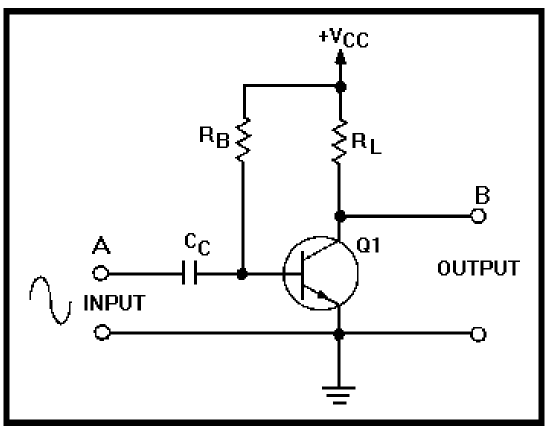


Figure 2B.—A basic transistor amplifier.

IN ANSWERING QUESTIONS 2-23 THROUGH 2-26, REFER TO FIGURE 2-B.

2-23. What is the purpose of CC?

1. To bypass ac signals to ground
2. To couple the input signal to the amplifier
3. To provide base bias to the preceding stage
4. To prevent ac variations on the base

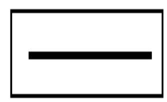
2-24. When the positive alternation of the input signal is applied at point A, what, if anything, happens to the bias on the base of the transistor?

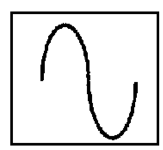
1. It increases
2. It decreases
3. It remains at its quiescent value
4. Nothing

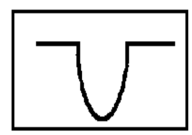
2-25. When the positive alternation of the input signal is applied at point A, what happens, if anything, to the current through RL?

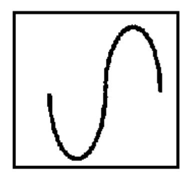
1. It increases
2. It decreases
3. It remains at its quiescent value
4. Nothing

2-26. When the signal shown at point A is applied to the base of Q1, which of the following signals will be at the output?

1. 

2. 

3. 

4. 

2-27. What type of bias keeps the base bias constant and improves thermal stability?

1. Self-bias
2. Fixed bias
3. Combination bias
4. Each of the above

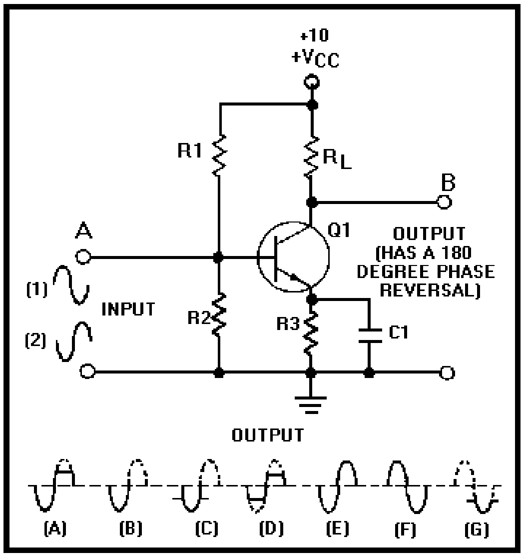


Figure 2C.—A transistor amplifier.

IN ANSWERING QUESTIONS 2-28 THROUGH 2-34, REFER TO FIGURE 2-C.

2-28. Which resistors are fixed biased resistors?

1. R1 and R_L
2. R1 and R3
3. R1 and R2
4. R2 and R3

2-29. Which resistor is self-biased?

1. R1
2. R2
3. R3
4. R_L

2-30. Which waveform is the output for a class AB amplifier with input number 1?

1. A
2. B
3. C
4. D

2-31. Which waveform is the output for a class A amplifier with input number 2?

1. C
2. D
3. E
4. F

2-32. Which waveform is the output for a class B amplifier with input number 1?

1. A
2. B
3. F
4. G

2-33. Which waveform is the output for a Class C amplifier with input number 2?

1. A
2. B
3. C
4. G

2-34. If the circuit is operating as a class A amplifier, but is being "overdriven," which output waveform is correct?

1. A
2. B
3. C
4. D

2-35. Which class of amplifier allows collector current to flow for a full 360 degrees of the input signal?

1. A
2. B
3. C
4. AB

2-36. Which class of amplifier allows collector current to flow for more than 180 degrees of the input signal but less than 360 degrees?

1. A
2. B
3. C
4. AB

2-37. Which class of amplifier has the highest fidelity and lowest efficiency?

1. A
2. B
3. C
4. AB

2-38. Which class of amplifier has the highest efficiency?

1. A
2. B
3. C
4. AB

2-39. What are the three transistor configurations?

1. Common base, common grid, and common output
2. Common anode, common collector, and common base
3. Common emitter, common base, and common collector
4. Common emitter, common base, and common base

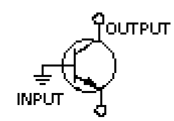
THIS SPACE LEFT BLANK INTENTIONALLY.

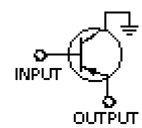
IN ANSWERING QUESTIONS 2-40 THROUGH 2-45, MATCH THE TRANSISTOR CONFIGURATIONS LISTED IN COLUMN B TO THE TRANSISTOR CIRCUITS SHOWN IN COLUMN A. (NOTE: ANSWERS IN COLUMN B MAY BE USED MORE THAN ONCE.)

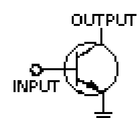
A. CIRCUITS

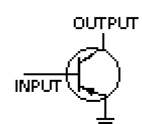
B. CONFIGURATIONS

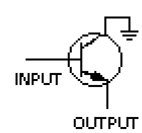
1. Common Emitter
2. Common Base
3. Common Collector

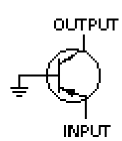
2-40. 

2-41. 

2-42. 

2-43. 

2-44. 

2-45. 

2-46. Which of the following is a transistor configuration that provides a phase reversal?

1. Common bias
2. Common input
3. Common emitter
4. Common collector

2-47. What is the symbol for input current in a common-emitter configuration?

1. I_E
2. I_B
3. I_C
4. I_T

2-48. What is the symbol for input current in a common-base configuration?

1. I_E
2. I_B
3. I_C
4. I_T

2-49. What term is used to indicate current gain in a common-emitter configuration?

1. Alpha
2. Beta
3. Gamma
4. X-ray

2-50. What term is used to indicate current gain in a common-collector configuration?

1. Alpha
2. Beta
3. Gamma
4. X-ray

2-51. What term is used to indicate current gain in a common-base configuration?

1. Alpha
2. Beta
3. Gamma
4. X-ray

2-52. Which of the following formulas is used to figure current gain in a common-emitter configuration?

1. $= \frac{\Delta I_C}{\Delta I_E}$
2. $= \frac{\Delta I_C}{\Delta I_B}$
3. $= \frac{\Delta I_E}{\Delta I_B}$
4. $= \frac{\Delta I_E}{\Delta I_C}$

2-53. Which of the following formulas is used to figure current gain in a common-base configuration?

1. $\alpha = \frac{\Delta I_C}{\Delta I_E}$
2. $\beta = \frac{\Delta I_C}{\Delta I_E}$
3. $\gamma = \frac{\Delta I_E}{\Delta I_B}$
4. $\gamma = \frac{\Delta I_E}{\Delta I_C}$

2-54. Which of the following formulas is used to figure current gain in a common-collector configuration?

1. $\gamma = \frac{\Delta I_E}{\Delta I_B}$
2. $\gamma = \frac{\Delta I_B}{\Delta I_E}$
3. $\beta = \frac{\Delta I_B}{\Delta I_C}$
4. $\alpha = \frac{\Delta I_C}{\Delta I_E}$

2-55. The common collector is also referred to by which of the following terms?

1. Low current gain amplifier
2. Voltage amplifier
3. Emitter follower
4. Grounded emitter

2-56. Which of the following conditions presents the greatest danger to a transistor?

1. Heat
2. High operating voltage
3. Excessive reverse current
4. Handling of the transistor

2-57. What method for checking transistors is cumbersome when more than one transistor is bad in a circuit?

1. Ohmmeter
2. Transistor checker
3. Voltage check
4. Substitution

A.	Hybrid IC
B.	Monolithic IC
C.	Microelectronics
D.	Modular Circuitry
E.	Integrated Circuit
F.	Printed Circuit Board
G.	Integrated Circuit Board

Figure 2D.—List of microelectronic terminology.

IN ANSWERING QUESTIONS 2-58 THROUGH 2-62, SELECT FROM FIGURE 2-D THE TERM DEFINED IN THE QUESTION.

2-58. A broad term used to describe the use of integrated circuits to miniaturize electronic equipment.

1. A
2. B
3. C
4. D

2-59. A flat insulating surface upon which printed wires and miniaturized components are connected in a predetermined design and attached to a common base.

1. A
2. B
3. F
4. G

2-60. An assembly technique in which printed circuit boards are stacked and connected to form a module.

1. B
2. C
3. D
4. E

2-61. A device that integrates both active and passive components of a complete electronic circuit in a single chip.

1. D
2. E
3. F
4. G

2-62. A plastic card on which integrated circuits are mounted.

1. A
2. B
3. F
4. G

ASSIGNMENT 3

Textbook assignment: Chapter 3, "Special Devices," pages 3-1 through 3-55.

3-1. What is the total number of connections in a diode?

1. One
2. Two
3. Three
4. Four

3-2. When the PN-junction diode is reversed biased, what happens to the majority carriers?

1. They combine with minority carriers at the junction
2. They move toward the junction
3. Both 1 and 2 above
4. They move away from the junction

3-3. What causes a small leakage current in a reverse-biased PN junction?

1. Holes
2. Electrons
3. Minority carriers
4. Majority carriers

3-4. At some potential, as you increase the reverse bias voltage on a PN junction, the reverse current increases very rapidly. What electronic term is given to this voltage potential?

1. Breakdown voltage
2. Reverse-bias
3. Forward-bias
4. Thermal runaway

THIS SPACE LEFT BLANK INTENTIONALLY.

3-5. Which of the following is a characteristic of the Zener diode?

1. A PN-junction diode that operates in the reverse-bias breakdown region
2. A PN-junction diode that uses the avalanche effect
3. A PN-junction diode that uses the Zener effect
4. Each of the above

3-6. What determines whether a solid material will act as a conductor, a semiconductor, or an insulator?

1. The energy level of the valence band
2. The energy level of the conductor band
3. The energy difference across the forbidden gap
4. The actual construction of the valence electrons

3-7. In comparing a conductor and an insulator, what is the relative dimension of the forbidden gap of (a) the conductor and (b) the insulator?

1. (a) Wide (b) wide
2. (a) Wide (b) narrow
3. (a) Narrow (b) narrow
4. (a) Narrow (b) wide

THIS SPACE LEFT BLANK INTENTIONALLY.

3-8. What is the "tunneling phenomenon" within the Zener diode?

1. An action where the minority carriers tunnel across the junction to form the current that occurs at breakdown
2. An action where the majority carriers tunnel across the junction to form the current that occurs at breakdown
3. An action that separates the conduction band and the valence band by a large gap
4. An action that removes all the electrons from the conduction band energy level

3-9. Which breakdown theory explains the action that takes place in a heavily doped PN junction with a reverse bias above 5 volts?

1. Zener effect
2. Avalanche breakdown
3. Energy band effect
4. Valence band gap crossing

3-10. Which breakdown theory explains the action that takes place in a heavily doped PN junction with a reverse bias below 5 volts?

1. Zener effect
2. Avalanche breakdown
3. Energy band effect
4. Valence band gap crossing

3-11. What happens to a Zener diode that has a reverse bias slightly higher than the breakdown voltage?

1. The Zener cuts off
2. The Zener acts like a short circuit
3. The Zener acts like an open circuit
4. The Zener conduction does not change

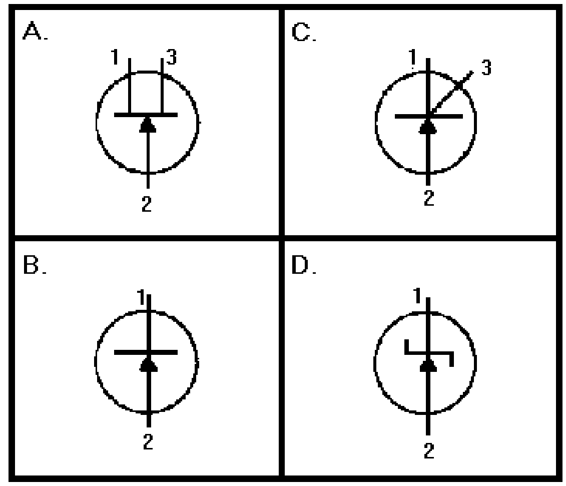


Figure 3A.—Schematic symbols.

IN ANSWERING QUESTIONS 3-12 AND 3-13, REFER TO FIGURE 3-A.

3-12. Which of the symbols represents a Zener diode?

1. A
2. B
3. C
4. D

3-13. In what direction does current flow in a Zener diode?

1. From point 1 to point 2
2. From point 2 to point 1
3. From point 1 to point 2 to point 3
4. From point 3 to point 2 to point 1

3-14. Why is the Zener diode an ideal voltage regulator?

1. It compensates for low supply voltage
2. It uses an unlimited number of carriers
3. Operating in the breakdown region does not harm it
4. The voltage across the diode remains almost constant after breakdown

3-15. In the construction of the tunnel diode, what is the ratio of impurity atoms to semiconductor atoms?

1. 10,000,000: 1000
2. 1,000:10,000,000
3. 10,000: 100,000
4. 100,000: 10,000

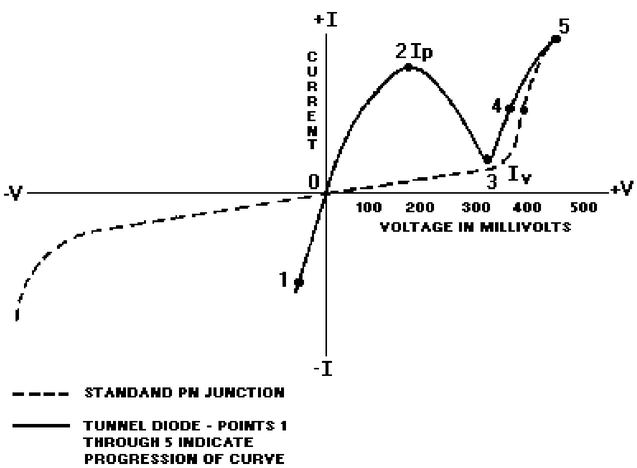


Figure 3B. —Tunnel diode characteristic curve.

IN ANSWERING QUESTIONS 3-16 THROUGH 3-18, REFER TO FIGURE 3-B.

3-16. What is/are the most important aspects(s) of the tunnel diode characteristic curve?

1. The forward-current increase to a peak with small applied forward bias
2. The decreasing forward current with an increasing forward bias to a minimum valley current (I_V)
3. The normal increasing forward current with further increase in the bias voltage
4. All of the above

3-17. What portion of the characteristic curve is the region of negative resistance?

1. From point 1 to point 2
2. From point 2 to point 3
3. From point 3 to point 4
4. From point 4 to point 5

3-18. At what area on the characteristic curve does the tunnel diode perform like a normal PN junction?

1. From point 0 to point 1
2. From point 1 to point 2
3. From point 2 to point 3
4. From point 3 to point 4

3-19. The varactor operates like which of the following electronic components?

1. A capacitor
2. An inductor
3. A variable capacitor
4. A variable inductor

3-20. An increase in reverse bias of a varactor will have what effect on the width of the depletion region?

1. It will stabilize
2. It will fluctuate
3. It will decrease
4. It will increase

3-21. What happens to the capacitance of a varactor diode as the reverse bias is increased?

1. It decreases
2. It increases
3. It remains the same

3-22. In electronic circuits, how is the varactor used?

1. As a tuning device
2. As a balancing device
3. As an amplifier
4. As a rectifier

3-23. What is/are the basic purpose(s) of the silicon controlled rectifier (SCR)?

1. To function as a switch
2. To function as a regulator
3. To function as a rectifier
4. All of the above

3-24. The SCR is equivalent to what electronic device?

1. Diode
2. Tetrode
3. Thyratron
4. Beam power tube

3-25. Which of the following circuits uses an SCR in its electronic circuitry?

1. Computer logic circuit
2. Voltage comparator circuit
3. Antenna power amplifier circuit
4. Each of the above

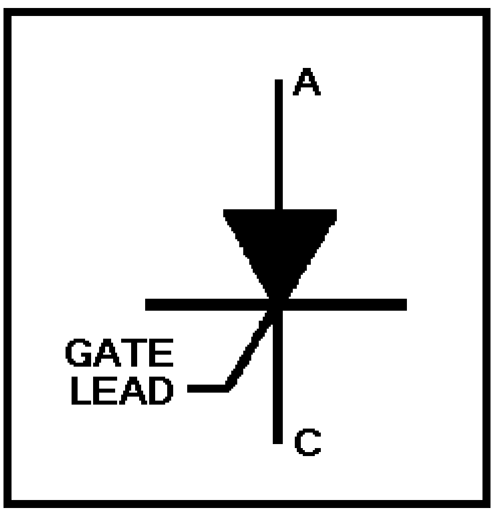


Figure 3C. —The schematic symbol for the SCR.

IN ANSWERING QUESTION 3-26, REFER TO FIGURE 3-C.

3-26. What is the impedance between points A and C (a) when the SCR is biased off and (b) when the SCR is at saturation?

1. (a) Low (b) low
2. (a) Low (b) high
3. (a) High (b) high
4. (a) High (b) low

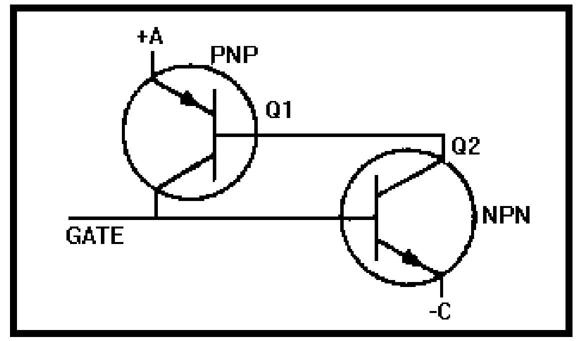


Figure 3D. —A two-transistor circuit.

IN ANSWERING QUESTION 3-27, REFER TO FIGURE 3-D.

3-27. When a positive current is applied to the gate lead, what happens to the collector current of Q1 and Q2?

1. Increases to a value limited only by the external circuit
2. Increases to a value limited only by the internal circuit
3. Decreases to a value limited only by the external circuit
4. Decreases to a value limited only by the internal circuit

3-28. Once an SCR is turned on by a positive pulse of current applied to the gate lead, what action turns the SCR off?

1. Removing the positive pulse from the gate lead
2. Inserting a negative pulse of current on the gate lead
3. Reducing the collector current to a value below that necessary to maintain conduction
4. Increasing the collector current to a point that the SCR will go into saturation and cut off

3-29. What is the total number of terminals in a TRIAC?

1. One
2. Two
3. Three
4. Four

3-30. What is the main difference between the TRIAC and the SCR?

1. The SCR requires a higher input voltage than the TRIAC
2. The TRIAC requires a higher input voltage than the SCR
3. The TRIAC controls and conducts current during both alternations of an ac cycle, while the SCR controls and conducts current during only one alternation
4. The SCR controls and conducts current during both alternations of an ac cycle, while the TRIAC controls and conducts currents during only one alternation

3-31. What name is given to a group of devices that either produce light or use light in their operation?

1. Optoelectronic
2. Ophthalmology
3. Optokenetic
4. Optometry

3-32. In optoelectronic devices, what do the initials LED stand for?

1. Low-emitting diode
2. Low-emitting device
3. Light-emitting diode
4. Light-emitting device

3-33. What determines the color of light emitted by an LED?

1. The type of incandescent bulb used
2. The type of material used
3. The type of bias used
4. The type of fluorescent bulb used

3-34. What is the standard schematic symbol used to designate LEDs?

1. An incandescent bulb with arrows pointing toward the light
2. An incandescent bulb with arrows pointing away from the light
3. A diode with two arrows pointing toward the cathode
4. A diode with two arrows pointing away from the cathode

3-35. The circuit symbols for all optoelectronic devices have arrows pointing either toward them or away from them. When the arrows point toward the symbol, what does this indicate?

1. The device produces light
2. The device uses light
3. The device requires current flow
4. The device produces current flow

THIS SPACE LEFT BLANK INTENTIONALLY.

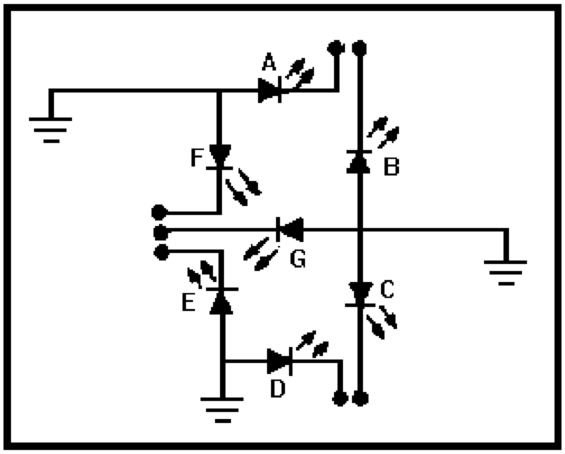


Figure 3E. —Light-emitting diode.

IN ANSWERING QUESTIONS 3-36 AND 3-37, REFER TO FIGURE 3-E.

3-36. When a negative voltage is applied to the proper cathodes in an LED display, a number is formed. To produce the numeral 5 on an LED display, which of the following diodes must have a negative potential on their cathodes?

1. ABCDE
2. ABGED
3. AFGCD
4. AFGBC

3-37. If an LED segment has an "8" displayed and the negative bias is removed from diodes F and C, what numeral will be displayed?

1. 1
2. 2
3. 3
4. 4

3-38. When replacing LED displays, which of the following methods should you use to ensure that the replacement display is of the same type as the faulty display?

1. A visual inspection
2. A check of the schematic symbols
3. Both 1 and 2 above
4. A check of the manufacturer's number

3-39. The photodiode acts as what type of electronic device?

1. Variable inductor
2. Variable resistor
3. Nonvariable inductor
4. Nonvariable resistor

3-40. When the photodiode is exposed to an external light, what happens to (a) resistance and (b) current?

1. (a) Increase (b) decreases
2. (a) Increases (b) increases
3. (a) Decreases (b) increases
4. (a) Decreases (b) decreases

3-41. To conduct, how must a photodiode be biased?

1. Reverse biased
2. Forward biased
3. Either 1 or 2 above, depending on light intensity

3-42. Photodiodes are useful in which of the following applications?

1. Computer card readers
2. Photographic light meters
3. Optic scanning equipment
4. Each of the above

3-43. Which of the following optoelectronic devices provides increased, conduction for a given light intensity?

1. LED
2. SCR
3. Phototransistor
4. Phototransformer

3-44. To compensate for ambient light, a phototransistor must have a total number of how many leads?

1. One
2. Two
3. Three
4. Four

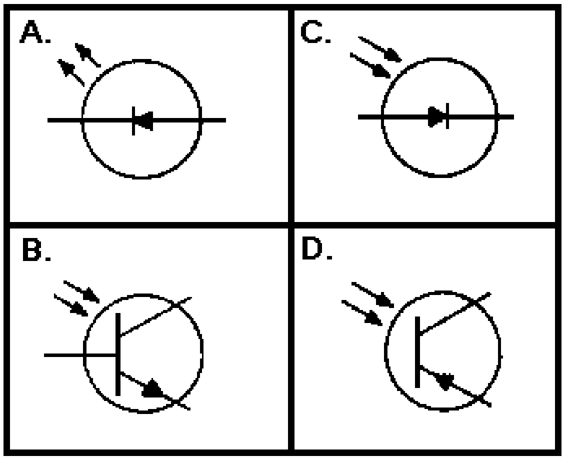


Figure 3F. —Solid-state schematic symbols.

IN ANSWERING QUESTIONS 3-45 THROUGH 3-47, SELECT FROM FIGURE 3-F THE SCHEMATIC SYMBOL IDENTIFIED IN THE QUESTION.

3-45. Schematic symbol for a photo-diode.

1. A
2. B
3. C
4. D

3-46. Schematic symbol for a two-terminal phototransistor.

1. A
2. B
3. C
4. D

3-47. Schematic symbol for a three-terminal phototransistor.

1. A
2. B
3. C
4. D

3-48. Which of the following devices is similar in operation to a photodiode?

1. Phototransistor
2. Photocell
3. LED
4. SCR

3-49. Which of the following is a typical light to dark resistance ratio of a photocell?

1. 1: 10
2. 1: 100
3. 1: 1000
4. 1:10,000

3-50. Photocells are used in which of the following circuits?

1. Controller
2. Oscillator
3. Amplifier
4. Detector

3-51. How should photovoltaic cells be coupled together to produce a relatively high voltage?

1. Series coupling
2. Parallel coupling
3. Inductive coupling
4. Mechanical coupling

3-52. What is the total number of terminals in a unijunction transistor (UJT)?

1. One
2. Two
3. Three
4. Four

3-53. The UJT has which of the following advantages over the conventional transistor?

1. Fewer terminals
2. Larger bandpass
3. Less bias is required
4. Increased temperature stability

3-54. How does the UJT differ from a conventional transistor?

1. The UJT has a second base instead of a collector
2. The UJT has a second emitter instead of a collector
3. The UJT has two collectors

3-55. When properly biased, what area(s) of the UJT act(s) as a resistor?

1. The area between base 1 and base 2
2. The area between emitter 1 and emitter 2
3. The area between collector 1 and collector 2
4. All of the above

3-56. The emitter of the UJT may be compared to what electronic component?

1. A fully charged capacitor
2. The wiper arm of a variable resistor
3. The collector of a conventional transistor
4. The secondary winding of a step-down transformer

3-57. What determines the level of voltage gradient at the emitter-base material contact point of a UJT?

1. The bias voltage
2. The manufacturer's specifications
3. The base area of the emitter
4. The voltage potential between base 2 and emitter

3-58. The UJT conducts from base 1 to (a) what point when it is forward biased and from (b) what point to bass 2 when it is reversed biased?

1. (a) Emitter (b) base 1
2. (a) Emitter (b) emitter
3. (a) Base 2 (b) base 1
4. (a) Base 1 (b) emitter

3-59. UJTs may be used in which of the following circuits?

1. Switching
2. Waveshaping
3. Oscillating
4. Each of the above

3-60. The field-effect transistor (FET) combines what desired characteristic of the vacuum tube with the many other advantages of the transistor?

1. Low output impedance
2. High output impedance
3. Low input impedance
4. High input impedance

3-61. What does the FET use to control the electrostatic field within the transistor?

1. Current
2. Voltage
3. Low input impedance
4. High input impedance

3-62. The junction field-effect transistor's (JFET) gate element corresponds very closely in operation with (a) what part of a conventional transistor and (b) what part of the vacuum tube?

1. (a) Emitter (b) cathode
2. (a) Base (b) grid
3. (a) Base (b) cathode
4. (a) Collector (b) plate

3-63. In the JFET, the portion of the bar between the deposit of gate material is of a smaller cross section than the rest of the bar. What does this cross section form?

1. A gate
2. A drain
3. A source
4. A channel

3-64. If a P-type material is used to construct the gate of a JFET, what material should be used to construct the remaining part of the JFET?

1. N-type
2. P-type
3. Mica type
4. Junction type

3-65. What is the key to FET operation?

1. The control of the effective cross-sectional area of the channel
2. The control of the effective cross-sectional area of the gate
3. Both 1 and 2 above
4. The low input impedance compared with the high output impedance

3-66. When reverse bias is applied to the gate lead of a JFET, what happens to (a) source-to-drain resistance of the device and (b) current flow?

1. (a) Decreases (b) decreases
2. (a) Decreases (b) increases
3. (a) Increases (b) decreases
4. (a) Increases (b) increases

3-67. What is the "pinch off" voltage of an FET?

1. The voltage required for the FET to conduct
2. The voltage required to overcome the FET reverse bias
3. The voltage required to reduce drain current to zero
4. The voltage required to reduce gate voltage to zero

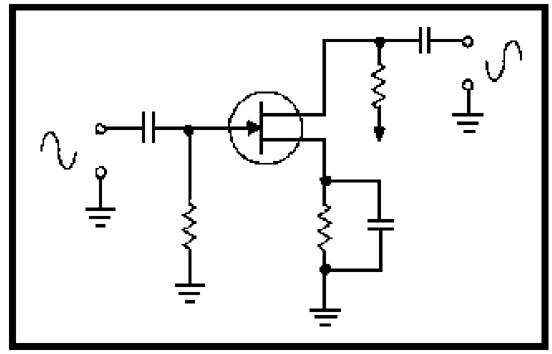


Figure 3G. —JFET common source amplifier.

IN ANSWERING QUESTIONS 3-68 AND 3-69, REFER TO FIGURE 3-G.

3-68. What type of transistor is shown in figure 3-G?

1. N channel JFET
2. P channel JFET
3. NPN/channel JFET
4. PNP/channel JFET

3-69. The circuit shown has which of the following characteristics?

1. Low impedance, high current gain
2. Low impedance, high voltage gain
3. High impedance, high current gain
4. High impedance, high voltage gain

3-70. The MOSFET has which of the following advantages over the JFET?

1. Less bias
2. Higher input impedance
3. Higher output impedance
4. All of the above

3-71. The MOSFET is normally constructed so that it operates in either the depletion mode or the enhancement mode. The depletion mode MOSFET (a) uses what type of bias and (b) has what type of doped channel to cause a depletion of current carriers in the channel?

1. (a) Reverse (b) lightly
2. (a) Forward (b) lightly
3. (a) Reverse (b) heavily
4. (a) Forward (b) heavily

3-72. The enhancement mode MOSFET (a) uses what type of bias and (b) has what type of doped channel to enhance the current carriers in the channel?

1. (a) Reverse (b) lightly
2. (a) Forward (b) lightly
3. (a) Reverse (b) heavily
4. (a) Forward (b) heavily

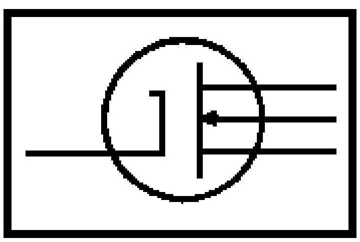


Figure 3H. —A MOSFET schematic symbol.

IN ANSWERING QUESTION 3-73, REFER TO FIGURE 3-H.

3-73. Which MOSFET element is identified by the arrow in the schematic symbol shown in figure 3-H?

1. Substrate
2. Source
3. Drain
4. Gate

3-74. What type metal is used in the construction of a MOSFET?

1. Oxide
2. Copper
3. Silver
4. Aluminum

3-75. What is the purpose of the shorting spring in a MOSFET?

1. To shunt the substrate to either the source or gate during operation
2. To protect the device from static electricity during replacement
3. To shunt the gates of a dual-gate MOSFET to make it operate like a single-gate MOSFET
4. To change the gain characteristics of the MOSFET

ASSIGNMENT 4

Textbook assignment: Chapter 4, "Solid-State Power Supplies," pages 4-1 through 4-62.

4-1. Which of the following is NOT one of the four sections of a basic power supply?

1. Transformer
2. Oscillator
3. Rectifier
4. Filter

4-2. The primary purpose of the transformer in an electronic power supply is to isolate the power supply from ground.

1. True
2. False

4-3. What is the primary function of the rectifier section?

1. To convert dc to ac
2. To convert ac to pulsating dc
3. To increase average voltage output
4. To decrease average voltage output

4-4. What is/are the functions of the filter section?

1. To eliminate dc voltage
2. To increase the amplitude of ac
3. To convert pulsating dc to steady dc
4. All of the above

4-5. The purpose of a center tap in a transformer is to provide

1. two separate dc voltages to the rectifier
2. a step-down voltage to the rectifier
3. pulsating dc to the rectifier
4. two equal voltages from one transformer

4-6. A diode is an ideal rectifier for which, if any, of the following reasons?

1. Current flows through the diode in one direction only
2. Current flows through the diode in both directions
3. Current will not flow through a diode
4. None of the above

4-7. When the anode of a diode is negative with respect to the cathode, the diode is said to be in what state?

1. Conduction
2. Saturation
3. Remission
4. Cutoff

4-8. In a simple half-way rectifier, the diode will conduct for a maximum of how many degrees of the 360-degree input signal?

1. 45
2. 90
3. 180
4. 270

4-9. What term is used to describe current pulses that flow in the same direction?

1. Average current
2. Secondary current
3. Pure direct current
4. Pulsating direct current

4-10. What is the ripple frequency of a half-wave rectifier with an input line frequency of 60 Hz?

1. 30 Hz
2. 60 Hz
3. 90 Hz
4. 120 Hz

4-11. In a half-wave rectifier, what is the average voltage output when the peak voltage is 300 volts?

1. 190.8 volts
2. 95.4 volts
3. 19.08 volts
4. 9.4 volts

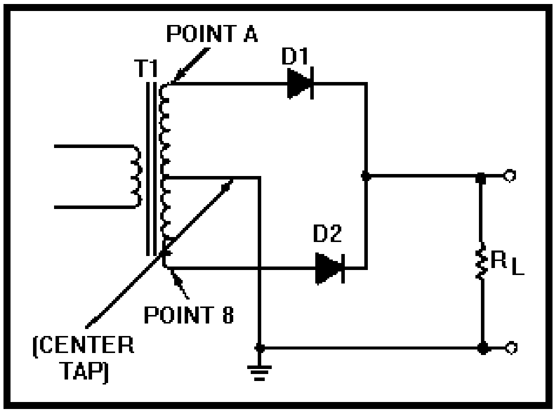


Figure 4A. —Complete full-wave rectifier.

IN ANSWERING QUESTIONS 4-12 AND 4-13, REFER TO FIGURE 4-A. ASSUME THAT THE VOLTAGE ACROSS THE TRANSFORMER SECONDARY (POINT A - POINT B) HAS AN RMS VALUE OF 480 VOLTS AC.

4-12. What is the peak value of the voltage pulse across the load?

1. 169.7 volts
2. 215.8 volts
3. 339.4 volts
4. 480 volts

4-13. What is the average load voltage?

1. 339.4 volts
2. 240 volts
3. 216 volts
4. 189.6 volts

4-14. What is the ripple frequency of a full-wave rectifier with an input line frequency of 60 Hz?

1. 30 Hz
2. 60 Hz
3. 90 Hz
4. 120 Hz

4-15. The full-wave rectifier has which of the following advantages over the half-wave?

1. Higher average voltage and current
2. Larger number of components
3. Higher value of voltage
4. Better regulation

4-16. What is the average voltage output of a full-wave rectifier that has an output of 100 volts peak?

1. 3.18 volts
2. 6.36 volts
3. 31.8 volts
4. 63.7 volts

4-17. The primary disadvantage of the conventional full-wave rectifier is that the peak output voltage is only one-half that of the half-wave rectifier.

1. True
2. False

THIS SPACE LEFT BLANK INTENTIONALLY.

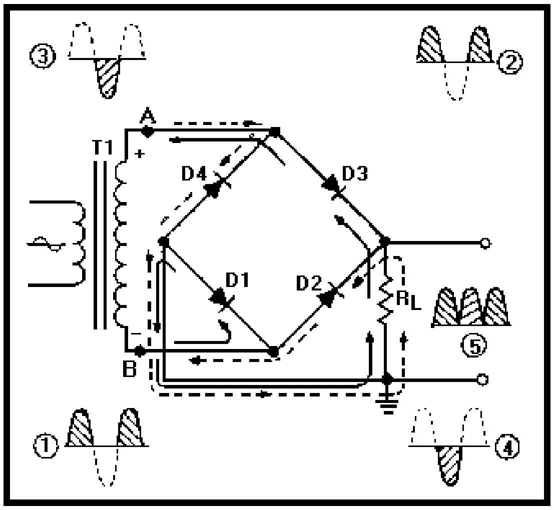


Figure 4B. —Bridge rectifier.

IN ANSWERING QUESTIONS 4-18 AND 4-19, REFER TO FIGURE 4-B.

4-18. When the voltage across the secondary of the transformer has the polarity shown, which of the diodes will conduct?

1. D1 and D3
2. D2 and D4
3. D1 and D2
4. D3 and D4

4-19. When the polarity reverses, which of the diodes will conduct?

1. D3 and D1
2. D4 and D2
3. D2 and D1
4. D4 and D3

4-20. In filter circuits, inductors are used as what type of impedance(s)?

1. Shunt impedance to oppose changes in current
2. Shunt impedances to oppose changes in voltage
3. Series impedances to oppose changes in current
4. Series impedances to oppose changes in voltage

4-21. To retain its charge, the capacitor in a simple capacitor filter must have a long charge time constant and a short discharge time constant.

1. True
2. False

4-22. If you increase the value of a capacitor, the X_C will increase.

1. True
2. False

4-23. To obtain a steady dc output in a simple capacitor circuit, the capacitor must charge almost instantaneously to the value of the applied voltage.

1. True
2. False

4-24. Which of the following factors, if any, determines the rate of discharge of the capacitor in a filter circuit?

1. The value of the load resistance
2. The amount of voltage
3. The type of capacitor
4. None of the above

4-25. A half-wave rectifier has an output frequency of 60 hertz, a filter capacitor value of 40 microfarads, and a load resistance of 10 kilohms. What is the value of X_C?

1. 133.3 ohms
2. 26.5 ohms
3. 66.3 ohms
4. 40.0 ohms

4-26. A full-wave rectifier has an output frequency of 120 hertz, a filter capacitor value of 25 microfarads, and a load resistance of 10 kilohms. What is the value of X_C?

1. 5.3 ohms
2. 53 ohms
3. 106 ohms
4. 1060 ohms

4-27. What type of filter is the most basic power supply filter?

1. Capacitor
2. LC choke-input
3. LC capacitor-input
4. RC capacitor-input

4-28. In a circuit with a capacitor filter, how is the capacitor connected?

1. In series with the load
2. In parallel with the load
3. In series with the input
4. Both 2 and 3 above

4-29. The LC choke-input filter is used primarily where which of the following types of regulation is/are important?

1. Frequency
2. Current only
3. Voltage only
4. Voltage and Current

4-30. In an LC choke-input filter circuit, the capacitor charges only to the average value of the input voltage. What component inhibits the capacitor from reaching the peak value of the input voltage?

1. The diode
2. The capacitor
3. The filter choke
4. The load resistor

4-31. In an LC choke-input filter, the larger the value of the filter capacitor, the better the filtering action. Which of the following factors represents the major limitation in obtaining the maximum value of the capacitor used?

1. Cost
2. Reliability
3. Availability
4. Physical size

4-32. What is the most common range of values, in henries, for a power supply choke?

1. 1 to 20
2. 5 to 25
3. 25 to 30
4. 10 to 200

4-33. If the impedance of the choke in an LC choke-input filter is increased, the ripple will

1. increase
2. decrease
3. oscillate
4. remain the same

4-34. A full-wave rectifier has an output frequency of 120 hertz, a filter choke with a value of 10 henries, and a load resistance of 10 kilohms. What is the value of X_L?

1. 75 ohms
2. 7.5 ohms
3. 75 kilohms
4. 7.5 kilohms

4-35. The filter capacitor in the LC choke-input filter is NOT subject to extreme voltage surges because of the protection provided by what component?

1. Shunt capacitor
2. Series resistor
3. Load resistor
4. Inductor

4-36. Shorted turns in the choke of an LC choke-input filter may reduce the value of inductance below the critical value. When this happens, which of the following problems may occur?

1. Poor voltage regulation
2. Excessive ripple amplitude
3. Abnormally high output voltage
4. Each of the above

4-37. The use of the RC capacitor-input filter is limited to which of the following situations?

1. When the load current is large
2. When the load current is small
3. When the load voltage is large
4. When the load voltage is small

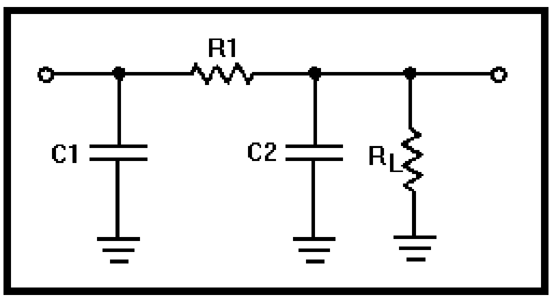


Figure 4C.—RC capacitor-input filter.

IN ANSWERING QUESTIONS 4-38 AND 4-39, REFER TO FIGURE 4-C.

4-38. Which of the components will have the highest failure rate?

1. C1
2. C2
3. R1
4. R_L

4-39. Which of the components provides protection against voltage surges in the circuit?

1. C1
2. C2
3. R1
4. R_L

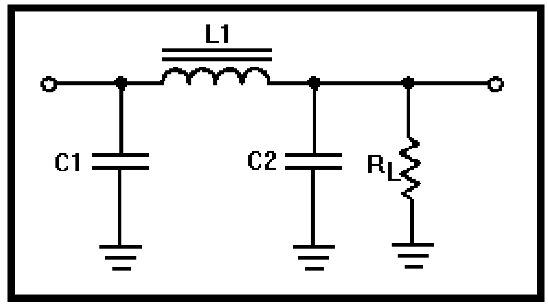


Figure 4D.—LC Capacitor-input filter.

IN ANSWERING QUESTIONS 4-40 AND 4-41, REFER TO FIGURE 4-D.

4-40. Components L1 and C2 form what type of circuit?

1. Ac voltage doubler
2. Dc voltage doubler
3. Ac voltage divider
4. Dc voltage divider

4-41. If component L1 shorts to the core, which of the following conditions will result?

1. Low output ripple frequency
2. Excessive ripple frequency
3. Excessively high output
4. No output

4-42. In a voltage regulator, what percent of regulation would be ideal?

1. 1 %
2. 5 %
3. 3 %
4. 0 %

IN ANSWERING QUESTIONS 4-43 THROUGH 4-45, REFER TO THE FOLLOWING FORMULA:

$$\text{Percent of Regulation} = \frac{E_{\text{no load}} - E_{\text{full load}}}{E_{\text{full load}}} \times 100$$

4-43. If a power supply produces 30 volts with no load and 25 volts under full load, what is the percent of regulation?

1. 5
2. 10
3. 20
4. 30

4-44. If a power supply produces 10 volts with no load and 9 volts under full load, what is the percent of regulation?

1. 8
2. 9
3. 10
4. 11

4-45. If a power supply produces 20 volts with no load and 20 volts under full load, what is the percent of regulation?

1. 1
2. 2
3. 3
4. 0

4-46. Basic voltage regulators are classified as either series or shunt. Their classification is determined by which of the following factors?

1. The type of regulating device used
2. The type of regulation required
3. The amount of regulation required
4. The position of the regulating device in relation to the load (R_L)

4-47. The simple series voltage regulator was designed to function as what type of resistance?

1. Fixed resistance in series with the load
2. Fixed resistance in parallel with the load
3. Variable resistance in series with the load
4. Variable resistance in parallel with the load

4-48. A series voltage regulator is designed so that what total percentage of current flows through the regulating device?

1. 25
2. 50
3. 75
4. 100

4-49. When a series voltage regulator is used to control output voltages, any increase in input voltage results in a/an

1. decrease in the voltage drop across the load resistance
2. increase in the voltage drop across the Zener diode
3. decrease in the resistance of the regulating device
4. increase in the resistance of the regulating device

THIS SPACE LEFT BLANK INTENTIONALLY.

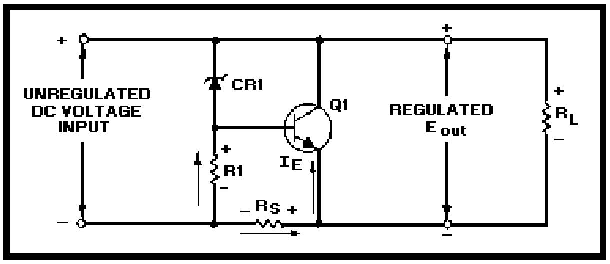


Figure 4E. —Simple shunt voltage regulator.

IN ANSWERING QUESTIONS 4-50 THROUGH 4-52, REFER TO FIGURE 4-E.

4-50. The base-emitter bias across Q1 is determined by which of the components?

1. R1 and CR1
2. R1 and R_S
3. R_S and CR1
4. R_S and R_L

4-51. What, if anything, happens to the forward bias of Q1 when the input voltage increases?

1. It increases
2. It decreases
3. Nothing, it remains the same

4-52. When the load current increases and the output voltage momentarily drops, what, if anything, happens to the resistance of Q1?

1. It increases to compensate for the drop
2. It decreases to compensate for the change
3. Nothing, it remains the same

4-53. What type of ammeter reading indicates that current regulator is functioning properly?

1. Constant
2. Deflection in the negative direction
3. Deflection in the positive direction
4. Fluctuation around the center line

4-54. A major disadvantage of having good current regulation is that good voltage regulation is lost.

1. True
2. False

4-55. To maintain a constant current flow when there is an increase in the load resistance (R_L), variable resistance (R_V) must compensate for this change by

1. increasing its resistance
2. decreasing its resistance
3. remaining the same

4-56. A decrease in the forward bias of a base-emitter junction has which of the following effects on the resistance of a transistor?

1. It increases
2. It deceases
3. It remains the same

4-57. Voltage multipliers are used primarily to develop what type of voltage?

1. Low voltage where low current is required
2. Low voltage where high current is required
3. High voltage where low current is required
4. High voltage where high current is required

4-58. The classification of voltage multipliers depends on which of the following ratios?

1. Input current to output current
2. Input current to output voltage
3. Output voltage to input voltage
4. Input voltage to output current

4-59. A half-wave voltage doubler consists of what total number of half-wave rectifiers?

1. One
2. Two
3. Three
4. Four

4-60. If a half-wave rectifier circuit is added to a half-wave voltage doubler circuit, what will be the resulting circuit?

1. A voltage doubler
2. A voltage tripler
3. A voltage quadruplet
4. A voltage quintuplet

4-61. Which of the following methods is used by manufacturers of electronic equipment to reduce the cost of extensive wiring?

1. Grounding the output of the power supply to the chassis
2. Grounding the return side of the power transformer to the chassis
3. Connecting all components in parallel
4. Connecting all components in series

4-62. When working on electronic equipment, the technician should observe which of the following safety precautions?

1. Make certain that the electronic equipment is properly grounded
2. Make certain that the test equipment is properly grounded
3. Make certain that the rubber mats are in good condition
4. All of the above

4-63. Which of the following is/are the most widely used check(s) for testing electronic equipment?

1. Smoke
2. Visual
3. Signal tracing
4. Both 2 and 3 above

4-64. Any connection that is located close to the chassis or to any other terminal should be examined for the possibility of which of the following problems?

1. An open
2. A short
3. A low resistance
4. A high resistance

4-65. Which of the following statements applies to a transformer that is discolored or leaking?

1. It is operational
2. It is cracked
3. It is shorted
4. It is open

4-66. As a technician, you notice that a resistor is discolored and charred. Which, if any, of the following conditions most likely caused the damage?

1. Overload
2. Open circuit
3. Ambient temperature
4. None of the above

4-67. You are in the process of energizing a power supply. You hear a boiling or sputtering noise and notice smoke coming from a section of the power supply. Which, if any, of the following actions should you take first?

1. Secure the power immediately
2. Examine the problem area
3. Remove the defective component
4. None of the above

4-68. Which, if any, of the following is the most rapid and accurate method for testing electronic circuits after completing visual inspection?

1. Smoke test
2. Current test
3. Signal tracing
4. None of the above

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 8—Introduction to Amplifiers

NAVEDTRA 14180

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 8 of a series.

History of the course:

Sep 1998: Original edition released.
Oct 2003: Administrative update released. Entered administrative updates. Technical content was not reviewed or revised.

NAVSUP Logistics Tracking Number
0504-LP-026-8330

TABLE OF CONTENTS

ASSIGNMENT QUESTIONS follow Index.

CHAPTER 1

AMPLIFIERS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the OCC/ECC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completion of this chapter, you will be able to:

1. Define amplification and list several common uses; state two ways in which amplifiers are classified.
2. List the four classes of operation of, four methods of coupling for, and the impedance characteristics of the three configurations of a transistor amplifier.
3. Define feedback and list the two types of feedback.
4. Describe and state one use for a phase splitter.
5. State a common use for and one advantage of a push-pull amplifier.

INTRODUCTION

This chapter is a milestone in your study of electronics. Previous modules have been concerned more with individual components of circuits than with the complete circuits as the subject. This chapter and the other chapters of this module are concerned with the circuitry of amplifiers. While components are discussed, the discussion of the components is not an explanation of the working of the component itself (these have been covered in previous modules) but an explanation of the component as it relates to the circuit.

The circuits this chapter is concerned with are AMPLIFIERS. Amplifiers are devices that provide AMPLIFICATION. That doesn't explain much, but it does describe an amplifier if you know what amplification is and what it is used for.

WHAT IS AMPLIFICATION?

Just as an amplifier is a device that provides amplification, amplification is the process of providing an increase in AMPLITUDE. Amplitude is a term that describes the size of a signal. In terms of a.c., amplitude usually refers to the amount of voltage or current. A 5-volt peak-to-peak a.c.signal would be larger in amplitude than a 4-volt peak-to-peak a.c. signal. "SIGNAL" is a general term used to refer to any a.c. or d.c. of interest in a circuit; e.g., input signal and output signal. A signal can be large or small, ac. or d.c., a sine wave or nonsinusoidal, or even nonelectrical such as sound or light. "Signal" is a very general term and, therefore, not very descriptive by itself, but it does sound more technical than the word "thing". It is not very impressive to refer to the "input thing" or the "thing that comes out of this circuit."

Perhaps the concept of the relationship of amplifier-amplification-amplitude will be clearer if you consider a parallel situation (an analogy). A magnifying glass is a magnifier. As such, it provides magnification which is an increase in the magnitude (size) of an object. This relationship of magnifier-magnification-magnitude is the same as the relationship of amplifier-amplification-amplitude. The analogy is true in one other aspect as well. The magnifier does not change the object that is being magnified; it is only the image that is larger, not the object itself. With the amplifier, the output signal differs in amplitude from the input signal, but the input signal still exists unchanged. So, the object (input signal) and the magnifier (amplifier) control the image (output signal).

An amplifier can be defined as a device that enables an input signal to control an output signal. The output signal will have some (or all) of the characteristics of the input signal but will generally be larger than the input signal in terms of voltage, current, or power.

USES OF AMPLIFICATION

Most electronic devices use amplifiers to provide various amounts of signal amplification. Since most signals are originally too small to control or drive the desired device, some amplification is needed.

For example, the audio signal taken from a record is too small to drive a speaker, so amplification is needed. The signal will be amplified several times between the needle of the record player and the speaker. Each time the signal is amplified it is said to go through a STAGE of amplification. The audio amplifier shown connected between the turntable and speaker system in figure 1-1 contains several stages of amplification.

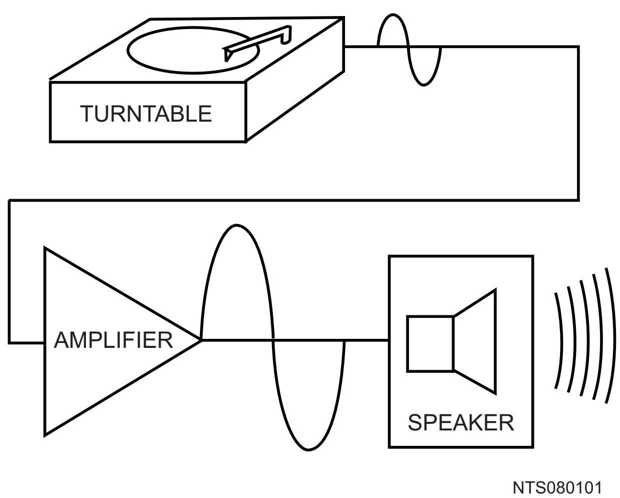


Figure 1-1.—Amplifier as used with turntable and speaker.

Notice the triangle used in figure 1-1 to represent the amplifier. This triangle is the standard block diagram symbol for an amplifier.

Another example of the use of an amplifier is shown in figure 1-2. In a radio receiver, the signal picked up by the antenna is too weak (small) to be used as it is. This signal must be amplified before it is sent to the detector. (The detector separates the audio signal from the frequency that was sent by the transmitter. The way in which this is done will be discussed later in this training series.)

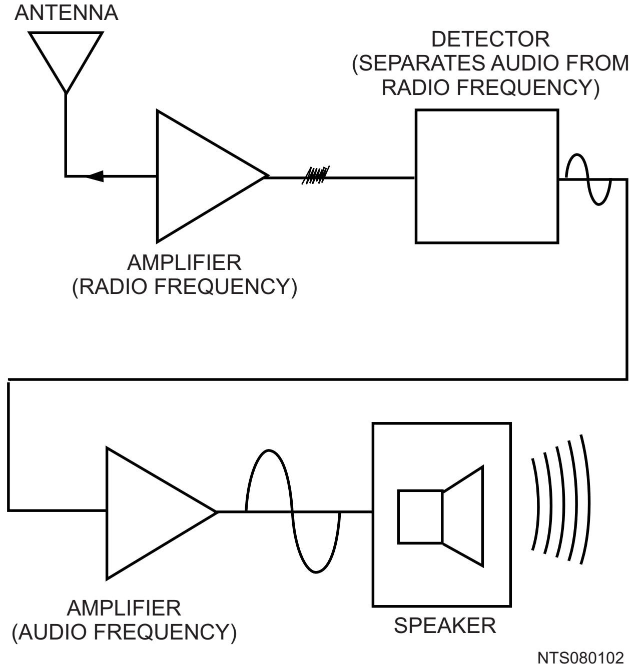


Figure 1-2.—Amplifiers as used in radio receiver.

The audio signal from the detector will then be amplified to make it large enough to drive the speaker of the radio.

Almost every electronic device contains at least one stage of amplification, so you will be seeing amplifiers in many devices that you work on. Amplifiers will also be used in most of the *NEETS* modules that follow this one.

Q-1. What is amplification?

Q-2. Does an amplifier actually change an input signal? Why or why not?

Q-3. Why do electronic devices use amplifiers?

CLASSIFICATION OF AMPLIFIERS

Most electronic devices use at least one amplifier, but there are many types of amplifiers. This module will not try to describe all the different types of amplifiers. You will be shown the general principles of amplifiers and some typical amplifier circuits.

Most amplifiers can be classified in two ways. The first classification is by their function. This means they are basically voltage amplifiers or power amplifiers. The second classification is by their frequency response. In other words what frequencies are they designed to amplify?

If you describe an amplifier by these two classifications (function and frequency response) you will have a good working description of the amplifier. You may not know what the exact circuitry is, but you will know what the amplifier does and the frequencies that it is designed to handle.

VOLTAGE AMPLIFIERS AND POWER AMPLIFIERS

All amplifiers are current-control devices. The input signal to an amplifier controls the current output of the amplifier. The connections of the amplifying device (electron tube, transistor, magnetic amplifier, etc.) and the circuitry of the amplifier determine the classification. Amplifiers are classified as voltage or power amplifiers.

A VOLTAGE AMPLIFIER is an amplifier in which the output signal voltage is larger than the input signal voltage. In other words, a voltage amplifier amplifies the voltage of the input signal.

A POWER AMPLIFIER is an amplifier in which the output signal power is greater than the input signal power. In other words, a power amplifier amplifies the power of the input signal. Most power amplifiers are used as the final amplifier (stage of amplification) and control (or drive) the output device. The output device could be a speaker, an indicating device, an antenna, or the heads on a tape recorder. Whatever the device, the power to make it work (or drive it) comes from the final stage of amplification which is a power amplifier.

Figure 1-3 shows a simple block diagram of a voltage amplifier with its input and output signals and a power amplifier with its input and output signals. Notice that in view (A) the output signal voltage is larger than the input signal voltage. Since the current values for the input and output signals are not shown, you cannot tell if there is a power gain in addition to the voltage gain.

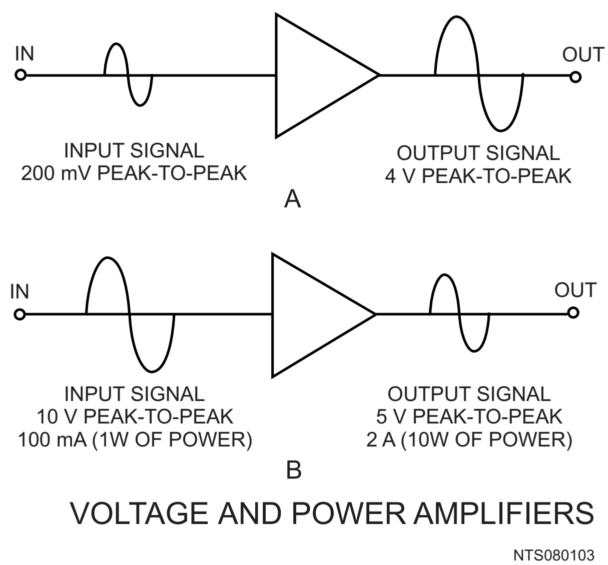


Figure 1-3.—Diagram of voltage and power amplifiers.

In view (B) of the figure the output signal voltage is less than the input signal voltage. As a voltage amplifier, this circuit has a gain of less than 1. The output power, however, is greater than the input power. Therefore, this circuit is a power amplifier.

The classification of an amplifier as a voltage or power amplifier is made by comparing the characteristics of the input and output signals. If the output signal is larger in voltage amplitude than the input signal, the amplifier is a voltage amplifier. If there is no voltage gain, but the output power is greater than the input power, the amplifier is a power amplifier.

FREQUENCY RESPONSE OF AMPLIFIERS

In addition to being classified by function, amplifiers are classified by frequency response. The frequency response of an amplifier refers to the band of frequencies or frequency range that the amplifier was designed to amplify.

You may wonder why the frequency response is important. Why doesn't an amplifier designed to amplify a signal of 1000 Hz work just as well at 1000 MHz? The answer is that the components of the amplifier respond differently at different frequencies. The amplifying device (electron tube, transistor, magnetic amplifier, etc.) itself will have frequency limitations and respond in different ways as the frequency changes. Capacitors and inductors in the circuit will change their reactance as the frequency changes. Even the slight amounts of capacitance and inductance between the circuit wiring and other components (interelectrode capacitance and self-inductance) can become significant at high frequencies. Since the response of components varies with the frequency, the components of an amplifier are selected to amplify a certain range or band of frequencies.

NOTE: For explanations of interelectrode capacitance and self-inductance see *NEETS Modules 2—Introduction to Alternating Current and Transformers;* 6—*Introduction to Electronic Emission, Tubes, and Power Supplies;* and 7—*Introduction to Solid-State Devices and Power Supplies.*

The three broad categories of frequency response for amplifiers are AUDIO AMPLIFIER, RF AMPLIFIER, and VIDEO AMPLIFIER.

An audio amplifier is designed to amplify frequencies between 15 Hz and 20 kHz. Any amplifier that is designed for this entire band of frequencies or any band of frequencies contained in the audio range is considered to be an audio amplifier.

In the term rf amplifier, the "rf" stands for radio frequency. These amplifiers are designed to amplify frequencies between 10 kHz and 100,000 MHz. A single amplifier will not amplify the entire rf range, but any amplifier whose frequency band is included in the rf range is considered an rf amplifier.

A video amplifier is an amplifier designed to amplify a band of frequencies from 10 Hz to 6 MHz. Because this is such a wide band of frequencies, these amplifiers are sometimes called WIDE-BAND AMPLIFIERS. While a video amplifier will amplify a very wide band of frequencies, it does not have the gain of narrower-band amplifiers. It also requires a great many more components than a narrow-band amplifier to enable it to amplify a wide range of frequencies.

Q-4. In what two ways are amplifiers classified?

Q-5. What type of amplifier would be used to drive the speaker system of a record player?

Q-6. What type of amplifier would be used to amplify the signal from a radio antenna?

TRANSISTOR AMPLIFIERS

A transistor amplifier is a current-control device. The current in the base of the transistor (which is dependent on the emitter-base bias) controls the current in the collector. A vacuum-tube amplifier is also a current-control device. The grid bias controls the plate current. These facts are expanded upon in *NEETS Module 6*, *Introduction to Electronic Emission, Tubes and Power Supplies,* and *Module 7*, *Introduction to Solid-State Devices and Power Supplies.*

You might hear that a vacuum tube is a voltage-operated device (since the grid does not need to draw current) while the transistor is a current-operated device. You might agree with this statement, but both the vacuum tube and the transistor are still current-control devices. The whole secret to understanding amplifiers is to remember that fact. Current control is the name of the game. Once current is controlled you can use it to give you a voltage gain or a power gain.

This chapter will use transistor amplifiers to present the concepts and principles of amplifiers. These concepts apply to vacuum-tube amplifiers and, in most cases, magnetic amplifiers as well as transistor amplifiers. If you wish to study the vacuum-tube equivalent circuits of the transistor circuits presented, an excellent source is the EIMB, NAVSEA 0967-LP-000-0120, *Electronics Circuits*.

The first amplifier concept that is discussed is the "class of operation" of an amplifier.

AMPLIFIER CLASSES OF OPERATION

The class of operation of an amplifier is determined by the amount of time (in relation to the input signal) that current flows in the output circuit. This is a function of the operating point of the amplifying device. The operating point of the amplifying device is determined by the bias applied to the device. There are four classes of operation for an amplifier. These are: A, AB, B and C. Each class of operation has certain uses and characteristics. No one class of operation is "better" than any other class. The selection of the "best" class of operation is determined by the use of the amplifying circuit. The best class of operation for a phonograph is not the best class for a radio transmitter.

Class A Operation

A simple transistor amplifier that is operated class A is shown in figure 1-4. Since the output signal is a 100% (or 360º) copy of the input signal, current in the output circuit must flow for 100% of the input signal time. This is the definition of a class A amplifier. Amplifier current flows for 100% of the input signal.

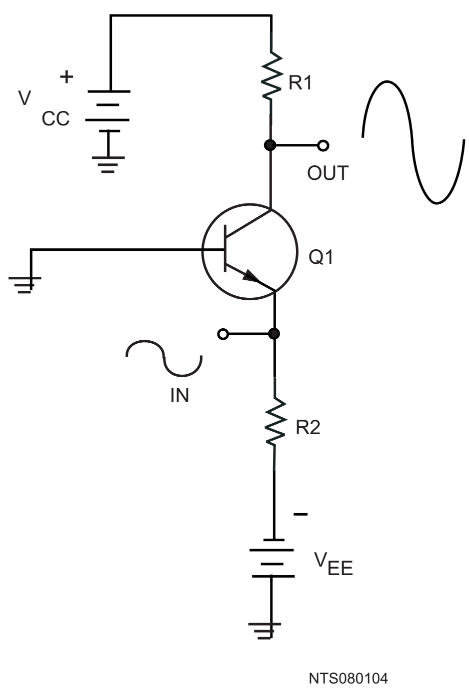


Figure 1-4.—A simple class A transistor amplifier.

The class A amplifier has the characteristics of good FIDELITY and low EFFICIENCY. Fidelity means that the output signal is just like the input signal in all respects except amplitude. It has the same shape and frequency. In some cases, there may be a phase difference between the input and output signal (usually 180º), but the signals are still considered to be "good copies." If the output signal is not like the input signal in shape or frequency, the signal is said to be DISTORTED. DISTORTION is any undesired change in a signal from input to output.

The efficiency of an amplifier refers to the amount of power delivered to the output compared to the power supplied to the circuit. Since every device takes power to operate, if the amplifier operates for 360º of input signal, it uses more power than if it only operates for 180º of input signal. If the amplifier uses more power, less power is available for the output signal and efficiency is lower. Since class A amplifiers operate (have current flow) for 360º of input signal, they are low in efficiency. This low efficiency is acceptable in class A amplifiers because they are used where efficiency is not as important as fidelity.

Class A B Operation

If the amplifying device is biased in such a way that current flows in the device for 51% - 99% of the input signal, the amplifier is operating class AB. A simple class AB amplifier is shown in figure 1-5.

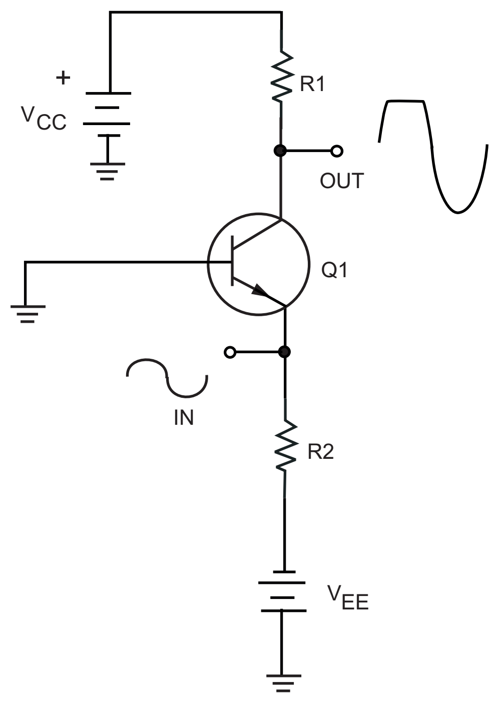


Figure 1-5.—A simple class AB transistor amplifier.

Notice that the output signal is distorted. The output signal no longer has the same shape as the input signal. The portion of the output signal that appears to be cut off is caused by the lack of current through the transistor. When the emitter becomes positive enough, the transistor cannot conduct because the base-to-emitter junction is no longer forward biased. Any further increase in input signal will not cause an increase in output signal voltage.

Class AB amplifiers have better efficiency and poorer fidelity than class A amplifiers. They are used when the output signal need not be a complete reproduction of the input signal, but both positive and negative portions of the input signal must be available at the output.

Class AB amplifiers are usually defined as amplifiers operating between class A and class B because class A amplifiers operate on 100% of input signal and class B amplifiers (discussed next) operate on 50% of the input signal. Any amplifier operating between these two limits is operating class AB.

Class B Operation

As was stated above, a class B amplifier operates for 50% of the input signal. A simple class B amplifier is shown in figure 1-6.

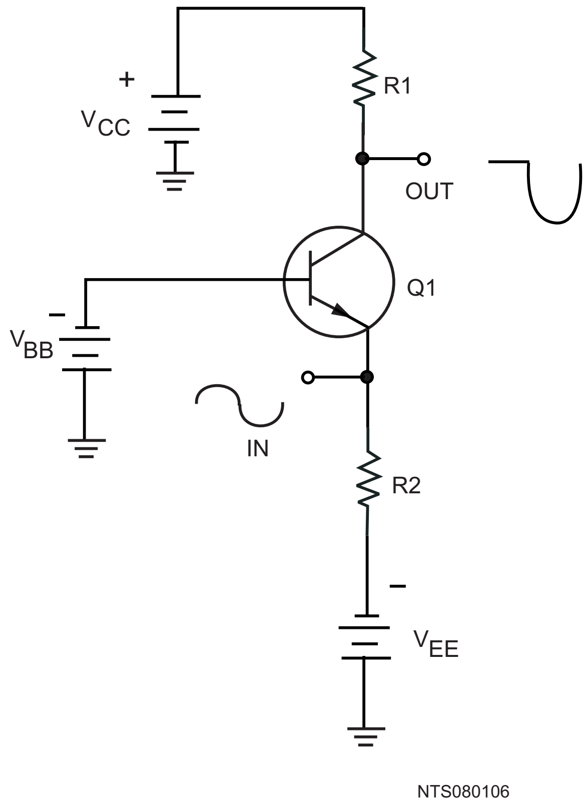


Figure 1-6.—A simple class B transistor amplifier.

In the circuit shown in figure 1-6, the base-emitter bias will not allow the transistor to conduct whenever the input signal becomes positive. Therefore, only the negative portion of the input signal is reproduced in the output signal. You may wonder why a class B amplifier would be used instead of a simple rectifier if only half the input signal is desired in the output. The answer to this is that the rectifier does not amplify. The output signal of a rectifier cannot be higher in amplitude than the input signal. The class B amplifier not only reproduces half the input signal, but amplifies it as well.

Class B amplifiers are twice as efficient as class A amplifiers since the amplifying device only conducts (and uses power) for half of the input signal. A class B amplifier is used in cases where exactly 50% of the input signal must be amplified. If less than 50% of the input signal is needed, a class C amplifier is used.

Class C Operation

Figure 1-7 shows a simple class C amplifier. Notice that only a small portion of the input signal is present in the output signal. Since the transistor does not conduct except during a small portion of the input signal, this is the most efficient amplifier. It also has the worst fidelity. The output signal bears very little resemblance to the input signal.

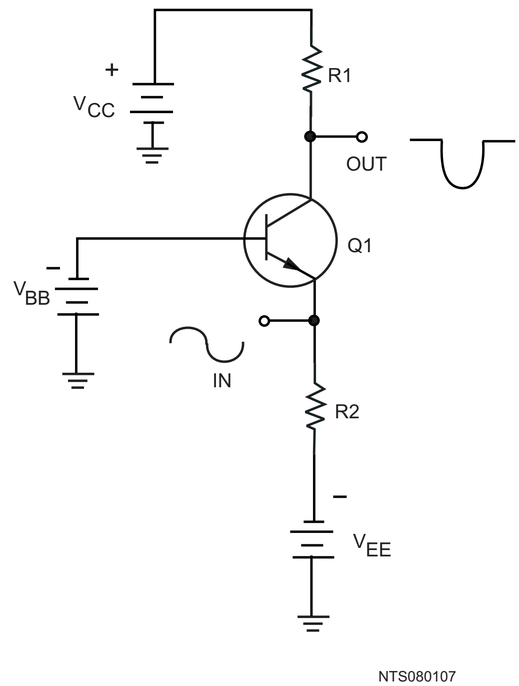


Figure 1-7.—A simple class C transistor amplifier.

Class C amplifiers are used where the output signal need only be present during part of one-half of the input signal. Any amplifier that operates on less than 50% of the input signal is operated class C.

Q-7. What determines the class of operation of an amplifier?

Q-8. What are the four classes of operation of a transistor amplifier?

Q-9. If the output of a circuit needs to be a complete representation of one-half of the input signal, what class of operation is indicated?

Q-10. Why is class C operation more efficient than class A operation?

Q-11. What class of operation has the highest fidelity?

AMPLIFIER COUPLING

Earlier in this module it was stated that almost every electronic device contains at least one stage of amplification. Many devices contain several stages of amplification and therefore several amplifiers. Stages of amplification are added when a single stage will not provide the required amount of amplification. For example, if a single stage of amplification will provide a maximum gain of 100 and the desired gain from the device is 1000, two stages of amplification will be required. The two stages might have gains of 10 and 100, 20 and 50, or 25 and 40. (The overall gain is the product of the individual stages-$10 \times 100 = 20 \times 50 = 25 \times 40 = 1000$.)

Figure 1-8 shows the effect of adding stages of amplification. As stages of amplification are added, the signal increases and the final output (from the speaker) is increased.

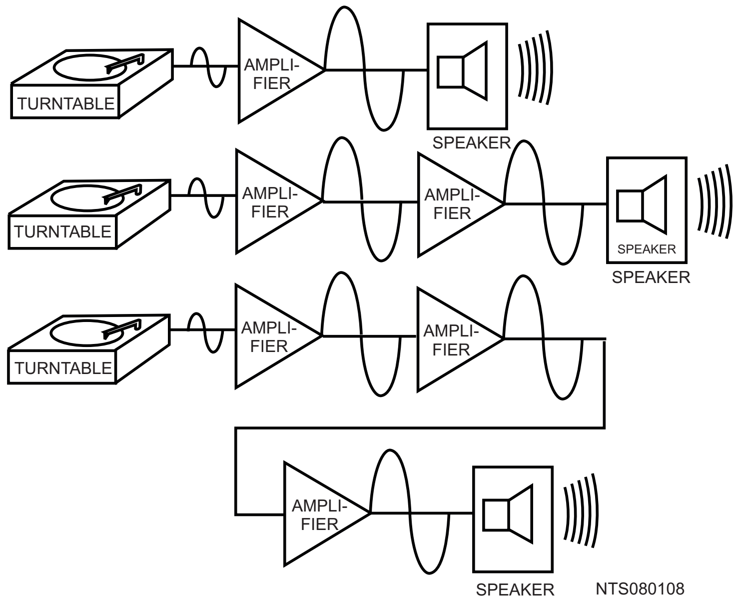


Figure 1-8.—Adding stages of amplification.

Whether an amplifier is one of a series in a device or a single stage connected between two other devices (top view, figure 1-8), there must be some way for the signal to enter and leave the amplifier. The process of transferring energy between circuits is known as COUPLING. There are various ways of coupling signals into and out of amplifier circuits. The following is a description of some of the more common methods of amplifier coupling.

Direct Coupling

The method of coupling that uses the least number of circuit elements and that is, perhaps, the easiest to understand is direct coupling. In direct coupling the output of one stage is connected directly to the input of the following stage. Figure 1-9 shows two direct-coupled transistor amplifiers.

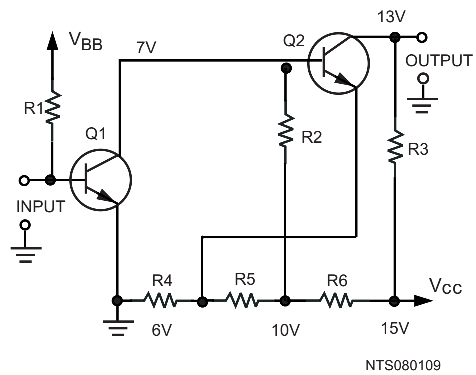


Figure 1-9.—Direct-coupled transistor amplifier.

Notice that the output (collector) of Q1 is connected directly to the input (base) of Q2. The network of R4, R5, and R6 is a voltage divider used to provide the bias and operating voltages for Q1 and Q2. The entire circuit provides two stages of amplification.

Direct coupling provides a good frequency response since no frequency-sensitive components (inductors and capacitors) are used. The frequency response of a circuit using direct coupling is affected only by the amplifying device itself.

Direct coupling has several disadvantages, however. The major problem is the power supply requirements for direct-coupled amplifiers. Each succeeding stage requires a higher voltage. The load and voltage divider resistors use a large amount of power and the biasing can become very complicated. In addition, it is difficult to match the impedance from stage to stage with direct coupling. (Impedance matching is covered a little later in this chapter.)

The direct-coupled amplifier is not very efficient and the losses increase as the number of stages increase. Because of the disadvantages, direct coupling is not used very often.

RC Coupling

The most commonly used coupling in amplifiers is RC coupling. An RC-coupling network is shown in figure 1-10.

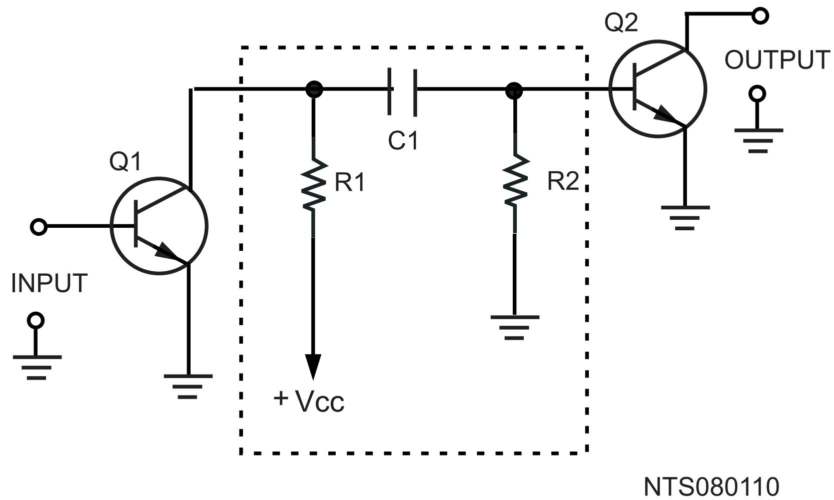


Figure 1-10.—RC-coupled transistor amplifier.

The network of R1, R2, and C1 enclosed in the dashed lines of the figure is the coupling network. You may notice that the circuitry for Q1 and Q2 is incomplete. That is intentional so that you can concentrate on the coupling network.

R1 acts as a load resistor for Q1 (the first stage) and develops the output signal of that stage. Do you remember how a capacitor reacts to ac and dc? The capacitor, C1, "blocks" the dc of Q1's collector, but "passes" the ac output signal. R2 develops this passed, or coupled, signal as the input signal to Q2 (the second stage). This arrangement allows the coupling of the signal while it isolates the biasing of each stage. This solves many of the problems associated with direct coupling.

RC coupling does have a few disadvantages. The resistors use dc power and so the amplifier has low efficiency. The capacitor tends to limit the low-frequency response of the amplifier and the amplifying device itself limits the high-frequency response. For audio amplifiers this is usually not a problem; techniques for overcoming these frequency limitations will be covered later in this module.

Before you move on to the next type of coupling, consider the capacitor in the RC coupling. You probably remember that capacitive reactance (X_C) is determined by the following formula:

$$X_C = \frac{1}{2\pi fC}$$

This explains why the low frequencies are limited by the capacitor. As frequency decreases, X_C increases. This causes more of the signal to be "lost" in the capacitor.

The formula for X_C also shows that the value of capacitance (C) should be relatively high so that capacitive reactance (X_C) can be kept as low as possible. So, when a capacitor is used as a coupling element, the capacitance should be relatively high so that it will couple the entire signal well and not reduce or distort the signal.

Impedance Coupling

Impedance coupling is very similar to RC coupling. The difference is the use of an impedance device (a coil) to replace the load resistor of the first stage.

Figure 1-11 shows an impedance-coupling network between two stages of amplification. L1 is the load for Q1 and develops the output signal of the first stage. Since the d.c. resistance of a coil is low, the efficiency of the amplifier stage is increased. The amount of signal developed in the output of the stage depends on the inductive reactance of L1. Remember the formula for inductive reactance:

$$X_L = 2\,\pi f L$$

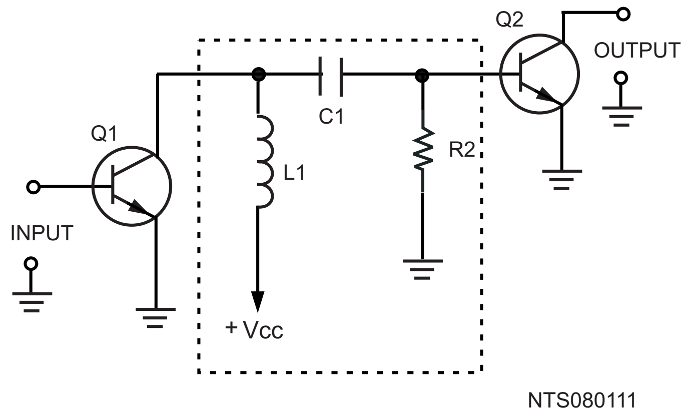


Figure 1-11.—Impedance-coupled transistor amplifier.

The formula shows that for inductive reactance to be large, either inductance or frequency or both must be high. Therefore, load inductors should have relatively large amounts of inductance and are most effective at high frequencies. This explains why impedance coupling is usually not used for audio amplifiers.

The rest of the coupling network (C1 and R1) functions just as their counterparts (C1 and R2) in the RC-coupling network. C1 couples the signal between stages while blocking the d.c. and R1 develops the input signal to the second stage (Q2).

Transformer Coupling

Figure 1-12 shows a transformer-coupling network between two stages of amplification. The transformer action of T1 couples the signal from the first stage to the second stage. In figure 1-12, the primary of T1 acts as the load for the first stage (Q1) and the secondary of T1 acts as the developing impedance for the second stage (Q2). No capacitor is needed because transformer action couples the signal between the primary and secondary of T1.

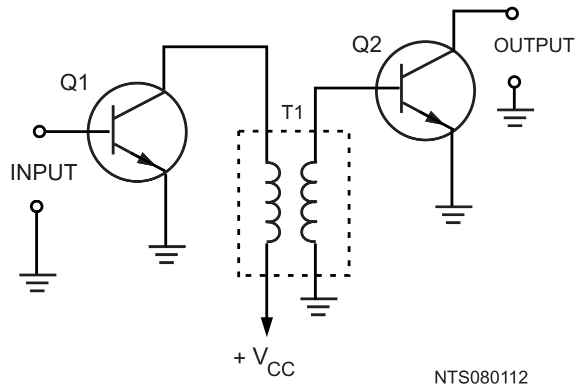


Figure 1-12.—Transformer-coupled transistor amplifier.

The inductors that make up the primary and secondary of the transformer have very little dc resistance, so the efficiency of the amplifiers is very high. Transformer coupling is very often used for the final output (between the final amplifier stage and the output device) because of the impedance-matching qualities of the transformer. The frequency response of transformer-coupled amplifiers is limited by the inductive reactance of the transformer just as it was limited in impedance coupling.

Q-12. What is the purpose of an amplifier-coupling network?

Q-13. What are four methods of coupling amplifier stages?

Q-14. What is the most common form of coupling?

Q-15. What type coupling is usually used to couple the output from a power amplifier?

Q-16. What type coupling would be most useful for an audio amplifier between the first and second stages?

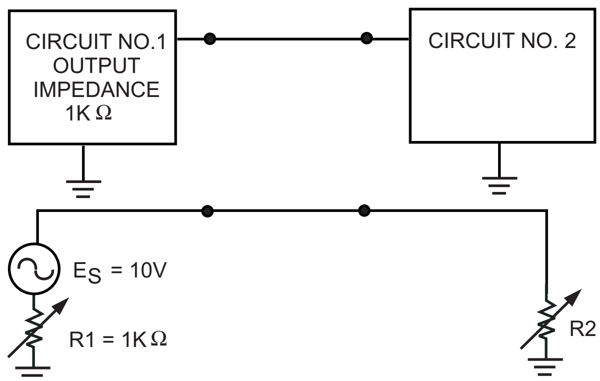
Q-17. What type of coupling is most effective at high frequencies?

IMPEDANCE CONSIDERATIONS FOR AMPLIFIERS

It has been mentioned that efficiency and impedance are important in amplifiers. The reasons for this may not be too clear. You have been shown that any amplifier is a current-control device. Now there are two other principles you should try to keep in mind. First, there is no such thing as "something for nothing" in electronics. That means every time you do something to a signal it costs something. It might mean a loss in fidelity to get high power. Some other compromise might also be made when a circuit is designed. Regardless of the compromise, every stage will require and use power. This brings up the second principle-do things as efficiently as possible. The improvement and design of electronic circuits is an attempt to do things as cheaply as possible, in terms of power, when all the other requirements (fidelity, power output, frequency range, etc.) have been met.

This brings us to efficiency. The most efficient device is the one that does the job with the least loss of power. One of the largest losses of power is caused by impedance differences between the output of one circuit and the input of the next circuit. Perhaps the best way to think of an impedance difference (mismatch) between circuits is to think of different-sized water pipes. If you try to connect a one-inch water pipe to a two-inch water pipe without an adapter you will lose water. You must use an adapter. A impedance-matching device is like that adapter. It allows the connection of two devices with different impedances without the loss of power.

Figure 1-13 shows two circuits connected together. Circuit number 1 can be considered as an a.c. source (E_S) whose output impedance is represented by a resistor (R1). It can be considered as an a.c. source because the output signal is an a.c. voltage and comes from circuit number 1 through the output impedance. The input impedance of circuit number 2 is represented by a resistor in series with the source. The resistance is shown as variable to show what will happen as the input impedance of circuit number 2 is changed.



R2	I	E_{R2}	P_{R1}	P_{R2}
100Ω	9mA	900mV	82.6mW	8.3mW
200 Ω	8.3mA	1.67V	69.4mW	13.9mW
500 Ω	6.7mA	3.33V	44.4mW	22.2mW
1kΩ	5mA	5V	25mW	25mW
2k Ω	3.3mA	6.67V	11.1mW	22.2mW
5k Ω	1.7mA	8.33V	2.7mW	13.9mW
10k Ω	.9mA	9.09V	.9mW	8.3mW

NTS0801I-11

Figure 1-13.—Effect of impedance matching in the coupling of two circuits.

The chart below the circuit shows the effect of a change in the input impedance of circuit number 2 (R2) on current (I), signal voltage developed at the input of circuit number 2 (E_{R2}), the power at the output of circuit number 1 (P_{R1}), and the power at the input to circuit number 2 (P_{R2}).

Two important facts are brought out in this chart. First, the power at the input to circuit number 2 is greatest when the impedances are equal (matched). The power is also equal at the output of circuit number 1 and the input of circuit number 2 when the impedance is matched. The second fact is that the largest <u>voltage</u> signal is developed at the input to circuit number 2 when its input impedance is much larger than the output impedance of circuit number 1. However, the power at the input of circuit number 2 is very low under these conditions. So you must decide what conditions you want in coupling two circuits together and select the components appropriately.

Two important points to remember about impedance matching are as follows. (1) Maximum power transfer requires matched impedance. (2) To get maximum voltage at the input of a circuit requires an intentional impedance mismatch with the circuit that is providing the input signal.

Impedance Characteristics of Amplifier Configurations

Now that you have seen the importance of impedance matching the stages in an electronic device, you may wonder what impedance characteristics an amplifier has. The input and output impedances of a transistor amplifier depend upon the configuration of the transistor. In *Module 7, Introduction to Solid-State Devices and Power Supplies*, you were introduced to the three transistor configurations; the common emitter, the common base, and the common collector. Examples of these configurations and their impedance characteristics are shown in figure 1-14.

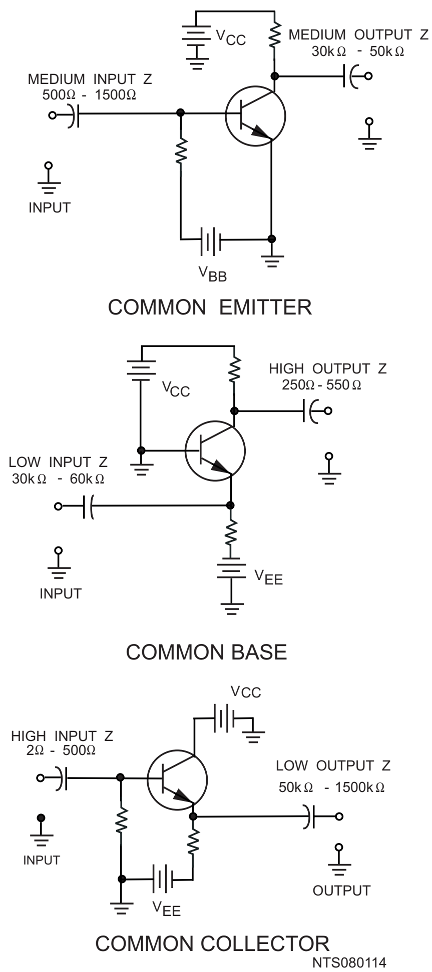


Figure 1-14.—Transistor amplifier configurations and their impedance characteristics.

NOTE: Only approximate impedance values are shown. This is because the exact impedance values will vary from circuit to circuit. The impedance of any particular circuit depends upon the device (transistor) and the other circuit components. The value of impedance can be computed by dividing the signal voltage by the signal current. Therefore:

$$\text{Input Signal Impedance} = \frac{\text{Input Signal Voltage}}{\text{Input Signal Current}}$$

and

$$\text{Output Signal Impedance} = \frac{\text{Output Signal Voltage}}{\text{Output Signal Current}}$$

The common-emitter configuration provides a medium input impedance and a medium output impedance. The common-base configuration provides a low input impedance and a high output impedance. The common-collector configuration provides a high input impedance and a low output impedance. The common-collector configuration is often used to provide impedance matching between a high output impedance and a low input impedance.

If the amplifier stage is transformer coupled, the turns ratio of the transformer can be selected to provide impedance matching. In *NEETS Module 2*, *Introduction to Alternating Current and Transformers,* you were shown the relationship between the turns ratio and the impedance ratio in a transformer. The relationship is expressed in the following formula:

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}}$$

Where:
N_P = Number of turns in the primary
N_S = Number of turns in the secondary
Z_P = Impedance of the primary
Z_S = Impedance of the secondary

As you can see, impedance matching between stages can be accomplished by a combination of the amplifier configuration and the components used in the amplifier circuit.

Q-18. What impedance relationship between the output of one circuit and the input of another circuit will provide the maximum power transfer?

Q-19. If maximum current is desired at the input to a circuit, should the input impedance of that circuit be lower than, equal to, or higher than the output impedance of the previous stage?

Q-20. What are the input- and output-impedance characteristics of the three transistor configurations?

Q-21. What transistor circuit configuration should be used to match a high output impedance to a low input impedance?

Q-22. What type of coupling is most useful for impedance matching?

AMPLIFIER FEEDBACK

Perhaps you have been around a public address system when a squeal or high-pitched noise has come from the speaker. Someone will turn down the volume and the noise will stop. That noise is an indication that the amplifier (at least one stage of amplification) has begun oscillating. Oscillation is covered in detail in *NEETS Module 9*, *Introduction to Wave-Generation and Wave-Shaping Circuits.* For now, you need only realize that the oscillation is caused by a small part of the signal from the amplifier output being sent back to the input of the amplifier. This signal is amplified and again sent back to the input where it is amplified again. This process continues and the result is a loud noise out of the speaker. The process of sending part of the output signal of an amplifier back to the input of the amplifier is called FEEDBACK.

There are two types of feedback in amplifiers. They are POSITIVE FEEDBACK, also called REGENERATIVE FEEDBACK, and NEGATIVE FEEDBACK, also called DEGENERATIVE FEEDBACK. The difference between these two types is whether the feedback signal is in phase or out of phase with the input signal.

Positive feedback occurs when the feedback signal is in phase with the input signal. Figure 1-15 shows a block diagram of an amplifier with positive feedback. Notice that the feedback signal is in phase with the input signal. This means that the feedback signal will add to or "regenerate" the input signal. The result is a larger amplitude output signal than would occur without the feedback. This type of feedback is what causes the public address system to squeal as described above.

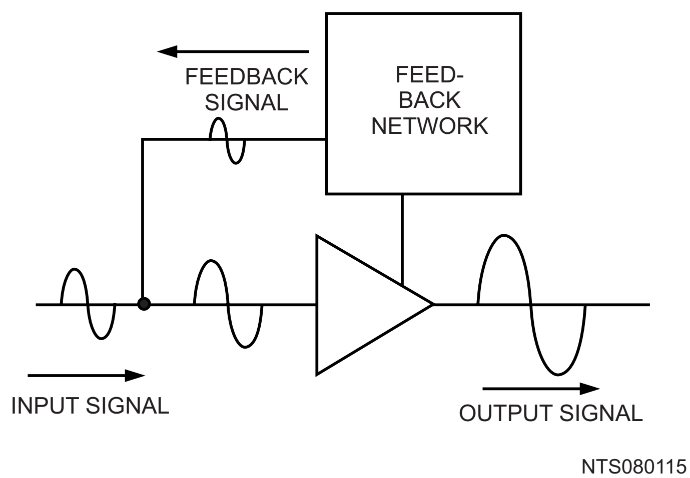


Figure 1-15.—Positive feedback in an amplifier.

Figure 1-16 is a block diagram of an amplifier with negative feedback. In this case, the feedback signal is out of phase with the input signal. This means that the feedback signal will subtract from or "degenerate" the input signal. This results in a lower amplitude output signal than would occur without the feedback.

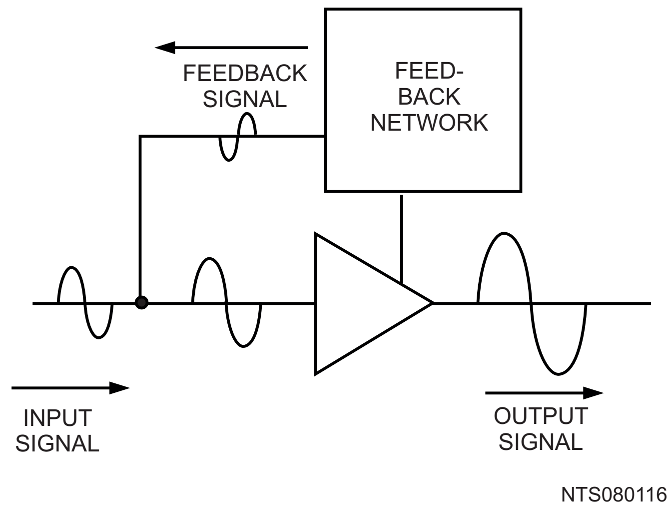


Figure 1-16.—Negative feedback in an amplifier.

Sometimes feedback that is not desired occurs in an amplifier. This happens at high frequencies and limits the high-frequency response of an amplifier. Unwanted feedback also occurs as the result of some circuit components used in the biasing or coupling network. The usual solution to unwanted feedback is a feedback network of the opposite type. For example, a positive feedback network would counteract unwanted, negative feedback.

Feedback is also used to get the ideal input signal. Normally, the maximum output signal is desired from an amplifier. The amount of the output signal from an amplifier is dependent on the amount of the input signal. However, if the input signal is too large, the amplifying device will be saturated and/or cut off during part of the input signal. This causes the output signal to be distorted and reduces the fidelity of the amplifier. Amplifiers must provide the proper balance of gain and fidelity.

Figure 1-17 shows the way in which feedback can be used to provide the maximum output signal without a loss in fidelity. In view A, an amplifier has good fidelity, but less gain than it could have. By adding some positive feedback, as in view B, the gain of the stage is increased. In view C, an amplifier has so much gain and such a large input signal that the output signal is distorted. This distortion is caused by the amplifying device becoming saturated and cutoff. By adding a negative feedback system, as in view D, the gain of the stage is decreased and the fidelity of the output signal improved.

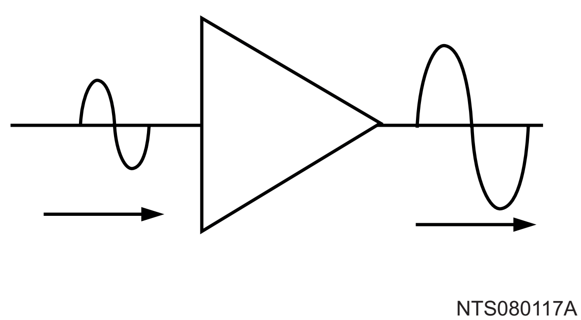


Figure 1-17A.-Feedback uses in amplifiers.

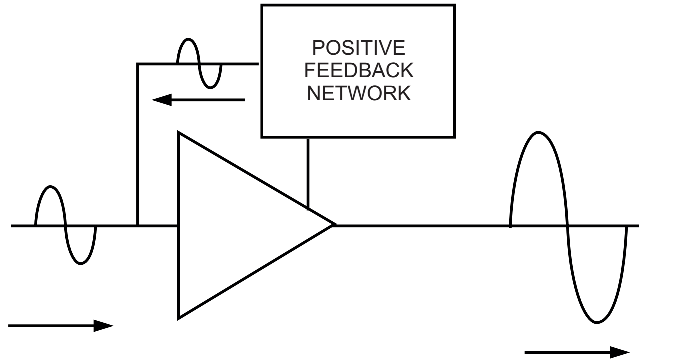


Figure 1-17B.—Feedback uses in amplifiers.

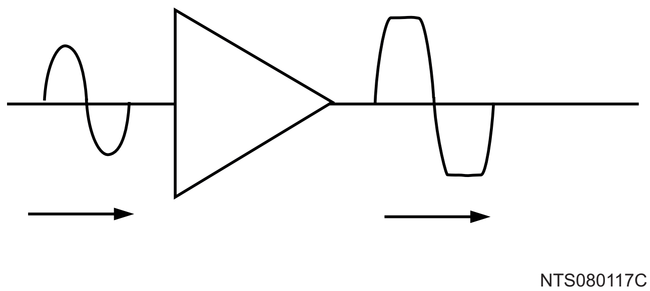


Figure 1-17C.—Feedback uses in amplifiers.

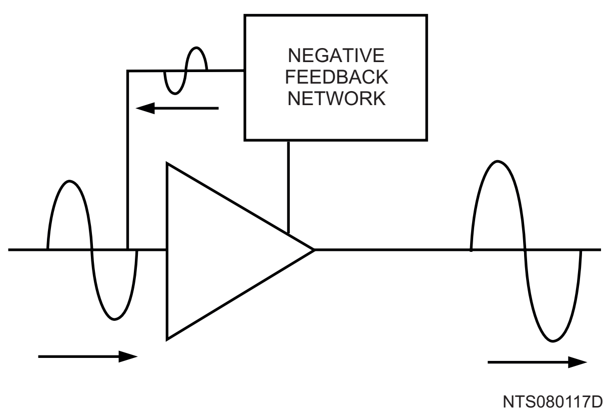


Figure 1-17D.—Feedback uses in amplifiers.

Positive and negative feedback are accomplished in many ways, depending on the reasons requiring the feedback. A few of the effects and methods of accomplishing feedback are presented next.

Positive Feedback

As you have seen, positive feedback is accomplished by adding part of the output signal in phase with the input signal. In a common-base transistor amplifier, it is fairly simple to provide positive feedback. Since the input and output signals are in phase, you need only couple part of the output signal back to the input. This is shown in figure 1-18.

The feedback network in this amplifier is made up of R2 and C2. The value of C2 should be large so that the capacitive reactance (X_C) will be low and the capacitor will couple the signal easily. (This is also the case with the input and output coupling capacitors C1 and C3.) The resistive value of R2 should be large to limit the amount of feedback signal and to ensure that the majority of the output signal goes on to the next stage through C3.

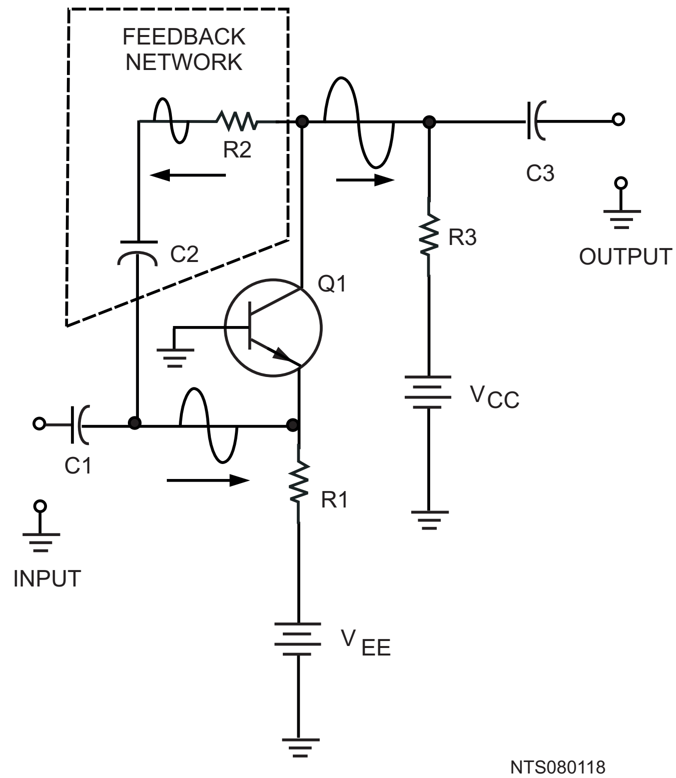


Figure 1-18.—Positive feedback in a transistor amplifier.

A more common configuration for transistor amplifiers is the common-emitter configuration. Positive feedback is a little more difficult with this configuration because the input and output signals are 180º out of phase. Positive feedback can be accomplished by feeding a portion of the output signal of the second stage back to the input of the first stage. This arrangement is shown in figure 1-19.

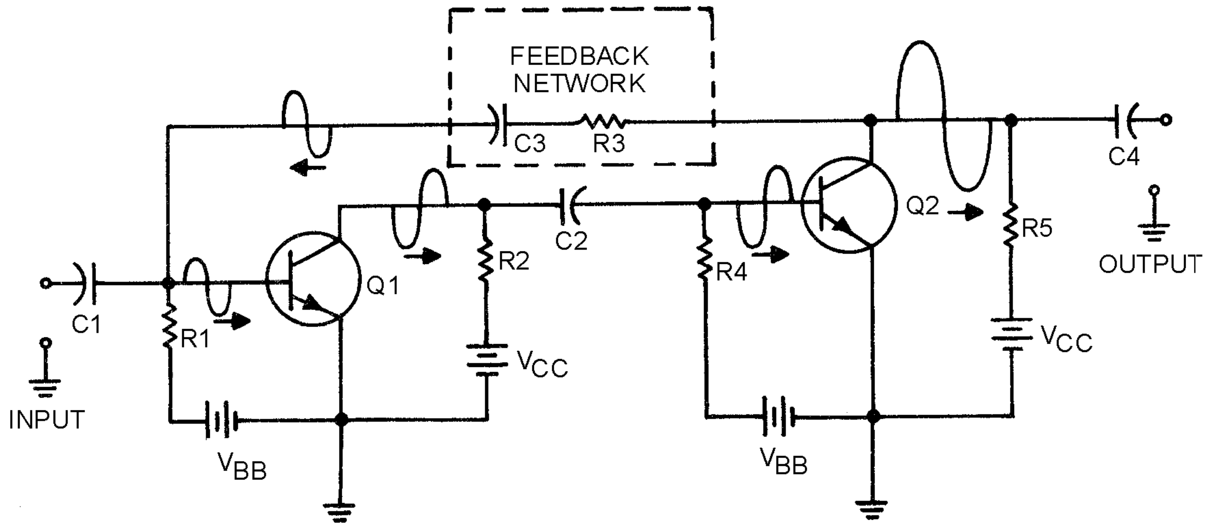


Figure 1-19.—Positive feedback in two stages of transistor amplification.

The figure shows that each stage of amplification has a 180º phase shift. This means that the output signal of Q2 will be in phase with the input signal to Q1. A portion of the output signal of Q2 is coupled back to the input of Q1 through the feedback network of C3 and R3. R3 should have a large resistance to limit the amount of signal through the feedback network. C3 should have a large capacitance so the capacitive reactance is low and the capacitor will couple the signal easily.

Sometimes positive feedback is used to eliminate the effects of negative feedback that are caused by circuit components. One way in which a circuit component can cause negative feedback is shown in figure 1-20.

In view (A) a common-emitter transistor amplifier is shown. An emitter resistor (R2) has been placed in this circuit to provide proper biasing and temperature stability. An undesired effect of this resistor is the development of a signal at the emitter in phase with the input signal on the base. This signal is caused by the changing current through the emitter resistor (R2) as the current through the transistor changes. You might think that this signal on the emitter is a form of positive feedback since it is in phase with the input signal. But the emitter signal is really negative feedback. Current through the transistor is controlled by the base-to-emitter bias. If both the base and emitter become more positive by the same amount at the same time, current will not increase. It is the difference between the base and emitter voltages that controls the current flow through the transistor.

To eliminate this negative feedback caused by the emitter resistor, some way must be found to remove the signal from the emitter. If the signal could be coupled to ground (decoupled) the emitter of the transistor would be unaffected. That is exactly what is done. A DECOUPLING CAPACITOR (C3 in view B) is placed between the emitter of Q1 and ground (across the emitter resistor). This capacitor should have a high capacitance so that it will pass the signal to ground easily. The decoupling capacitor (C3) should have the same qualities as the coupling capacitors (C1 and C2) of the circuit. Decoupling capacitors are also called bypass capacitors.

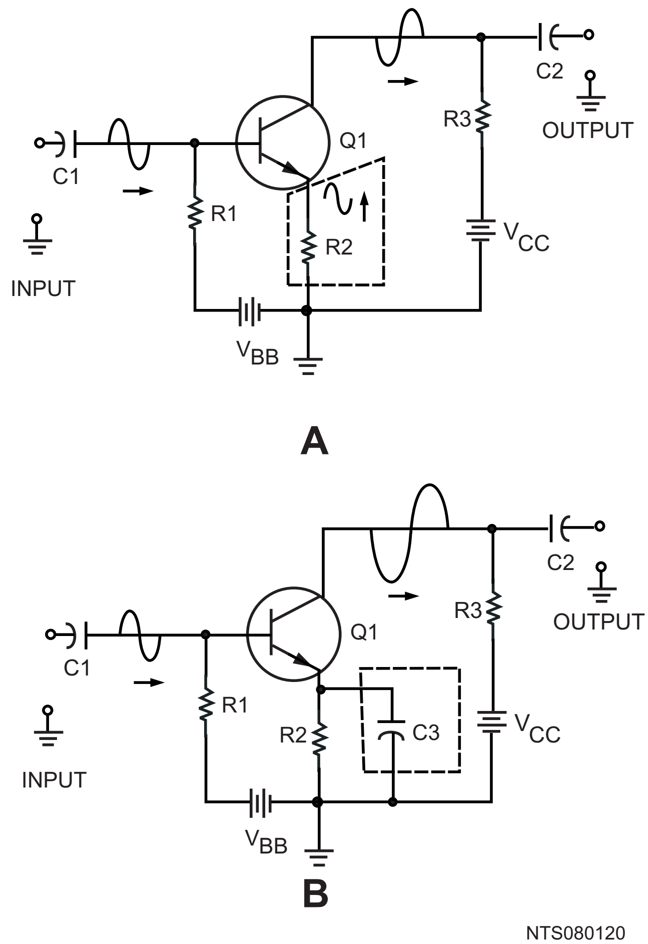


Figure 1-20A-B.—Decoupling (bypass) capacitor in a transistor amplifier.

Regardless of the method used to provide positive feedback in a circuit, the purpose is to increase the output signal amplitude.

Negative Feedback

Negative feedback is accomplished by adding part of the output signal out of phase with the input signal. You have seen that an emitter resistor in a common-emitter transistor amplifier will develop a negative feedback signal. Other methods of providing negative feedback are similar to those methods used to provide positive feedback. The phase relationship of the feedback signal and the input signal is the only difference.

Figure 1-21 shows negative feedback in a common-emitter transistor amplifier. The feedback network of C2 and R2 couples part of the output signal of Q1 back to the input. Since the output signal is 180º out of phase with the input signal, this causes negative feedback.

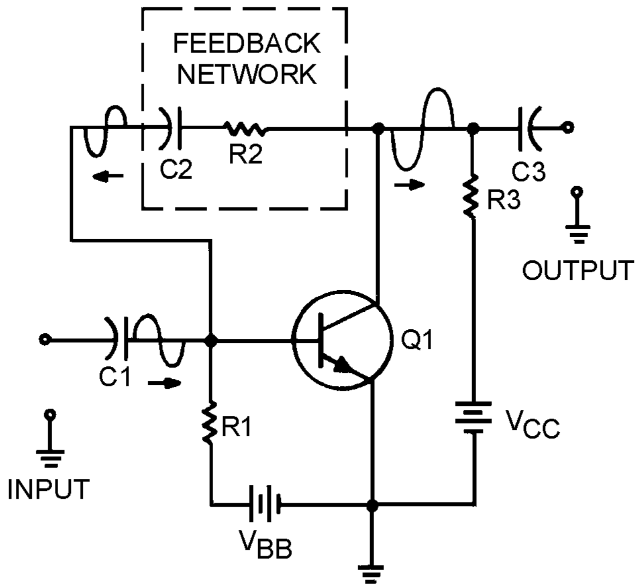


Figure 1-21.—Negative feedback in a transistor amplifier.

Negative feedback is used to improve fidelity of an amplifier by limiting the input signal. Negative feedback can also be used to increase the frequency response of an amplifier. The gain of an amplifier decreases when the limit of its frequency response is reached. When negative feedback is used, the feedback signal decreases as the output signal decreases. At the limits of frequency response of the amplifier the smaller feedback signal means that the effective gain (gain with feedback) is increased. This will improve the frequency response of the amplifier.

Q-23. What is feedback?

Q-24. What are the two types of feedback?

Q-25. What type feedback provides increased amplitude output signals?

Q-26. What type feedback provides the best fidelity?

Q-27. If the feedback signal is out of phase with the input signal, what type feedback is provided?

Q-28. What type feedback is provided by an unbypassed emitter resistor in a common-emitter transistor amplifier?

AUDIO AMPLIFIERS

An audio amplifier has been described as an amplifier with a frequency response from 15 Hz to 20 kHz. The frequency response of an amplifier can be shown graphically with a frequency response curve. Figure 1-22 is the ideal frequency response curve for an audio amplifier. This curve is practically "flat" from 15 Hz to 20 kHz. This means that the gain of the amplifier is equal between 15 Hz and 20 kHz. Above 20 kHz or below 15 Hz the gain decreases or "drops off" quite rapidly. The frequency response of an amplifier is determined by the components in the circuit.

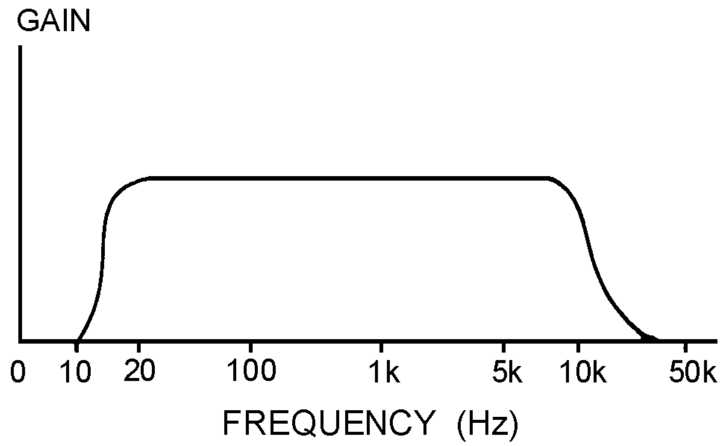


Figure 1-22.—Ideal frequency response curve for an audio amplifier.

The difference between an audio amplifier and other amplifiers is the frequency response of the amplifier. In the next chapter of this module you will be shown the techniques and components used to change and extend the frequency response of an amplifier.

The transistor itself will respond quite well to the audio frequency range. No special components are needed to extend or modify the frequency response.

You have already been shown the purpose of all the components in a transistor audio amplifier. In this portion of the chapter, schematic diagrams of several audio amplifiers will be shown and the functions of each of the components will be discussed.

SINGLE-STAGE AUDIO AMPLIFIERS

The first single-stage audio amplifier is shown in figure 1-23. This circuit is a class A, common-emitter, RC-coupled, transistor, audio amplifier. C1 is a coupling capacitor that couples the input signal to the base of Q1. R1 is used to develop the input signal and provide bias for the base of Q1. R2 is used to bias the emitter and provide temperature stability for Q1. C2 is used to provide decoupling (positive feedback) of the signal that would be developed by R2. R3 is the collector load for Q1 and develops the output signal. C3 is a coupling capacitor that couples the output signal to the next stage. V_{CC} represents the collector-supply voltage. Since the transistor is a common-emitter configuration, it provides voltage amplification. The input and output signals are 180º out of phase. The input and output impedance are both medium.

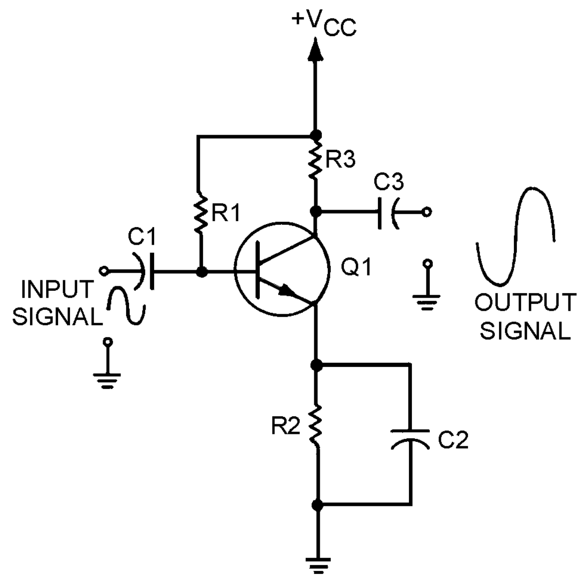


Figure 1-23.—Transistor audio amplifier.

There is nothing new presented in this circuit. You should understand all of the functions of the components in this circuit. If you do not, look back at the various sections presented earlier in this chapter.

The second single-stage audio amplifier is shown in figure 1-24. This circuit is a class A, common-source, RC-coupled, FET, audio amplifier. C1 is a coupling capacitor which couples the input signal to the gate of Q1. R1 is used to develop the input signal for the gate of Q1. R2 is used to bias the source of Q1. C2 is used to decouple the signal developed by R2 (and keep it from affecting the source of Q1). R3 is the drain load for Q1 and develops the output signal. C3 couples the output signal to the next stage. V_{DD} is the supply voltage for the drain of Q1. Since this is a common-source configuration, the input and output signals are 180º out of phase.

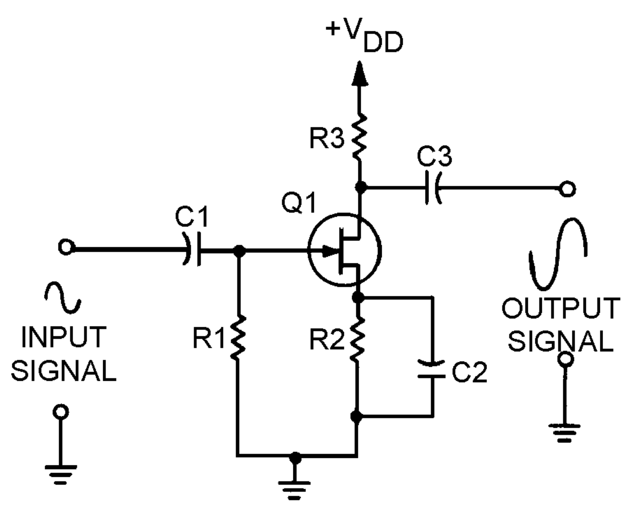


Figure 1-24.—FET audio amplifier.

If you do not remember how a FET works, refer to *NEETS Module 7 Introduction to Solid-State Devices and Power Supplies.*

The third single-stage audio amplifier is shown in figure 1-25. This is a class A, common-emitter, transformer-coupled, transistor, audio amplifier. The output device (speaker) is shown connected to the secondary winding of the transformer. C1 is a coupling capacitor which couples the input signal to the base of Q1. R1 develops the input signal. R2 is used to bias the emitter of Q1 and provides temperature stability. C2 is a decoupling capacitor for R2. R3 is used to bias the base of Q1. The primary of T1 is the collector load for Q1 and develops the output signal. T1 couples the output signal to the speaker and provides impedance matching between the output impedance of the transistor (medium) and the impedance of the speaker (low).

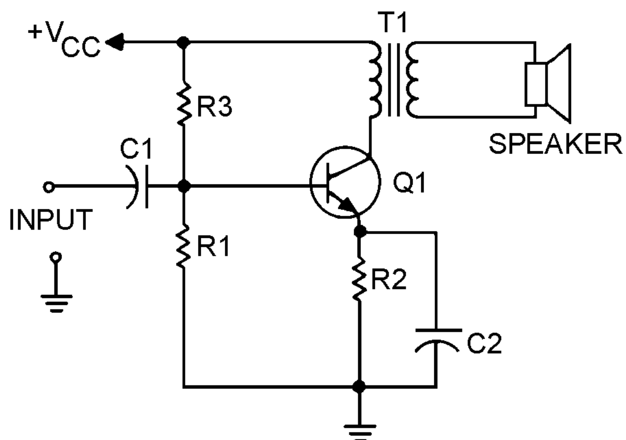


Figure 1-25.—Single-stage audio amplifier.

PHASE SPLITTERS

Sometimes it is necessary to provide two signals that are equal in amplitude but 180º out of phase with each other. (You will see one use of these two signals a little later in this chapter.) The two signals can be provided from a single input signal by the use of a PHASE SPLITTER. A phase splitter is a device that produces two signals that differ in phase from each other from a single input signal. Figure 1-26 is a block diagram of a phase splitter.

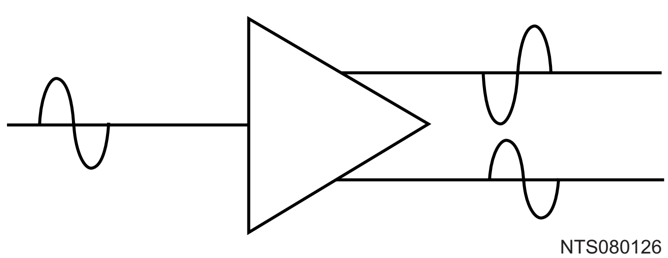


Figure 1-26.—Block diagram of a phase splitter.

One way in which a phase splitter can be made is to use a center-tapped transformer. As you may remember from your study of transformers, when the transformer secondary winding is center-tapped, two equal amplitude signals are produced. These signals will be 180º out of phase with each other. So a transformer with a center-tapped secondary fulfills the definition of a phase splitter.

A transistor amplifier can be configured to act as a phase splitter. One method of doing this is shown in figure 1-27.

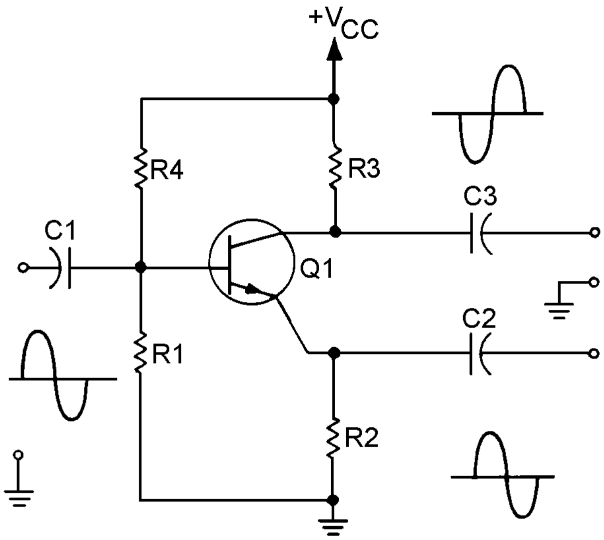


Figure 1-27.—Single-stage transistor phase splitter.

C1 is the input signal coupling capacitor and couples the input signal to the base of Q1. R1 develops the input signal. R2 and R3 develop the output signals. R2 and R3 are equal resistances to provide equal amplitude output signals. C2 and C3 couple the output signals to the next stage. R4 is used to provide proper bias for the base of Q1.

This phase splitter is actually a single transistor combining the qualities of the common-emitter and common-collector configurations. The output signals are equal in amplitude of the input signal, but are 180º out of phase from each other.

If the output signals must be larger in amplitude than the input signal, a circuit such as that shown in figure 1-28 will be used.

Figure 1-28 shows a two-stage phase splitter. C1 couples the input signal to the base of Q1. R1 develops the input signal and provides bias for the base of Q1. R2 provides bias and temperature stability for Q1. C2 decouples signals from the emitter of Q1. R3 develops the output signal of Q1. Since Q1 is configured as a common-emitter amplifier, the output signal of Q1 is 180º out of phase with the input signal and larger in amplitude. C3 couples this output signal to the next stage through R4. R4 allows only a small portion of this output signal to be applied to the base of Q2. R5 develops the input signal and provides bias for the base of Q2. R6 is used for bias and temperature stability for Q2. C4 decouples signals from the emitter of Q2. R7 develops the output signal from Q2. Q2 is configured as a common-emitter amplifier, so the output signal is 180º out of phase with the input signal to Q2 (output signal from Q1). The input signal to Q2 is 180º out of phase with the original input signal, so the output from Q2 is in phase with the original input signal. C5 couples this output signal to the next stage. So the circuitry shown provides two output signals that are 180º out of phase with each other. The output signals are equal in amplitude with each other but larger than the input signal.

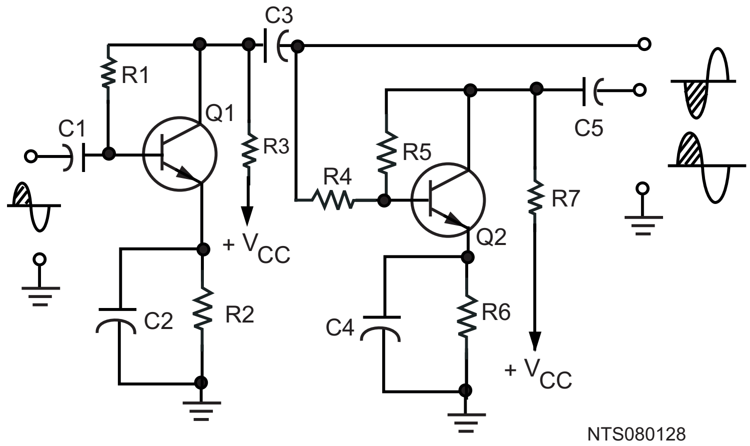


Figure 1-28.—Two-stage transistor phase splitter.

Q-29. What is a phase splitter?

PUSH-PULL AMPLIFIERS

One use of phase splitters is to provide input signals to a single-stage amplifier that uses two transistors. These transistors are configured in such a way that the two outputs, 180º out of phase with each other, combine. This allows more gain than one transistor could supply by itself. This "push-pull" amplifier is used where high power output and good fidelity are needed: receiver output stages, public address amplifiers, and AM modulators, for example.

The circuit shown in figure 1-29 is a class A transistor push-pull amplifier, but class AB or class B operations can be used. Class operations were discussed in an earlier topic. The phase splitter for this amplifier is the transformer T1, although one of the phase splitters shown earlier in this topic could be used. R1 provides the proper bias for Q1 and Q2. The tapped secondary of T1 develops the two input signals for the bases of Q1 and Q2. Half of the original input signal will be amplified by Q-1, the other half by Q-2. T2 combines (couples) the amplified output signal to the speaker and provides impedance matching.

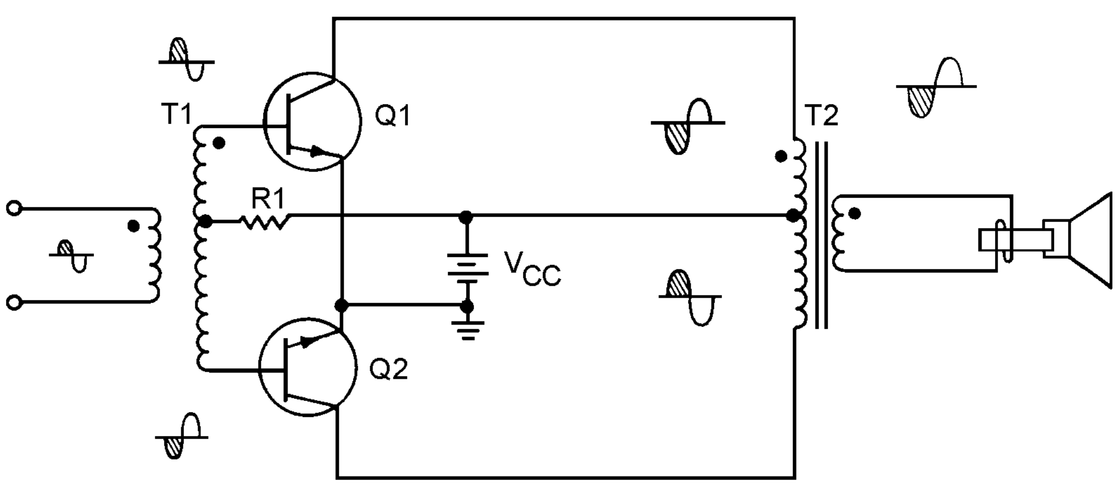


Figure 1-29.—Class A transistor push-pull amplifier.

Q-30. What is one use for a splitter?

Q-31. What is a common use for a push-pull amplifier?

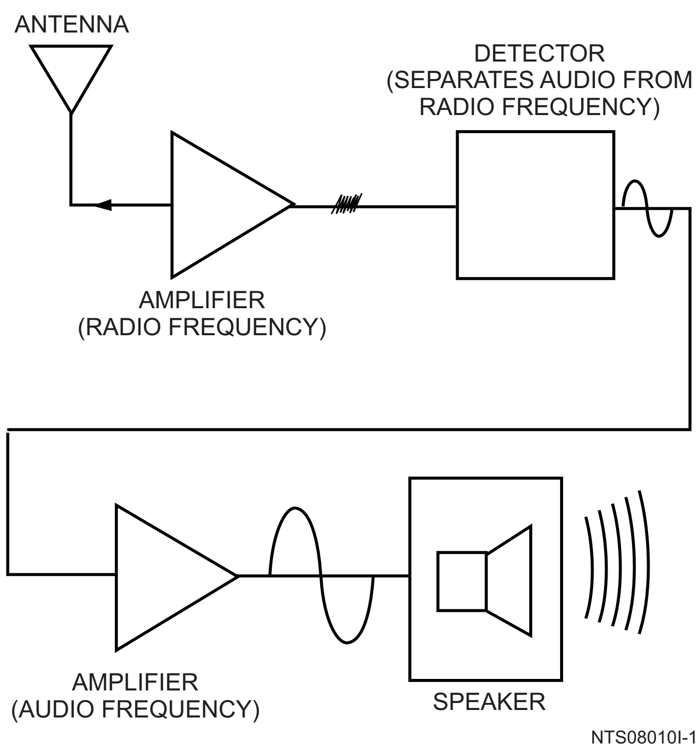
Q-32. What is the advantage of a push-pull amplifier?

Q-33. What class of operation can be used with a push-pull amplifier to provide good fidelity output signals?

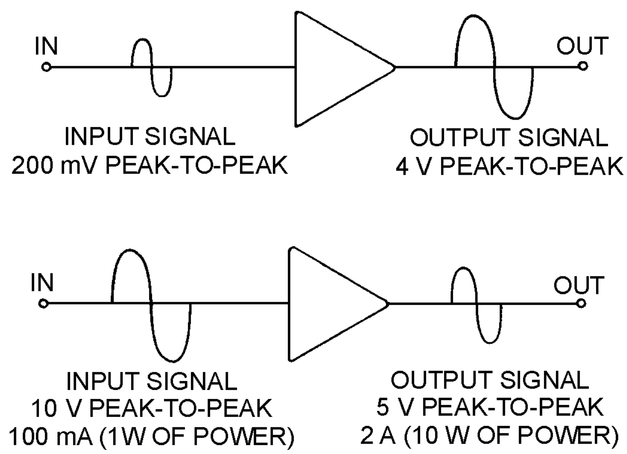
SUMMARY

This chapter has presented some general information that applies to all amplifiers, as well as some specific information about transistor and audio amplifiers. All of this information will be useful to you in the next chapter of this module and in your future studies of electronics.

An **AMPLIFIER** is a device that enables an input signal to control an output signal. The output signal will have some (or all) of the characteristics of the input signal but will generally be larger than the input signal in terms of voltage, current, or power. A basic line diagram of an amplifier is shown below.

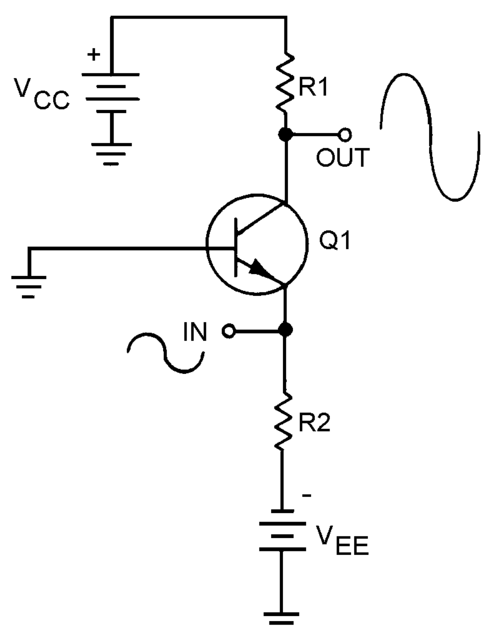


Amplifiers are classified by FUNCTION and FREQUENCY RESPONSE. Function refers to an amplifier being a VOLTAGE AMPLIFIER or a POWER AMPLIFIER. Voltage amplifiers provide voltage amplification and power amplifiers provide power amplification. The frequency response of an amplifier can be described by classifying the amplifier as an AUDIO AMPLIFIER, RF AMPLIFIER, or VIDEO (WIDE-BAND) AMPLIFIER. Audio amplifiers have frequency response in the range of 15 Hz to 20 kHz. An rf amplifier has a frequency response in the range of 10 kHz to 100,000 MHz. A video (wide-band) amplifier has a frequency response of 10 Hz to 6 MHz.

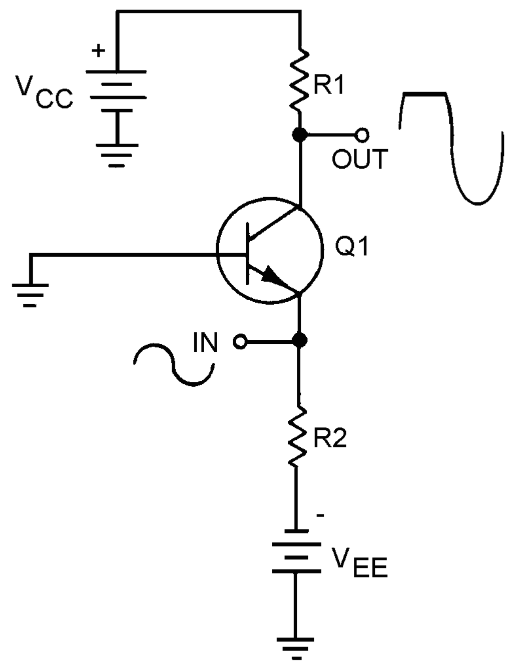


The **CLASS OF OPERATION** of a transistor amplifier is determined by the percent of time that current flows through the transistor in relation to the input signal.

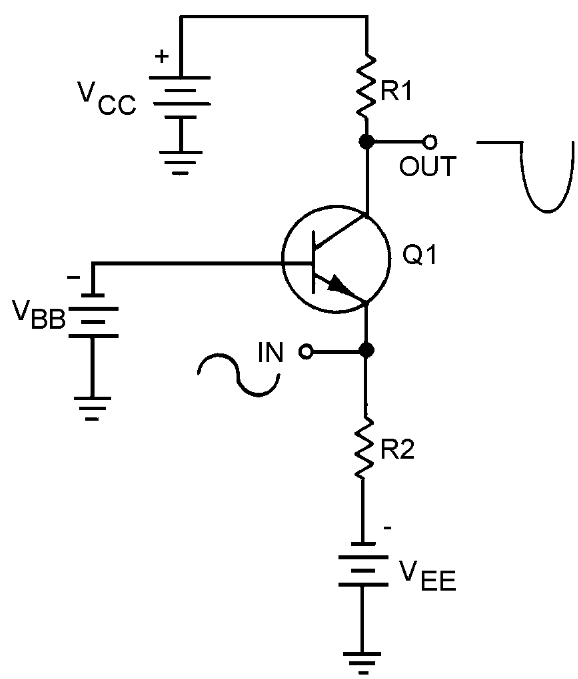
In **CLASS A OPERATION**, transistor current flows for 100% (360º) of the input signal. Class A operation is the least efficient class of operation, but provides the best fidelity.



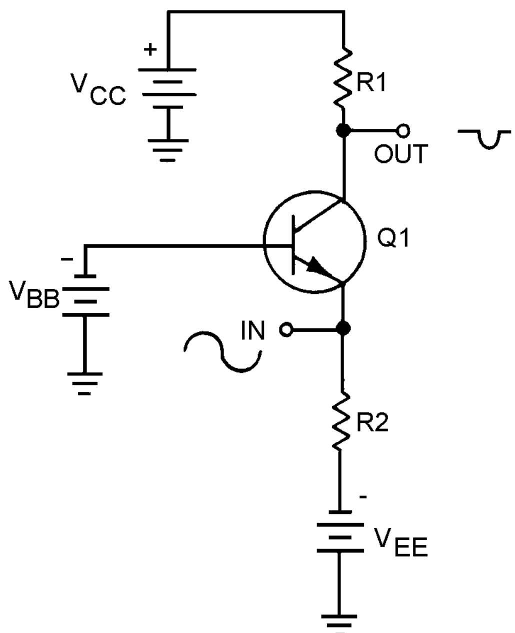
In **CLASS AB OPERATION**, transistor current flows for more than 50% but less than 100% of the input signal.



In **CLASS B OPERATION**, transistor current flows for 50% of the input signal.

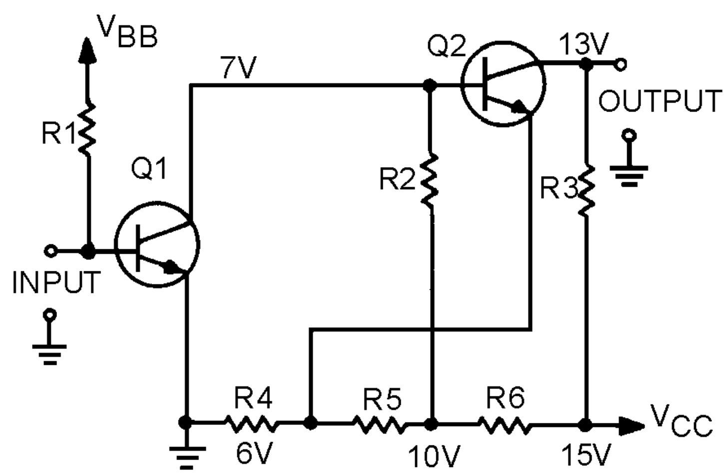


In **CLASS C OPERATION**, transistor current flows for less than 50% of the input signal. Class C operation is the most efficient class of operation.

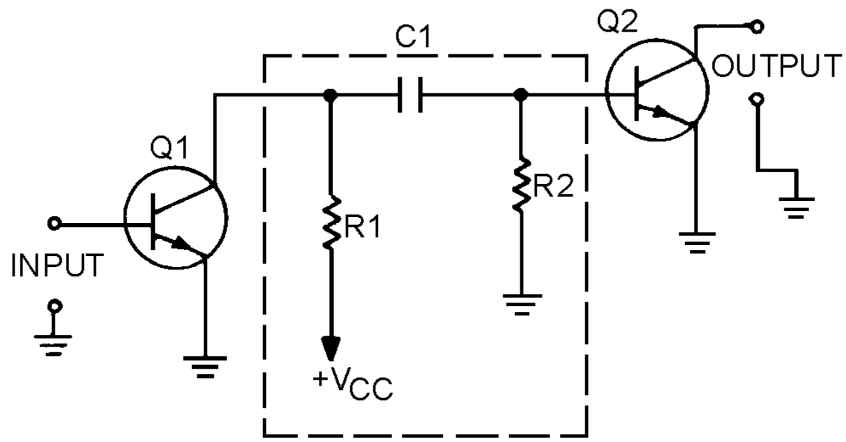


COUPLING is used to transfer a signal from one stage to another.

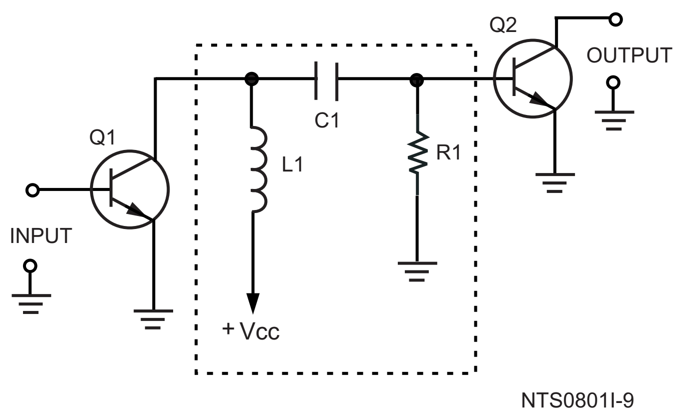
DIRECT COUPLING is the connection of the output of one stage directly to the input of the next stage. This method is not used very often due to the complex power supply requirements and impedance-matching problems.



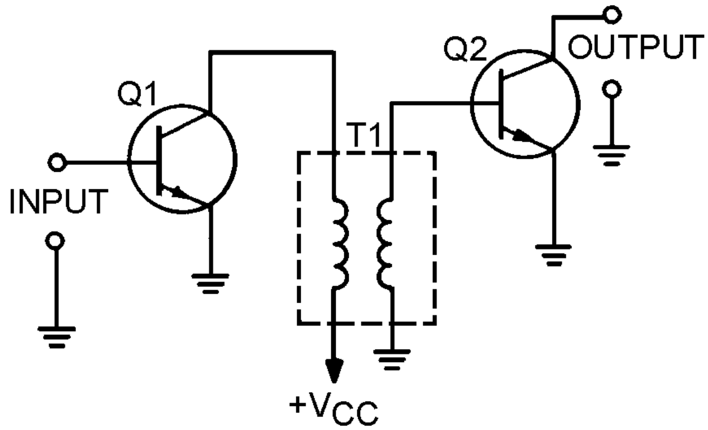
RC COUPLING is the most common method of coupling and uses a coupling capacitor and signal-developing resistors.



IMPEDANCE COUPLING uses a coil as a load for the first stage but otherwise functions just as RC coupling. Impedance coupling is used at high frequencies.

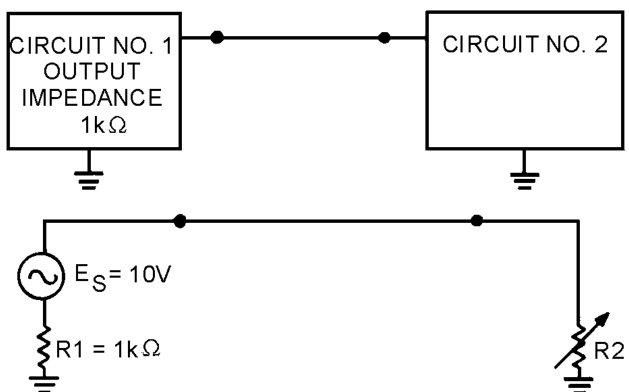


TRANSFORMER COUPLING uses a transformer to couple the signal from one stage to the next. Transformer coupling is very efficient and the transformer can aid in impedance matching.



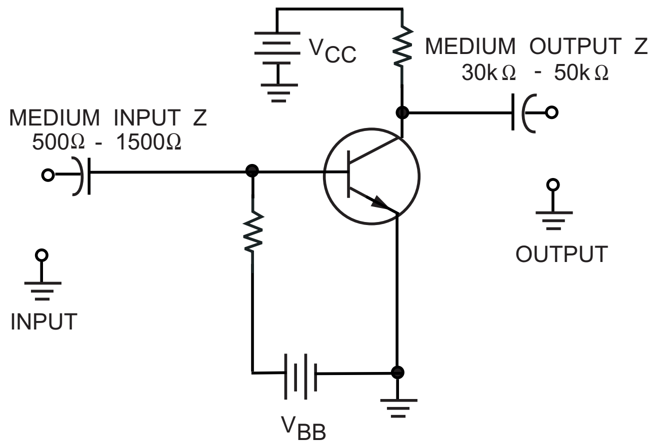
MAXIMUM POWER TRANSFER occurs between two circuits when the output impedance of the first circuit matches the input impedance of the second circuit.

A **MAXIMUM VOLTAGE INPUT SIGNAL** is present when the input impedance of the second circuit is larger than the output impedance of the first circuit (mismatched).



R2	I	E_{R2}	P_{R1}	P_{R2}
100Ω	9mA	900mV	82.6mW	8.3mW
200Ω	8.3mA	1.67V	69.4mW	13.9mW
500Ω	6.7mA	3.33V	44.4mW	22.2mW
1kΩ	5mA	5V	25mW	25mW
2kΩ	3.3mA	6.67V	11.1mW	22.2mW
5kΩ	1.7mA	8.33V	2.7mW	13.9mW
10kΩ	.9mA	9.09V	.9mW	8.3mW

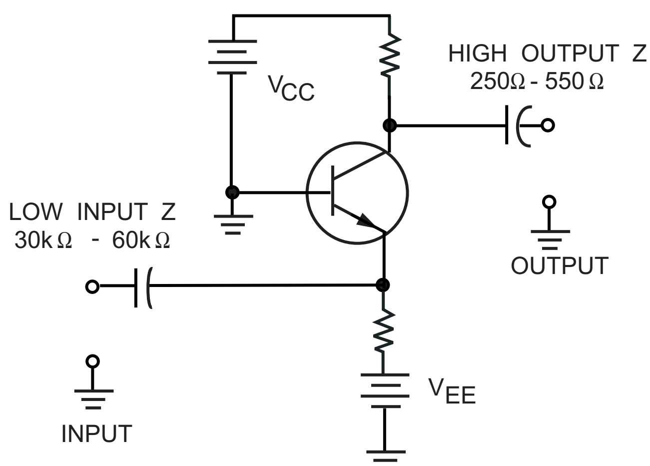
The **COMMON-EMITTER** configuration of a transistor amplifier has MEDIUM INPUT and MEDIUM OUTPUT IMPEDANCE.



COMMON EMITTER

NTS0801I-12

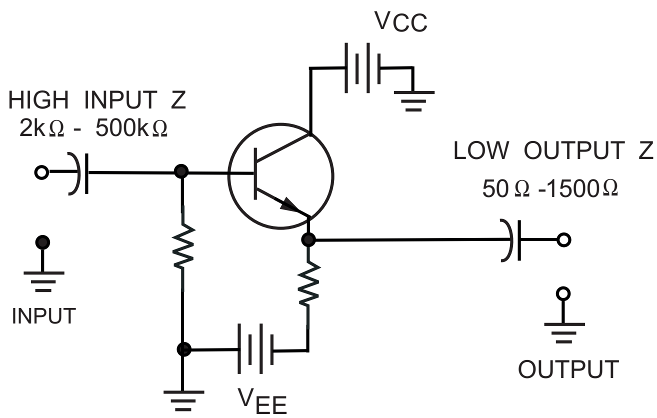
The **COMMON-BASE** configuration of a transistor amplifier has LOW INPUT and HIGH OUTPUT IMPEDANCE.



COMMON BASE

NTS0801I-13

The **COMMON-COLLECTOR (EMITTER FOLLOWER)** configuration of a transistor amplifier as HIGH INPUT and LOW OUTPUT IMPEDANCE.

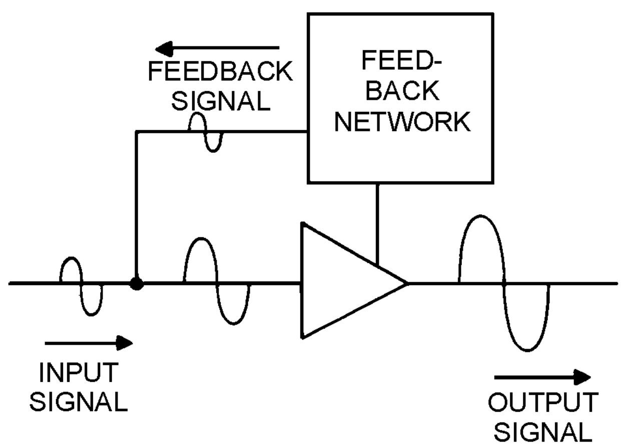


COMMON COLLECTOR

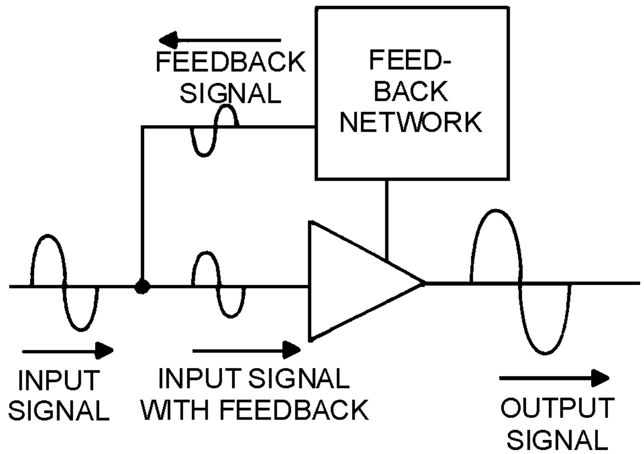
NTS0801I-14

FEEDBACK is the process of coupling a portion of the output signal back to the input of an amplifier.

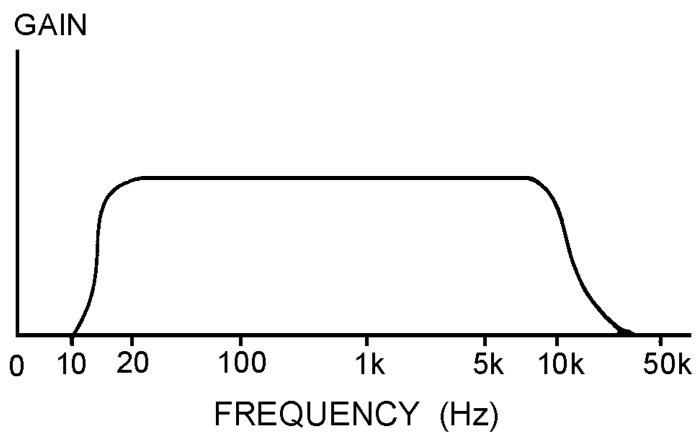
POSITIVE (REGENERATIVE) FEEDBACK is provided when the feedback signal is in phase with the input signal. Positive feedback increases the gain of an amplifier.



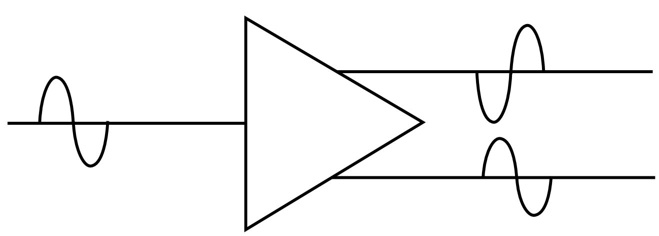
NEGATIVE (DEGENERATIVE) FEEDBACK is provided when the feedback signal is 180º out of phase with the input signal. Negative feedback decreases the gain of an amplifier but improves fidelity and may increase the frequency response of the amplifier.



The **IDEAL FREQUENCY RESPONSE** of an audio amplifier is equal gain from 15 Hz to 20 kHz with very low gain outside of those limits.

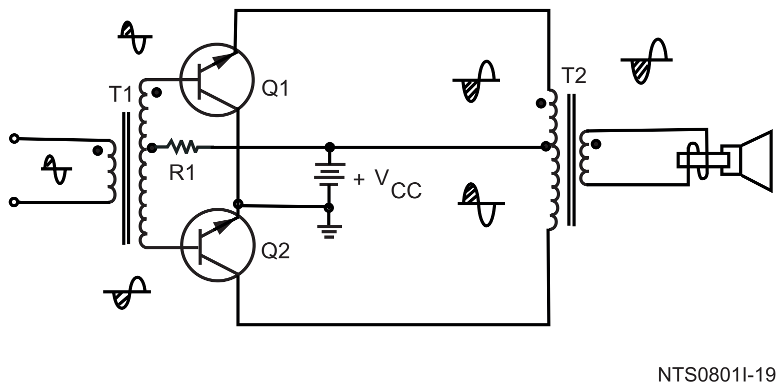


A **PHASE SPLITTER** provides two output signals that are equal in amplitude but different in phase from a single input signal. Phase splitters are often used to provide input signals to a push-pull amplifier.



NTS0801I-18

A **PUSH-PULL AMPLIFIER** uses two transistors whose output signals are added together to provide a larger gain (usually a power gain) than a single transistor could provide. Push-pull amplifiers can be operated class A, class AB or class B.



ANSWERS TO QUESTIONS Q1. THROUGH Q33.

A-1. *Amplification is the control of an output signal by an input signal so that the output signal has some (or all) of the characteristics of the input signal. The output signal is generally larger than the input signal in terms of voltage, current, or power.*

A-2. *No, the input signal is unchanged, the output signal is controlled by the input signal but does not effect the actual input signal.*

A-3. *To amplify the input signal to a usable level.*

A-4. *By function and frequency response.*

A-5. *An audio power amplifier.*

A-6. *An rf voltage amplifier.*

A-7. *The amount of time (in relation to the input signal) in which current flows in the output circuit.*

A-8. *A, AB, B, C.*

A-9. *Class B operation.*

A-10. *The amplifier operates (and therefore uses power) for less time in class C than in class A.*

A-11. *Class A operation.*

A-12. *To transfer energy (a signal) from one stage to another.*

A-13. *Direct, RC, impedance, and transformer coupling.*

A-14. *RC coupling.*

A-15. *Transformer coupling.*

A-16. *RC coupling.*

A-17. Impedance coupling.

A-18. Equal impedance.

A-19. Lower than.

A-20. Common emitter-medium input, medium output; common base-low input, high output; common collector-high input, low output.

A-21. Common collector.

A-22. Transformer coupling.

A-23. The process of coupling a portion of the output of a circuit back to the circuit input.

A-24. Positive and negative or regenerative and degenerative.

A-25. Positive (regenerative) feedback.

A-26. Negative (degenerative) feedback.

A-27. Negative (degenerative) feedback.

A-28. Negative (degenerative) feedback.

A-29. A device that provides two output signals that differ in phase from a single input signal.

A-30. A phase splitter is used to provide the input signals to a push-pull amplifier.

A-31. A push-pull amplifier is used when high power output and good fidelity are needed.

A-32. A push-pull amplifier provides more gain than a single transistor amplifier.

A-33. Class A, Class AB or Class B operation.

CHAPTER 2

VIDEO AND RF AMPLIFIERS

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to:

1. Define the term "bandwidth of an amplifier."
2. Determine the upper and lower frequency limits of an amplifier from a frequency-response curve.
3. List the factors that limit frequency response in an amplifier.
4. List two techniques used to increase the high-frequency response for a video amplifier.
5. State one technique used to increase the low-frequency response of a video amplifier.
6. Identify the purpose of various components on a schematic of a complete typical video amplifier circuit.
7. State the purpose of a frequency-determining network in an rf amplifier.
8. State one method by which an rf amplifier can be neutralized.
9. Identify the purpose of various components on a schematic of a complete typical rf amplifier.

INTRODUCTION

In this chapter you will be given information on the frequency response of amplifiers as well as specific information on video and rf amplifiers. For all practical purposes, all the general information you studied in chapter 1 about audio amplifiers will apply to the video and rf amplifiers which you are about to study.

You may be wondering why you need to learn about video and rf amplifiers. You need to understand these circuits because, as a technician, you will probably be involved in working on equipment in which these circuits are used. Many of the circuits shown in this and the next chapter are incomplete and would not be used in actual equipment. For example, the complete biasing network may not be shown. This is done so you can concentrate on the concepts being presented without being overwhelmed by an abundance of circuit elements. With this idea in mind, the information that is presented in this chapter is real, practical information about video and rf amplifiers. It is the sort of information that you will use in working with these circuits. Engineering information (such as design specifications) will not be presented because it is not needed to understand the concepts that a technician needs to perform the job of circuit analysis and repair. Before you are given the specific information on video and rf amplifiers, you may be wondering how these circuits are used.

Video amplifiers are used to amplify signals that represent video information. (That's where the term "video" comes from.) Video is the "picture" portion of a television signal. The "sound" portion is audio.

Although the Navy uses television in many ways, video signals are used for more than television. Radar systems (discussed later in this training series) use video signals and, therefore, video amplifiers. Video amplifiers are also used in video recorders and some communication and control devices. In addition to using video amplifiers, televisions use rf amplifiers. Many other devices also use rf amplifiers, such as radios, navigational devices, and communications systems. Almost any device that uses broadcast, or transmitted, information will use an rf amplifier.

As you should recall, rf amplifiers are used to amplify signals between 10 kilohertz (10 kHz) and 100,000 megahertz (100,000 MHz) (not this entire band of frequencies, but any band of frequencies within these limits). Therefore, any device that uses frequencies between 10 kilohertz and 100,000 megahertz will most likely use an rf amplifier.

Before you study the details of video and rf amplifiers, you need to learn a little more about the frequency response of an amplifier and frequency-response curves.

AMPLIFIER FREQUENCY RESPONSE

In chapter 1 of this module you were shown the frequency-response curve of an audio amplifier. Every amplifier has a frequency-response curve associated with it. Technicians use frequency-response curves because they provide a "picture" of the performance of an amplifier at various frequencies. You will probably never have to draw a frequency-response curve, but, in order to use one, you should know how a frequency-response curve is created. The amplifier for which the frequency-response curve is created is tested at various frequencies. At each frequency, the input signal is set to some predetermined level of voltage (or current). This same voltage (or current) level for all of the input signals is used to provide a standard input and to allow evaluation of the output of the circuit at each of the frequencies tested. For each of these frequencies, the output is measured and marked on a graph. The graph is marked "frequency" along the horizontal axis and "voltage" or "current" along the vertical axis. When points have been plotted for all of the frequencies tested, the points are connected to form the frequency-response curve. The shape of the curve represents the frequency response of the amplifier.

Some amplifiers should be "flat" across a band of frequencies. In other words, for every frequency within the band, the amplifier should have equal gain (equal response). For frequencies outside the band, the amplifier gain will be much lower.

For other amplifiers, the desired frequency response is different. For example, perhaps the amplifier should have high gain at two frequencies and low gain for all other frequencies. The frequency-response curve for this type of amplifier would show two "peaks." In other amplifiers the frequency-response curve will have one peak indicating high gain at one frequency and lower gain at all others.

Note the frequency-response curve shown in figure 2-1. This is the frequency-response curve for an audio amplifier as described in chapter 1. It is "flat" from 15 hertz (15 Hz) to 20 kilohertz (20 kHz).

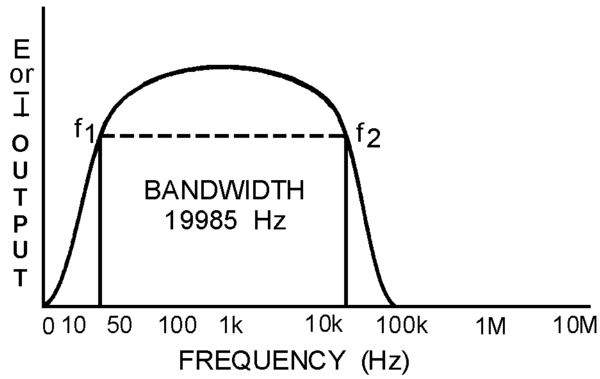


Figure 2-1.—Frequency response curve of audio amplifier.

Notice in the figure that the lower frequency limit is labeled f_1 and the upper frequency limit is labeled f_2. Note also the portion inside the frequency-response curve marked "BANDWIDTH." You may be wondering just what a "bandwidth" is.

BANDWIDTH OF AN AMPLIFIER

The bandwidth represents the amount or "width" of frequencies, or the "band of frequencies," that the amplifier is MOST effective in amplifying. However, the bandwidth is NOT the same as the band of frequencies that is amplified. The bandwidth (BW) of an amplifier is the difference between the frequency limits of the amplifier. For example, the band of frequencies for an amplifier may be from 10 kilohertz (10 kHz) to 30 kilohertz (30 kHz). In this case, the bandwidth would be 20 kilohertz (20 kHz). As another example, if an amplifier is designed to amplify frequencies between 15 hertz (15 Hz) and 20 kilohertz (20 kHz), the bandwidth will be equal to 20 kilohertz minus 15 hertz or 19,985 hertz (19,985 Hz). This is shown in figure 2-1.

Mathematically:

$$BW = f_2 - f_1$$

$$BW = 20\text{ kHz} - 15\text{ Hz}$$

$$BW = 20{,}000\text{ Hz} - 15\text{ Hz}$$

$$BW = 19{,}985\text{ Hz}$$

You should notice on the figure that the frequency-response curve shows output voltage (or current) against frequency. The lower and upper frequency limits (f_1 and f_2) are also known as HALF-POWER POINTS. The half-power points are the points at which the output voltage (or current) is 70.7 percent of the maximum output voltage (or current). Any frequency that produces less than 70.7 percent of the maximum output voltage (or current) is outside the bandwidth and, in most cases, is not considered a useable output of the amplifier.

The reason these points are called "half-power points" is that the true output power will be half (50 percent) of the maximum true output power when the output voltage (or current) is 70.7 percent of the maximum output voltage (or current), as shown below. (All calculations are rounded off to two decimal places.)

As you learned in NEETS, Module 2, in an a.c. circuit true power is calculated using the resistance (R) of the circuit, NOT the impedance (Z). If the circuit produces a maximum output voltage of 10 volts across a 50-ohm load, then:

$$\text{True Power} = \frac{E^2}{R}$$

$$\text{True Power} = \frac{(10V)^2}{50\Omega}$$

$$\text{True Power} = \frac{100}{50}\text{ watts}$$

$$\text{True Power} = 2\text{ watts}$$

When the output voltage drops to 70.7 percent of the maximum voltage of 10 volts, then:

$$\text{True Power} = \frac{E^2}{R}$$

$$\text{True Power} = \frac{(7.07V)^2}{50\Omega}$$

$$\text{True Power} = \frac{50}{50}\text{ watts}$$

$$\text{True Power} = 1\text{ watts}$$

As you can see, the true power is 50 percent (half) of the maximum true power when the output voltage is 70.7 percent of the maximum output voltage. If, instead, you are using the output current of the above circuit, the maximum current is

$$.2\text{amp}\left(\frac{10V}{50\Omega} = .2A\right).$$

The calculations are:

$$\text{True Power} = I^2R$$

$$\text{True Power} = (.2A)^2\,(50\Omega)$$

$$\text{True Power} = (.04)\ \ (50)\text{ watts}$$

$$\text{True Power} = 2\text{ watts}$$

At 70.7 percent of the output current (.14 A):

$$\text{True Power} = I^2 R$$

$$\text{True Power} = (.14A)^2 (50\Omega)$$

$$\text{True Power} = (.02 \times 50) \text{ watts}$$

$$\text{True Power} = 1 \text{ watts}$$

On figure 2-1, the two points marked f_1 and f_2 will enable you to determine the frequency-response limits of the amplifier. In this case, the limits are 15 hertz (15 Hz) and 20 kilohertz (20 kHz). You should now see how a frequency-response curve can enable you to determine the frequency limits and the bandwidth of an amplifier.

READING AMPLIFIER FREQUENCY-RESPONSE CURVES

Figure 2-2 shows the frequency-response curves for four different amplifiers. View (A) is the same frequency-response curve as shown on page 5-6, figure 2-1. View (B) is the frequency-response curve of an amplifier that would also be classified as an audio amplifier, even though the curve is not "flat" from 15 hertz to 20 kilohertz and does not drop off sharply at the frequency limits. From the curve, you can see that the lower frequency limit of this amplifier (f_1) is 100 hertz. The upper frequency limit (f_2) is 10 kilohertz. Therefore, the bandwidth of this amplifier must be 10 kilohertz minus 100 hertz or 9900 hertz. Most amplifiers will have a frequency-response curve shaped like view (B) if nothing is done to modify the frequency-response characteristics of the circuit. (The factors that affect frequency response and the methods to modify the frequency response of an amplifier are covered a little later in this chapter.)

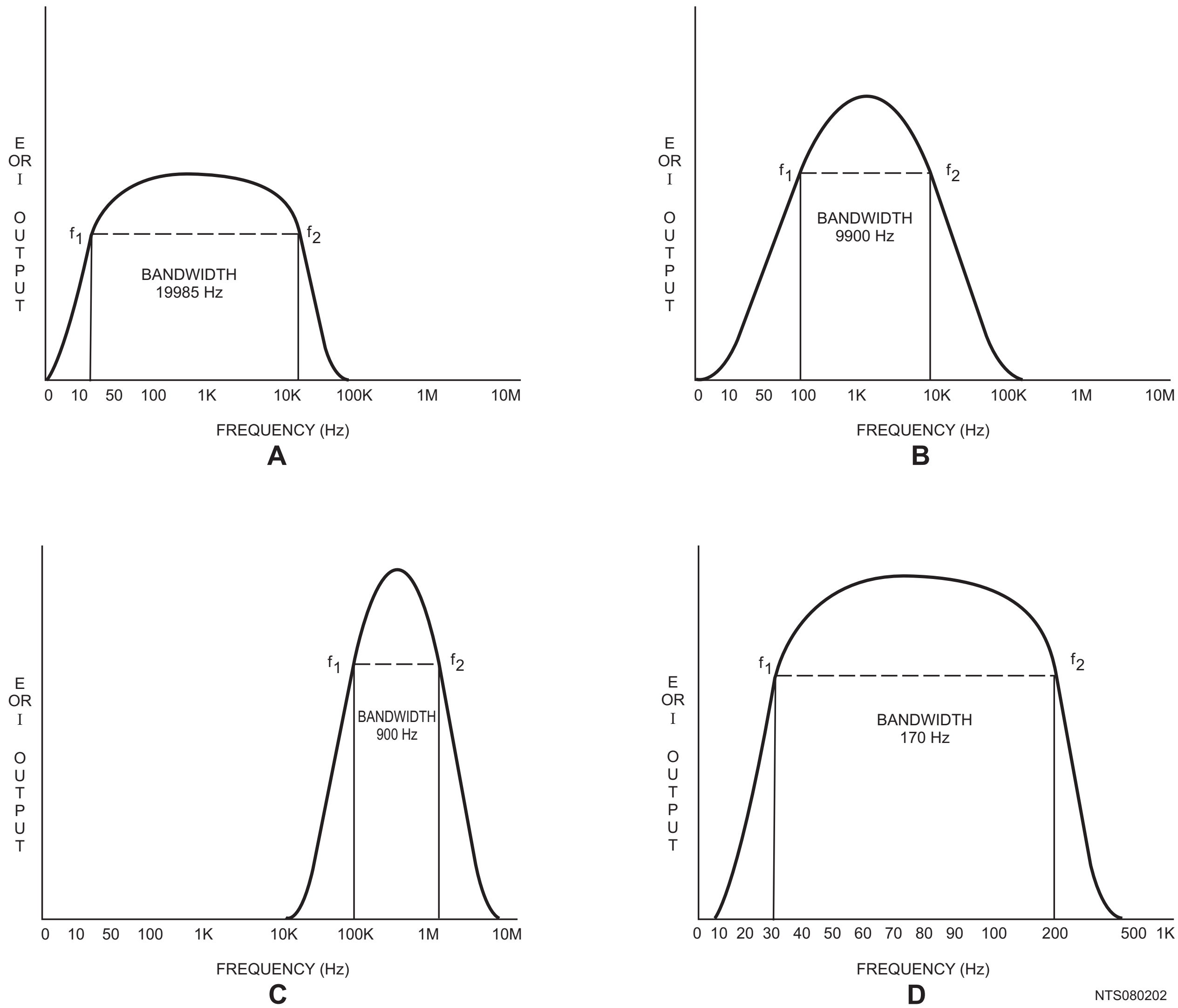


Figure 2-2.—Frequency response curves, A THROUGH D.

Now look at view (C). This frequency-response curve is for an rf amplifier. The frequency limits of this amplifier are 100 kilohertz (f_1) and 1 megahertz (f_2); therefore, the bandwidth of this amplifier is 900 kilohertz (kHz).

View (D) shows another audio amplifier. This time the frequency limits are 30 hertz (f_1) and 200 hertz (f_2). The bandwidth of this amplifier is only 170 hertz. The important thing to notice in view (D) is that the frequency scale is different from those used in other views. Any frequency scale can be used for a frequency-response curve. The scale used would be determined by what frequencies are most useful in presenting the frequency-response curve for a particular amplifier.

Q-1. What is the bandwidth of an amplifier?

Q-2. What are the upper and lower frequency limits of an amplifier?

Q-3. What are the upper and lower frequency limits and the bandwidth for the amplifiers that have frequency-response curves as shown in figure 2-3?

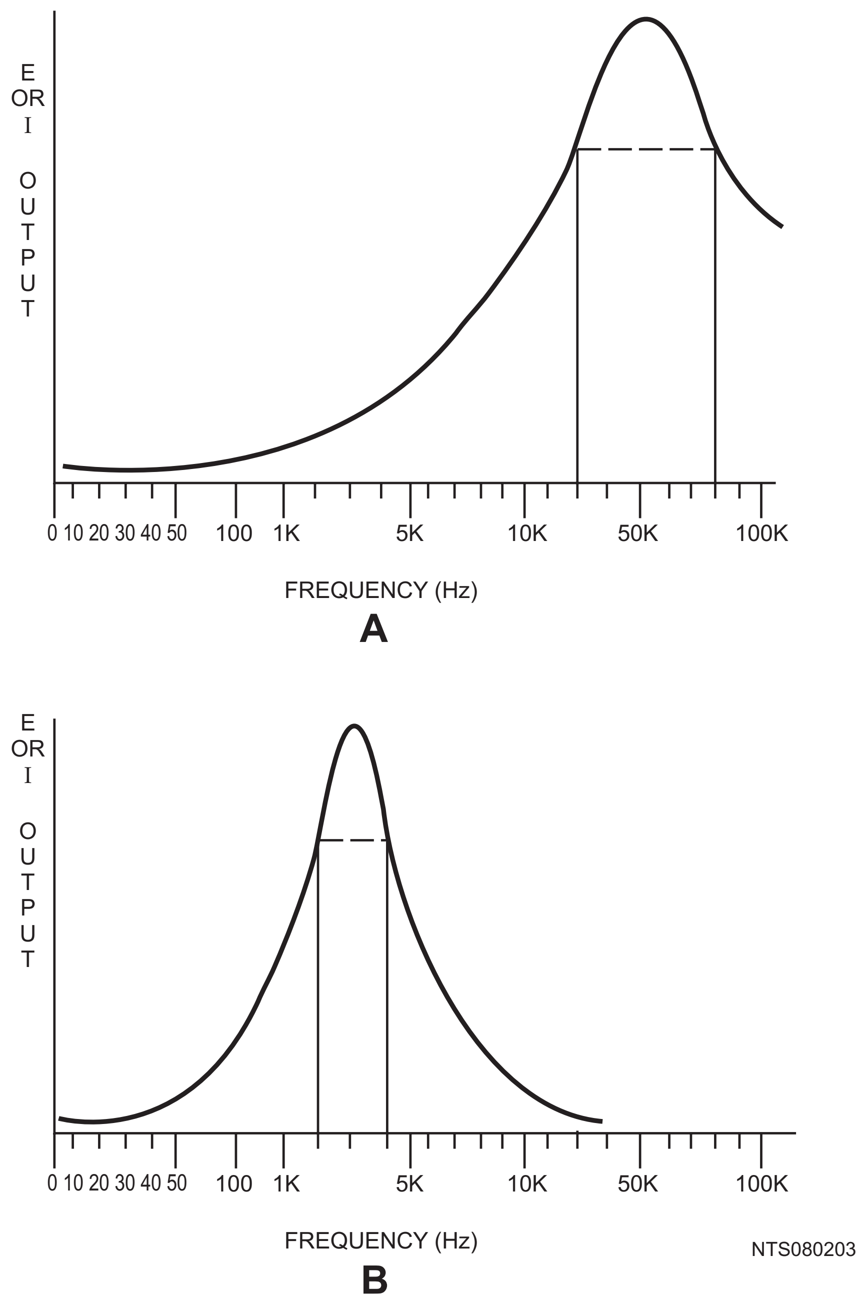


Figure 2-3B.—Frequency-response curves for Q3.

FACTORS AFFECTING FREQUENCY RESPONSE OF AN AMPLIFIER

In chapter 1 of this module, the fact was mentioned that an audio amplifier is limited in its frequency response. Now you will see why this is true.

You should recall that the frequency response of an a.c. circuit is limited by the reactive elements (capacitance and inductance) in the circuit. As you know, this is caused by the fact that the capacitive and inductive reactances vary with the frequency. In other words, the value of the reactance is determined, in part, by frequency. Remember the formulas:

$$X_C = \frac{1}{2\,\pi f C}$$

$$X_L = 2\,\pi f L$$

If you ignore the amplifying device (transistor, electron tube, etc.), and if the amplifier circuit is made up of resistors only, there should be no limits to the frequency response. In other words, a totally resistive circuit would have no frequency limits. However, there is no such thing as a totally resistive circuit because circuit components almost always have some reactance. In addition to the reactance of other components in the circuit, most amplifiers use RC coupling. This means that a capacitor is used to couple the signal in to and out of the circuit. There is also a certain amount of capacitance and inductance in the wiring of the circuit. The end result is that all circuits are reactive. To illustrate this point, figure 2-4 shows amplifier circuits with the capacitance and inductance of the wiring represented as "phantom" capacitors and inductors. The reactances of the capacitors (X_C) and the inductors (X_L) are shown as "phantom" variable resistors. View (A) shows the circuit with a low-frequency input signal, and view (B) shows the circuit with a high-frequency input signal.

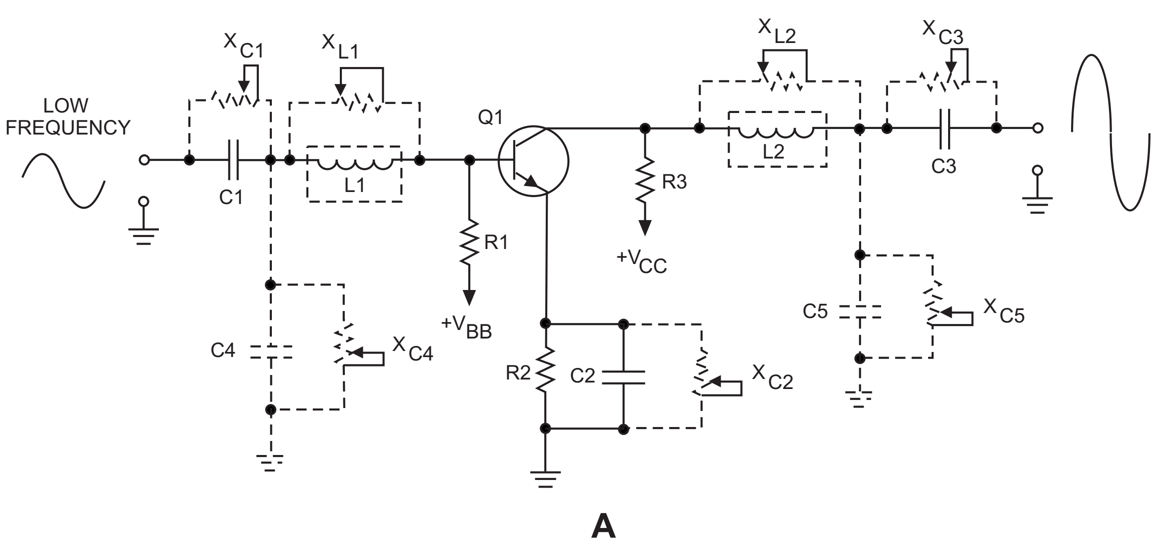


Figure 2-4A.—Amplifiers showing reactive elements and reactance.

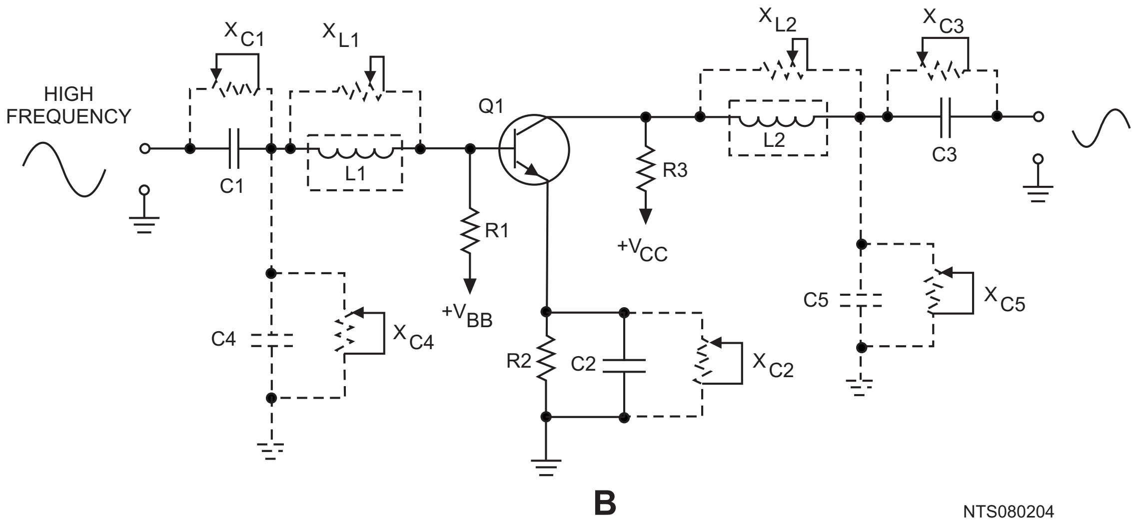


Figure 2-4B.—Amplifiers showing reactive elements and reactance.

The actual circuit components are: C1, C2, C3, R1, R2, R3, and Q1. C1 is used to couple the input signal. R1 develops the input signal. R2, the emitter resistor, is used for proper biasing and temperature stability. C2 is a decoupling capacitor for R2. R3 develops the output signal. C3 couples the output signal to the next stage. Q1 is the amplifying device.

The phantom circuit elements representing the capacitance and inductance of the wiring are: L1, L2, C4, and C5. L1 represents the inductance of the input wiring. L2 represents the inductance of the output wiring. C4 represents the capacitance of the input wiring. C5 represents the capacitance of the output wiring.

In view (A) the circuit is shown with a low-frequency input signal. Since the formulas for capacitive reactance and inductive reactance are:

$$X_C = \frac{1}{2\,\pi f C}$$

$$X_L = 2\,\pi f L$$

You should remember that if frequency is low, capacitive reactance will be high and inductive reactance will be low. This is shown by the position of the variable resistors that represent the reactances. Notice that X_{L1} and X_{L2} are low; therefore, they do not "drop" very much of the input and output signals. X_{C4} and X_{C5} are high; these reactances tend to "block" the input and output signals and keep them from going to the power supplies (V_{BB} and V_{CC}). Notice that the output signal is larger in amplitude than the input signal.

Now look at view (B). The input signal is a high-frequency signal. Now X_C is low and X_L is high. X_{L1} and X_{L2} now drop part of the input and output signals. At the same time X_{C4} and X_{C5} tend to "short" or "pass" the input and output signals to signal ground. The net effect is that both the input and output signals are reduced. Notice that the output signal is smaller in amplitude than the input signal.

Now you can see how the capacitance and inductance of the wiring affect an amplifier, causing the output of an amplifier to be less for high-frequency signals than for low-frequency signals.

In addition to the other circuit components, an amplifying device (transistor or electronic tube), itself, reacts differently to high frequencies than it does to low frequencies. In earlier *NEETS* modules you were told that transistors and electronic tubes have interelectrode capacitance. Figure 2-5 shows a portion of the interelectrode capacitance of a transistor and the way in which this affects high- and low-frequency signals.

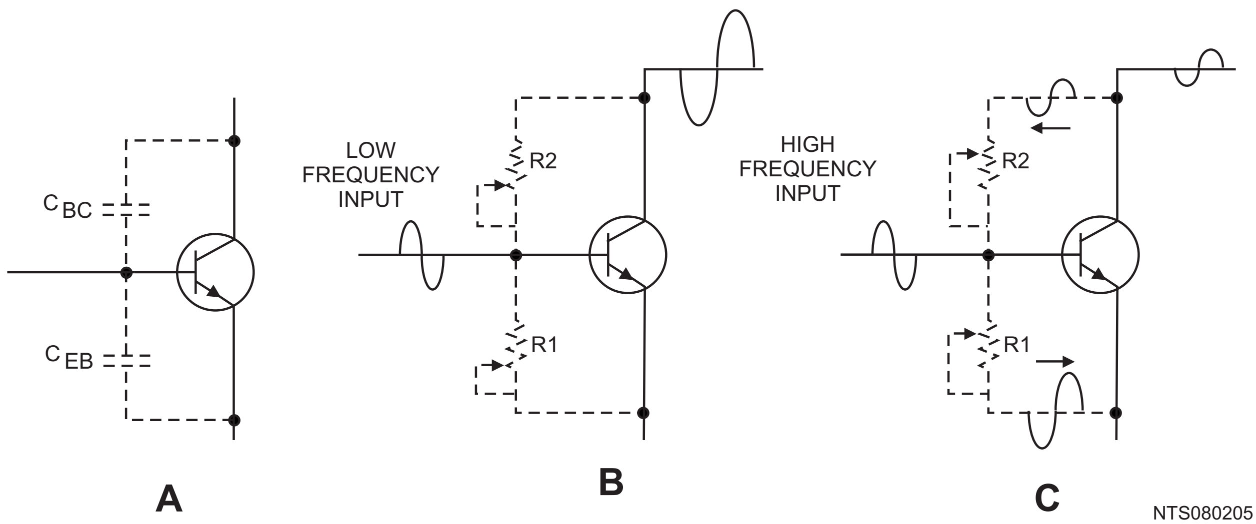


Figure 2-5.—Interelectrode capacitance of a transistor.

In view (A) a transistor is shown with phantom capacitors connected to represent the interelectrode capacitance. C_{EB} represents the emitter-to-base capacitance. C_{BC} represents the base-to-collector capacitance.

For simplicity, in views (B) and (C) the capacitive reactance of these capacitors is shown by variable resistors R1 (for C_{EB}) and R2 (for C_{BC}). View (B) shows the reactance as high when there is a low-frequency input signal. In this case there is very little effect from the reactance on the transistor. The transistor amplifies the input signal as shown in view (B). However, when a high-frequency input signal is applied to the transistor, as in view (C), things are somewhat different. Now the capacitive reactance is low (as shown by the settings of the variable resistors). In this case, as the base of the transistor attempts to go positive during the first half of the input signal, a great deal of this positive signal is felt on the emitter (through R1). If both the base and the emitter go positive at the same time, there is no change in emitter-base bias and the conduction of the transistor will not change. Of course, a small amount of change does occur in the emitter-base bias, but not as much as when the capacitive reactance is higher (at low frequencies). As an output signal is developed in the collector circuit, part of this signal is fed back to the base through R2. Since the signal on the collector is 180 degrees out of phase with the base signal, this tends to drive the base negative. The effect of this is to further reduce the emitter-base bias and the conduction of the transistor.During the second half of the input signal, the same effect occurs although the polarity is reversed. The net effect is a reduction in the gain of the transistor as indicated by the small output signal. This decrease in the amplifier output at higher frequencies is caused by the interelectrode capacitance. (There are certain special cases in which the feedback signal caused by the interelectrode capacitance is in phase with the base signal. However, in most cases, the feedback caused by interelectrode capacitance is degenerative and is 180 degrees out of phase with the base signal as explained above.)

Q-4. What are the factors that limit the frequency response of a transistor amplifier?

Q-5. What type of feedback is usually caused by interelectrode capacitance?

Q-6. What happens to capacitive reactance as frequency increases?

Q-7. What happens to inductive reactance as frequency increases?

VIDEO AMPLIFIERS

As you have seen, a transistor amplifier is limited in its frequency response. You should also remember from chapter 1 that a VIDEO AMPLIFIER should have a frequency response of 10 hertz (10 Hz) to 6 megahertz (6 MHz). The question has probably occurred to you: How is it possible to "extend" the range of frequency response of an amplifier?

HIGH-FREQUENCY COMPENSATION FOR VIDEO AMPLIFIERS

If the frequency-response range of an audio amplifier must be extended to 6 megahertz (6 MHz) for use as a video amplifier, some means must be found to overcome the limitations of the audio amplifier. As you have seen, the capacitance of an amplifier circuit and the interelectrode capacitance of the transistor (or electronic tube) cause the higher frequency response to be limited.

In some ways capacitance and inductance can be thought of as opposites. As stated before, as frequency increases, capacitive reactance decreases, and inductive reactance increases. Capacitance opposes changes in voltage, and inductance opposes changes in current. Capacitance causes current to lead voltage, and inductance causes voltage to lead current.

Since frequency affects capacitive reactance and inductive reactance in opposite ways, and since it is the capacitive reactance that causes the problem with high-frequency response, inductors are added to an amplifier circuit to improve the high-frequency response. This is called HIGH-FREQUENCY COMPENSATION. Inductors (coils), when used for high-frequency compensation, are called PEAKING COILS. Peaking coils can be added to a circuit so they are in series with the output signal path or in parallel to the output signal path. Instead of only in series or parallel, a combination of peaking coils in series and parallel with the output signal path can also be used for high-frequency compensation.

As in all electronic circuits, nothing comes free. The use of peaking coils WILL increase the frequency response of an amplifier circuit, but it will ALSO lower the gain of the amplifier.

Series Peaking

The use of a peaking coil in series with the output signal path is known as SERIES PEAKING. Figure 2-6 shows a transistor amplifier circuit with a series peaking coil. In this figure, R1 is the input-signal-developing resistor. R2 is used for bias and temperature stability of Q1. C1 is the bypass capacitor for R2. R3 is the load resistor for Q1 and develops the output signal. C2 is the coupling capacitor which couples the output signal to the next stage. "Phantom" capacitor C_{OUT} represents the output capacitance of the circuit, and "phantom" capacitor C_{IN} represents the input capacitance of the next stage.

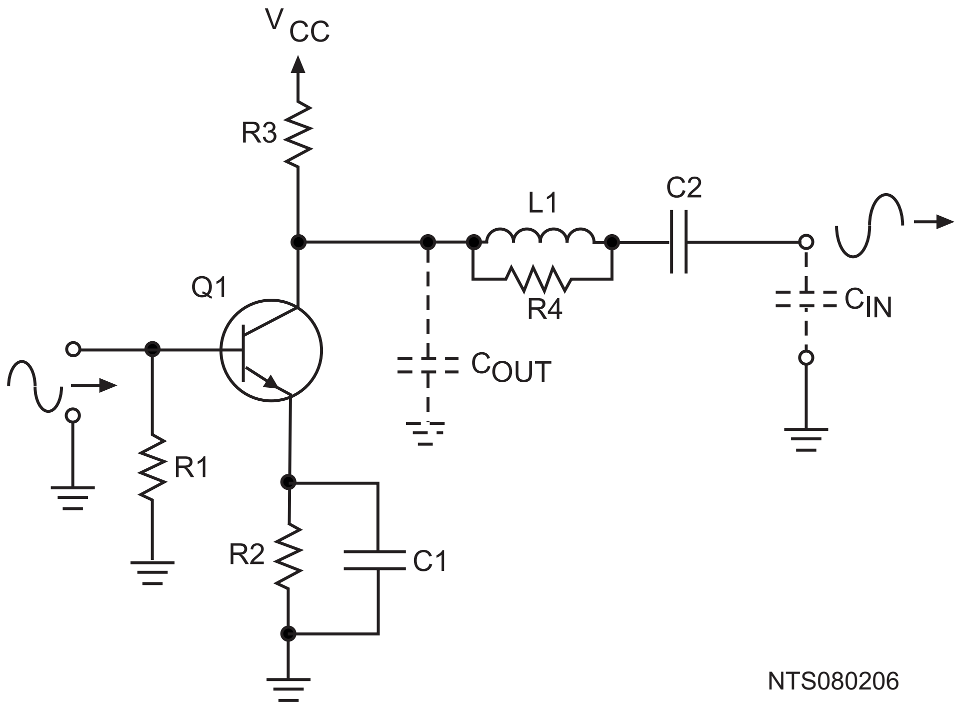


Figure 2-6.—Series peaking coil.

You know that the capacitive reactance of C_{OUT} and C_{IN} will limit the high-frequency response of the circuit. L1 is the series peaking coil. It is in series with the output-signal path and isolates C_{OUT} from C_{IN}. R4 is called a "swamping" resistor and is used to keep L1 from overcompensating at a narrow range of frequencies. In other words, R4 is used to keep the frequency-response curve flat. If R4 were not used with L1, there could be a "peak" in the frequency-response curve. (Remember, L1 is called a peaking coil.)

Shunt Peaking

If a coil is placed in parallel (shunt) with the output signal path, the technique is called SHUNT PEAKING. Figure 2-7 shows a circuit with a shunt peaking coil. With the exceptions of the "phantom" capacitor and the inductor, the components in this circuit are the same as those in figure 2-6. R1 is the input-signal-developing resistor. R2 is used for bias and temperature stability. C1 is the bypass capacitor for R2. R3 is the load resistor for Q1 and develops the output signal. C2 is the coupling capacitor which couples the output signal to the next stage.

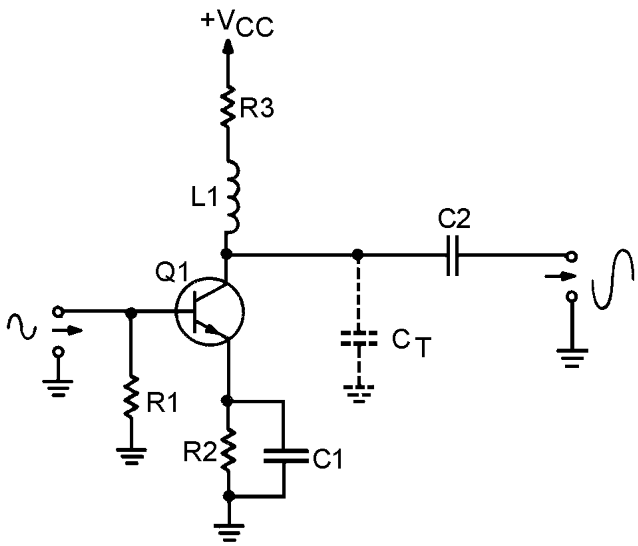


Figure 2-7.—Shunt peaking coil.

The "phantom" capacitor, C_T, represents the total capacitance of the circuit. Notice that it tends to couple the output signal to ground.

L1 is the shunt peaking coil. While it is in series with the load resistor (R3), it is in parallel (shunt) with the output-signal path.

Since inductive reactance increases as frequency increases, the reactance of L1 develops more output signal as the frequency increases. At the same time, the capacitive reactance of C_T is decreasing as frequency increases. This tends to couple more of the output signal to ground. The increased inductive reactance counters the effect of the decreased capacitive reactance and this increases the high-frequency response of the amplifier.

Combination Peaking

You have seen how a series peaking coil isolates the output capacitance of an amplifier from the input capacitance of the next stage. You have also seen how a shunt peaking coil will counteract the effects of the total capacitance of an amplifier. If these two techniques are used together, the combination is more effective than the use of either one alone. The use of both series and shunt peaking coils is known as COMBINATION PEAKING. An amplifier circuit with combination peaking is shown in figure 2-8. In figure 2-8 the peaking coils are L1 and L2. L1 is a shunt peaking coil, and L2 is a series peaking coil.

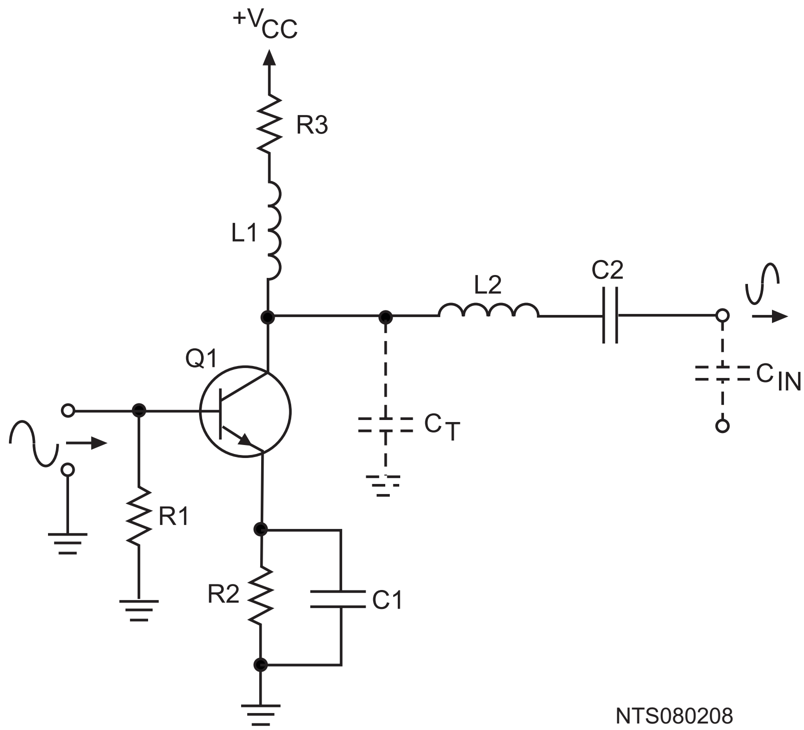


Figure 2-8.—Combination peaking.

The "phantom" capacitor C_T represents the total capacitance of the amplifier circuit. "Phantom" capacitor C_{IN} represents the input capacitance of the next stage. Combination peaking will easily allow an amplifier to have a high-frequency response of 6 megahertz (6 MHz).

Q-8. What is the major factor that limits the high-frequency response of an amplifier circuits?

Q-9. What components can be used to increase the high-frequency response of an amplifier?

Q-10. What determines whether these components are considered series or shunt?

Q-11. What is the arrangement of both series and shunt components called?

LOW-FREQUENCY COMPENSATION FOR VIDEO AMPLIFIERS

Now that you have seen how the high-frequency response of an amplifier can be extended to 6 megahertz (6 MHz), you should realize that it is only necessary to extend the low-frequency response to 10 hertz (10 Hz) in order to have a video amplifier.

Once again, the culprit in low-frequency response is capacitance (or capacitive reactance). But this time the problem is the coupling capacitor between the stages.

At low frequencies the capacitive reactance of the coupling capacitor (C2 in figure 2-8) is high. This high reactance limits the amount of output signal that is coupled to the next stage. In addition, the RC network of the coupling capacitor and the signal-developing resistor of the next stage cause a phase shift in the output signal. (*Refer to NEETS, Module 2, for a discussion of phase shifts in RC networks.)* Both of these problems (poor low-frequency response and phase shift) can be solved by adding a parallel RC network in series with the load resistor. This is shown in figure 2-9.

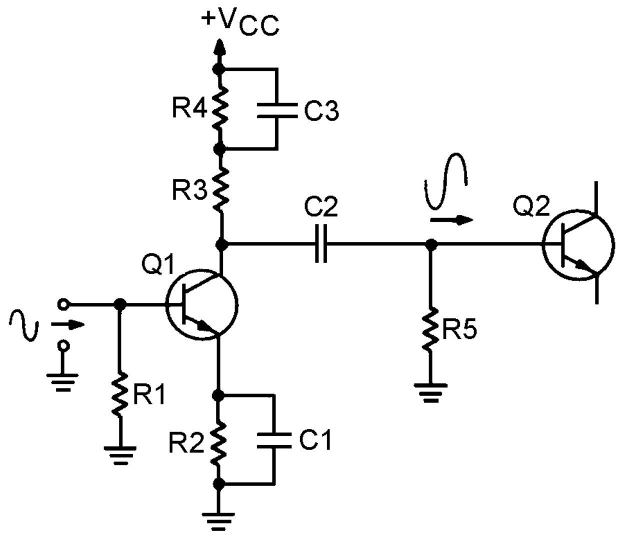


Figure 2-9.—Low frequency compensation network.

The complete circuitry for Q2 is not shown in this figure, as the main concern is the signal-developing resistor (R5) for Q2. The coupling capacitor (C2) and the resistor (R5) limit the low-frequency response of the amplifier and cause a phase shift. The amount of the phase shift will depend upon the amount of resistance and capacitance. The RC network of R4 and C3 compensates for the effects of C2 and R5 and extends the low-frequency response of the amplifier.

At low frequencies, R4 adds to the load resistance (R3) and increases the gain of the amplifier. As frequency increases, the reactance of C3 decreases. C3 then provides a path around R4 and the gain of the transistor decreases. At the same time, the reactance of the coupling capacitor (C2) decreases and more signal is coupled to Q2.

Because the circuit shown in figure 2-9 has no high-frequency compensation, it would not be a very practical video amplifier.

TYPICAL VIDEO-AMPLIFIER CIRCUIT

There are many different ways in which video amplifiers can be built. The particular configuration of a video amplifier depends upon the equipment in which the video amplifier is used. The circuit shown in figure 2-10 is only one of many possible video-amplifier circuits. Rather than reading about what each component does in this circuit, you can see how well you have learned about video amplifiers by answering the following questions. You should have no problem identifying the purpose of the components because similar circuits have been explained to you earlier in the text.

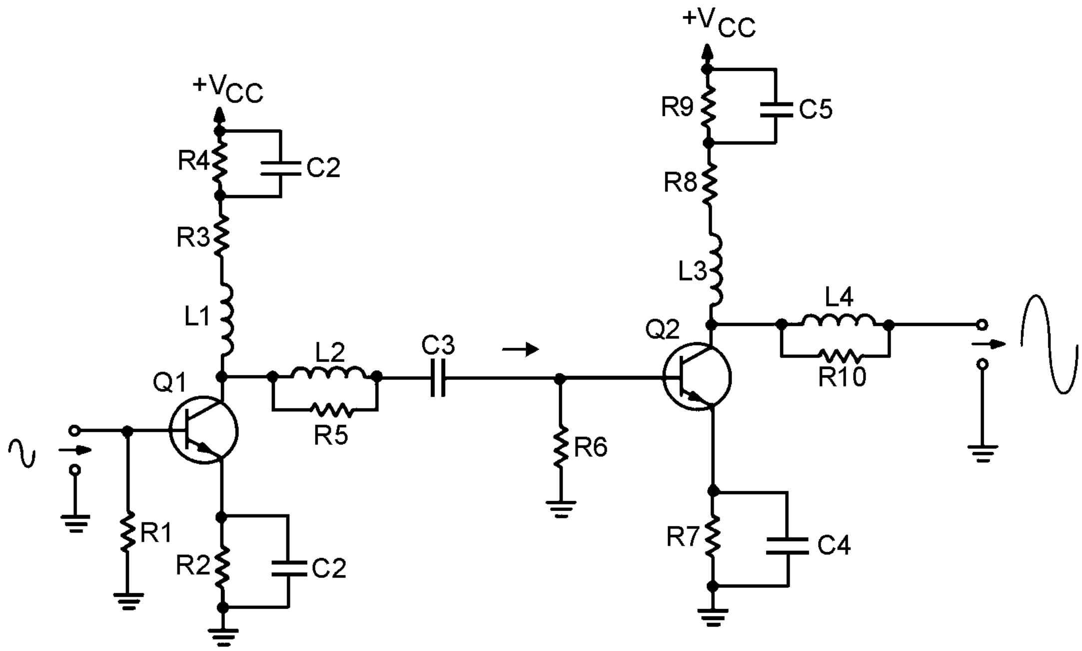


Figure 2-10.—Video amplifier circuit.

The following questions refer to figure 2-10.

Q-12. What component in an amplifier circuit tends to limit the low-frequency response of the amplifier?

Q-13. What is the purpose of L3?

Q-14. What is the purpose of C1?

Q-15. What is the purpose of R4?

Q-16. What is the purpose of L2?

Q-17. What is the purpose of R5?

Q-18. What component(s) is/are used for high-frequency compensation for Q1?

Q-19. What component(s) is/are used for low-frequency compensation for Q2?

RADIO-FREQUENCY AMPLIFIERS

Now that you have seen the way in which a broadband, or video, amplifier can be constructed, you may be wondering about radio-frequency (rf) amplifiers. Do they use the same techniques? Are they just another type of broadband amplifier?

The answer to both questions is "no." Radio-frequency amplifiers use different techniques than video amplifiers and are very different from them.

Before you study the specific techniques used in rf amplifiers, you should review some information on the relationship between the input and output impedance of an amplifier and the gain of the amplifier stage.

AMPLIFIER INPUT/OUTPUT IMPEDANCE AND GAIN

You should remember that the gain of a stage is calculated by using the input and output signals. The formula used to calculate the gain of a stage is:

$$\text{gain} = \frac{\text{output signal}}{\text{input signal}}$$

Voltage gain is calculated using input and output voltage; current gain uses input and output current; and power gain uses input and output power. For the purposes of our discussion, we will only be concerned with voltage gain.

Figure 2-11 shows a simple amplifier circuit with the input- and output-signal-developing impedances represented by variable resistors. In this circuit, C1 and C2 are the input and output coupling capacitors. R1 represents the impedance of the input circuit. R2 represents the input-signal-developing impedance, and R3 represents the output impedance.

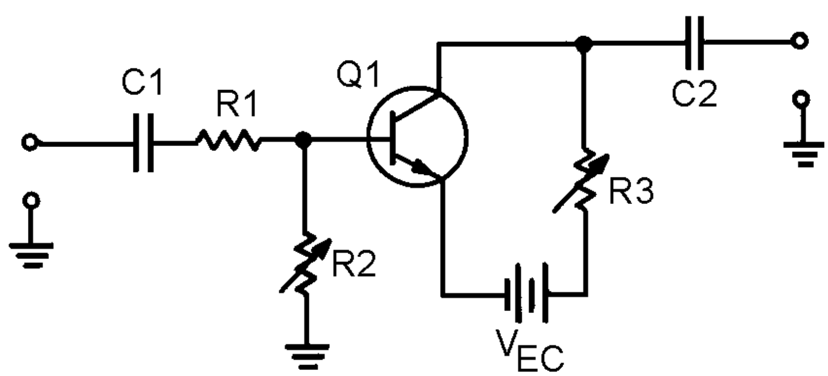


Figure 2-11.—Variable input and output impedances.

R1 and R2 form a voltage-divider network for the input signal. When R2 is increased in value, the input signal to the transistor (Q1) increases. This causes a larger output signal, and the gain of the stage increases.

Now look at the output resistor, R3. As R3 is increased in value, the output signal increases. This also increases the gain of the stage.

As you can see, increasing the input-signal-developing impedance, the output impedance, or both will increase the gain of the stage. Of course there are limits to this process. The transistor must not be overdriven with too high an input signal or distortion will result.

With this principle in mind, if you could design a circuit that had maximum impedance at a specific frequency (or band of frequencies), that circuit could be used in an rf amplifier. This FREQUENCY-DETERMINING NETWORK could be used as the input-signal-developing impedance, the output impedance, or both. The rf amplifier circuit would then be as shown in figure 2-12.

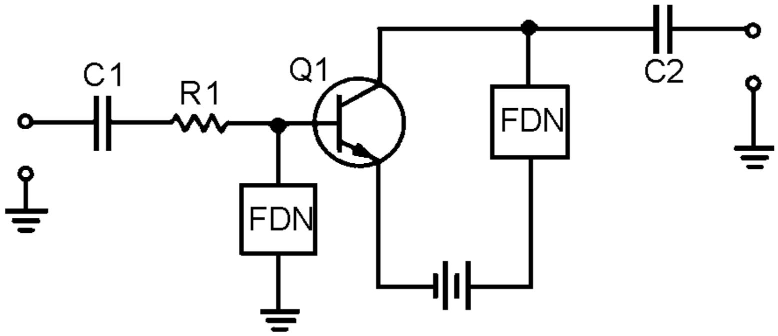


Figure 2-12.—Semiblock diagram of rf amplifier.

In this "semi-block" diagram, C1 and C2 are the input and output coupling capacitors. R1 represents the impedance of the input circuit. The blocks marked FDN represent the frequency-determining networks. They are used as input-signal-developing and output impedances for Q1.

FREQUENCY-DETERMINING NETWORK FOR AN RF AMPLIFIER

What kind of circuit would act as a frequency-determining network? In general, a frequency-determining network is a circuit that provides the desired response at a particular frequency. This response could be maximum impedance or minimum impedance; it all depends on how the frequency-determining network is used. You will see more about frequency-determining networks in *NEETS, Module 9—Introduction to Wave-Generation and Shaping Circuits*. As you have seen, the frequency-determining network needed for an rf amplifier should have maximum impedance at the desired frequency.

Before you are shown the actual components that make up the frequency-determining network for an rf amplifier, look at figure 2-13, which is a simple parallel circuit. The resistors in this circuit are variable and are connected together (ganged) in such a way that as the resistance of R1 increases, the resistance of R2 decreases, and vice versa.

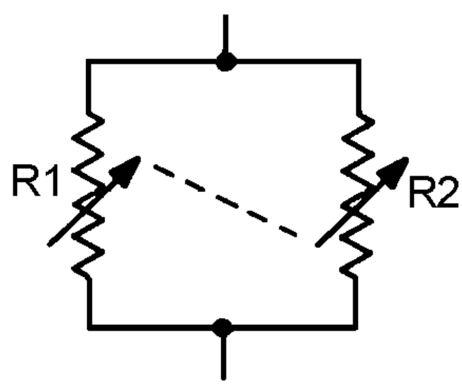


Figure 2-13.—Parallel variable resistors (ganged).

If each resistor has a range from 0 to 200 ohms, the following relationship will exist between the individual resistances and the resistance of the network (R_T). (All values are in ohms, R_T rounded off to two decimal places. These are selected values; there are an infinite number of possible combinations.)

R1	R2	RT
0	200	0.00
10	190	9.50
25	175	21.88
50	150	37.50
75	125	46.88
100	100	50.00
125	75	46.88
150	50	37.50
175	25	21.88
190	10	9.50
200	0	0.00

As you can see, this circuit has maximum resistance (R_T) when the individual resistors are of equal value. If the variable resistors represented impedances and if components could be found that varied their impedance in the same way as the ganged resistors in figure 2-13, you would have the frequency-determining network needed for an rf amplifier.

There are components that will vary their impedance (reactance) like the ganged resistors. As you know, the reactance of an inductor and a capacitor vary as frequency changes. As frequency increases, inductive reactance increases, and capacitive reactance decreases.

At some frequency, inductive and capacitive reactance will be equal. That frequency will depend upon the value of the inductor and capacitor. If the inductor and capacitor are connected as a parallel LC circuit, you will have the ideal frequency-determining network for an rf amplifier.

The parallel LC circuit used as a frequency-determining network is called a TUNED CIRCUIT. This circuit is "tuned" to give the proper response at the desired frequency by selecting the proper values of inductance and capacitance. A circuit using this principle is shown in figure 2-14 which shows an rf amplifier with parallel LC circuits used as frequency-determining networks. This rf amplifier will only be effective in amplifying the frequency determined by the parallel LC circuits.

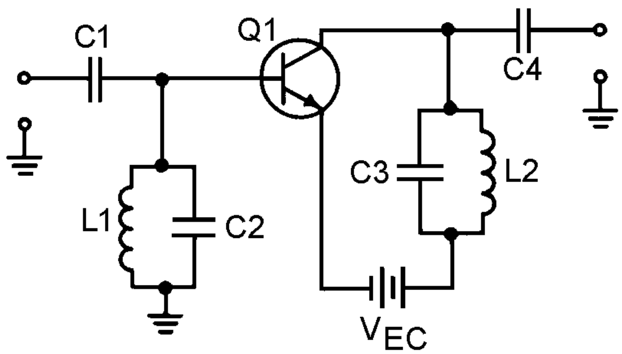


Figure 2-14.—Simple rf amplifier.

In many electronic devices, such as radio or television receivers or radar systems, a particular frequency must be selected from a band of frequencies. This could be done by using a separate rf amplifier for each frequency and then turning on the appropriate rf amplifier. It would be more efficient if a single rf amplifier could be "tuned" to the particular frequency as that frequency is needed. This is what happens when you select a channel on your television set or tune to a station on your radio. To accomplish this "tuning," you need only change the value of inductance or capacitance in the parallel LC circuits (tuned circuits).

In most cases, the capacitance is changed by the use of variable capacitors. The capacitors in the input and output portions of all the rf amplifier stages are ganged together in order that they can all be changed at one time with a single device, such as the tuning dial on a radio. (This technique will be shown on a schematic a little later in this chapter.)

Q-20. If the input-signal-developing impedance of an amplifier is increased, what is the effect on the gain?

Q-21. If the output impedance of an amplifier circuit is decreased, what is the effect on the gain?

Q-22. What is the purpose of a frequency-determining network in an rf amplifier?

Q-23. Can a parallel LC circuit be used as the frequency-determining network for an rf amplifier?

Q-24. How can the frequency be changed in the frequency-determining network?

RF AMPLIFIER COUPLING

Figure 2-14 and the other circuits you have been shown use capacitors to couple the signal in to and out of the circuit (C1 and C4 in figure 2-14). As you remember from chapter 1, there are also other methods of coupling signals from one stage to another. Transformer coupling is the most common method used to couple rf amplifiers. Transformer coupling has many advantages over RC coupling for rf amplifiers; for example, transformer coupling uses fewer components than capacitive coupling. It can also provide a means of increasing the gain of the stage by using a step-up transformer for voltage gain. If a current gain is required, a step-down transformer can be used.

You should also remember that the primary and secondary windings of a transformer are inductors. With these factors in mind, an rf amplifier could be constructed like the one shown in figure 2-15.

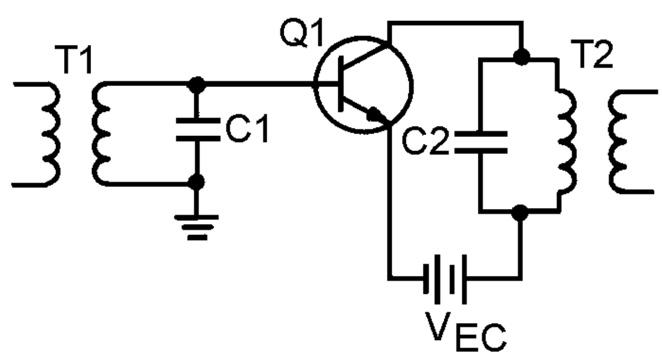


Figure 2-15.—Transformer-coupled rf amplifier.

In this circuit, the secondary of T1 and capacitor C1 form a tuned circuit which is the input-signal-developing impedance. The primary of T2 and capacitor C2 are a tuned circuit which acts as the output impedance of Q1. (Both T1 and T2 must be rf transformers in order to operate at rf frequencies.)

The input signal applied to the primary of T1 could come from the previous stage or from some input device, such as a receiving antenna. In either case, the input device would have a capacitor connected across a coil to form a tuned circuit. In the same way, the secondary of T2 represents the output of this circuit. A capacitor connected across the secondary of T2 would form a parallel LC network. This network could act as the input-signal-developing impedance for the next stage, or the network could represent some type of output device, such as a transmitting antenna.

The tuned circuits formed by the transformer and capacitors may not have the bandwidth required for the amplifier. In other words, the bandwidth of the tuned circuit may be too "narrow" for the requirements of the amplifier. (For example, the rf amplifiers used in television receivers usually require a bandwidth of 6 MHz.)

One way of "broadening" the bandpass of a tuned circuit is to use a swamping resistor. This is similar to the use of the swamping resistor that was shown with the series peaking coil in a video amplifier. A swamping resistor connected in parallel with the tuned circuit will cause a much broader bandpass. (This technique and the theory behind it are discussed in more detail in *NEETS, Module 9.)*

Another technique used to broaden the bandpass involves the amount of coupling in the transformers. For transformers, the term "coupling" refers to the amount of energy transferred from the primary to the secondary of the transformer. This depends upon the number of flux lines from the primary that intersect, or cut, the secondary. When more flux lines cut the secondary, more energy is transferred.

Coupling is mainly a function of the space between the primary and secondary windings. A transformer can be loosely coupled (having little transfer of energy), optimumly coupled (just the right amount of energy transferred), or overcoupled (to the point that the flux lines of primary and secondary windings interfere with each other).

Figure 2-16, (view A) (view B) (view C), shows the effect of coupling on frequency response when parallel LC circuits are made from the primary and secondary windings of transformers.

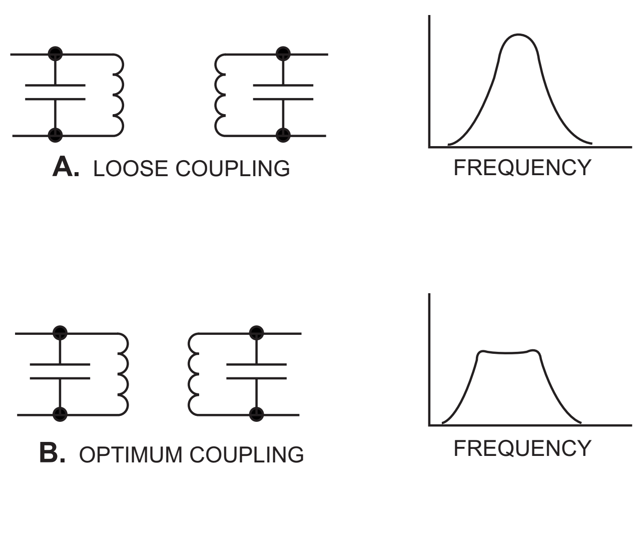


Figure 2-16 .—Effect of coupling on frequency response: A. LOOSE COUPLING B. OPTIMUM COUPLING

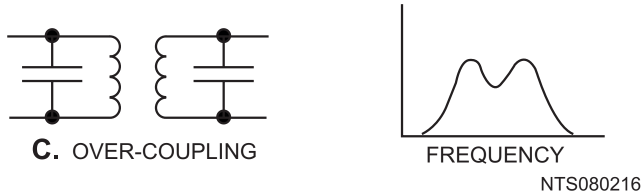


Figure 2-16C.—Effect of coupling on frequency response. OVER-COUPLING

In view (A) the transformer is loosely coupled; the frequency response curve shows a narrow bandwidth. In view (B) the transformer has optimum coupling; the bandwidth is wider and the curve is relatively flat. In view (C) the transformer is overcoupled; the frequency response curve shows a broad bandpass, but the curve "dips" in the middle showing that these frequencies are not developed as well as others in the bandwidth.

Optimum coupling will usually provide the necessary bandpass for the frequency-determining network (and therefore the rf amplifier). For some uses, such as rf amplifiers in a television receiver, the bandpass available from optimum coupling is not wide enough. In these cases, a swamping resistor (as mentioned earlier) will be used with the optimum coupling to broaden the bandpass.

COMPENSATION OF RF AMPLIFIERS

Now you have been shown the way in which an rf amplifier is configured to amplify a band of frequencies and the way in which an rf amplifier can be "tuned" for a particular band of frequencies. You have also seen some ways in which the bandpass of an rf amplifier can be adjusted. However, the frequencies at which rf amplifiers operate are so high that certain problems exist.

One of these problems is the losses that can occur in a transformer at these high frequencies. Another problem is with interelectrode capacitance in the transistor. The process of overcoming these problems is known as COMPENSATION.

Transformers in RF Amplifiers

As you recall from *NEETS, Module 1, t*he losses in a transformer are classified as copper loss, eddy-current loss, and hysteresis loss. Copper loss is not affected by frequency, as it depends upon the resistance of the winding and the current through the winding. Similarly, eddy-current loss is mostly a function of induced voltage rather than the frequency of that voltage. Hysteresis loss, however, increases as frequency increases.

Hysteresis loss is caused by the realignment of the magnetic domains in the core of the transformer each time the polarity of the magnetic field changes. As the frequency of the a.c. increases, the number of shifts in the magnetic field also increases (two shifts for each cycle of a.c.); therefore, the "molecular friction" increases and the hysteresis loss is greater. This increase in hysteresis loss causes the efficiency of the transformer (and therefore the amplifier) to decrease. The energy that goes into hysteresis loss is taken away from energy that could go into the signal.

RF TRANSFORMERS, specially designed for use with rf, are used to correct the problem of excessive hysteresis loss in the transformer of an rf amplifier. The windings of rf transformers are wound onto a tube of nonmagnetic material and the core is either powdered iron or air. These types of cores also reduce eddy-current loss.

Neutralization of RF Amplifiers

The problem of interelectrode capacitance in the transistor of an rf amplifier is solved by NEUTRALIZATION. Neutralization is the process of counteracting or "neutralizing" the effects of interelectrode capacitance.

Figure 2-17 shows the effect of the base-to-collector interelectrode capacitance in an rf amplifier. The "phantom" capacitor (C_{BC}) represents the interelectrode capacitance between the base and the collector of Q1. This is the interelectrode capacitance that has the most effect in an rf amplifier. As you can see, C_{BC} causes a degenerative (negative) feedback which decreases the gain of the amplifier. (There are some special cases in which C_{BC} can cause regenerative (positive) feedback. In this case, the technique described below will provide negative feedback which will accomplish the neutralization of the amplifier.)

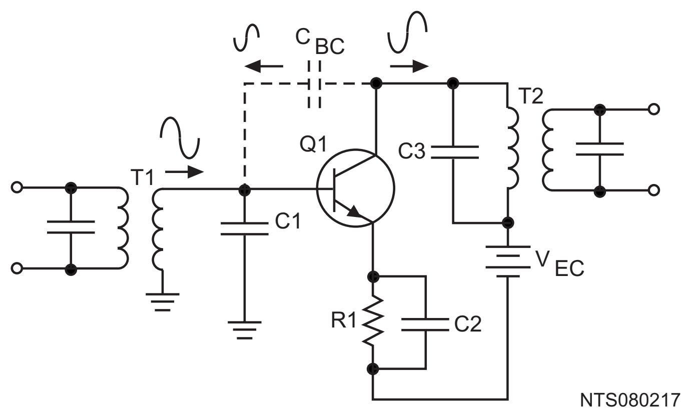


Figure 2-17.—Interelectrode capacitance in an rf amplifier.

As you may recall, unwanted degenerative feedback can be counteracted (neutralized) by using positive feedback. This is exactly what is done to neutralize an rf amplifier.

Positive feedback is accomplished by the use of a feedback capacitor. This capacitor must feed back a signal that is in phase with the signal on the base of Q1. One method of doing this is shown in figure 2-18.

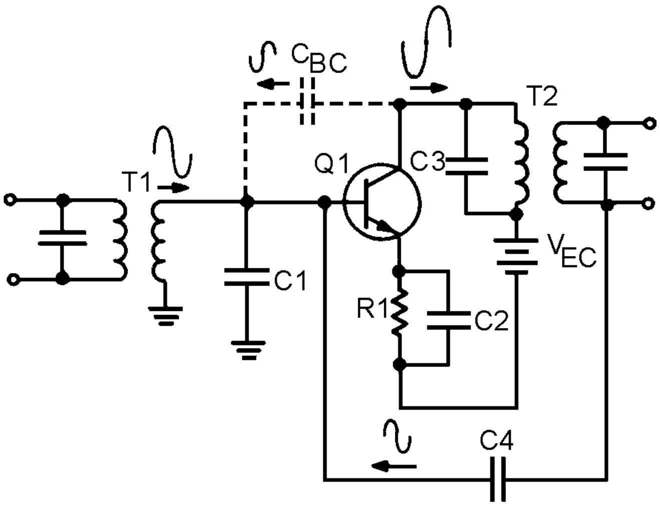


Figure 2-18.—Neutralized rf amplifier.

In figure 2-18, a feedback capacitor (C4) has been added to neutralize the amplifier. This solves the problem of unwanted degenerative feedback. Except for capacitor C4, this circuit is identical to the circuit shown in figure 2-17. (When C_{BC} causes regenerative feedback, C4 will still neutralize the amplifier. This is true because C4 always provides a feedback signal which is 180 degrees out of phase with the feedback signal caused by C_{BC}.)

Q-25. What is the most common form of coupling for an rf amplifier?

Q-26. What are two advantages of this type of coupling?

Q-27. If current gain is required from an rf amplifier, what type of component should be used as an output coupling element?

Q-28. What problem is caused in an rf amplifier by a loosely coupled transformer?

Q-29. How is this problem corrected?

Q-30. What problem is caused by overcoupling in a transformer?

Q-31. What method provides the widest bandpass?

Q-32. What two methods are used to compensate for the problems that cause low gain in an rf amplifier?

Q-33. What type of feedback is usually caused by the base-to-collector interelectrode capacitance?

Q-34. How is this compensated for?

TYPICAL RF AMPLIFIER CIRCUITS

As a technician, you will see many different rf amplifiers in many different pieces of equipment. The particular circuit configuration used for an rf amplifier will depend upon how that amplifier is used. In the final part of this chapter, you will be shown some typical rf amplifier circuits.

Figure 2-19 is the schematic diagram of a typical rf amplifier that is used in an AM radio receiver. In figure 2-19, the input circuit is the antenna of the radio (L1-a coil) which forms part of an LC circuit which is tuned to the desired station by variable capacitor C1. L1 is wound on the same core as L2, which couples the input signal through C2 to the transistor (Q1). R1 is used to provide proper bias to Q1 from the base power supply (V_{BB}). R2 provides proper bias to the emitter of Q1, and C3 is used to bypass R2. The primary of T1 and capacitor C4 form a parallel LC circuit which acts as the load for Q1. This LC circuit is tuned by C4, which is ganged to C1 allowing the antenna and the LC circuit to be tuned together. The primary of T1 is center-tapped to provide proper impedance matching with Q1.

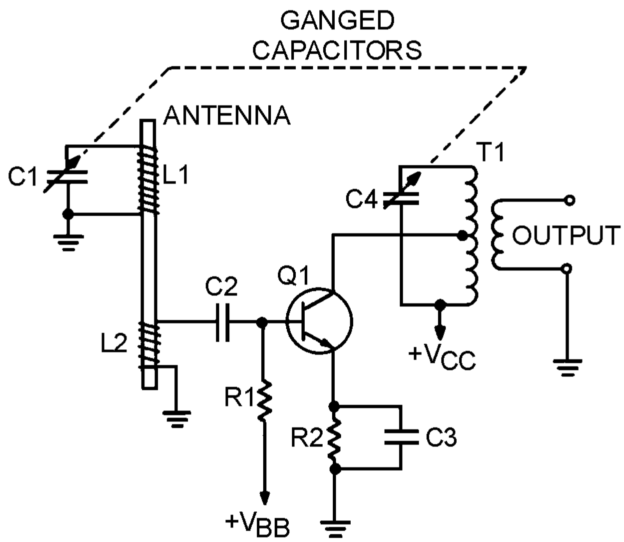


Figure 2-19.—Typical AM radio rf amplifier.

You may notice that no neutralization is shown in this circuit. This circuit is designed for the AM broadcast band (535 kHz - 1605 kHz).

At these relatively low rf frequencies the degenerative feedback caused by base-to-collector interelectrode capacitance is minor and, therefore, the amplifier does not need neutralization.

Figure 2-20 is a typical rf amplifier used in a vhf television receiver. The input-signal-developing circuit for this amplifier is made up of L1, C1, and C2. The inductor tunes the input-signal-developing circuit for the proper TV channel. (L1 can be switched out of the circuit and another inductor switched in to the circuit by the channel selector.) R1 provides proper bias to Q1 from the base supply voltage (V_{BB}). Q1 is the transistor. Notice that the case of Q1 (the dotted circle around the transistor symbol) is shown to be grounded. The case must be grounded because of the high frequencies (54 MHz - 217 MHz) used by the circuit. R2 provides proper bias from the emitter of Q1, and C3 is used to bypass R2. C5 and L2 are a parallel LC circuit which acts as the load for Q1. The LC circuit is tuned by L2 which is switched in to and out of the LC circuit by the channel selector. L3 and C6 are a parallel LC circuit which develops the signal for the next stage. The parallel LC circuit is tuned by L3 which is switched in to and out of the LC circuit by the channel selector along with L1 and L2. (L1, L2, and L3 are actually part of a bank of inductors. L1, L2, and L3 are in the circuit when the channel selector is on channel 2. For other channels, another group of three inductors would be used in the circuit.) R3 develops a signal which is fed through C4 to provide neutralization. This counteracts the effects of the interelectrode capacitance from the base to the collector of Q1. C7 is used to isolate the rf signal from the collector power supply (V_{CC}).

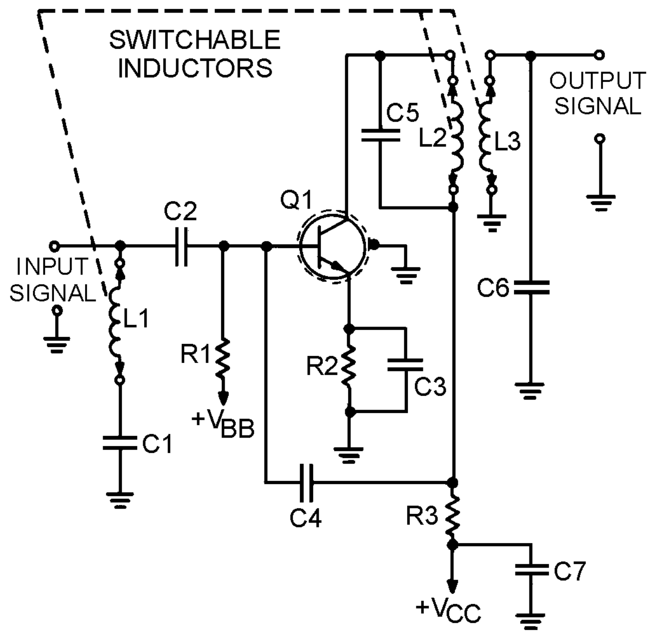


Figure 2-20.—Typical vhf television rf amplifier.

The following questions refer to figure 2-21.

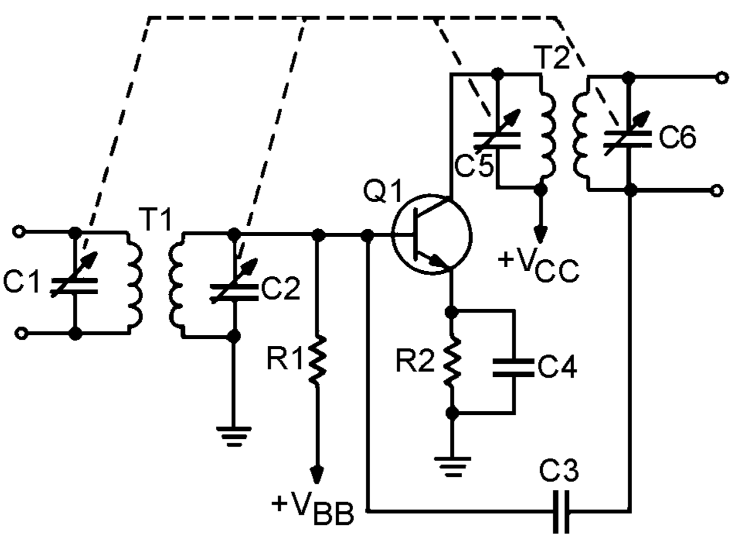


Figure 2-21.—Typical rf amplifier.

Q-35. What components form the input-signal-developing impedance for the amplifier?

Q-36. What is the purpose of R1?

Q-37. What is the purpose of R2?

Q-38. If C4 were removed from the circuit, what would happen to the output of the amplifier?

Q-39. What components form the load for Q1?

Q-40. How many tuned parallel LC circuits are shown in this schematic?

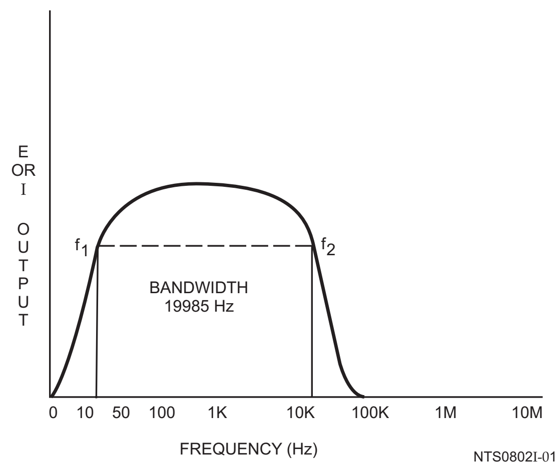
Q-41. What do the dotted lines connecting C1, C2, C5, and C6 indicate?

Q-42. What is the purpose of C3?

SUMMARY

This chapter has presented information on video and rf amplifiers. The information that follows summarizes the important points of this chapter.

A **FREQUENCY-RESPONSE CURVE** will enable you to determine the BANDWIDTH and the UPPER and LOWER FREQUENCY LIMITS of an amplifier.



The **BANDWIDTH** of an amplifier is determined by the formula:

$$BW = f_2 - f_1$$

Where:

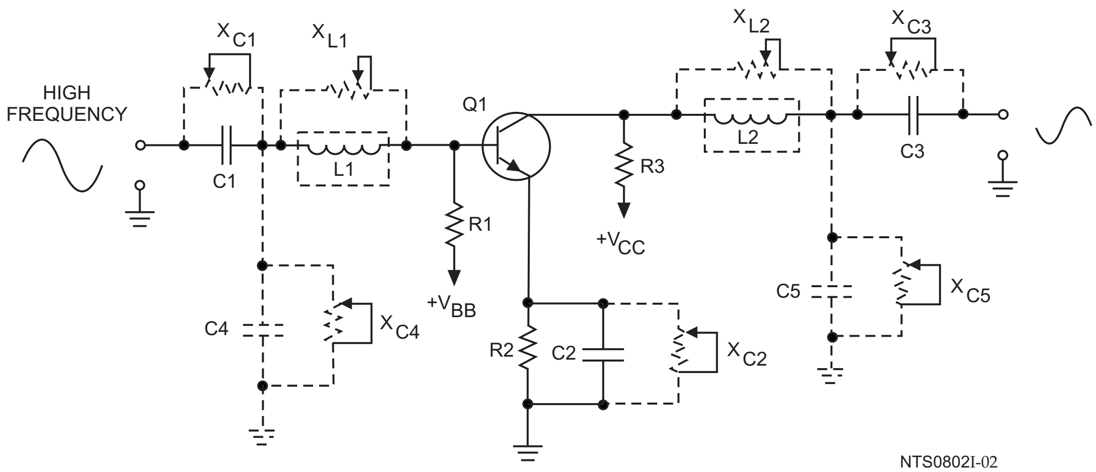
BW is the bandwidth

f_2 is the upper-frequency limit

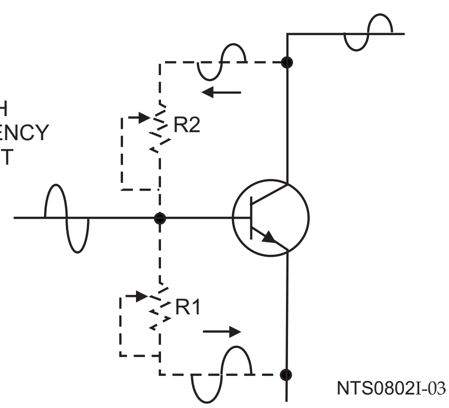
and

f_1 is the lower-frequency limit

The **UPPER-FREQUENCY RESPONSE** of an amplifier is limited by the inductance and capacitance of the circuit.



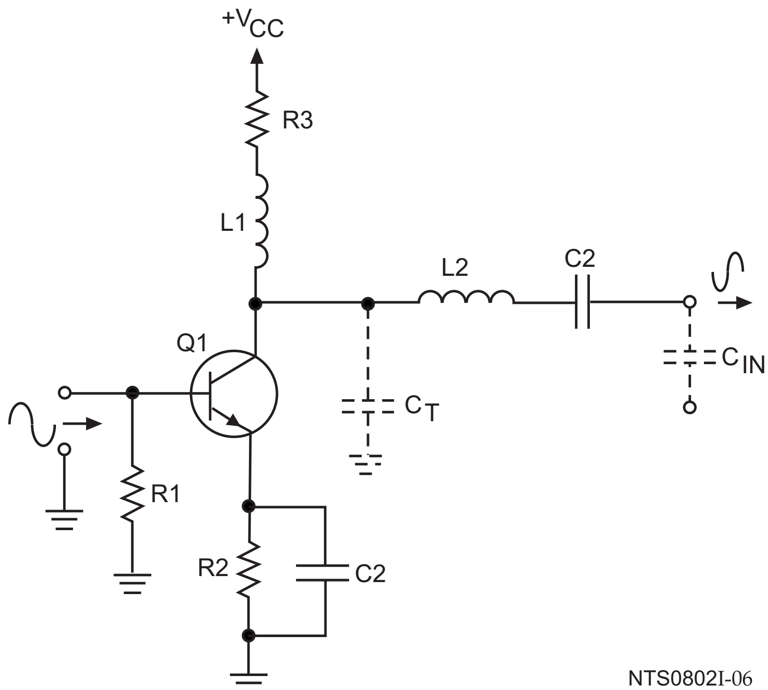
The **INTERELECTRODE CAPACITANCE** of a transistor causes DEGENERATIVE FEEDBACK at high frequencies.



VIDEO AMPLIFIERS must have a frequency response of 10 hertz to 6 megahertz (10 Hz – 6 MHz). To provide this frequency response, both high- and low-frequency compensation must be used.

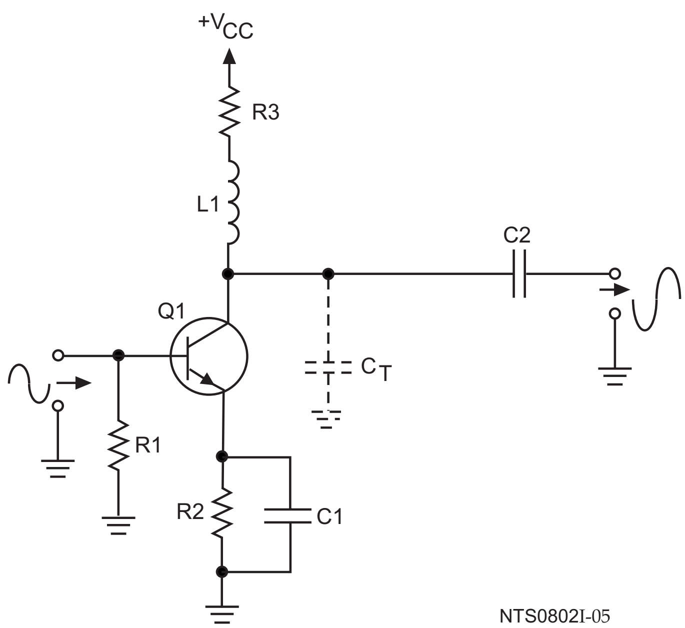
PEAKING COILS are used in video amplifiers to overcome the high-frequency limitations caused by the capacitance of the circuit.

SERIES PEAKING is accomplished by a peaking coil in series with the output-signal path.



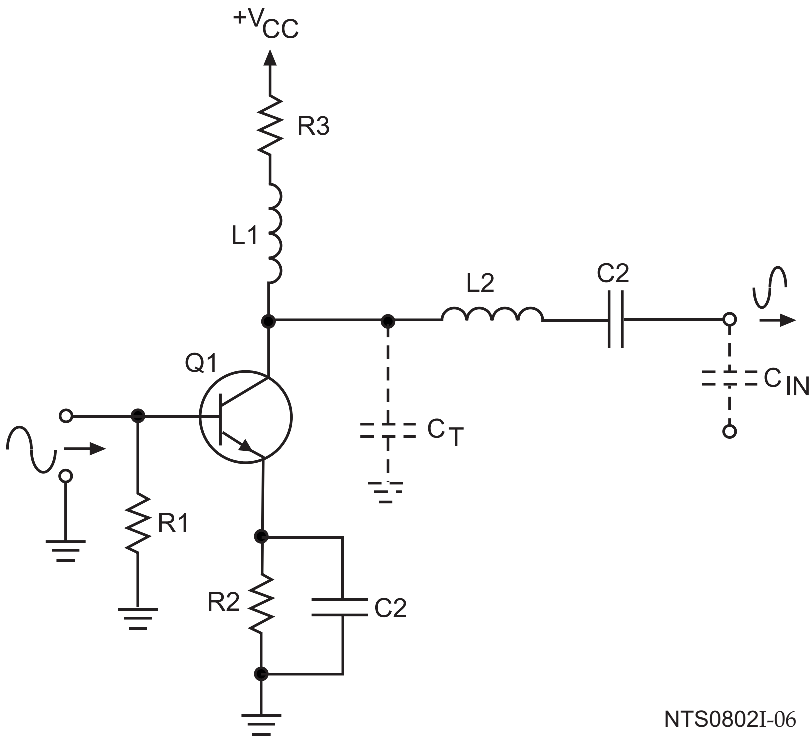
NTS0802I-06

SHUNT PEAKING is accomplished by a peaking coil in parallel (shunt) with the output-signal path.



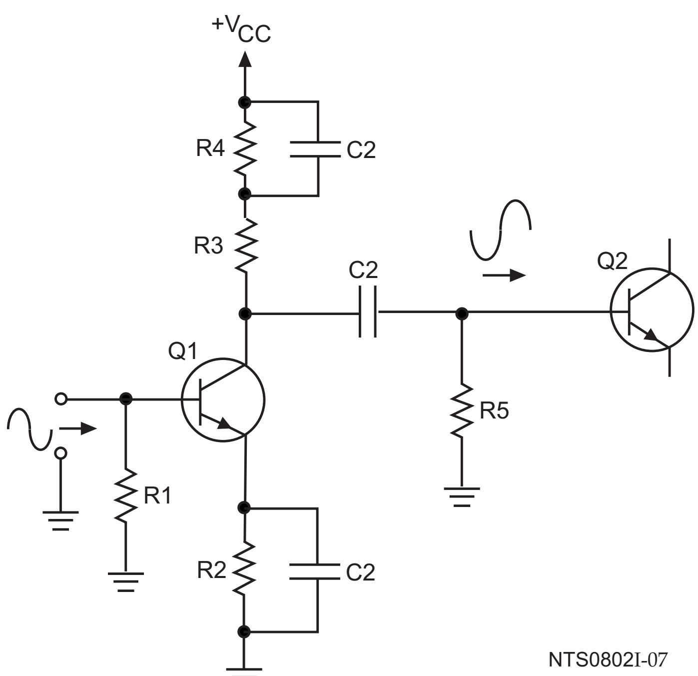
NTS0802I-05

COMBINATION PEAKING is accomplished by using both series and shunt peaking.



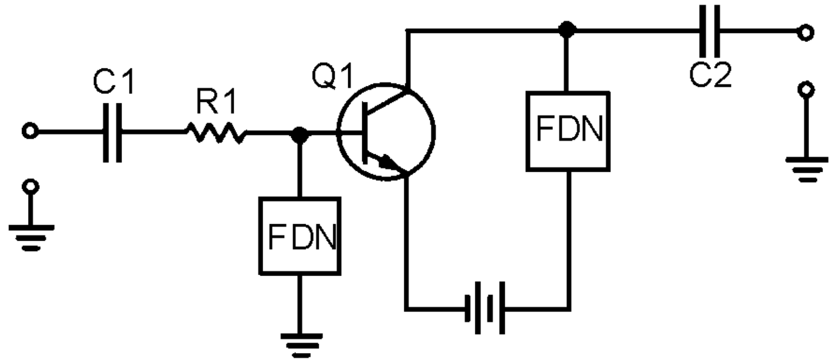
NTS0802I-06

LOW-FREQUENCY COMPENSATION is accomplished in a video amplifier by the use of a parallel RC circuit in series with the load resistor.

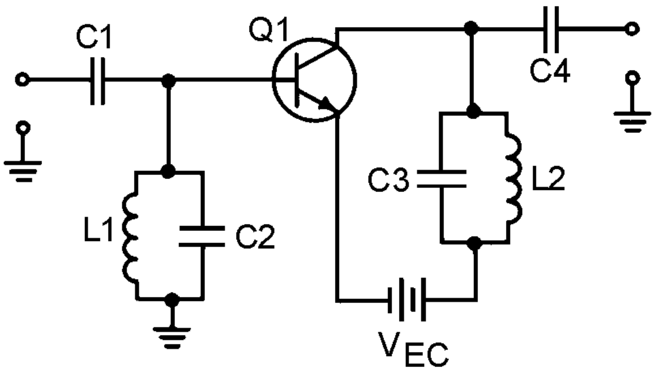


NTS0802I-07

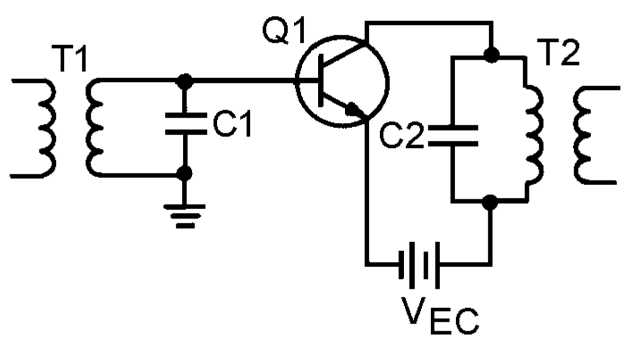
A **RADIO-FREQUENCY (RF) AMPLIFIER** uses FREQUENCY-DETERMINING NETWORKS to provide the required response at a given frequency.



The **FREQUENCY-DETERMINING NETWORK** in an rf amplifier provides maximum impedance at the desired frequency. It is a parallel LC circuit which is called a TUNED CIRCUIT.

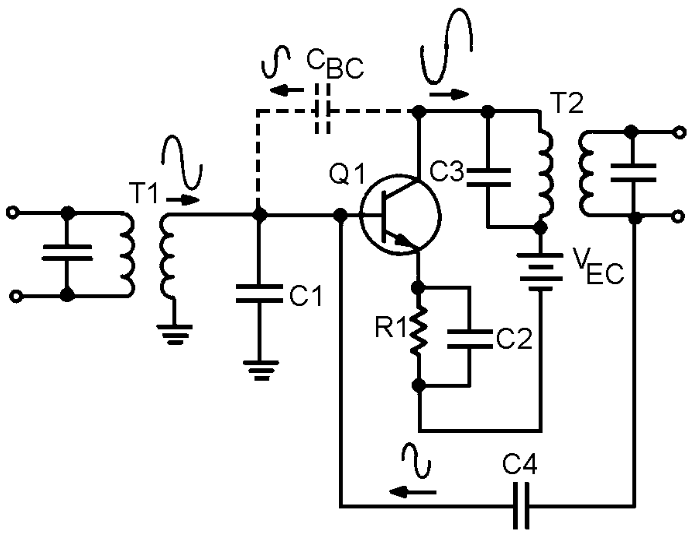


TRANSFORMER COUPLING is the most common form of coupling in rf amplifiers. This coupling is accomplished by the use of rf transformers as part of the frequency-determining network for the amplifier.



ADEQUATE BANDPASS is accomplished by optimum coupling in the rf transformer or by the use of a SWAMPING RESISTOR.

NEUTRALIZATION in an rf amplifier provides feedback (usually positive) to overcome the effects caused by the base-to-collector interelectrode capacitance.



ANSWERS TO QUESTIONS Q1. THROUGH Q42.

A-1. *The difference between the upper and lower frequency limits of an amplifier.*

A-2. *The half-power points of a frequency-response curve. The upper and lower limits of the band f frequencies for which the amplifier is most effective.*

A-3. *(A) f_2 = 80 kHz, f_1 = 30 kHz, BW = 50 kHz (B) f_2 = 4 kHz, f_1 = 2 kHz, BW = 2 kHz*

A-4. *The capacitance and inductance of the circuit and the interelectrode capacitance of the transistor.*

A-5. *Negative (degenerative) feedback.*

A-6. *It decreases.*

A-7. *It increases.*

A-8. *The capacitance of the circuit.*

A-9. *Peaking coils.*

A-10. *The relationship of the components to the output-signal path.*

A-11. *Combination peaking.*

A-12. *The coupling capacitor (C3).*

A-13. A shunt peaking coil for Q2.

A-14. A decoupling capacitor for the effects of R2.

A-15. A part of the low-frequency compensation network for Q1.

A-16. A series peaking coil for Q1.

A-17. A swamping resistor for L2.

A-18. L1, L2, and R5.

A-19. R9 and C5.

A-20. The gain increases.

A-21. The gain decreases.

A-22. To provide maximum impedance at the desired frequency.

A-23. Yes.

A-24. By changing the value.

A-25. Transformer coupling.

A-26. It uses fewer components than capacitive coupling and can provide an increase in gain.

A-27. A step-down transformer.

A-28. A too-narrow bandpass.

A-29. By using an optimumly-coupled transformer.

A-30. Low gain at the center frequency.

A-31. A swamping resistor in parallel with the tuned circuit.

A-32. RF transformers are used and the transistor is neutralized.

A-33. Degenerative or negative.

A-34. By neutralization such as the use of a capacitor to provide regenerative (positive) feedback.

A-35. C2 and the secondary of T1.

A-36. R1 provides the proper bias to the base of Q1 from V_{BB}.

A-37. R2 provides the proper bias to the emitter of Q1.

A-38. The output would decrease. (C4 decouples R2 preventing degenerative feedback from R2.)

A-39. C5 and the primary of T2.

A-40. Four.

A-41. The dotted lines indicate that these capacitors are "ganged" and are tuned together with a single control.

A-42. C3 provides neutralization for Q1.

CHAPTER 3

SPECIAL AMPLIFIERS

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to:

1. Describe the basic operation of a differential amplifier.
2. Describe the operation of a differential amplifier under the following conditions:
   a. Single Input, Single Output
   b. Single input, differential output
   c. Differential input, differential output
3. List the characteristics of an operational amplifier.
4. Identify the symbol for an operational amplifier.
5. Label the blocks on a block diagram of an operational amplifier.
6. Describe the operation of an operational amplifier with inverting and noninverting configurations.
7. Describe the bandwidth of a typical operational amplifier and methods to modify the bandwidth.
8. Identify the following applications of operational amplifiers:
   a. Adder
   b. Subtractor
9. State the common usage for a magnetic amplifier.
10. Describe the basic operation of a magnetic amplifier.
11. Describe various methods of changing inductance.
12. Identify the purpose of components in a simple magnetic amplifier.

INTRODUCTION

If you were to make a quick review of the subjects discussed in this module up to this point, you would see that you have been given a considerable amount of information about amplifiers. You have been shown what amplification is and how the different classes of amplifiers affect amplification. You also have been shown that many factors must be considered when working with amplifiers, such as impedance, feedback, frequency response, and coupling. With all this information behind you, you might ask yourself "what more can there be to know about amplifiers?"

There is a great deal more to learn about amplifiers. Even after you finish this chapter you will have only "scratched the surface" of the study of amplifiers. But, you will have prepared yourself for the remainder of the *NEETS*. This, in turn, should prepare you for further study and, perhaps, a career in electronics.

As in chapter 2, the circuits shown in this chapter are intended to present particular concepts to you. Therefore, the circuits may be incomplete or not practical for use in an actual piece of electronic equipment. You should keep in mind the fact that this text is intended to teach certain facts about amplifiers, and in order to simplify the illustrations used, complete operational circuits are not always shown.

In this chapter three types of special amplifiers are discussed. These are: DIFFERENTIAL AMPLIFIERS, OPERATIONAL AMPLIFIERS, and MAGNETIC AMPLIFIERS. These are called special amplifiers because they are used only in certain types of equipment.

The names of each of these special amplifiers describe the operation of the amplifier, NOT what is amplified. For example, a magnetic amplifier does not amplify magnetism but uses magnetic effects to produce amplification of an electronic signal.

A differential amplifier is an amplifier that can have two input signals and/or two output signals. This amplifier can amplify the difference between two input signals. A differential amplifier will also "cancel out" common signals at the two inputs.

One of the more interesting aspects of an operational amplifier is that it can be used to perform mathematical operations electronically. Properly connected, an operational amplifier can add, subtract, multiply, divide, and even perform the calculus operations of integration and differentiation. These amplifiers were originally used in a type of computer known as the "analog computer" but are now used in many electronic applications.

The magnetic amplifier uses a device called a "saturable core reactor" to control an a.c.output signal. The primary use of magnetic amplifiers is in power control systems.

These brief descriptions of the three special amplifiers are intended to provide you with a general idea of what these amplifiers are and how they can be used. The remaining sections of this chapter will provide you with more detailed information on these special amplifiers.

DIFFERENTIAL AMPLIFIERS

A differential amplifier has two possible inputs and two possible outputs. This arrangement means that the differential amplifier can be used in a variety of ways. Before examining the three basic configurations that are possible with a differential amplifier, you need to be familiar with the basic circuitry of a differential amplifier.

BASIC DIFFERENTIAL AMPLIFIER CIRCUIT

Before you are shown the operation of a differential amplifier, you will be shown how a simpler circuit works. This simpler circuit, known as the DIFFERENCE AMPLIFIER, has one thing in common with the differential amplifier: It operates on the difference between two inputs. However, the difference amplifier has only one output while the differential amplifier can have two outputs.

By now, you should be familiar with some amplifier circuits, which should give you an idea of what a difference amplifier is like. In *NEETS, Module 7*, you were shown the basic configurations for transistor amplifiers. Figure 3-1 shows two of these configurations: the common emitter and the common base.

In view (A) of figure 3-1 a common-emitter amplifier is shown. The output signal is an amplified version of the input signal and is 180 degrees out of phase with the input signal. View (B) is a common-base amplifier. In this circuit the output signal is an amplified version of the input signal and is in phase with the input signal. In both of these circuits, the output signal is controlled by the base-to-emitter bias. As this bias changes (because of the input signal) the current through the transistor changes. This causes the output signal developed across the collector load (R2) to change. None of this information is new, it is just a review of what you have already been shown regarding transistor amplifiers.

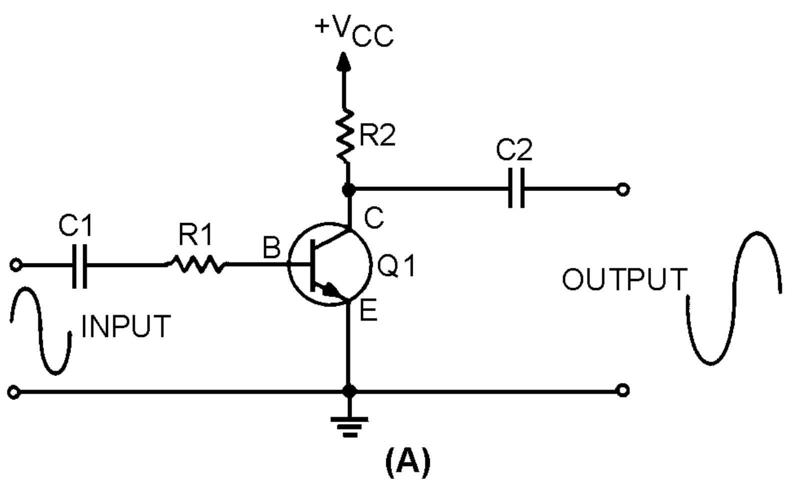


Figure 3-1A.—Common-emitter and common-base amplifiers.

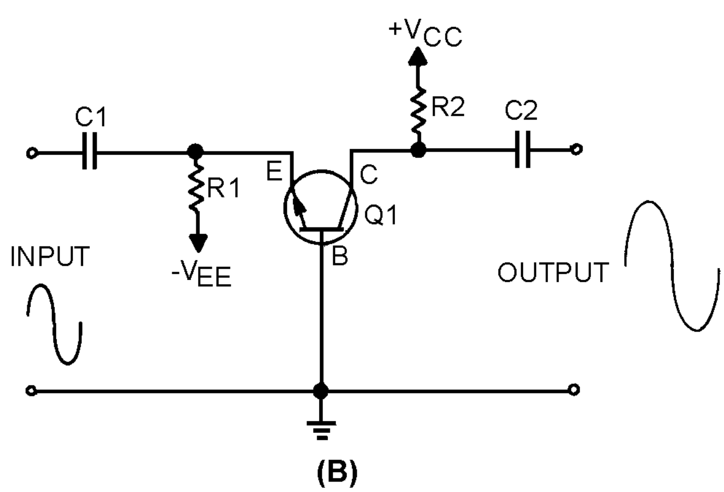


Figure 3-1B.—Common-emitter and common-base amplifiers.

NOTE: Bias arrangements for the following explanations will be termed base-to-emitter. In other publications you will see the term emitter-to-base used to describe the same bias arrangement.

THE TWO-INPUT, SINGLE-OUTPUT, DIFFERENCE AMPLIFIER

If you combine the common-base and common-emitter configurations into a single transistor amplifier, you will have a circuit like the one shown in figure 3-2. This circuit is the two-input, single-output, difference amplifier.

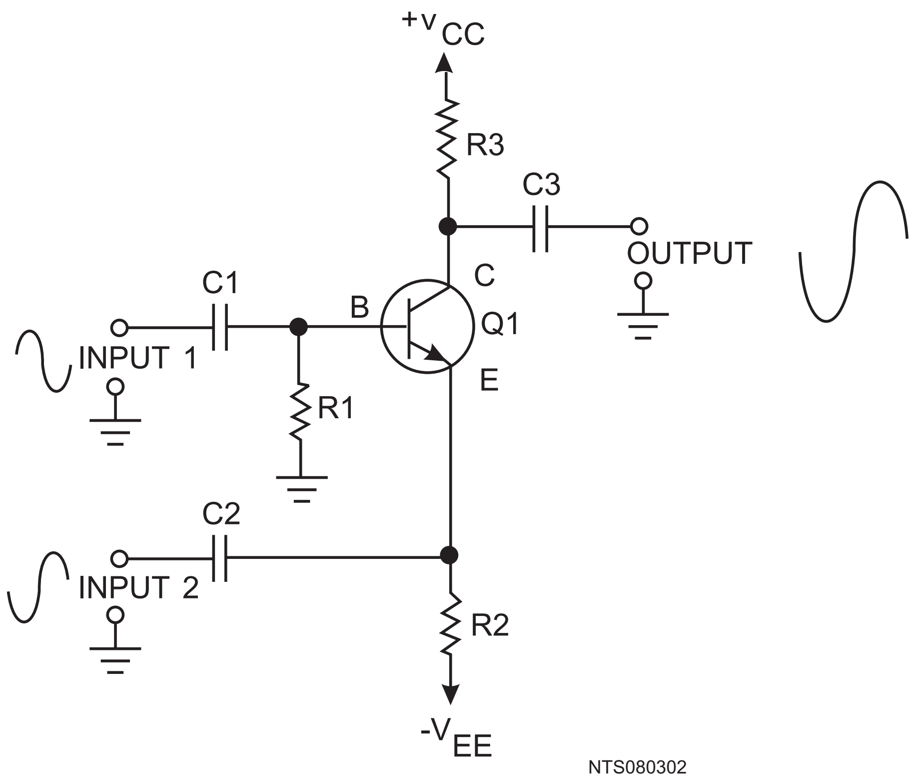


Figure 3-2.—Two-input, single-output, difference amplifier.

In figure 3-2, the transistor has two inputs (the emitter and the base) and one output (the collector). Remember, the current through the transistor (and therefore the output signal) is controlled by the base-to-emitter bias. In the circuit shown in figure 3-2, the combination of the two input signals controls the output signal. In fact, the DIFFERENCE BETWEEN THE INPUT SIGNALS determines the base-to-emitter bias.

For the purpose of examining the operation of the circuit shown in figure 3-2, assume that the circuit has a gain of -10. This means that for each 1-volt change in the base-to-emitter bias, there would be a 10-volt change in the output signal. Assume, also, that the input signals will peak at 1-volt levels (+1 volt for the positive peak and -1 volt for the negative peak). The secret to understanding this circuit (or any transistor amplifier circuit) is to realize that the collector current is controlled by the base-to-emitter bias. In other words, in this circuit the output signal (the voltage developed across R3) is <u>determined by the difference between the voltage on the base and the voltage on the emitter</u>.

Figure 3-3 shows this two-input, single-output amplifier with input signals that are equal in amplitude and 180 degrees out of phase. Input number one has a positive alternation when input number two has a negative alternation and vice versa.

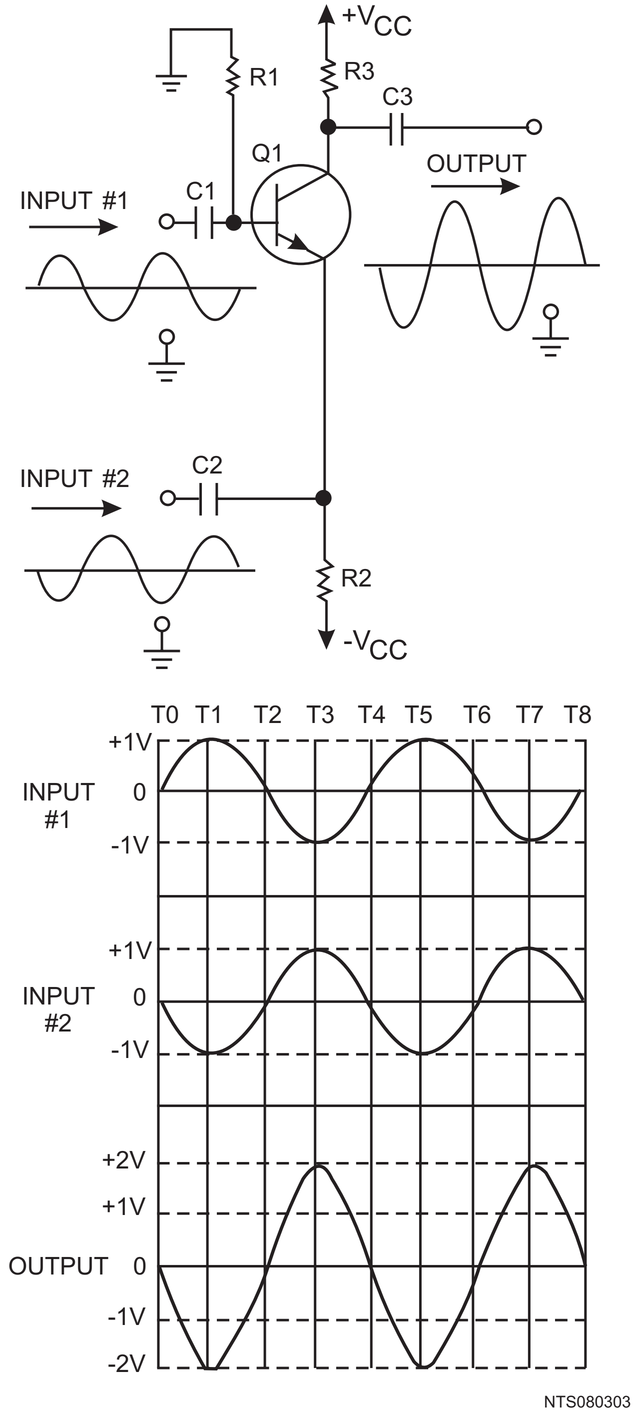


Figure 3-3.—Input signals 180º out of phase.

The circuit and the input and output signals are shown at the top of the figure. The lower portion of the figure is a comparison of the input signals and the output signal. Notice the vertical lines marked "T0" through "T8." These represent "time zero" through "time eight." In other words, these lines provide a way to examine the two input signals and the output signal at various instants of time.

In figure 3-3 at time zero (T0) both input signals are at 0 volts. The output signal is also at 0 volts. Between time zero (T0) and time one (T1), input signal number one goes positive and input signal number two goes negative. Each of these voltage changes causes an increase in the base-to-emitter bias which causes current through Q1 to increase. Increased current through Q1 results in a greater voltage drop across the collector load (R3) which causes the output signal to go negative.

By time one (T1), input signal number one has reached +1 volt and input signal number two has reached –1 volt. This is an overall increase in base-to-emitter bias of 2 volts. Since the gain of the circuit is –10, the output signal has decreased by 20 volts. As you can see, the output signal has been determined by the <u>difference between the two input signals</u>. In fact, the base-to-emitter bias can be found by subtracting the value of input signal number two from the value of input signal number one.

Mathematically:

Bias = (input signal #1) - (input signal #2)
Bias = (+1V) - (1V)
Bias = +1V + 1V
Bias = +2V

Between time one (T1) and time two (T2), input signal number one goes from +1 volt to 0 volts and input signal number two goes from –1 volt to 0 volts. At time two (T2) both input signals are at 0 volts and the base-to-emitter bias has returned to 0 volts. The output signal is also 0 volts.

Mathematically:

Bias = (input signal #1) - (input signal #2)
Bias = (0V) - (0V)
Bias = 0V

Between time two (T2) and time three (T3), input signal number one goes negative and input signal number two goes positive. At time three (T3), the value of the base-to-emitter bias is –2 volts.

Mathematically:

Bias = (input signal #1) - (input signal #2)
Bias = (-1V) - (+1V)
Bias = (-1V) + (-1V)
Bias = -2V

This causes the output signal to be +20 volts at time three (T3).

Between time three (T3) and time four (T4), input signal #1 goes from –1 volt to 0 volts and input signal #2 goes from +1 volt to 0 volts. At time four (T4) both input signals are 0 volts, the bias is 0 volts, and the output is 0 volts.

During time four (T4) through time eight (T8), the circuit repeats the sequence of events that took place from time zero (T0) through time four (T4).

You can see that when the input signals are equal in amplitude and 180 degrees out of phase, the output signal is twice as large (40 volts peak to peak) as it would be from either input signal alone (if the other input signal were held at 0 volts).

Figure 3-4 shows the two-input, single-output, difference amplifier with two input signals that are equal in amplitude and in phase.

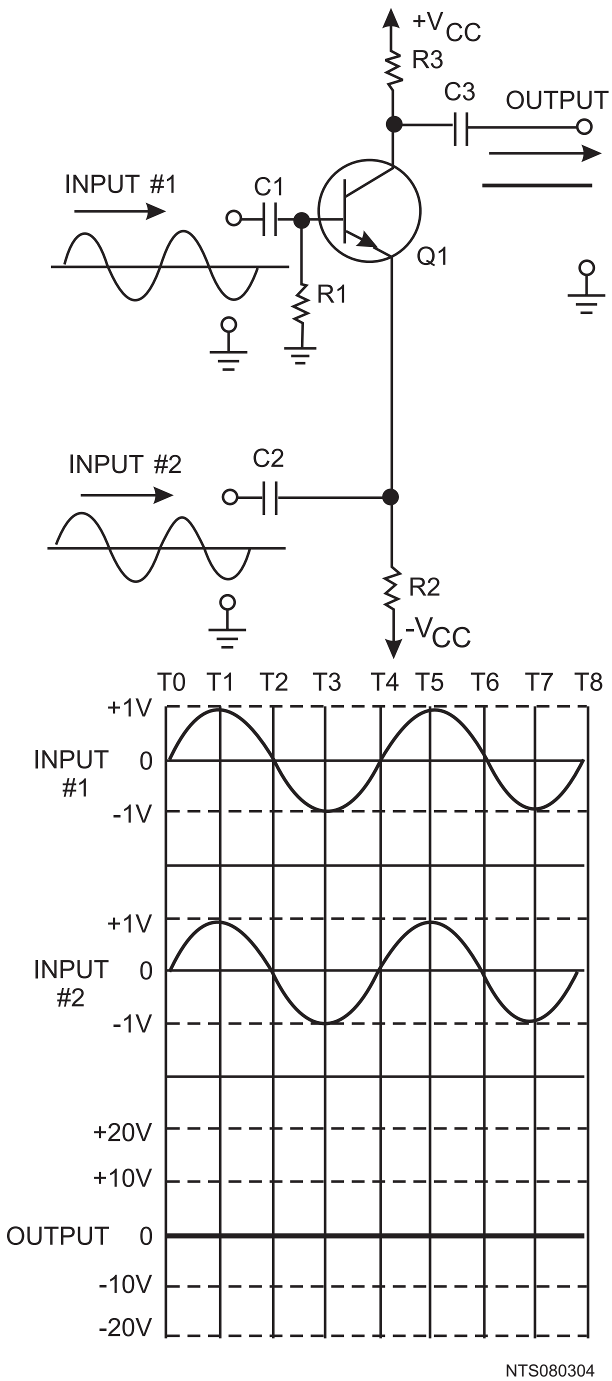


Figure 3-4.—Input signals in phase.

Notice, that the output signal remains at 0 volts for the entire time (T0 - T8). Since the two input signals are equal in amplitude and in phase, the difference between them (the base-to-emitter bias) is always 0 volts. This causes a 0-volt output signal.

If you compute the bias at any time period (T0 - T8), you will see that the output of the circuit remains at a constant zero.

For example:

Bias = (input signal #1) - (input signal #2)
T1 Bias = (+1V) - (+1V) = 0
and so forth

From the above example, you can see that when the input signals are equal in amplitude and in phase, there is no output from the difference amplifier because there is no difference between the two inputs. You also know that when the input signals are equal in amplitude but 180 degrees out of phase, the output looks just like the input except for amplitude and a 180-degree phase reversal with respect to input signal number one. What happens if the input signals are equal in amplitude but different in phase by something other than 180 degrees? This would mean that sometimes one signal would be going negative while the other would be going positive; sometimes both signals would be going positive; and sometimes both signals would be going negative. Would the output signal still look like the input signals? The answer is "no," because figure 3-5 shows a difference amplifier with two input signals that are equal in amplitude but 90 degrees out of phase. From the figure you can see that at time zero (T0) input number one is at 0 volts and input number two is at -1 volt. The base-to-emitter bias is found to be +1 volt.

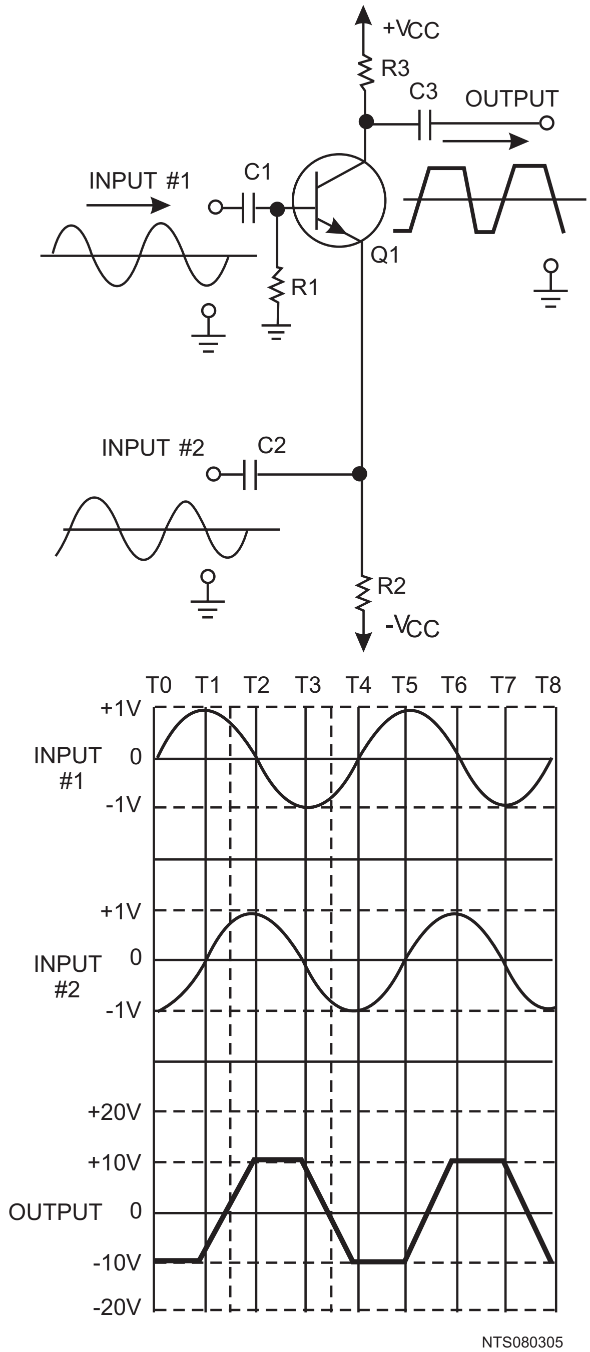


Figure 3-5.—Input signals 90° out of phase.

This +1-volt bias signal causes the output signal to be –10 volts at time zero (T0). Between time zero (T0) and time one (T1), both input signals go positive. The difference between the input signals stays constant. The effect of this is to keep the bias at +1 volt for the entire time between T0 and T1. This, in turn, keeps the output signal at -10 volts.

Between time one (T1) and time two (T2), input signal number one goes in a negative direction but input signal number two continues to go positive. Now the difference between the input signals decreases rapidly from +1 volt. Halfway between T1 and T2 (the dotted vertical line), input signal number one and input signal number two are equal in amplitude. The difference between the input signals is 0 volts and this causes the output signal to be 0 volts. From this point to T2 the difference between the input signals is a negative value. At T2:

Bias = (input signal #1) - (input signal #2)
Bias = (0V) - (+1V)
Bias = +1V

From time two (T2) to time three (T3), input signal number one goes negative and input signal number two goes to zero. The difference between them stays constant at -1 volt. Therefore, the output signal stays at a +10-volt level for the entire time period from T2 to T3. At T3 the bias condition will be:

Bias = (input signal #1) - (input signal #2)
Bias = (0V) - (-1V)
Bias = +1V

Between T3 and T4 input signal number one goes to zero while input signal number two goes negative. This, again, causes a rapid change in the difference between the input signals. Halfway between T3 and T4 (the dotted vertical line) the two input signals are equal in amplitude; therefore, the difference between the input signals is 0 volts, and the output signal becomes 0 volts. From that point to T4, the difference between the input signals becomes a positive voltage. At T4:

Bias = (input signal #1) - (input signal #2)
Bias = (0V) - (-1V)
Bias = +1V

(The sequence of events from T4 to T8 are the same as those of T0 to T4.)

As you have seen, this amplifier amplifies the difference between two input signals. But this is NOT a differential amplifier. A differential amplifier has two inputs and two outputs. The circuit you have just been shown has only one output. Well then, how does a differential amplifier schematic look?

TYPICAL DIFFERENTIAL AMPLIFIER CIRCUIT

Figure 3-6 is the schematic diagram of a typical differential amplifier. Notice that there are two inputs and two outputs. This circuit requires two transistors to provide the two inputs and two outputs. If you look at one input and the transistor with which it is associated, you will see that each transistor is a common-emitter amplifier for that input (input one and Q1; input two and Q2). R1 develops the signal at input one for Q1, and R5 develops the signal at input two for Q2. R3 is the emitter resistor for both Q1 and Q2. Notice that R3 is NOT bypassed. This means that when a signal at input one affects the current through Q1, that signal is developed by R3. (The current through Q1 must flow through R3; as this current changes, the voltage developed across R3 changes.) When a signal is developed by R3, it is applied to the emitter of Q2. In the same way, signals at input two affect the current of Q2, are developed by R3, and are felt on the emitter of Q1. R2 develops the signal for output one, and R4 develops the signal for output two.

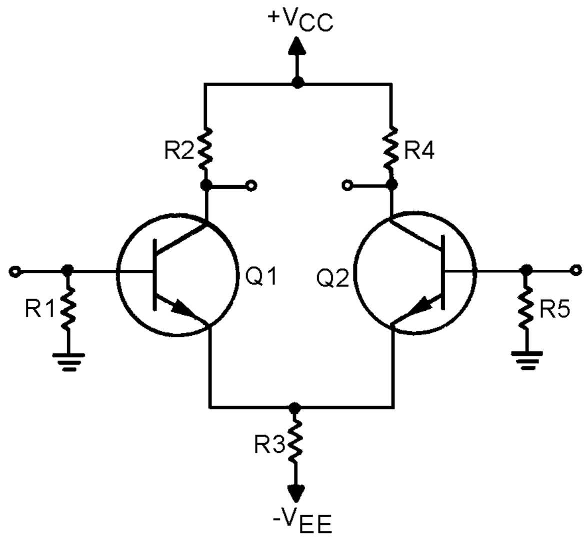


Figure 3-6.—Differential amplifier.

Even though this circuit is designed to have two inputs and two outputs, it is not necessary to use both inputs and both outputs. (Remember, a differential amplifier was defined as having two possible inputs and two possible outputs.) A differential amplifier can be connected as a single-input, single-output device; a single-input, differential-output device; or a differential-input, differential-output device.

Q-1. How many inputs and outputs are possible with a differential amplifier?

Q-2. What two transistor amplifier configurations are combined in the single-transistor, two-input, single-output difference amplifier?

Q-3. If the two input signals of a difference amplifier are in phase and equal in amplitude, what will the output signal be?

Q-4. If the two input signals to a difference amplifier are equal in amplitude and 180 degrees out of phase, what will the output signal be?

Q-5. If only one input signal is used with a difference amplifier, what will the output signal be?

Q-6. If the two input signals to a difference amplifier are equal in amplitude but neither in phase nor 180 degrees out of phase, what will the output signal be?

SINGLE-INPUT, SINGLE-OUTPUT, DIFFERENTIAL AMPLIFIER

Figure 3-7 shows a differential amplifier with one input (the base of Q1) and one output (the collector of Q2). The second input (the base of Q2) is grounded and the second output (the collector of Q1) is not used.

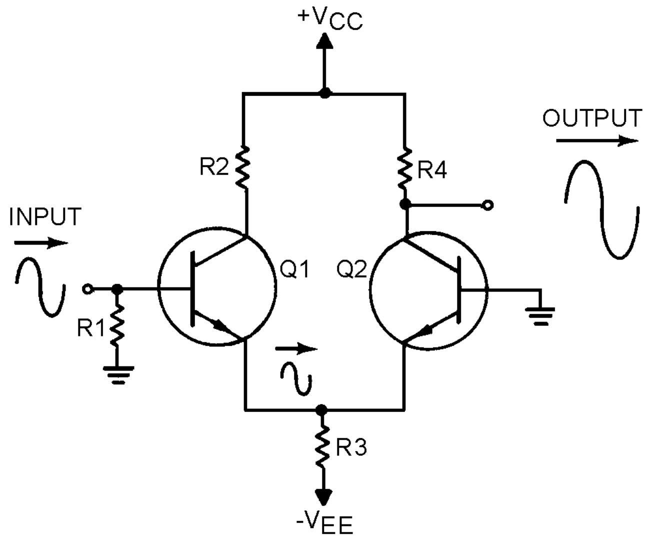


Figure 3-7.—Single-input, single-output differential amplifier.

When the input signal developed by R1 goes positive, the current through Q1 increases. This increased current causes a positive-going signal at the top of R3. This signal is felt on the emitter of Q2. Since the base of Q2 is grounded, the current through Q2 decreases with a positive-going signal on the emitter. This decreased current causes less voltage drop across R4. Therefore, the voltage at the bottom of R4 increases and a positive-going signal is felt at the output.

When the input signal developed by R1 goes negative, the current through Q1 decreases. This decreased current causes a negative-going signal at the top of R3. This signal is felt on the emitter of Q2. When the emitter of Q2 goes negative, the current through Q2 increases. This increased current causes more of a voltage drop across R4. Therefore, the voltage at the bottom of R4 decreases and a negative-going signal is felt at the output.

This single-input, single-output, differential amplifier is very similar to a single-transistor amplifier as far as input and output signals are concerned. This use of a differential amplifier does provide amplification of a.c. or d.c. signals but does not take full advantage of the characteristics of a differential amplifier.

SINGLE-INPUT, DIFFERENTIAL-OUTPUT, DIFFERENTIAL AMPLIFIER

In chapter one of this module you were shown several phase splitters. You should remember that a phase splitter provides two outputs from a single input. These two outputs are 180 degrees out of phase with each other. The single-input, differential-output, differential amplifier will do the same thing.

Figure 3-8 shows a differential amplifier with one input (the base of Q1) and two outputs (the collectors of Q1 and Q2). One output is in phase with the input signal, and the other output is 180 degrees out of phase with the input signal. The outputs are differential outputs.

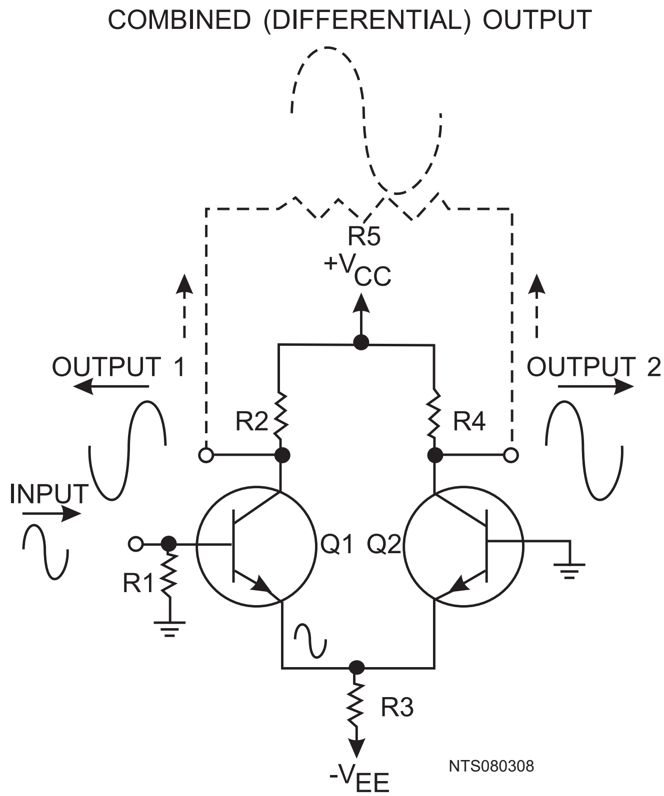


Figure 3-8.—Single-input, differential-output differential amplifier.

This circuit's operation is the same as for the single-input, single-output differential amplifier just described. However, another output is obtained from the bottom of R2. As the input signal goes positive, thus causing increased current through Q1, R2 has a greater voltage drop. The output signal at the bottom of R2 therefore is negative going. A negative-going input signal will decrease current and reverse the polarities of both output signals.

Now you see how a differential amplifier can produce two amplified, differential output signals from a single-input signal. One further point of interest about this configuration is that if a combined output signal is taken between outputs number one and two, this single output will be twice the amplitude of the individual outputs. In other words, you can double the gain of the differential amplifier (single output) by taking the output signal between the two output terminals. This single-output signal will be in phase with the input signal. This is shown by the phantom signal above R5 (the phantom resistor connected between outputs number one and two would be used to develop this signal).

DIFFERENTIAL-INPUT, DIFFERENTIAL-OUTPUT, DIFFERENTIAL AMPLIFIER

When a differential amplifier is connected with a differential input and a differential output, the full potential of the circuit is used. Figure 3-9 shows a differential amplifier with this type of configuration (differential-input, differential-output).

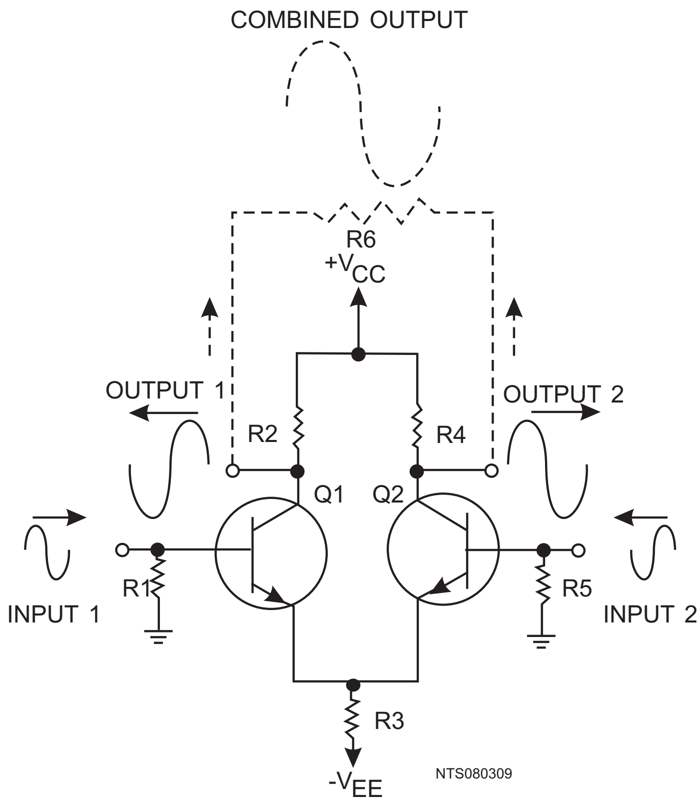


Figure 3-9.—Differential-input, differential-output differential amplifier.

Normally, this configuration uses two input signals that are 180 degrees out of phase. This causes the difference (differential) signal to be twice as large as either input alone. (This is just like the two-input, single-output difference amplifier with input signals that are 180 degrees out of phase.)

Output number one is a signal that is in phase with input number two, and output number two is a signal that is in phase with input number one. The amplitude of each output signal is the input signal multiplied by the gain of the amplifier. With 180-degree-out-of-phase input signals, each output signal is greater in amplitude than either input signal by a factor of the gain of the amplifier.

When an output signal is taken between the two output terminals of the amplifier (as shown by the phantom connections, resistor, and signal), the combined output signal is twice as great in amplitude as either signal at output number one or output number two. (This is because output number one and output number two are 180 degrees out of phase with each other.) When the input signals are 180 degrees out of phase, the amplitude of the combined output signal is equal to the amplitude of one input signal multiplied by two times the gain of the amplifier.

When the input signals are not 180 degrees out of phase, the combined output signal taken across output one and output two is similar to the output that you were shown for the two-input, single-output, difference amplifier. The differential amplifier can have two outputs (180 degrees out of phase with each other), or the outputs can be combined as shown in figure 3-9.

In answering Q7 through Q9 use the following information: All input signals are sine waves with a peak-to-peak amplitude of 10 millivolts. The gain of the differential amplifier is 10.

Q-7. If the differential amplifier is configured with a single input and a single output, what will the peak-to-peak amplitude of the output signal be?

Q-8. If the differential amplifier is configured with a single input and differential outputs, what will the output signals be?

Q-9. If the single-input, differential-output, differential amplifier has an output signal taken between the two output terminals, what will the peak-to-peak amplitude of this combined output be?

In answering Q10 through Q14 use the following information: A differential amplifier is configured with a differential input and a differential output. All input signals are sine waves with a peak-to-peak amplitude of 10 millivolts. The gain of the differential amplifier is 10.

Q-10. If the input signals are in phase, what will be the peak-to-peak amplitude of the output signals?

Q-11. If the input signals are 180 degrees out of phase with each other, what will be the peak-to-peak amplitude of the output signals?

Q-12. If the input signals are 180 degrees out of phase with each other, what will the phase relationship be between (a) the output signals and (b) the input and output signals?

Q-13. If the input signals are 180 degrees out of phase with each other and a combined output is taken between the two output terminals, what will the amplitude of the combined output signal be?

Q-14. If the input signals are 90 degrees out of phase with each other and a combined output is taken between the two output terminals, (a) what will the peak-to-peak amplitude of the combined output signal be, and (b) will the combined output signal be a sine wave?

OPERATIONAL AMPLIFIERS

An OPERATIONAL AMPLIFIER (OP AMP) is an amplifier which is designed to be used with other circuit components to perform either computing functions (addition, subtraction) or some type of transfer operation, such as filtering. Operational amplifiers are usually high-gain amplifiers with the amount of gain determined by feedback.

Operational amplifiers have been in use for some time. They were originally developed for analog (non-digital) computers and used to perform mathematical functions. Operational amplifiers were not used in other devices very much because they were expensive and more complicated than other circuits.

Today many devices use operational amplifiers. Operational amplifiers are used as d.c. amplifiers, a.c. amplifiers, comparators, oscillators (which are covered in *NEETS, Module 9*), filter circuits, and many other applications. The reason for this widespread use of the operational amplifier is that it is a very versatile and efficient device. As an integrated circuit (chip) the operational amplifier has become an inexpensive and readily available "building block" for many devices. In fact, an operational amplifier in integrated circuit form is no more expensive than a good transistor.

CHARACTERISTICS OF AN OPERATIONAL AMPLIFIER

The schematic symbols for an operational amplifier are shown in figure 3-10. View (A) shows the power supply requirements while view (B) shows only the input and output terminals. An operational amplifier is a special type of high-gain, d.c. amplifier. To be classified as an operational amplifier, the circuit must have certain characteristics. The three most important characteristics of an operational amplifier are:

1. Very high gain
2. Very high input impedance
3. Very low output impedance

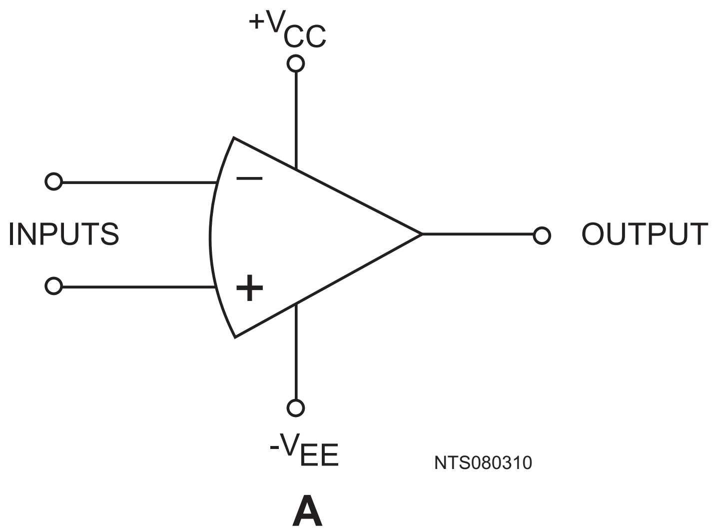


Figure 3-10A.—Schematic symbols of an operational amplifier.

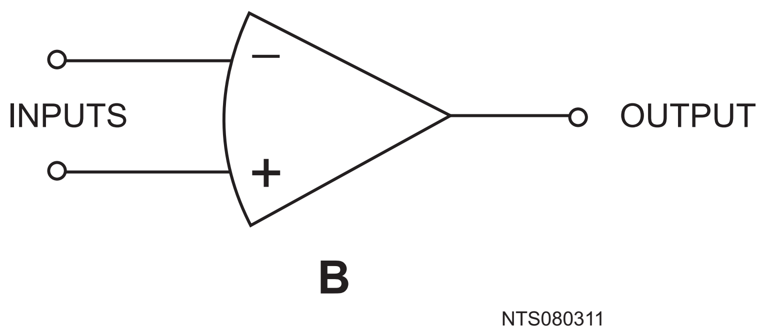


Figure 3-10B.—Schematic symbols of an operational amplifier.

Since no single amplifier stage can provide all these characteristics well enough to be considered an operational amplifier, various amplifier stages are connected together. The total circuit made up of these individual stages is called an operational amplifier. This circuit (the operational amplifier) can be made up of individual components (transistors, resistors, capacitors, etc.), but the most common form of the operational amplifier is an integrated circuit. The integrated circuit (chip) will contain the various stages of the operational amplifier and can be treated and used as if it were a single stage.

BLOCK DIAGRAM OF AN OPERATIONAL AMPLIFIER

Figure 3-11 is a block diagram of an operational amplifier. Notice that there are three stages within the operational amplifier.

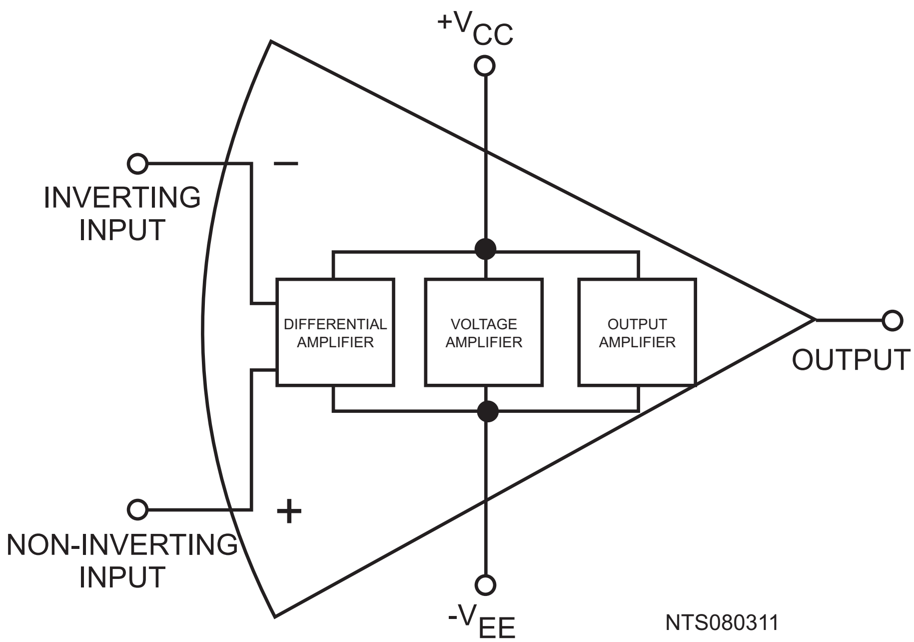


Figure 3-11.—Block diagram of an operational amplifier.

The input stage is a differential amplifier. The differential amplifier used as an input stage provides differential inputs and a frequency response down to d.c. Special techniques are used to provide the high input impedance necessary for the operational amplifier.

The second stage is a high-gain voltage amplifier. This stage may be made from several transistors to provide high gain. A typical operational amplifier could have a voltage gain of 200,000. Most of this gain comes from the voltage amplifier stage.

The final stage of the OP AMP is an output amplifier. The output amplifier provides low output impedance. The actual circuit used could be an emitter follower. The output stage should allow the operational amplifier to deliver several milliamperes to a load.

Notice that the operational amplifier has a positive power supply ($+V_{CC}$) and a negative power supply ($-V_{EE}$). This arrangement enables the operational amplifier to produce either a positive or a negative output.

The two input terminals are labeled "inverting input" (−) and "noninverting input" (+). The operational amplifier can be used with three different input conditions (modes). With differential inputs (first mode), both input terminals are used and two input signals which are 180 degrees out of phase with each other are used. This produces an output signal that is in phase with the signal on the noninverting input. If the noninverting input is grounded and a signal is applied to the inverting input (second mode), the output signal will be 180 degrees out of phase with the input signal (and one-half the amplitude of the first mode output). If the inverting input is grounded and a signal is applied to the noninverting input (third mode), the output signal will be in phase with the input signal (and one-half the amplitude of the first mode output).

Q-15. What are the three requirements for an operational amplifier?

Q-16. What is the most commonly used form of the operational amplifier?

Q-17. Draw the schematic symbol for an operational amplifier.

Q-18. Label the parts of the operational amplifier shown in figure 3-12.

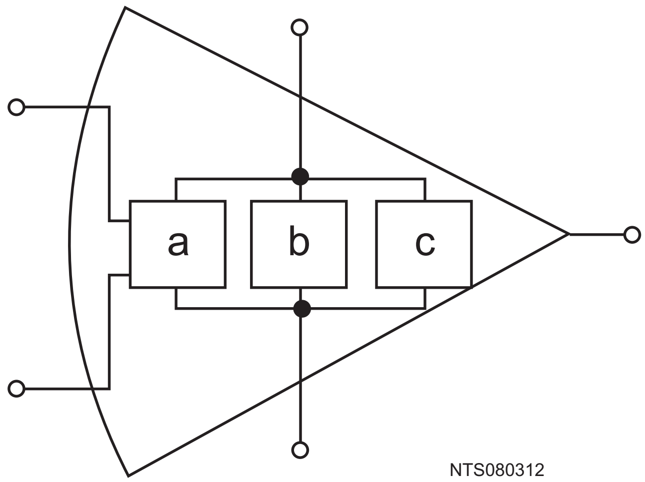


Figure 3-12.—Operational amplifier.

CLOSED-LOOP OPERATION OF AN OPERATIONAL AMPLIFIER

Operational amplifiers can have either a closed-loop operation or an open-loop operation. The operation (closed-loop or open-loop) is determined by whether or not feedback is used. Without feedback the operational amplifier has an open-loop operation. This open-loop operation is practical only when the operational amplifier is used as a comparator (a circuit which compares two input signals or compares an input signal to some fixed level of voltage). As an amplifier, the open-loop operation is not practical because the very high gain of the operational amplifier creates poor stability. (Noise and other unwanted signals are amplified so much in open-loop operation that the operational amplifier is usually not used in this way.) Therefore, most operational amplifiers are used with feedback (closed-loop operation).

Operational amplifiers are used with degenerative (or negative) feedback which reduces the gain of the operational amplifier but greatly increases the stability of the circuit. In the closed-loop configuration, the output signal is applied back to one of the input terminals. This feedback is always degenerative (negative). In other words, the feedback signal always opposes the effects of the original input signal. One result of degenerative feedback is that the inverting and noninverting inputs to the operational amplifier will be kept at the same potential.

Closed-loop circuits can be of the inverting configuration or noninverting configuration. Since the inverting configuration is used more often than the noninverting configuration, the inverting configuration will be shown first.

Inverting Configuration

Figure 3-13 shows an operational amplifier in a closed-loop, inverting configuration. Resistor R2 is used to feed part of the output signal back to the input of the operational amplifier.

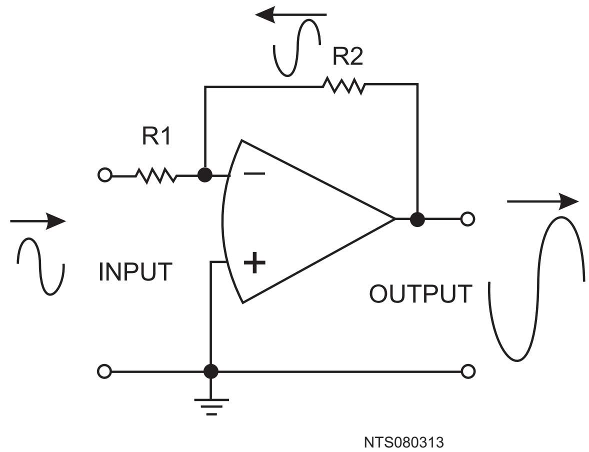


Figure 3-13.—Inverting configuration.

At this point it is important to keep in mind the difference between the entire circuit (or operational circuit) and the operational amplifier. The operational amplifier is represented by the triangle-like symbol while the operational circuit includes the resistors and any other components as well as the operational amplifier. In other words, the input to the circuit is shown in figure 3-13, but the signal at the inverting input of the operational amplifier is determined by the feedback signal as well as by the circuit input signal.

As you can see in figure 3-13, the output signal is 180 degrees out of phase with the input signal. The feedback signal is a portion of the output signal and, therefore, also 180 degrees out of phase with the input signal. Whenever the input signal goes positive, the output signal and the feedback signal go negative. The result of this is that the inverting input to the operational amplifier is always very close to 0 volts with this configuration. In fact, with the noninverting input grounded, the voltage at the inverting input to the operational amplifier is so small compared to other voltages in the circuit that it is considered to be VIRTUAL GROUND. (Remember, in a closed-loop operation the inverting and noninverting inputs are at the same potential.)

Virtual ground is a point in a circuit which is at ground potential (0 volts) but is NOT connected to ground. Figure 3-14, (view A) (view B) and (view C), shows an example of several circuits with points at virtual ground.

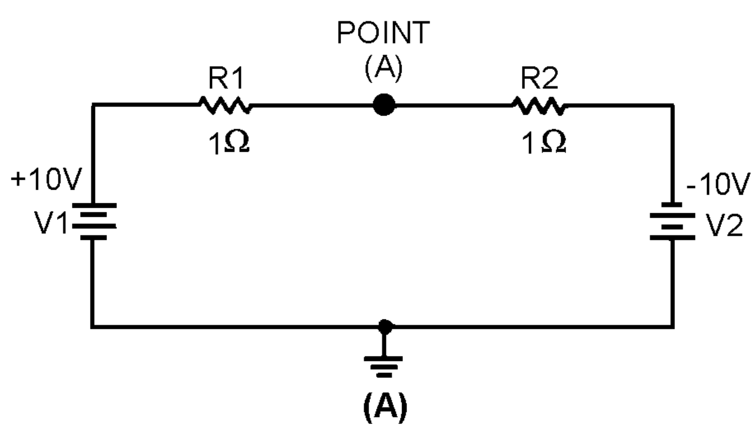


Figure 3-14A.—Virtual ground circuits.

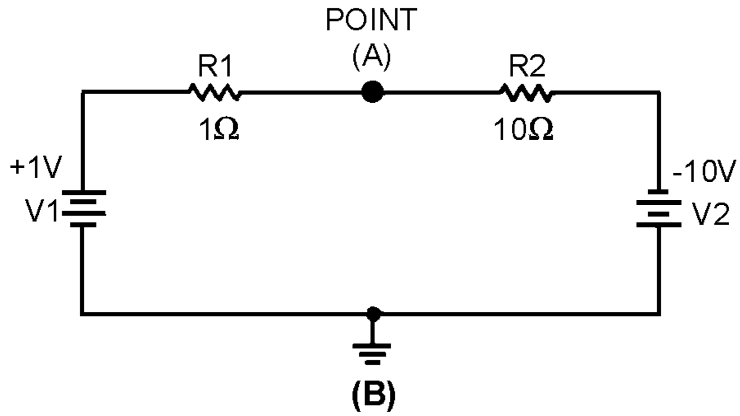


Figure 3-14B.—Virtual ground circuits.

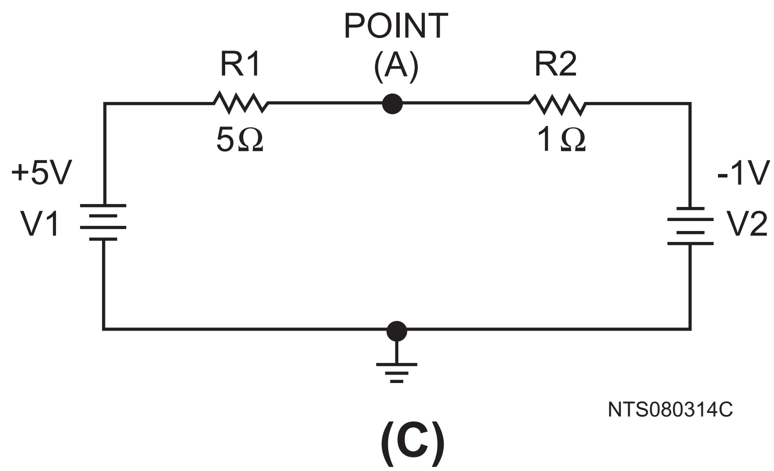


Figure 3-14C.—Virtual ground circuits.

In view (A), V1 (the left-hand battery) supplies +10 volts to the circuit while V2 (the right-hand battery) supplies -10 volts to the circuit. The total difference in potential in the circuit is 20 volts.

The total resistance of the circuit can be calculated:

$$R_T = R_1 + R_2$$

$$R_T = 1\Omega + 1\Omega$$

$$RT = 2\Omega$$

Now that the total resistance is known, the circuit current can be calculated:

$$I_T = \frac{E_T}{R_T}$$

$$I_T = \frac{20V}{2\Omega}$$

$$I_T = 10A$$

The voltage drop across R1 can be computed:

$$E_{R1} = R_1 \text{ x } I_T$$

$$E_{R1} = 1\Omega \text{ x } 10A$$

$$E_{R1} = 10V$$

The voltage at point A would be equal to the voltage of V1 minus the voltage drop of R1.

$$\text{Voltage at point A} = V1 - E_{R1}$$

$$\text{Voltage at point A} = +10V - 10V$$

$$\text{Voltage at point A} = 0V$$

To check this result, compute the voltage drop across R2 and subtract this from the voltage at point A. The result should be the voltage of V2.

$$E_{R2} = R2 \text{ x } I_T$$

$$E_{R2} = 1\Omega \text{ x } 10A$$

$$E_{R2} = 10V$$

$$V2 = (\text{voltage at point A}) - (E_{R2})$$

$$V2 = (0V) - (10V)$$

$$V2 = -10V$$

It is not necessary that the voltage supplies be equal to create a point of virtual ground. In view (B) V1 supplies +1 volt to the circuit while V2 supplies -10 volts. The total difference in potential is 11 volts. The total resistance of this circuit (R1 + R2) is 11 ohms. The total current (I_T) is 1 ampere. The voltage drop across R1 ($E_{R1} = R_1 \times I_T$) is 1 volt. The voltage drop across R2 ($E_{R2} = R2 \times I_T$) is 10 volts. The voltage at point A can be computed:

$$\text{Voltage at point A} = V1 - E_{R1}$$

$$\text{Voltage at point A} = +10V - 10V$$

$$\text{Voltage at point A} = 0V$$

So point A is at virtual ground in this circuit also. To check the results, compute the voltage at V2.

$$V2 = (\text{voltage at point A}) - E_{R1}$$

$$V2 = (0V) - (+10V)$$

$$V2 = -10V$$

You can compute the values for view (C) and prove that point A in that circuit is also at virtual ground.

The whole point is that the inverting input to the operational amplifier shown in figure 3-13 is at virtual ground since it is at 0 volts (for all practical purposes). Because the inverting input is at 0 volts, there will be no current (for all practical purposes) flowing into the operational amplifier from the connection point of R1 and R2.

Given these conditions, the characteristics of this circuit are determined almost entirely by the values of R1 and R2. Figure 3-15 should help show how the values of R1 and R2 determine the circuit characteristics.

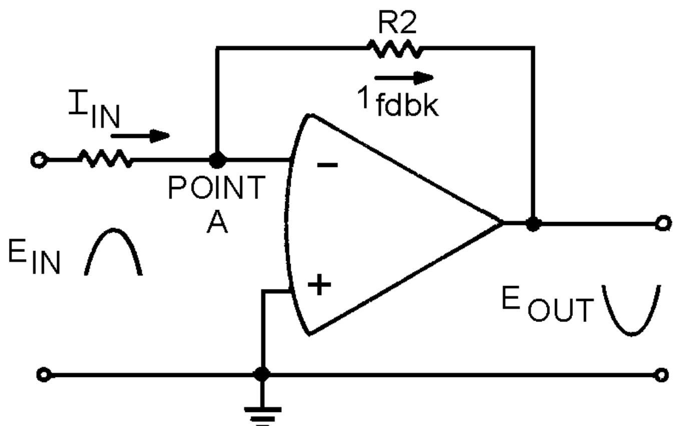


Figure 3-15.—Current flow in the operational circuit.

NOTE: It should be stressed at this point that for purpose of explanation the operational amplifier is a theoretically perfect amplifier. In actual practice we are dealing with less than perfect. In the practical operational amplifier there will be a slight input current with a resultant power loss. This small signal can be measured at the theoretical point of virtual ground. This does not indicate faulty operation.

The input signal causes current to flow through R1. (Only the positive half cycle of the input signal is shown and will be discussed.) Since the voltage at the inverting input of the operational amplifier is at 0 volts, the input current (I_{in}) is computed by:

$$I_{in} = \frac{E_{in}}{R_1}$$

The output signal (which is opposite in phase to the input signal) causes a feedback current (I_{fdbk}) to flow through R2. The left-hand side of R2 is at 0 volts (point A) and the right-hand side is at E_{out}. Therefore, the feedback current is computed by:

$$I_{fdbk} = \frac{-E_{out}}{R_2}$$

(The minus sign indicates that E_{out} is 180 degrees out of phase with E_{in} and should not be confused with output polarity.)

Since no current flows into or out of the inverting input of the operational amplifier, any current reaching point A from R1 must flow out of point A through R2. Therefore, the input current (I_{in}) and the feedback current (I_{fdbk}) must be equal. Now we can develop a mathematical relationship between the input and output sigABCDABCDnals and R1 and R2.

Mathematically:

$$I_{in} = I_{fdbk}$$

By substitution:

$$\frac{E_{in}}{R_1} = \frac{-E_{out}}{R_2}$$

If you multiply both sides of the equation by R1:

$$E_{in} = \frac{-(E_{out})\ (R1)}{R2}$$

If you divide both sides of the equation by E_{out}:

$$\frac{E_{in}}{E_{out}} = -\frac{R_1}{R_2}$$

By inverting both sides of the equation:

$$\frac{E_{in}}{E_{out}} = -\frac{R_2}{R_1}$$

You should recall that the voltage gain of a stage is defined as the output voltage divided by the input voltage:

$$\left(\frac{E_{out}}{E_{in}}\right)$$

Therefore, the voltage gain of the inverting configuration of the operational amplifier is expressed by the equation:

$$-\frac{R_1}{R_2}$$

(As stated earlier, the minus sign indicates that the output signal is 180 degrees out of phase with the input signal.)

Noninverting Configuration

Figure 3-16 shows a noninverting configuration using an operational amplifier. The input signal (E_{in}) is applied directly to the noninverting (+) input of the operational amplifier. Feedback is provided by coupling part of the output signal (E_{out}) back to the inverting (-) input of the operational amplifier. R1 and R2 act as voltage divider that allows only a part of the output signal to be applied as feedback (E_{fdbk}).

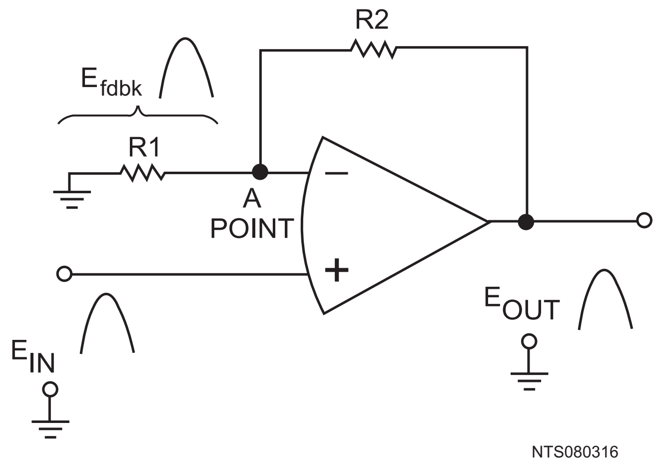


Figure 3-16.—Noninverting configuration.

Notice that the input signal, output signal, and feedback signal are all in phase. (Only the positive alternation of the signal is shown.) It may appear as if the feedback is regenerative (positive) because the feedback and input signals are in phase. The feedback is, in reality, degenerative (negative) because the input signals is applied to the noninverting input and the feedback signal is applied to the inverting input, (Remember, that the operational amplifier will react to the difference between the two inputs.)

Just as in the inverting configuration, the feedback signal is equal to the input signal (for all practical purposes). This time, however, the feedback signal is in phase with the input signal.

Therefore:

$$E_{in} = E_{fdbk}$$

Given this condition, you can calculate the gain of the stage in terms of the resistors (R1 and R2).

The gain of the stage is defined as:

$$\text{Gain} = \frac{E_{out}}{E_{in}}$$

$$\text{Since: } E_{in} = E_{fdbk}$$

$$\text{Then: Gain} = \frac{E_{out}}{E_{fdbk}}$$

The feedback signal (E_{fdbk}) can be shown in terms of the output signal (E_{out}) and the voltage divider (R1 and R2). The voltage divider has the output signal on one end and ground (0 volts) on the other end. The feedback signal is that part of the output signal developed by R1 (at point A). Another way to look at it is that the feedback signal is the amount of output signal left (at point A) after part of the output signal has been dropped by R2. In either case, the feedback signal (E_{fdbk}) is the ratio of R1 to the entire voltage divider (R1 + R2) multiplied by the output signal (E_{out}).

Mathematically, the relationship of the output signal, feedback signal, and voltage divider is:

$$E_{fdbk} = \frac{R_1}{R_1 + R_2}(E_{out})$$

If you divide both sides of the equation by E_{out}:

$$\frac{E_{fdbk}}{E_{out}} = \frac{R_1}{R_1 + R_2}$$

By inverting both sides of the equation:

$$\frac{E_{fdbk}}{E_{out}} = \frac{R_1 + R_2}{R_1}$$

Separating the right-hand side:

$$\frac{E_{fdbk}}{E_{out}} = \frac{R_1}{R_1} + \frac{R_2}{R_1}$$

Remember:

$$\text{Gain} = \frac{E_{out}}{E_{fdbk}}$$

Therefore, by substitution:

$$\text{Gain} = \frac{R_2}{R_1} + 1$$

You can now see that the gain of the noninverting configuration is determined by the resistors. The formula is different from the one used for the inverting configuration, but the gain is still determined by the values of R1 and R2.

BANDWIDTH LIMITATIONS

As with most amplifiers, the gain of an operational amplifier varies with frequency. The specification sheets for operational amplifiers will usually state the open-loop (no feedback) gain for d.c. (or 0 hertz). At higher frequencies, the gain is much lower. In fact, for an operational amplifier, the gain decreases quite rapidly as frequency increases.

Figure 3-17 shows the open-loop (no feedback) frequency-response curve for a typical operational amplifier. As you should remember, bandwidth is measured to the half-power points of a frequency-response curve. The frequency-response curve shows that the bandwidth is only 10 hertz with this configuration. The UNITY GAIN POINT, where the signal out will have the same amplitude as the signal in (the point at which the gain of the amplifier is 1), is 1 megahertz for the amplifier. As you can see, the frequency response of this amplifier drops off quite rapidly.

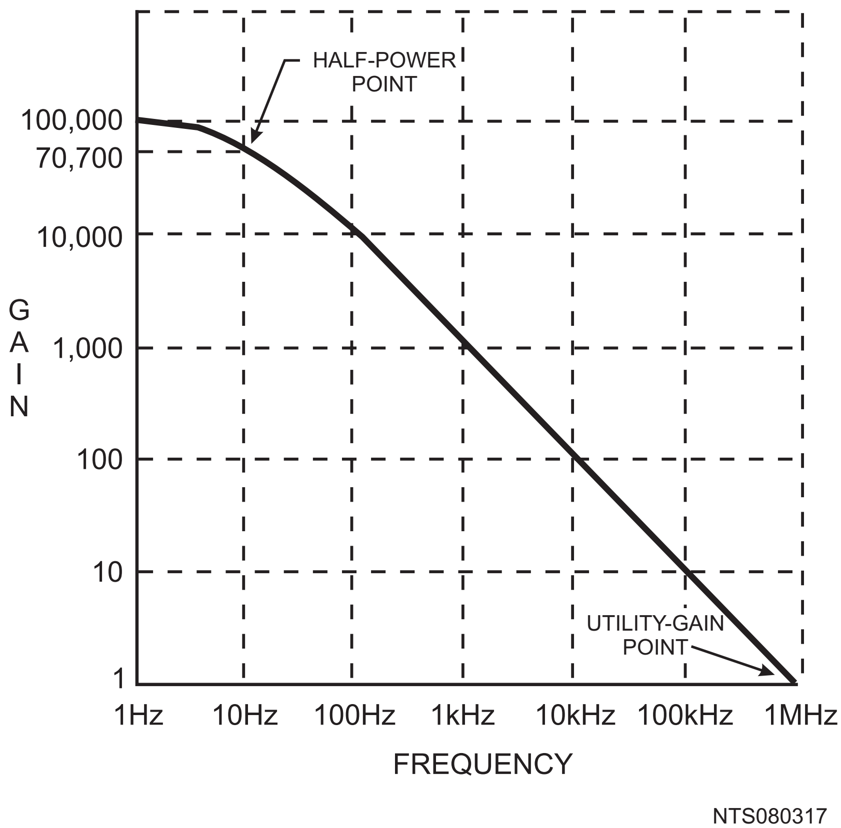


Figure 3-17.—Open-loop frequency-response curve.

Figure 3-17 is the open-loop frequency-response curve. You have been told that most operational amplifiers are used in a closed-loop configuration. When you look at the frequency-response curve for a closed-loop configuration, one of the most interesting and important aspects of the operational amplifier becomes apparent: <u>The use of degenerative feedback increases the bandwidth of an operational amplifier circuit</u>.

This phenomenon is another example of the difference between the operational amplifier itself and the operational-amplifier circuit (which includes the components in addition to the operational amplifier). You should also be able to see that the external resistors not only affect the gain of the circuit, but the bandwidth as well.

You might wonder exactly how the gain and bandwidth of a closed-loop, operational-amplifier circuit are related. Figure 3-18 should help to show you the relationship. The frequency-response curve shown in figure 3-18 is for a circuit in which degenerative feedback has been used to decrease the circuit gain to 100 (from 100,000 for the operational amplifier). Notice that the half-power point of this curve is just slightly above 10 kilohertz.

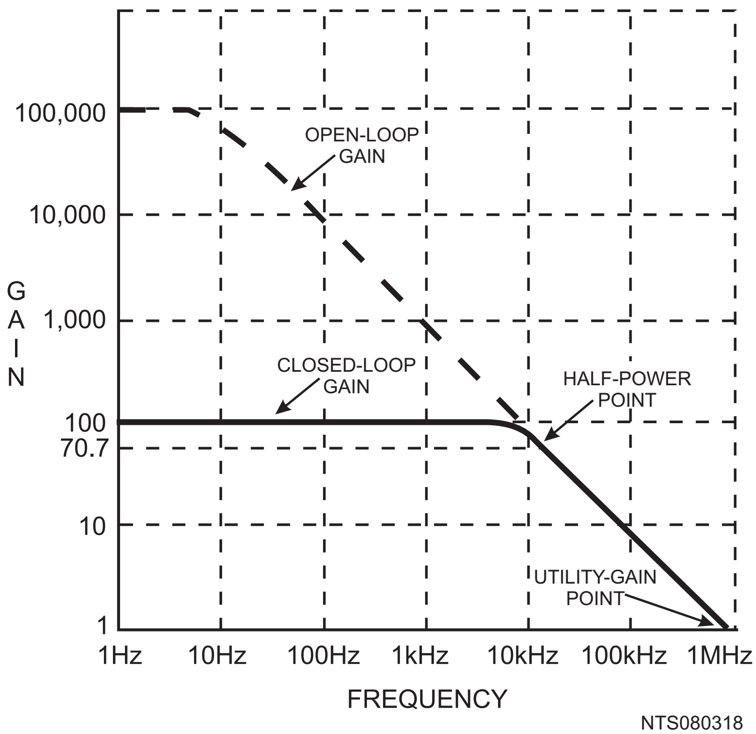


Figure 3-18.—Closed-loop frequency-response curve for gain of 100.

Now look at figure 3-19. In this case, more feedback has been used to decrease the gain of the circuit to 10. Now the bandwidth of the circuit is extended to about 100 kilohertz.

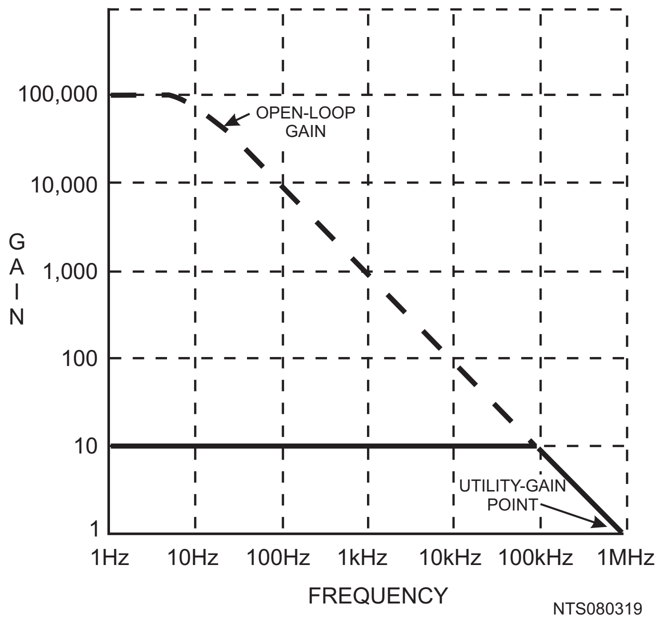


Figure 3-19.—Closed-loop frequency-response curve for gain of 10.

The relationship between circuit gain and bandwidth in an operational-amplifier circuit can be expressed by the GAIN-BANDWIDTH PRODUCT (GAIN × BANDWIDTH = UNITY GAIN POINT). In other words, for operational-amplifier circuits, the gain times the bandwidth for one configuration of an operational amplifier will equal the gain times the bandwidth for any other configuration of the same operational amplifier. In other words, when the gain of an operational-amplifier circuit is changed (by changing the value of feedback or input resistors), the bandwidth also changes. But the gain times the bandwidth of the first configuration will equal the gain times the bandwidth of the second configuration. The following example should help you to understand this concept.

The frequency-response curves shown in figures 3-17, 3-18, and 3-19 have a gain-bandwidth product of 1,000,000. In figure 3-17, the gain is 100,000 and the bandwidth is 10 hertz. The gain-bandwidth product is 100,000 times 10 (Hz), or 1,000,000. In figure 3-18, the gain has been reduced to 100 and the bandwidth increases to 10 kilohertz. The gain-bandwidth product is 100 times 10,000 (Hz) which is also equal to 1,000,000. In figure 3-19 the gain has been reduced to 10 and the bandwidth is 100 kilohertz. The gain-bandwidth product is 10 times 100,000 (Hz), which is 1,000,000. If the gain were reduced to 1, the bandwidth would be 1 megahertz (which is shown on the frequency-response curve as the unity-gain point) and the gain-bandwidth product would still be 1,000,000.

Q-19. What does the term "closed-loop" mean in the closed-loop configuration of an operational amplifier?

In answering Q20, Q21, and Q23, select the correct response from the choices given in the parentheses.

Q-20. In a closed-loop configuration the output signal is determined by (the input signal, the feedback signal, both).

Q-21. In the inverting configuration, the input signal is applied to the (a) (inverting, noninverting) input and the feedback signal is applied to the (b) (inverting, noninverting) input.

Q-22. In the inverting configuration, what is the voltage (for all practical purposes) at the inverting input to the operational amplifier if the input signal is a 1-volt, peak-to-peak sine wave?

Q-23. In the inverting configuration when the noninverting input is grounded, the inverting input is at (signal, virtual) ground.

Q-24. In a circuit such as that shown in figure 3-15, if R1 has a value of 100 ohms and R2 has a value of 1 kilohm and the input signal is at a value of + 5 millivolts, what is the value of the output signal?

Q-25. If the unity-gain point of the operational amplifier used in question 24 is 500 kilohertz, what is the bandwidth of the circuit?

Q-26. In a circuit such as that shown in figure 3-16, if R1 has a value of 50 ohms and R2 has a value of 250 ohms and the input signal has a value of +10 millivolts, what is the value of the output signal?

Q-27. If the open-loop gain of the operational amplifier used in question 26 is 200,000 and the open-loop bandwidth is 30 hertz, what is the closed loop bandwidth of the circuit?

APPLICATIONS OF OPERATIONAL AMPLIFIERS

Operational amplifiers are used in so many different ways that it is not possible to describe all of the applications. Entire books have been written on the subject of operational amplifiers. Some books are devoted entirely to the applications of operational amplifiers and are not concerned with the theory of operation or other circuits at all. This module, as introductory material on operational amplifiers, will show you only two common applications of the operational amplifier: the summing amplifier and the difference amplifier. For ease of explanation the circuits shown for these applications will be explained with d.c. inputs and outputs, but the circuit will work as well with a.c. signals.

Summing Amplifier (Adder)

Figure 3-20 is the schematic of a two-input adder which uses an operational amplifier. The output level is determined by adding the input signals together (although the output signal will be of opposite polarity compared to the sum of the input signals).

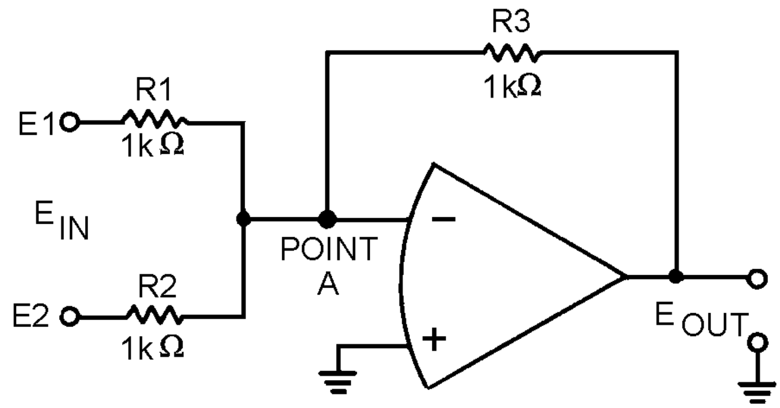


Figure 3-20.—Two-input adder.

If the signal on input number one (E1) is +3 volts and the signal on input number two (E2) is +4 volts, the output signal (E_{out}) should be −7 volts [(+3 V) + (+4 V) = +7 V and change the polarity to get −7 V].

With +3 volts at E1 and 0 volts at point A (which is at virtual ground), the current through R1 must be 3 milliamperes.

Mathematically:

$$I_{R1} = \frac{E_1}{R_1}$$

$$I_{R1} = \frac{+3V}{1k\Omega}$$

$$I_{R1} = +3mA$$

(The + sign indicates a current flow from right to left.)

By the same sort of calculation, with +4 volts at E2 and 0 volts at point A the current through R2 must be 4 milliamps.

This means that a total of 7 milliamps is flowing from point A through R1 and R2. If 7 milliamps is flowing from point A, then 7 milliamps must be flowing into point A. The 7 milliamps flowing into point A flows through R3 causing 7 volts to be developed across R3. With point A at 0 volts and 7 volts developed across R3, the voltage potential at E_{out} must be a −7 volts. Figure 3-21 shows these voltages and currents.

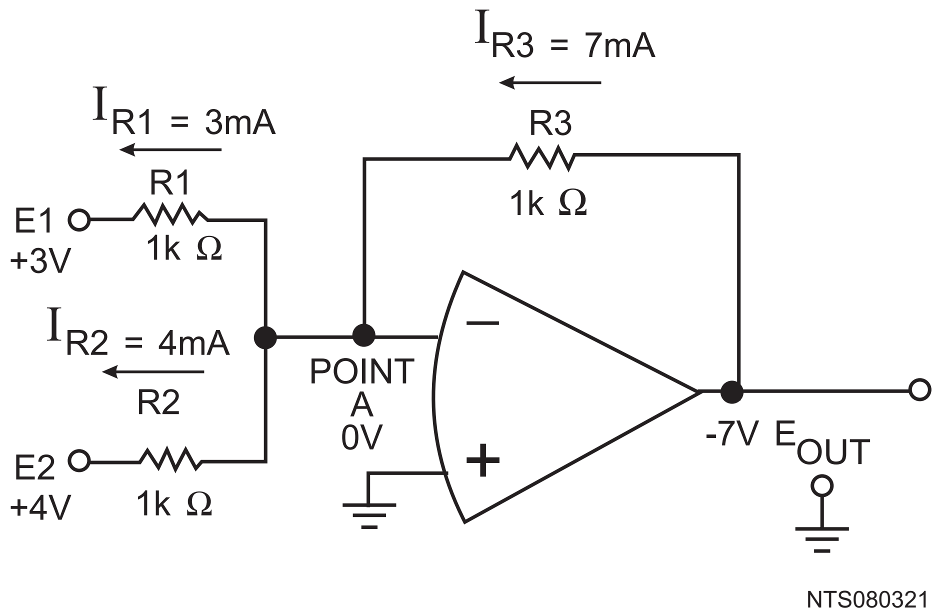
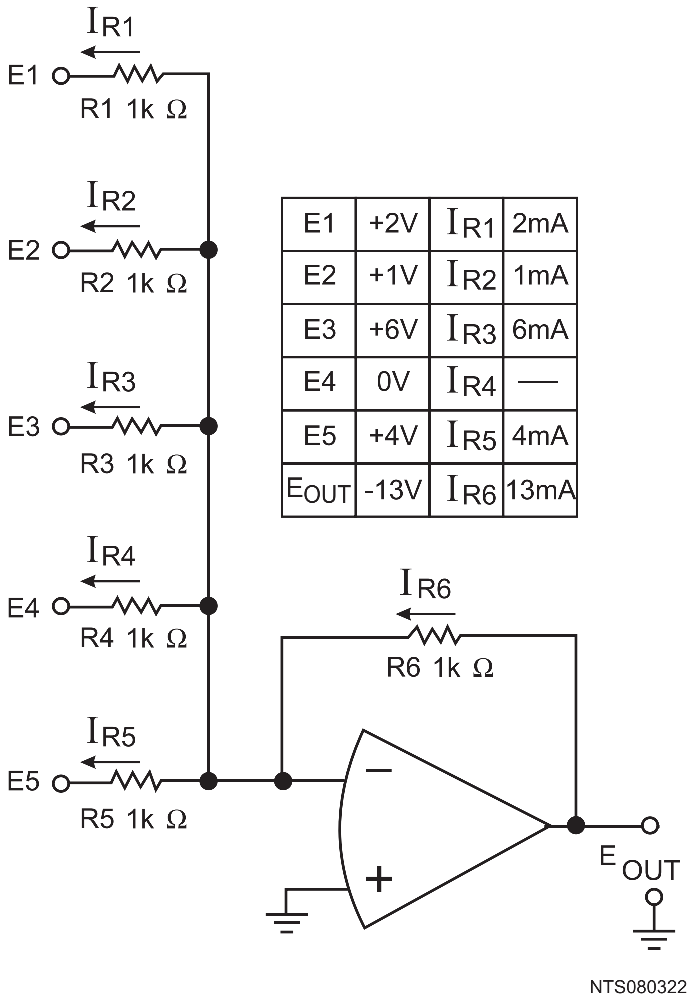


Figure 3-21.—Current and voltage in a two-input adder.

An adder circuit is not restricted to two inputs. By adding resistors in parallel to the input terminals, any number of inputs can be used. The adder circuit will always produce an output that is equal to the sum of the input signals but opposite in polarity. Figure 3-22 shows a five-input adder circuit with voltages and currents indicated.



E1	+2V	I_{R1}	2mA
E2	+1V	I_{R2}	1mA
E3	+6V	I_{R3}	6mA
E4	0V	I_{R4}	—
E5	+4V	I_{R5}	4mA
E_{OUT}	-13V	I_{R6}	13mA

Figure 3-22.—Five-input adder.

The previous circuits have been adders, but there are other types of summing amplifiers. A summing amplifier can be designed to amplify the results of adding the input signals. This type of circuit actually multiplies the sum of the inputs by the gain of the circuit.

Mathematically (for a three-input circuit):

$$E_{out} = \text{gain}\ (E1 + E2 + E3)$$

If the circuit gain is -10:

$$E_{out} = -10\ (E1 + E2 + E3)$$

The gain of the circuit is determined by the ratio between the feedback resistor and the input resistors. To change figure 3-20 to a summing amplifier with a gain of -10, you would replace the feedback resistor (R3) with a 10-kilohm resistor. This new circuit is shown in figure 3-23.

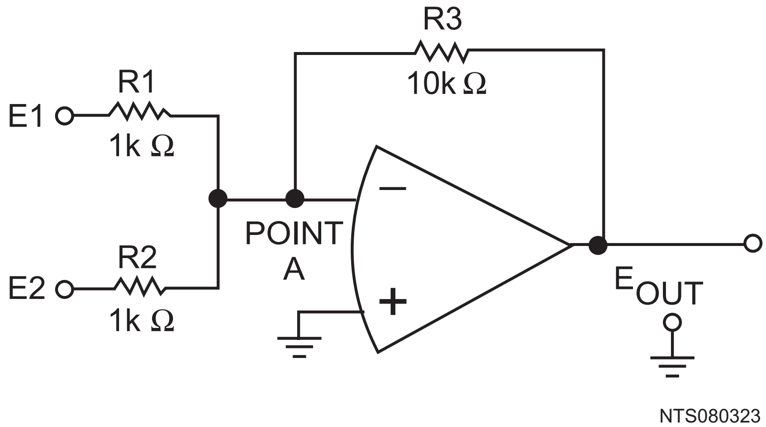


Figure 3-23.—Summing amplifier.

If this ciABCDrcuit is designed correctly and the input voltages (E1 and E2) are +2 volts and +3 volts, respectively, the output voltage (E_{out}) should be:

$$E_{out} = \text{gain}\ (E1 + E2)$$

$$E_{out} = -10\ [(+2V) + (+3V)]$$

$$E_{out} = -10\ (+5)$$

$$E_{out} = -50V$$

To see if this output (–50 V) is what the circuit will produce with the inputs given above, start by calculating the currents through the input resistors, R1 and R2 (remember that point A is at virtual ground):

$$I_{R1} = \frac{E_1}{R_1}$$

$$I_{R1} = \frac{2V}{1k\Omega}$$

$$I_{R1} = 2mA$$

$$I_{R2} = \frac{E_2}{R_2}$$

$$I_{R2} = \frac{3V}{1k\Omega}$$

$$I_{R2} = 3mA$$

Next, calculate the current through the feedback resistor (R3):

$$I_{R3} = -(I_{R1} + I_{R2})$$

$$I_{R3} = -(2mA + 3\ mA)$$

$$I_{R3} = -5mA$$

(The minus sign indicates current flow from left to right.)

Finally, calculate the voltage dropped across R3 (which must equal the output voltage):

$$E_{OUT} = (I_{R3} \text{ x } R3)$$

$$E_{OUT} = (-5\text{ mA x } 10\text{ k}\Omega)$$

$$E_{OUT} = -50V$$

As you can see, this circuit performs the function of adding the inputs together and multiplying the result by the gain of the circuit.

One final type of summing amplifier is the SCALING AMPLIFIER. This circuit multiplies each input by a factor (the factor is determined by circuit design) and then adds these values together. The factor that is used to multiply each input is determined by the ratio of the feedback resistor to the input resistor. For example, you could design a circuit that would produce the following output from three inputs (E1, E2, E3):

$$-[(2 \text{ x } E1) + (4 \text{ x } E2) + (3 \text{ x } E3)]$$

Using input resistors R1 for input number one (E1), R2 for input number two (E2), R3 for input number three (E3), and R4 for the feedback resistor, you could calculate the values for the resistors:

$$2 = \frac{R_4}{R_1}$$

$$4 = \frac{R_4}{R_2}$$

$$3 = \frac{R_4}{R_3}$$

Any resistors that will provide the ratios shown above could be used. If the feedback resistor (R4) is a 12-kilohm resistor, the values of the other resistors would be:

$$2 = \frac{12\ k\Omega}{R1}$$

$$2\ (R1) = 12\ k\Omega$$

$$R1 = 6\ k\Omega$$

$$4 = \frac{12\ k\Omega}{R2}$$

$$4(R2) = 12\ k\Omega$$

$$R2 = 3\ k\Omega$$

$$3 = \frac{12\ k\Omega}{R3}$$

$$3(R3) = 12\ k\Omega$$

$$R3 = 4\ k\Omega$$

Figure 3-24 is the schematic diagram of a scaling amplifier with the values calculated above.

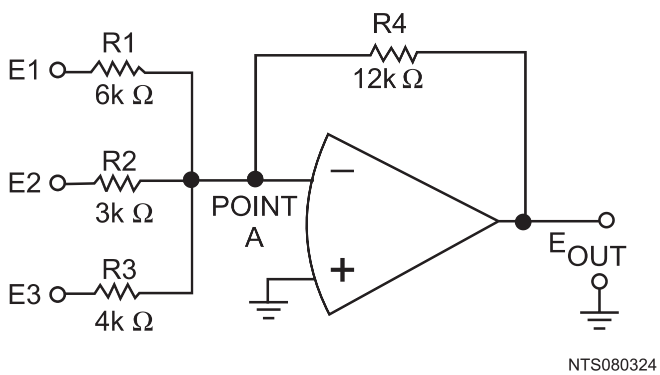


Figure 3-24.—Scaling amplifier.

To see if the circuit will produce the desired output, calculate the currents and voltages as done for the previous circuits.

With:

$$E1 = +12V$$

$$E2 = +3V$$

$$E3 = +8V$$

the output should be:

$$E_{out} = -[(2 \times E1) + (4 \times E2) + (3 \times E3)]$$

$$E_{out} = -[(2 \times 12V) + (4 \times +3V) + (3 \times +8V)]$$

$$E_{out} = -[(+24V) + (+12V) + (+24V)]$$

$$E_{out} = -60V$$

Calculate the current for each input:

$$I_{R1} = \frac{E_1}{R_1}$$

$$I_{R1} = \frac{+12V}{6k\Omega}$$

$$I_{R1} = +2mA$$

$$I_{R2} = \frac{E_2}{R_2}$$

$$I_{R2} = \frac{+3V}{3k\Omega}$$

$$I_{R2} = +1mA$$

$$I_{R3} = \frac{E_3}{R_3}$$

$$I_{R3} = \frac{+8V}{4k\Omega}$$

$$I_{R3} = +2mA$$

$$I_{R4} = -(I_{R1} + I_{R2} + I_{R3})$$

$$I_{R4} = -(2mA + 1mA + 2mA)$$

$$I_{R4} = -5mA$$

Calculate the output voltage:

$$E_{out} = E_{R4}$$

$$E_{out} = I_{R4} \times R_4$$

$$E_{out} = (-5mA \times 12k\Omega)$$

$$E_{out} = -60V$$

You have now seen how an operational amplifier can be used in a circuit as an adder, a summing amplifier, and a scaling amplifier.

Difference Amplifier (Subtractor)

A difference amplifier will produce an output based on the difference between the input signals. The subtractor circuit shown in figure 3-25 will produce the following output:

$$E_{OUT} = 5\ (E2 - E1)$$

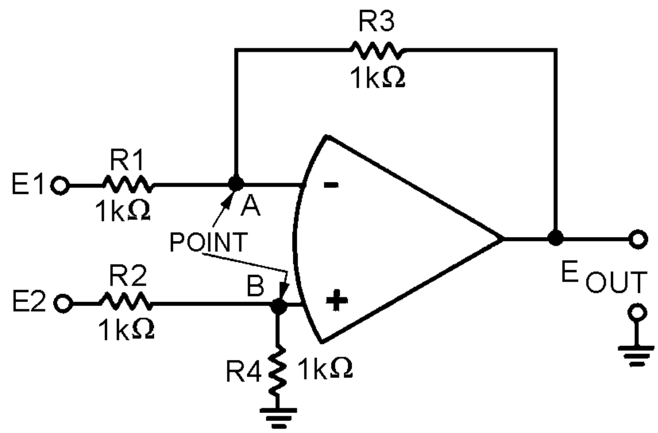


Figure 3-25.—Subtractor circuit.

Normally, difference amplifier circuits have the ratio of the inverting input resistor to the feedback resistor equal to the ratio of the noninverting input resistors. In other words, for figure 3-25:

$$\frac{R1}{R3} = \frac{R2}{R4}$$

and, by inverting both sides:

$$\frac{R3}{R1} = \frac{R4}{R2}$$

For ease of explanation, in the circuit shown in figure 3-25 all the resistors have a value of 1 kilohm, but any value could be used as long as the above ratio is true. For a subtractor circuit, the values of R1 and R3 must also be equal, and therefore, the values of R2 and R4 must be equal. It is NOT necessary that the value of R1 equal the value of R2.

Using figure 3-25, assume that the input signals are:

$$E1 = +3V$$
$$E2 = +12V$$

The output signal should be:

$$E_{OUT} = E2 - E1$$

$$E_{OUT} = (+12V) - (+3V)$$

$$E_{OUT} = +9V$$

To check this output, first compute the value of R2 plus R4:

$$R_2 + R_4 = 1k\Omega + 1k\Omega$$

$$R_2 + R_4 = 2k\Omega$$

With this value, compute the current through R2 (I_{R2}):

$$I_{R2} = \frac{E2}{R_2 + R_4}$$

$$I_{R2} = \frac{+12V}{2k\Omega}$$

$$I_{R2} = +6mA$$

(indicating current flow from left to right)

Next, compute the voltage drop across R2 (E_{R2}):

$$E_{R2} = R_2 \times I_{R2}$$

$$E_{R2} = 1k\Omega \times (+6mA)$$

$$E_{R2} = +6V$$

Then compute the voltage at point B:

Then compute the voltage at point B:

Voltage at point B = $E2 - E_{R2}$
Voltage at point B = (+12V) - (+6V)
Voltage at point B = +6V

Since point B and point A will be at the same potential in an operational amplifier:

Voltage at point A = +6V

Now compute the voltage developed by R1 (E_{R1}):

E_{R1} = (voltage at point A) - (E1)

$$E_{R1} = (+6V) - (+3V)$$

$$E_{R1} = +3V$$

Compute the current through R1 (I_{R1}):

$$I_{R1} = \frac{E_{R1}}{R_1}$$

$$I_{R1} = \frac{+3V}{1k\Omega}$$

$$I_{R1} = +3mA$$

Since: $I_{R1} = I_{R3}$
Then: $I_{R3} = +3mA$

Compute the voltage developed by R3 (E_{R3}):

$$E_{R3} = (R_3) \text{ x } (I_{R3})$$
$$E_{R3} = (1k\,\Omega) \text{ x } (+3mA)$$
$$E_{R3} = +3V$$

Add this to the voltage at point A to compute the output voltage (E_{out}):

$$E_{OUT} = (E_{R3}) + (\text{voltage at point A})$$

$$E_{OUT} = (+3V) \text{ x } (+6V)$$

$$E_{OUT} = +9V$$

As you can see, the circuit shown in figure 3-25 functions as a subtractor. But just as an adder is only one kind of summing amplifier, a subtractor is only one kind of difference amplifier. A difference amplifier can amplify the difference between two signals. For example, with two inputs (E1 and E2) and a gain of five, a difference amplifier will produce an output signal which is:

$$E_{OUT} = 5\ (E2 - E1)$$

The difference amplifier that will produce that output is shown in figure 3-26. Notice that this circuit is the same as the subtractor shown in figure 3-25 except for the values of R3 and R4. The gain of this difference amplifier is:

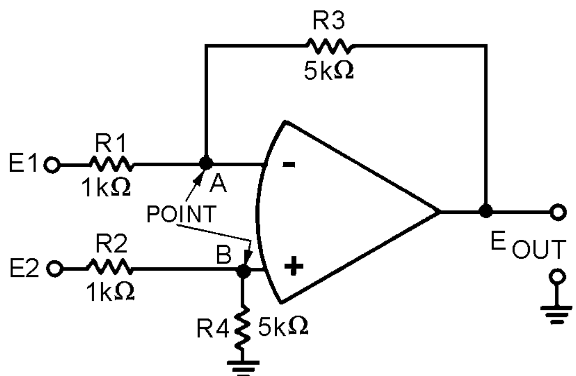


Figure 3-26.—Difference amplifier.

$$\text{Gain} = \frac{R3}{R1}$$

$$\text{Gain} = \frac{5\ k\Omega}{1\ k\Omega}$$

$$\text{Gain} = 5$$

Then, for a difference amplifer:

$$\text{Gain} = \frac{R3}{R1} = \frac{R4}{R2}$$

With the same inputs that were used for the subtractor, (E1 = + 3 V; E2 = + 12 V) the output of the difference amplifier should be five times the output of the subtractor (E_{out} = + 45 V).

Following the same steps used for the subtractor:

First compute the value of R2 plus R4:

$$R2 + R4 = 1\ k\Omega + 5\ k\Omega$$
$$R2 + R4 = 4\ k\Omega$$

With this value, compute the current through R2 (I_{R2}):

$$I_{R2} = \frac{E2}{R_2 + R_4}$$

$$I_{R2} = \frac{+12V}{6k\Omega}$$

$$I_{R2} = +2mA$$

Next, compute the voltage drop across R2 (E_{R2}):

$$E_{R2} = (R_2) \times (I_{R2})$$

$$E_{R2} = (1k\Omega) \times (+2mA)$$

$$E_{R2} = +2V$$

Then, compute the voltage at point B:

$$\text{Voltage at point B} = E2 - E_{R2}$$
$$\text{Voltage at point B} = (+12V) - (+2V)$$
$$\text{Voltage at point B} = +10V$$

Since point A and point B will be at the same potential in an operational amplifier:

$$\text{Voltage at point A} = +10V$$

Now compute the voltage developed by R1 (E_{R1}):

$$E_{R1} = (\text{voltage at point A}) - (E1)$$

$$E_{R1} = (+10V) - (+3V)$$

$$E_{R1} = +7V$$

Compute the current through R1 (I_{R1}):

$$I_{R1} = \frac{E_{R1}}{R_1}$$

$$I_{R1} = \frac{+7V}{k\Omega}$$

$$I_{R1} = +7mA$$

$$\text{Since: } I_{R1} = I_{R3}$$

$$\text{Then: } I_{R3} = +7mA$$

Compute the voltage developed by R3 (E_{R3}):

$$E_{R3} = R3 \times I_{R3}$$
$$E_{R3} = (5k\Omega) \times (+7mA)$$
$$E_{R3} = +35V$$

Add this voltage to the voltage at point A to compute the output voltage (E_{out}):

$$E_{OUT} = (E_{R3}) + (\text{voltage at point A})$$

$$E_{OUT} = (+35V) + (+10V)$$

$$E_{OUT} = +45V$$

This was the output desired, so the circuit works as a difference amplifier.

Q-28. What is the difference between a summing amplifier and an adder circuit?

Q-29. Can a summing amplifier have more than two inputs?

Q-30. What is a scaling amplifier?

Refer to figure 3-27 in answering Q31 through Q33.

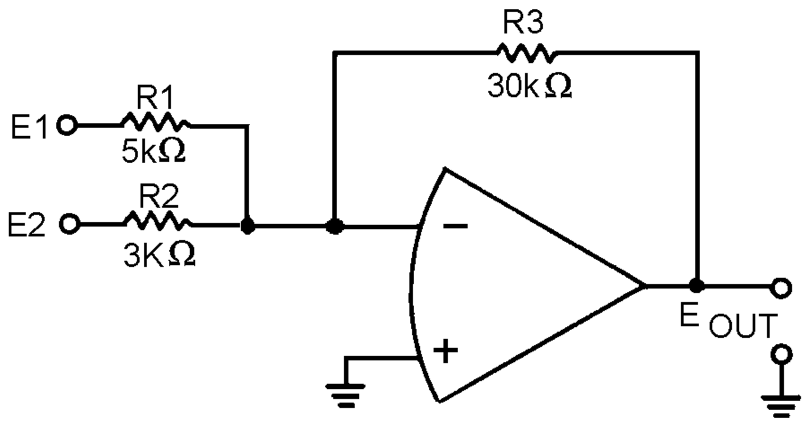


Figure 3-27.—Circuit for Q31 through Q33.

Q-31. What type of circuit is figure 3-27?

Q-32. If: E1 = +2V, and: E2 = +6V, then E_{out} = ?

Q-33. What is the difference in potential between the inverting (-) and noninverting (+) inputs to the operational amplifier when: E1 = +6V, and E2 = +2V

Q-34. What is the difference between a subtractor and a difference amplifier?

Q-35. Can a difference amplifier have more than two inputs?

Refer to figure 3-28 in answering Q36 through Q38.

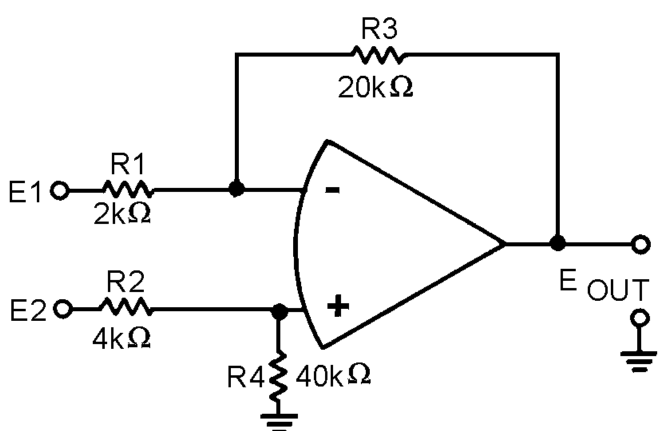


Figure 3-28.—Circuit for Q36 through Q38.

Q-36. What type of circuit is figure 3-28?

Q-37. If: E1 = +5V, and: E2 = +11V, then E_{out} = ?

Q-38. What is the difference in potential between the inverting (−) and noninverting (+) inputs to the operational amplifier when: E1 = +2V, and E2 = +4V

MAGNETIC AMPLIFIERS

You have now been shown various ways that electron tubes (*NEETS*, *Module 6*) and transistors (*NEETS*, *Module 7)* can be used to amplify signals. You have also been shown the way in which this is done. There is another type of amplifier in use—the MAGNETIC AMPLIFIER, sometimes called the MAG AMP.

The magnetic amplifier has certain advantages over other types of amplifiers. These include (1) high efficiency (up to 90 percent); (2) reliability (long life, freedom from maintenance, reduction of spare parts inventory); (3) ruggedness (shock and vibration resistance, high overload capability, freedom from effects of moisture); and (4) no warm-up time. The magnetic amplifier has no moving parts and can be hermetically sealed within a case similar to the conventional dry-type transformer.

However, the magnetic amplifier has a few disadvantages. For example, it cannot handle low-level signals; it is not useful at high frequencies; it has a time delay associated with the magnetic effects; and the output waveform is not an exact reproduction of the input waveform (poor fidelity).

The magnetic amplifier is important, however, to many phases of naval engineering because it provides a rugged, trouble-free device that has many applications aboard ship and in aircraft. These applications include throttle controls on the main engines of ships; speed, frequency, voltage, current, and temperature controls on auxiliary equipment; and fire control, servomechanisms, and stabilizers for guns, radar, and sonar equipment.

As stated earlier, the magnetic amplifier does not amplify magnetism, but uses electromagnetism to amplify a signal. It is a power amplifier with a very limited frequency response. Technically, it falls into the classification of an audio amplifier; but, since the frequency response is normally limited to 100 hertz and below, the magnetic amplifier is more correctly called a low-frequency amplifier.

The basic principle of a magnetic amplifier is very simple. (Remember, all amplifiers are current-control devices.) A magnetic amplifier uses a changing inductance to control the power delivered to a load.

BASIC OPERATION OF A MAGNETIC AMPLIFIER

Figure 3-29 shows a simple circuit with a variable inductor in series with a resistor (representing a load). The voltage source is 100 volts at 60 hertz.

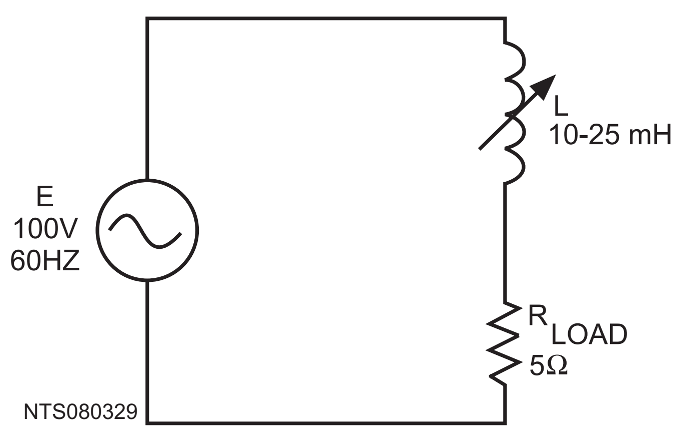


Figure 3-29.—Variable inductor in series with a load.

What happens when the inductance <u>decreases</u>? The end result is that the power in the load (true power) <u>increases</u>. Why? Look at the following formulas and see how each is affected by a decrease in inductance.

$$X_L = 2\pi f L \quad \text{(inductive reactance in the circuit)}$$

$$Z = \sqrt{X_L^2 + R^2} \quad \text{(impedance in the circuit)}$$

$$I = \frac{E}{Z} \quad \text{(current in the circuit)}$$

$$\text{True power} = I^2R \quad \text{(true power or power in the load)}$$

(True power is covered in *NEETS, Module 2—Introduction to Alternating Current and Transformers*.)

As inductance (L) decreases, X_L decreases. As X_L decreases, Z decreases. As Z decreases, I increases. Finally, as I increases, true power increases.

This general conclusion can be confirmed by using some actual values of inductance in the formulas along with other values from figure 3-29.

If the value of inductance is 23 millihenries, the formulas yield the following values:

$$X_L = 2\pi f L$$

$$X_L = (2)\,(3.14)\,(60\text{Hz})\,(23\text{mH})$$

$$X_L = 8.67\Omega \text{ (rounded off)}$$

$$Z = \sqrt{X_L^2 + R^2}$$

$$Z = \sqrt{(8.67\Omega)^2 + (5\Omega)^2}$$

$$Z = \sqrt{100.1689\Omega}$$

$$Z = 10\Omega \text{ (rounded off)}$$

$$I = \frac{E}{Z}$$

$$I = \frac{100\text{v}}{10\Omega}$$

$$I = 10\text{A}$$

$$\text{True power} = I^2R$$

$$\text{True power} = (10\text{A})^2 + (5\Omega)$$

$$\text{True power} = 500 \text{ watts}$$

Now, if the value of inductance is decreased to 11.7 millihenries, the formulas yield the following values:

$$X_L = 2\pi f L$$

$$X_L = (2)(3.14)(60Hz)(117mH)$$

$$X_L = 4.41\,\Omega \text{ (rounded off)}$$

$$Z = \sqrt{X_L^2 + R^2}$$

$$Z = \sqrt{(4.41\Omega)^2 + (5\Omega)^2}$$

$$Z = \sqrt{144.4481\Omega}$$

$$Z = 6.67\,\Omega \text{ (rounded off)}$$

$$I = \frac{E}{Z}$$

$$I = \frac{100v}{6.67\Omega}$$

$$I = 15A \text{ (rounded off)}$$

$$\text{True power} = I^2R$$
$$\text{True power} = (15A)^2 + (5\Omega)$$
$$\text{True power} = 1125 \text{ watts}$$

So a decrease in inductance of 11.3 millihenries (23 mH—11.7 mH) causes an increase in power to the load (true power) of 625 watts (1125 W—500 W). If it took 1 watt of power to change the inductance by 11.3 millihenries (by some electrical or mechanical means), figure 3-29 would represent a power amplifier with a gain of 625.

Q-39. What is the frequency classification of a magnetic amplifier?

Q-40. What is the basic principle of a magnetic amplifier?

Q-41. If inductance increases in a series LR circuit, what happens to true power?

METHODS OF CHANGING INDUCTANCE

Since changing the inductance of a coil enables the control of power to a load, what methods are available to change the inductance? Before answering that question, you should recall a few things about *magnetism and inductors from NEETS, Module 1—Introduction to Matter, Energy, and Direct Current,* chapter 1—*Matter, Energy, and Electricity*; and *Module 2 —Introduction to Alternating Current and Transformers*, chapter 2—*Inductance*.

Permeability was defined as the measure of the ability of a material to act as a path for additional magnetic lines of force. Soft iron was presented as having high permeability compared with air. In fact, the permeability of unmagnetized iron is 5000 while air has a permeability of 1. A nonmagnetized piece of iron has high permeability because the tiny molecular magnets (Weber's Theory) or the directions of electron spin (Domain Theory) are able to be aligned by a magnetic field. As they align, they act as a path for the magnetic lines of force.

Earlier *NEETS* modules state that the inductance of a coil increases directly as the permeability of the core material increases. If a coil is wound around an iron core, the permeability of the core is 5000. Now, if the iron is pulled part way out of the coil of wire, the core is part iron and part air. The permeability of the core decreases. As the permeability of the core decreases, the inductance of the coil decreases. This increases the power delivered to the load (true power). This relationship is shown in figure 3-30.

The system shown in figure 3-30 is not too practical. Even if a motor were used in place of the hand that is shown, the resulting amplifier would be large, expensive, and not easily controlled. If the permeability of a core could be changed by electrical means rather than mechanical, a more practical system would result.

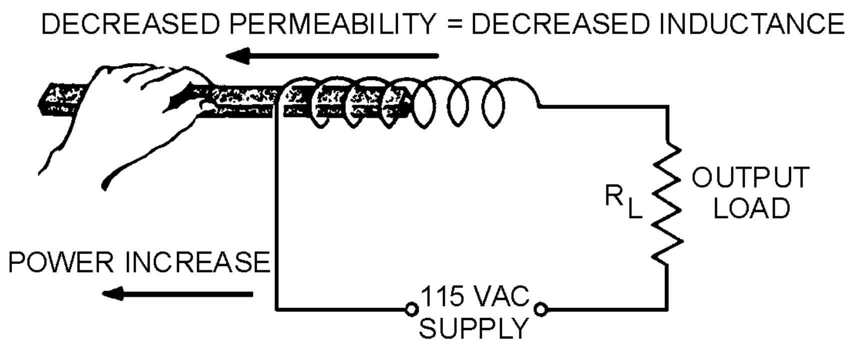


Figure 3-30.—Varying coil inductance with a movable coil.

High permeability depends on there being many molecular magnets (or electron spin directions) that can be aligned to provide a path for magnetic lines of force. If almost all of these available paths are already being used, the material is magnetized and there are no more paths for additional lines of force. The "flux density" (number of lines of force passing through a given area) is as high as it can be. This means that the permeability of the material has decreased. When this condition is reached, the core is said to be SATURATED because it is saturated (filled) with all the magnetic lines of force it can pass. At this point, the core has almost the same value of permeability as air (1) instead of the much higher value of permeability (5000) that it had when it was unmagnetized.

Of course, the permeability does not suddenly change from 5000 to 1. The permeability changes as the magnetizing force changes until saturation is reached. At saturation, permeability remains very low no matter how much the magnetizing force increases. If you were to draw a graph of the flux density compared to the magnetizing force, you would have something similar to the graph shown in figure 3-31. Figure 3-31 also includes a curve representing the value of permeability as the magnetizing force increases. Point "s" in figure 3-31 is the point of saturation. The flux density does not increase above point "s," and the permeability is at a steady, low value.

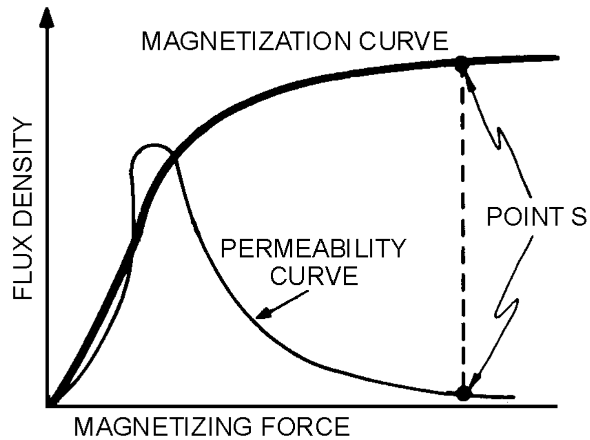


Figure 3-31.—Magnetization and permeability curves.

You have now seen how a change in the magnetizing force causes a change in permeability. The next question is, how do you change the magnetizing force? Magnetizing force is a function of AMPERE-TURNS. (An ampere-turn is the magnetomotive force developed by 1 ampere of current flowing in a coil of one turn.) If you increase the ampere-turns of a coil, the magnetizing force increases. Since it is not practical to increase the number of turns, the easiest way to accomplish this is to increase the current through the coil.

If you increase the current through a coil, you increase the ampere-turns. By increasing the ampere-turns you increase the magnetizing force. At some point, this causes a decrease in the permeability of the core. With the permeability of the core decreased, the inductance of the coil decreases. As said before, a decrease in the inductance causes an increase in power through the load. A device that uses this arrangement is called a SATURABLE-CORE REACTOR or SATURABLE REACTOR.

SATURABLE-CORE REACTOR

A saturable-core reactor is a magnetic-core reactor (coil) whose reactance is controlled by changing the permeability of the core. The permeability of the core is changed by varying a unidirectional flux (flux in one direction) through the core.

Figure 3-32 shows a saturable-core reactor that is used to control the intensity of a lamp. Notice that two coils are wound around a single core. The coil on the left is connected to a rheostat and a battery. This coil is called the control coil because it is part of the control circuit. The coil on the right is connected to a lamp (the load) and an a.c. source. This coil is called the load coil because it is part of the load circuit.

As the wiper (the movable connection) of the rheostat is moved toward the right, there is less resistance in the control circuit. With less resistance, the control-circuit current increases. This causes the amount of magnetism in the core to increase and the inductance of the coil in the load circuit to decrease (because the core is common to both coils). With less inductance in the load circuit, load current increases and the lamp gets brighter.

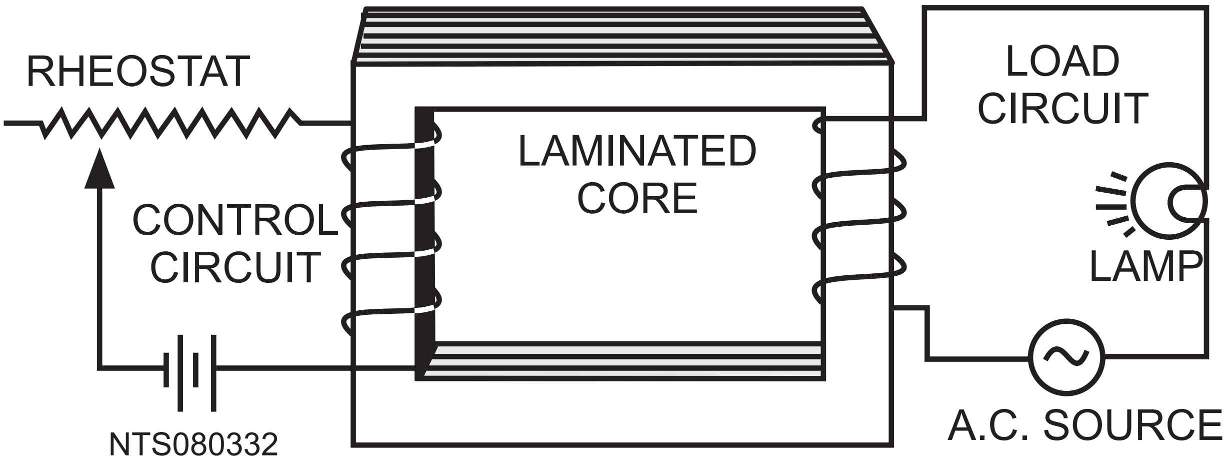


Figure 3-32.—A simple saturable-core reactor circuit.

The schematic diagram of this circuit is shown in figure 3-33. L1 is the schematic symbol for a saturable-core reactor. The control winding is shown with five loops, and the load winding is shown with three loops. The double bar between the inductors stands for an iron core, and the symbol that cuts across the two windings is a saturable-core symbol indicating that the two windings share a saturable core.

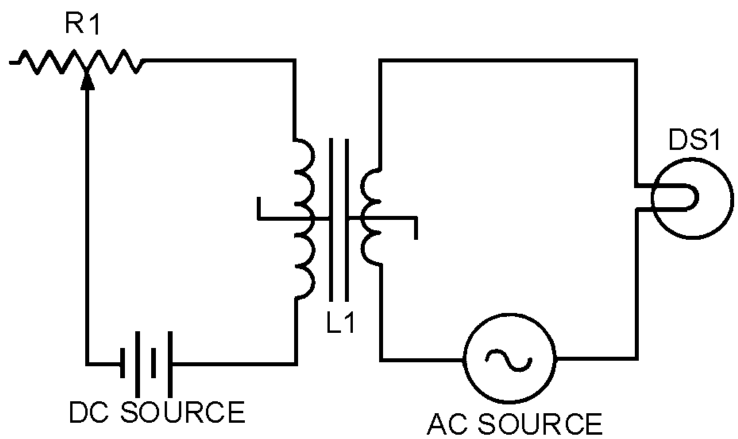


Figure 3-33.—Schematic diagram of a simple saturable-core reactor.

Now that you have seen the basic operation of a saturable-core reactor, there is one other idea to discuss before moving on to the circuitry of a magnetic amplifier. There is a point upon the magnetization curve where the saturable-core reactor should be operated. <u>The ideal operating point is the place in which a small increase in control current will cause a large increase in output power and a small decrease in control current will cause a large decrease in output power.</u> This point is on the flattest portion of the permeability curve (after its peak).

Figure 3-34 shows the magnetization and permeability curves for a saturable-core reactor with the ideal operating point (point "O") indicated. Notice point "O" on the magnetization curve. The portion of the magnetization curve where point "O" is located is called the KNEE OF THE CURVE. The knee of the curve is the point of maximum curvature. It is called the "knee" because it looks like the knee of a leg that is bent. Saturable-core reactors and magnetic amplifiers should be operated on the knee of the magnetization curve.

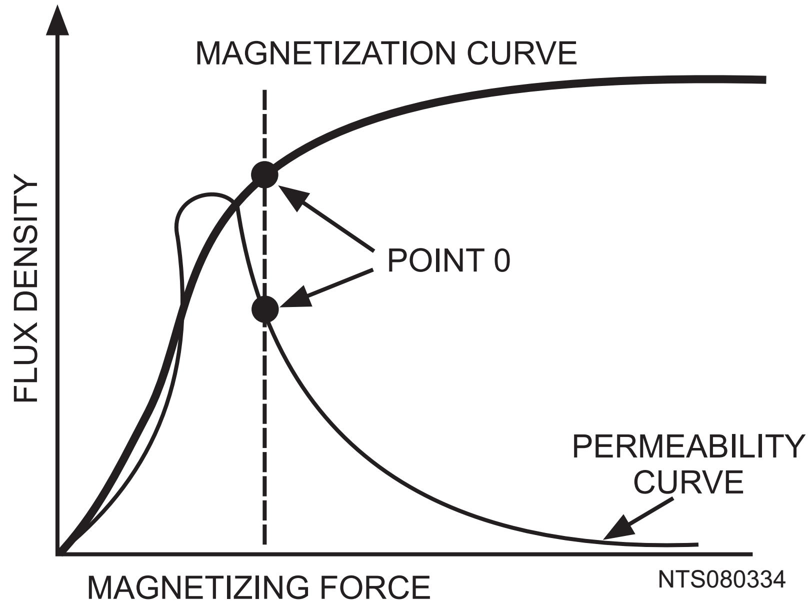


Figure 3-34.—Magnetization and permeability curves with operating point.

When the saturable-core reactor is set at the knee of the magnetization curve, any small increase in control current will cause a large increase in load current. Any small decrease in control current will cause a large decrease in load current. That is why point "O" is the ideal operating point-because small changes in control current will cause large changes in load current. In other words, the saturable-core reactor can amplify the control current. However, a saturable-core reactor is NOT a magnetic amplifier. You will find out a little later how a magnetic amplifier differs from a saturable-core reactor. First you should know a few more things about the saturable-core reactor.

If a d.c. voltage is applied to the control winding of a saturable-core reactor and an a.c. voltage is applied to the load windings, the a.c. flux will aid the d.c. flux on one half cycle and oppose the d.c. flux on the other half cycle. This is shown in figure 3-35. Load flux is indicated by the dashed-line arrows, and control flux is indicated by the solid-line arrows. View (A) shows the load and control flux adding during one half cycle of the a.c. View (B) of the figure shows the load and control flux opposing during the other half cycle of the a.c.

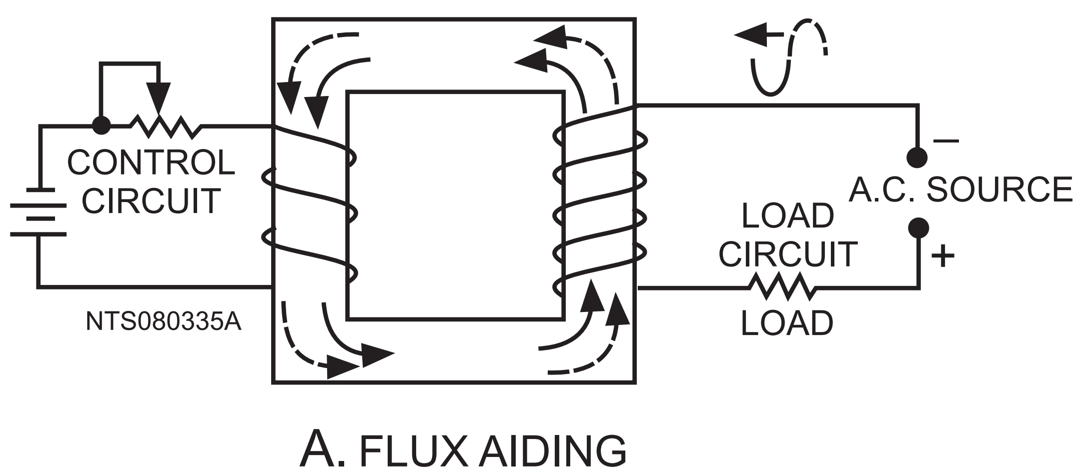


Figure 3-35A.—Flux paths in a saturable-core reactor. FLUX AIDING

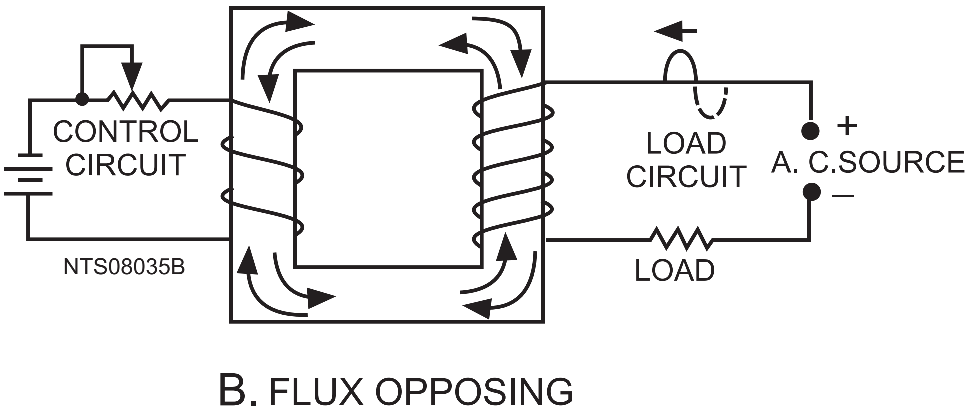


Figure 3-35B.—Flux paths in a saturable-core reactor. FLUX OPPOSING

This situation causes the operating point of the saturable-core reactor to shift with the applied a.c. However, the situation would be better if the load flux was not an influence on the control flux. Figure 3-36 shows a circuit in which this is accomplished.

During the first half cycle, the load circuit flux (dashed-line arrows) cancels in the center leg of the core. This is shown in figure 3-36, view (A). As a result, there is no effect upon the flux from the control circuit. During the second half cycle, the polarity of the a.c. (and therefore the polarity of the flux) reverses as shown in view (B). The result is the same as it was during the first half cycle. There is no effect upon the control circuit flux.

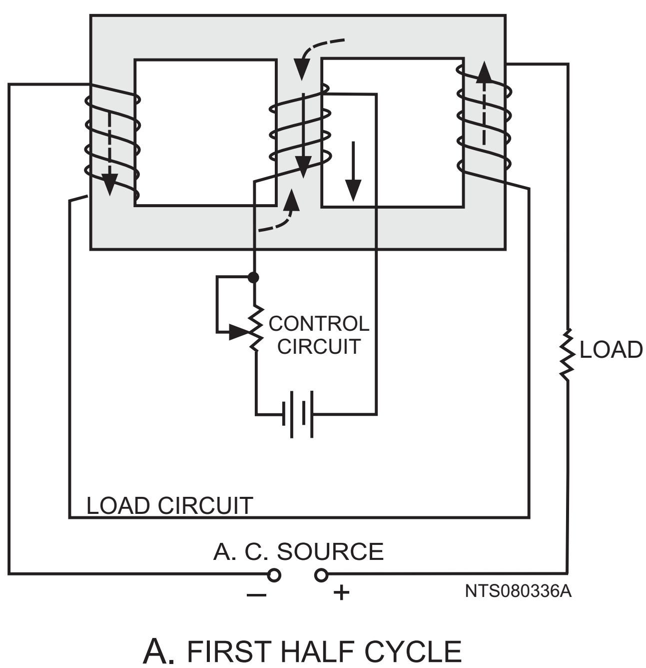


Figure 3-36A.—Three-legged, saturable-core reactor. FIRST HALF CYCLE

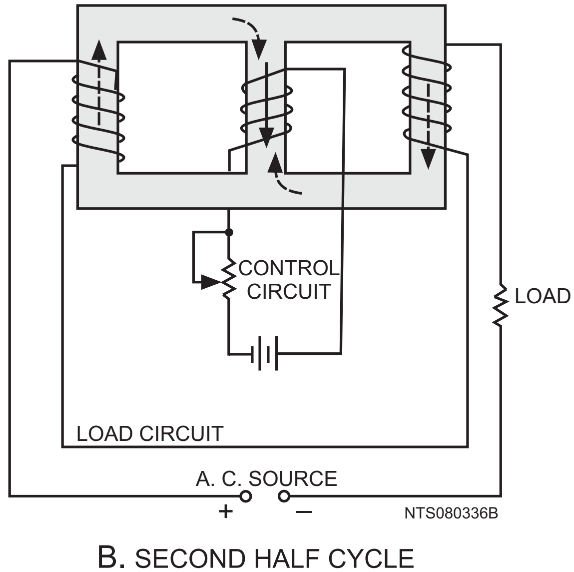


Figure 3-36B.—Three-legged, saturable-core reactor. SECOND HALF CYCLE

Another approach to solving the problem of load flux affecting control flux is shown in figure 3-37. Figure 3-37 shows a toroidal saturable-core reactor. The shape of these cores is a toroid (donut shape). The windings are wound around the cores so that the load flux aids the control flux in one core and opposes the control flux in the other core.

During the first half cycle, the flux aids in the left core and opposes in the right core, as shown in figure 3-37, view (A). During the second half cycle, the flux opposes in the left core and aids in the right core, as shown in view (B). Regardless of the amount of load flux or polarity of the load voltage, there is no net effect of load flux on control flux.

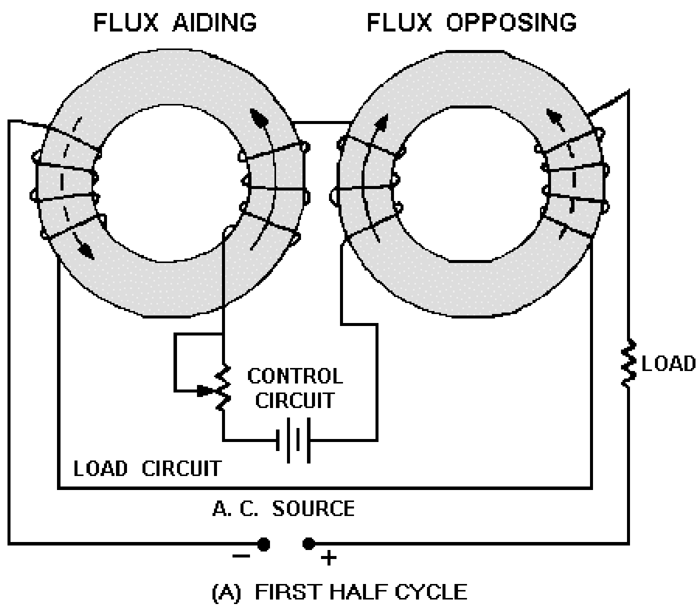


Figure 3-37A.—Toroidal saturable-core reactor. FIRST HALF CYCLE

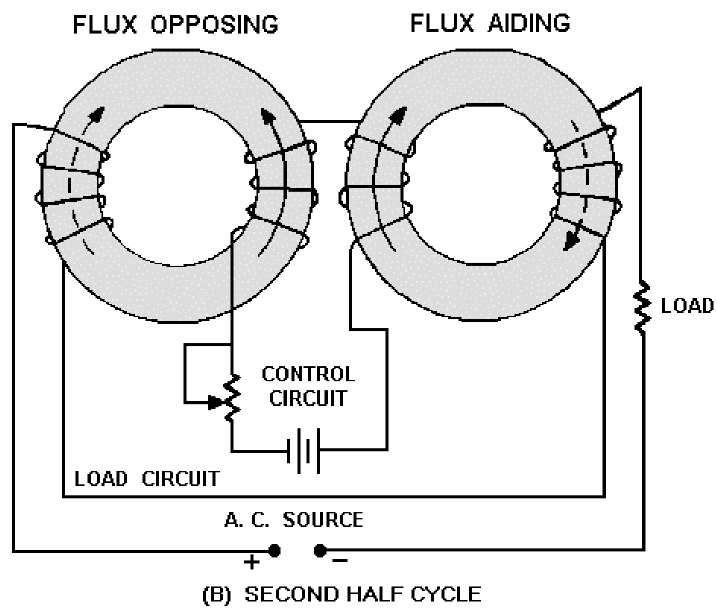


Figure 3-37B.—Toroidal saturable-core reactor. SECOND HALF CYCLE

Figures 3-36 and 3-37 both represent practical, workable saturable-core reactors. Circuits similar to these are actually used to control lighting in auditoriums or electric industrial furnaces. These circuits are sometimes referred to as magnetic amplifiers, but that is NOT technically correct. A magnetic amplifier differs from a saturable-core reactor in one important aspect: A magnetic amplifier has a rectifier in addition to a saturable-core reactor.

Q-42. If the permeability of the core of a coil increases, what happens to (a) inductance and (b) true power in the circuit?

Q-43. What happens to the permeability of an iron core as the current increases from the operating point to a large value?

Q-44. If two coils are wound on a single iron core, what will a change in current in one coil cause in the other coil?

Q-45. What symbol in figure 3-33 indicates a saturable core connecting two windings?

SIMPLIFIED MAGNETIC AMPLIFIER CIRCUITRY

If the saturable-core reactor works, why do we need to add a rectifier to produce a magnetic *amplifier? To answer this question, recall that in NEETS, Module 2—Introduction to Alternating Current and Transformers*, you were told about hysteresis loss. Hysteresis loss occurs because the a.c. applied to a coil causes the tiny molecular magnets (or electron-spin directions) to realign as the polarity of the a.c. changes. This realignment uses up power. The power that is used for realignment is a loss as far as the rest of the circuit is concerned. Because of this hysteresis loss in the saturable-core reactor, the power gain is relatively low. A rectifier added to the load circuit will eliminate the hysteresis loss and increase the gain. This is because the rectifier allows current to flow in only one direction through the load coils.

A simple half-wave magnetic amplifier is shown in figure 3-38. This is a half-wave magnetic amplifier because it uses a half-wave rectifier. During the first half cycle of the load voltage, the diode conducts and the load windings develop load flux as shown in view (A) by the dashed-line arrows. The load flux from the two load coils cancels and has no effect on the control flux. During the second half cycle, the diode does not conduct and the load coils develop no flux, as shown in view (B). The load flux never has to reverse direction as it did in the saturable-core reactor, so the hysteresis loss is eliminated.

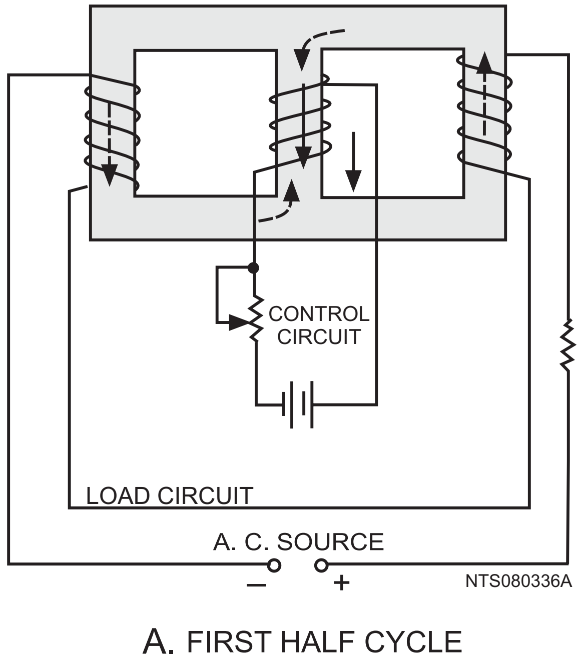


Figure 3-38A.—Simple half-wave magnetic amplifier. FIRST HALF CYCLE

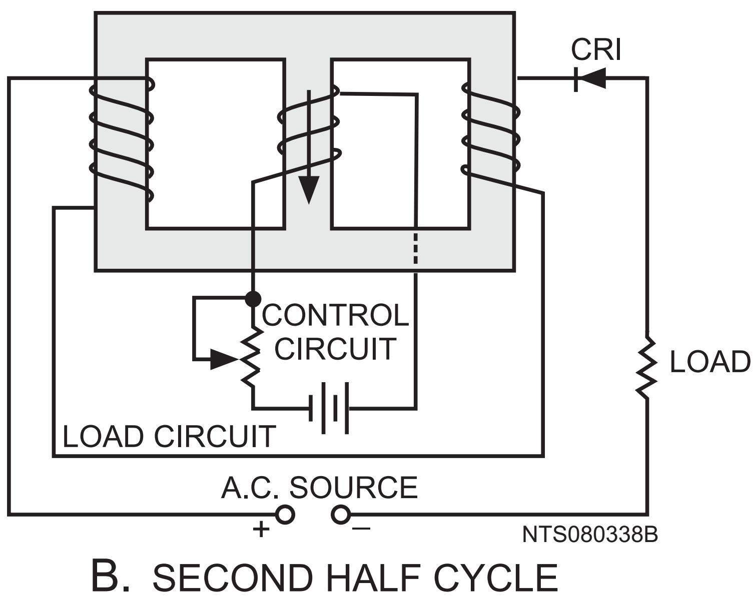


Figure 3-38B.—Simple half-wave magnetic amplifier. SECOND HALF CYCLE

The circuit shown in figure 3-38 is only able to use half of the load voltage (and therefore half the possible load power) since the diode blocks current during half the load-voltage cycle. A full-wave rectifier used in place of CR1 would allow current flow during the entire cycle of load voltage while still preventing hysteresis loss.

Figure 3-39 shows a simple full-wave magnetic amplifier. The bridge circuit of CR1, CR2, CR3, CR4 allows current to flow in the load circuit during the entire load voltage cycle, but the load current is always in the same direction. This current flow in one direction prevents hysteresis loss.

View (A) shows that during the first half cycle of load voltage, current flows through CR1, the load coils, and CR3. View (B) shows that during the second half cycle, load current flows through CR2, the load coils, and CR4.

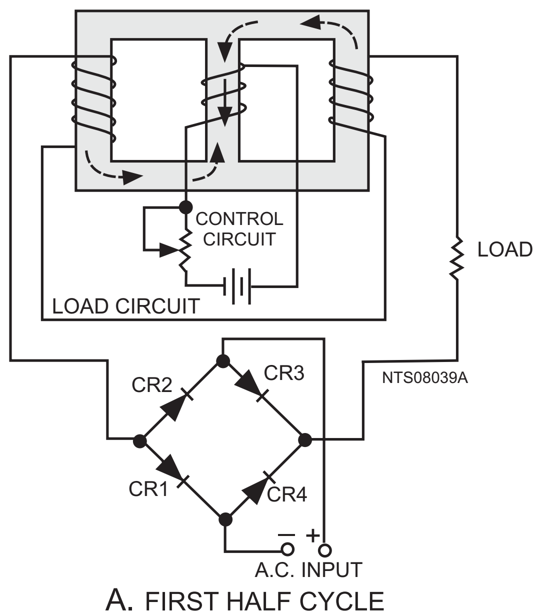


Figure 3-39A.—Simple full-wave magnetic amplifier. FIRST HALF CYCLE

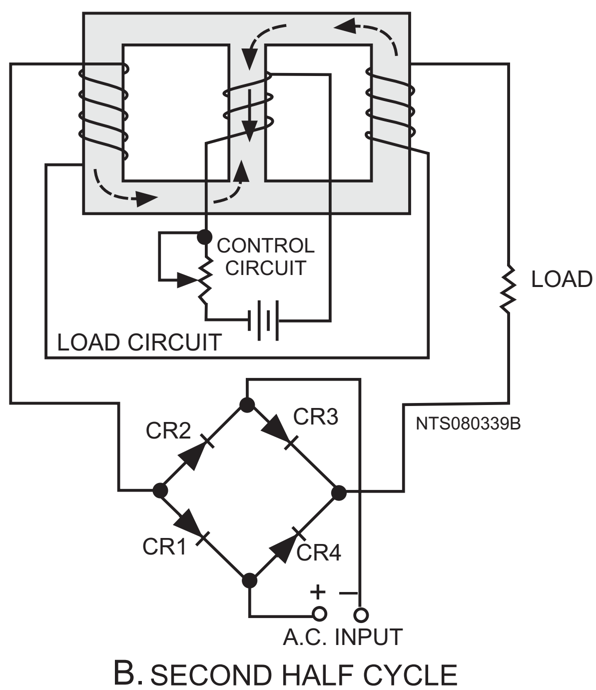


Figure 3-39B.—Simple full-wave magnetic amplifier. SECOND HALF CYCLE

Up to this point, the control circuit of the magnetic amplifier has been shown with d.c. applied to it. Magnetic-amplifier control circuits should accept a.c. input signals as well as d.c. input signals. As shown earlier in figure 3-34, a saturable-core reactor has an ideal operating point. Some d.c. must always be applied to bring the saturable core to that operating point. This d.c. is called BIAS. the most effective way to apply bias to the saturable core and also allow a.c. input signals to control the magnetic amplifier is to use a bias winding. A full-wave magnetic amplifier with a bias winding is shown in figure 3-40.

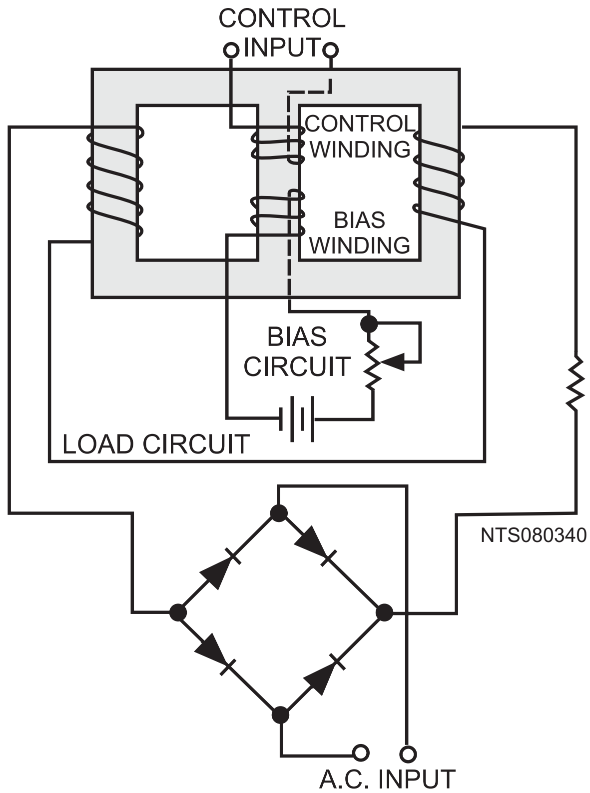


Figure 3-40.—Full-wave magnetic amplifier with bias winding.

In the circuit shown in figure 3-40, the bias circuit is adjusted to set the saturable-core reactor at the ideal operating point. Input signals, represented by the a.c. source symbol, are applied to the control input. The true power of the load circuit is controlled by the control input signal (a.c.)

The block diagram symbol for a magnetic amplifier is shown in figure 3-41. The triangle is the general symbol for an amplifier. The saturable-core reactor symbol in the center of the triangle identifies the amplifier as a magnetic amplifier. Notice the input and output signals shown. The input signal is a small-amplitude, low-power a.c. signal. The output signal is a pulsating d.c. with an amplitude that varies. This variation is controlled by the input signal and represents a power gain of 1000.

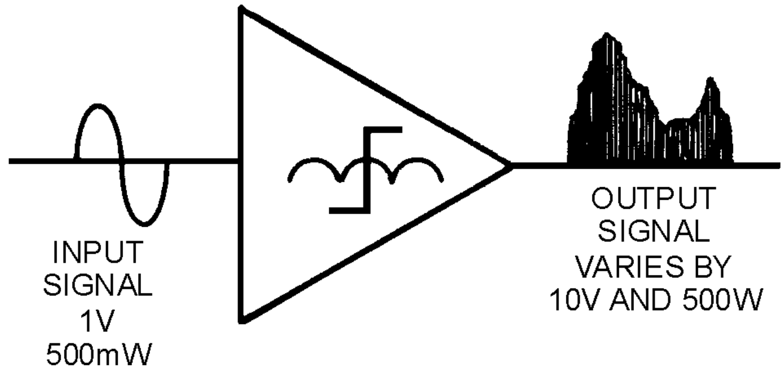


Figure 3-41.—Magnetic amplifier input and output signals.

Some magnetic amplifiers are designed so a.c. goes through the load rather than pulsating d.c. This is done by placing the load in a different circuit position with respect to the rectifier. The principle of the magnetic amplifier remains the same: Control current still controls load current.

Magnetic amplifiers provide a way of accurately controlling large amounts of power. They are used in servosystems (which are covered later in this training series), temperature or pressure indicators, and power supplies.

This chapter has presented only the basic operating theory of saturable-core reactors and magnetic amplifiers. For your convenience, simple schematic diagrams have been used to illustrate this material. When magnetic amplifiers and saturable-core reactors are used in actual equipment, the schematics may be more complex than those you have seen here. Also, you may find coils used in addition to those presented in this chapter. The technical manual for the equipment in question should contain the information you need to supplement what you have read in this chapter.

Q-46. At what portion of the magnetization curve should a magnetic amplifier be operated?

Q-47. How is the effect of load flux on control flux eliminated in a saturable-core reactor?

Q-48. What is the purpose of the rectifier in a magnetic amplifier?

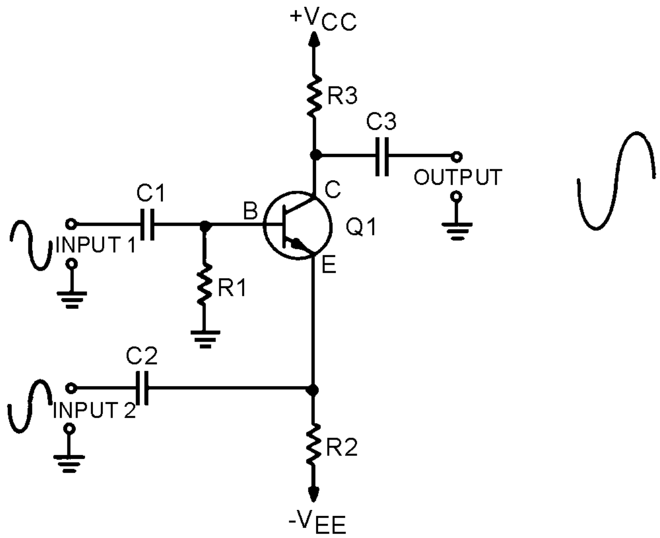
Q-49. What is used to bias a magnetic amplifier so that the control winding remains free to accept control (input) signals?

Q-50. List two common usages of magnetic amplifiers.

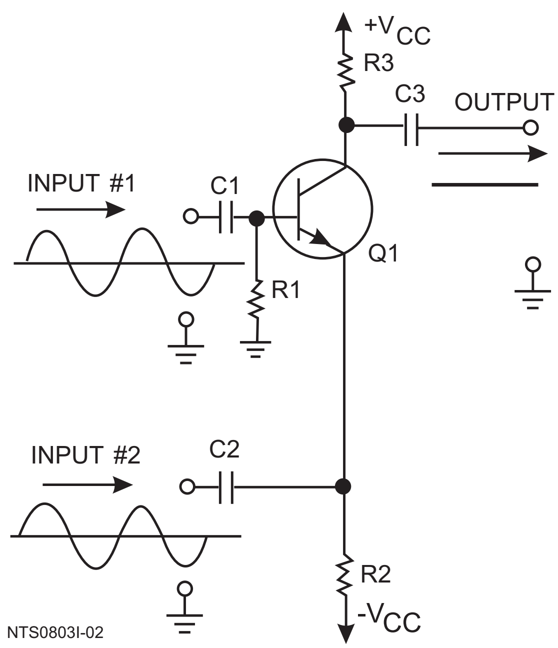
SUMMARY

This chapter has presented information on differential amplifiers, operational amplifiers, and magnetic amplifiers. The information that follows summarizes the important points of this chapter.

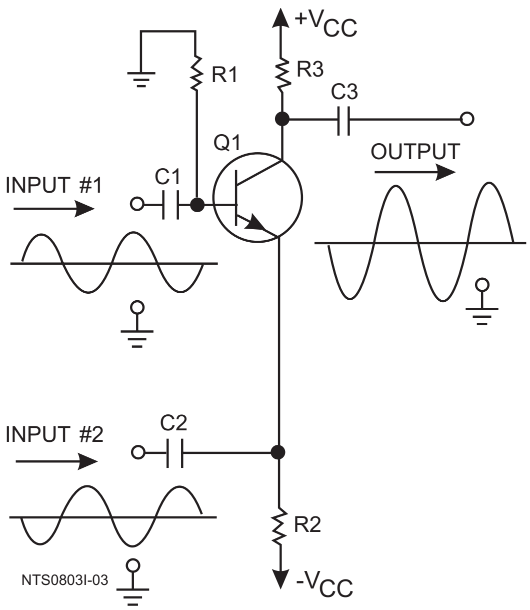
A **DIFFERENCE AMPLIFIER** is any amplifier with an output signal dependent upon the difference between the input signals. A two-input, single-output difference amplifier can be made by combining the common-emitter and common-base configurations in a single transistor.



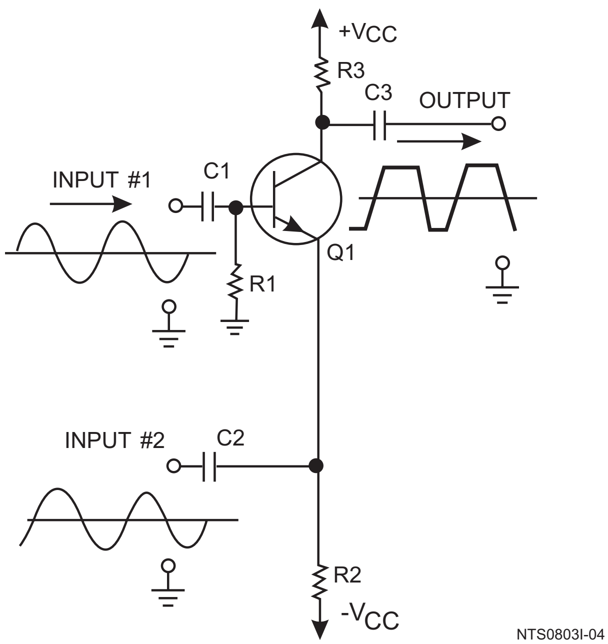
A difference amplifier can have input signals that are IN PHASE with each other, view (A), 180 DEGREES OUT OF PHASE with each other, view (B), or OUT OF PHASE BY SOMETHING OTHER THAN 180 DEGREES with each other, view (C).



A. DIFFERENTIAL AMPLIFIER IN PHASE

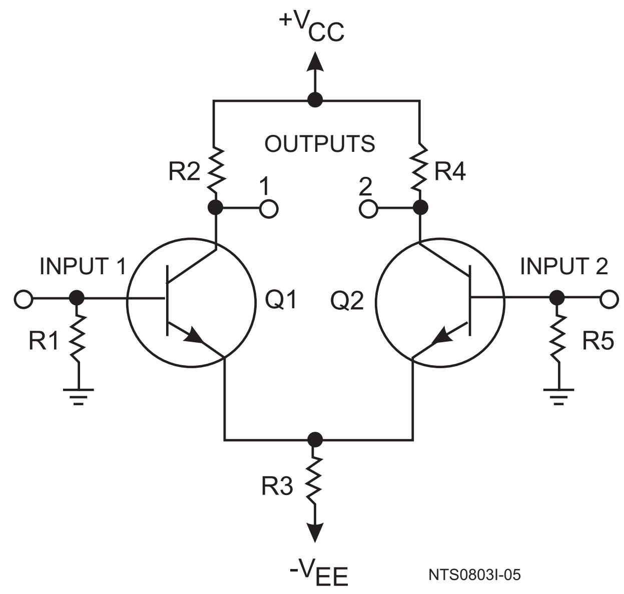


B. DIFFERENTIAL AMPLIFIER 180° OUT OF PHASE

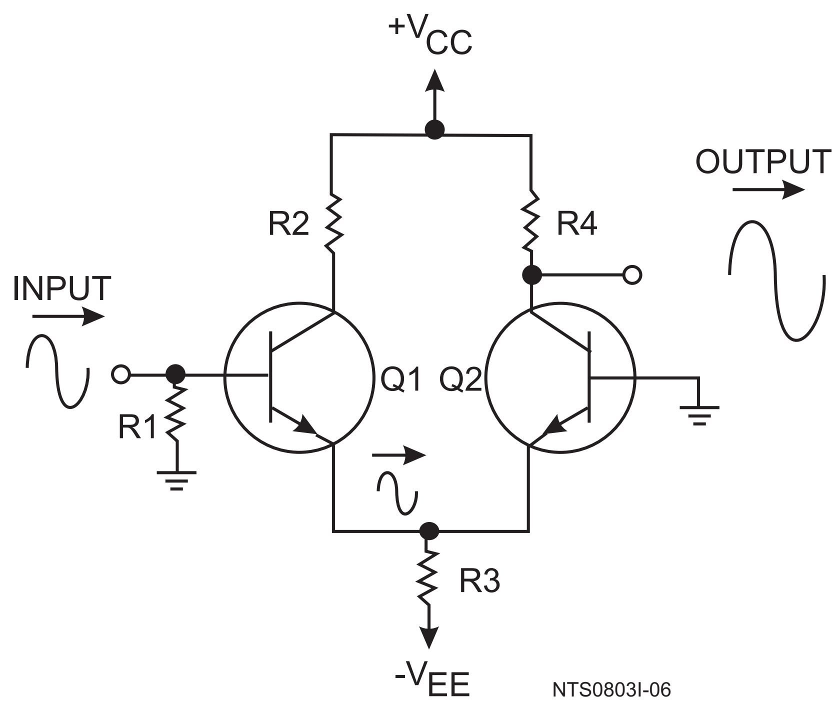


C. DIFFERENTIAL AMPLIFIER OUT OF PHASE OTHER THAN 180°

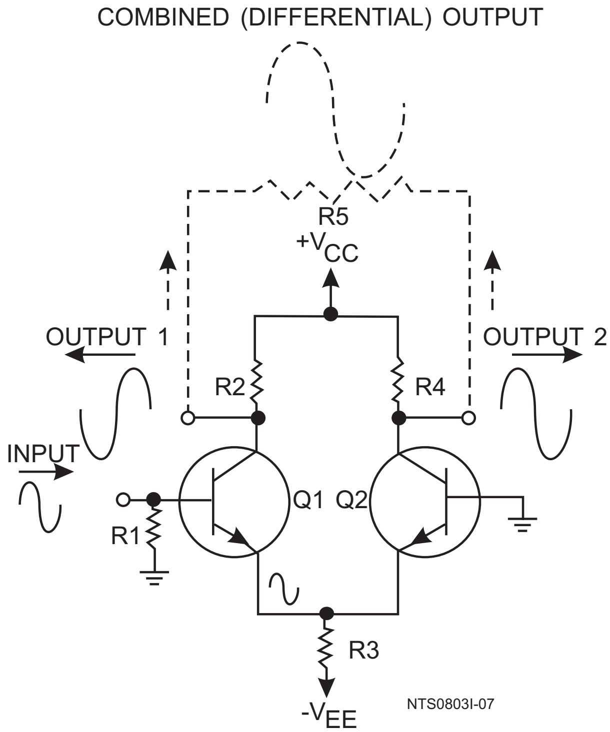
A **DIFFERENTIAL AMPLIFIER** has two possible inputs and two possible outputs. The combined output signal is dependent upon the difference between the input signals.



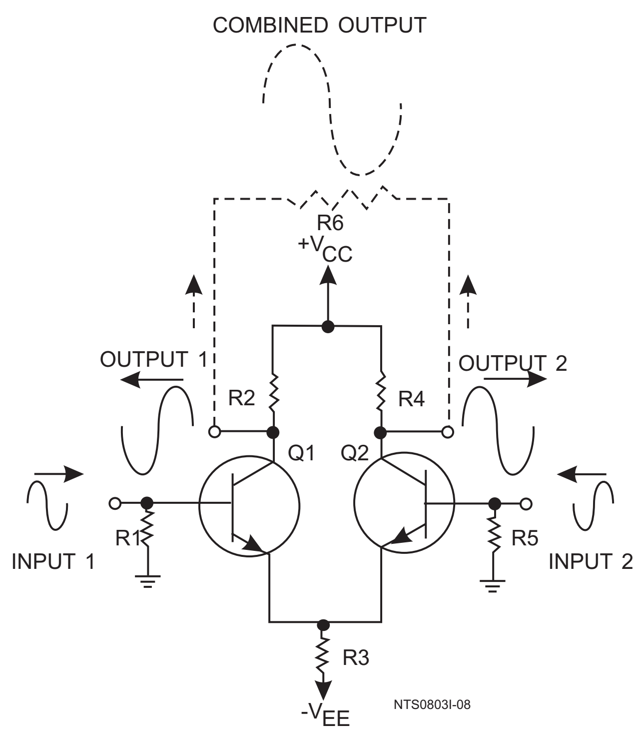
A differential amplifier can be configured with a SINGLE INPUT and a SINGLE OUTPUT



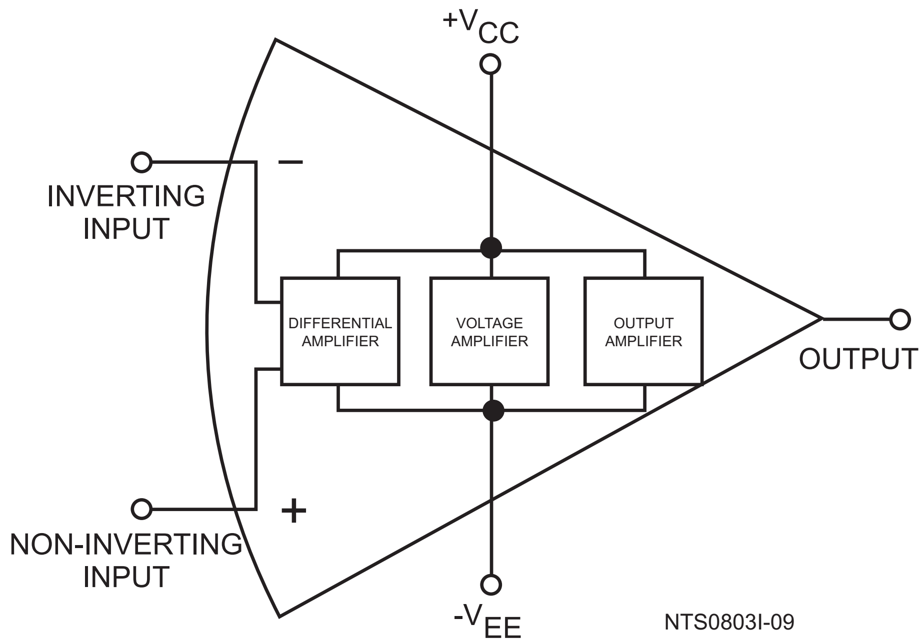
a SINGLE INPUT and a DIFFERENTIAL OUTPUT



or a DIFFERENTIAL INPUT and a DIFFERENTIAL OUTPUT.

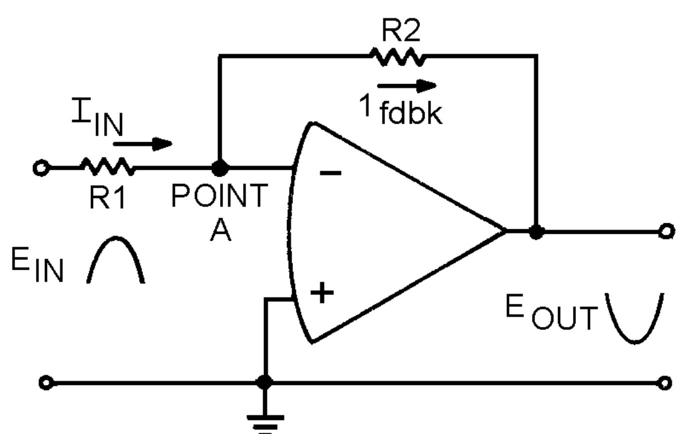


An **OPERATIONAL AMPLIFIER** is an amplifier which has very high gain, very high input impedance, and very low output impedance. An OP AMP is made from three stages: (1) a differential amplifier, (2) a high-gain voltage amplifier, and (3) an output amplifier.

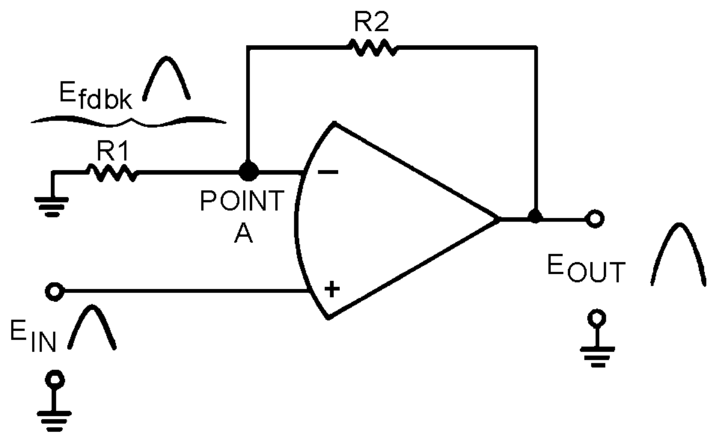


Operational amplifiers are usually used in a CLOSED-LOOP OPERATION. This means that degenerative feedback is used to lower the gain and increase the stability of the operational amplifier.

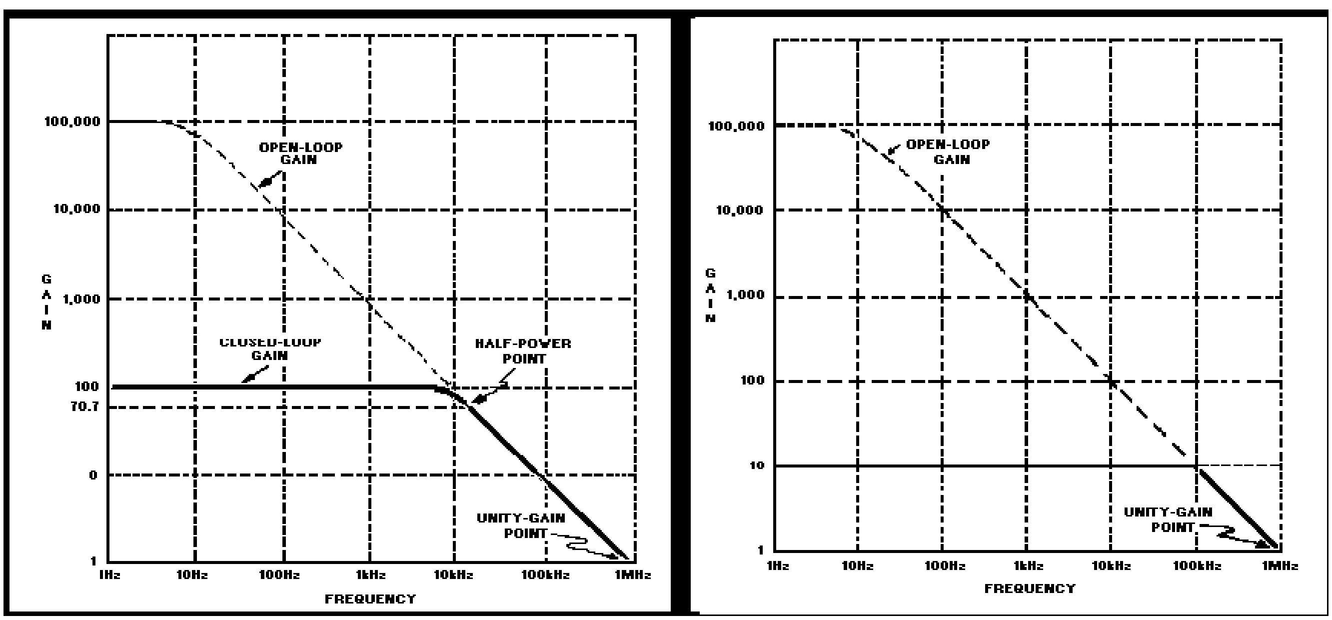
An operational amplifier circuit can be connected with an INVERTING CONFIGURATION



or a NONINVERTING CONFIGURATION.

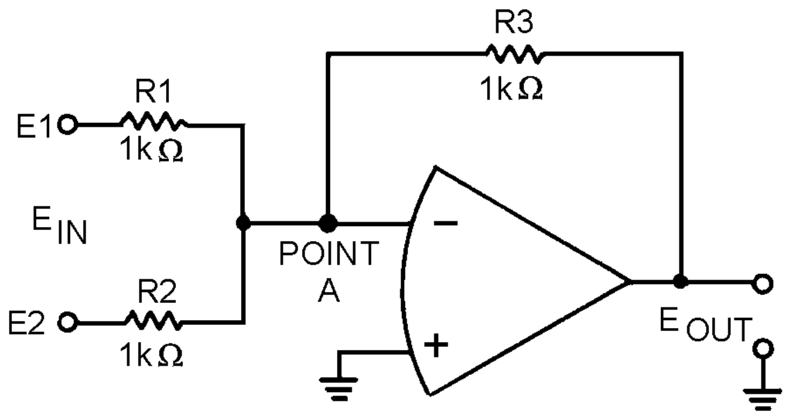


The **GAIN-BANDWIDTH PRODUCT** for an operational amplifier is computed by multiplying the gain by the bandwidth (in hertz). For any given operational amplifier, the gain-bandwidth product will remain the same regardless of the amount of feedback used.



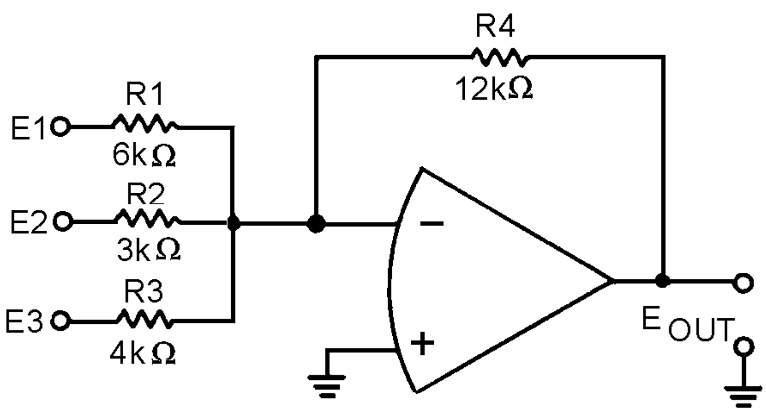
A **SUMMING AMPLIFIER** is an application of an operational amplifier in which the output signal is determined by the sum of the input signals multiplied by the gain of the amplifier:

$$E_{OUT} = \text{gain}\ (E1 + E2...)$$



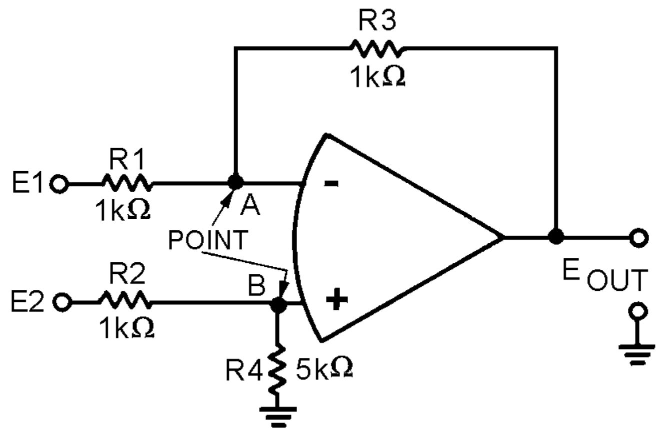
A **SCALING AMPLIFIER** is a special type of summing amplifier with the output signal determined by multiplying each input signal by a different factor (determined by the ratio of the input-signal resistor and feedback resistor) and then adding these products:

$$E_{OUT} = \left[\left(\frac{R_{fdbk}}{R_{in1}} \times E1 \right) + \left(\frac{R_{fdbk}}{R_{in2}} \times E2 \right) + \frac{R_{fdbk}}{R_{in3}} \times E3 \ldots) \right]$$

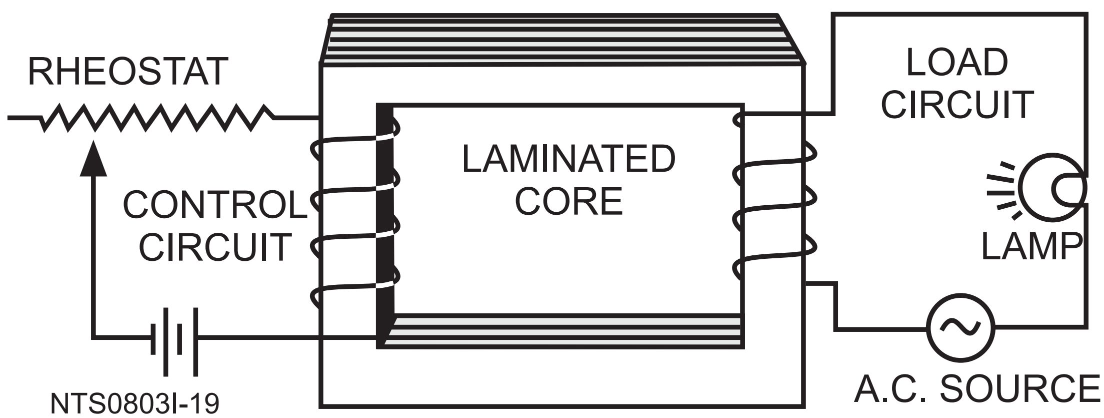


A **DIFFERENCE AMPLIFIER** is an application of an operational amplifier in which the output signal is determined by the difference between the input signals multiplied by the gain of the amplifier:

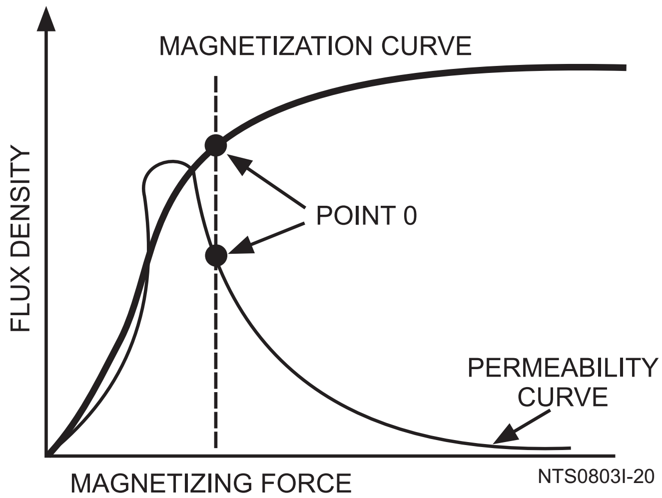
$$E_{OUT} = \text{gain} \ (E2 - E1)$$



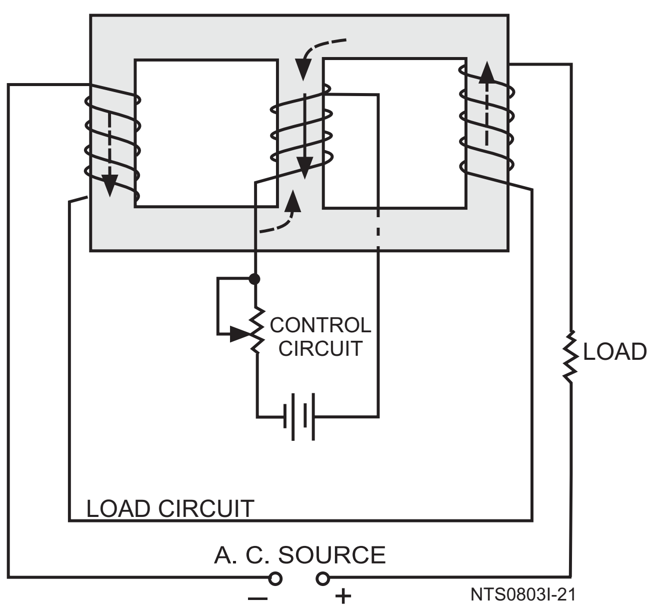
A **SATURABLE-CORE REACTOR** works upon the principle that increasing the current through a coil decreases the permeability of the core; the decreased permeability decreases the inductance of the coil which causes an increase in current (power) through the load.



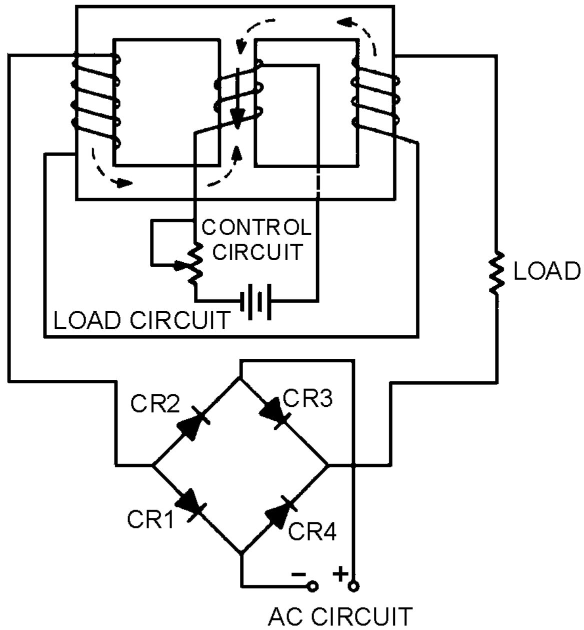
The **IDEAL OPERATING POINT** of a saturable-core reactor is on the KNEE OF THE MAGNETIZATION CURVE. At this point, small changes in control current will cause large changes in load current (power).



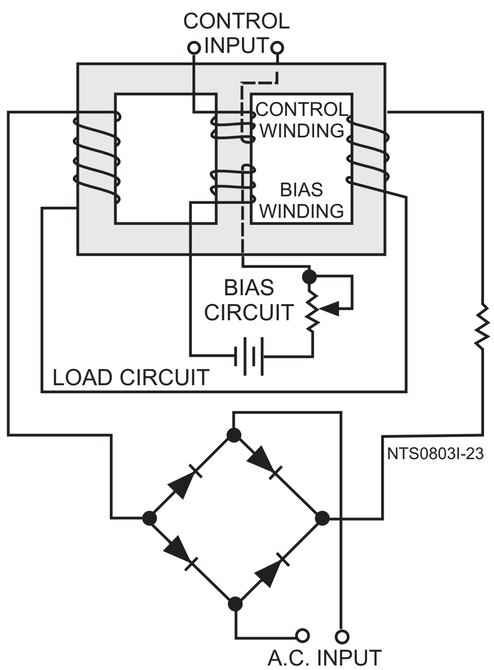
THREE-LEGGED and **TOROIDAL** saturable-core reactors solve the problem of load flux aiding and opposing control flux during alternate half cycles of the a.c. load current.



MAGNETIC AMPLIFIERS use the principle of electromagnetism to amplify signals. They are power amplifiers with a frequency response normally limited to 100 hertz or below. Magnetic amplifiers use a saturable-core reactor. A magnetic amplifier uses a RECTIFIER to solve the problem of HYSTERESIS LOSS in a saturable-core reactor.



A **BIAS WINDING** allows a d.c. bias voltage to be applied to the saturable-core reactor while a.c. control signals are applied to a separate control winding. In this way a magnetic amplifier can be set to the proper operating point.



ANSWERS TO QUESTIONS Q1. THROUGH Q50.

A-1. Two inputs, two outputs.

A-2. Common emitter (CE) and common base (CB).

A-3. No output (the signals will "cancel out").

A-4. Equal in shape and frequency to each input signal and larger in amplitude by two times than either input signal.

A-5. Equal in shape and frequency to the input signal; larger in amplitude than the input signal; half as large in amplitude as when two input signals were used that were 180 degrees out of phase.

A-6. A different shape than the input signals but larger in amplitude.

A-7. 100 millivolts.

A-8. Each output will be a sine wave with a peak-to-peak amplitude of 100 millivolts. The output signals will be 180 degrees out of phase with each other.

A-9. 200 millivolts.

A-10. 0 volts (the input signals will "cancel out").

A-11. Each output signal will be 100 millivolts.

A-12.

a. 180 degrees out of phase with each other.

b. Output signal number one will be in phase with input signal number two; output signal number two will be in phase with input signal number one.

A-13. 200 millivolts.

A-14.

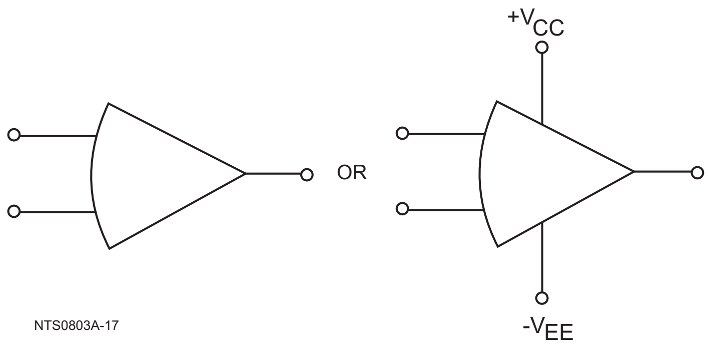
a. 100 millivolts.

b. No.

A-15. Very high gain, very high input impedance, very low output impedance.

A-16. An integrated circuit (chip).

A-17.



A-18.

a. Differential amplifier.

b. Voltage amplifier.

c. Output amplifier.

A-19. The use of degenerative (negative) feed-back.

A-20. Both the input signal and the feedback signal.

A-21.

a. Inverting.

b. Inverting.

A-22. 0 volts.

A-23. Virtual.

A-24. -50 millivolts.

A-25. 50 kilohertz (Gain = 10; Gain-Bandwidth Product = 500,000;

$$\text{BW} = \frac{500{,}000\ (\text{Hz})}{10} = 50\text{kHz}$$

NTS0803A-25

A-26. 60 millivolts.

A-27. 1 megahertz.

Open-loop Gain-Bandwidth Product = Closed-loop Gain-Bandwidth Prod.

Open-loop Gain-Bandwidth Product = 200,000 $\times$ 30 (Hz)

Open-loop Gain Bandwidth Product = 600,000

Closed-loop Gain Bandwidth Product = 6 $\times$ Bandwidth

6,000,000 = 6 $\times$ Bandwidth

1,000,000 (Hz) = Bandwidth

A-28. The adder simply adds the input signals together while the summing amplifier multiplies the sum of the input signals by the gain of circuit.

A-29. Yes, a summing amplifier can have as many inputs as desired.

A-30. A summing amplifier that applies a factor to each input signal beforeadding the results.

A-31. A scaling amplifier.

A-32. $E_{out} = -72V$

Solution:

$$E_{out} = -(+2V \text{ x } \frac{30k\Omega}{5k\Omega}) + (+6V \text{ x } \frac{30k\Omega}{3k\Omega})$$

$$E_{out} = - (+2V \text{ x } 6) + (+6V \text{ x } 10)$$

NTS0803A-32

A-33. 0 volts. (The two inputs to the operational amplifier are both at 0 volts.)

A-34. The difference amplifier multiplies the difference between the two inputs by the gain of the circuit while the subtractor merely subtracts one input signal from the other.

A-35. No.

A-36. A difference amplifier.

A-37.

$E_{out} = +60V$

Solution:

$$E_{out} = [(+11V) - (+5V)] \; x \; \frac{R_3}{R_1}$$

$$E_{out} = (+6V) + \frac{20k\Omega}{2k\Omega}$$

$$E_{out} = (+6V) \; x \; 10$$

NTS0803A-37

A-38. *0 volts. (The two inputs to the operational amplifier are both at the same potential.)*

A-39. *An audio (or low) frequency power amplifier.*

A-40. *A change in inductance in a series LR circuit causes a change in true power.*

A-41. *It decreases.*

A-42. *(a) Inductance increases; (b) true power decreases.*

A-43. *Permeability decreases.*

A-44. *A change in inductance.*

A-45.

A-46. *The knee of the curve.*

A-47. *Use two load windings whose flux effects cancel in the core of the reactor or use two load windings on two toroidal cores so that load flux always aids control flux in one core and opposes control flux in the other core.*

A-48. *The rectifier eliminates hysteresis loss.*

A-49. *A bias winding and associated circuitry.*

A-50. Servosystems, temperature recorders, or power supplies.

APPENDIX I

GLOSSARY

AMPERE-TURN—The magnetomotive force developed by one ampere of current flowing through a coil of one turn.

AMPLIFICATION—The process of enlarging a signal in amplitude (as of voltage or current).

AMPLIFIER—A device that enables an input signal to control an output signal. The output signal will have some (or all) of the characteristics of the input signal but will generally be larger than the input signal in terms of voltage, current, or power.

AMPLITUDE—The size of a signal. Generally used to describe voltage, current, or power.

AUDIO AMPLIFIER—An amplifier designed to amplify frequencies between 15 hertz (15 Hz) and 20 kilohertz (20 kHz).

BANDWIDTH—The difference between the highest usable frequency of a device (upper frequency limit) and the lowest usable frequency of the device (lower frequency limit)—measured at the half-power points.

BYPASS CAPACITOR—A capacitor used to transfer unwanted signals out of a circuit; e.g., coupling an unwanted signal to ground. Also called a DECOUPLING CAPACITOR.

COMBINATION PEAKING—A technique in which a combination of peaking coils in series <u>and</u> parallel (shunt) with the output signal path is used to improve high-frequency response.

COMPENSATION—The process of overcoming the problems associated with frequencies in an amplifier.

COUPLING—The process of transferring energy from one point in a circuit to another point or from one circuit to another.

COUPLING CAPACITOR—A capacitor used to couple signals.

DECOUPLING CAPACITOR—A capacitor used to transfer unwanted signals out of a circuit; e.g., coupling an unwanted signal to ground. Also called a BYPASS CAPACITOR.

DEGENERATIVE FEEDBACK—Feedback in which the feedback signal is out of phase with the input signal, also called NEGATIVE FEEDBACK.

DIFFERENTIAL AMPLIFIER—An amplifier with an output which is determined by the difference between two input signals and which can provide up to two output signals.

DISTORTION—Any unwanted change between an input signal and an output signal.

DRIVER—An electronic circuit that supplies the input to another circuit.

FEEDBACK—The process of sending part of an output signal of an amplifier back to the input of the amplifier.

FIDELITY—The quality of reproducing an output signal exactly like the input signal except for amplitude (and somtimes phase); i.e., output and input signals exactly alike in terms of frequency and shape.

FREQUENCY-DETERMINING NETWORK—A circuit that provides the desired response (maximum or minimum impedance) at a specific frequency.

FREQUENCY-RESPONSE CURVE—A curve showing the output of an amplifier (or any other device) in terms of voltage or current plotted against frequency with a fixed-amplitude input signal.

GAIN-BANDWIDTH PRODUCT—The number that results when the gain of a circuit is multiplied by the bandwidth of that circuit. For an operational amplifier, the gain-bandwidth product for one configuration will always equal the gain-bandwidth product for any other configuration of the same amplifier.

HALF-POWER POINTS—The points on a frequency-response curve at which the output power is one-half of the maximum power out.

HIGH-FREQUENCY COMPENSATION—See peaking coil.

KNEE OF THE CURVE—The point of maximum curvature. (Shaped like the knee of a leg that is bent.)

MAGNETIC AMPLIFIER (MAG AMP)—An amplifier that uses electromagnetic effects to provide amplification of a signal. The magnetic amplifier uses a changing inductance to control the power delivered to a load.

NEGATIVE FEEDBACK—Feedback in which the feedback signal is out of phase with the input signal. Also called DEGENERATIVE FEEDBACK.

NEUTRALIZATION—The process of counteracting or "neutralizing" the effects of interelectrode capacitance.

OPERATIONAL AMPLIFIER (OP AMP)—An amplifier designed to perform computing or transfer operations and which has the following characteristics: (1) very high gain, (2) very high input impedance, and (3) very low out put impedance.

PEAKING COIL—An inductor used in an amplifier to provide high-frequency compensation which extends the high-frequency response of the amplifier.

PERMEABILITY—The measure of the ability of a material to act as a path for additional magnetic lines of force.

PHASE SPLITTER—A device that provides two output signals from a single input signal. The two output signals will differ from each other in phase.

POSITIVE FEEDBACK—Feedback in which the feedback signal is in phase with the input signal. Also called REGENERATIVE FEEDBACK.

POWER AMPLIFIER—An amplifier in which the output-signal power is greater than the input-signal power.

PUSH-PULL AMPLIFIER—An amplifler which uses two transistors (or electron tubes) whose output signals are combined to provide a larger gain (usually a power gain) than a single transistor (or electron tube) can provide.

REGENERATIVE FEEDBACK—Feedback in which the feedback signal is in phase with the input signal. Also called POSITIVE FEEDBACK

RF (RADIO FREQUENCY) AMPLIFIER—An amplifier designed to amplify signals with frequencies between 10 kilohertz (10 kHz) and 100,000 megahertz (100,000 MHz).

RF (RADIO FREQUENCY) TRANSFORMER—A transformer specially designed for use with rf (radio frequencies). An rf transformer is wound onto a tube of non- magnetic material and has a core of either powdered iron or air.

SATURATION (MAGNETIC CORE)—The condition in which a magnetic material has reached a maximum flux density and the permeability has decreased to a value of (approximately) 1.

SATURABLE-CORE REACTOR—A coil whose reactance is controlled by changing the permeability of the core.

SERIES PEAKING—A technique used to improve high-frequency response in which a peaking coil is placed in series with the output signal path.

SHUNT PEAKING—A technique used to improve high-frequency response in which a peaking coil is placed in parallel (shunt) with the output signal path.

SIGNAL—A general term used to describe any a.c. or d.c. of interest in a circuit; e.g., input signal.

STAGE—One of a series of circuits within a single device; e.g., first stage of amplification.

SWAMPING RESISTOR—A resistor used to increase or "broaden" the bandwidth of a circuit.

TUNED CIRCUIT—An LC circuit used as frequency-determining network.

VIDEO AMPLIFIER—An amplifier designed to amplify the entire band of frequencies from 10 hertz (10 Hz) to six megahertz (6 MHz). Also called a WIDE-BAND AMPLIFIER.

VIRTUAL GROUND—A point in a circuit which is at ground potential (0 V) but is not connected to ground.

VOLTAGE AMPLIFIER—An amplifier in which the output-signal voltage is greater than the input-signal voltage.

WIDE-BAND AMPLIFIER—An amplifier designed to pass an extremely wide band of frequencies, i.e., video amplifier.

INDEX-1

P

R

S

T

V

ASSIGNMENT 1

Textbook assignment: Chapter 1, “Amplifiers,” pages 1-1 through 1-40. Chapter 2, “Video and RF Amplifiers,” pages 2-1 through 2-34.

1-1. The control of an output signal by an input signal resulting in the output signal having some (or all) of the characteristics of the input signal is known by which of the following terms?

1. Multiplication
2. Magnification
3. Amplification
4. Addition

1-2. Which of the following statements describes the relationship of input and output signals in a amplifier?

1. The input signal is actually changed into the output signal
2. Both the input and output signal are unchanged; neither is affected by the other
3. The input signal is controlled by the output signal and the output signal remains unchanged
4. The input signal remains unchanged and the output signal is controlled by the input signal

1-3. Why are amplifiers used in electronic devices?

1. To provide signals of usable amplitude
2. To "pick up" broadcast signals
3. To select the proper broadcast signal
4. To change the broadcast signal to an audio signal

1-4. Most amplifiers can be classified in which of the following ways?

1. Function and size
2. Power requirements and size
3. Function and frequency response
4. Frequency response and power requirements

1-5. The speaker system of a record player should be driven by which of the following types of amplifier?

1. An audio power amplifier
2. A video voltage amplifier
3. A direct-current voltage amplifier
4. An alternating-current rf amplifier

1-6. The signal from a radio antenna should be amplified by which of the following types of amplifier?

1. An rf voltage amplifier
2. A video power amplifier
3. A direct-current audio amplifier
4. An alternating-current power amplifier

1-7. The class of operation of an amplifier is determined by which of the following factors?

1. The gain of the stage
2. The efficiency of the amplifier
3. The amount of time (in relation to the input signal) that current flows in the output circuit
4. The amount of current (in relation to the input-signal current) that flows in the output circuit

1-8. Which of the following is NOT a class of operation for an amplifier?

1. A
2. C
3. AB
4. AC

1-9. If the output of a circuit should be a representation of less than one-half of the input signal, what class of operation should be used?

1. A
2. C
3. AB
4. AC

1-10. What class of operation is the most efficient?

1. A
2. C
3. AB
4. AC

1-11. What class of operation has the highest fidelity?

1. A
2. C
3. AB
4. AC

1-12. What is the purpose of an amplifier-coupling network?

1. To "block" d.c.
2. To provide gain between stages
3. To separate one stage from another
4. To transfer energy from one stage to another

1-13. Which of the following is NOT a method of coupling amplifier stages?

1. RC
2. Resistor
3. Impedance
4. Transformer

1-14. What is the most common form of coupling?

1. RC
2. Resistor
3. Impedance
4. Transformer

1-15. Which of the following types of coupling is usually used to couple the output from a power amplifier?

1. RC
2. Resistor
3. Impedance
4. Transformer

1-16. Which of the following types of amplifiers have both high and low frequency response limitations?

1. RC
2. Resistor
3. Impedance
4. Transformer

1-17. Which of the following types of coupling is most effective at high frequencies?

1. RC
2. Resistor
3. Impedance
4. Direct

THIS SPACE LEFT BLANK INTENTIONALLY.

1-18. For maximum power transfer between circuits, what impedance relationship should there be between the two circuits?

1. The output impedance of circuit number one should be higher
2. The input impedance of circuit number one should be higher than the output impedance of circuit .number two
3. The output impedance of circuit number one should be lower than the input impedance of circuit number two
4. The output impedance of circuit number one should be equal to the input impedance of circuit number two

1-19. For maximum current at the input to a circuit, what should the relationship of the input impedance be to the output impedance of the previous stage?

1. Higher than
2. Lower than
3. Equal to
4. The impedance relationship is immaterial

1-20. What is the (a) input impedance and (b) output impedance of a common-base transistor configuration?

1. (a) Low (b) low
2. (a) Low (b) high
3. (a) High (b) low
4. (a) High (b) high

1-21. What transistor configuration should be used to match a high output impedance to a low input impedance?

1. Common collector
2. Common emitter
3. Common gate
4. Common base

1-22. What type of coupling is most useful for impedance matching?

1. RC
2. Resistor
3. Impedance
4. Transformer

1-23. What is feedback?

1. The control of a circuit output signal by the input signal
2. The control of a circuit input signal by the output signal
3. The coupling of a portion of the output signal to the input of the circuit
4. The coupling of a portion of the input signal to the output of the circuit

1-24. Which of the following terms describe the two types of feedback?

1. Positive and negative
2. Degenerative and regenerative
3. Both 1 and 2 above
4. Bypassed and unbypassed

1-25. What type of feedback provide an increased amplitude output signal?

1. Positive
2. Negative
3. Bypassed
4. Unbypassed

1-26. Distortion caused by amplifier saturation can be reduced by using which of the following types of feedback?

1. Positive
2. Negative
3. Regenerative
4. Unbypassed

1-27. What type feedback is provided if the feedback signal is out of phase with the input signal?

1. Unbypassed
2. Bypassed
3. Negative
4. Positive

1-28. What type of feedback is provided by a capacitor connected across the emitter-resistor in a common-emitter transistor amplifier?

1. Bypassed
2. Positive
3. Negative
4. Unbypassed

1-29. What are the (a) inputs and (b) outputs of a phase splitter?

1. (a) Two signals in phase
   (b) One signal
2. (a) Two signals out of phase
   (b) One signal
3. (a) One signal
   (b) Two signals in phase
4. (a) One signal
   (b) Two signals out of phase

1-30. A single-stage, two transistor amplifier that uses a phase splitter input is classified as what type of amplifier?

1. Inverse
2. Push-pull
3. Phase splitter
4. Regenerative

1-31. Which of the following is a common use for a push-pull amplifier?

1. The first stage of a video amplifier
2. The amplifier stage connected directly to an antenna
3. The second stage in a four stage rf amplifier
4. The final stage in an audio amplifier

1-32. What is the advantage of a push-pull amplifier as compared to a single transistor amplifier?

1. Lower cost
2. Fewer parts
3. Higher gain
4. Less power usage

1-33. To provide good fidelity output signals, which of the following classes of operation CANNOT be used by a push-pull amplifier?

1. A
2. B
3. C
4. AB

1-34. What is the bandwidth of an amplifier?

1. The actual frequencies the amplifier is effective in amplifying
2. The difference between the high and low frequencies seen at the input of the amplifier
3. The width, in inches, between the half-power points on a frequency-response curve
4. The difference between the lowest and highest frequency shown on a frequency-response curve

THIS SPACE LEFT BLANK INTENTIONALLY.

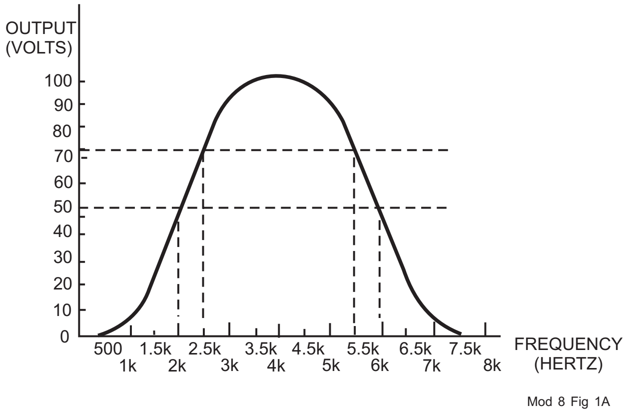


Figure 1A.—Frequency-response curve.

IN ANSWERING QUESTIONS 1-35 AND 1-36 REFER TO FIGURE 1A.

1-35. What are the (a) upper and (b) lower frequency limits shown?

1. (a) 8 kHz (b) 0 Hz
2. (a) 7.5 kHz (b) 0 Hz
3. (a) 6.0 kHz (b) 2 kHz
4. (a) 5.5 kHz (b) 2.5 kHz

1-36. What is the bandwidth shown?

1. 1 inch
2. 8 kHz
3. 3 kHz
4. 2 kHz to 8 kHz

1-37. Which of the following limit(s) the frequency response of a transistor amplifier?

1. The inductance
2. The transistor
3. The capacitance
4. All of the above

1-38. What type of feedback is caused by interelectrode capacitance?

1. Bypassed
2. Negative
3. Positive
4. Regenerative

1-39. What happens to capacitive reactance as frequency decreases?

1. It increases
2. It decreases
3. It remains the same
4. It cannot be determined

1-40. What happens to inductive reactance as frequency increases?

1. It increases
2. It decreases
3. It remains the same
4. It cannot be determined

1-41. What is the major factor that limits the high frequency response of an amplifier?

1. Inductance
2. Resistance
3. Capacitance
4. Transformer reactance

1-42. What components can be used to increase the high-frequency response of an amplifier?

1. Diodes
2. Inductors
3. Resistors
4. Capacitors

1-43. What determines whether a peaking component is considered "series" or "shunt"?

1. The relationship of the component to the power supply
2. The relationship of the component to the input signal path
3. The relationship of the component to the amplifying device
4. The relationship of the component to the output signal path

1-44. What is the arrangement of both "series" and "shunt" peaking components called?

1. Coordinated
2. Combination
3. Combined
4. Complex

1-45. Which of the following components in a transistor amplifier circuit tends to limit the low-frequency response of the amplifier?

1. The transistor
2. The load resistor
3. The coupling capacitor
4. The input-signal-developing resistor

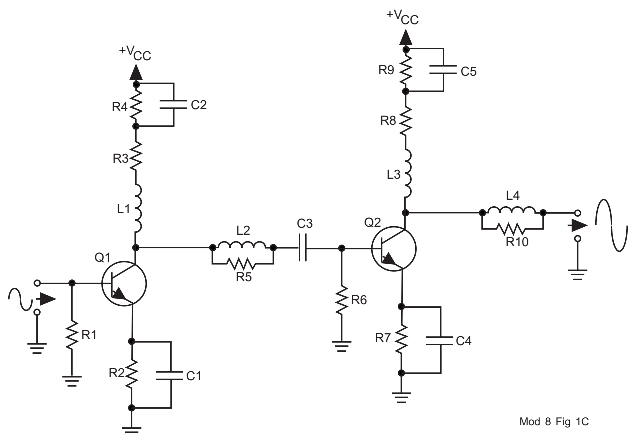


Figure 1B.—Video Amplifier.

IN ANSWERING QUESTIONS 1-46 THROUGH 1-52, REFER TO FIGURE 1B.

1-46. What is the purpose of L1 in relation to Q1?

1. Decoupling
2. Shunt peaking
3. Series peaking
4. Input-signal developing

1-47. What is the purpose of C4 in relation to Q2?

1. Decoupling (Bypass)
2. Shunt peaking
3. Series peaking
4. Input-signal developing

1-48. What is the purpose of R4 in relation to Q1?

1. Coupling resistor
2. Input-signal developing
3. Low-frequency compensation
4. High-frequency compensation

1-49. What is the purpose of L4 in relation to Q2?

1. Coupling
2. Decoupling
3. Shunt peaking
4. Series peaking

1-50. What is the purpose of R10 in relation to Q2?

1. Swamping
2. Input-signal developing
3. Low-frequency compensation
4. High-frequency compensation

1-51. Which of the following components is/are used for high-frequency compensation for Q2?

1. C4
2. L3
3. R9
4. All of the above

1-52. Which of the following components is/are used for low-frequency compensation for Q1?

1. C1
2. C2
3. R3
4. All of the above

1-53. What is the effect of the gain of an amplifier if the input-signal developing impedance is decreased?

1. It decreases
2. It increases
3. It remains the same
4. It cannot be determined

1-54. What is the effect on the gain of an amplifier if the output-signal-developing impedance is increased?

1. It decreases
2. It increases
3. It remains the same
4. It cannot be determined

1-55. What is/are the purpose(s) of a frequency-determining network in an rf amplifier?

1. To create a large bandpass
2. To compensate for low-frequency losses
3. To provide maximum impedance at a given frequency
4. All of the above

1-56. Of the following networks, which could be used as a frequency-determining network for an rf amplifier?

1. A parallel-resistor network
2. A series-resistor network
3. A parallel RC network
4. A parallel LC network

1-57. Which of the following methods may be used to tune an LRC frequency-determining network to a different frequency?

1. Vary the capacitance
2. Vary the inductance
3. Both 1 and 2 above
4. Vary the resistance

1-58. What is the most common form of coupling for an rf amplifier?

1. RC
2. Resistor
3. Impedance
4. Transformer

1-59. Which of the following advantages are provided by transformer coupling?

1. Simpler power supplies can be used
2. The circuit is not affected by frequency
3. Low-frequency response is improved
4. Fewer parts are used

1-60. If a current gain is desired, which of the following elements/networks should be used as an output-coupling device?

1. An RC network
2. A resistive network
3. A step-up transformer
4. A step-down transformer

1-61. Which of the following techniques would cause a too-narrow bandpass in an rf amplifier?

1. An overcoupled transformer
2. A loosely coupled transformer
3. The use of a swamping resistor
4. The use of a frequency-determining network

1-62. Which of the following techniques would cause low gain at the center frequency of an rf amplifier?

1. An overcoupled transformer
2. A loosely coupled transformer
3. The use of a swamping resistor
4. The use of a frequency-determining network

1-63. What type of transformer coupling should be used in an rf amplifier?

1. Ideal
2. Loose
3. Optimum
4. Overcoupling

1-64. Which of the following methods provides the widest band-pass in an rf amplifier?

1. A swamping resistor
2. A loosely coupled amplifier
3. A large input-signal-developing resistor
4. A small output-signal-developing resistor

1-65. Which of the following methods will compensate for the problem that cause low gain in an rf amplifier?

1. Using rf transformers
2. Taking advantage of the interelectrode capacitance
3. Both 1 and 2 above
4. Using audio transformers

1-66. Which of the following types of feedback is usually caused by the base-to-collector interelectrode capacitance?

1. Regenerative
2. Decoupled
3. Positive
4. Negative

1-67. In an rf amplifier an unwanted signal is coupled through the base-to-collector interelectrode capacitance. This problem can be solved by providing feedback out of phase with the unwanted signal. What is this technique called?

1. Neutralization
2. Compensating
3. Decoupling
4. Swamping

THIS SPACE LEFT BLANK INTENTIONALLY.

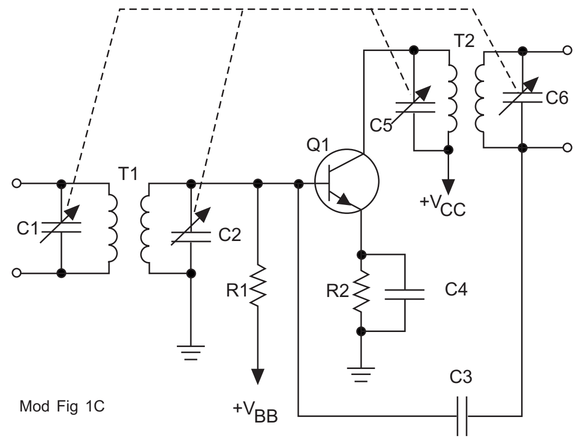


Figure 1C.—RF amplifier.

IN ANSWERING QUESTIONS 1-68 THROUGH 1-75, REFER TO FIGURE 1C.

1-68. Which of the following components is/are part of the input-signal-developing impedance for Q1?

1. C1
2. T1
3. C3
4. All of the above

1-69. What is the purpose of R1?

1. To provide swamping for the secondary of T1
2. To act as an output-signal-developing resistor
3. To provide proper bias to the base of Q1
4. To develop the signal coupled by C3

1-70. What is the purpose of R2?

1. To provide swamping for C4
2. To develop the input signal for Q1
3. To provide bias to the emitter of Q1
4. To act as the output-signal-developing resistor

1-71. If C4 were removed from the circuit, what would happen to the output?

1. It would increase
2. It would decrease
3. It would remain the same
4. It cannot be determined

1-72. Which of the following components is/are part of the load for Q1?

1. C6
2. T2
3. Both 1 and 2 above
4. C3

1-73. How many tuned parallel LC circuits are shown in the schematic?

1. One
2. Two
3. Three
4. Four

1-74. What do the dotted lines connecting C1, C2, C5, and C6 indicate?

1. The components are in a different physical location
2. That the components are "phantom" components
3. The components are variable capacitors
4. The components are ganged together

1-75. What is the purpose of C3?

1. To couple the input signal of Q1 to the secondary of T2
2. To tune the parallel LC circuit of C3, C6, and T2
3. To provide neutralization for Q1
4. To bypass R1

ASSIGNMENT 2

Textbook assignment: Chapter 3, “Special Amplifiers,” pages 3-1 through 3-70.

2-1. What is the maximum number of possible inputs in a differential amplifier?

1. One
2. Two
3. Three
4. Four

2-2. What is the maximum number of possible outputs in an differential amplifier?

1. One
2. Two
3. Three
4. Four

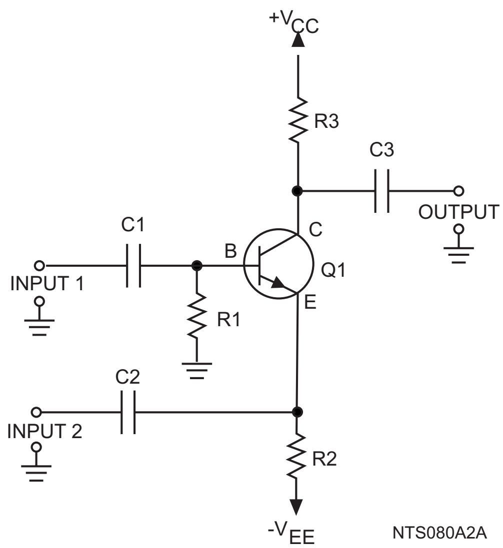


Figure 2A.—Difference amplifier.

IN ANSWERING QUESTIONS 2-3 THROUGH 2-10, REFER TO FIGURE 2A AND THE FOLLOWING INFORMATION: THE AMPLIFIER HAS A GAIN OF 6 AND THE INPUT SIGNALS ARE SINE WAVES WHICH ARE EQUAL IN AMPLITUDE AND VARY BETWEEN +2 VOLTS AND -2 VOLTS.

2-3. If the input signals are in phase with each other, what will the peak-to-peak amplitude of the output signal be?

1. 0 volts
2. 12 volts
3. 24 volts
4. 48 volts

2-4. Which of the following statements describes the output signal if the input signals are in phase with each other?

1. A sine wave in phase with the input signals
2. A sine wave 180 degrees out of phase with the input signals
3. A sine wave 90 degrees out of phase with the input signals
4. Not a sine wave

2-5. If the input signals are 180 degrees out of phase with each other, what will the peak-to-peak amplitude of the output signal be?

1. 0 volts
2. 12 volts
3. 24 volts
4. 48 volts

2-6. Which of the following statements describes the output signal if the input signals are 180 degrees out of phase with each other?

1. A sine wave 90 degrees out of phase with each input signal
2. A sine wave in phase with input signal number one
3. A sine wave in phase with input signal number two
4. Not a sine wave

2-7. If input signal number one is 90 degrees out of phase with input signal number two, what will the peak-to-peak amplitude of the output signal be?

1. 0 volts
2. 12 volts
3. 24 volts
4. 48 volts

2-8. Which of the following statements describes the output signal if input signal number one is 90 degrees out of phase with input signal number two?

1. A sine wave 90 degrees out of phase with input number two
2. A sine wave 180 degrees out of phase with input number one
3. A sine wave 180 degrees out of phase with input number two
4. Not a sine wave

2-9. If input number two is the only input signal applied to the amplifier, what will the peak-to-peak amplitude of the output signal be?

1. 0 volts
2. 12 volts
3. 24 volts
4. 48 volts

2-10. Which of the following statements describes the output signal if input number two is the only input signal applied to the amplifier?

1. A sine wave in phase with the input signal
2. A sine wave 90 degrees out of phase with the input signal
3. A sine wave 180 degrees out of phase with the input signal
4. Not a sine wave

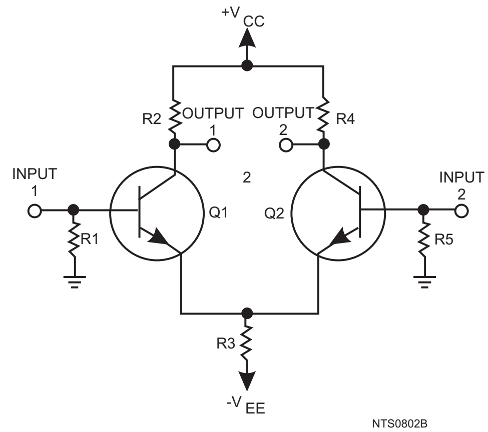


Figure 2B.—Differential amplifier.

IN ANSWERING QUESTIONS 2-11 THROUGH 2-24, REFER TO FIGURE 2B AND THE FOLLOWING INFORMATION: ALL INPUT SIGNALS ARE SINE WAVES WITH A PEAK-TO-PEAK AMPLITUDE OF 5 MILLIVOLTS. THE GAIN OF THE AMPLIFIER IS 100.

2-11. If input number one and output number one are the only terminals used, what will the peak-to-peak amplitude of the output signal be?

1. 1 volt
2. 2 volts
3. 500 millivolts
4. 0 volt

2-12. Which of the following statements describes the output signal if input number one, and output number one are the only terminals used?

1. A sine wave in phase with the input signal
2. A sine wave 90 degrees out of phase with the input signal
3. A sine wave 180 degrees out of phase with the input signal
4. Not a sine wave

2-13. If input number one is the only input and both output terminals are used, what will the peak-to-peak amplitude of each output signal be? (Assume base of Q2 grounded)

1. 1 volt
2. 2 volts
3. 500 millivolts
4. 0 volts

2-14. If input one is the only input used and an output signal is taken between output number one and output number two, what will the peak-to-peak amplitude of the output signal be? (Assume base of Q2 grounded)

1. 1 volt
2. 2 volts
3. 500 millivolts
4. 0 volts

2-15. Which of the following statements describes the output signal if input one is the only input used and an output signal is taken between output number one and output number two?

1. A sine wave twice the amplitude of output number two
2. A sine wave 90 degrees out of phase with the input signal
3. A sine wave one-half the amplitude of output number one
4. Not a sine wave

2-16. If the input signals are in phase with each other what will the amplitude of each output signal be?

1. 1 volt
2. 2 volts
3. 500 millivolts
4. 0 volts

2-17. If the input signals are in phase with each other and an output signal is taken between the two output terminals, what will the amplitude of the output signal be?

1. 1 volt
2. 2 volts
3. 500 millivolts
4. 0 volts

2-18. Which of the following statements describes output signal number one if the input signals are in phase with each other?

1. A sine wave in phase with input signal number one
2. A sine wave in phase with input signal number two
3. A sine wave 90 degrees out of phase with input signal number one
4. Not a sine wave

2-19. If the input signals are 180 degrees out of phase with each other, what will the peak-to-peak output of each output signal be?

1. 1 volt
2. 2 volts
3. 500 millivolts
4. 0 volts

2-20. If the input signals are 180 degrees out of phase with each other and the output signal is taken between the two output terminals, what will the peak-to-peak amplitude of the output signal be?

1. 1 volt
2. 2 volts
3. 500 millivolts
4. 0 volts

2-21. Which of the following statements describes output signal number one if the input signals are 180 degrees out of phase with each other?

1. A sine wave in phase with input signal number one
2. A sine wave in phase with input signal number two
3. A sine wave in phase with output signal number two
4. Not a sine wave

2-22. Which of the following statements describes the output signal if the input signals are 180 degrees out of phase with each other and the output signal is taken between the output terminals?

1. A sine wave
2. Not a sine wave
3. A sine wave 90 degrees out of phase with input signal number one
4. A sine wave in phase with input signals number one and two

2-23. If the input amplitudes are increased to 15 millivolts and are 180 degrees out of phase, what will be the peak-to-peak amplitude of the combined output?

1. 1 volt
2. 2 volts
3. 3 volts
4. 1.5 volts

2-24. What will be the peak-to-peak amplitude of the combined output if the inputs are 6 millivolts peak-to-peak, 180 degrees out of phase, and the gain is 20?

1. 2 volts
2. 2.4 volts
3. 0.12 volts
4. 0.24 volts

2-25. Which of the following is NOT a requirement for an operational amplifier?

1. Very high gain
2. Very high input impedance
3. Very high output impedance
4. Very low output impedance

2-26. Which of the following types of components are used in most operational amplifiers?

1. Transistor circuits
2. Electron tube circuits
3. Both 1 and 2 above
4. Integrated circuits

THIS SPACE LEFT BLANK INTENTIONALLY.

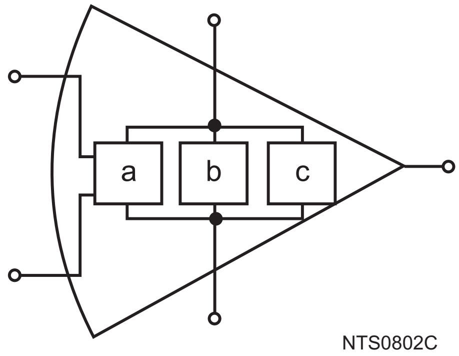


Figure 2C.—Operational Amplifier.

IN ANSWERING QUESTIONS 2-27 THROUGH 2-29, REFER TO FIGURE 2C.

2-27. What is part A of the figure?

1. An input amplifier
2. A power amplifier
3. A voltage amplifier
4. A differential amplifier

2-28. What is part B of the figure?

1. A differential amplifier
2. A voltage amplifier
3. A power amplifier
4. A video amplifier

2-29. What is part C of the figure?

1. An output amplifier
2. A voltage amplifier
3. A differential amplifier
4. A high-impedance amplifier

2-30. If degenerative feedback is used in an operational-amplifier circuit, which of the following terms describes the circuit configuration?

1. Open loop
2. Closed loop
3. Full circle
4. Neutralized

2-31. Which of the following signals determines the stability of the output signal from an operational-amplifier circuit in which degenerative feedback is used?

1. The input signal only
2. The feedback signal only
3. Both 1 and 2 above
4. The detected signal

THIS SPACE LEFT BLANK INTENTIONALLY.

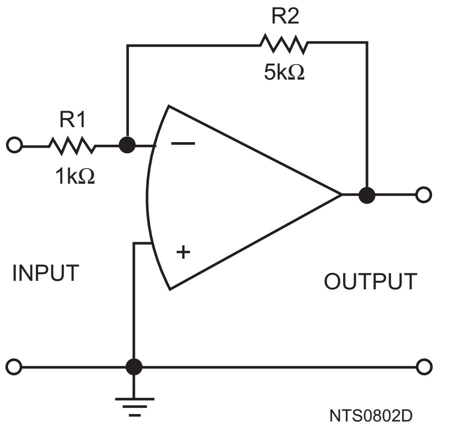


Figure 2D.—Inverting Configuration.

IN ANSWERING QUESTIONS 2-32 THROUGH 2-36, REFER TO FIGURE 2D.

2-32. In the inverting configuration of an operational-amplifier circuit where are the (a) input signal and (b) feedback signal applied?

1. (a) Inverting input
   (b) Inverting input
2. (a) Inverting input
   (b) Noninverting input
3. (a) Noninverting input
   (b) Inverting input
4. (a) Noninverting input
   (b) Noninverting input

2-33. In the inverting configuration of an operational-amplifier circuit with feedback applied and a 1-volt, peak-to-peak, sine wave as an input signal, what is the amplitude of the signal at the inverting input of the operational amplifier?

1. 1 volt
2. 2 volts
3. 10 volts
4. 0 volts

2-34. In the inverting configuration of an operational-amplifier circuit, when the noninverting input of the operational amplifier is grounded, what is the term that describes the potential at the inverting input of the operational amplifier?

1. Feedback-signal voltage
2. Input-signal voltage
3. Signal ground
4. Virtual ground

2-35. If the amplitude of the input signal to the circuit is +2 millivolts, what will the amplitude of the output signal be?

1. -10 mV
2. -2 mV
3. +10 mv
4. +2 mV

2-36. If the unity gain point of the operational amplifier is 1 mega-hertz, what is the bandwidth of the circuit?

1. 100 kHz
2. 200 kHz
3. 300 kHz
4. 400 kHz

THIS SPACE LEFT BLANK INTENTIONALLY.

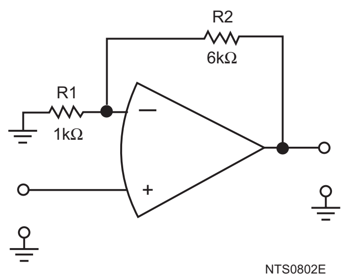


Figure 2E.—Noninverting Configuration.

IN ANSWERING QUESTIONS 2-37 AND 2-38, REFER TO FIGURE 2E.

2-37. If the amplitude of the input signal is +10 millivolts, what is the amplitude of the output signal?

1. +70 mV
2. +60 mV
3. -70 mV
4. -60 mV

2-38. The open-loop gain of the operational amplifier is 100,000 and the open-loop bandwidth is 10 hertz. If we make it a closed-loop with a gain of 10, what is the bandwidth of the circuit?

1. 100 kHz
2. 350 kHz
3. 500 kHz
4. 583 kHz

2-39. Which of the following is a difference between a summing amplifier and an adder circuit?

1. The amount of gain
2. The number of inputs
3. The type of operational amplifier
4. The placement of resistors in the circuit

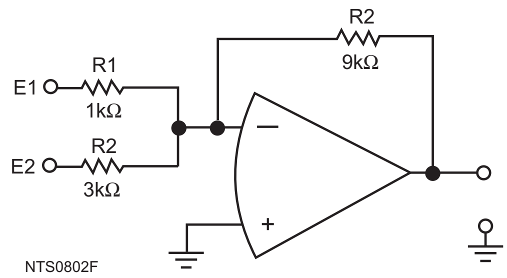


Figure 2F.—Scaling Amplifier.

IN ANSWERING QUESTIONS 2-40 AND 2-41, REFER TO FIGURE 2F.

2-40. THIS QUESTION HAS BEEN DELETED.

2-41. If the amplitude of the signal at E1 is +3 volts and the amplitude of the signal at E2 is +4 volts, what is the amplitude of the output signal?

1. +39 volts
2. +45 volts
3. –39 volts
4. –45 volts

2-42. If the amplitude of the signal at E1 is +5 volts and the amplitude of the signal at E2 is +2 volts, what is the amplitude of the signal at the inverting (–) input of the operational amplifier?

1. 0 volts
2. +7 volts
3. +21 volts
4. +54 volts

2-43. Which of the following is a difference between a difference amplifier and a subtractor?

1. The amount of gain
2. The number of inputs
3. The type of operational amplifier
4. The placement of resistors in the circuit

2-44. How many inputs can a (a) difference amplifier and (b) summing amplifier have?

1. (a) Two only
   (b) Two only
2. (a) Two only
   (b) More than two
3. (a) More than two
   (b) Two only
4. (a) More than two
   (b) More than two

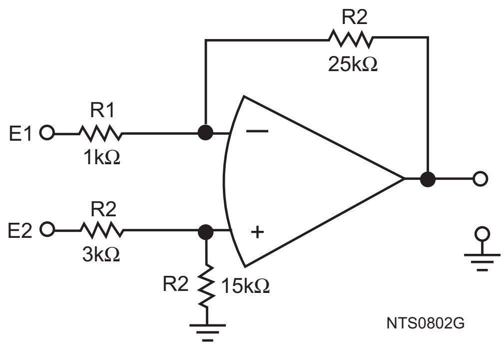


Figure 2G.—Difference Amplifier.

IN ANSWERING QUESTIONS 2-45 THROUGH 2-48, REFER TO FIGURE 2G.

2-45. THIS QUESTION HAS BEEN DELETED.

2-46. If the amplitude of the signal at E1 is +3 volts and the amplitude of the signal at E2 is +9 volts, what will the amplitude of the output signal be?

1. +6 volts
2. +12 volts
3. +30 volts
4. +60 volts

2-47. If the amplitude of the signal at E1 is +6 volts and the amplitude of the signal at E2 is +2 volts, what will the amplitude of the output signal be?

1. +8 volts
2. +20 volts
3. −20 volts
4. −40 volts

2-48. The gain of the operational amplifier shown in figure 2G can be determined by using the ratio of resistance. Which of the following ratios is correct for determining this gain.

1. $\frac{R1}{R2}=\frac{R3}{R4}$
2. $\frac{R1}{R4}=\frac{R2}{R3}$
3. $\frac{R3}{R2}=\frac{R4}{R1}$
4. $\frac{R3}{R1}=\frac{R4}{R2}$

2-49. A magnetic amplifier can be classified as which of the following types of amplifier?

1. RF amplifier
2. Audio amplifier
3. Video amplifier
4. Voltage amplifier

2-50. Which of the following statements describes the basic operating principle of a magnetic amplifier?

1. Any power amplifier will create a magnetic field which can be used to increase the gain of the power amplifier
2. The inductance of an air-core inductor will change as the power used by the load changes
3. A changing inductance can be used to control the current in a load
4. Magnetism can be increased (amplified) by changing the voltage amplitude

2-51. What happens to the true power in a series LR circuit if the inductance is decreased?

1. It increases
2. It decreases
3. It remains the same
4. It increases initially and then decreases rapidly

2-52. If the permeability of the core of a coil decreases, what happens to the (a) inductance and (b) true power in the circuit?

1. (a) Increases (b) increases
2. (a) Increases (b) decreases
3. (a) Decreases (b) increases
4. (a) Decreases (b) decreases

2-53. If the current in an iron-core coil is increased to a large value (from the operating point) what happens to the permeability of the core?

1. It increases
2. It decreases
3. It remains the same
4. It increases initially and then decreases rapidly

2-54. If two coils are wound on a single iron core, a change in current in one coil (a) will or will not cause a change in inductance and (b) will or will not cause a change in current in the other coil.

1. (a) Will (b) will
2. (a) Will (b) will not
3. (a) Will not (b) will
4. (a) Will not (b) will not

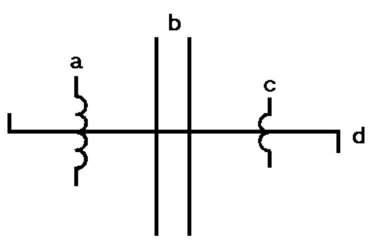


Figure 2H.—Saturable-core reactor.

IN ANSWERING QUESTION 2-55, REFER TO FIGURE 2H.

2-55. What portion of the schematic diagram indicates a saturable core?

1. a
2. b
3. c
4. d

2-56. A magnetic amplifier should be operated on what portion of the magnetization curve?

1. The positive peak
2. The negative peak
3. The mid-point
4. The knee

2-57. A toroidal core is used in a saturable-core reactor to counteract which of the following effects?

1. Hysteresis
2. Copper loss
3. Both 1 and 2 above
4. The effect of load flux on control flux

2-58. Why is a rectifier used in a magnetic amplifier?

1. To decrease current
2. To eliminate hysteresis loss
3. To increase the power-handling capability
4. To convert the magnetic amplifier from an a.c. device to a d.c. device.

2-59. What can be used to set a magnetic amplifier to the proper operating point and leave the control winding free to accept input signals?

1. A filter
2. A bias winding
3. A d.c. power source
4. A feedback network

2-60. A magnetic amplifier would not be used in which of the following devices?

1. A servo system
2. A d.c. power supply
3. Temperature indicators
4. A wideband audio power amplifier system

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 9—Introduction to Wave Generation and Wave Shaping

NAVEDTRA 14181

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its references to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 9 of a series.

History of the course:

Sep 1998: Original edition released.
May 2003: Administrative update released. No significant content changes were made.

NAVSUP Logistics Tracking Number
0504-LP-026-8340

TABLE OF CONTENTS

ASSIGNMENT QUESTIONS follow Index.

CHAPTER 1

TUNED CIRCUITS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the OCC/ECC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completion of this chapter, you will be able to:

1. State the applications of a resonant circuit.
2. Identify the conditions that exist in a resonant circuit.
3. State and apply the formula for resonant frequency of an a.c. circuit.
4. State the effect of changes in inductance (L) and capacitance (C) on resonant frequency ($_{fr}$).
5. Identify the characteristics peculiar to a series resonant circuit.
6. Identify the characteristics peculiar to a parallel resonant circuit.
7. State and apply the formula for Q.
8. State what is meant by the bandwidth of a resonant circuit and compute the bandwidth for a given circuit.
9. Identify the four general types of filters.
10. Identify how the series- and parallel-resonant circuit can be used as a bandpass or a band-reject filter.

INTRODUCTION TO TUNED CIRCUITS

When your radio or television set is turned on, many events take place within the "receiver" before you hear the sound or see the picture being sent by the transmitting station.

Many different signals reach the antenna of a radio receiver at the same time. To select a station, the listener adjusts the tuning dial on the radio receiver until the desired station is heard. Within the radio or TV receiver, the actual "selecting" of the desired signal and the rejecting of the unwanted signals are accomplished by what is called a TUNED CIRCUIT. A tuned circuit consists of a coil and a capacitor connected in series or parallel. Later in this chapter you will see the application and advantages of both series- and parallel-tuned circuits. Whenever the characteristics of inductance and capacitance are found in a tuned circuit, the phenomenon as RESONANCE takes place.

You learned earlier in the *Navy Electricity and Electronics Training Series, Module 2*, chapter 4, that inductive reactance (X_L) and capacitive reactance (X_C) have opposite effects on circuit impedance (Z).

You also learned that if the frequency applied to an LCR circuit causes X_L and X_C to be equal, the circuit is RESONANT.

If you realize that X_L and X_C can be equal ONLY at ONE FREQUENCY (the resonant frequency), then you will have learned the most important single fact about resonant circuits. This fact is the principle that enables tuned circuits in the radio receiver to select one particular frequency and reject all others. This is the reason why so much emphasis is placed on X_L and X_C in the discussions that follow.

Examine figure 1-1. Notice that a basic tuned circuit consists of a coil and a capacitor, connected either in series, view (A), or in parallel, view (B). The resistance (R) in the circuit is usually limited to the inherent resistance of the components (particularly the resistance of the coil). For our purposes we are going to disregard this small resistance in future diagrams and explanations.

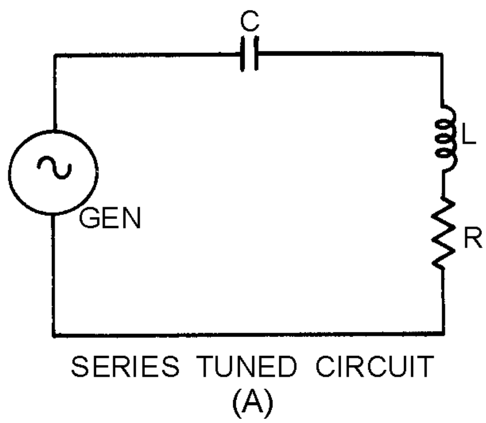


Figure 1-1A.—Basic tuned circuits. SERIES TUNED CIRCUIT

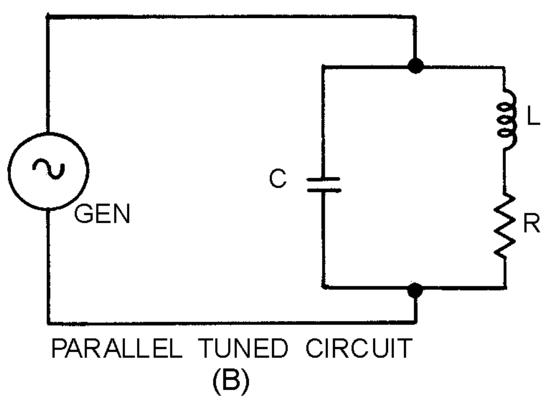


Figure 1-1B.—Basic tuned circuits. PARALLEL TUNED CIRCUIT

You have already learned how a coil and a capacitor in an a.c. circuit perform. This action will be the basis of the following discussion about tuned circuits.

Why should you study tuned circuits? Because the tuned circuit that has been described above is used in just about every electronic device, from remote-controlled model airplanes to the most sophisticated space satellite.

You can assume, if you are going to be involved in electricity or electronics, that you will need to have a good working knowledge of tuned circuits and how they are used in electronic and electrical circuits.

REVIEW OF SERIES/PARALLEL A.C. CIRCUITS

First we will review the effects of frequency on a circuit which contains resistance, inductance, and capacitance. This review recaps what you previously learned in the Inductive and Capacitive Reactance chapter in *module 2* of the *NEETS*.

FREQUENCY EFFECTS ON RLC CIRCUITS

Perhaps the most often used control of a radio or television set is the station or channel selector. Of course, the volume, tone, and picture quality controls are adjusted to suit the individual's taste, but very often they are not adjusted when the station is changed. What goes on behind this station selecting? In this chapter, you will learn the basic principles that account for the ability of circuits to "tune" to the desired station.

Effect of Frequency on Inductive Reactance

In an a.c. circuit, an inductor produces inductive reactance which causes the current to lag the voltage by 90 degrees. Because the inductor "reacts" to a changing current, it is known as a reactive component. The opposition that an inductor presents to a.c. is called inductive reactance (X_L). This opposition is caused by the inductor "reacting" to the changing current of the a.c. source. Both the inductance and the frequency determine the magnitude of this reactance. This relationship is stated by the formula:

$$X_L = 2\pi fL$$

Where:

X_L = the inductive reactance in ohms
f = the frequency in hertz
L = the inductance in henries
π = 3.1416

As shown in the equation, any increase in frequency, or "f," will cause a corresponding increase of inductive reactance, or "X_L." Therefore, the INDUCTIVE REACTANCE VARIES DIRECTLY WITH THE FREQUENCY. As you can see, the higher the frequency, the greater the inductive reactance; the lower the frequency, the less the inductive reactance for a given inductor. This relationship is illustrated in figure 1-2. Increasing values of X_L are plotted in terms of increasing frequency. Starting at the lower left corner with zero frequency, the inductive reactance is zero. As the frequency is increased (reading to the right), the inductive reactance is shown to increase in direct proportion.

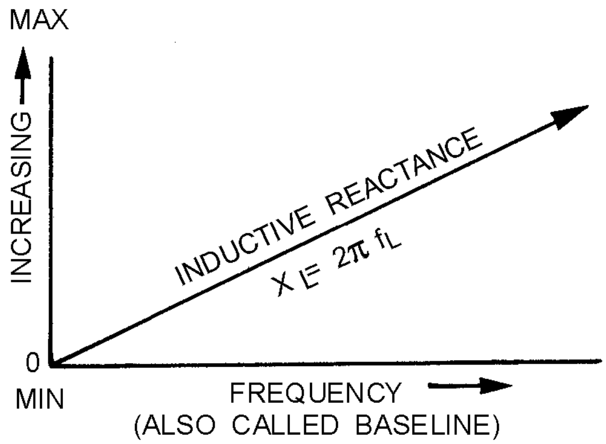


Figure 1-2.—Effect of frequency on inductive reactance.

Effect of Frequency on Capacitive Reactance

In an a.c. circuit, a capacitor produces a reactance which causes the current to lead the voltage by 90 degrees. Because the capacitor "reacts" to a changing voltage, it is known as a reactive component. The opposition a capacitor presents to a.c. is called capacitive reactance (X_C). The opposition is caused by the capacitor "reacting" to the changing voltage of the a.c. source. The formula for capacitive reactance is:

$$X_C = \frac{1}{2\pi fC}$$

Where:

X_C = the capacitive reactance in ohms
f = the frequency in hertz
C = the capacitance in farads
π = 3.1416

In contrast to the inductive reactance, this equation indicates that the CAPACITIVE REACTANCE VARIES INVERSELY WITH THE FREQUENCY. When $f = 0$, X_C is infinite (∞) and decreases as frequency increases. That is, the lower the frequency, the greater the capacitive reactance; the higher the frequency, the less the reactance for a given capacitor.

As shown in figure 1-3, the effect of capacitance is opposite to that of inductance. Remember, capacitance causes the current to lead the voltage by 90 degrees, while inductance causes the current to lag the voltage by 90 degrees.

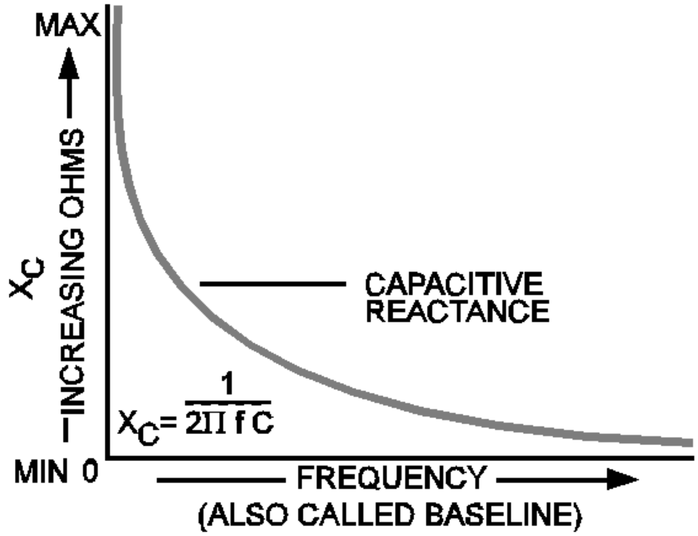


Figure 1-3.—Effect of frequency on capacitive reactance.

Effect of Frequency on Resistance

In the expression for inductive reactance, $X_L = 2\pi fL$, and in the expression for capacitive reactance,

$$X_C = \frac{1}{2\pi fC}$$

both contain "f" (frequency). Any change of frequency changes the reactance of the circuit components as already explained. So far, nothing has been said about the effect of frequency on resistance. In an Ohm's law relationship, such as R = E/I no "f" is involved. Thus, for all practical purposes, a change of frequency does not affect the resistance of the circuit. If a 60-hertz a.c. voltage causes 20 milliamperes of current in a resistive circuit, then the same voltage at 2000 hertz, for example, would still cause 20 milliamperes to flow.

NOTE: Remember that the total opposition to a.c. is called impedance (Z). Impedance is the combination of inductive reactance (X_L), capacitive reactance (X_C), and resistance (R). When dealing with a.c. circuits, the impedance is the factor with which you will ultimately be concerned. But, as you have just been shown, the resistance (R) is not affected by frequency. Therefore, the remainder of the discussion of a.c. circuits will only be concerned with the reactance of inductors and capacitors and will ignore resistance.

A.c. Circuits Containing Both Inductive and Capacitive Reactances

A.c. circuits that contain both an inductor and a capacitor have interesting characteristics because of the opposing effects of L and C. X_L and X_C may be treated as reactors which are 180 degrees out of phase. As shown in figure 1-2, the vector for X_L should be plotted above the baseline; vector for X_C, figure 1-3, should be plotted below the baseline. In a series circuit, the effective reactance, or what is termed the RESULTANT REACTANCE, is the difference between the individual reactances. As an equation, the resultant reactance is:

$$X = X_L - X_C$$

Suppose an a.c. circuit contains an X_L of 300 ohms and an X_C of 250 ohms. The resultant reactance is:

$$X = X_L - X_C = 300 - 250 = 50 \text{ ohms (inductive)}$$

In some cases, the X_C may be larger than the X_L. If $X_L = 1200$ ohms and $X_C = 4000$ ohms, the difference is: $X = X_L - X_C = 1200 - 4000 = -2800$ ohms (capacitive). The total carries the sign (+ or –) of the greater number (factor).

Q-1. What is the relationship between frequency and the values of (a) X_L, (b) X_C, and (c) R?

Q-2. In an a.c. circuit that contains both an inductor and a capacitor, what term is used for the difference between the individual reactances?

RESONANCE

For every combination of L and C, there is only ONE frequency (in both series and parallel circuits) that causes X_L to exactly equal X_C; this frequency is known as the RESONANT FREQUENCY. When the resonant frequency is fed to a series or parallel circuit, X_L becomes equal to X_C, and the circuit is said to be RESONANT to that frequency. The circuit is now called a RESONANT CIRCUIT; resonant circuits are tuned circuits. The circuit condition wherein X_L becomes equal to X_C is known as RESONANCE.

Each LCR circuit responds to resonant frequency differently than it does to any other frequency. Because of this, an LCR circuit has the ability to separate frequencies. For example, suppose the TV or radio station you want to see or hear is broadcasting at the resonant frequency. The LC "tuner" in your set can divide the frequencies, picking out the resonant frequency and rejecting the other frequencies. Thus, the tuner selects the station you want and rejects all other stations. If you decide to select another station, you can change the frequency by tuning the resonant circuit to the desired frequency.

RESONANT FREQUENCY

As stated before, the frequency at which X_L equals X_C (in a given circuit) is known as the resonant frequency of that circuit. Based on this, the following formula has been derived to find the exact resonant frequency when the values of circuit components are known:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

There are two important points to remember about this formula. First, the resonant frequency found when using the formula will cause the reactances (X_L and X_C) of the L and C components to be equal. Second, any change in the value of either L or C will cause a change in the resonant frequency.

An <u>increase</u> in the value of either L or C, or both L and C, will <u>lower</u> the resonant frequency of a given circuit. A <u>decrease</u> in the value of L or C, or both L and C, will <u>raise</u> the resonant frequency of a given circuit.

The symbol for resonant frequency used in this text is f. Different texts and references may use other symbols for resonant frequency, such as f_o, F_r, and fR. The symbols for many circuit parameters have been standardized while others have been left to the discretion of the writer. When you study, apply the rules given by the writer of the text or reference; by doing so, you should have no trouble with nonstandard symbols and designations.

The resonant frequency formula in this text is:

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

Where:

f_r = the resonant frequency in hertz
L = the inductance in henries
C = the capacitance in farads
π = 3.1416

By substituting the constant .159 for the quantity

$$\frac{1}{2\pi}$$

the formula can be simplified to the following:

$$f_r = \frac{.159}{\sqrt{LC}}$$

Let's use this formula to figure the resonant frequency (f_r). The circuit is shown in the practice tank circuit of figure 1-4.

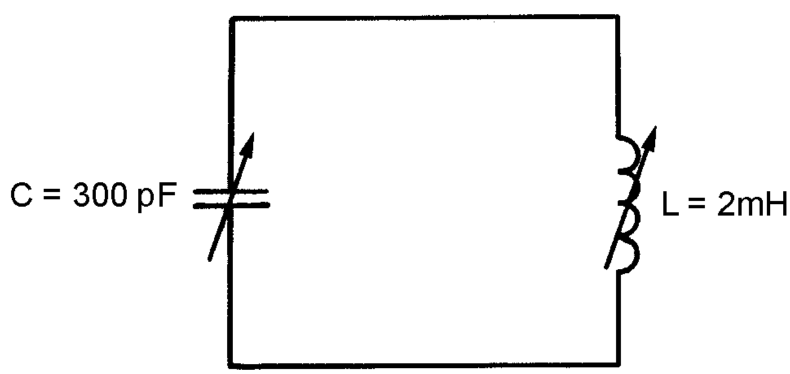


Figure 1-4.—Practice tank circuit.

Given:

$L = 2\text{mH}\ (2 \times 10^{-3}\ \text{H})$

$C = 300\text{pF}\ (300 \times 10^{-12}\ \text{F})$

Solution:

$$f_r = \frac{.159}{\sqrt{LC}}$$

$$f_r = \frac{.159}{\sqrt{(2 \times 10^{-3}\ \text{H}) \times (300 \times 10^{-12}\ \text{F})}}$$

$$f_r = \frac{.159}{\sqrt{600 \times 10^{-15}}}$$

(F and H are shown in this step to show units)

$$f_r = \frac{.159}{60 \times 10^{-14}}$$

(Solving for square root $60 = 7.75 \quad 10^{-14} = 10^{-7}$)

$$f_r = \frac{.159}{7.75 \times 10^{-7}}$$

$$f_r = \frac{.159 \times 10^{7}}{7.75}$$

$$f_r = \frac{.159 \times 10^{4}}{7.75}$$

$f_r = 20.5 \times 10^{4}$ (rounded off)

$f_r = 205{,}000$ Hz or 205 kHz

The important point here is not the formula nor the mathematics. In fact, you may never have to compute a resonant frequency. The important point is for you to see that any given combination of L and C can be resonant at only one frequency; in this case, 205 kHz.

The universal reactance curves of figures 1-2 and 1-3 are joined in figure 1-5 to show the relative values of X_L and X_L at resonance, below resonance, and above resonance.

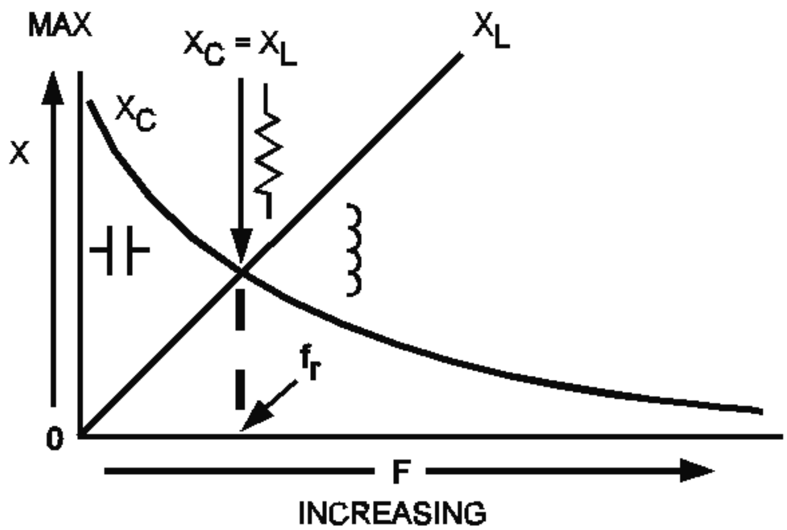


Figure 1-5.—Relationship between X_L and X_C as frequency increases.

First, note that f_r, (the resonant frequency) is that frequency (or point) where the two curves cross. At this point, and ONLY this point, X_L equals X_C. Therefore, the frequency indicated by f_r is the one and only frequency of resonance. Note the resistance symbol which indicates that at resonance all reactance is cancelled and the circuit impedance is effectively purely resistive. Remember, a.c. circuits that are resistive have no phase shift between voltage and current. Therefore, at resonance, phase shift is cancelled. The phase angle is effectively zero.

Second, look at the area of the curves to the left of f_r. This area shows the relative reactances of the circuit at frequencies BELOW resonance. To these LOWER frequencies, X_C will always be greater than X_L. There will always be some capacitive reactance left in the circuit after all inductive reactance has been cancelled. Because the impedance has a reactive component, there will be a phase shift. We can also state that below f_r the circuit will appear capacitive.

Lastly, look at the area of the curves to the right of f. This area shows the relative reactances of the circuit at frequencies ABOVE resonance. To these HIGHER frequencies, X_L will always be greater than X_C. There will always be some inductive reactance left in the circuit after all capacitive reactance has been cancelled. The inductor symbol shows that to these higher frequencies, the circuit will always appear to have some inductance. Because of this, there will be a phase shift.

RESONANT CIRCUITS

Resonant circuits may be designed as series resonant or parallel resonant. Each has the ability to discriminate between its resonant frequency and all other frequencies. How this is accomplished by both series- and parallel-LC circuits is the subject of the next section.

NOTE: Practical circuits are often more complex and difficult to understand than simplified versions. Simplified versions contain all of the basic features of a practical circuit, but leave out the nonessential features. For this reason, we will first look at the IDEAL SERIES-RESONANT CIRCUIT—a circuit that really doesn't exist except for our purposes here.

THE IDEAL SERIES-RESONANT CIRCUIT

The ideal series-resonant circuit contains no resistance; it consists of only inductance and capacitance in series with each other and with the source voltage. In this respect, it has the same characteristics of the series circuits you have studied previously. Remember that current is the same in all parts of a series circuit because there is only one path for current.

Each LC circuit responds differently to different input frequencies. In the following paragraphs, we will analyze what happens internally in a series-LC circuit when frequencies at resonance, below resonance, and above resonance are applied. The L and C values in the circuit are those used in the problem just studied under resonant-frequency. The frequencies applied are the three inputs from figure 1-6. Note that the resonant frequency of each of these components is 205 kHz, as figured in the problem.

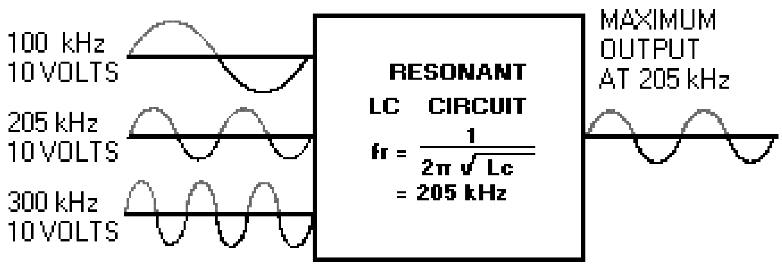


Figure 1-6.—Output of the resonant circuit.

How the Ideal Series-LC Circuit Responds to the Resonant Frequency (205 kHz)

Given:

$L = 2\text{ mH } (2 \times 10^{-3}\text{ H})$

$C = 300\text{ pF } (300 \times 10^{-12}\text{ F})$

$f_r = 205\text{ kHz (rounded off)}$

$$f_r = \frac{.159}{\sqrt{LC}}$$

$X_L = 2580\text{ ohms } (2\pi fL)$

$X_c = 2580\text{ ohms } \left(\frac{1}{2\pi FC}\right)$

$E_S = 10$ volts at a frequency 205 kHz

Note: You are given the values of X_L, X_C, and f_r but you can apply the formulas to figure them. The values given are rounded off to make it easier to analyze the circuit.

First, note that X_L and X_C are equal. This shows that the circuit is resonant to the applied frequency of 205 kHz. X_L and X_C are opposite in effect; therefore, they subtract to zero. (2580 ohms − 2580 ohms = zero.) At resonance, then, X = zero. In our theoretically perfect circuit with zero resistance and zero reactance, the total opposition to current (Z) must also be zero.

Now, apply Ohm's law for a.c. circuits:

$$I = \frac{E}{Z}$$

$$I = \frac{10 \text{ volts}}{0 \text{ ohms}}$$

$$I = \text{ INFINITELY HIGH}$$

Don't be confused by this high value of current. Our perfect, but impossible, circuit has no opposition to current. Therefore, current flow will be extremely high. The important points here are that AT RESONANCE, impedance is VERY LOW, and the resulting current will be comparatively HIGH.

If we apply Ohm's law to the individual reactances, we can figure relative values of voltage across each reactance.

$$E_L = I \times X_L$$

$$E_C = I \times X_C$$

These are reactive voltages that you have studied previously. The voltage across each reactance will be comparatively high. A comparatively high current times 2580 ohms yields a high voltage. At any given instant, this voltage will be of opposite polarity because the reactances are opposite in effect. $E_L + E_C =$ zero volts

WARNING

THE INDIVIDUAL VOLTAGES MAY REACH QUITE HIGH VALUES. ALTHOUGH LITTLE POWER IS PRESENT, THE VOLTAGE IS REAL AND CARE SHOULD BE TAKEN IN WORKING WITH IT.

Let's summarize our findings so far. In a series-LC circuit with a resonant-frequency voltage applied, the following conditions exist:

- X_L and X_C are equal and subtract to zero.
- Resultant reactance is zero ohms.
- Impedance (Z) is reduced to a MINIMUM value.
- With minimum Z, current is MAXIMUM for a given voltage.
- Maximum current causes maximum voltage drops across the individual reactances.

All of the above follow in sequence from the fact that $X_L = X_C$ at the resonant frequency.

How the Ideal Series-LC Circuit Respond to a Frequency Below Resonance (100 kHz)

Given:

$L = 2$ mH (2×10^{-3}H)

$C = 300$ pF (300×10^{-12}F)

$f_r = 205$ kHz (at resonant frequency)

$$f_r = \frac{.159}{\sqrt{LC}}$$

$X_L =$ 1260 ohms (rounded off) (at 100 kHz)

$X_c =$ 5300 ohms (rounded off) (at 100 kHz)

$E_s =$ 10 volts (at 100 kHz)

(As in the previous analysis, you are given values that are possible for you to compute. If you do the computations, remember that most values are rounded off.)

First, note that X_L and X_C are no longer equal. X_C is larger than it was at resonance; X_L is smaller. By applying the formulas you have learned, you know that a lower frequency produces a higher capacitive reactance and a lower inductive reactance. The reactances subtract but do not cancel ($X_L - X_C = 1260 - 5300 = 4040$ ohms (capacitive)). At an input frequency of 100 kHz, the circuit (still resonant to 205 kHz) has a net reactance of 4040 ohms. In our theoretically perfect circuit, the total opposition (Z) is equal to X, or 4040 ohms.

As before, let's apply Ohm's law to the new conditions.

$$I = \frac{E}{Z}$$

$$I = \frac{10 \text{ volts}}{4040 \text{ ohms}}$$

$$I = .00248 \text{ ampere}$$
(approximately 2.5 mA)

The voltage drops across the reactances are as follows:

$$E_L = I \times X_L$$

$$E_L = .0025\,\text{A} \times 1260\,\Omega$$

$$E_L = 3 \text{ volts (approximately)}$$

$$E_C = I \times X_C$$

$$E_C = .0025\,\text{A} \times 5300\,\Omega$$

$$E_C = 13 \text{ volts (approximately)}$$

In summary, in a series-LC circuit with a source voltage that is below the resonant frequency (100 kHz in the example), the resultant reactance (X), and therefore impedance, is higher than at resonance. In addition current is lower, and the voltage drops across the reactances are lower. All of the above follow in sequence due to the fact that X_C is greater than X_L at any frequency lower than the resonant frequency.

How the Ideal Series-LC Circuit Responds to a Frequency Above Resonance (300 kHz)

Given:

$$L = 2\text{ mH } (2 \times 10^{-3}\text{ H})$$

$$C = 300\text{ pF } (300 \times 10^{-12}\text{ F})$$

$$f_r = 205\text{ kHz (at resonant frequency)}$$

$$X_L = 3770\text{ ohms (rounded off) (at 300 kHz)}$$

$$X_c = 1770\text{ ohms (rounded off) (at 300 kHz)}$$

$$E_s = 10\text{ volts (at 300 kHz)}$$

Again, X_L and X_C are not equal. This time, X_L is larger than X_C. (If you don't know why, apply the formulas and review the past several pages.) The resultant reactance is 2000 ohms ($X_L - X_C = 3770 - 1770 = 2000$ ohms.) Therefore, the resultant reactance (X), or the impedance of our perfect circuit at 300 kHz, is 2000 ohms.

By applying Ohm's law as before:

$$I = 5\text{ milliamperes}$$

$$E_L = 19\text{ volts (rounded off)}$$

$$E_C = 9\text{ volts (rounded off)}$$

In summary, in a series-LC circuit with a source voltage that is above the resonant frequency (300 kHz in this example), impedance is higher than at resonance, current is lower, and the voltage drops across the reactances are lower. All of the above follow in sequence from the fact that X_L is greater than X_C at any frequency higher than the resonant frequency.

Summary of the Response of the Ideal Series-LC Circuit to Frequencies Above, Below, and at Resonance

The ideal series-resonant circuit has zero impedance. The impedance increases for frequencies higher and lower than the resonant frequency. The impedance characteristic of the ideal series-resonant circuit results because resultant reactance is zero ohms at resonance and ONLY at resonance. All other frequencies provide a resultant reactance greater than zero.

Zero impedance at resonance allows maximum current. All other frequencies have a reduced current because of the increased impedance. The voltage across the reactance is greatest at resonance because voltage drop is directly proportional to current. All discrimination between frequencies results from the fact that X_L and X_C completely counteract ONLY at the resonant frequency.

How the Typical Series-LC Circuit Differs From the Ideal

As you learned much earlier in this series, resistance is always present in practical electrical circuits; it is impossible to eliminate. A typical series-LC circuit, then, has R as well as L and C.

If our perfect (ideal) circuit has zero resistance, and a typical circuit has "some" resistance, then a circuit with a very small resistance is closer to being perfect than one that has a large resistance. Let's list what happens in a series-resonant circuit because resistance is present. This is not new to you - just a review of what you have learned previously.

In a series-resonant circuit that is basically L and C, but that contains "some" R, the following statements are true:

- X_L, X_C, and R components are all present and can be shown on a vector diagram, each at right angles with the resistance vector (baseline).

- At resonance, the resultant reactance is zero ohms. Thus, at resonance, The circuit impedance equals only the resistance (R). The circuit impedance can never be less than R because the original resistance will always be present in the circuit.

- At resonance, a practical series-RLC circuit ALWAYS has MINIMUM impedance. The actual value of impedance is that of the resistance present in the circuit ($Z = R$).

Now, if the designers do their very best (and they do) to keep the value of resistance in a practical series-RLC circuit LOW, then we can still get a fairly high current at resonance. The current is NOT "infinitely" high as in our ideal circuit, but is still higher than at any other frequency. The curve and vector relationships for the practical circuit are shown in figure 1-7.

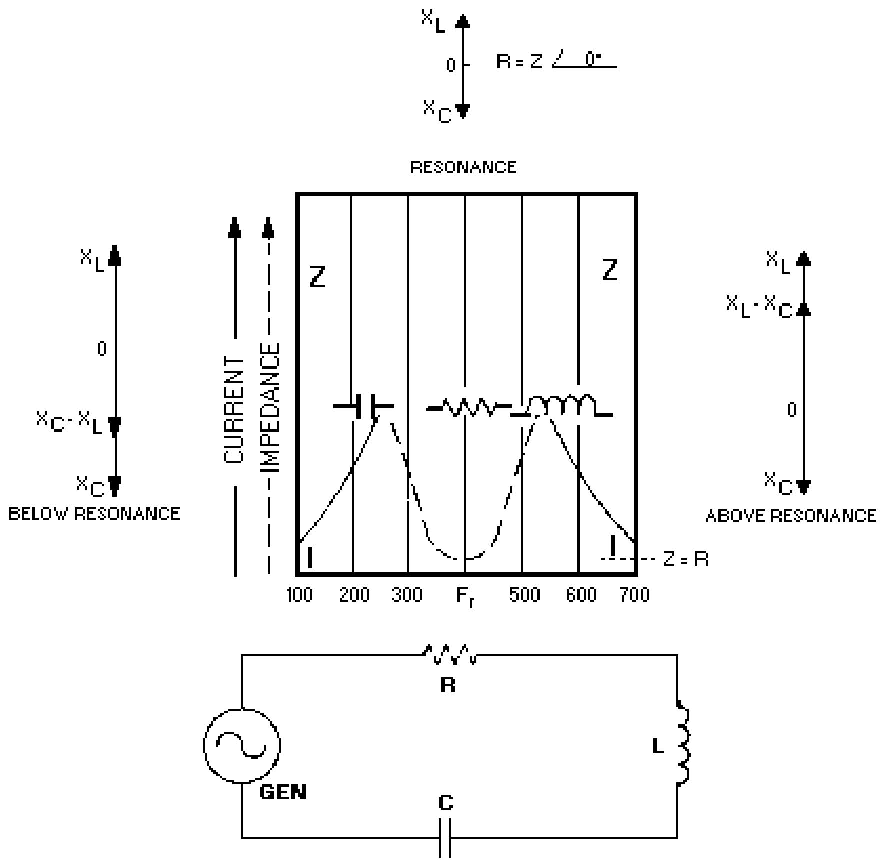


Figure 1-7.—Curves of impedance and current in an RLC series resonant circuit.

Note that the impedance curve does not reach zero at its minimum point. The vectors above and below resonance show that the phase shift of the circuit at these frequencies is less than 90 degrees because of the resistance.

The horizontal width of the curve is a measure of how well the circuit will pick out (discriminate) the one desired frequency. The width is called BANDWIDTH, and the ability to discriminate between frequencies is known as SELECTIVITY. Both of these characteristics are affected by resistance. Lower resistance allows narrower bandwidth, which is the same as saying the circuit has better selectivity. Resistance, then, is an unwanted quantity that cannot be eliminated but can be kept to a minimum by the circuit designers.

More on bandwidth, selectivity, and measuring the effects of resistance in resonant circuits will follow the discussion of parallel resonance.

Q-3. State the formula for resonant frequency.

Q-4. If the inductor and capacitor values are increased, what happens to the resonant frequency?

Q-5. In an "ideal" resonant circuit, what is the relationship between impedance and current?

Q-6. In a series-RLC circuit, what is the condition of the circuit if there is high impedance, low current, and low reactance voltages?

How the Parallel-LC Circuit Stores Energy

A parallel-LC circuit is often called a TANK CIRCUIT because it can store energy much as a tank stores liquid. It has the ability to take energy fed to it from a power source, store this energy alternately in the inductor and capacitor, and produce an output which is a continuous a.c. wave. You can understand how this is accomplished by carefully studying the sequence of events shown in figure 1-8. You must thoroughly understand the capacitor and inductor action in this figure before you proceed further in the study of parallel-resonant circuits.

In each view of figure 1-8, the waveform is of the charging and discharging CAPACITOR VOLTAGE. In view (A), the switch has been moved to position C. The d.c. voltage is applied across the capacitor, and the capacitor charges to the potential of the battery.

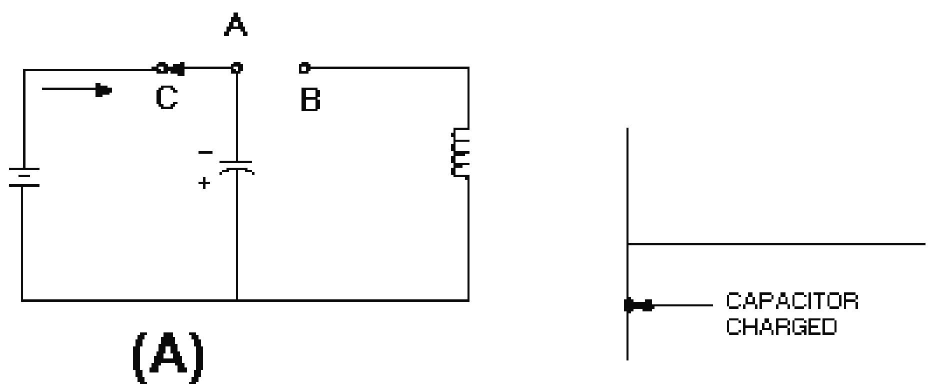


Figure 1-8A.—Capacitor and inductor action in a tank circuit.

In view (B), moving the switch to the right completes the circuit from the capacitor to the inductor and places the inductor in series with the capacitor. This furnishes a path for the excess electrons on the upper plate of the capacitor to flow to the lower plate, and thus starts neutralizing the capacitor charge. As these electrons flow through the coil, a magnetic field is built up around the coil. The energy which was first stored by the electrostatic field of the capacitor is now stored in the electromagnetic field of the inductor.

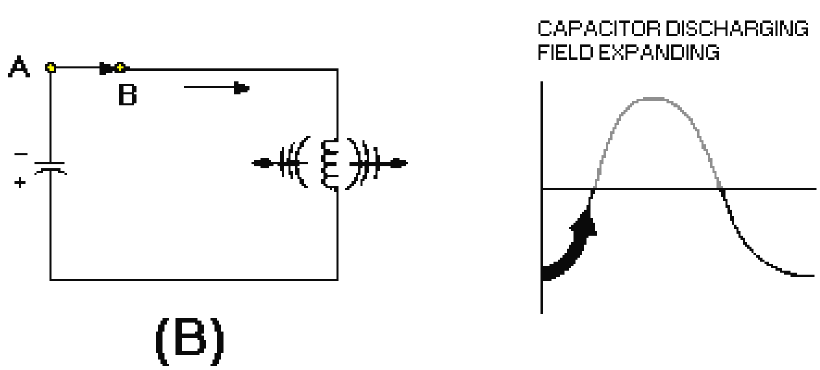


Figure 1-8B.—Capacitor and inductor action in a tank circuit.

View (C) shows the capacitor discharged and a maximum magnetic field around the coil. The energy originally stored in the capacitor is now stored entirely in the magnetic field of the coil.

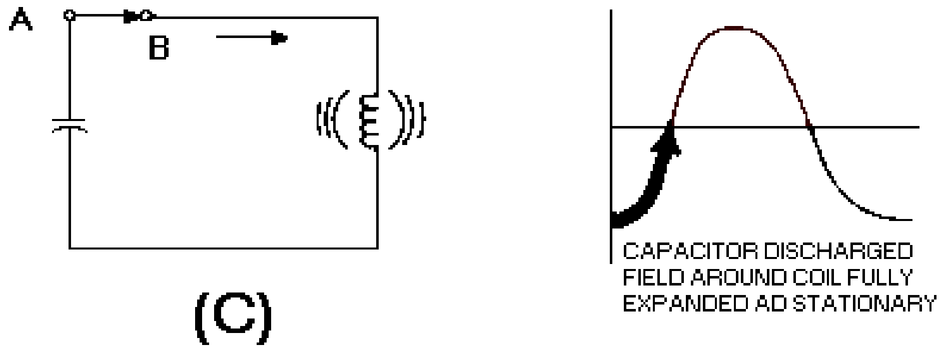


Figure 1-8C.—Capacitor and inductor action in a tank circuit.

Since the capacitor is now completely discharged, the magnetic field surrounding the coil starts to collapse. This induces a voltage in the coil which causes the current to continue flowing in the same direction and charges the capacitor again. This time the capacitor charges to the opposite polarity, view (D).

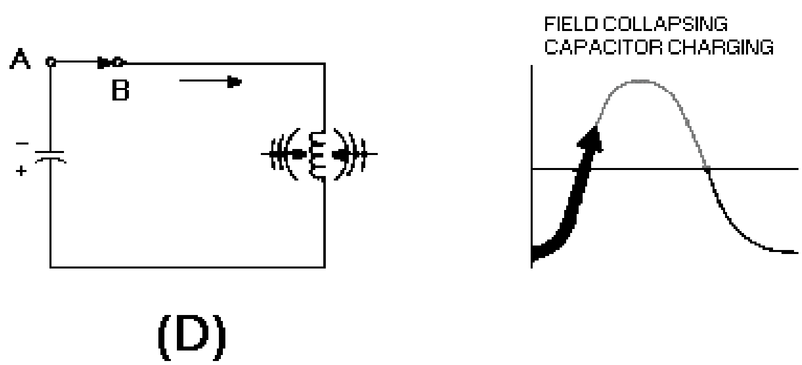


Figure 1-8D.—Capacitor and inductor action in a tank circuit.

In view (E), the magnetic field has completely collapsed, and the capacitor has become charged with the opposite polarity. All of the energy is again stored in the capacitor.

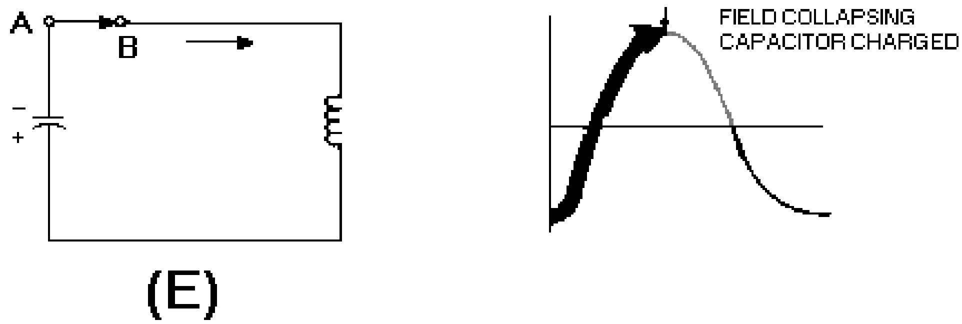


Figure 1-8E.—Capacitor and inductor action in a tank circuit.

In view (F), the capacitor now discharges back through the coil. This discharge current causes the magnetic field to build up again around the coil.

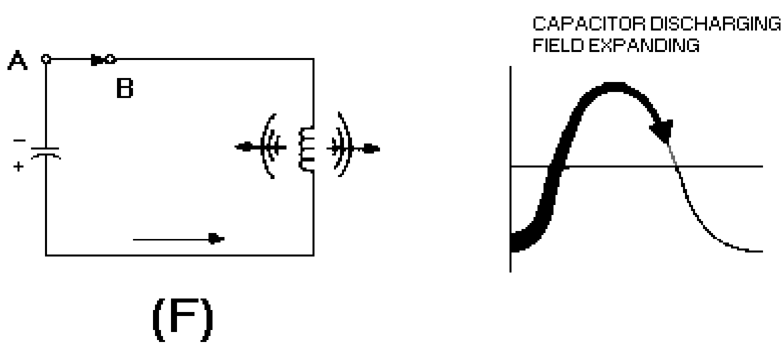


Figure 1-8F.—Capacitor and inductor action in a tank circuit.

In view (G), the capacitor is completely discharged. The magnetic field is again at maximum.

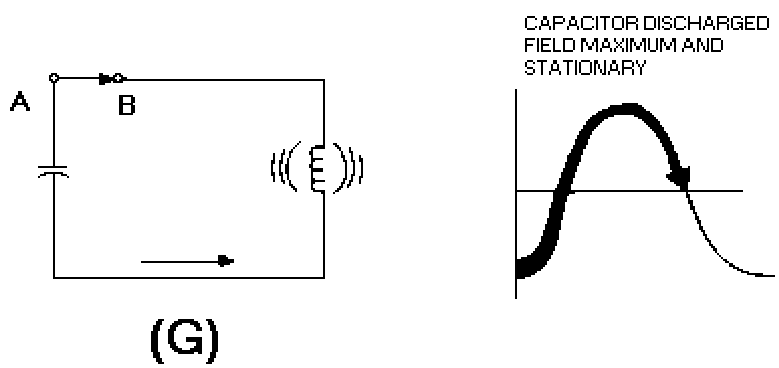


Figure 1-8G.—Capacitor and inductor action in a tank circuit.

In view (H), with the capacitor completely discharged, the magnetic field again starts collapsing. The induced voltage from the coil maintains current flowing toward the upper plate of the capacitor.

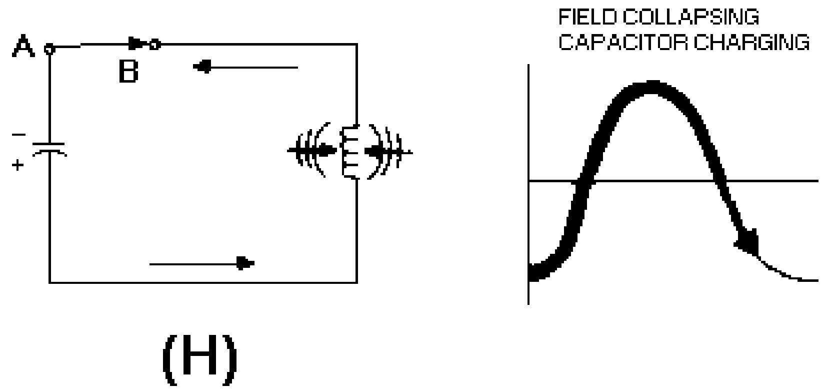


Figure 1-8H.—Capacitor and inductor action in a tank circuit.

In view (I), by the time the magnetic field has completely collapsed, the capacitor is again charged with the same polarity as it had in view (A). The energy is again stored in the capacitor, and the cycle is ready to start again.

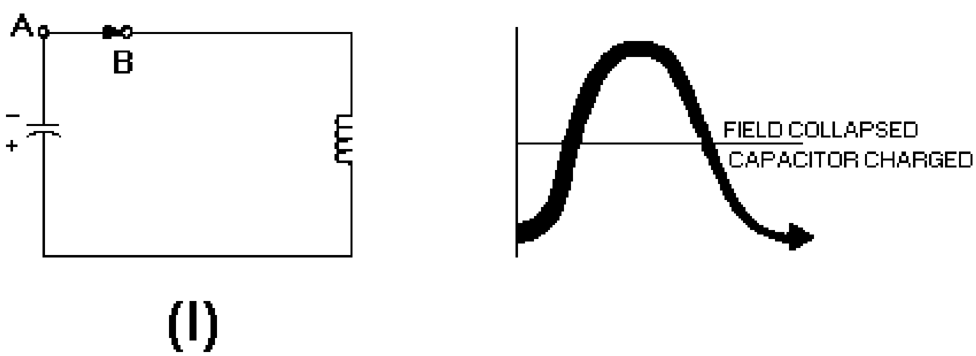


Figure 1-8I.—Capacitor and inductor action in a tank circuit.

The number of times per second that these events in figure 1-8 take place is called NATURAL FREQUENCY or RESONANT FREQUENCY of the circuit. Such a circuit is said to oscillate at its resonant frequency.

It might seem that these oscillations could go on forever. You know better, however, if you apply what you have already learned about electric circuits.

This circuit, as all others, has some resistance. Even the relatively small resistance of the coil and the connecting wires cause energy to be dissipated in the form of heat (I^2R loss). The heat loss in the circuit resistance causes the charge on the capacitor to be less for each subsequent cycle. The result is a DAMPED WAVE, as shown in figure 1-9. The charging and discharging action will continue until all of the energy has been radiated or dissipated as heat.

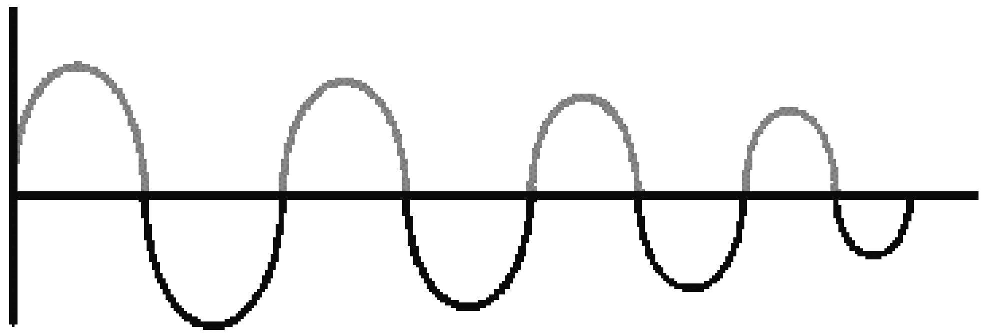


Figure 1-9.—Damped wave.

If it were possible to have a circuit with absolutely no resistance, there would be no heat loss, and the oscillations would tend to continue indefinitely. You have already learned that tuned circuits are designed to have very little resistance. Reducing I^2R losses is still another reason for having low resistance.

A "perfect" tuned circuit would produce the continuous sine wave shown in figure 1-10. Its frequency would be that of the circuit.

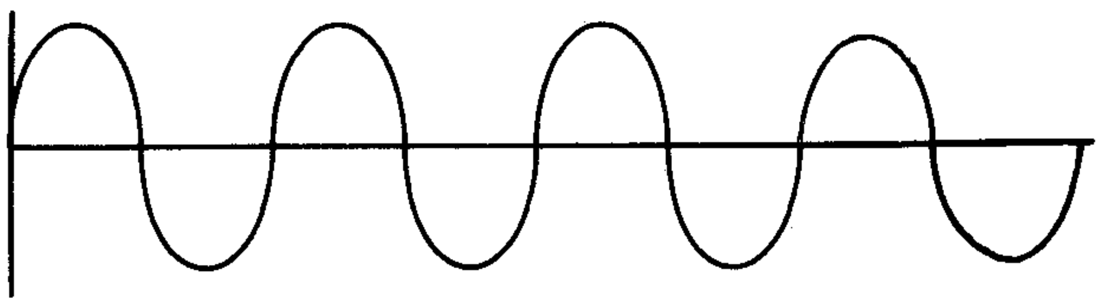


Figure 1-10.—Sine wave-resonant frequency.

Because we don't have perfection, another way of causing a circuit to oscillate indefinitely would be to apply a continuous a.c. or pulsing source to the circuit. If the source is at the resonant frequency of the circuit, the circuit will oscillate as long as the source is applied.

The reasons why the circuit in figure 1-8 oscillates at the resonant frequency have to do with the characteristics of resonant circuits. The discussion of parallel resonance will not be as detailed as that for series resonance because the idea of resonance is the same for both circuits. Certain characteristics differ as a result of L and C being in parallel rather than in series. These differences will be emphasized.

Q-7. When the capacitor is completely discharged, where is the energy of the tank circuit stored?

Q-8. When the magnetic field of the inductor is completely collapsed, where is the energy of the tank circuit stored?

PARALLEL RESONANCE

Much of what you have learned about resonance and series-LC circuits can be applied directly to parallel-LC circuits. The purpose of the two circuits is the same — to select a specific frequency and reject all others. X_L still equals X_C at resonance. Because the inductor and capacitor are in parallel, however, the circuit has the basic characteristics of an a.c. parallel circuit. The parallel hookup causes frequency selection to be accomplished in a different manner. It gives the circuit different characteristics. The first of these characteristics is the ability to store energy.

The Characteristics of a Typical Parallel-Resonant Circuit

Look at figure 1-11. In this circuit, as in other parallel circuits, the voltage is the same across the inductor and capacitor. The currents through the components vary inversely with their reactances in accordance with Ohm's law. The total current drawn by the circuit is the vector sum of the two individual component currents. Finally, these two currents, I_L and I_C, are 180 degrees out of phase because the effects of L and C are opposite. There is not a single fact new to you in the above. It is all based on what you have learned previously about parallel a.c. circuits that contain L and C.

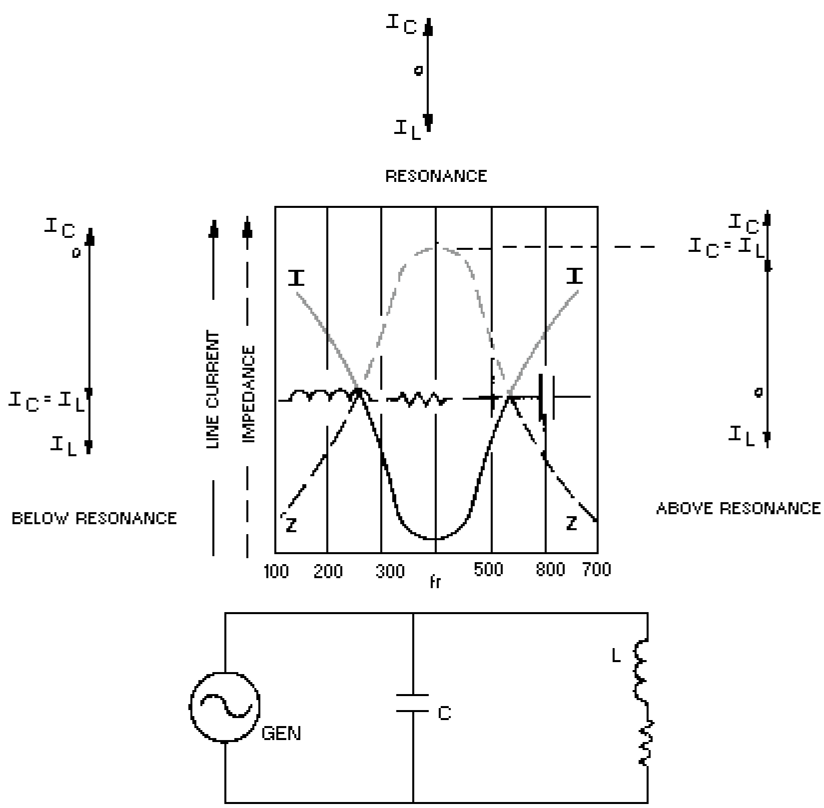


Figure 1-11.—Curves of impedance and current in an RLC parallel-resonant circuit.

Now, at resonance, X_L is still equal to X_C. Therefore, I_L must equal I_C. Remember, the voltage is the same; the reactances are equal; therefore, according to Ohm's law, the currents must be equal. But, don't forget, even though the currents are equal, they are still opposites. That is, if the current is flowing "up" in the capacitor, it is flowing "down" in the coil, and vice versa. In effect, while the one component draws current, the other returns it to the source. The net effect of this "give and take action" is that zero current is drawn from the source at resonance. The two currents yield a total current of zero amperes because they are exactly equal and opposite at resonance.

A circuit that is completed and has a voltage applied, but has zero current, must have an INFINITE IMPEDANCE (apply Ohm's law — any voltage divided by zero yields infinity).

By now you know that we have just ignored our old friend resistance from previous discussions. In an actual circuit, at resonance, the currents will not quite counteract each other because each component will have different resistance. This resistance is kept extremely low, but it is still there. The result is that a relatively small current flows from the source at resonance instead of zero current. Therefore, a basic characteristic of a practical parallel-LC circuit is that, at resonance, the circuit has MAXIMUM impedance which results in MINIMUM current from the source. This current is often called line current. This is shown by the peak of the waveform for impedance and the valley for the line current, both occurring at f_r the frequency of resonance in figure 1-11.

There is little difference between the circuit pulsed by the battery in figure 1-8 that oscillated at its resonant (or natural) frequency, and the circuit we have just discussed. The equal and opposite currents in the two components are the same as the currents that charged and discharged the capacitor through the coil.

For a given source voltage, the current oscillating between the reactive parts will be stronger at the resonant frequency of the circuit than at any other frequency. At frequencies below resonance, capacitive current will decrease; above the resonant frequency, inductive current will decrease. Therefore, the oscillating current (or circulating current, as it is sometimes called), being the lesser of the two reactive currents, will be maximum at resonance.

If you remember, the basic resonant circuit produced a "damped" wave. A steady amplitude wave was produced by giving the circuit energy that would keep it going. To do this, the energy had to be at the same frequency as the resonant frequency of the circuit.

So, if the resonant frequency is "timed" right, then all other frequencies are "out of time" and produce waves that tend to buck each other. Such frequencies cannot produce strong oscillating currents.

In our typical parallel-resonant (LC) circuit, the line current is minimum (because the impedance is maximum). At the same time, the internal oscillating current in the tank is maximum. Oscillating current may be several hundred times as great as line current at resonance.

In any case, this circuit reacts differently to the resonant frequency than it does to all other frequencies. This makes it an effective frequency selector.

Summary of Resonance

Both series- and parallel-LC circuits discriminate between the resonant frequency and all other frequencies by balancing an inductive reactance against an equal capacitive reactance.

In series, these reactances create a very low impedance. In parallel, they create a very high impedance. These characteristics govern how and where designers use resonant circuits. A low-impedance requirement would require a series-resonant circuit. A high-impedance requirement would require the designer to use a parallel-resonant circuit.

Tuning a Band of Frequencies

Our resonant circuits so far have been tuned to a single frequency - the resonant frequency. This is fine if only one frequency is required. However, there are hundreds of stations on many different frequencies.

Therefore, if we go back to our original application, that of tuning to different radio stations, our resonant circuits are not practical. The reason is because a tuner for each frequency would be required and this is not practical.

What is a practical solution to this problem? The answer is simple. Make either the capacitor or the inductor variable. Remember, changing either L or C changes the resonant frequency.

Now you know what has been happening all of these years when you "pushed" the button or "turned" the dial. You have been changing the L or C in the tuned circuits by the amount necessary to adjust the tuner to resonate at the desired frequency. No matter how complex a unit, if it has LC tuners, the tuners obey these basic laws.

Q-9. What is the term for the number of times per second that tank circuit energy is either stored in the inductor or capacitor?

Q-10. In a parallel-resonant circuit, what is the relationship between impedance and current?

Q-11. When is line current minimum in a parallel-LC circuit?

RESONANT CIRCUITS AS FILTER CIRCUITS

The principle of series- or parallel-resonant circuits have many applications in radio, television, communications, and the various other electronic fields throughout the Navy. As you have seen, by making the capacitance or inductance variable, the frequency at which a circuit will resonate can be controlled.

In addition to station selecting or tuning, resonant circuits can separate currents of certain frequencies from those of other frequencies.

Circuits in which resonant circuits are used to do this are called FILTER CIRCUITS.

If we can select the proper values of resistors, inductors, or capacitors, a FILTER NETWORK, or "frequency selector," can be produced which offers little opposition to one frequency, while BLOCKING or ATTENUATING other frequencies. A filter network can also be designed that will "pass" a band of frequencies and "reject" all other frequencies.

Most electronic circuits require the use of filters in one form or another. You have already studied several in modules 6, 7, and 8 of the NEETS.

One example of a filter being applied is in a rectifier circuit. As you know, an alternating voltage is changed by the rectifier to a direct current. However, the d.c. voltage is not pure; it is still pulsating and fluctuating. In other words, the signal still has an a.c. component in addition to the d.c. voltage. By feeding the signal through simple filter networks, the a.c. component is reduced. The remaining d.c. is as pure as the designers require.

Bypass capacitors, which you have already studied, are part of filter networks that, in effect, bypass, or shunt, unwanted a.c. components to ground.

THE IDEA OF "Q"

Several times in this chapter, we have discussed "ideal" or theoretically perfect circuits. In each case, you found that resistance kept our circuits from being perfect. You also found that low resistance in tuners was better than high resistance. Now you will learn about a factor that, in effect, measures just how close to perfect a tuner or tuner component can be. This same factor affects BANDWIDTH and SELECTIVITY. It can be used in figuring voltage across a coil or capacitor in a series-resonant circuit and the amount of circulating (tank) current in a parallel-resonant circuit. This factor is very important and useful to designers. Technicians should have some knowledge of the factor because it affects so many things. The factor is known as Q. Some say it stands for quality (or merit). The higher the Q, the better the circuit; the lower the losses (I^2R), the closer the circuit is to being perfect.

Having studied the first part of this chapter, you should not be surprised to learn that resistance (R) has a great effect on this figure of merit or quality.

Q Is a Ratio

Q is really very simple to understand if you think back to the tuned-circuit principles just covered. Inductance and capacitance are in all tuners. Resistance is an impurity that causes losses. Therefore, components that provide the reactance with a minimum of resistance are "purer" (more perfect) than those with higher resistance. The actual measure of this purity, merit, or quality must include the two basic quantities, X and R.

The ratio

$$\frac{X}{R}$$

does the job for us. Let's take a look at it and see just why it measures quality.

First, if a perfect circuit has zero resistance, then our ratio should give a very high value of Q to reflect the high quality of the circuit. Does it?

Assume any value for X and a zero value for R.

Then:

$$Q = \frac{X}{R} = \frac{\text{Some Value}}{0} = \text{Infinity}$$

Remember, any value divided by zero equals infinity. Thus, our ratio is infinitely high for a theoretically perfect circuit.

With components of higher resistance, the Q is reduced. Dividing by a larger number always yields a smaller quantity. Thus, lower quality components produce a lower Q. Q, then, is a direct and accurate measure of the quality of an LC circuit.

Q is just a ratio. It is always just a number — no units. The higher the number, the "better" the circuit. Later as you get into more practical circuits, you may find that low Q may be desirable to provide certain characteristics. For now, consider that higher is better.

Because capacitors have much, much less resistance in them than inductors, the Q of a circuit is very often expressed as the Q of the coil or:

$$Q = \frac{X_L}{R}$$

The answer you get from using this formula is very near correct for most purposes. Basically, the Q of a capacitor is so high that it does not limit the Q of the circuit in any practical way. For that reason, the technician may ignore it.

The Q of a Coil

Q is a feature that is designed into a coil. When the coil is used within the frequency range for which it is designed, Q is relatively constant. In this sense, it is a physical characteristic.

Inductance is a result of the physical makeup of a coil - number of turns, core, type of winding, etc. Inductance governs reactance at a given frequency. Resistance is inherent in the length, size, and material of the wire. Therefore, the Q of a coil is mostly dependent on physical characteristics.

Values of Q that are in the hundreds are very practical and often found in typical equipment.

Application of Q

For the most part, Q is the concern of designers, not technicians. Therefore, the chances of you having to figure the Q of a coil are remote. However, it is important for you to know some circuit relationships that are affected by Q.

Q Relationships in Series Circuits

Q can be used to determine the "gain" of series-resonant circuits. <u>Gain refers to the fact that at resonance, the voltage drop across the reactances are greater than the applied voltage</u>. Remember, when we applied Ohm's law in a series-resonant circuit, it gave us the following characteristics:

- Low impedance, high current.
- High current; high voltage across the comparatively high reactances.

This high voltage is usable where little power is required, such as in driving the grid of a vacuum tube or the gate of a field effect transistor (F.E.T.). The gain of a properly designed series-resonant circuit may be as great or greater than the amplification within the amplifier itself. The gain is a function of Q, as shown in the following example:

E = the input voltage to the tuned circuit

E_L = the voltage drop across the coil at resonance Q.

Q = the Q of the coil

Then:

$$E_L = EQ$$

If the Q of the coil were 100, then the gain would be 100; that is, the voltage of the coil would be 100 times that of the input voltage to the series circuit.

Resistance affects the resonance curve of a series circuit in two ways — the lower the resistance, the higher the current; also, the lower the resistance, the sharper the curve. Because low resistance causes high Q, these two facts are usually expressed as functions of Q. That is, the <u>higher</u> the Q, the <u>higher</u> and <u>sharper</u> the curve and the more <u>selective</u> the circuit.

The <u>lower</u> the Q (because of higher resistance), the <u>lower</u> the current curve; therefore, <u>the broader the curve, the less selective</u> the circuit. A summary of the major characteristics of series RLC-circuits at resonance is given in table 1-1.

Table 1-1.—Major Characteristics of Series RLC Circuits at Resonance

QUANTITY	SERIES CIRCUIT
At resonance: Reactance ($X_L - X_C$)	Zero, because $X_L = X_C$
Resonant frequency	$f_r = \frac{1}{2\pi\sqrt{LC}}$
Impedance	Minimum: $Z = R$
I_{LINE}	Maximum value
I_L	I_{LINE}
I_C	I_{LINE}
E_L	$Q \cdot E_{LINE}$
E_C	$Q \cdot E_{LINE}$
Phase angle between E_{LINE} and I_{LINE}	0°
Angle between E_L & E_C	180°
Angle between I_L & I_C	0°
Desired value of Q	10 or more
Desired value of R	Low
Highest selectivity	High Q, low R, high $\frac{L}{C}$
When f is greater than f_r Reactance	Inductive
Phase angle between I_{LINE} and E_{LINE}	Lagging current
When f is less than f_r Reactance	Capacitive
Phase angle between I_{LINE} and E_{LINE}	Leading current

Q Relationships in a Parallel-Resonant Circuit

There is no voltage gain in a parallel-resonant circuit because voltage is the same across all parts of a parallel circuit. However, Q helps give us a measure of the current that circulates in the tank.

Given:

I_{LINE} = current drawn from the source

I_L = current through the coil (or circulating current)

Q = the Q of the coil

Then:

$$I_L = I_{LINE}\, Q$$

Again, if the Q were 100, the circulating current would be 100 times the value of the line current. This may help explain why some of the wire sizes are very large in high-power amplifying circuits.

The impedance curve of a parallel-resonant circuit is also affected by the Q of the circuit in a manner similar to the current curve of a series circuit. The Q of the circuit determines how much the impedance is increased across the parallel-LC circuit. ($Z = Q \times X_L$)

The higher the Q, the greater the impedance at resonance and the sharper the curve. The lower the Q, the lower impedance at resonance; therefore, the broader the curve, the less selective the circuit. The major characteristics of parallel-RLC circuits at resonance are given in table 1-2.

Table 1-2.—Major Characteristics of Parallel RLC Circuits at Resonance

QUANTITY	PARALLEL CIRCUIT
At resonance: Reactance $(X_L - X_C)$	Zero; because nonenergy currents are equal
Resonant frequency	$f_r = \frac{1}{2\pi\sqrt{LC}}$
Impedance	Maximum: $Z = \frac{L}{CR}$
I_{LINE}	Minimum value
I_L	$Q \cdot I_{LINE}$
I_C	$Q \cdot I_{LINE}$
E_L	E_{LINE}
E_C	E_{LINE}
Phase angle between E_{LINE} and I_{LINE}	0°
Angle between E_L & E_C	0°
Angle between I_L & I_C	180°
Desired value of **Q**	**10** or more
Desired value of **R**	Low
Highest selectivity	High **Q**, low **R**, $\frac{L}{C}$
When **f** is greater than f_r Reactance	Capacitive
Phase angle between I_{LINE} and E_{LINE}	Leading current
When **f** is less than f_r Reactance	Inductive
Phase angle between I_{LINE} and E_{LINE}	Lagging current

Summary of Q

The ratio that is called Q is a measure of the quality of resonant circuits and circuit components. Basically, the value of Q is an inverse function of electrical power dissipated through circuit resistance. Q is the ratio of the power stored in the reactive components to the power dissipated in the resistance. That is, high power loss is low Q; low power loss is high Q.

Circuit designers provide the proper Q. As a technician, you should know what can change Q and what quantities in a circuit are affected by such a change.

BANDWIDTH

If circuit Q is low, the gain of the circuit at resonance is relatively small. The circuit does not discriminate sharply (reject the unwanted frequencies) between the resonant frequency and the frequencies on either side of resonance, as shown by the curve in figure 1-12, view (A). The range of frequencies included between the two frequencies (426.4 kHz and 483.6 kHz in this example) at which the current drops to 70 percent of its maximum value at resonance is called the BANDWIDTH of the circuit.

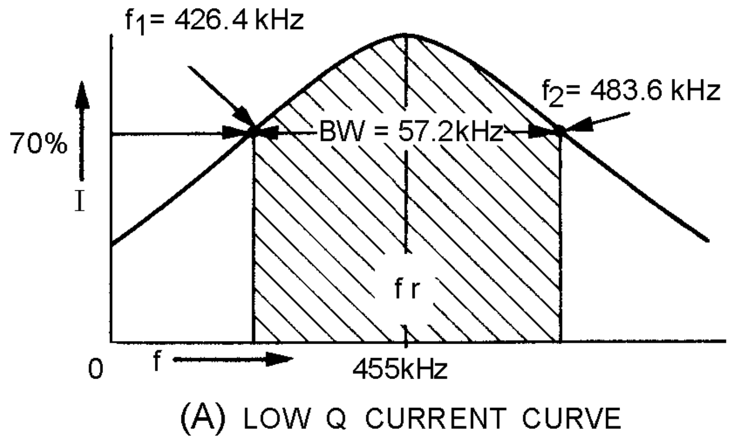


Figure 1-12A.—Bandwidth for high- and low-Q series circuit. LOW Q CURRENT CURVE.

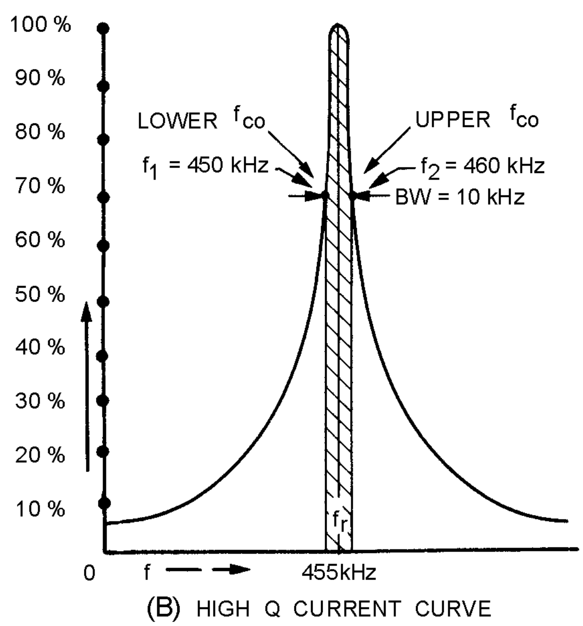


Figure 1-12B.—Bandwidth for high- and low-Q series circuit. HIGH Q CURRENT CURVE.

It is often necessary to state the band of frequencies that a circuit will pass. The following standard has been set up: the limiting frequencies are those at either side of resonance at which the curve falls to a point of .707 (approximately 70 percent) of the maximum value. This point is called the HALF-POWER point. Note that in figure 1-12, the series-resonant circuit has two half-power points, one above and one below the resonant frequency point. The two points are designated upper frequency cutoff (f_{co}) and lower frequency cutoff (f_{co}) or simply f_1 and f_2. The range of frequencies between these two points comprises the bandwidth. Views (A) and (B) of figure 1-12 illustrate the bandwidths for low- and high-Q resonant circuits. The bandwidth may be determined by use of the following formulas:

$$BW = \frac{f_r}{Q}$$

or

$$BW = f_2 - f_1$$

Where:

BW = bandwidth of a circuit in units of frequency

f_r = resonant frequency

f_2 = the upper cutoff frequency

f_1 = the lower cutoff frequency

For example, by applying the formula we can determine the bandwidth for the curve shown in figure 1-12, view (A).

Solution:

$$BW = f_2 - f_1$$

$$BW = 483.6 \text{ kHz} - 426.4 \text{ kHz}$$

$$BW = 57.2 \text{ kHz}$$

If the Q of the circuit represented by the curve in figure 1-12, view (B), is 45.5, what would be the bandwidth?

Solution:

$$BW = \frac{f_r}{Q}$$

$$BW = \frac{455 \text{ kHz}}{45.5}$$

$$BW = 10 \text{ kHz}$$

If Q equals 7.95 for the low-Q circuit as in view (A) of figure 1-12, we can check our original calculation of the bandwidth.

Solution:

$$BW = \frac{f_r}{Q}$$

$$BW = \frac{455\ kHz}{7.95}$$

$$BW = 57.2\ kHz$$

The Q of the circuit can be determined by transposing the formula for bandwidth to:

$$Q = \frac{f_r}{BW}$$

To find the Q of the circuit using the information found in the last example problem:

Given:

$$f_r = 455\ kHz$$

$$BW = 57.2\ kHz$$

Solution:

$$Q = \frac{f_r}{BW}$$

$$Q = \frac{455\ kHz}{57.2\ kHz}$$

$$Q = 7.95$$

Q-12. What is the relationship of the coil to the resistance of a circuit with high "Q"?

Q-13. What is the band of frequencies called that is included between the two points at which current falls to 70 percent of its maximum value in a resonant circuit?

FILTERS

In many practical applications of complex circuits, various combinations of direct, low-frequency, audio-frequency, and radio-frequency currents may exist. It is frequently necessary to have a means for separating these component currents at any desired point. An electrical device for accomplishing this separation is called a FILTER.

A filter circuit consists of inductance, capacitance, and resistance used singularly or in combination, depending upon the purpose. It may be designed so that it will separate alternating current from direct current, or so that it will separate alternating current of one frequency (or a band of frequencies) from other alternating currents of different frequencies.

The use of resistance by itself in filter circuits does not provide any filtering action, because it opposes the flow of any current regardless of its frequency. What it does, when connected in series or parallel with an inductor or capacitor, is to decrease the "sharpness," or selectivity, of the filter. Hence, in some particular application, resistance might be used in conjunction with inductance or capacitance to provide filtering action over a wider band of frequencies.

Filter circuits may be divided into four general types: LOW-PASS, HIGH-PASS, BANDPASS, AND BAND-REJECT filters.

Electronic circuits often have currents of different frequencies. The reason is that a source produces current with the same frequency as the applied voltage. As an example, the a.c. signal input to an audio amplifier can have high- and low-audio frequencies; the input to an rf amplifier can have a wide range of radio frequencies.

In such applications where the current has different frequency components, it is usually necessary for the filter either to accept or reject one frequency or a group of frequencies. The electronic filter that can pass on the higher-frequency components to a load or to the next circuit is known as a HIGH-PASS filter. A LOW-PASS filter can be used to pass on lower-frequency components.

Before discussing filters further, we will review and apply some basic principles of the frequency-response characteristics of the capacitor and the inductor. Recall the basic formula for capacitive reactance and inductive reactance:

$$X_C = \frac{1}{2\pi fC}$$

and

$$X_L = 2\pi fL$$

Assume any given value of L and C. If we increase the applied frequency, X_C decreases and X_L increases. If we increase the frequency enough, the capacitor acts as a short and the inductor acts as an open. Of course, the opposite is also true. Decreasing frequency causes X_C to increase and X_L to decrease. Here again, if we make a large enough change, X_C acts as an open and X_L acts as a short. Figure 1-13 gives a pictorial representation of these two basic components and how they respond to low and high frequencies.

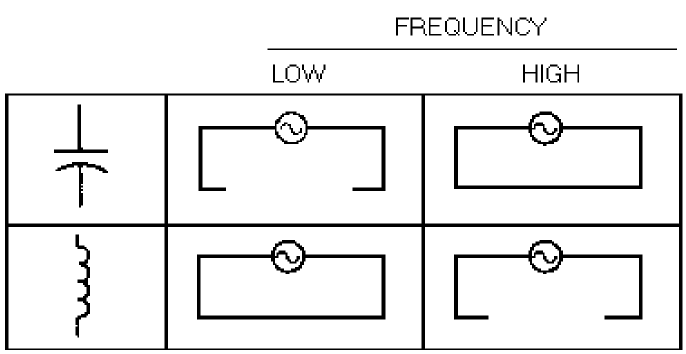


Figure 1-13.—Effect of frequency on capacitive and inductive reactance.

If we apply these same principles to simple circuits, such as the ones in figure 1-14, they affect input signals as shown. For example, in view (A) of the figure, a low frequency is blocked by the capacitor which acts as an open and at a high frequency the capacitor acts as a short. By studying the figure, it is easy to see how the various components will react in different configurations with a change in frequency.

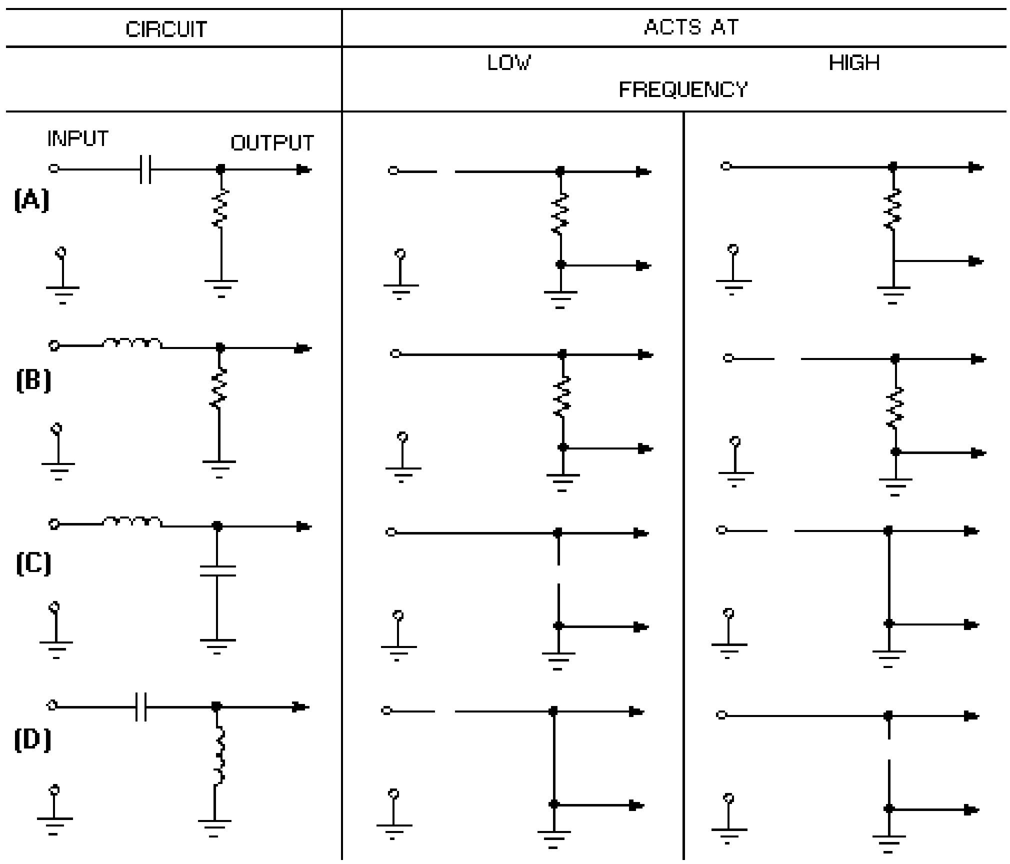


Figure 1-14.—Reaction to circuit by change in frequency.

As mentioned before, high-pass and low-pass filters pass the specific frequencies for which circuits are designed.

There can be a great deal of confusion when talking about high-pass, low-pass, discrimination, attenuation, and frequency cutoff, unless the terms are clearly understood. Since these terms are used widely throughout electronics texts and references, you should have a clear understanding before proceeding further.

- HIGH-PASS FILTER. A high-pass filter passes on a majority of the high frequencies to the next circuit and rejects or attenuates the lower frequencies. Sometimes it is called a low-frequency discriminator or low-frequency attenuator.

- LOW-PASS FILTER. A low-pass filter passes on a majority of the low frequencies to the next circuit and rejects or attenuates the higher frequencies. Sometimes it is called a high-frequency discriminator or high-frequency attenuator.

- DISCRIMINATION. The ability of the filter circuit to distinguish between high and low frequencies and to eliminate or reject the unwanted frequencies.

- ATTENUATION. The ability of the filter circuit to reduce the amplitude of the unwanted frequencies below the level of the desired output frequency.

- FREQUENCY CUTOFF (f_{co}). The frequency at which the filter circuit changes from the point of rejecting the unwanted frequencies to the point of passing the desired frequency; OR the point at which the filter circuit changes from the point of passing the desired frequency to the point of rejecting the undesired frequencies.

LOW-PASS FILTER

A low-pass filter passes all currents having a frequency below a specified frequency, while opposing all currents having a frequency above this specified frequency. This action is illustrated in its ideal form in view (A) of figure 1-15. At frequency cutoff, known as f_c the current decreases from maximum to zero. At all frequencies above f_c the filter presents infinite opposition and there is no current. However, this sharp division between no opposition and full opposition is impossible to attain. A more practical graph of the current is shown in view (B), where the filter gradually builds up opposition as the cutoff frequency (f) is approached. Notice that the filter cannot completely block current above the cutoff frequency.

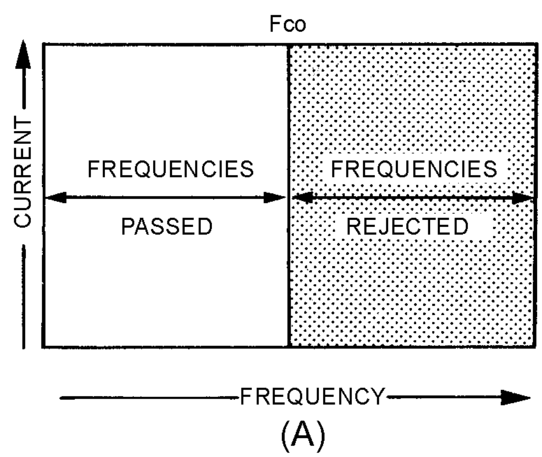


Figure 1-15A.—Low-pass filter.

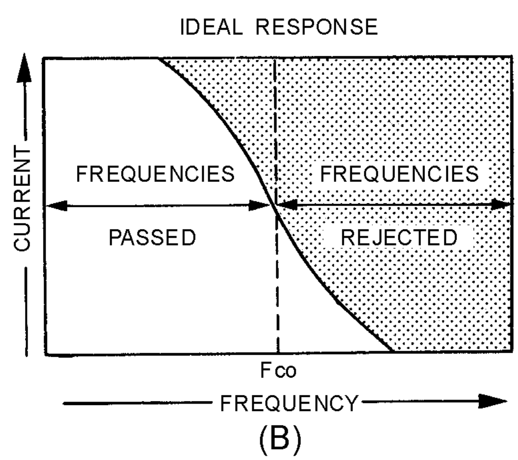


Figure 1-15B.—Low-pass filter.

View (A) of figure 1-16 shows the electrical construction of a low-pass filter with an inductor inserted in series with one side of a line carrying both low and high frequencies. The opposition offered by the reactance will be small at the lower frequencies and great at the higher frequencies. In order to divert the undesired high frequencies back to the source, a capacitor must be added across the line to bypass the higher frequencies around the load, as shown in view (B).

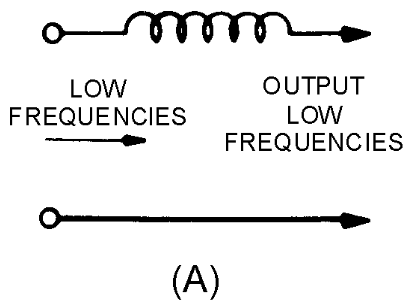


Figure 1-16A.—Components of a simple low-pass filter.

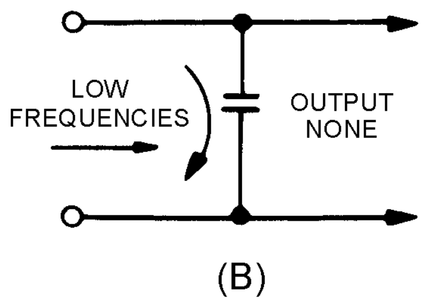


Figure 1-16B.—Components of a simple low-pass filter.

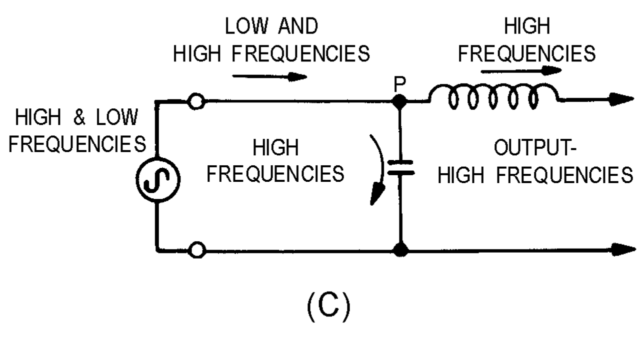


Figure 1-16C.—Components of a simple low-pass filter.

The capacitance of the capacitor must be such that its reactance will offer little opposition to frequencies above a definite value, and great opposition to frequencies below this value. By combining the series inductance and bypass capacitance, as shown in view (C), the simplest type of low-pass filter is obtained. At point P, a much higher opposition is offered to the low frequencies by the capacitor than by the inductor, and most of the low-frequency current takes the path of least opposition. On the other hand, the least amount of opposition is offered to the high frequencies by the capacitor, and most of the high-frequency energy returns to the source through the capacitor.

HIGH-PASS FILTER

A high-pass filter circuit passes all currents having a frequency higher than a specified frequency, while opposing all currents having a frequency lower than its specified frequency. This is illustrated in figure 1-17. A capacitor that is used in series with the source of both high and low frequencies, as shown in view (A) of figure 1-18, will respond differently to high-frequency, low-frequency, and direct currents. It will offer little opposition to the passage of high-frequency currents, great opposition to the passage of low-frequency currents, and completely block direct currents. The value of the capacitor must be chosen so that it allows the passage of all currents having frequencies above the desired value, and opposes those having frequencies below the desired value. Then, in order to shunt the undesired low-frequency currents back to the source, an inductor is used, as shown in view (B). This inductor must have a value that will allow it to pass currents having frequencies below the frequency cutoff point, and reject currents having frequencies above the frequency cutoff point, thus forcing them to pass through the capacitor. By combining inductance and capacitance, as shown in view (C), you obtain the simplest type of high-pass filter. At point P most of the high-frequency energy is passed on to the load by the capacitor, and most of the low-frequency energy is shunted back to the source through the inductor.

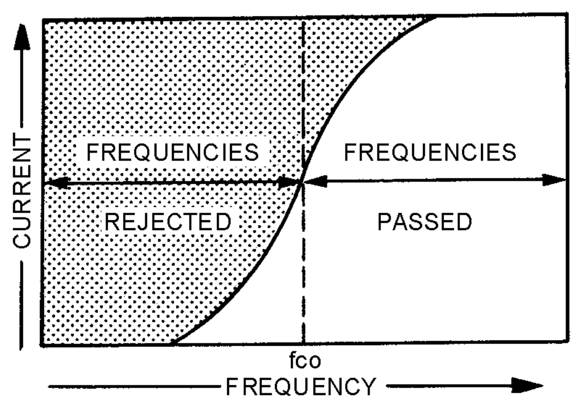


Figure 1-17.—High-pass filter response curve.

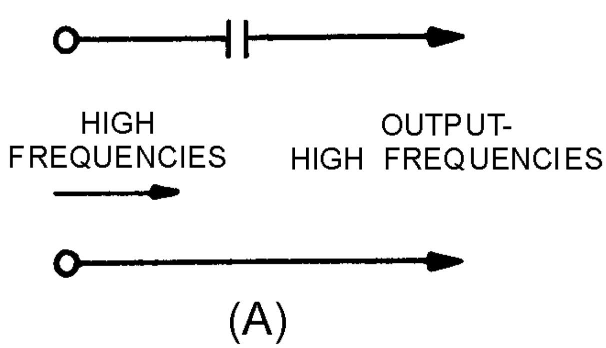


Figure 1-18A.—Components of a simple high-pass filter.

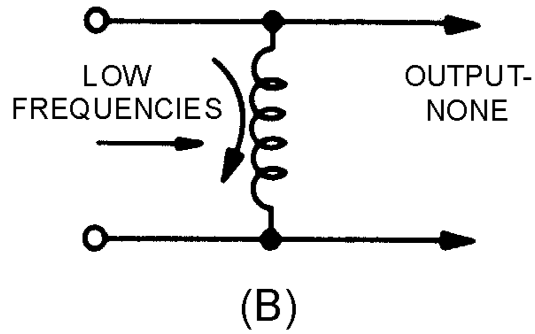


Figure 1-18B.—Components of a simple high-pass filter.

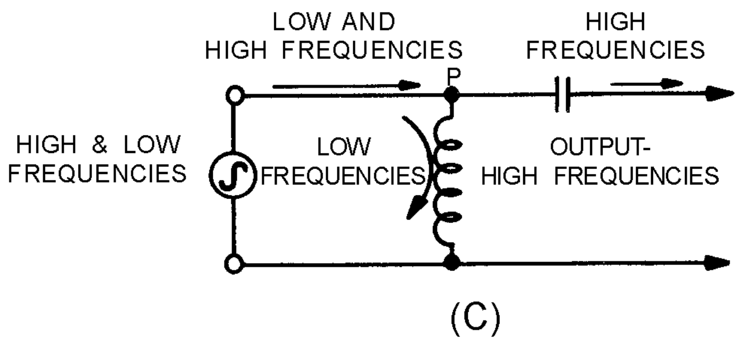


Figure 1-18C.—Components of a simple high-pass filter.

RESONANT CIRCUITS AS FILTERS

Resonant circuits can be made to serve as filters in a manner similar to the action of individual capacitors and inductors. As you know, the series-LC circuit offers minimum opposition to currents that have frequencies at or near the resonant frequency, and maximum opposition to currents of all other frequencies.

You also know that a parallel-LC circuit offers a very high impedance to currents that have frequencies at or near the resonant frequency, and a relatively low impedance to currents of all other frequencies.

If you use these two basic concepts, the BANDPASS and BAND-REJECT filters can be constructed. The bandpass filter and the band-reject filter are two common types of filters that use resonant circuits.

Bandpass Filter

A bandpass filter passes a narrow band of frequencies through a circuit and attenuates all other frequencies that are higher or lower than the desired band of frequencies. This is shown in figure 1-19 where the greatest current exists at the center frequency (f_r). Frequencies below resonance (f_1) and frequencies above resonance (f_2) drop off rapidly and are rejected.

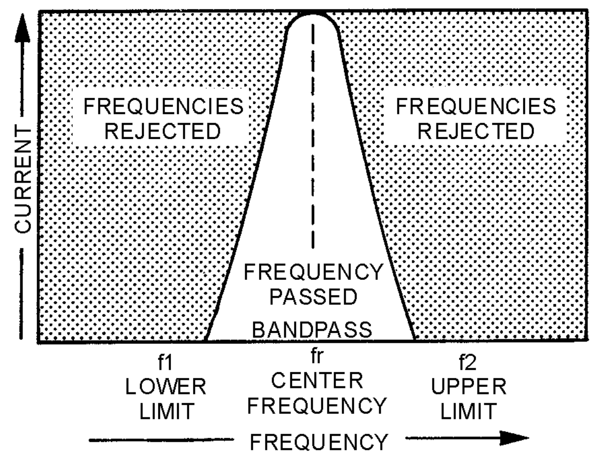


Figure 1-19.—Bandpass filter response curve.

In the circuit of figure 1-20, view (A), the series-LC circuit replaces the inductor of figure 1-16, view (A), and acts as a BANDPASS filter. It passes currents having frequencies at or near its resonant frequency, and opposes the passage of all currents having frequencies outside this band.

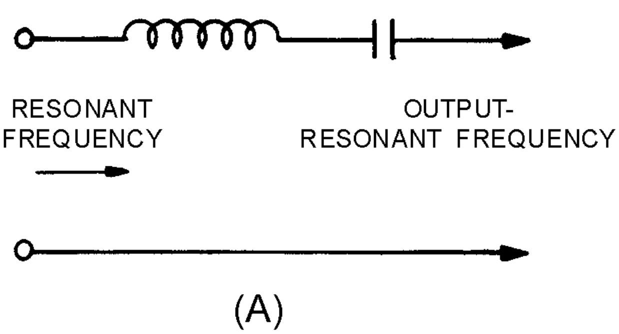


Figure 1-20A.—Components of a simple bandpass filter.

Thus, in the circuit of figure 1-20, view (B), the parallel-LC circuit replaces the capacitor of figure 1-16, view (B). If this circuit is tuned to the same frequency as the series-LC circuit, it will provide a path for all currents having frequencies outside the limits of the frequency band passed by the series-resonant circuit. The simplest type of bandpass filter is formed by connecting the two LC circuits as shown in figure 1-20, view (C). The upper and lower frequency limits of the filter action are filter cutoff points.

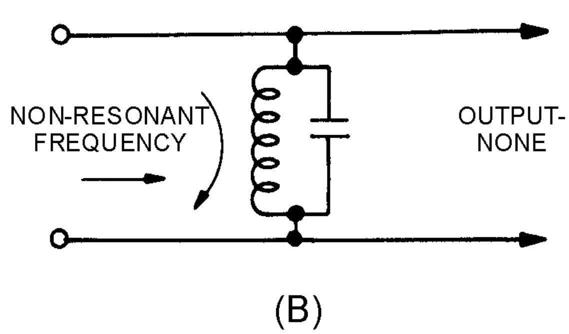


Figure 1-20B.—Components of a simple bandpass filter.

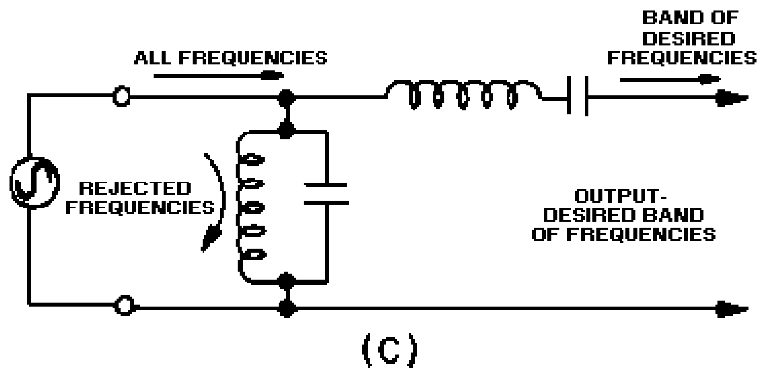


Figure 1-20C.—Components of a simple bandpass filter.

Band-Reject Filter

A band-reject filter circuit is used to block the passage of current for a narrow band of frequencies, while allowing current to flow at all frequencies above or below this band. This type of filter is also known as a BAND-SUPPRESSION or BAND-STOP filter. The way it responds is shown by the response curve of figure 1-21. Since the purpose of the band-reject filter is directly opposite to that of a bandpass filter, the relative positions of the resonant circuits in the filter are interchanged. The parallel-LC circuit shown in figure 1-22, view (A), replaces the capacitor of figure 1-18, view (A). It acts as a band-reject filter, blocking the passage of currents having frequencies at or near resonant frequency and passing all currents having frequencies outside this band. The series-LC circuit shown in figure 1-22, view (B), replaces the inductor of figure 1-18, view (B). If this series circuit is tuned, to the same frequency as the parallel circuit, it acts as a bypass for the band of rejected frequencies. Then, the simplest type of band-reject filter is obtained by connecting the two circuits as shown in figure 1-22, view (C).

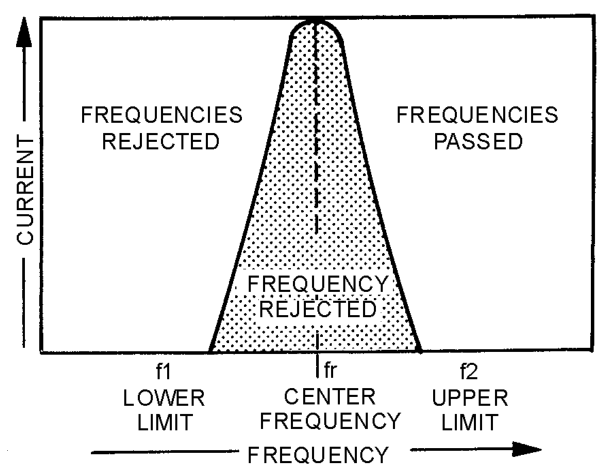


Figure 1-21.—Band-reject filter response curve.

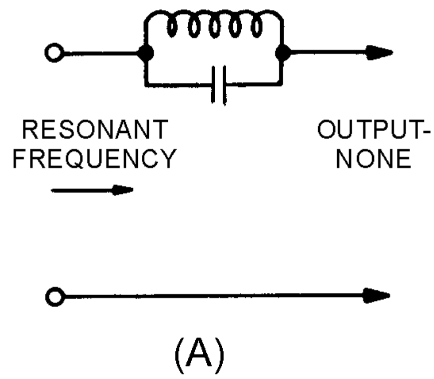


Figure 1-22A.—Components of a simple band-reject filter.

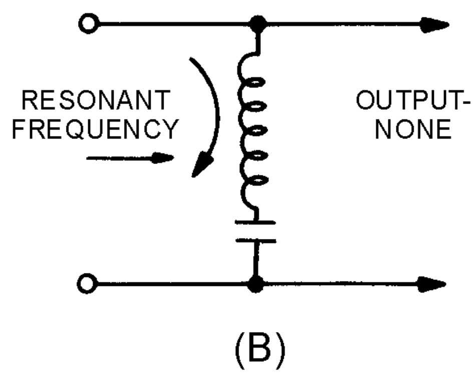


Figure 1-22B.—Components of a simple band-reject filter.

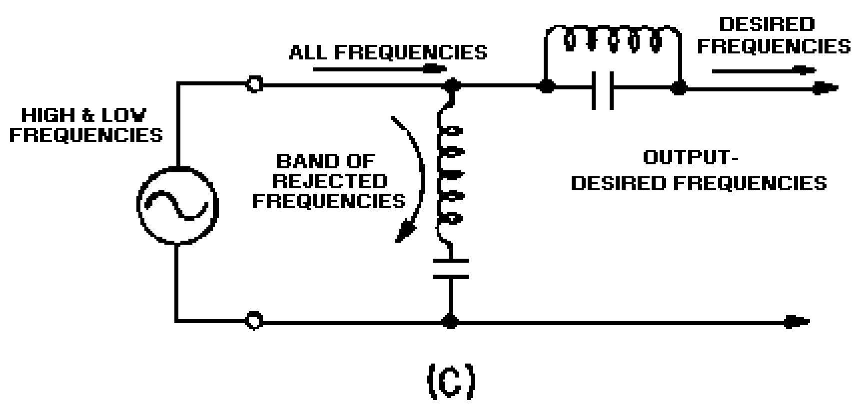


Figure 1-22C.—Components of a simple band-reject filter.

Q-14. What is the device called that will separate alternating current from direct current, or that will separate alternating current of one frequency from other alternating currents of different frequencies?

Q-15. What are the four general types of filters?

Q-16. What is the filter called in which the low frequencies do not produce a useful voltage?

Q-17. What is the filter called that passes low frequencies but rejects or attenuates high frequencies?

Q-18. How does a capacitor and an inductor react to (a) low frequency and (b) high frequency?

Q-19. What term is used to describe the frequency at which the filter circuit changes from the point of rejecting the unwanted frequencies to the point of passing the desired frequencies?

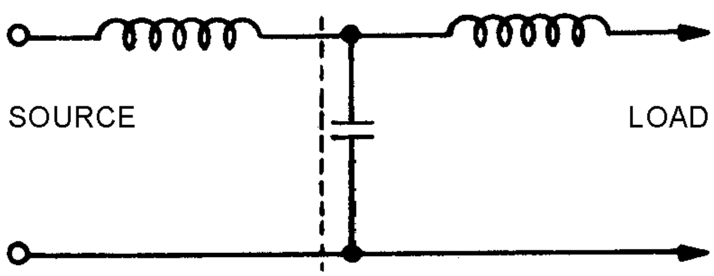
Q-20. What type filter is used to allow a narrow band of frequencies to pass through a circuit and attenuate all other frequencies above or below the desired band?

Q-21. What type filter is used to block the passage of current for a narrow band of frequencies, while allowing current to flow at all frequencies above or below this band?

MULTISECTION FILTERS

All of the various types of filters we have discussed so far have had only one section. In many cases, the use of such simple filter circuits does not provide sufficiently sharp cutoff points. But by adding a capacitor, an inductor, or a resonant circuit in series or in parallel (depending upon the type of filter action required), the ideal effect is more nearly approached. When such additional units are added to a filter circuit, the form of the resulting circuit will resemble the letter T, or the Greek letter π (pi). They are, therefore, called T- or π-type filters, depending upon which symbol they resemble. Two or more T- or π–type filters may be connected together to produce a still sharper cutoff point.

Figure 1-23, (view A) (view B) and (view C), and figure 1-24, (view A) (view B) and (view C) depict some of the common configurations of the T- and π-type filters. Further discussion about the theory of operation of these circuits is beyond the intended scope of this module. If you are interested in learning more about filters, a good source of information to study is the *Electronics Installation and Maintenance Handbook* (EIMB), section 4 (Electronics Circuits), NAVSEA 0967-LP-000-0120.



(A)

Figure 1-23A.—Formation of a T-type filter.

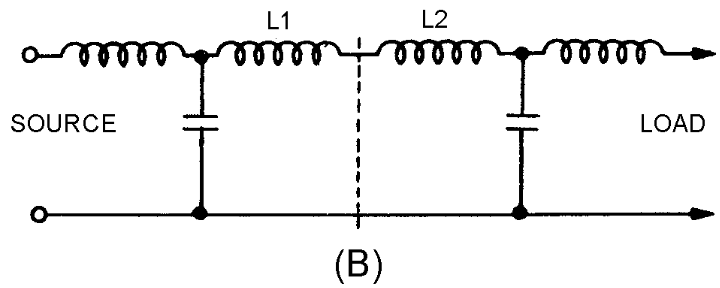


Figure 1-23B.—Formation of a T-type filter.

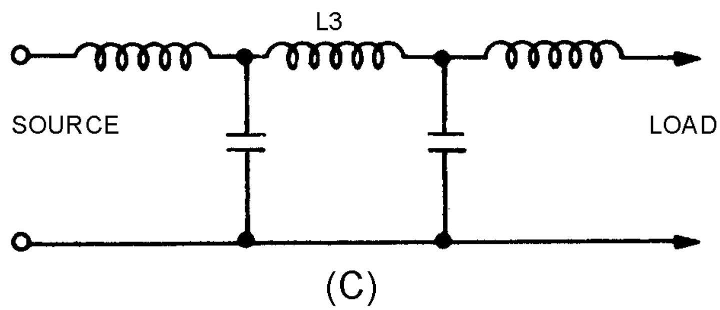


Figure 1-23C.—Formation of a T-type filter.

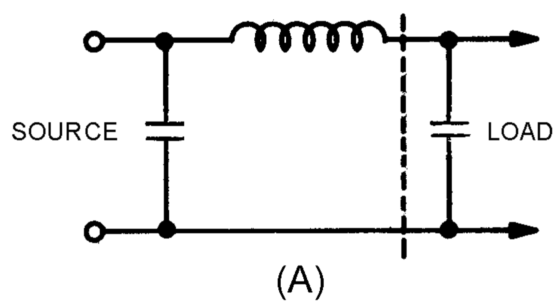


Figure 1-24A.—Formation of a π-type filter.

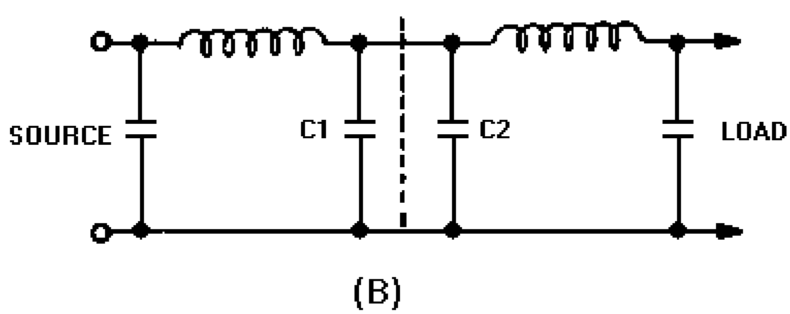


Figure 1-24B.—Formation of a π-type filter.

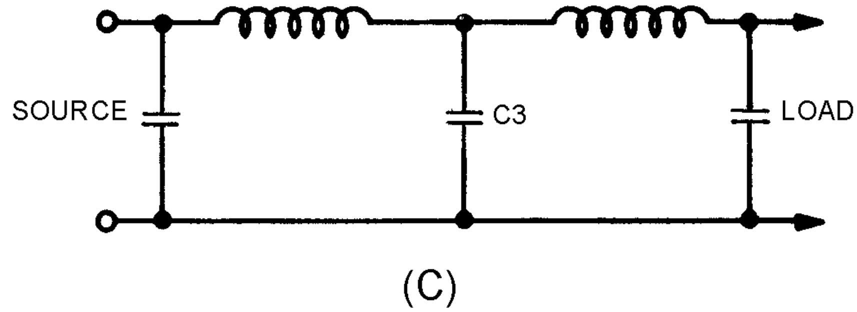


Figure 1-24C.—Formation of a π-type filter.

SAFETY PRECAUTIONS

When working with resonant circuits, or electrical circuits, you must be aware of the potentially high voltages. Look at figure 1-25. With the series circuit at resonance, the total impedance of the circuit is 5 ohms.

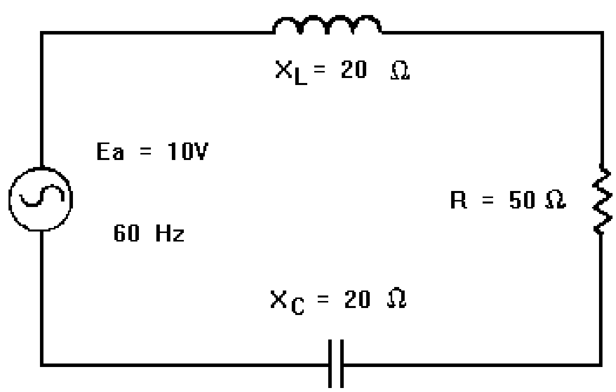


Figure 1-25.—Series RLC circuit at resonance.

Remember, the impedance of a series-RLC circuit at resonance depends on the resistive element. At resonance, the impedance (Z) equals the resistance (R). Resistance is minimum and current is maximum. Therefore, the current at resonance is:

$$I_T = \frac{E_a}{Z} = \frac{10\text{ V}}{5\Omega} = 2\text{ A}$$

The voltage drops around the circuit with 2 amperes of current flow are:

$$E_C = I_T \times X_C$$

$$E_C = 2 \times 20$$

$$E_C = 40 \text{ volts a.c.}$$

$$E_L = I_T \times X_L$$

$E_L = 2 \times 20$

$E_L = 40$ volts a.c.

$E_R = I_T \times R$

$E_R = 2 \times 5$

$E_R = 10$ volts a.c.

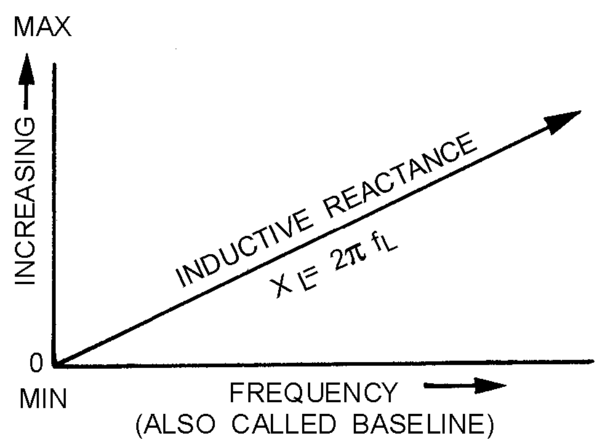
You can see that there is a voltage gain across the reactive components at resonance.

If the frequency was such that X_L and X_C were equal to 1000 ohms at the resonant frequency, the reactance voltage across the inductor or capacitor would increase to 2000 volts a.c. with 10 volts a.c. applied. Be aware that potentially high voltage can exist in series-resonant circuits.

SUMMARY

This chapter introduced you to the principles of tuned circuits. The following is a summary of the major subjects of this chapter.

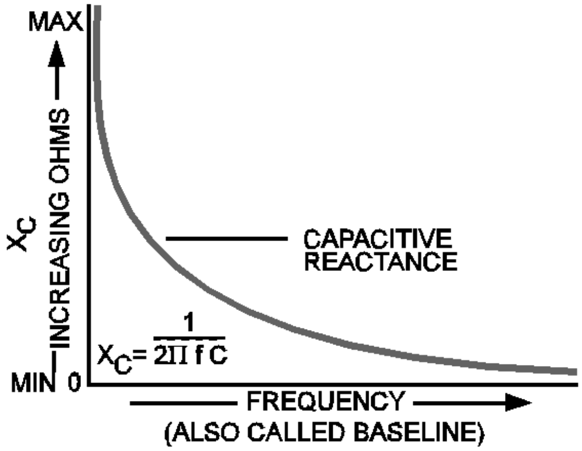
THE **EFFECT OF FREQUENCY** on an **INDUCTOR** is such that an increase in frequency will cause an increase in inductive reactance. Remember that $X_L = 2\pi fL$; therefore, X_L varies directly with frequency.



THE **EFFECT OF FREQUENCY** on a **CAPACITOR** is such that an increase in frequency will cause a decrease in capacitive reactance. Remember that

$$X_C = \frac{1}{2\pi fC}$$

therefore, the relationship between X_C and frequency is that X_C varies inversely with frequency.

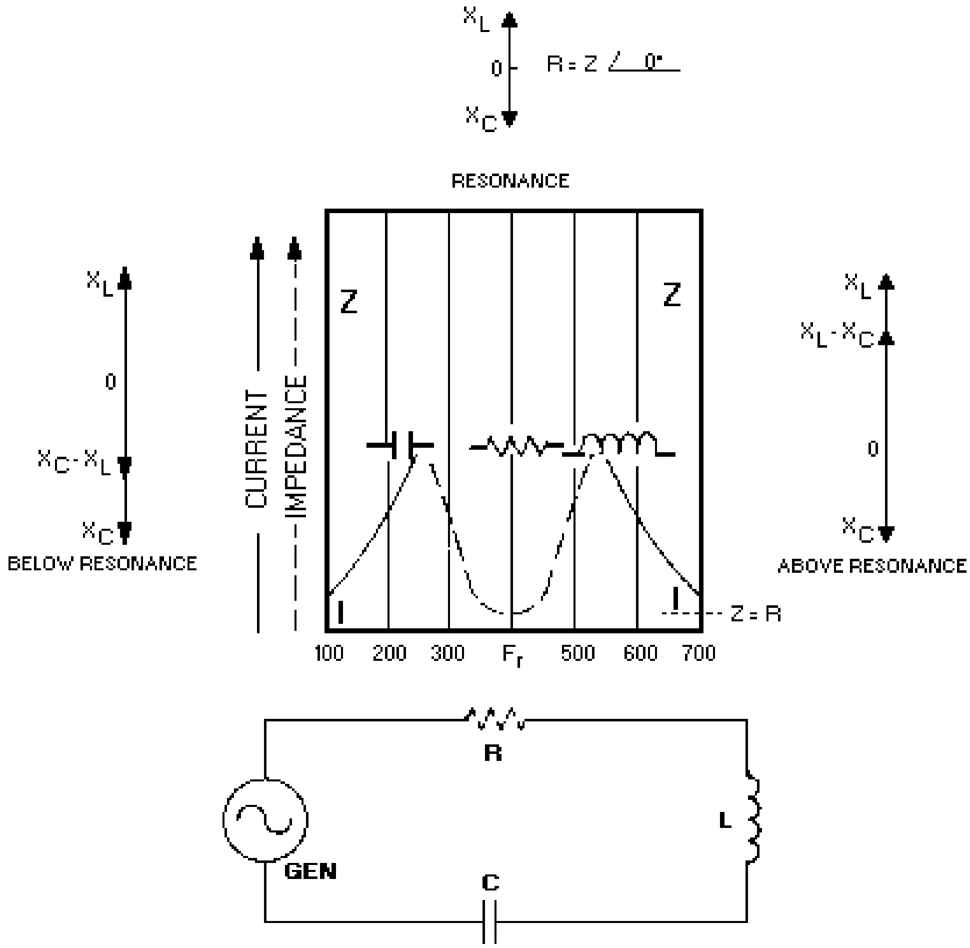


RESULTANT REACTANCE $X = (X_L - X_C)$ or $X = (X_C - X_L)$. X_L is usually plotted above the reference line and X_C below the reference line. Inductance and capacitance have opposite effects on the current in respect to the voltage in a.c. circuits. Below resonance, X_C is larger than X_L, and the series circuit appears capacitive. Above resonance, X_L is larger than X_C, and the series circuit appears inductive. At resonance, $X_L = X_C$, and the total impedance of the circuit is resistive.

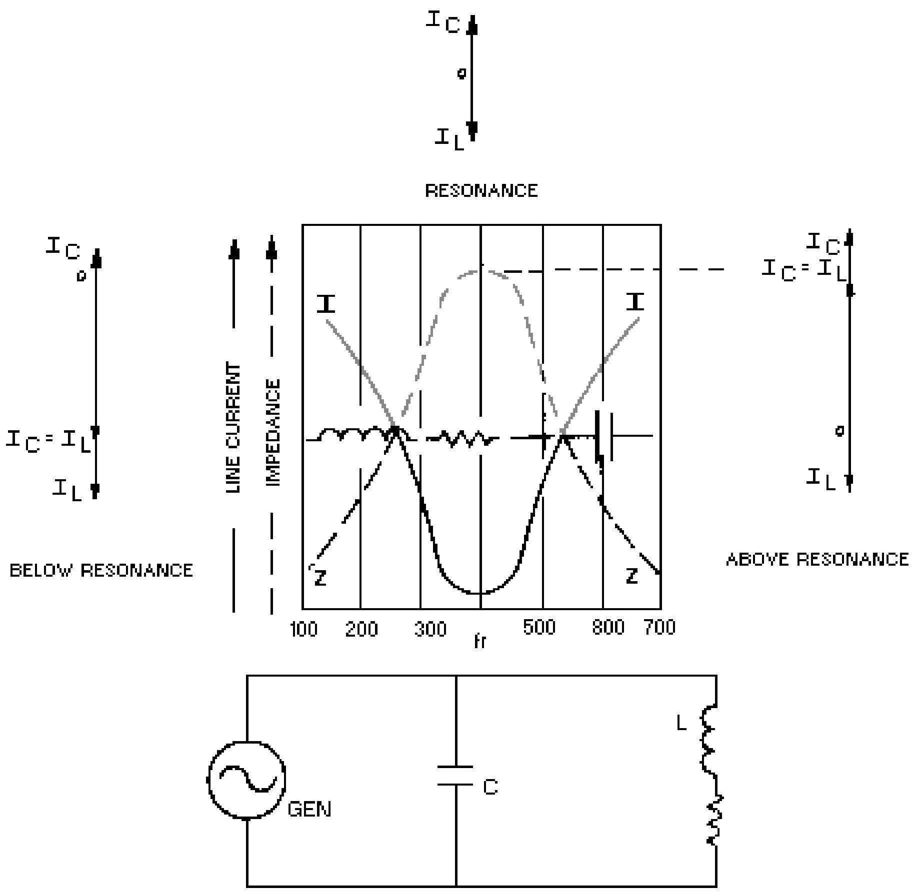
A **RESONANT CIRCUIT** is often called a **TANK CIRCUIT**. It has the ability to take energy fed from a power source, store the energy alternately in the inductor and capacitor, and produce an output which is a continuous a.c. wave. The number of times this set of events occurs per second is called the resonant frequency of the circuit. The actual frequency at which a tank circuit will oscillate is determined by the formula:

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

IN A **SERIES-LC CIRCUIT** impedance is minimum and current is maximum. Voltage is the variable, and voltage across the inductor and capacitor will be equal but of opposite phases at resonance. Above resonance it acts inductively, and below resonance it acts capacitively.



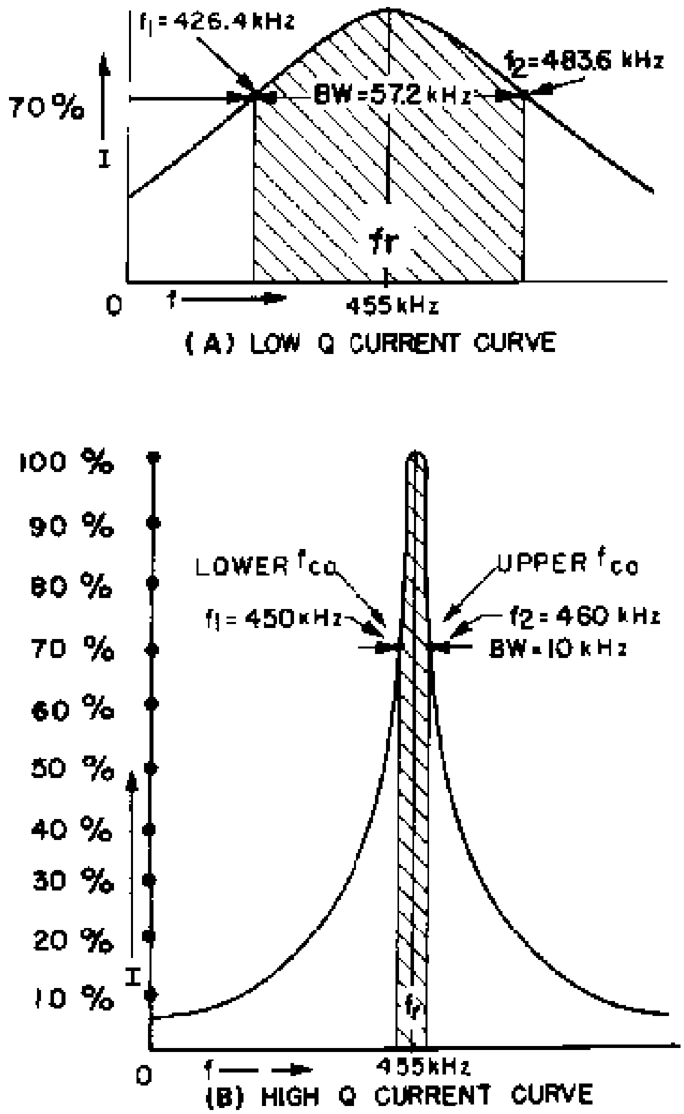
IN A **PARALLEL-LC CIRCUIT** impedance is maximum and current is minimum. Current is the variable and at resonance the two currents are 180 degrees out of phase with each other. Above resonance the current acts capacitively, and below resonance the current acts inductively.



THE **"Q"** OR **FIGURE OF MERIT** of a circuit is the ratio of X_L to R. Since the capacitor has negligible losses, the circuit Q becomes equivalent to the Q of the coil.

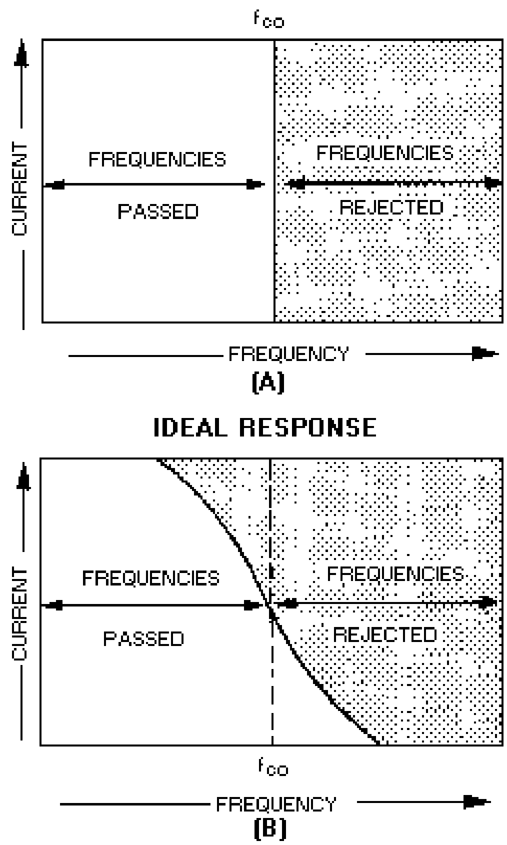
$$(Q = \frac{X_L}{R})$$

THE **BANDWIDTH** of a circuit is the range of frequencies between the half-power points. The limiting frequencies are those at either side of resonance at which the curve falls to .707 of the maximum value. If circuit Q is low, you will have a wide bandpass. If circuit Q is high, you will have a narrow bandpass.

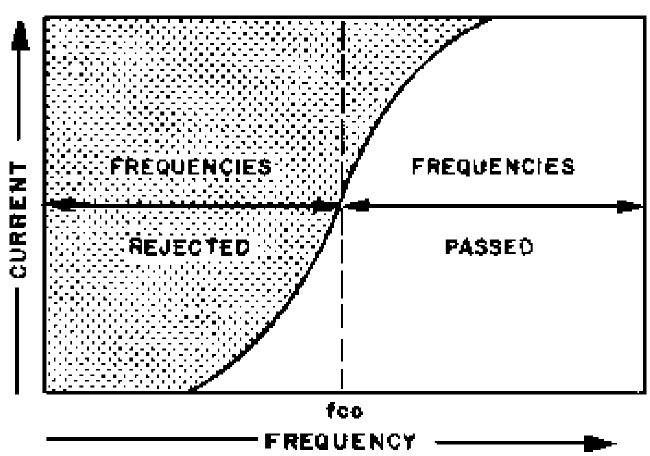


A **FILTER CIRCUIT** consists of a combination of capacitors, inductors, and resistors connected so that the filter will either permit or prevent passage of a certain band of frequencies.

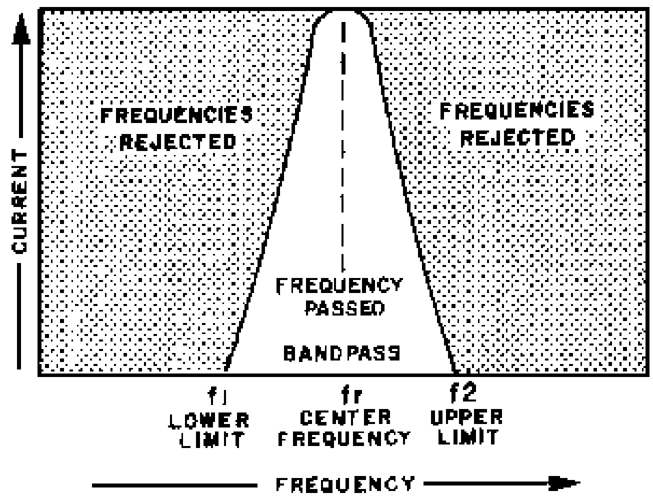
A **LOW-PASS FILTER** passes low frequencies and attenuates high frequencies.



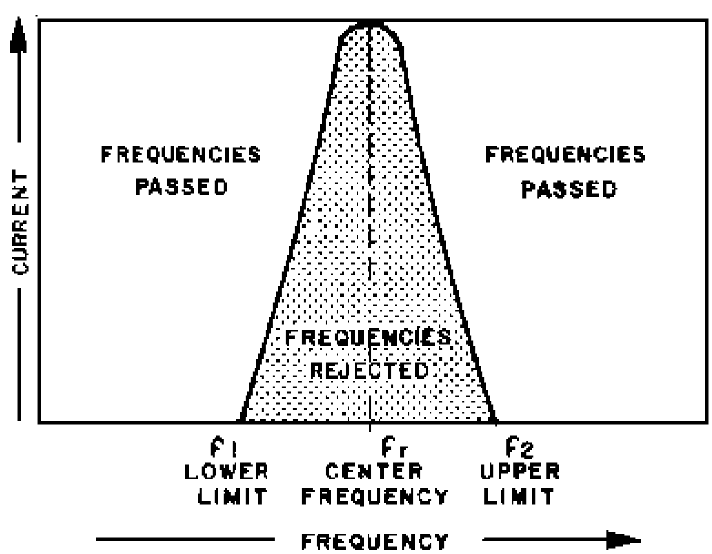
A **HIGH-PASS FILTER** passes high frequencies and attenuates low frequencies.



A **BANDPASS FILTER** will permit a certain band of frequencies to be passed.



A **BAND-REJECT FILTER** will reject a certain band of frequencies and pass all others.



A **SAFETY PRECAUTION** concerning series resonance: Very high reactive voltage can appear across L and C. Care must be taken against possible shock hazard.

ANSWERS TO QUESTIONS Q1. THROUGH Q21.

A-1.

a. X_L *varies directly with frequency.*

$X_L = 2\pi fL$

b. X_C *varies inversely with frequency.*

$$X_C = \frac{1}{2\pi fC}$$

c. Frequency has no affect on resistance.

A-2. Resultant reactance.

A-3.

$$f_r = \frac{1}{2\pi\sqrt{LC}} \quad \text{or} \quad f_r = \frac{.159}{\sqrt{LC}}$$

A-4. Decreases.

A-5. Impedance low Current high.

A-6. Nonresonant (circuit is either above or below resonance).

A-7. Inductor magnetic field.

A-8. Capacitor.

A-9. Natural frequency or resonant frequency (f r).

A-10. Maximum impedance, minimum current.

A-11. At the resonant frequency.

A-12.

$$(Q = \frac{X_L}{R})\ (\text{high } X_L,\ \text{low } R)$$

A-13. Bandwidth of the circuit.

A-14. A filter.

A-15.

a. Low-pass.

b. High-pass

c. Bandpass.

d. Band-reject.

A-16. High-pass filter, low-frequency discriminator, or low-frequency attenuator.

A-17. Low-pass filter, high-frequency discriminator or high-frequency attenuator.

A-18. At low-frequency, a capacitor acts as an open and an inductor acts as a short. At high-frequency, a capacitor acts as a short and an inductor acts as an open.

A-19. Frequency cutoff (f_{co}).

A-20. Bandpass.

A-21. Band-reject.

CHAPTER 2

OSCILLATORS

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. List the two broad classifications of oscillators (wave generators).
2. Identify the three frequency-determining devices for sine-wave oscillators.
3. Describe the differences between series-fed and shunt-fed oscillators.
4. Explain how the crystal is equivalent to the series and parallel LC circuit.
5. Identify the Armstrong oscillator.
6. Identify the Hartley oscillator.
7. Identify the Colpitts oscillator.
8. Identify the resistive-capacitive oscillator.
9. Determine the frequency of a resistive-capacitive oscillator.
10. Explain the operation of a pulsed oscillator.
11. Determine how many cycles are present in the output of a pulsed oscillator.
12. Explain how frequency multiplication takes place.

INTRODUCTION

WAVE GENERATORS play a prominent role in the field of electronics. They generate signals from a few hertz to several gigahertz (10^9 hertz). Modern wave generators use many different circuits and generate such outputs as SINUSOIDAL, SQUARE, RECTANGULAR, SAWTOOTH, and TRAPEZOIDAL waveshapes. These waveshapes serve many useful purposes in the electronic circuits you will be studying. For example, they are used extensively throughout the television receiver to reproduce both picture and sound.

One type of wave generator is known as an OSCILLATOR. An oscillator can be regarded as an amplifier which provides its own input signal. Oscillators are classified according to the waveshapes they produce and the requirements needed for them to produce oscillations.

CLASSIFICATION OF OSCILLATORS (GENERATORS)

Wave generators can be classified into two broad categories according to their output waveshapes, SINUSOIDAL and NONSINUSOIDAL.

Sinusoidal Oscillators

A sinusoidal oscillator produces a sine-wave output signal. Ideally, the output signal is of constant amplitude with no variation in frequency. Actually, something less than this is usually obtained. The degree to which the ideal is approached depends upon such factors as class of amplifier operation, amplifier characteristics, frequency stability, and amplitude stability.

Sine-wave generators produce signals ranging from low audio frequencies to ultrahigh radio and microwave frequencies. Many low-frequency generators use resistors and capacitors to form their frequency-determining networks and are referred to as RC OSCILLATORS. They are widely used in the audio-frequency range.

Another type of sine-wave generator uses inductors and capacitors for its frequency-determining network. This type is known as the LC OSCILLATOR. LC oscillators, which use tank circuits, are commonly used for the higher radio frequencies. They are not suitable for use as extremely low-frequency oscillators because the inductors and capacitors would be large in size, heavy, and costly to manufacture.

A third type of sine-wave generator is the CRYSTAL-CONTROLLED OSCILLATOR. The crystal-controlled oscillator provides excellent frequency stability and is used from the middle of the audio range through the radio frequency range.

Nonsinusoidal Oscillators

Nonsinusoidal oscillators generate complex waveforms, such as square, rectangular, trigger, sawtooth, or trapezoidal. Because their outputs are generally characterized by a sudden change, or relaxation, they are often referred to as RELAXATION OSCILLATORS. The signal frequency of these oscillators is usually governed by the charge or discharge time of a capacitor in series with a resistor. Some types, however, contain inductors that affect the output frequency. Thus, like sinusoidal oscillators, both RC and LC networks are used for determining the frequency of oscillation. Within this category of nonsinusoidal oscillators are MULTIVIBRATORS, BLOCKING OSCILLATORS, SAWTOOTH GENERATORS, and TRAPEZOIDAL GENERATORS.

THE BASIC OSCILLATOR

An oscillator can be thought of as an amplifier that provides itself (through feedback) with an input signal. By definition, it is a nonrotating device for producing alternating current, the output frequency of which is determined by the characteristics of the device. The primary purpose of an oscillator is to generate a given waveform at a constant peak amplitude and specific frequency and to maintain this waveform within certain limits of amplitude and frequency.

An oscillator must provide amplification. Amplification of signal power occurs from input to output. In an oscillator, a portion of the output is fed back to sustain the input, as shown in figure 2-1. Enough power must be fed back to the input circuit for the oscillator to drive itself as does a signal generator. To cause the oscillator to be self-driven, the feedback signal must also be

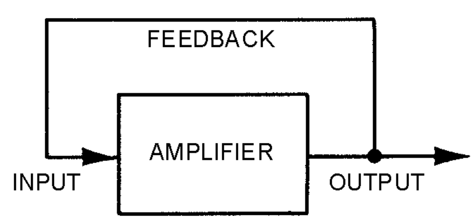


Figure 2-1.—Basic oscillator block diagram.

REGENERATIVE (positive). Regenerative signals must have enough power to compensate for circuit losses and to maintain oscillations.

Since a practical oscillator must oscillate at a predetermined frequency, a FREQUENCY-DETERMINING DEVICE (fdd), sometimes referred to as a FREQUENCY-DETERMINING NETWORK (fdn), is needed. This device acts as a filter, allowing only the desired frequency to pass. Without a frequency-determining device, the stage will oscillate in a random manner, and a constant frequency will not be maintained.

Before discussing oscillators further, let's review the requirements for an oscillator. First, amplification is required to provide the necessary gain for the signal. Second, sufficient regenerative feedback is required to sustain oscillations. Third, a frequency-determining device is needed to maintain the desired output frequency.

The basic oscillator requirements, in addition to the application, determine the type of oscillator to be used. Let's consider some factors that account for the complexity and unique characteristics of oscillators.

Virtually every piece of equipment that uses an oscillator has two stability requirements, AMPLITUDE STABILITY and FREQUENCY STABILITY. Amplitude stability refers to the ability of the oscillator to maintain a constant amplitude in the output waveform. The more constant the amplitude of the output waveform, the better the amplitude stability. Frequency stability refers to the ability of the oscillator to maintain its operating frequency. The less the oscillator varies from its operating frequency, the better the frequency stability.

A constant frequency and amplitude can be achieved by taking extreme care to prevent variations in LOAD, BIAS, and COMPONENT CHARACTERISTICS. Load variations can greatly affect the amplitude and frequency stability of the output of an oscillator. Therefore, maintaining the load as constant as possible is necessary to ensure a stable output.

As you should know from your study of transistor biasing, bias variations affect the operating point of the transistor. These variations may alter the amplification capabilities of the oscillator circuits as well. A well-regulated power supply and a bias-stabilizing circuit are required to ensure a constant, uniform signal output.

As a result of changing temperature and humidity conditions, the value or characteristics of components such as capacitors, resistors, and transistors can change. The changes in these components also cause changes in amplitude and frequency.

Output power is another consideration in the use of oscillators. Generally speaking, high power is obtained at some sacrifice to stability. When both requirements are to be met, a low-power, stable oscillator can be followed by a higher-power BUFFER AMPLIFIER. The buffer provides isolation between the oscillator and the load to prevent changes in the load from affecting the oscillator.

If the oscillator stage must develop high power, efficiency becomes important. Many oscillators use class C bias to increase efficiency. Other types of oscillators may use class A bias when a high efficiency is not required but distortion must be kept at a minimum. Other classes of bias may also be used with certain oscillators.

SINE-WAVE OSCILLATOR

RC networks, LC tanks, and crystals may appear in sine-wave oscillator circuits. An amplifier can be made into a sine-wave oscillator by providing regenerative feedback through an RC network.

RC Network

Figure 2-2, view (A), shows the block diagram of an amplifier with an RC network through which regenerative feedback is provided. The RC network also acts as the frequency-determining device. View (B) shows a vector analysis of the signal E at various points in the circuit.

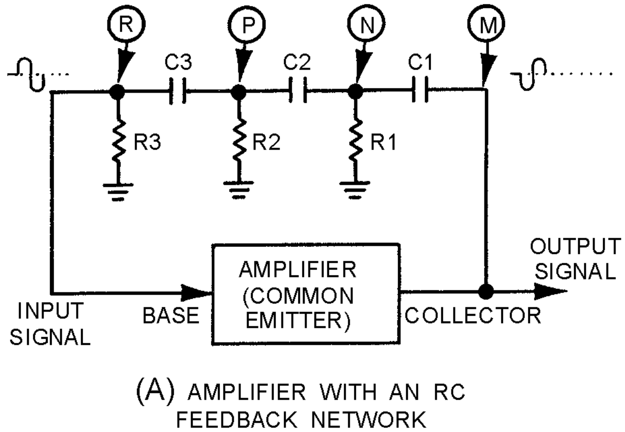


Figure 2-2A.—RC oscillator. AMPLIFIER WITH AND RC FEEDBACK NETWORK.

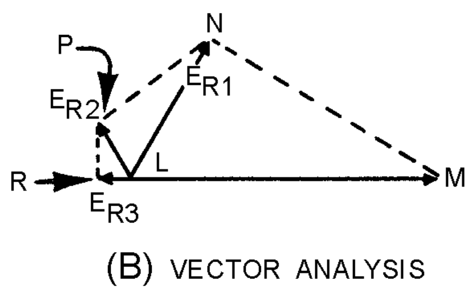


Figure 2-2B.—RC oscillator. VECTOR ANALYSIS

To analyze the operation of the circuit in view (A), assume that the amplifier is a common-emitter configuration. The signal on the collector (M) is 180 degrees out of phase with the signal (input) on the base (R). For the circuit to produce regenerative feedback, the RC network must provide a 180-degree phase shift of the collector signal. When power is applied to the circuit, a noise voltage (noise contains many different frequencies) will appear on the collector. This noise signal is represented by vector LM in view (B). As the signal couples through C1 and across R1 (view (A)), a phase shift occurs. The voltage across R1 (E_{R1}), represented by vector LN, has been shifted in phase (about 60 degrees) and reduced in amplitude. The signal at point N (view (A)) is then coupled to the next RC section (R2 and C2). Using the same size resistor and capacitor as before will cause another 60-degree phase shift to take place. The signal at point P is the voltage across R2, represented by vector LP. Now the signal at point P has been shifted about 120 degrees and its amplitude is reduced still further. The same actions occur for the last section (R3 and C3). This signal experiences another 60-degree phase shift and has further amplitude reduction. The signal at point R (E_{R3}) has been shifted 180 degrees and is represented by vector LR.

Notice that point R is the input to the base of the common-emitter amplifier. Also, vector LR shows that the signal on the base is regenerative (aiding the circuit operation). This meets the regenerative feedback requirement. An exact 60-degree phase shift per stage is not required, but the sum of the three phase shifts must equal 180 degrees.

For a given RC network, only one frequency of the initial noise signal will be shifted exactly 180 degrees. In other words, the network is frequency selective. Therefore, the RC network is the frequency-determining device since the lengths of the vectors and their phase relationships depend on frequency. The frequency of oscillations is governed by the values of resistance and capacitance in these sections. Variable resistors and capacitors may be used to provide tuning in the feedback network to allow for minor variations in phase shift. For an RC phase-shift oscillator, the amplifier is biased for class A operation to minimize distortion of the wave or signal.

LC Network

Some sine-wave oscillators use resonant circuits consisting of inductance and capacitance. For example, recall the tank circuit in which a resonant circuit stores energy alternately in the inductor and capacitor, producing a sine wave. You studied this action of the tank circuit in chapter 1.

If there were absolutely no internal resistances in a tank circuit, oscillations would continue indefinitely, as shown in figure 2-3, view (A). Each resonant circuit does, however, contain some resistance which dissipates power. This power loss causes the amplitude to decrease, as shown in views (B) and (C). The reduction of amplitude in an oscillator circuit is referred to as DAMPING. Damping is caused by both tank and load resistances. The larger the tank resistance, the greater the amount of damping. Loading the tank causes the same effect as increasing the internal resistance of the tank. The effect of this damping can be overcome by applying regenerative feedback.

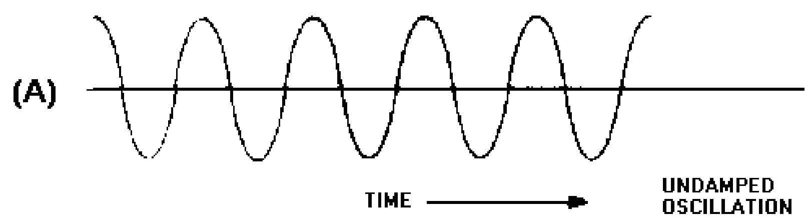


Figure 2-3A.—Effects of damping.

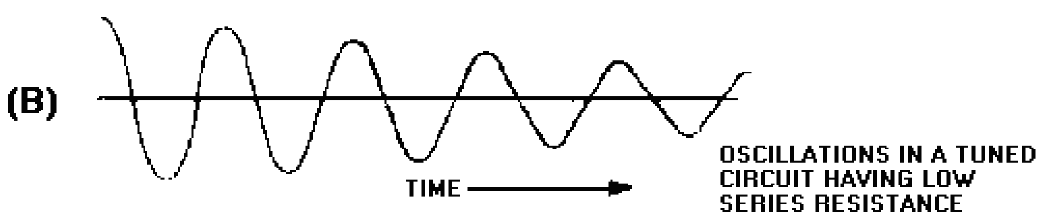


Figure 2-3B.—Effects of damping.

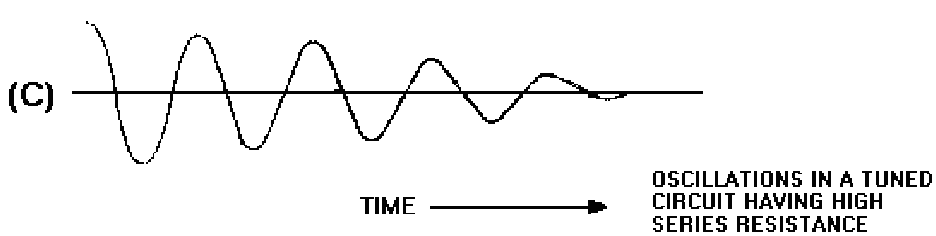


Figure 2-3C.—Effects of damping.

Figure 2-4 shows a block diagram of a typical LC oscillator. Notice that the oscillator contains the three basic requirements for sustained oscillations: amplification, a frequency-determining device, and regenerative feedback.

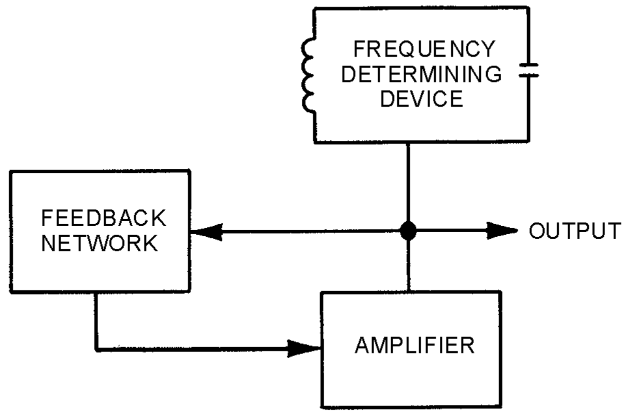


Figure 2-4.—LC oscillator.

The amplifier supplies energy to begin what is known as the FLYWHEEL EFFECT. The flywheel effect is the maintenance of oscillations in a circuit in the intervals between pulses of excitation energy. Recall that in chapter 1 the tank circuit alternately stored energy in the inductor and capacitor. The LC network provides initial oscillations. A portion of the output of the LC network is then returned to the input of the amplifier through the regenerative-feedback network to sustain the oscillations.

When a tank circuit is used to develop oscillations in an oscillator, the output frequency of the oscillator is primarily the resonant frequency of the tank circuit and can be found by the formula:

$$f_r = \frac{1}{2\pi\sqrt{LC}}$$

Crystals

Another frequency-determining device is the CRYSTAL. The crystal may be used with a tank circuit, or it may perform alone. Crystals exhibit a characteristic known as the PIEZOELECTRIC EFFECT. The piezoelectric effect is the property of a crystal by which mechanical forces produce electrical charges and, conversely, electrical charges produce mechanical forces. This effect is a form of oscillation similar to the flywheel effect of a tank circuit.

The piezoelectric effect can be seen in a number of crystal substances. The most important of these are the minerals quartz and Rochelle salt. Although quartz does not exhibit the piezoelectric effect to the degree that Rochelle salt does, quartz is used for frequency control in oscillators because of its greater mechanical strength. Another mineral, tourmaline, is physically strong like quartz; but because it is more expensive, it is not used extensively as an fdd. This discussion will deal only with the quartz crystal.

The crystals used in oscillator circuits are thin sheets, or wafers, cut from natural or synthetic quartz and ground to a specific thickness to obtain the desired resonant frequency. The crystals are mounted in holders, which support them physically and provide electrodes by which voltage is applied. The holder must allow the crystals freedom for vibration. There are many different types of holders. One type is shown in figure 2-5.

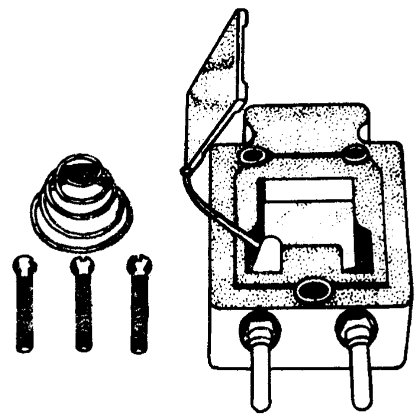


Figure 2-5.—Crystal holder.

The frequency for which a crystal is ground is referred to as the NATURAL RESONANT FREQUENCY of the crystal. Voltage applied to the crystal produces mechanical vibrations which, in turn, produce an output voltage at the natural resonant frequency of the crystal. A vibrating crystal can be represented by an equivalent electrical circuit composed of capacitance, inductance, and resistance.

Figure 2-6, view (A), illustrates the symbol of a crystal; view (B) shows an equivalent circuit for the crystal. View (C) shows an equivalent circuit for the crystal and the holder; C1 represents the capacitance between the metal plates of the holder.

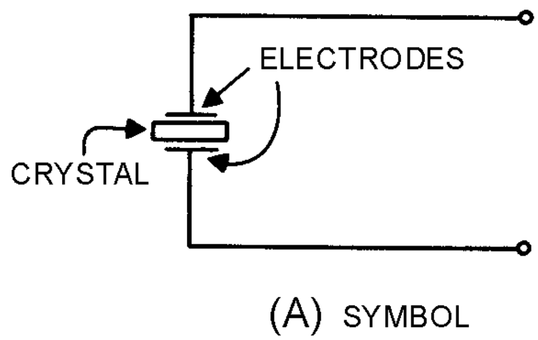


Figure 2-6A.—Crystal symbol and equivalent circuits. SYMBOL

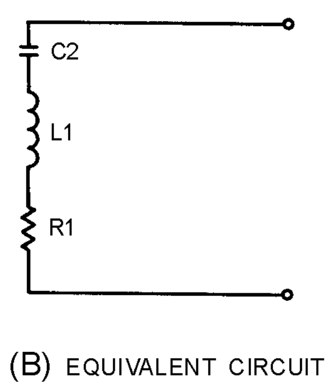


Figure 2-6B.—Crystal symbol and equivalent circuits. EQUIVALENT CIRCUIT.

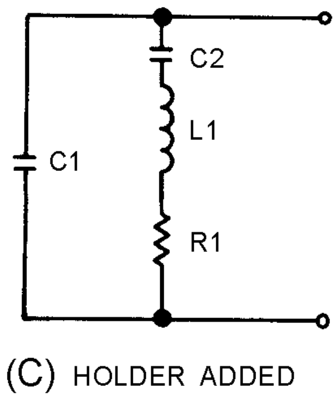


Figure 2-6C.—Crystal symbol and equivalent circuits. HOLDER ADDED

The Q (discussed in chapter 1) of a crystal is many times greater than that of an LC tank circuit. The high Q is present because the resistance in the crystal is extremely small. Commercially produced crystals range in Q from 5,000 to 30,000. The high Q causes the frequency stability to be much greater than that of an ordinary LC tank circuit. This is the reason a crystal is used in many sine-wave generator circuits.

Q-1. What are the two classifications of wave generators according to their output waveshapes?

Q-2. What are the three networks used for frequency-determining devices?

Q-3. What is another name for nonsinusoidal oscillators?

Q-4. What is a nonrotating device that produces alternating current?

Q-5. What are the three requirements necessary for oscillations to exist in a circuit?

SOLID-STATE LC OSCILLATORS

As you have just studied, a basic oscillator can be broken down into three main sections: a frequency-determining device, an amplifier, and a feedback circuit. The frequency-determining device in an LC oscillator is usually an LC tank circuit. Although the tank circuit is normally found in the input circuit of an oscillator (both electron tube and transistor), it sometimes appears in the output circuit. The differences in magnitude of plate and collector currents and shunting impedances are considerations in the designed locations of such tank circuits. In both solid-state and electron tube circuits, oscillations take place in the tuned circuit. Both the electron tube and the transistor function primarily as electrical valves that amplify and automatically deliver to the input circuit the proper amount of energy to sustain oscillations. In both tube and transistor oscillators, the feedback circuit couples energy of the proper amount and of the correct phase from the output to the input circuit to sustain oscillations.

FEEDBACK

Let's review what you have studied up to this point concerning feedback. Feedback is the process of transferring energy from a high-level point in a system to a low-level point in a system. This means transferring energy from the output of an amplifier back to its input. If the output feedback signal opposes the input signal, the signal is DEGENERATIVE or NEGATIVE FEEDBACK. However, if the feedback aids the input signal, the feedback is REGENERATIVE or POSITIVE FEEDBACK. Regenerative or positive feedback is one of the requirements to sustain oscillations in an oscillator. This feedback can be applied in any of several ways to produce a practical oscillator circuit.

TYPES OF FEEDBACK

Chapter 1 described the resonant or tank circuit and how a sinusoidal signal is generated by the action of an inductor and a capacitor. The feedback signal is coupled from this circuit by either of two means. The first method is to take some of the energy from the inductor. This can be done by any one of the three ways shown in figure 2-7, views (A), (B), and (C). When an oscillator uses a TICKLER COIL, as shown in view (A), it is referred to as an ARMSTRONG OSCILLATOR. When an oscillator uses a tapped coil (view (B)) or a split coil (view (C)), it is referred to as a HARTLEY OSCILLATOR. The second method of coupling the feedback signal is to use two capacitors in the tank circuit and tap the feedback signal between them. This is shown in view (D). An oscillator using this method is referred to as a COLPITTS OSCILLATOR. Each of these particular oscillators is named after the person who originally designed them.

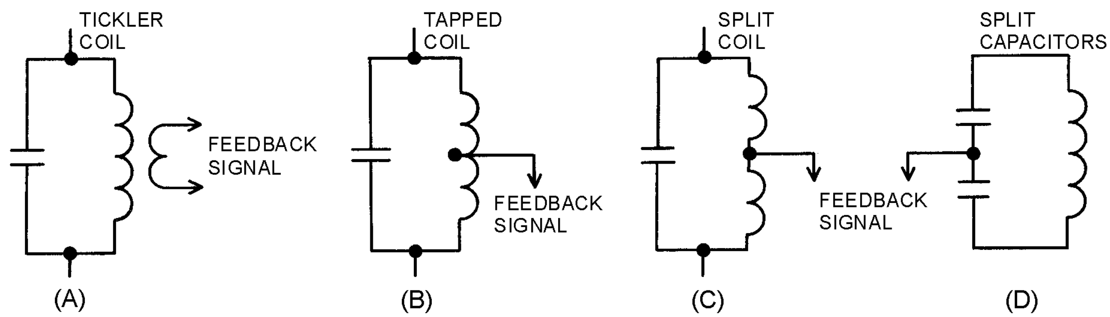


Figure 2-7.—Feedback signals.

CONFIGURATION OF OSCILLATORS

Any of the three basic amplifier configurations (common collector, common base, or common emitter) described in NEETS, Module 7, *Introduction to Solid-State Devices and Power Supplies*, Chapter 2, may be used for the oscillator circuit. However, certain considerations in the application of the circuit, such as the operating frequency and output power required, usually determine which of the three configurations is to be used. The three basic configurations are shown in figure 2-8, views (A), (B), and (C).

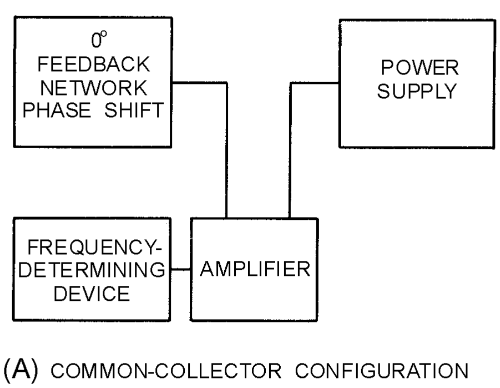


Figure 2-8A.—Basic configurations. COMMON-COLLECTOR CONFIGURATION

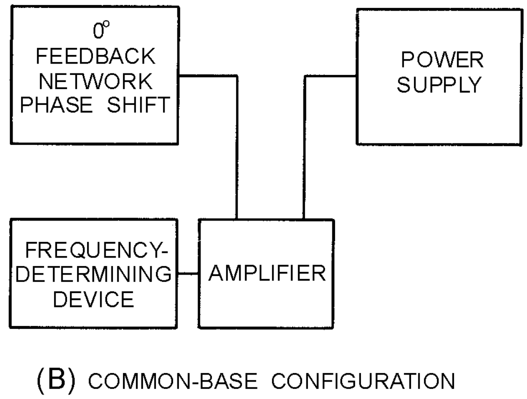


Figure 2-8B.—Basic configurations. COMMON-BASE CONFIGURATION.

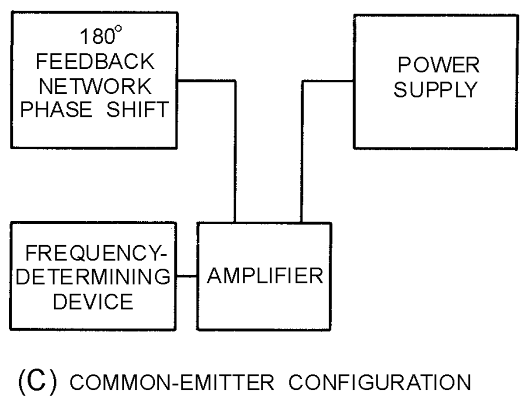


Figure 2-8C.—Basic configurations. COMMON-EMITTER CONFIGURATION.

COMMON-COLLECTOR CONFIGURATION

Since there is no phase reversal between the input and output circuits of a common-collector configuration, the feedback network does not need to provide a phase shift. However, since the voltage gain is less than unity and the power gain is low, the common-collector configuration is very seldom used in oscillator circuits.

COMMON-BASE CONFIGURATION

The power gain and voltage gain of the common-base configuration are high enough to give satisfactory operation in an oscillator circuit. The wide range between the input resistance and the output resistance make impedance matching slightly harder to achieve in the common-base circuit than in the common-emitter circuit. An advantage of the common-base configuration is that it exhibits better high-frequency response than does the common-emitter configuration.

COMMON-EMITTER CONFIGURATION

The common-emitter configuration has high power gain and is used in low-frequency applications. For the energy which is fed back from the output to be in phase with the energy at the input, the feedback network of a common-emitter oscillator must provide a phase shift of approximately 180 degrees. An advantage of the common-emitter configuration is that the medium resistance range of the input and output simplifies the job of impedance matching.

Q-6. What type of feedback aids an input signal?

Q-7. What are the two methods used for feedback coupling?

Q-8. Which oscillator uses a tickler coil for feedback?

Q-9. Which oscillator uses a tapped inductor for feedback?

Q-10. Which oscillator uses tapped capacitors for feedback?

Q-11. What are the three basic configurations of transistor oscillators?

OSCILLATOR CIRCUITS

Oscillators may be classified by name, such as Armstrong, Hartley, Colpitts, or by the manner in which dc power is applied. An oscillator in which dc power is supplied to the transistor through the tank circuit, or a portion of the tank circuit, is said to be SERIES FED. An oscillator which receives its dc power for the transistor through a path separate and parallel to the tank circuit is said to be PARALLEL FED OR SHUNT FED. All the oscillators in this chapter can be constructed either way, series or shunt fed. The construction depends on the characteristics of the oscillator circuit the designer is interested in.

A SERIES-FED, TUNED-COLLECTOR ARMSTRONG OSCILLATOR is illustrated in figure 2-9, view (A). The dc path is from the negative side (ground) of V_{CC} through R_E, Q1, T1, and back to the positive side of V_{CC}. The figure clearly illustrates that both the ac and dc components flow through the tank circuit.

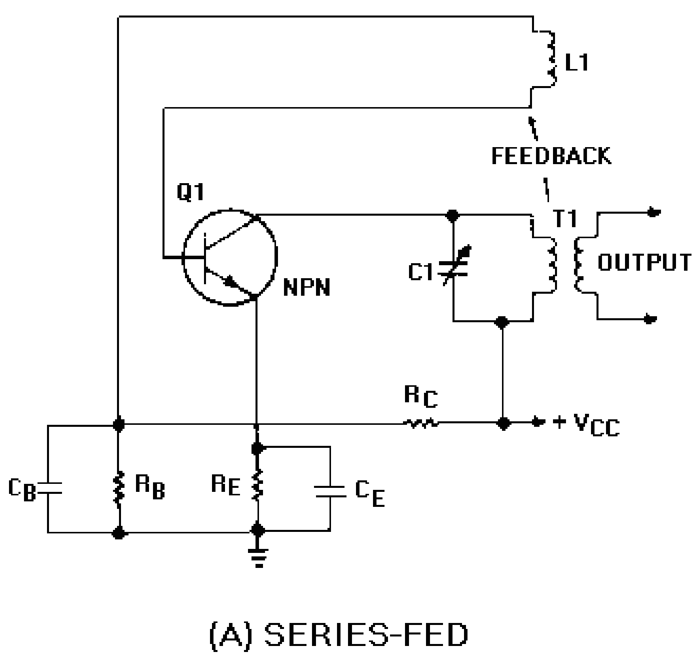


Figure 2-9A.—Series- and shunt-fed, tuned-collector Armstrong oscillators. SERIES-FED.

By modifying the circuit slightly, it becomes a SHUNT-FED, TUNED-COLLECTOR ARMSTRONG OSCILLATOR as shown in view (B). The dc component flows from ground through R_E to Q1 to positive V_{CC}. The dc is blocked from the tank circuit by capacitor C2. Only the ac component flows in the tank circuit.

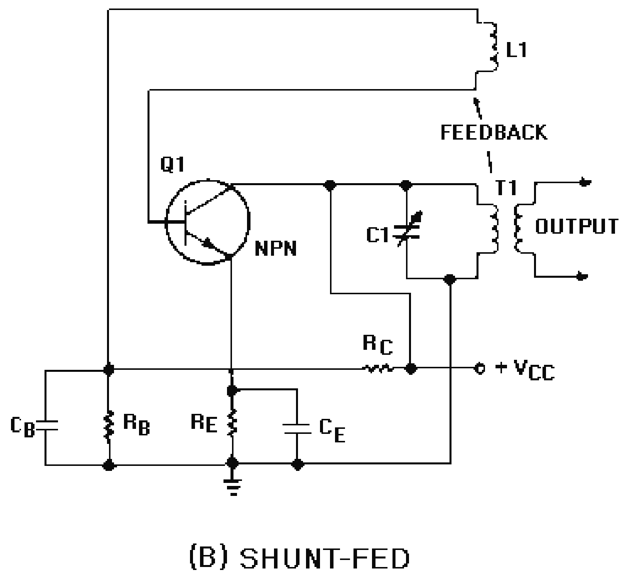


Figure 2-9B.—Series- and shunt-fed, tuned-collector Armstrong oscillators. SHUNT-FED.

The function of an oscillator is to produce a sinusoidal waveshape of a specific frequency and amplitude. In doing so, the stability of an oscillator is very important. Depending on its application, an oscillator may be required to have either good frequency stability or amplitude stability; in many circumstances, both are required. Of the two, good frequency stability is usually considered more important.

FREQUENCY STABILITY

The FREQUENCY STABILITY of an oscillator is a measure of the degree to which a constant frequency output is approached. The better the frequency stability, the closer the output will be to a constant frequency.

Frequency INSTABILITY (variations above and below the desired output frequency) may be caused by transistor characteristics or by variations in the external circuit elements.

As stated before, when output power is not of prime importance, transistor oscillators may be biased class A to ensure stability and minimize distortion. When this is done, the dc operating point established by the power supply is chosen so that the operation of the transistor oscillator occurs over the most linear portion of the transistor's characteristic curve. When the operation of the circuit falls into the nonlinear portion of the characteristic curve, the transistor's parameters (voltages and currents) vary. These parameters are basic to the stable frequency of the transistor oscillator. Operating frequency variations may occur with changes in these bias voltages. Thus, a constant supply voltage is a prime requirement for good frequency stability.

The use of a common bias source for both collector and emitter electrodes results in a relatively constant ratio of the two voltages. In effect, a change in one voltage is somewhat counteracted by the change in the other. This counteraction takes place because an increase in collector voltage causes an increase in the oscillating frequency, and an increase in emitter voltage causes a decrease in the oscillating frequency. This is a result of the change in capacitance between the junctions of the transistor. However, a common bias source does not completely compensate since the effects on other circuit parameters of each bias voltage differ.

Just as in any transistor circuit, changes in the transistor operating point and changes in temperature are encountered in the transistor oscillator. The effects of changes in temperature are to cause collector current to increase if the transistor is not stabilized. The increase in collector current can be prevented by reducing the forward bias.

AMPLITUDE STABILITY

The AMPLITUDE STABILITY of a transistor oscillator indicates the amount by which the actual output amplitude varies from the desired output amplitude.

The same parameters (voltages and currents) that affect frequency stability also affect amplitude stability. Output amplitude may be kept relatively constant by ensuring that the feedback is large enough that the collector current is maintained at the proper level. Feedback used in this manner makes the output voltage directly proportional to the supply voltage. Thus, regulation of the supply voltage ensures good amplitude stability.

ARMSTRONG OSCILLATOR

The ARMSTRONG OSCILLATOR is used to produce a sine-wave output of constant amplitude and of fairly constant frequency within the rf range. It is generally used as a local oscillator in receivers, as a source in signal generators, and as a radio-frequency oscillator in the medium- and high-frequency range.

The identifying characteristics of the Armstrong oscillator are that (1) it uses an LC tuned circuit to establish the frequency of oscillation, (2) feedback is accomplished by mutual inductive coupling between the tickler coil and the LC tuned circuit, and (3) it uses a class C amplifier with self-bias. Its frequency is fairly stable, and the output amplitude is relatively constant.

Views (A), (B), and (C) shown in figure 2-10 can be used to build the basic Armstrong oscillator. View (A) shows a conventional amplifier. R2 provides the forward bias for Q1, C2 is a coupling capacitor, and L1 and R1 form the collector load impedance. This is a common-emitter configuration which provides the 180-degree phase shift between the base and collector.

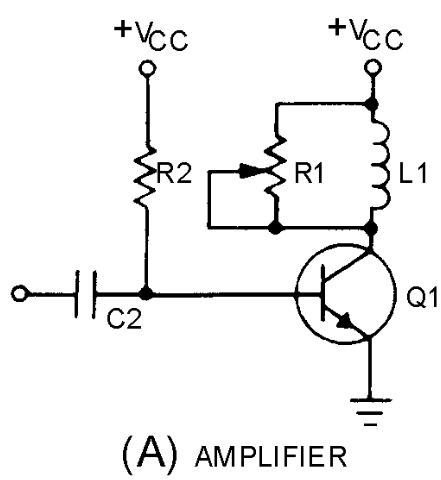
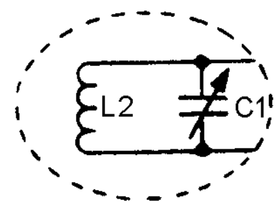
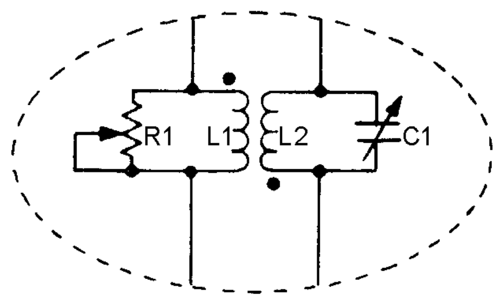


Figure 2-10A.—Basic Armstrong oscillator circuit. AMPLIFIER



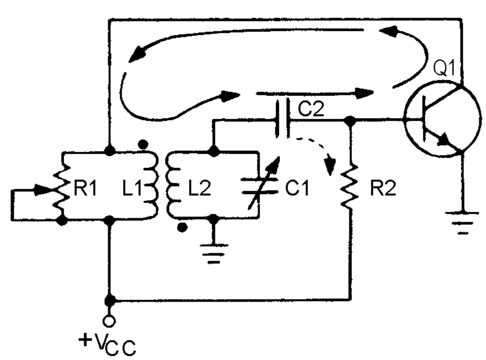
(B) FREQUENCY-DETERMINING DEVICE

Figure 2-10B.—Basic Armstrong oscillator circuit. FREQUENCY-DETERMINING DEVICE.



(C) FEEDBACK NETWORK

Figure 2-10C.—Basic Armstrong oscillator circuit. FEEDBACK NETWORK.



(D) OSCILLATOR

Figure 2-10D.—Basic Armstrong oscillator circuit. OSCILLATOR.

View (B) shows the frequency-determining device composed of inductor L2 and capacitor C1. C1 is a variable tuning capacitor which is used to adjust the resonant frequency to the desired value.

View (C) is the feedback network which uses L1 (the collector load) as the primary and L2 as the secondary winding of a coupling transformer to provide a 180-degree phase shift. Variable resistor R1 controls the amount of current through L1. When R1 is adjusted for maximum resistance, most of the current flows through L1. The transformer now couples a maximum signal which represents a large feedback amplitude into the tank circuit (L2, C1). If R1 is adjusted for a smaller resistance, less current flows through L1, and less energy is coupled to the tank circuit; therefore, feedback amplitude decreases. R1 is normally adjusted so that the L1 current is adequate to sustain tank oscillations.

View (D) shows the complete oscillator circuit. Connecting the feedback network through coupling capacitor C2 to the base of Q1 forms a "closed loop" for feedback (shown by the solid arrows). Let's verify that the feedback is regenerative. Assume a positive signal on the base of Q1. The transistor conducts heavily when forward biased. This current flow through L1 and R1 causes the voltage across L1 to increase. The voltage increase is inductively coupled to L2 and inverted. This action ensures that the voltage is positive at the base end of L2 and C1 and in phase with the base voltage. The positive signal is now coupled through C2 to the base of Q1. The regenerative feedback offsets the damping in the frequency-determining network and has sufficient amplitude to provide unity circuit gain.

The circuit in view (D) has all three requirements for an oscillator: (1) amplification, (2) a frequency-determining device, and (3) regenerative feedback. The oscillator in this schematic drawing is a tuned-base oscillator, because the fdd is in the base circuit. If the fdd were in the collector circuit, it would be a tuned-collector oscillator. The circuit in view (D) is basically an Armstrong oscillator.

Refer to figure 2-10, view (D), for the following discussion of the circuit operation of the Armstrong oscillator. When V_{CC} is applied to the circuit. a small amount of base current flows through R2 which sets the forward bias on Q1. This forward bias causes collector current to flow from ground through Q1, R1, and L1 to $+V_{CC}$. The current through L1 develops a magnetic field which induces a voltage into the tank circuit. The voltage is positive at the top of L2 and C1. At this time, two actions occur. First, resonant tank capacitor C1 charges to this voltage; the tank circuit now has stored energy. Second, coupling capacitor C2 couples the positive signal to the base of Q1. With a positive signal on its base, Q1 conducts harder. With Q1 conducting harder, more current flows through L1, a larger voltage is induced into L2, and a larger positive signal is coupled back to the base of Q1. While this is taking place, the frequency-determining device is storing more energy and C1 is charging to the voltage induced into L2.

The transistor will continue to increase in conduction until it reaches saturation. At saturation, the collector current of Q1 is at a maximum value and cannot increase any further. With a steady current through L1, the magnetic fields are not moving and no voltage is induced into the secondary.

With no external voltage applied, C1 acts as a voltage source and discharges. As the voltage across C1 decreases, its energy is transferred to the magnetic field of L2. Now, let's look at C2.

The coupling capacitor, C2, has charged to approximately the same voltage as C1. As C1 discharges, C2 discharges. The primary discharge path for C2 is through R2 (shown by the dashed arrow). As C2 discharges, the voltage drop across R2 opposes the forward bias on Q1 and collector current begins to decrease. This is caused by the decreasing positive potential at the base of Q1.

A decrease in collector current allows the magnetic field of L1 to collapse. The collapsing field of L1 induces a negative voltage into the secondary which is coupled through C2 and makes the base of Q1 more negative. This, again, is regenerative action; it continues until Q1 is driven into cutoff.

When Q1 is cut off, the tank circuit continues to flywheel, or oscillate. The flywheel effect not only produces a sine-wave signal, but it aids in keeping Q1 cut off. Without feedback, the oscillations of L2 and C1 would dampen out after several cycles.

To ensure that the amplitude of the signal remains constant, regenerative feedback is supplied to the tank once each cycle, as follows: As the voltage across C1 reaches maximum negative, C1 begins discharging toward 0 volts. Q1 is still below cutoff. C1 continues to discharge through 0 volts and becomes positively charged. The tank circuit voltage is again coupled to the base of Q1, so the base voltage becomes positive and allows collector current to flow. The collector current creates a magnetic field in L1, which is coupled into the tank. This feedback action replaces any lost energy in the tank circuit and drives Q1 toward saturation. After saturation is reached, the transistor is again driven into cutoff.

The operation of the Armstrong oscillator is basically this: Power applied to the transistor allows energy to be applied to the tank circuit causing it to oscillate. Once every cycle, the transistor conducts for a short period of time (class C operation) and returns enough energy to the tank to ensure a constant amplitude output signal.

Class C operation has high efficiency and low loading characteristics. The longer Q1 is cut off, the less the loading on the frequency-determining device.

Figure 2-11 shows a tuned-base Armstrong oscillator as you will probably see it. R3 has been added to improve temperature stability. Bypass capacitor C3 prevents degeneration. C4 is an output coupling capacitor, and impedance-matching transformer T2 provides a method of coupling the output signal. T2 is usually a loosely coupled rf transformer which reduces undesired reflected impedance from the load back to the oscillator.

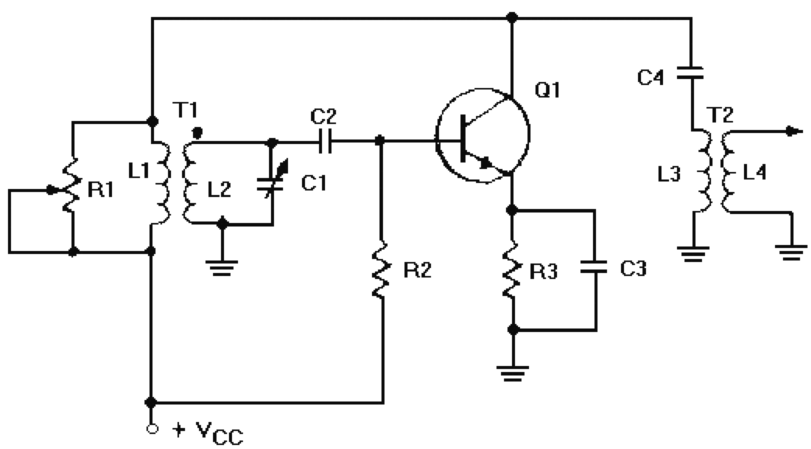


Figure 2-11.—Tuned-base Armstrong oscillator.

The Armstrong oscillator is an example of how a class C amplifier can produce a sine-wave output that is not distorted. Although class C operation is nonlinear and many harmonic frequencies are generated, only one frequency receives enough gain to cause the circuit to oscillate. This is the frequency of the resonant tank circuit. Thus, high efficiency and an undistorted output signal can be obtained.

The waveforms in figure 2-12 illustrate the relationship between the collector voltage and collector current. Notice that collector current (I_C) flows for only a short time during each cycle. While the tank circuit is oscillating (figure 2-11), L2 acts as the primary of the transformer and L1 acts as the secondary. The signal from the tank is, therefore, coupled through T1 to coupling capacitor C4, and the output voltage across L4 is a sine wave.

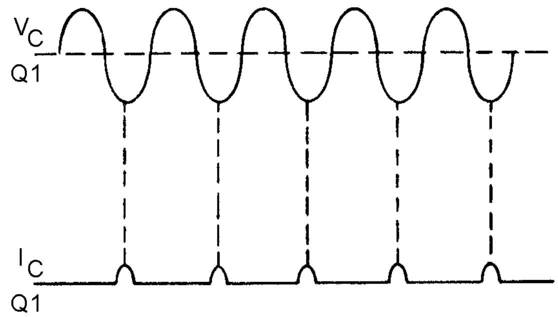


Figure 2-12.—Collector current and voltage waveforms of a class C oscillator.

HARTLEY OSCILLATOR

The HARTLEY OSCILLATOR is an improvement over the Armstrong oscillator. Although its frequency stability is not the best possible of all the oscillators, the Hartley oscillator can generate a wide range of frequencies and is very easy to tune. The Hartley will operate class C with self-bias for ordinary operation. It will operate class A when the output waveform must be of a constant voltage level or of a linear waveshape. The two versions of this oscillator are the series-fed and the shunt-fed. The main difference between the Armstrong and the Hartley oscillators lies in the design of the feedback (tickler) coil. A separate coil is not used. Instead, in the Hartley oscillator, the coil in the tank circuit is a split inductor. Current flow through one section induces a voltage in the other section to develop a feedback signal.

Series-Fed Hartley Oscillator

One version of a SERIES-FED HARTLEY OSCILLATOR is shown in figure 2-13. The tank circuit consists of the tapped coil (L1 and L2) and capacitor C2. The feedback circuit is from the tank circuit to the base of Q1 through the coupling capacitor C1. Coupling capacitor C1 prevents the low dc resistance of L2 from placing a short across the emitter-to-base junction and resistor R_E. Capacitor C3 bypasses the sine-wave signal around the battery, and resistor R_E is used for temperature stabilization to prevent thermal runaway. Degeneration is prevented by C_E in parallel with R_E. The amount of bias is determined by the values of R_B, the emitter-to-base resistance, the small amount of dc resistance of coil L1, and the resistance of R_E.

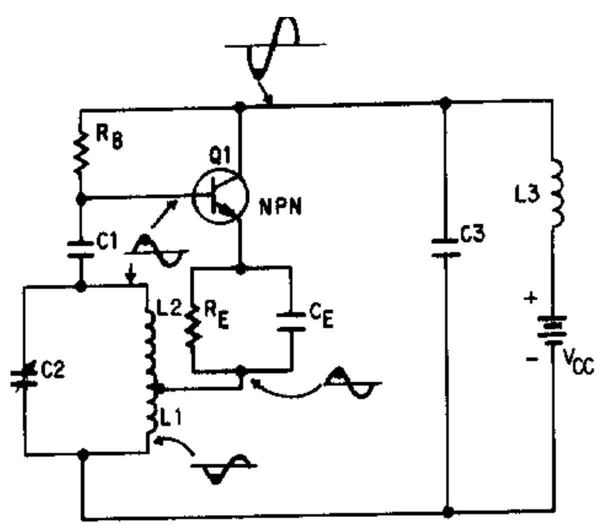


Figure 2-13.—Series-fed, tuned-base Hartley oscillator.

When a voltage is applied to the circuit, current from the battery flows through coil L1 and to the emitter through R_E. Current then flows from the emitter to the collector and back to the battery. The surge of current through coil L1 induces a voltage in coil L2 to start oscillations within the tank circuit.

When current first starts to flow through coil L1, the bottom of L1 is negative with respect to the top of L2. The voltage induced into coil L2 makes the top of L2 positive. As the top of L2 becomes positive, the positive potential is coupled to the base of Q1 by capacitor C1. A positive potential on the base results in an increase of the forward bias of Q1 and causes collector current to increase. The increased collector current also increases the emitter current flowing through coil L1. Increased current through L1 results in more energy being supplied to the tank circuit, which, in turn, increases the positive potential at the top of the tank (L2) and increases the forward bias of Q1. This action continues until the rate of current change through coil L1 can no longer increase. The current through coil L1 and the transistor cannot continue increasing indefinitely, or the coil and transistor will burn up. The circuit must be designed, by proper selection of the transistor and associated parts, so that some point is reached when the current can no longer continue to increase. At this point C2 has charged to the potential across L1 and L2. This is shown as the heavy dot on the base waveform. As the current through L1 decreases, the voltage induced in L2 decreases. The positive potential across the tank begins to decrease and C2 starts discharging through L1 and L2. This action maintains current flow through the tapped coil and causes a decrease in the forward bias of Q1. In turn, this decrease in the forward bias of Q1 causes the collector and emitter current to decrease. At the instant the potential across the tank circuit decreases to 0, the energy of the tank circuit is contained in the magnetic field of the coil. The oscillator has completed a half cycle of operation.

Next, the magnetic field around L2 collapses as the current from C2 stops. The action of the collapsing magnetic field causes the top of L2 to become negative at this instant. The negative charge causes capacitor C2 to begin to charge in the opposite direction. This negative potential is coupled to the base of Q1, opposing its forward bias. Most transistor oscillators are operated class A; therefore, the positive and negative signals applied to the base of Q1 will not cause it to go into saturation or cutoff. When the tank circuit reaches its maximum negative value, the collector and the emitter currents will still be present but at a minimum value. The magnetic field will have collapsed and the oscillator will have completed 3/4 cycle.

At this point C2 begins to discharge, decreasing the negative potential at the top of L2 (potential will swing in the positive direction). As the negative potential applied to the base of Q1 decreases, the opposition to the forward bias also decreases. This, in effect, causes the forward bias to begin increasing, resulting in increased emitter current flowing through L1. The increase in current through L1 causes additional energy to be fed to the tank circuit to replace lost energy. If the energy lost in the tank is replaced with an equal or larger amount of energy, oscillations will be sustained. The oscillator has now completed 1 cycle and will continue to repeat it over and over again.

Shunt-Fed Hartley Oscillator

A version of a SHUNT-FED HARTLEY OSCILLATOR is shown in figure 2-14. The parts in this circuit perform the same basic functions as do their counterparts in the series-fed Hartley oscillator. The difference between the series-fed and the shunt-fed circuit is that dc does not flow through the tank circuit. The shunt-fed circuit operation is essentially the same as the series-fed Hartley oscillator. When voltage is applied to the circuit, Q1 starts conducting. As the collector current of Q1 increases, the change (increase) is coupled through capacitor C3 to the tank circuit, causing it to oscillate. C3 also acts as an isolation capacitor to prevent dc from flowing through the feedback coil. The oscillations at the collector will be coupled through C3 (feedback) to supply energy lost within the tank.

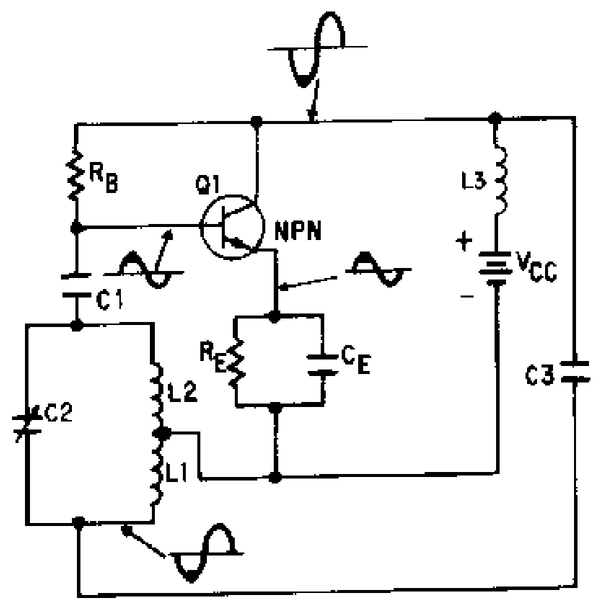


Figure 2-14.—Shunt-fed, tuned-base Hartley oscillator.

Q-12. What is the main difference between the Armstrong oscillator and the Hartley oscillator?

Q-13. What is the difference between the series-fed and the shunt-fed Hartley oscillator?

COLPITTS OSCILLATOR

Both the Armstrong and the Hartley oscillators have a tendency to be unstable in frequency because of junction capacitance. In comparison, the COLPITTS OSCILLATOR has fairly good frequency stability, is easy to tune, and can be used for a wide range of frequencies. The large value of split capacitance is in parallel with the junctions and minimizes the effect on frequency stability.

The Colpitts oscillator is very similar to the shunt-fed Hartley oscillator, except that two capacitors are used in the tank circuit instead of a tapped coil (figure 2-15). The Hartley oscillator has a tap between two coils, while the Colpitts has a tap between two capacitors. You can change the frequency of the Colpitts either by varying the inductance of the coil or by varying the capacitance of the two capacitors in the tank circuit. Notice that no coupling capacitor is used between the tank circuit and the base of Q1. Capacitors C1 and C2 of the tank circuit are in parallel with the input and the output interelement capacitance (capacitance between emitter, base, and collector) of the transistor. Thus the input and the output capacitive effect can be minimized on the tank circuit and better frequency stability can be obtained than with the Armstrong or the Hartley oscillator.

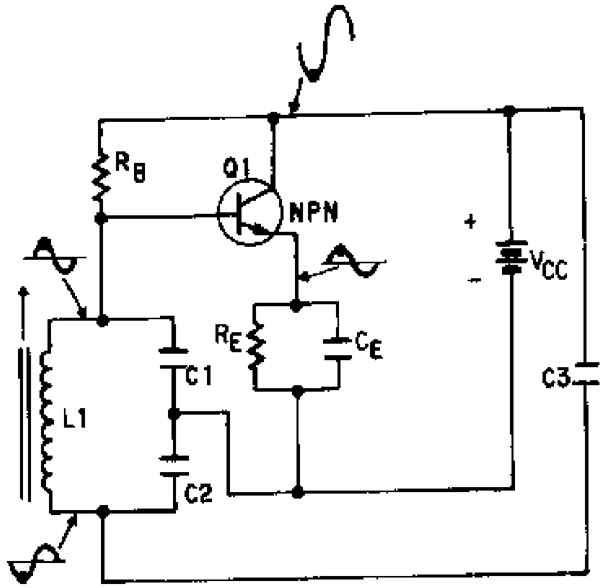


Figure 2-15.—Colpitts oscillator.

Figure 2-16 shows a common-base Colpitts oscillator using a pnp transistor as the amplifying device. Notice in this version of the Colpitts oscillator that regenerative feedback is obtained from the tank circuit and applied to the emitter. Base bias is provided by resistor R_B and R_F. Resistor R_C is the collector load resistor. Resistor R_E develops the input signal and also acts as the emitter swamping resistor. The tuned circuit consists of C1 and C2 in parallel with the primary winding of transformer T1. The voltage developed across C2 is the feedback voltage. Either or both capacitors may be adjusted to control the frequency. In the common-base configuration there is no phase difference between the signal at the collector and the emitter signal. Therefore, the phase of the feedback signal does not have to be changed. When the emitter swings negative, the collector also swings negative and C2 charges negatively at the junction of C1 and C2. This negative charge across C2 is fed back to the emitter. This increases the reverse bias on Q1. The collector of Q1 becomes more negative and C2 charges to a negative potential. This feedback effect continues until the collector of Q1 is unable to become any more negative. At that time the primary of T1 will act as a source because of normal tank circuit operation. As its field collapses, the tank potential will reverse and C1 and C2 will begin to discharge. As C2 becomes less negative, the reverse bias on Q1 decreases and its collector voltage swings in the positive direction. C1 and C2 will continue to discharge and then charge in a positive direction. This positive-going voltage across C2 will be fed back to the emitter as regenerative feedback. This will continue until the field around the primary of T1 collapses. At that time the collector of Q1 will be at a maximum positive value. C1 and C2 will begin to discharge and the potential at their junction will become less positive. This increases the reverse bias on Q1 and drives the collector negative, causing C1 and C2 to charge in a negative direction and to repeat the cycle.

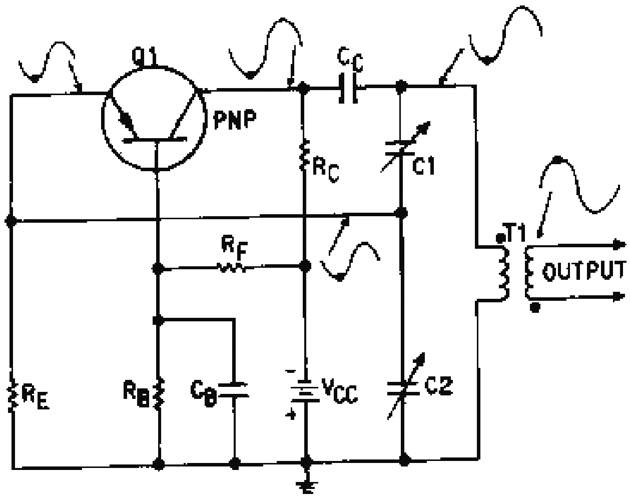


Figure 2-16.—Common-base Colpitts oscillator.

Q-14. What is the identifying feature of a Colpitts oscillator?

RESISTIVE-CAPACITIVE (RC) FEEDBACK OSCILLATOR

As mentioned earlier, resistive-capacitive (RC) networks provide regenerative feedback and determine the frequency of operation in RESISTIVE-CAPACITIVE (RC) OSCILLATORS.

The oscillators presented in this chapter have used resonant tank circuits (LC). You should already know how the LC tank circuit stores energy alternately in the inductor and capacitor.

The major difference between the LC and RC oscillator is that the frequency-determining device in the RC oscillator is not a tank circuit. Remember, the LC oscillator can operate with class A or C biasing because of the oscillator action of the resonant tank. The RC oscillator, however, must use class A biasing because the RC frequency-determining device doesn't have the oscillating ability of a tank circuit.

An RC FEEDBACK or PHASE-SHIFT oscillator is shown in figure 2-17. Components C1, R1, C2, R2, C3, and R_B are the feedback and frequency-determining network. This RC network also provides the needed phase shift between the collector and base.

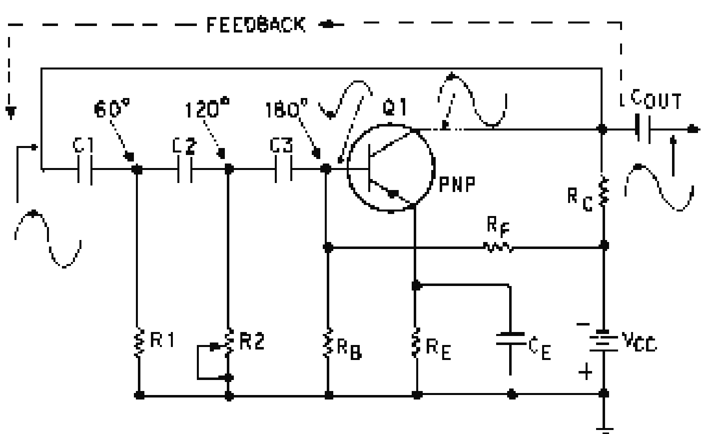


Figure 2-17.—Phase-shift oscillator.

Phase-Shift Oscillators

The PHASE-SHIFT OSCILLATOR, shown in figure 2-17, is a sine-wave generator that uses a resistive-capacitive (RC) network as its frequency-determining device.

As discussed earlier in the common-emitter amplifier configuration (figure 2-17), there is a 180-degree phase difference between the base and the collector signal. To obtain the regenerative feedback in the phase-shift oscillator, you need a phase shift of 180 degrees between the output and the input signal. An RC network consisting of three RC sections provides the proper feedback and phase inversion to provide this regenerative feedback. Each section shifts the feedback signal 60 degrees in phase.

Since the impedance of an RC network is capacitive, the current flowing through it leads the applied voltage by a specific phase angle. The phase angle is determined by the amount of resistance and capacitance of the RC section.

If the capacitance is a fixed value, a change in the resistance value will change the phase angle. If the resistance could be changed to zero, we could get a maximum phase angle of 90 degrees. But since a voltage cannot be developed across zero resistance, a 90-degree phase shift is not possible.

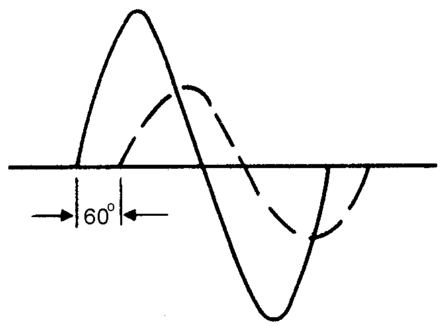
With a small value of resistance, however, the phase angle or phase shift is less than 90 degrees. In the phase-shift oscillator, therefore, at least three RC sections are needed to give the required 180-degree phase shift for regenerative feedback. The values of resistance and capacitance are generally chosen so that each section provides about a 60-degree phase shift.

Resistors R_B, R_F, and R_C provide base and collector bias. Capacitor C_E bypasses ac variations around the emitter resistor R_E. Capacitors C1, C2, and C3 and resistors R1, R2, and R_B form the feedback and phase-shifting network. Resistor R2 is variable for fine tuning to compensate for any small changes in value of the other components of the phase-shifting network.

When power is applied to the circuit, oscillations are started by any random noise (random electrical variations generated internally in electronic components). A change in the flow of base current results in an amplified change in collector current which is phase-shifted the 180 degrees. When the signal is returned to the base, it has been shifted 180 degrees by the action of the RC network, making the circuit regenerative. View (A) of figure 2-18 shows the amount of phase shift produced by C1 and R1. View (B) shows the amount of phase shift produced by C2 and R2 (signal received from C1 and R1), and view (C) shows the complete phase shift as the signal leaves the RC network. With the correct amount of resistance and capacitance in the phase-shifting network, the 180-degree phase shift occurs at only one frequency. At any other than the desired frequency, the capacitive reactance increases or decreases and causes an incorrect phase relationship (the feedback becomes degenerative). Thus, the oscillator works at only one frequency. To find the resonant frequency (f_r) of an RC phase shift oscillator, use the following formula:

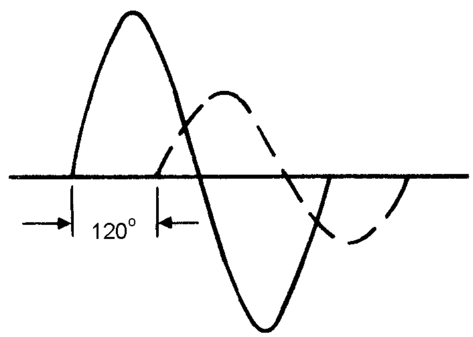
$$f_r = \frac{1}{2\pi RC\sqrt{2n}}$$

where n is the number of RC sections.



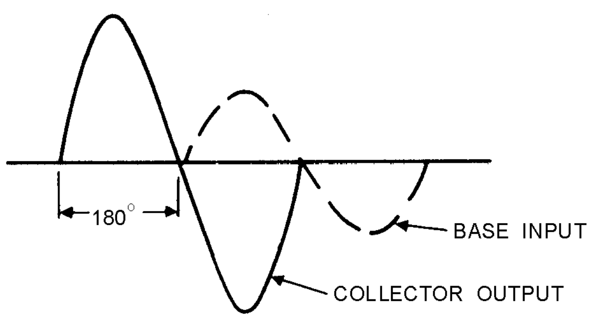
(A) PHASE-SHIFT NETWORK C1 AND R1

Figure 2-18A.—Three-section, phase-shifting RC network. PHASE-SHIFT NETWORK C1 AND R1.



(B) PHASE-SHIFT NETWORK C2 AND R2

Figure 2-18B.—Three-section, phase-shifting RC network. PHASE-SHIFT NETWORK C2 AND R2.



(C) PHASE-SHIFT NETWORK C3 AND R_B

Figure 2-18C.—Three-section, phase-shifting RC network. PHASE-SHIFT NETWORK C3 AND R_B.

A high-gain transistor must be used with the three-section RC network because the losses in the network are high. Using more than three RC sections actually reduces the overall signal loss within the network. This is because additional RC sections reduce the phase shift necessary for each section, and the loss for each section is lowered as the phase shift is reduced. In addition, an oscillator that uses four or more RC networks has more stability than one that uses three RC networks. In a four-part RC network, each part shifts the phase of the feedback signal by approximately 45 degrees to give the total required 180-degree phase shift.

Q-15. Which components provide the regenerative feedback signal in the phase-shift oscillator?

Q-16. Why is a high-gain transistor used in the phase-shift oscillator?

Q-17. Which RC network provides better frequency stability, three-section or four-section?

CRYSTAL OSCILLATORS

Crystal oscillators are those in which a specially-cut crystal controls the frequency. CRYSTAL-CONTROLLED OSCILLATORS are the standard means used for maintaining the frequency of radio transmitting stations within their assigned frequency limits. A crystal-controlled oscillator is usually used to produce an output which is highly stable and at a very precise frequency.

As stated earlier, crystals used in electrical circuits are thin sheets cut from the natural crystal and are ground to the proper thickness for the desired resonant frequency. For any given crystal cut, the thinner the crystal, the higher the resonant frequency. The "cut" (X, Y, AT, and so forth) of the crystal means the precise way in which the usable crystal is cut from the natural crystal. Some typical crystal cuts may be seen in figure 2-19.

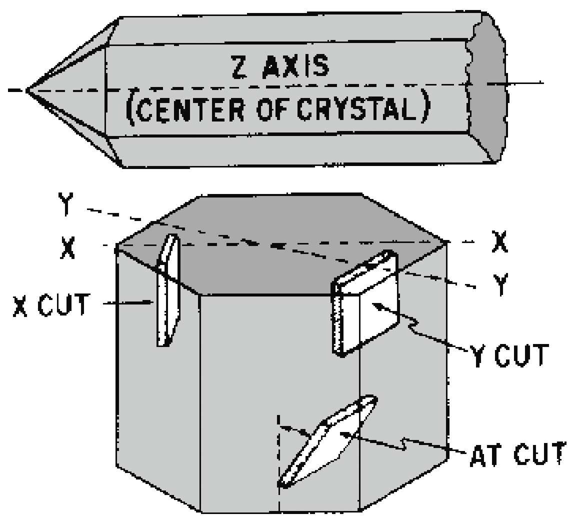


Figure 2-19.—Quartz crystal cuts.

Transmitters which require a very high degree of frequency stability, such as a broadcast transmitter, use temperature-controlled ovens to maintain a constant crystal temperature. These ovens are thermostatically controlled containers in which the crystals are placed.

The type of cut also determines the activity of the crystal. Some crystals vibrate at more than one frequency and thus will operate at harmonic frequencies. Crystals which are not of uniform thickness may have two or more resonant frequencies. Usually one resonant frequency is more pronounced than the others. The other less pronounced resonant frequencies are referred to as SPURIOUS frequencies. Sometimes such a crystal oscillates at two frequencies at the same time.

The amount of current that can safely pass through a crystal ranges from 50 to 200 milliamperes. When the rated current is exceeded, the amplitude of mechanical vibration becomes too great, and the crystal may crack. Overloading the crystal affects the frequency of vibration because the power dissipation and crystal temperature increase with the amount of load current.

Crystals as Tuned Circuits

A quartz crystal and its equivalent circuit are shown in figure 2-20, views (A) and (B). Capacitor C2, inductor L1, and resistor R1 in view (B) represent the electrical equivalent of the quartz crystal in view (A). Capacitance C1 in (view B) represents the capacitance between the crystal electrodes in view (A). Depending upon the circuit characteristics, the crystal can act as a capacitor, an inductor, a series-tuned circuit, or a parallel-tuned circuit.

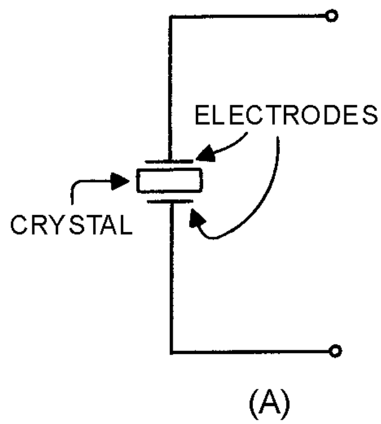


Figure 2-20A.—Quartz crystal and equivalent circuit.

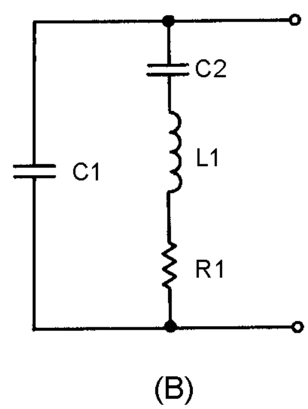


Figure 2-20B.—Quartz crystal and equivalent circuit.

At some frequency, the reactances of equivalent capacitor C1 and inductor L will be equal and the crystal will act as a series-tuned circuit. A series-tuned circuit has a minimum impedance at resonance (figure 2-21). Above resonance the series-tuned circuit acts INDUCTIVELY, and below resonance it acts CAPACITIVELY. In other words, the crystal unit has its lowest impedance at the series-resonance frequency. The impedance increases as the frequency is lowered because the unit acts as a capacitor. The impedance of the crystal unit also increases as the frequency is raised above the series-resonant point because the unit acts as an inductor. Therefore, the crystal unit reacts as a series-tuned circuit.

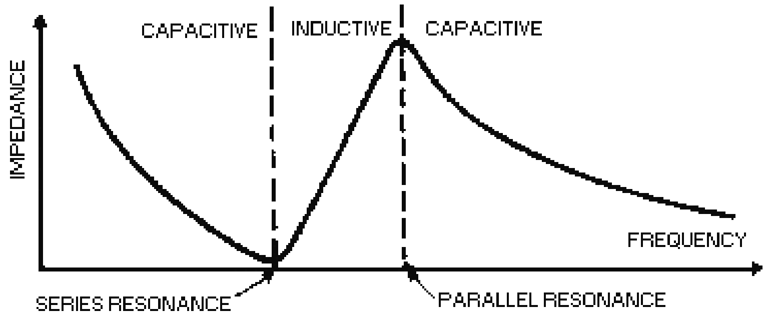


Figure 2-21.—Frequency response of a crystal.

Since the series-tuned circuit acts as an inductor above the resonant point, the crystal unit becomes equivalent to an inductor and is parallel with the equivalent capacitor C1 (view (B) of figure 2-20). At some frequency above the series-resonant point, the crystal unit will act as a parallel-tuned circuit. A parallel-tuned circuit has a MAXIMUM impedance at the parallel-resonant frequency and acts inductively below parallel resonance (figure 2-21). Therefore, at some frequency, depending upon the cut of the crystal, the crystal unit will act as a parallel-tuned circuit.

The frequency stability of crystal-controlled oscillators depends on the Q of the crystal. The Q of a crystal is very high. It may be more than 100 times greater than that obtained with an equivalent electrical circuit. The Q of the crystal is determined by the cut, the type of holder, and the accuracy of grinding. Commercially produced crystals range in Q from 5,000 to 30,000 while some laboratory experiment crystals range in Q up to 400,000.

Crystal-Controlled Armstrong Oscillator

The crystal-controlled Armstrong oscillator (figure 2-22) uses the series-tuned mode of operation. It works much the same as the Hartley oscillator except that frequency stability is improved by the crystal (in the feedback path). To operate the oscillator at different frequencies, you simply change crystals (each crystal operates at a different frequency).

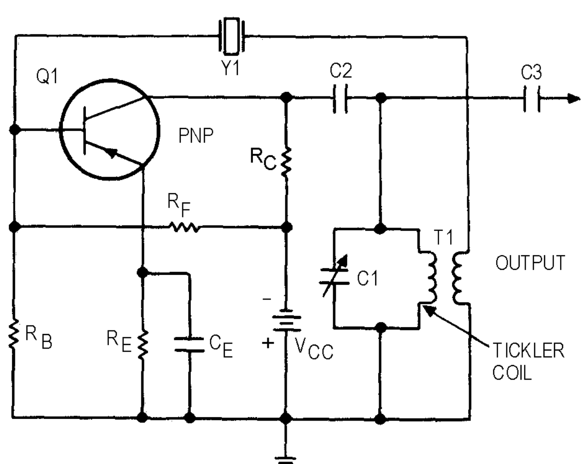


Figure 2-22.—Crystal-controlled Armstrong oscillator.

Variable capacitor C1 makes the circuit tunable to the selected crystal frequency. C1 is capable of tuning to a wide band of selected crystal frequencies. Regenerative feedback from the collector to base is through the mutual inductance between the transformer windings of T1. This provides the necessary 180-degree phase shift for the feedback signal. Resistors R_B, R_F, and R_C provide the base and collector bias voltage. Capacitor C_E bypasses ac variations around emitter resistor R_E.

At frequencies above and below the series-resonant frequency of the selected crystal, the impedance of the crystal increases and reduces the amount of feedback signal. This, in turn, prevents oscillations at frequencies other than the series-resonant frequency.

Crystal-Controlled Pierce Oscillator

The crystal-controlled PIERCE OSCILLATOR uses a crystal unit as a parallel-resonant circuit. The Pierce oscillator is a modified Colpitts oscillator. They operate in the same way except that the crystal unit replaces the parallel-resonant circuit of the Colpitts.

Figure 2-23 shows the common-base configuration of the Pierce oscillator. Feedback is supplied from the collector to the emitter through capacitor C1. Resistors R_B, R_C, and R_F provide the proper bias conditions for the circuit and resistor R_E is the emitter resistor. Capacitors C1 and C_E form a voltage divider connected across the output. Since no phase shift occurs in the common-base circuit, capacitor C1 feeds back a portion of the output signal to the emitter without a phase shift. The oscillating frequency is determined not only by the crystal but also by the parallel capacitance caused by capacitors C1 and C_E. This parallel capacitance affects the oscillator frequency by lowering it. Any change in capacitance of either C1 or C_E changes the frequency of the oscillator.

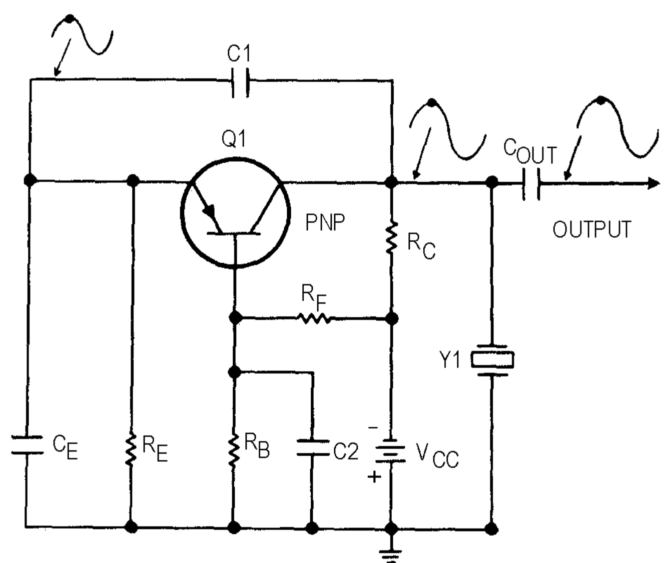


Figure 2-23.—Pierce oscillator, common-base configuration.

Figure 2-24 shows the common-emitter configuration of the Pierce oscillator. The resistors in the circuit provide the proper bias and stabilization conditions. The crystal unit and capacitors C1 and C2 determine the output frequency of the oscillator. The signal developed at the junction between Y1 and C1 is 180 degrees out of phase with the signal at the junction between Y1 and C2. Therefore, the signal at the Y1-C1 junction can be coupled back to the base of Q1 as a regenerative feedback signal to sustain oscillations.

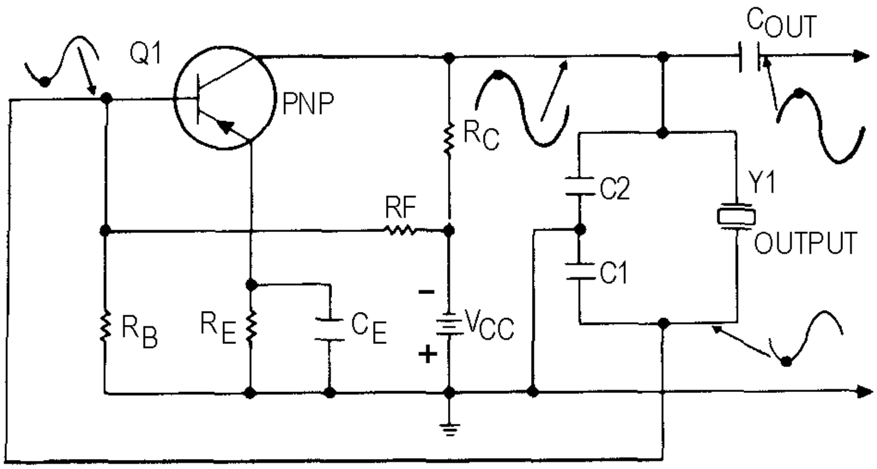


Figure 2-24.—Pierce oscillator, common-emitter configuration.

Q-18. What is the impedance of a crystal at its resonant frequency when it is used in the parallel mode?

Q-19. What is the impedance of a crystal at its resonant frequency when it is used in the series mode?

PULSED OSCILLATORS

A sinusoidal (sine-wave) oscillator is one that will produce output pulses at a predetermined frequency for an indefinite period of time; that is, it operates continuously. Many electronic circuits in equipment such as radar require that an oscillator be turned on for a specific period of time and that it remain in an off condition until required at a later time. These circuits are referred to as PULSED OSCILLATORS or RINGING OSCILLATORS. They are nothing more than sine-wave oscillators that are turned on and off at specific times.

Figure 2-25, view (A), shows a pulsed oscillator with the resonant tank in the emitter circuit. A positive input makes Q1 conduct heavily and current flow through L1; therefore no oscillations can take place. A negative-going input pulse (referred to as a gate) cuts off Q1, and the tank oscillates until the gate ends or until the ringing stops, whichever comes first.

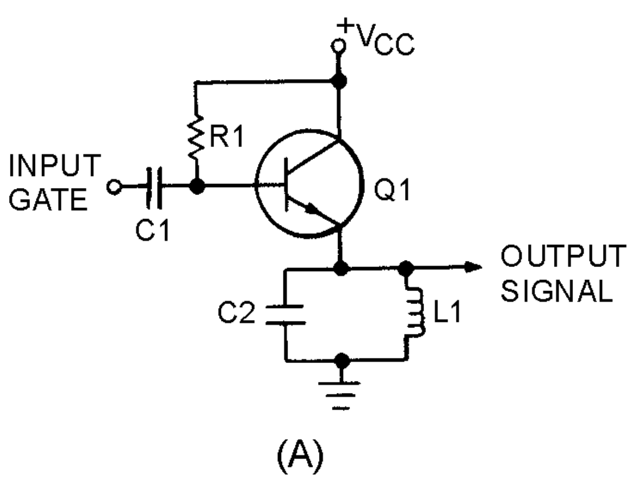
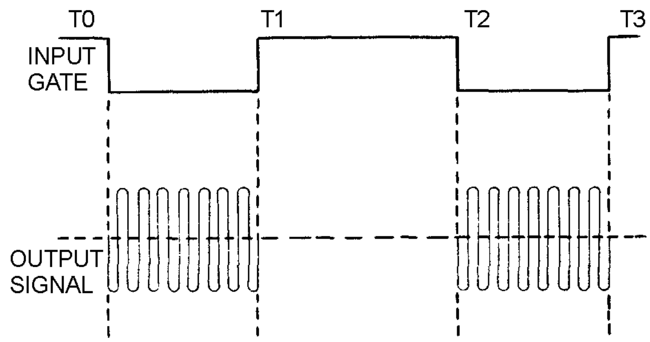


Figure 2-25A.—Pulsed oscillator.



(B)

Figure 2-25B.—Pulsed oscillator.

The waveforms in view (B) show the relationship of the input gate and the output signal from the pulsed oscillator. To see how this circuit operates, assume that the Q of the LC tank circuit is high enough to prevent damping. An output from the circuit is obtained when the input gate goes negative (T0 to T1 and T2 to T3). The remainder of the time (T1 to T2) the transistor conducts heavily and there is no output from the circuit. The width of the input gate controls the time for the output signal. Making the gate wider causes the output to be present (or ring) for a longer time.

Frequency of a Pulsed Oscillator

The frequency of a pulsed oscillator is determined by both the input gating signal and the resonant frequency of the tank circuit. When a sinusoidal oscillator is resonant at 1 megahertz, the output is 1 million cycles per second. In the case of a pulsed oscillator, the number of cycles present in the output is determined by the gating pulse width.

If a 1-megahertz oscillator is cut off for 1/2 second, or 50 percent of the time, then the output is 500,000 cycles at the 1 -megahertz rate. In other words, the frequency of the tank circuit is still 1 megahertz, but the oscillator is only allowed to produce 500,000 cycles each second.

The output frequency can be determined by controlling how long the tank circuit will oscillate. For example, suppose a negative input gate of 500 microseconds and a positive input gate of 999,500 microseconds (total of 1 second) are applied. The transistor will be cut off for 500 microseconds and the tank circuit will oscillate for that 500 microseconds, producing an output signal. The transistor will then conduct for 999,500 microseconds and the tank circuit will not oscillate during that time period. The 500 microseconds that the tank circuit is allowed to oscillate will allow only 500 cycles of the 1-megahertz tank frequency.

You can easily check this frequency by using the following formula:

$$t = \frac{1}{f}(\text{one cycle of resonant frequency})$$

$$t = \text{time}$$

$$f = \text{resonant frequency of tank circuit}$$

One cycle of the 1-megahertz resonant frequency is equal to 1 microsecond.

$$\frac{1}{1{,}000{,}000} = .000001 \text{ or } 1 \times 10^{-6} \text{ seconds}$$

Then, by dividing the time for 1 cycle (1 microsecond) into gate length (500 microseconds), you will get the number of cycles (500).

There are several different varieties of pulsed oscillators for different applications. The schematic diagram shown in figure 2-25, view (A), is an emitter-loaded pulsed oscillator. The tank circuit can be placed in the collector circuit, in which case it is referred to as a collector-loaded pulsed oscillator. The difference between the emitter-loaded and the collector-loaded oscillator is in the output signal. The first alternation of an emitter-loaded npn pulsed oscillator is negative. The first alternation of the collector-loaded pulsed oscillator is positive. If a pnp is used, the oscillator will reverse the first alternation of both the emitter-loaded and the collector-loaded oscillator.

You probably have noticed by now that feedback has not been mentioned in this discussion. Remember that regenerative feedback was a requirement for sustained oscillations. In the case of the pulsed oscillator, oscillations are only required for a very short period of time. You should understand, however, that as the width of the input gate (which cuts off the transistor) is increased, the amplitude of the sine wave begins to decrease (dampen) near the end of the gate period because of the lack of feedback. If a long period of oscillation is required for a particular application, a pulsed oscillator with regenerative feedback is used. The principle of operation remains the same except that the feedback network sustains the oscillation period for the desired amount of time.

Q-20. Oscillators that are turned on and off at a specific time are known as what type of oscillators?

Q-21. What is the polarity of the first alternation of the tank circuit in an emitter-loaded npn pulsed oscillator?

HARMONICS

From your study of oscillators, you should know that the oscillator will oscillate at the resonant frequency of the tank circuit. Although the tank circuit is resonant at a particular frequency, many other frequencies other than the resonant frequency are present in the oscillator. These other frequencies are referred to as HARMONICS. A harmonic is defined as a sinusoidal wave having a frequency that is a multiple of the fundamental frequency. In other words, a sine wave that is twice that fundamental frequency is referred to as the SECOND HARMONIC.

What you must remember is that the current in circuits operating at the resonant frequency is relatively large in amplitude. The harmonic frequency amplitudes are relatively small. For example, the second harmonic of a fundamental frequency has only 20 percent of the amplitude of the resonant frequency. A third harmonic has perhaps 10 percent of the amplitude of the fundamental frequency.

One useful purpose of harmonics is that of frequency multiplication. It can be used in circuits to multiply the fundamental frequency to a higher frequency. The need for frequency-multiplier circuits results from the fact that the frequency stability of most oscillators decreases as frequency increases. Relatively good stability can be achieved at the lower frequencies. Thus, to achieve optimum stability, an oscillator is operated at a low frequency, and one or more stages of multiplication are used to raise the signal to the desired operating frequency.

FREQUENCY MULTIPLICATION

FREQUENCY MULTIPLIERS are special class C amplifiers that are biased at 3 to 10 times the normal cutoff bias. They are used to generate a frequency that is a multiple (harmonic) of a lower frequency. Such circuits are called frequency multipliers or harmonic generators.

Figure 2-26 illustrates a frequency multiplier known as a FREQUENCY DOUBLER or SECOND HARMONIC GENERATOR. As illustrated, the input is 1 megahertz and the output is 2 megahertz, or twice the input frequency. In other words, the second harmonic of 1 megahertz is 2 megahertz. The third harmonic (frequency tripler) would be 3 megahertz, or 3 times the input signal. The fourth harmonic (quadruplet) would be 4 megahertz, or 4 times the 1-megahertz input signal. The fourth harmonic generator (frequency quadruplet) is normally as high in multiplication as is practical, because at harmonics higher than the fourth, the output diminishes to a very weak output signal.

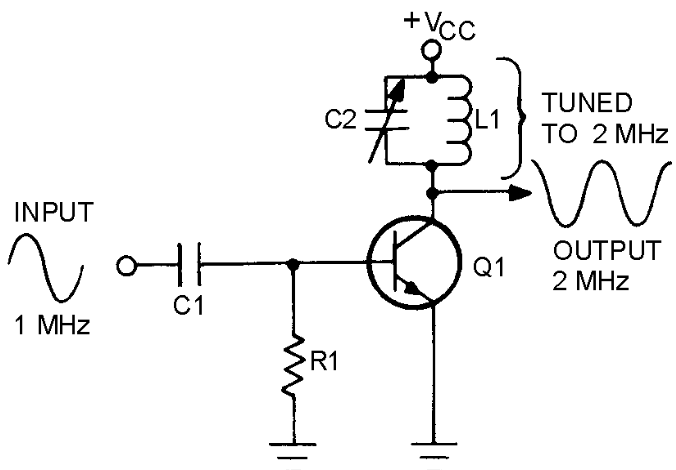


Figure 2-26.—Frequency doubler.

Frequency multipliers are operated by the pulses of collector current produced by a class C amplifier. Although the collector current flows in pulses, the alternating collector voltage is sinusoidal because of the action of the tank circuit. When the output tank circuit is tuned to the required harmonic, the tank circuit acts as a filter, accepting the desired frequency and rejecting all others.

Figure 2-27 illustrates the waveforms in a typical doubler circuit. You can see that the pulses of collector current are the same frequency as the input signal. These pulses of collector current energize the tank circuit and cause it to oscillate at twice the base signal frequency. Between the pulses of collector current, the tank circuit continues to oscillate. Therefore, the tank circuit receives a current pulse for every other cycle of its output.

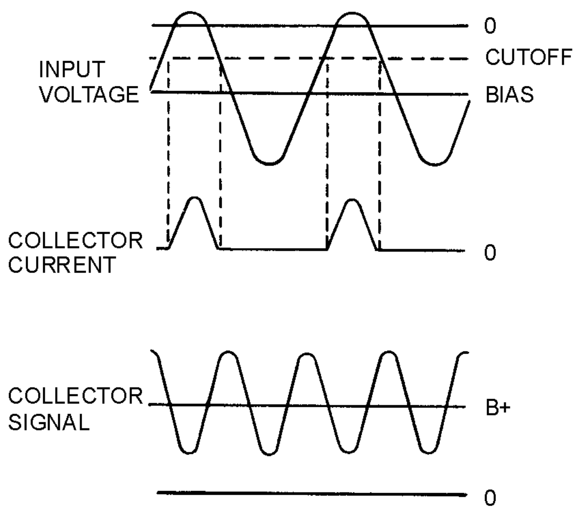


Figure 2-27.—Frequency doubler waveforms.

Buffer Amplifier

Coupling the resonant frequency from the oscillator by different coupling methods also affects the oscillator frequency and amplitude. A BUFFER AMPLIFIER decreases the loading effect on the oscillator by reducing the interaction (matching impedance) between the load and the oscillator.

Figure 2-28 is the schematic diagram of a buffer amplifier. This circuit is a common-collector amplifier. A common-collector amplifier has a high input impedance and a low output impedance. Since the output of an oscillator is connected to the high impedance of the common-collector amplifier, the buffer has little effect on the operation of the oscillator. The output of the common-collector buffer is then connected to an external load; therefore, the changes in the output load cannot reflect back to the oscillator circuit. Thus, the buffer amplifier reduces interaction between the load and the oscillator. Figure 2-29 illustrates a shunt-fed Hartley oscillator with a buffer amplifier. This is "one-way" coupling since the oscillator signal is coupled forward, but load changes are not coupled back to the oscillator.

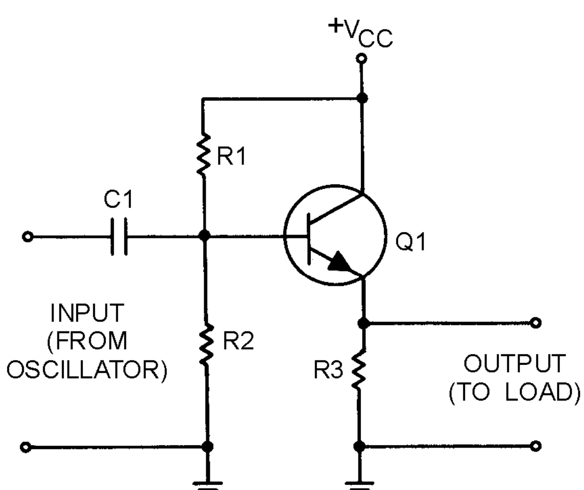


Figure 2-28.—Buffer amplifier.

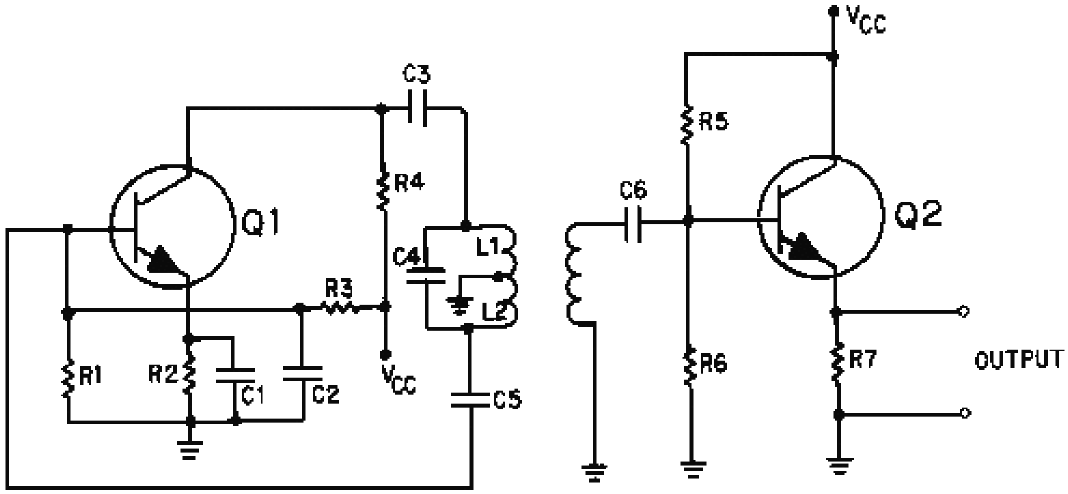


Figure 2-29.—Shunt-fed Hartley oscillator with buffer amplifier.

Q-22. What is the frequency that is twice the fundamental frequency?

Q-23. What is the purpose of the buffer amplifier?

SUMMARY

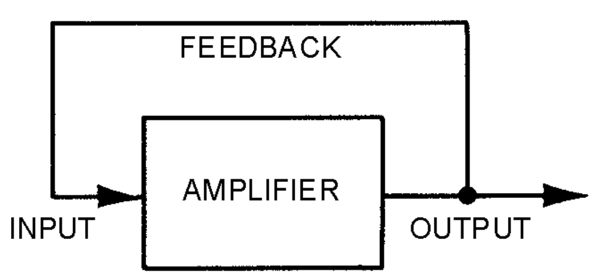
This chapter has presented information on oscillators. The information that follows summarizes the important points of this chapter.

WAVE GENERATORS can be classified according to the SINUSOIDAL or NONSINUSOIDAL waveforms produced.

SINUSOIDAL WAVE GENERATORS (oscillators) produce a sine wave of constant amplitude and frequency. There are three ways to control the frequency of sine-wave generators: (1) RC NETWORKS, (2) LC NETWORKS, and (3) CRYSTALS.

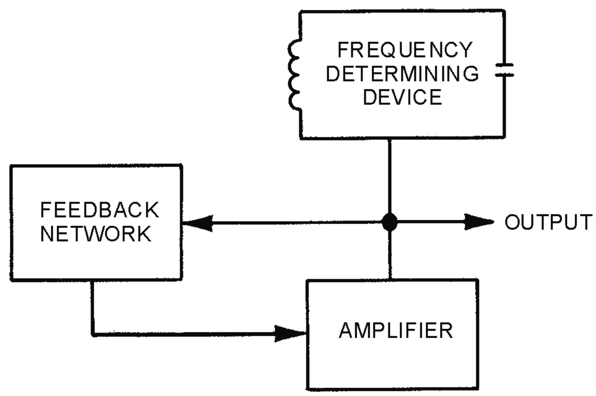
NONSINUSOIDAL WAVE GENERATORS (oscillators) generate complex waveforms such as SQUARE WAVES, RECTANGULAR WAVES, SAWTOOTH WAVES, TRAPEZOIDAL WAVES, and TRIGGERS. Nonsinusoidal wave generators are often called RELAXATION OSCILLATORS.

A **BASIC OSCILLATOR** can be thought of as an amplifier that provides itself with a signal input.



An **OSCILLATOR** is a device that converts dc power to ac power at a predetermined frequency.

The requirements for an oscillator are AMPLIFICATION, REGENERATIVE FEEDBACK, and a FREQUENCY-DETERMINING NETWORK.



An oscillator has two stability requirements, AMPLITUDE STABILITY and FREQUENCY STABILITY.

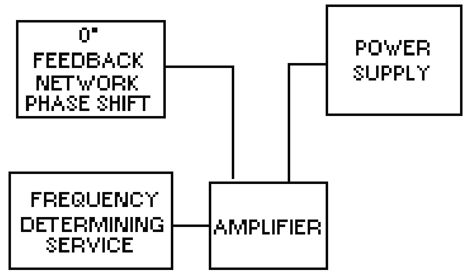
FEEDBACK is the process of transferring energy from a high-level point in a system to a low-level point. Feedback that aids the input signal is REGENERATIVE or POSITIVE. Feedback that opposes the input signal is DEGENERATIVE or NEGATIVE.

The three basic circuit configurations used for oscillators are **COMMON COLLECTOR**, **COMMON BASE**, and **COMMON EMITTER**.

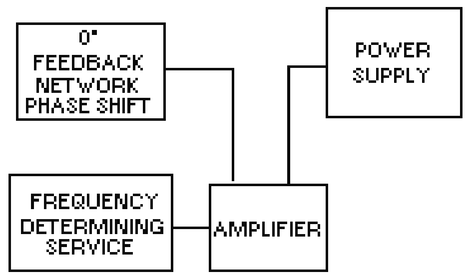
In the **COMMON-COLLECTOR** configuration there is no PHASE SHIFT between input and output. It is not necessary for the feedback network to provide a phase shift. Voltage gain is less than unity (one) and power gain is low so it is very seldom used as an oscillator.

In the **COMMON-BASE** configuration, there is no PHASE SHIFT between input and output. It is not necessary for the feedback network to provide a phase shift. Voltage and power gain are high enough to give satisfactory operation in an oscillator circuit.

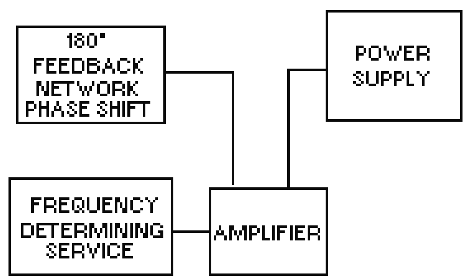
In the **COMMON-EMITTER** configuration, there is a 180-degree PHASE SHIFT between input and output. The feedback network must provide another phase shift of 180 degrees. It has a high power gain.



(A) COMMON- COLECTOR CONFIGURATION



(B) COMMON-BASE CONFIGURATION



(C) COMMON-EMITTER CONFIGURATION

The **ARMSTRONG OSCILLATOR** is used to produce a sine-wave output of constant amplitude and fairly constant frequency.

An oscillator in which dc power is supplied to the transistor through the tank circuit, or a portion of the tank circuit, is SERIES FED.

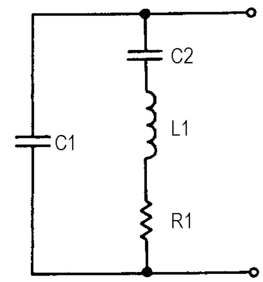
An oscillator in which dc power is supplied to the transistor through a path separate and parallel to the tank circuit is PARALLEL or SHUNT FED.

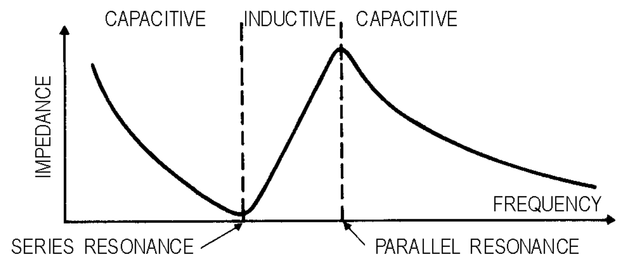
The **HARTLEY OSCILLATOR** is used to produce a sine-wave output of constant amplitude and fairly constant frequency.

The **COLPITTS OSCILLATOR** is used to produce a sine-wave output of constant amplitude and fairly constant frequency within the rf range. The identifying features of the Colpitts oscillator are the split capacitors.

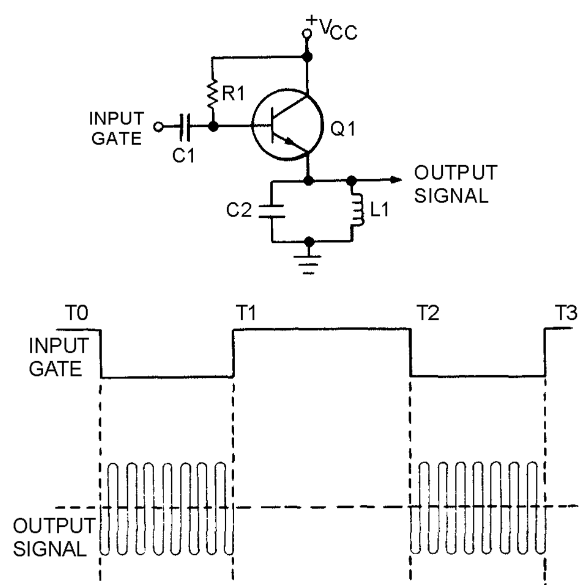
The **RESISTIVE-CAPACITIVE (RC) FEEDBACK OSCILLATOR** is used to produce a sine-wave output of relatively constant amplitude and frequency. It uses RC networks to produce feedback and eliminate the need for inductors in the resonant circuit.

CRYSTAL OSCILLATORS are those oscillators that use a specially cut crystal to control the frequency. The crystal can act as either a capacitor or inductor, a series-tuned circuit, or a parallel-tuned circuit.



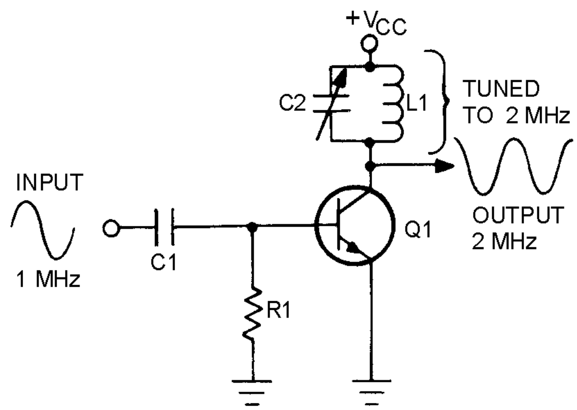


PULSED OSCILLATORS are sinusoidal oscillators that are turned on and off for a specific time duration. The frequency of a pulsed oscillator is determined by both the input gating pulse and the resonant frequency of the tank circuit.

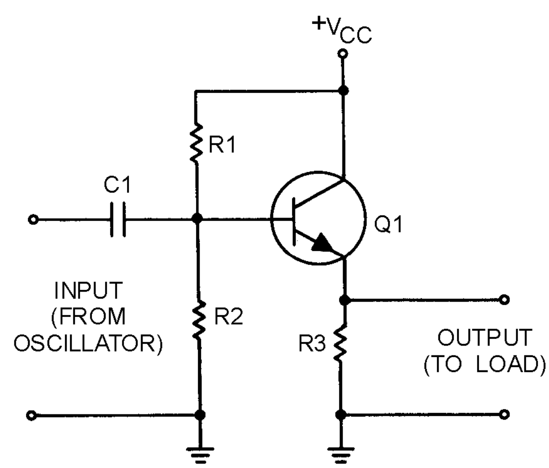


A **HARMONIC** is a sinusoidal wave having a frequency that is a multiple of the fundamental frequency.

FREQUENCY MULTIPLIERS (HARMONIC GENERATORS) are special class C amplifiers that are biased at 3 to 10 times the normal cutoff. They are used to generate a frequency that is a multiple or harmonic of a lower frequency.



A **BUFFER AMPLIFIER** decreases the loading effect on the oscillator by reducing the interaction between the load and the oscillator.



ANSWERS TO QUESTIONS Q1. THROUGH Q23.

A-1. Sinusoidal and nonsinusoidal.

A-2. RC, LC, and crystal.

A-3. Relaxation oscillators.

A-4. Oscillator.

A-5. Amplification, regenerative feedback, and frequency-determining device.

A-6. Regenerative or positive.

A-7. Inductive and capacitive.

A-8. Armstrong.

A-9. Hartley.

A-10. Colpitts.

A-11. Common collector (CC), common emitter (CE), and common base (CB).

A-12. Feedback coil. Armstrong uses a separate coil. Hartley uses a tapped coil.

A-13. In the series-fed Hartley oscillator, dc flows through the tank circuit.

A-14. Split capacitors.

A-15. Resistor-capacitor networks.

A-16. Because of the losses encountered in the RC networks.

A-17. Four-section.

A-18. Maximum.

A-19. Minimum.

A-20. Pulsed oscillators.

A-21. Negative.

A-22. Second harmonic.

A-23. Reduce interaction between oscillator and load.

CHAPTER 3

WAVEFORMS AND WAVE GENERATORS

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Explain the operation of a stable, monostable, and bistable multivibrators.
2. Explain the operation of a blocking oscillator.
3. Explain the operation of a sawtooth generator.
4. Explain the operation of a trapezoidal wave generator.
5. Explain how the jump voltage is produced in a trapezoidal wave generator.

WAVEFORMS

This chapter will present methods of generating waveforms. Before you begin to study how waveforms are generated, you need to know the basic characteristics of waveforms. This section will discuss basic periodic waveforms.

PERIODIC WAVEFORMS

A waveform which undergoes a pattern of changes, returns to its original pattern, and repeats the same pattern of changes is called a PERIODIC waveform. Periodic waveforms are nonsinusoidal except for the sine wave. Periodic waveforms which will be discussed are the sine wave, square wave, rectangular wave, sawtooth wave, trapezoidal wave, and trigger pulses.

Sine Wave

Each completed pattern of a periodic waveform is called a CYCLE, as shown by the SINE WAVE in figure 3-1, view (A). Sine waves were presented in NEETS, Module 2, *Alternating Current and Transformers*, Chapter 1.

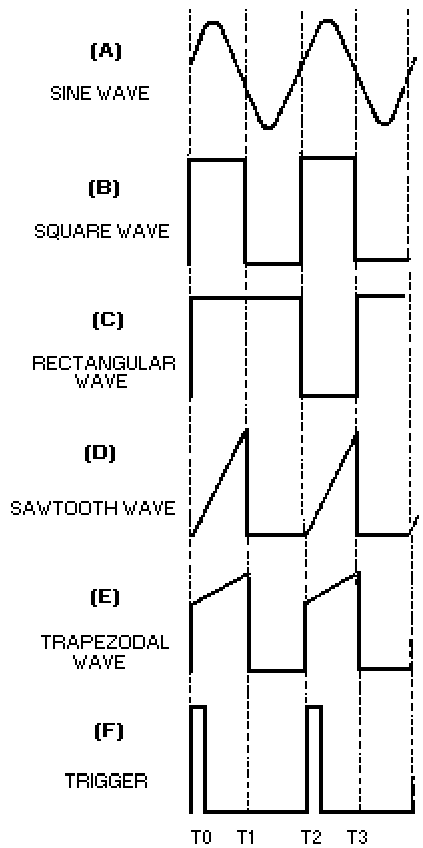


Figure 3-1.—Periodic waveforms.

Square Wave

A SQUARE WAVE is shown in figure 3-1, view (B). As shown, it has two alternations of equal duration and a square presentation for each complete cycle. Figure 3-2 shows a breakdown of the square wave and is the figure you should view throughout the square wave discussion. The amplitude is measured vertically. The time for a complete cycle is measured between corresponding points on the wave (T0 to T2, or T1 to T3).

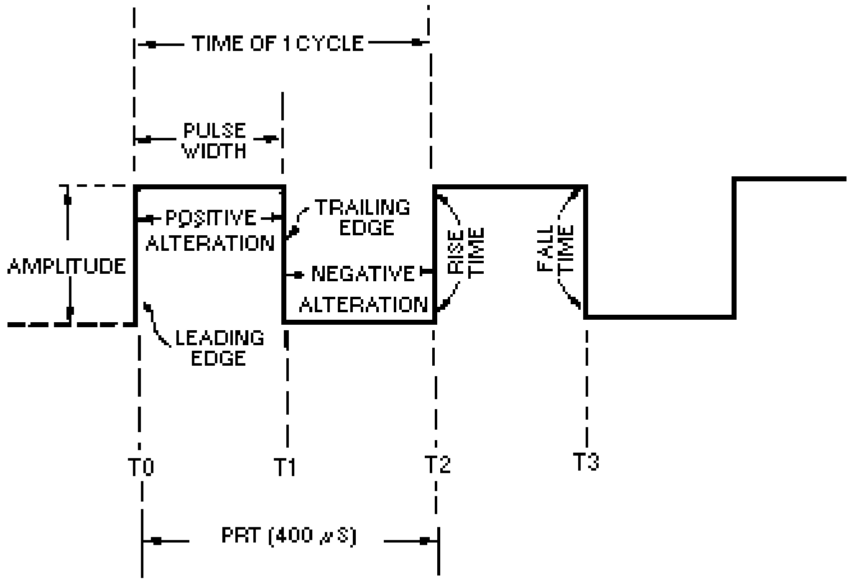


Figure 3-2.—Square wave.

One alternation is called a PULSE. The time for one complete cycle is called the PULSE-REPETITION TIME (prt). The number of times in 1 second that the cycle repeats itself is called the PULSE-REPETITION FREQUENCY (prf) or PULSE-REPETITION RATE (prr). If each alternation in figure 3-2 is 200 microseconds (µs), the prt will be 400 microseconds, and the prf will be 2,500 hertz. The following examples are provided to illustrate the mathematical relationship between prf and prt:

Given:

$$prf = \frac{1}{prt}$$

Where:

$$prf = 400\mu s$$

Solution:

$$prf = \frac{1}{prt}$$

$$prf = \frac{1}{400\mu s}$$

$$prf = 2{,}500 \text{ Hz}$$

You should readily see that prt is just the inverse of prf. Therefore: Given:

$$prt = \frac{1}{prf}$$

Where:

$$prf = 2{,}500\ Hz$$

Solution:

$$prt = \frac{1}{prf}$$

$$prt = \frac{1}{2{,}500\ Hz}$$

$$prt = 400\ \text{microseconds}$$

The length of the pulse measured in time (T0 to T1) is referred to as the PULSE WIDTH (pw). The left side of the pulse is called the LEADING EDGE and the right side is called the TRAILING EDGE.

Time is required for a voltage or current to change in amplitude. The interval of time needed for the voltage to go from 0 to 100 percent (or from 100 to 0 percent) of its maximum value is called the TRANSIENT INTERVAL. The two types of transient intervals are RISE TIME and FALL TIME. Rise time is more accurately defined as the time required for the voltage to build up from 10 percent to 90 percent of the maximum amplitude point. Fall time is the time required for the voltage to drop from 90 percent to 10 percent of the maximum amplitude point.

In this text you will be presented with information in which waveforms appear to have instantaneous rise and fall times. This is done to simplify the presentation of the material. In reality these waveforms do have rise and fall times (transient intervals).

Rectangular Wave

A rectangular wave is similar to the square wave. The difference is that in the rectangular waveform, the two alternations of the waveform are of unequal time duration. Figure 3-1, view (C), shows that the negative alternation (pulse) is shorter (in time) than the positive alternation. The negative alternation could be represented as the longer of the two alternations. Either way, the appearance is that of a rectangle.

Sawtooth Wave

The SAWTOOTH waveform is shown in figure 3-1, view (D). A sawtooth wave resembles the teeth of a saw blade. There is a rapid vertical rise of voltage from T0 to T1, which is linear (straight). At T1 this voltage abruptly falls (essentially no time used) to its previous static value. The voltage remains at this value until T2 when it again has a linear rise. You can see this action in an oscilloscope where there are two voltage input locations, vertical and horizontal. If you apply a linear voltage to the vertical input, the electron beam will be forced to move in a vertical direction on the crt. A linear voltage applied to the horizontal input will cause the electron beam to move horizontally across the crt. The application of two linear voltages, one to the vertical input and one to the horizontal input at the same time, will cause the beam to move in both a vertical and horizontal (diagonal) direction at the same time. This then is how a sawtooth wave is made to appear on an oscilloscope. You should refer to NEETS, Module 6, *Electronic Emission, Tubes, and Power Supplies*, Chapter 2, for a review of oscilloscopes.

Trapezoidal Wave

A TRAPEZOIDAL wave looks like a sawtooth wave on top of a square or rectangular wave, as shown in figure 3-1, view (E). The leading edge of a trapezoidal wave is called the JUMP voltage. The next portion of the wave is the linear rise or SLOPE. The trailing edge is called the FALL or DECAY. A trapezoidal wave is used to furnish deflection current in the electromagnetic cathode ray tube and is found in television and radar display systems. Electromagnetic cathode ray tubes use coils for the deflection system, and a linear rise in current is required for an accurate horizontal display. The square or rectangular wave portion provides the jump voltage for a linear rise in current through the resistance of the coil. This will be explained further in a discussion of the trapezoidal sweep generator.

Triggers

A trigger is a very narrow pulse, as shown in figure 3-1, view (F). Trigger pulses are normally used to turn other circuits on or off.

WAVEFORM GENERATOR

Nonsinusoidal oscillators generate complex waveforms such as those just discussed. Because the outputs of these oscillators are generally characterized by a sudden change, or relaxation, these oscillators are often called RELAXATION OSCILLATORS. The pulse repetition rate of these oscillators is usually governed by the charge and discharge timing of a capacitor in series with a resistor. However, some oscillators contain inductors that, along with circuit resistance, affect the output frequency. These RC and LC networks within oscillator circuits are used for frequency determination. Within this category of relaxation oscillators are MULTIVIBRATORS, BLOCKING OSCILLATORS, and SAWTOOTH- and TRAPEZOIDAL-WAVE GENERATORS.

Many electronic circuits are not in an "on" condition all of the time. In computers, for example, waveforms must be turned on and off for specific lengths of time. The time intervals vary from tenths of microseconds to several thousand microseconds. Square and rectangular waveforms are normally used to turn such circuits on and off because the sharp leading and trailing edges make them ideal for timing purposes.

MULTIVIBRATORS

The type of circuit most often used to generate square or rectangular waves is the multivibrator. A multivibrator, as shown in figure 3-3, is basically two amplifier circuits arranged with regenerative feedback. One of the amplifiers is conducting while the other is cut off.

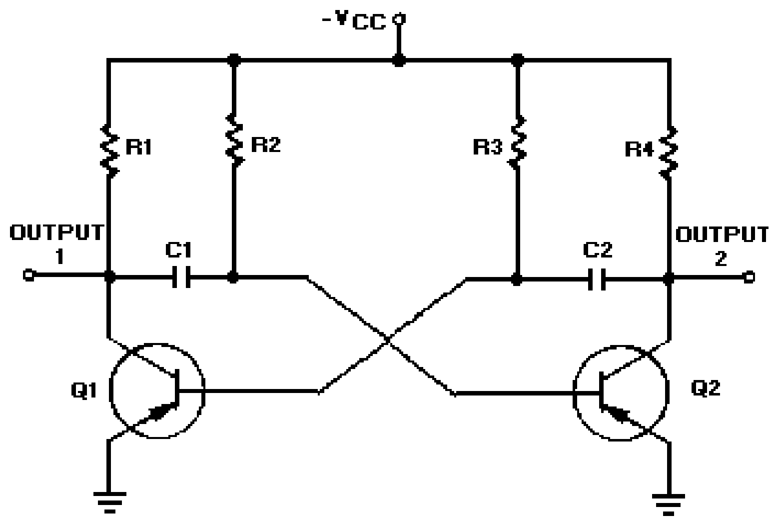


Figure 3-3.—Astable Multivibrator.

When an input signal to one amplifier is large enough, the transistor can be driven into cutoff, and its collector voltage will be almost V_{CC}. However, when the transistor is driven into saturation, its collector voltage will be about 0 volts. A circuit that is designed to go quickly from cutoff to saturation will produce a square or rectangular wave at its output. This principle is used in multivibrators.

Multivibrators are classified according to the number of steady (stable) states of the circuit. A steady state exists when circuit operation is essentially constant; that is, one transistor remains in conduction and the other remains cut off until an external signal is applied. The three types of multivibrators are the ASTABLE, MONOSTABLE, and BISTABLE.

The astable circuit has no stable state. With no external signal applied, the transistors alternately switch from cutoff to saturation at a frequency determined by the RC time constants of the coupling circuits.

The monostable circuit has one stable state; one transistor conducts while the other is cut off. A signal must be applied to change this condition. After a period of time, determined by the internal RC components, the circuit will return to its original condition where it remains until the next signal arrives.

The bistable multivibrator has two stable states. It remains in one of the stable states until a trigger is applied. It then FLIPS to the other stable condition and remains there until another trigger is applied. The multivibrator then changes back (FLOPS) to its first stable state.

Q1. What type circuit is used to produce square or rectangular waves?

Q2. What type of multivibrator does not have a stable state?

Q3. What type of multvibrator has one stable state?

Q4. What type of multivibrator has two stable states?

Astable Multivibrator

An astable multivibrator is also known as a FREE-RUNNING MULTIVIBRATOR. It is called free-running because it alternates between two different output voltage levels during the time it is on. The output remains at each voltage level for a definite period of time. If you looked at this output on an oscilloscope, you would see continuous square or rectangular waveforms. The astable multivibrator has two outputs, but NO inputs.

Let's look at the multivibrator in figure 3-3 again. This is an astable multivibrator. The astable multivibrator is said to oscillate. To understand why the astable multivibrator oscillates, assume that transistor Q1 saturates and transistor Q2 cuts off when the circuit is energized. This situation is shown in figure 3-4. We assume Q1 saturates and Q2 is in cutoff because the circuit is symmetrical; that is, R1 = R4, R2 = R3, C1 = C2, and Q1 = Q2. It is impossible to tell which transistor will actually conduct when the circuit is energized. For this reason, either of the transistors may be assumed to conduct for circuit analysis purposes.

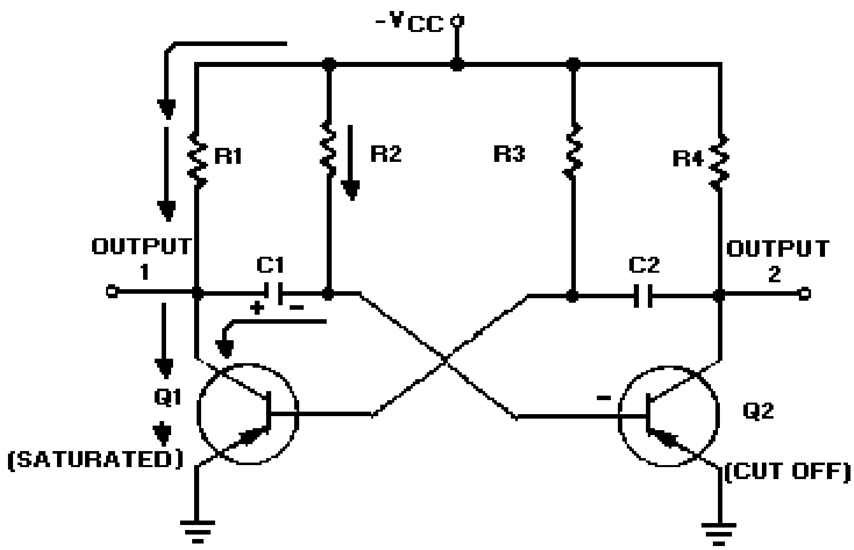


Figure 3-4.—Astable multivibrator (Q1 saturated).

Essentially, all the current in the circuit flows through Q1; Q1 offers almost no resistance to current flow. Notice that capacitor C1 is charging. Since Q1 offers almost no resistance in its saturated state, the rate of charge of C1 depends only on the time constant of R2 and C1 (recall that TC = RC). Notice that the right-hand side of capacitor C1 is connected to the base of transistor Q2, which is now at cutoff.

Let's analyze what is happening. The right-hand side of capacitor C1 is becoming increasingly negative. If the base of Q2 becomes sufficiently negative, Q2 will conduct. After a certain period of time, the base of Q2 will become sufficiently negative to cause Q2 to change states from cutoff to conduction. The time necessary for Q2 to become saturated is determined by the time constant R2C1.

The next state is shown in figure 3-5. The negative voltage accumulated on the right side on capacitor C1 has caused Q2 to conduct. Now the following sequence of events takes place almost instantaneously. Q2 starts conducting and quickly saturates, and the voltage at output 2 changes from approximately $-V_{CC}$ to approximately 0 volts. This change in voltage is coupled through C2 to the base of Q1, forcing Q1 to cutoff. Now Q1 is in cutoff and Q2 is in saturation. This is the circuit situation shown in figure 3-6.

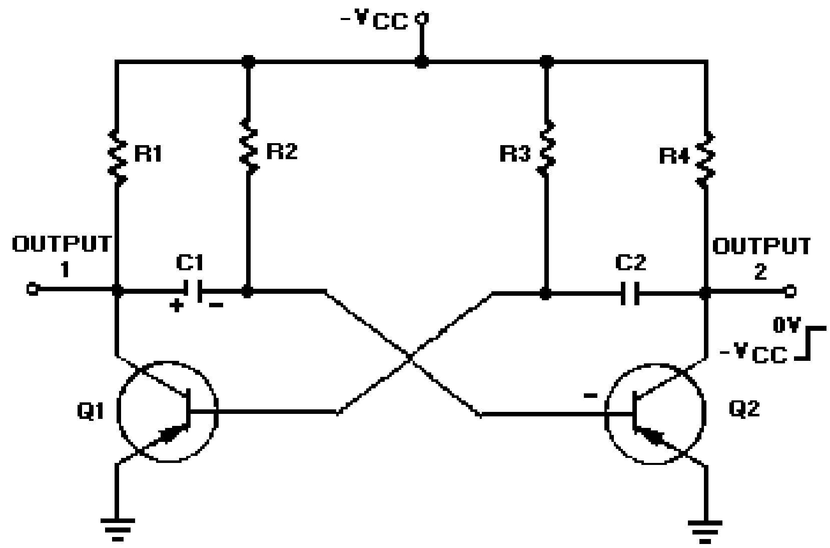


Figure 3-5.—Astable multivibrator.

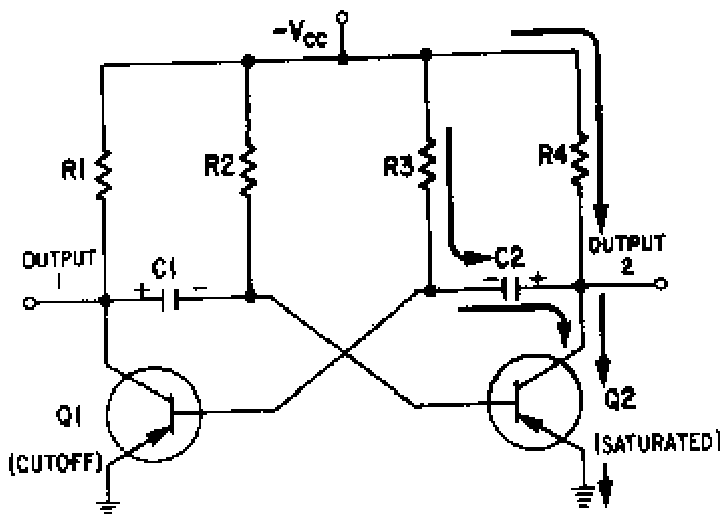


Figure 3-6.—Astable multivibrator. (Q2 saturated).

Notice that figure 3-6 is the mirror image of figure 3-4. In figure 3-6 the left side of capacitor C2 becomes more negative at a rate determined by the time constant R3C2. As the left side of C2 becomes more negative, the base of Q1 also becomes more negative. When the base of Q1 becomes negative enough to allow Q1 to conduct, Q1 will again go into saturation. The resulting change in voltage at output 1 will cause Q2 to return to the cutoff state.

Look at the output waveform from transistor Q2, as shown in figure 3-7. The output voltage (from either output of the multivibrator) alternates from approximately 0 volts to approximately $-V_{CC}$, remaining in each state for a definite period of time. The time may range from a microsecond to as much as a second or two. In some applications, the time period of higher voltage ($-V_{CC}$) and the time period of lower voltage (0 volts) will be equal. Other applications require differing higher- and lower-voltage times. For example, timing and gating circuits often have different pulse widths as shown in figure 3-8.

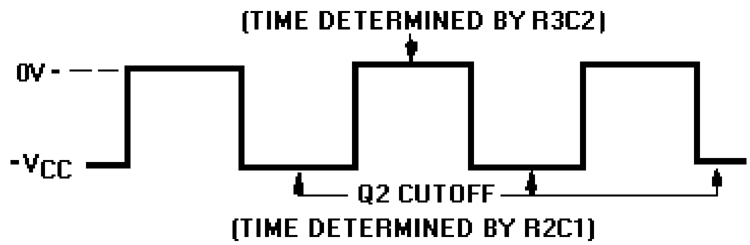


Figure 3-7.—Square wave output from Q2.

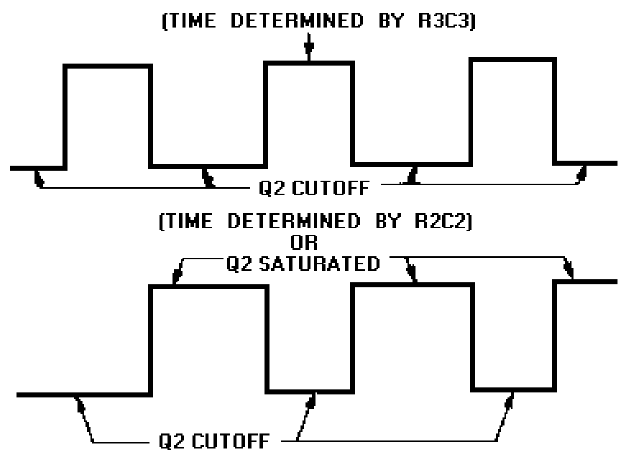


Figure 3-8.—Rectangular waves.

FREQUENCY STABILITY.—Some astable multivibrators must have a high degree of frequency stability. One way to obtain a high degree of frequency stability is to apply triggers. Figure 3-9, view (A), shows the diagram of a triggered, astable multivibrator. At time T0, a negative input trigger to the base of Q1 (through C1) causes Q1 to go into saturation, which drives Q2 to cutoff. The circuit will remain in this condition as long as the base voltage of Q2 is positive. The length of time the base of Q2 will remain positive is determined by C3, R3, and R6. Observe the parallel paths for C3 to discharge.

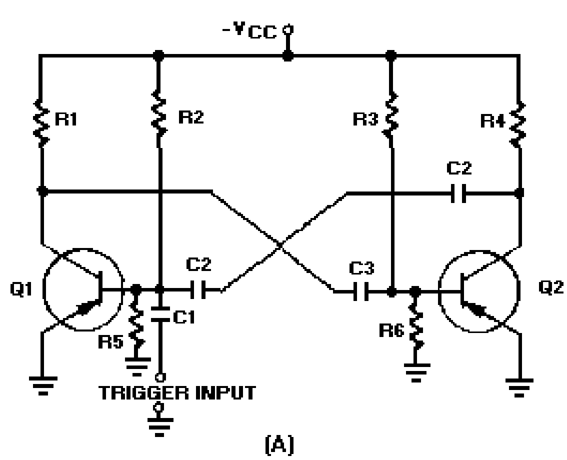


Figure 3-9A.—Triggered astable multivibrator and output.

View (B) of figure 3-9 shows the waveforms associated with the circuit. At time T1, Q2 comes out of cutoff and goes into saturation. Also, Q1 is caused to come out of saturation and is cut off. The base voltage waveform of Q1 shows a positive potential that is holding Q1 at cutoff. This voltage would normally hold Q1 at cutoff until a point between T2 and T3. However, at time T2 another trigger is applied to the base of Q1, causing it to begin conducting. Q1 goes into saturation and Q2 is caused to cut off. This action repeats each time a trigger (T2, T4, T6) is applied.

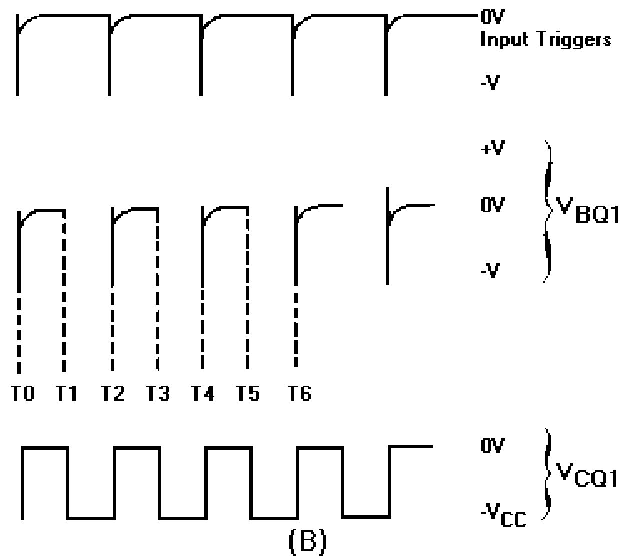


Figure 3-9B.—Triggered astable multivibrator and output.

The prt of the input triggers must be shorter than the natural free-running prt of the astable multivibrator, or the trigger prf must be slightly higher than the free-running prf of the circuit. This is to make certain the triggers control the prt of the output.

Monostable Multivibrator

The monostable multivibrator (sometimes called a ONE-SHOT MULTIVIBRATOR) is a square- or rectangular-wave generator with just one stable condition. With no input signal (quiescent condition) one amplifier conducts and the other is in cutoff. The monostable multivibrator is basically used for pulse stretching. It is used in computer logic systems and communication navigation equipment.

The operation of the monostable multivibrator is relatively simple. The input is triggered with a pulse of voltage. The output changes from one voltage level to a different voltage level. The output remains at this new voltage level for a definite period of time. Then the circuit automatically reverts to its original condition and remains that way until another trigger pulse is applied to the input. The monostable multivibrator actually takes this series of input triggers and converts them to uniform square pulses, as shown in figure 3-10. All of the square output pulses are of the same amplitude and time duration.

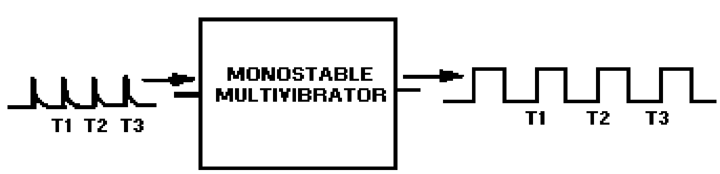


Figure 3-10.—Monostable multivibrator block diagram.

The schematic for a monostable multivibrator is shown in figure 3-11. Like the astable multivibrator, one transistor conducts and the other cuts off when the circuit is energized.

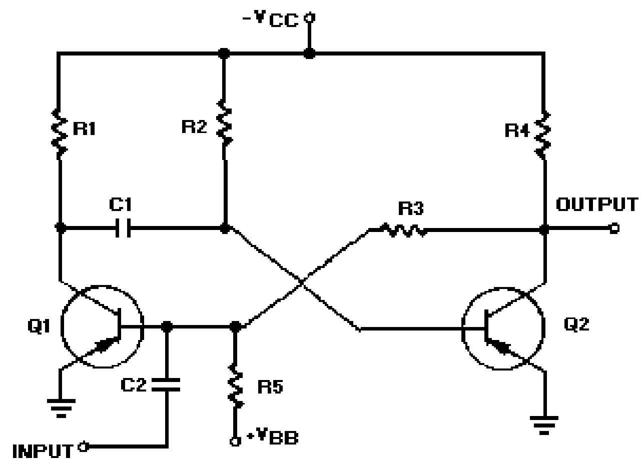


Figure 3-11.—Monostable multivibrator schematic.

Recall that when the astable multivibrator was first energized, it was impossible to predict which transistor would initially go to cutoff because of circuit symmetry. The one-shot circuit is not symmetrical like the astable multivibrator. Positive voltage V_{BB} is applied through R5 to the base of Q1. This positive voltage causes Q1 to cut off. Transistor Q2 saturates because of the negative voltage applied from $-V_{CC}$ to its base through R2. Therefore, Q1 is cut off and Q2 is saturated before a trigger pulse is applied, as shown in figure 3-12. The circuit is shown in its stable state.

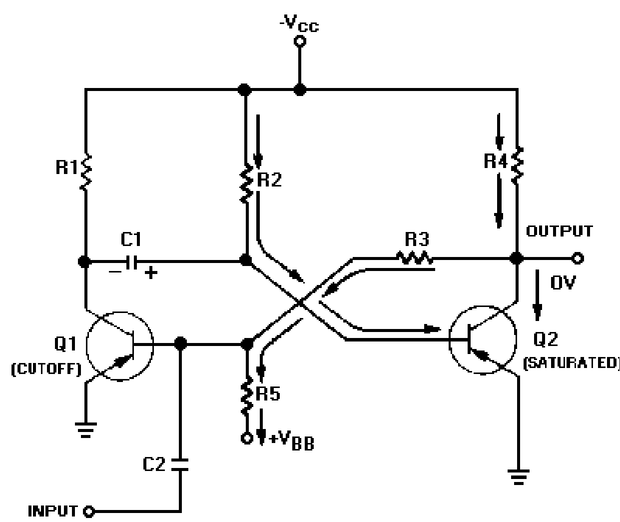


Figure 3-12.—Monostable multivibrator (stable state).

Let's take a more detailed look at the circuit conditions in this stable state (refer to figure 3-12). As stated above, Q1 is cut off, so no current flows through R1, and the collector of Q1 is at $-V_{CC}$. Q2 is saturated and has practically no voltage drop across it, so its collector is essentially at 0 volts. R5 and R3 form a voltage divider from V_{BB} to the ground potential at the collector of Q2. The tie point between these two resistors will be positive. Thus, the base of Q1 is held positive, ensuring that Q1 remains cutoff. Q2 will remain saturated because the base of Q2 is very slightly negative as a result of the voltage drop across R2. If the collector of Q1 is near $-V_{CC}$ and the base of Q2 is near ground, C1 must be charged to nearly V_{CC} volts with the polarity shown.

Now that all the components and voltages have been described for the stable state, let us see how the circuit operates (see figure 3-13). Assume that a negative pulse is applied at the input terminal. C2 couples this voltage change to the base of Q1 and starts Q1 conducting. Q1 quickly saturates, and its collector voltage immediately rises to ground potential. This sharp voltage increase is coupled through C1 to the base of Q2, causing Q2 to cut off; the collector voltage of Q2 immediately drops to V_{CC}. The voltage divider formed by R5 and R3 then holds the base of Q1 negative, and Q1 is locked in saturation.

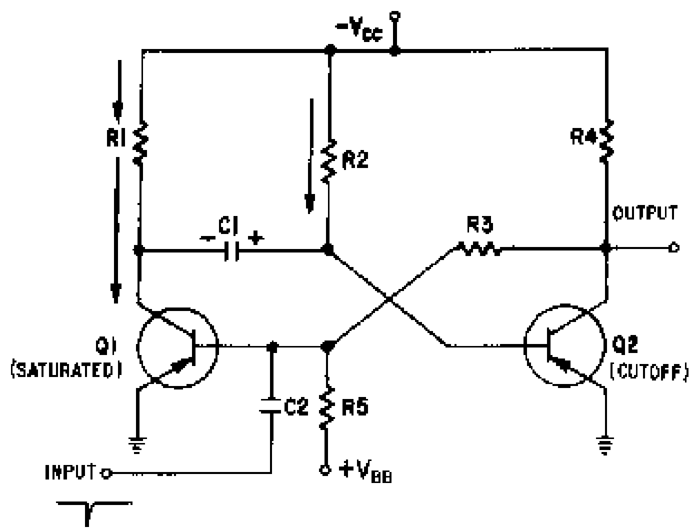


Figure 3-13.—Monostable multivibrator (triggered).

The one-shot multivibrator has now been turned on by applying a pulse at the input. It will turn itself off after a period of time. To see how it does this, look at figure 3-13 again. Q1 is held in saturation by the negative voltage applied through R3 to its base, so the circuit cannot be turned off here. Notice that the base of Q2 is connected to C1. The positive charge on C1 keeps Q2 cutoff. Remember that a positive voltage change (essentially a pulse) was coupled from the collector of Q1 when it began conducting to the base of Q2, placing Q2 in cutoff. When the collector of Q1 switches from $-V_{CC}$ volts to 0 volts, the charge on C1 acts like a battery with its negative terminal on the collector of Q1, and its positive terminal connected to the base of Q2. This voltage is what cuts off Q2. C1 will now begin to discharge through Q1 to ground, back through $-V_{CC}$, through R2 to the other side of C1. The time required for C1 to discharge depends on the RC time constant of C1 and R2. Figure 3-14 is a timing diagram that shows the negative input pulse and the resultant waveforms that you would expect to see for this circuit description.

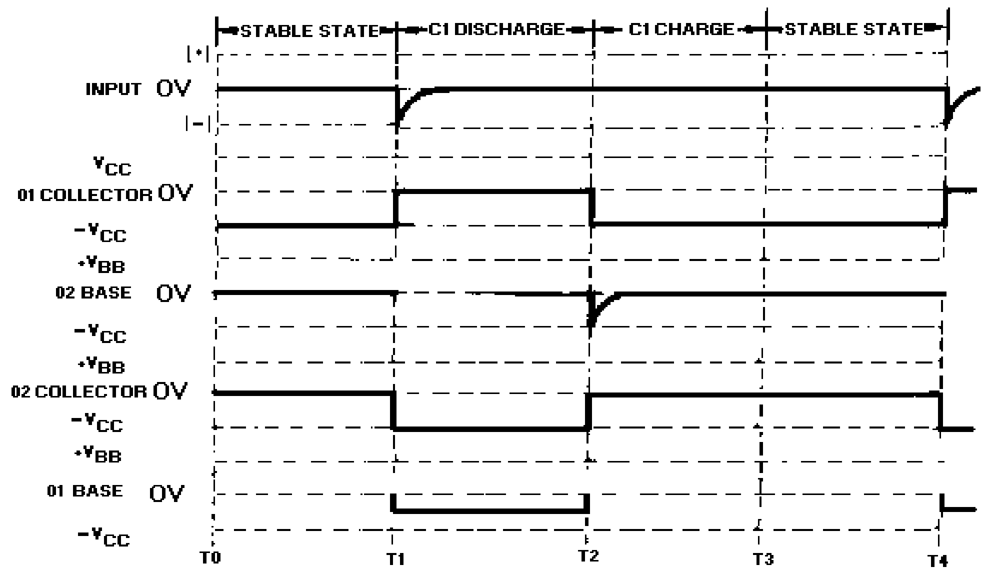


Figure 3-14.—Waveforms of a monostable multivibrator (triggered).

The only part of the operation not described so far is the short C1 charge time that occurs right after Q1 and Q2 return to their stable states. This is simply the time required for C1 to gain electrons on its left side. This charge time is determined by the R1C1 time constant.

Another version of the monostable multivibrator is shown in figure 3-15. View (A) is the circuit and view (B) shows the associated waveforms. In its stable condition (T0), Q1 is cut off and Q2 is conducting. The input trigger (positive pulse at T1) is applied to the collector of Q1 and coupled by C1 to the base of Q2 causing Q2 to be cut off. The collector voltage of Q2 then goes $-V_{CC}$. The more negative voltage at the collector of Q2 forward biases Q1 through R4. With the forward bias, Q1 conducts, and the collector voltage of Q1 goes to about 0 volts. C1 now discharges and keeps Q2 cut off. Q2 remains cut off until C1 discharges enough to allow Q2 to conduct again (T2). When Q2 conducts again, its collector voltage goes toward 0 volts and Q1 is cut off. The circuit returns to its quiescent state and has completed a cycle. The circuit remains in this stable state until the next trigger arrives (T3).

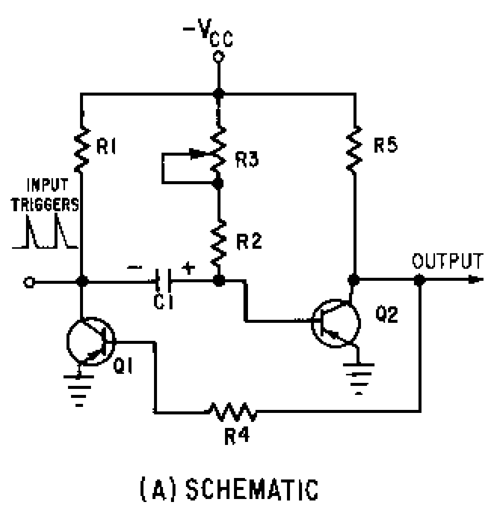


Figure 3-15A.—Monostable miltivibrator and waveshapes. Schematic.

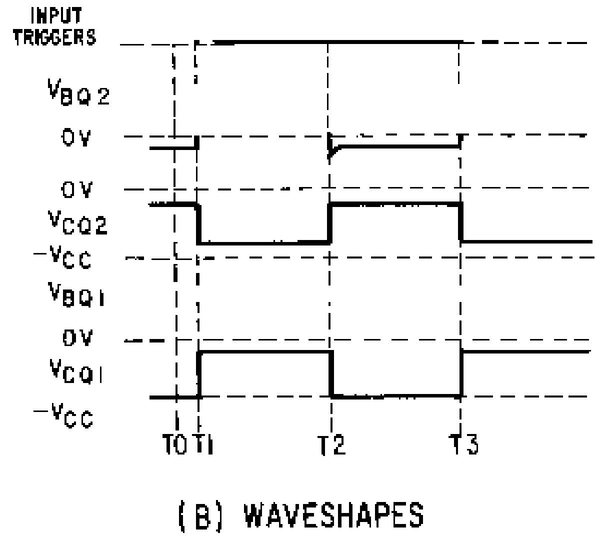


Figure 3-15B.—Monostable miltivibrator and waveshapes. Waveshapes

Note that R3 is variable to allow adjustment of the gate width. Increasing R3 increases the discharge time for C1 which increases the cutoff time for Q2. Increasing the value of R3 widens the gate. To decrease the gate width, decrease the value of R3. Figure 3-16 shows the relationships between the trigger and the output signal. View (A) of the figure shows the input trigger; views (B) and (C) show the different gate widths made available by R3. Although the durations of the gates are different, the duration of the complete cycle remains the same as the pulse repetition time of the triggers. View (D) of the figure illustrates that the trailing edge of the positive alternation is variable.

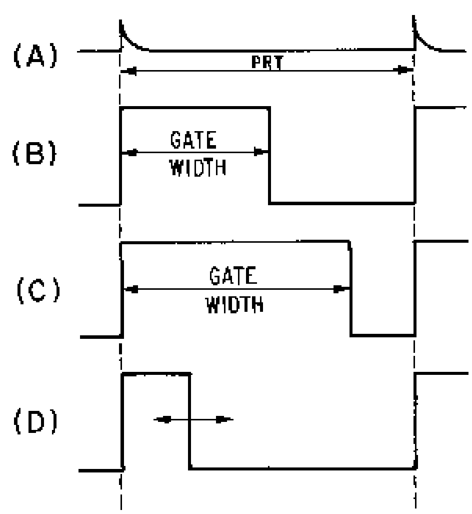


Figure 3-16.—Monostable multivibrator waveforms with a variable gate.

The reason the monostable multivibrator is also called a one-shot multivibrator can easily be seen. For every trigger pulse applied to the multivibrator, a complete cycle, or a positive and negative alternation of the output, is completed.

Q5. In an astable multivibrator, which components determine the pulse repetition frequency?

Q6. What is another name for the monostable multivibrator?

Bistable Multivibrator

As the name implies, the bistable multivibrator has two stable states. If a trigger of the correct polarity and amplitude is applied to the circuit, it will change states and remain there until triggered again. The trigger need not have a fixed prf; in fact, triggers from different sources, occurring at different times, can be used to switch this circuit.

The bistable multivibrator circuit and the associated waveforms are shown in figure 3-17, views (A) and (B), respectively. In this circuit, R1 and R7 are the collector load resistors. Voltage dividers R1, R2, and R5 provide forward bias for Q2; R7, R6, and R3 provide forward bias for Q1. These resistors also couple the collector signal from one transistor to the base of the other. Observe that this is direct coupling of feedback. This type of coupling is required because the circuit depends on input triggers for operation, not on RC time constants inside the circuit. Both transistors use common emitter resistor R4 which provides emitter coupling. C1 and C2 couple the input triggers to the transistor bases.

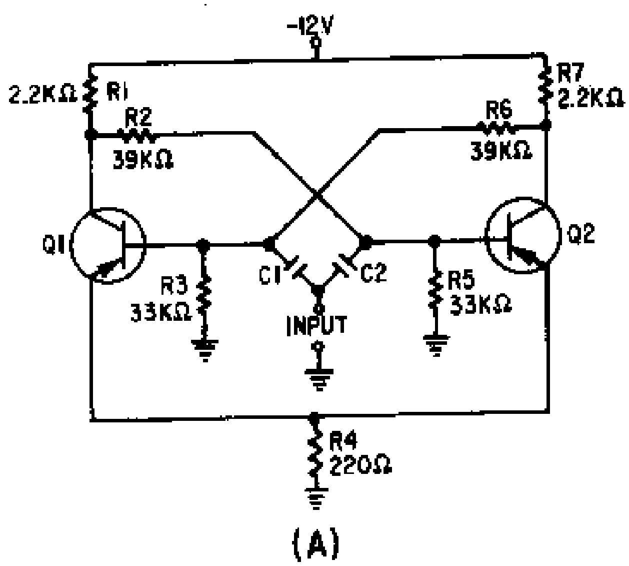


Figure 3-17A.—Bistable multivibrator and waveforms.

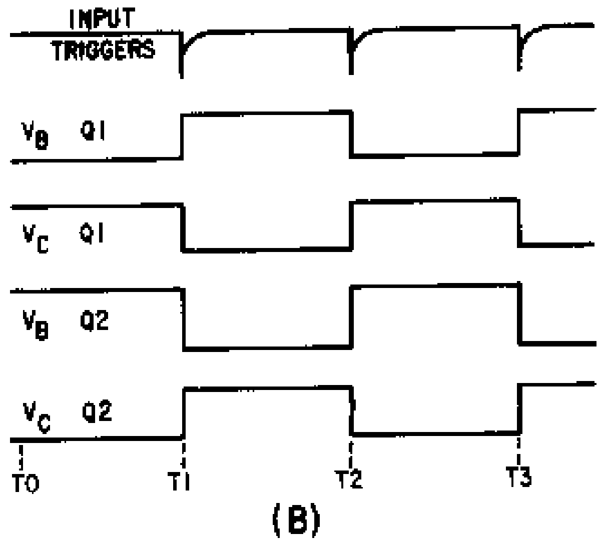


Figure 3-17B.—Bistable multivibrator and waveforms.

Notice that the circuit is symmetrical; that is, each transistor amplifier has the same component values. When power is first applied, the voltage divider networks place a negative voltage at the bases of Q1 and Q2. Both transistors have forward bias and both conduct.

Due to some slight difference between the two circuits, one transistor will conduct more than the other. Assume that Q1 conducts more than Q2. The increased conduction of Q1 causes the collector voltage of Q1 to be less negative (more voltage drop across R1). This decreases the forward bias of Q2 and decreases the conduction of Q2. When Q2 conducts less, its collector voltage becomes more negative. The negative-going change at the collector of Q2 is coupled to the base of Q1 and causes Q1 to conduct even more heavily. This regenerative action continues until Q2 is cut off and Q1 is saturated. The circuit is in a stable state and will remain there until a trigger is applied to change the state.

At T1, a negative trigger is applied to both bases through C1 and C2. The trigger does not affect Q1 since it is already conducting. The trigger overcomes cutoff bias on Q2 and causes it to conduct. As Q2 goes into conduction, its collector voltage becomes positive. The positive-going change at the Q2 collector causes a reverse bias on the base of Q1. As the conduction of Q1 decreases to the cutoff point, the collector voltage becomes negative. This switching action causes a very rapid change of state with Q2 now conducting and Q1 cut off.

At T2, a negative trigger is again applied to both bases. This time, Q1 is brought into conduction and the regenerative switching action cuts off Q2. The bistable multivibrator will continue to change states as long as triggers are applied. Notice that two input triggers are required to produce one gate; one to turn it on and the other to turn it off. The input trigger frequency is twice the output frequency.

The bistable multivibrator that most technicians know is commonly known by other names: the ECCLES-JORDAN circuit and, more commonly, the FLIP-FLOP circuit (figure 3-18). The flip-flop is a bistable multivibrator, "bi" meaning two; that is, the flip-flop has two stable states. The flip-flop (f/f) can rapidly flip from one state to the other and then flop back to its original state. If a voltmeter were connected to the output of a flip-flop, it would measure either a small positive or negative voltage, or a particularly low voltage (essentially 0 volts). No matter which voltage is measured, the flip-flop would be stable. Remember, stable means that the flip-flop will remain in a particular state indefinitely. It will not change states unless the proper type of trigger pulse is applied.

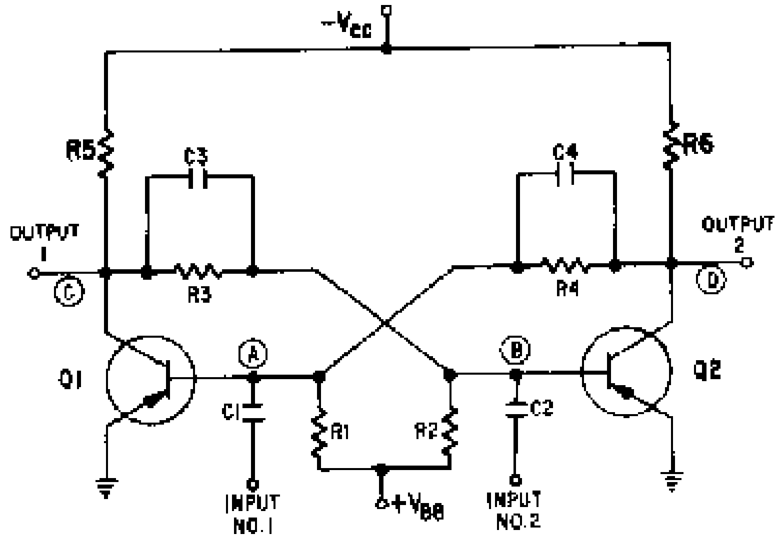


Figure 3-18.—Basic flip-flop.

Flip-flops are used in switching-circuit applications (computer logic operations) as counters, shift registers, clock pulse generators, and in memory circuits. They are also used for relay-control functions and for a variety of similar applications in radar and communications systems.

Notice that the basic flip-flop, illustrated in figure 3-18, has two inputs and two outputs. The inputs are coupled to the bases of the transistors and the outputs are coupled from the collectors of the transistors. Think of the flip-flop as two common-emitter amplifier circuits, where the output of one amplifier is connected to the input of the other amplifier, and vice-versa. Point (D) is connected through R4 to C4 to point (A). Point (A) is the input to transistor Q1. By the same token, point (C), which is the output of Q1, is connected through R3 and C3 to the input (point (B)) of transistor Q2.

Taking a close look at the flip-flop circuit, you should be able to see how it maintains its stable condition. Typical values for the resistors and applied voltages are shown in figure 3-19. The capacitors have been removed for simplicity.

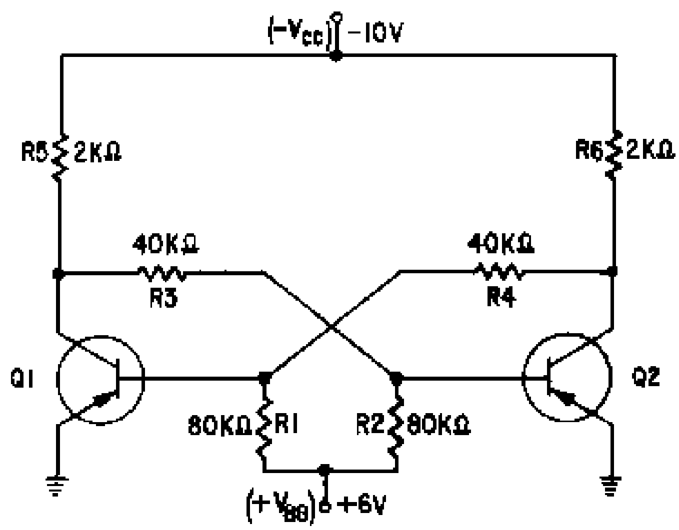


Figure 3-19.—Flip-flop (capacitors removed).

Two voltage-divider networks extend from -10 volts (V_{CC}) to +6 volts (V_{BB}). One voltage divider consisting of resistors R1, R4, and R6 supplies the bias voltage to the base of Q1. The other voltage divider consists of R2, R3, and R5 and supplies the bias voltage to the base of Q2.

Assume that Q1 (figure 3-20) is initially saturated and Q2 is cut off. Recall that the voltage drop from the base to the emitter of a saturated transistor is essentially 0 volts. In effect, this places the base of Q1 at ground potential. The voltages developed in the voltage divider, $-V_{CC}$, R6, R4, R1, and $+V_{BB}$, are shown in the figure.

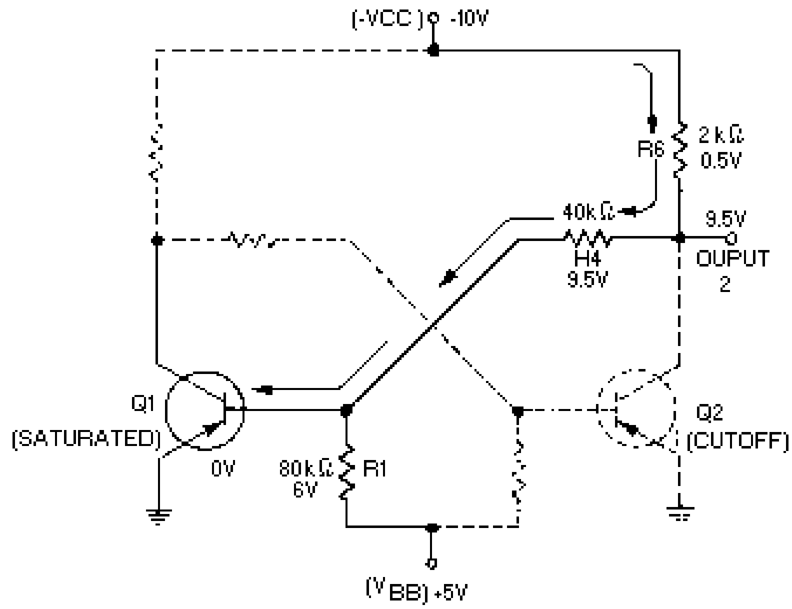


Figure 3-20.—Flip-flop (Q1 voltage divider).

Since no current flows through Q2, very little voltage is dropped across R6 (approximately 0.5 volt). The voltage at output 2 would measure -9.5 volts to ground (approximately $-V_{CC}$).

This voltage (–9.5 volts) is considered to be a HIGH output. Figure 3-21 shows the values of the other voltage-divider network.

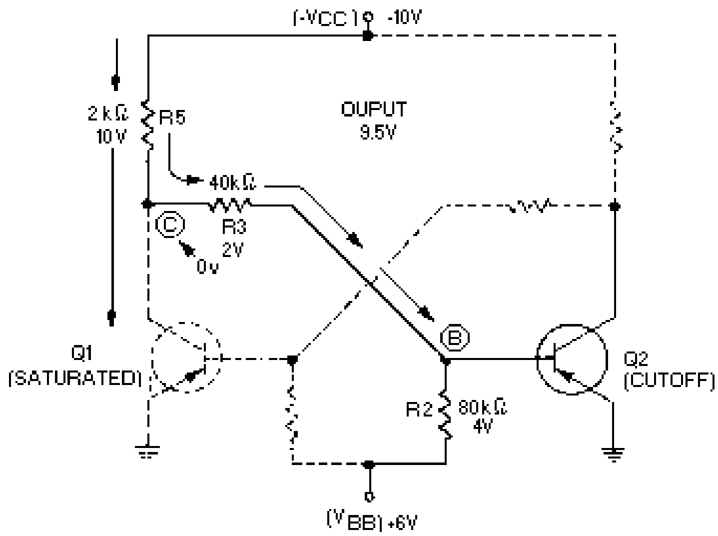


Figure 3-21.—Flip-flop (Q2 voltage divider).

With Q1 saturated, a large current flows through R5. The meter would measure approximately 0 volts (ground potential) at point (C). Notice that point (B) is located between point (C) (at 0 volts) and $+V_{BB}$ (at +6 volts). The meter would measure a positive voltage (between 0 volts and +6 volts) at the base of Q2 (point (B)).

A positive voltage on the base of a pnp transistor will cause that transistor to cut off. If one transistor is saturated, the other must be cut off. The flip-flop is stable in this state.

The capacitors that were removed from figure 3-19 must be returned to the flip-flop as shown in figure 3-22 to change the state of the flip-flop from one condition to the other. Capacitors C3 and C4 transmit almost instantaneously any changes in voltage from the collector of one transistor to the base of the other. Capacitors C1 and C2 are input coupling capacitors.

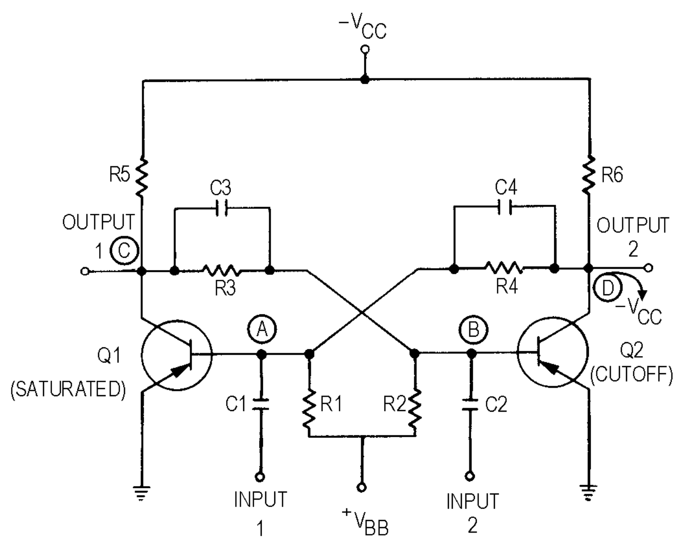


Figure 3-22.—Flip-flop.

As before, assume that transistor Q1 is saturated and transistor Q2 is cut off. Two methods are available to cause the flip-flop to change states. First, a positive-going pulse can be applied to input 1 to cause Q1 to change from saturation to cutoff. Second, the same result can be achieved by applying a negative-going pulse to input 2. Transistor Q2 would then change from Cutoff to saturation. Normally, a pulse is applied to the saturated transistor causing it to cut off. An input pulse which is of the correct polarity to change the state of the flip-flop is, as before, called a trigger pulse.

In figure 3-23 a positive-going trigger pulse has been applied to input 1. The flip-flop has now changed states; Q1 is cut off and Q2 is saturated. If a second positive-going trigger pulse is applied to input 1, it has no effect. This is because Q1 is already cut off; therefore, a positive pulse on its base has no effect. But if a positive-going trigger pulse were applied to input 2, the flip-flop would change back to its original state as shown in figure 3-24.

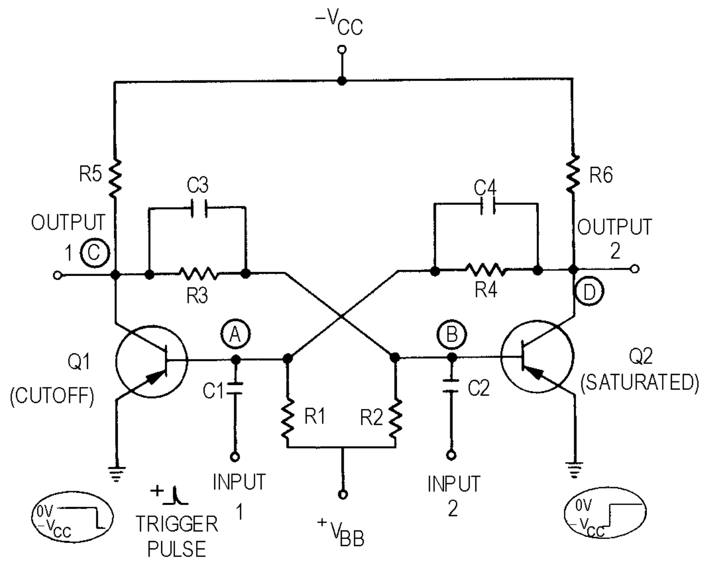


Figure 3-23.—Bistable multivibrator (flip-flop).

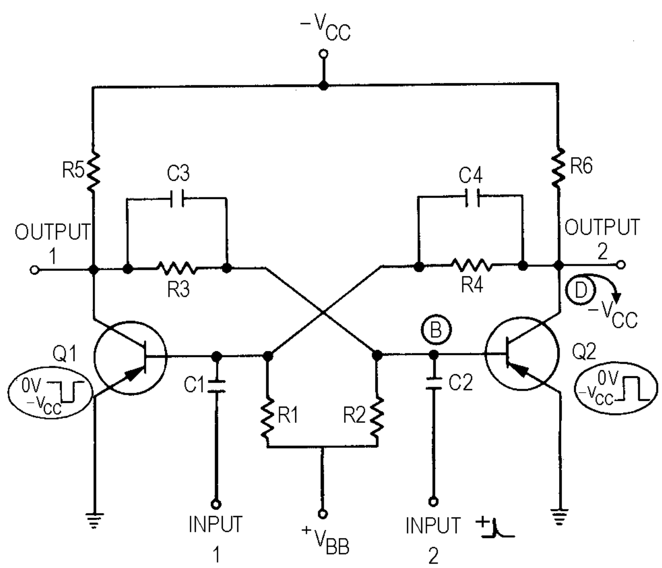


Figure 3-24.—Flip-flop (original state).

So far, the basic flip-flop has used only pnp transistors. It could have just as easily used npn transistors. The functional operation would not change; only the polarities required for conduction and cutoff change. As a technician, you may see either type of transistor used, npn or pnp. A symbolic block diagram is sometimes used to avoid confusion about voltage polarities.

A special kind of block diagram has been adopted as a standard symbol for the flip-flop and is shown in figures 3-25 and 3-26. The two inputs are represented by the lines on the left and the outputs by the lines on the right. INPUTS to a flip-flop are S (SET) and C (CLEAR) and OUTPUTS from a flip-flop are "1" and "0." A trigger pulse applied to the SET input causes the "1" output to be a positive or negative voltage, depending on the type of transistor. At the same time, the "0" output equals 0 volts. This condition is called the SET STATE.

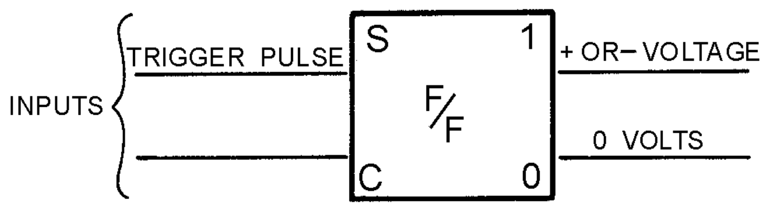


Figure 3-25.—Flip-flop (SET state).

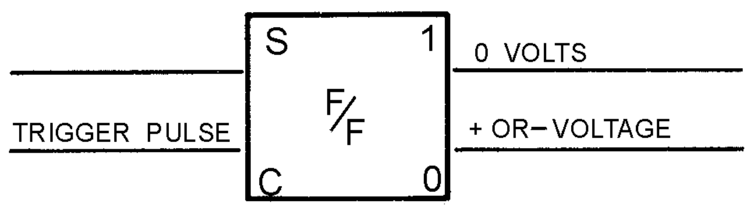


Figure 3-26.—Flip-flop (CLEAR state).

If a trigger pulse is applied to the CLEAR input, a positive or negative voltage is produced at the "0" output. The "1" output goes to 0 volts. This condition is called the CLEAR STATE, as shown in figure 3-26.

To determine what state the flip-flop is in, you can measure either the "1" or the "0" output. Measuring 0 volts at the "1" output indicates that the flip-flop is in the CLEAR state. If the "0" output is measured, a positive or negative voltage would also indicate that the flip-flop is in the CLEAR state. Either way, only one reading is necessary.

In figure 3-27, the flip-flop is in the SET state prior to T0 (negative voltage on the "1" output). Now compare the changes in output voltage at each point in time (T0, T1, T2, and T3) with the input pulse. Studying this figure should help you understand how the flip-flop works. The positive pulse at T0 on the CLEAR input shifts the f/f to the CLEAR state (negative voltage at the "0" output). At T1 a positive pulse on the SET input drives the "1" output to the SET state. At T2 a positive pulse on the CLEAR input drives the "0" output to a CLEAR state. At T3 another positive pulse is applied to the CLEAR input. This input has no effect since the f/f is already in the CLEAR state.

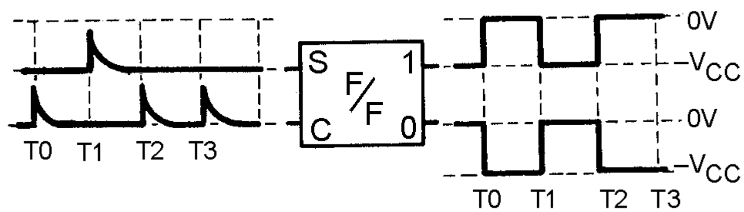


Figure 3-27.—Flip-flop with trigger pulse on SET and inputs.

Some flip-flops use a third input lead, as shown in figure 3-28. This third input lead is called a TOGGLE (T) input. Every time a pulse is applied to the T input, the flip-flop will change states from whatever its state was previously. The two diodes (CR1, CR2) form a STEERING NETWORK. This steering network directs a positive input pulse to the saturated transistor, causing it to cut off. Negative pulses are blocked by the diodes. Note that if npn transistors were used, the diodes would have to be reversed and the TOGGLE signal would have to be negative. For example, assume that Q1 is saturated, Q2 is cut off, and a positive pulse is applied the at T input. The input pulse will be directed to both transistors. The positive pulse will not affect Q2 since it is already in cutoff. Q1 however, which is conducting, will cut off and will cause Q2 to become saturated. The transistors have reversed states. A block diagram which represents a multivibrator and its outputs with only a TOGGLE input signal is shown in figure 3-29. Studying this figure should help you understand how this flip-flop works. Each TOGGLE input causes the output to change states. Figure 3-30 shows what happens when triggers are applied to all three inputs of the flip-flop shown in figure 3-28. Assume that the flip-flop in figure 3-30 is in the CLEAR state ("1" output is 0 volts, "0" output is high) prior to T0. At T0 a trigger is applied to the set input and the flip-flop changes states. Next, the CLEAR input is triggered and the flip-flop returns to the CLEAR state at T1. A TOGGLE at T2 causes the flip-flop to change state, so it is once again SET. Another TOGGLE changes the flip-flop to the CLEAR state at T3 (notice that TOGGLE triggers flip the multivibrator regardless of its previous state). Now, a SET input trigger at T4 sets the flip-flop. The CLEAR input pulse at T5 causes the circuit to CLEAR, and the CLEAR input at T6 has no effect on the flip-flop, for it is already in the CLEAR state.

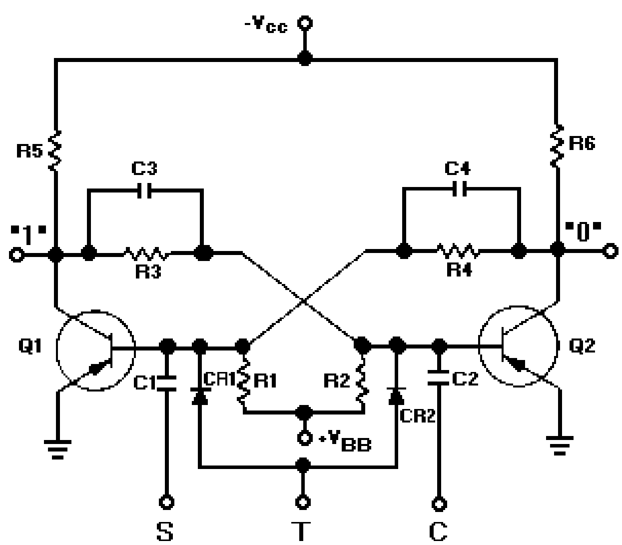


Figure 3-28.—Flip-flop with three inputs.

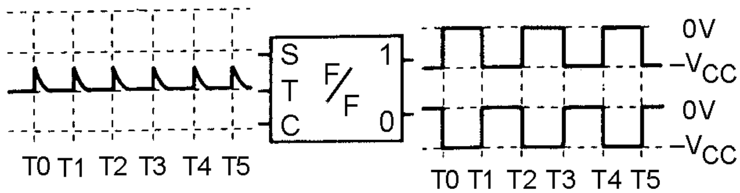


Figure 3-29.—Block diagram of a flip-flop with a toggle input.

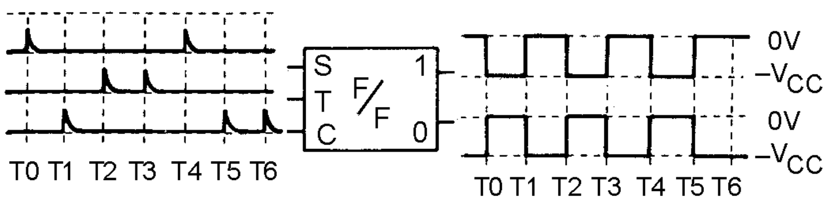


Figure 3-30.—Flip-flop with three inputs (block diagram).

Remember, a SET input will SET the flip-flop if it is in the CLEAR state, otherwise, it will not do anything; a trigger at the CLEAR input can only CLEAR the circuit if it is SET; and a trigger applied to the TOGGLE input will cause the bistable multivibrator to change states regardless of what state it is in.

Q7. In a bistable multivibrator, how many trigger pulses are needed to produce one complete cycle in the output?

Q8. How many stable states are there for a flip-flop?

Q9. If a voltage (positive or negative) is measured on the "1" output of a flip-flop, what state is it in?

BLOCKING OSCILLATOR

The BLOCKING OSCILLATOR is a special type of wave generator used to produce a narrow pulse, or trigger. Blocking oscillators have many uses, most of which are concerned with the timing of some other circuit. They can be used as frequency dividers or counter circuits and for switching other circuits on and off at specific times.

In a blocking oscillator the pulse width (pw), pulse repetition time (prt), and pulse repetition rate (prr) are all controlled by the size of certain capacitors and resistors and by the operating characteristics of the transformer. The transformer primary determines the duration and shape of the output. Because of their importance in the circuit, transformer action and series RL circuits will be discussed briefly. You may want to review transformer action in NEETS, Module 2, *Introduction to Alternating Current and Transformers* before going to the next section.

Transformer Action

Figure 3-31, view (A), shows a transformer with resistance in both the primary and secondary circuits. If S1 is closed, current will flow through R1 and L1. As the current increases in L1, it induces a voltage into L2 and causes current flow through R2. The voltage induced into L2 depends on the ratio of turns between L1 and L2 as well as the current flow through L1.

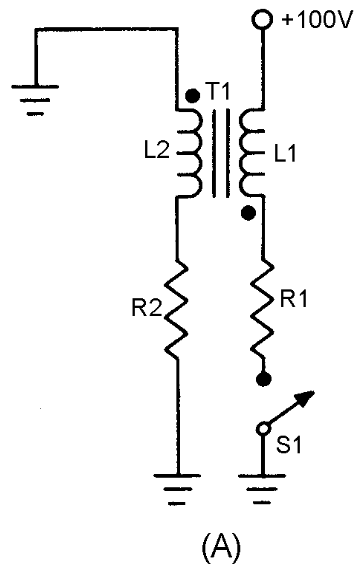


Figure 3-31A.—RL circuit.

The secondary load impedance, R2, affects the primary impedance through reflection from secondary to primary. If the load on the secondary is increased (R2 decreased), the load on the primary is also increased and primary and secondary currents are increased.

T1 can be shown as an inductor and R1-R2 as a combined or equivalent series resistance (R_E) since T1 has an effective inductance and any change in R1 or R2 will change the current. The equivalent circuit is shown in figure 3-31, view (B). It acts as a series RL circuit and will be discussed in those terms.

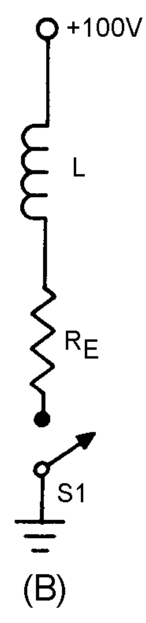


Figure 3-31B.—RL circuit.

Simple Series RL Circuit

When S1 is closed in the series RL circuit (view (B) of figure 3-31) L acts as an open at the first instant as source voltage appears across it. As current begins to flow, E_L decreases and E_R and I increase, all at exponential rates. Figure 3-32, view (A), shows these exponential curves. In a time equal to 5 time constants the resistor voltage and current are maximum and E_L is zero. This relationship is shown in the following formula:

$$5TC = \frac{L}{R_E} \times 5$$

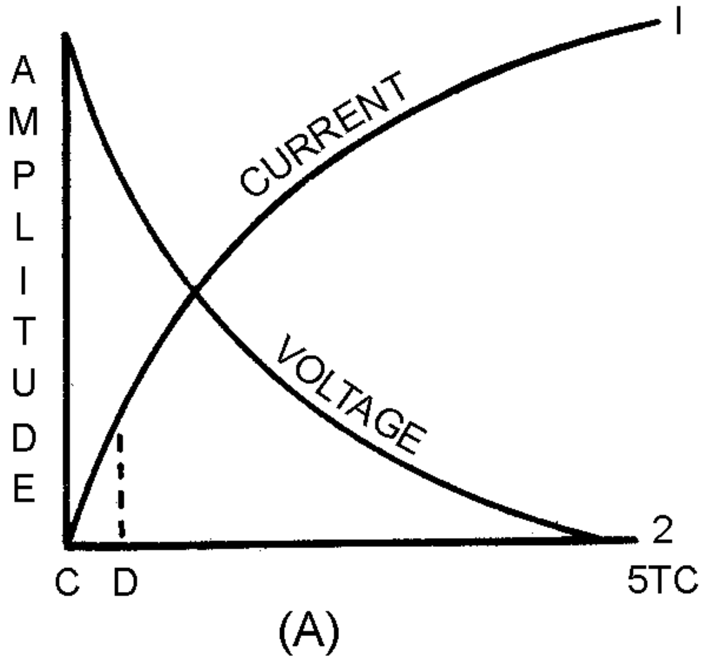


Figure 3-32A.—Voltage across a coil.

If S1 is closed, as shown in figure 3-31, view (B), the current will follow curve 1 as shown in figure 3-32, view (A). The time required for the current to reach maximum depends on the size of L and R_E. If R_E is small, then the RL circuit has a long time constant. If only a small portion of curve 1 (C to D of view (A)) is used, then the current increase will have maximum change in a given time period. Further, the smaller the time increment the more nearly linear is the current rise. A constant current increase through the coil is a key factor in a blocking oscillator.

Blocking Oscillator Applications

A basic principle of inductance is that if the increase of current through a coil is linear; that is, the rate of current increase is constant with respect to time, then the induced voltage will be constant. This is true in both the primary and secondary of a transformer. Figure 3-32, view (B), shows the voltage waveform across the coil when the current through it increases at a constant rate. Notice that this waveform is similar in shape to the trigger pulse shown earlier in figure 3-1, view (E). By definition, a blocking oscillator is a special type of oscillator which uses inductive regenerative feedback. The output duration and frequency of such pulses are determined by the characteristics of a transformer and its relationship to the circuit. Figure 3-33 shows a blocking oscillator. This is a simplified form used to illustrate circuit operation.

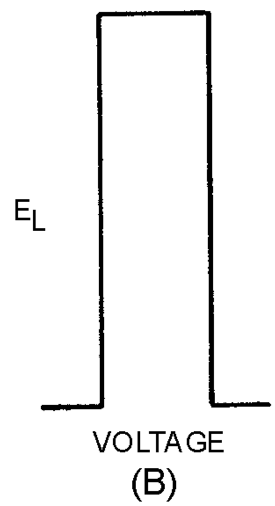


Figure 3-32B.—Voltage across a coil.

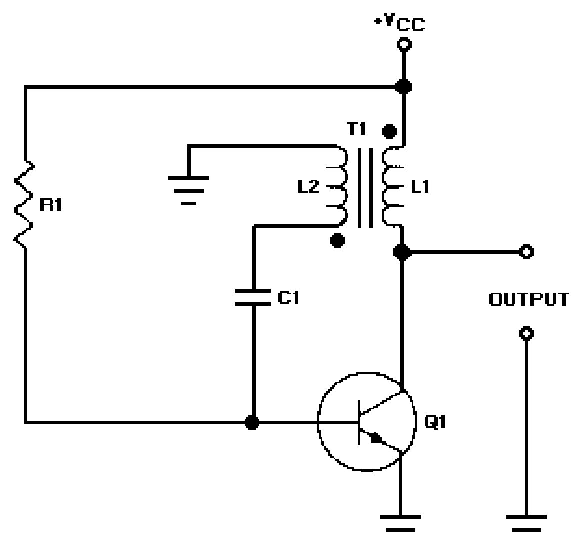


Figure 3-33.—Blocking oscillator.

When power is applied to the circuit, R1 provides forward bias and transistor Q1 conducts. Current flow through Q1 and the primary of T1 induces a voltage in L2. The phasing dots on the transformer indicate a 180-degree phase shift. As the bottom side of L1 is going negative, the bottom side of L2 is going positive. The positive voltage of L2 is coupled to the base of the transistor through C1, and Q1 conducts more. This provides more collector current and more current through L1. This action is regenerative feedback. Very rapidly, sufficient voltage is applied to saturate the base of Q1. Once the base becomes saturated, it loses control over collector current. The circuit now can be compared to a small resistor (Q1) in series with a relatively large inductor (L1), or a series RL circuit.

The operation of the circuit to this point has generated a very steep leading edge for the output pulse. Figure 3-34 shows the idealized collector and base waveforms. Once the base of Q1 (figure 3-33) becomes saturated, the current increase in L1 is determined by the time constant of L1 and the total series resistance. From T0 to T1 in figure 3-34 the current increase (not shown) is approximately linear. The voltage across L1 will be a constant value as long as the current increase through L1 is linear.

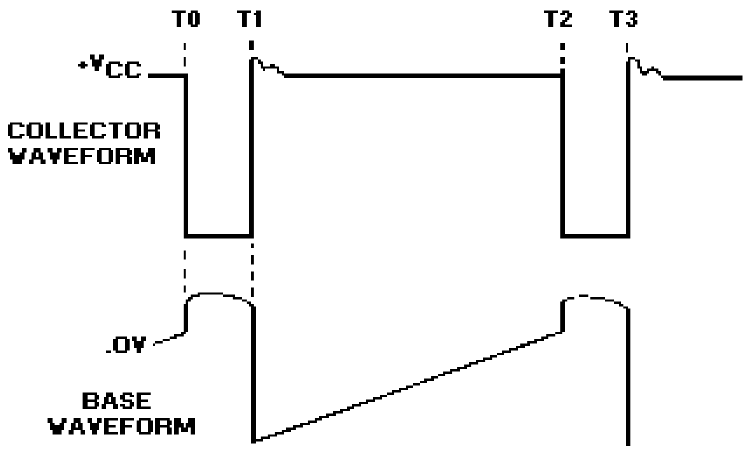


Figure 3-34.—Blocking oscillator idealized waveforms.

At time T1, L1 saturates. At this time, there is no further change in magnetic flux and no coupling from L1 to L2. C1, which has charged during time TO to T1, will now discharge through R1 and cut off Q1. This causes collector current to stop, and the voltage across L1 returns to 0.

The length of time between T0 and T1 (and T2 to T3 in the next cycle) is the pulse width, which depends mainly on the characteristics of the transformer and the point at which the transformer saturates. A transformer is chosen that will saturate at about 10 percent of the total circuit current. This ensures that the current increase is nearly linear. The transformer controls the pulse width because it controls the slope of collector current increase between points T0 and T1. Since $TC = L \div R$, the greater the L, the longer the TC. The longer the time constant, the slower the rate of current increase. When the rate of current increase is slow, the voltage across L1 is constant for a longer time. This primarily determines the pulse width.

From T1 to T2 (figure 3-34), transistor Q1 is held at cutoff by C1 discharging through R1 (figure 3-33). The transistor is now said to be "blocked." As C1 gradually loses its charge, the voltage on the base of Q1 returns to a forward-bias condition. At T2, the voltage on the base has become sufficiently positive to forward bias Q1, and the cycle is repeated.

The collector waveform may have an INDUCTIVE OVERSHOOT (PARASITIC OSCILLATIONS) at the end of the pulse. When Q1 cuts off, current through L1 ceases, and the magnetic field collapses, inducing a positive voltage at the collector of Q1. These oscillations are not desirable, so some means must be employed to reduce them. The transformer primary may be designed to have a high dc resistance resulting in a low Q; this resistance will decrease the amplitude of the oscillations. However, more damping may be necessary than such a low-Q transformer primary alone can achieve. If so, a DAMPING resistor can be placed in parallel with L1, as shown in figure 3-35.

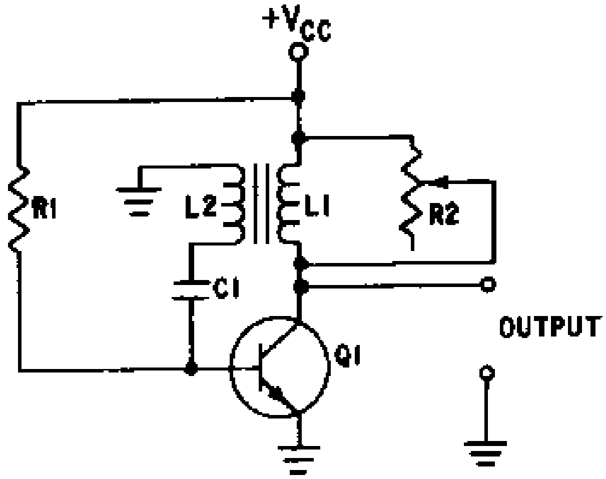


Figure 3-35.—Circuit damping.

When an external resistance is placed across a tank, the formula for the Q of the tank circuit is $Q = R/X_L$, where R is the equivalent total circuit resistance in parallel with L. You should be able to see from the equation that the Q is directly proportional to the damping resistance (R). In figure 3-35, damping resistor R2 is used to adjust the Q which reduces the amplitude of overshoot parasitic oscillations. As R2 is varied from infinity toward zero, the decreasing resistance will load the transformer to the point that pulse amplitude, pulse width, and prf are affected. If reduced enough, the oscillator will cease to function. By varying R2, varying degrees of damping can be achieved, three of which are shown in figure 3-36, view (A), view (B and view (C).

Figure 3-36A.—Waveform damping. CRITICAL DAMPING.

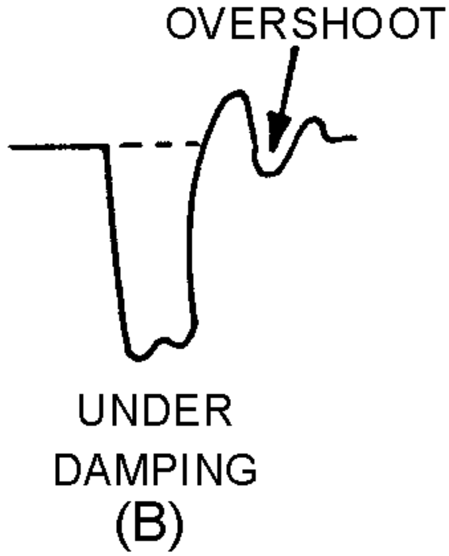


Figure 3-36B.—Waveform damping. UNDER DAMPING.

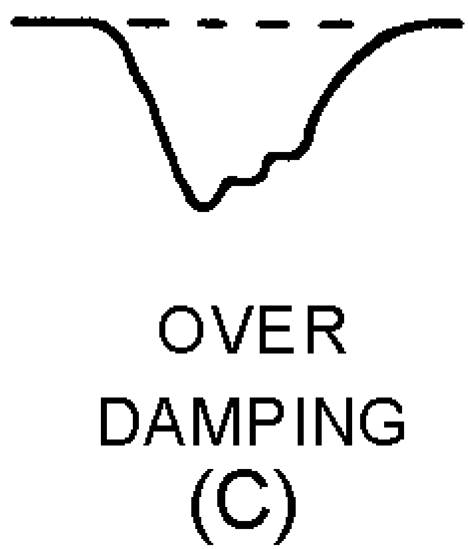


Figure 3-36C.—Waveform damping. OVER DAMPING.

CRITICAL DAMPING gives the most rapid transient response without overshoot. This is accomplished by adjusting R2 to achieve a waveform as shown in figure 3-36, view (A). The resistance of R2 depends upon the Q of the transformer. View (A) shows that oscillations, including the overshoot, are damping out.

UNDERDAMPING gives rapid transient response with overshoot caused by high or infinite resistance as shown in figure 3-36, view (B). OVERDAMPING is caused by very low resistance and gives a slow transient response. It may reduce the pulse amplitude as shown in figure 3-36, view (C).

The blocking oscillator discussed is a free-running circuit. For a fixed prf, some means of stabilizing the frequency is needed. One method is to apply external synchronization triggers (figure 3-37), view (A) and view (B). Coupling capacitor C2 feeds input synchronization (sync) triggers to the base of Q1.

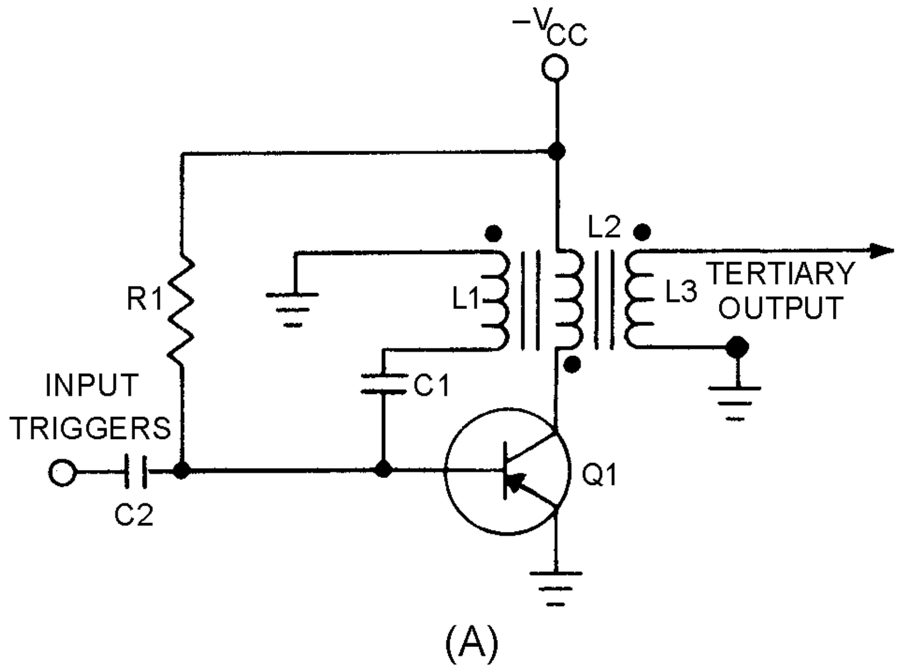


Figure 3-37A.—Blocking oscillator (synchronized).

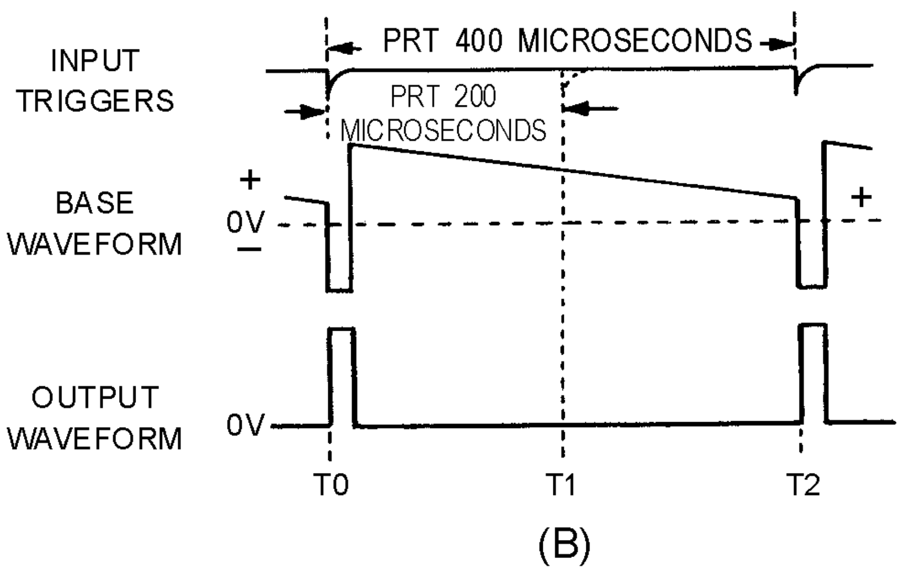


Figure 3-37B.—Blocking oscillator (synchronized).

If the trigger frequency is made slightly higher than the free-running frequency, the blocking oscillator will "lock in" at the higher frequency. For instance, assume the free-running frequency of this blocking oscillator is 2 kilohertz, with a prt of 500 microseconds. If sync pulses with a prt of 400 microseconds, or 2.5 kilohertz, are applied to the base, the blocking oscillator will "lock in" and run at 2.5 kilohertz. If the sync prf is too high, however, frequency division will occur. This means that if the sync prt is too short, some of the triggers occur when the base is far below cutoff. The blocking oscillator may then synchronize with every second or third sync pulse.

For example, in figure 3-37, view (A) and view (B) if trigger pulses are applied every 200 microseconds (5 kilohertz), the trigger that appears at T1 is not of sufficient amplitude to overcome the cutoff bias and turn on Q1. At T2, capacitor C1 has nearly discharged and the trigger causes Q1 to conduct. Note that with a 200-microsecond input trigger, the output prt is 400 microseconds. The output frequency is one-half the input trigger frequency and the blocking oscillator becomes a frequency divider.

Q10. What component in a blocking oscillator controls pulse width?

TIME-BASE GENERATORS

Radar sets, oscilloscopes, and computer circuits all use sawtooth (voltage or current) waveforms. A sawtooth waveshape must have a linear rise. The sawtooth waveform is often used to produce a uniform, progressive movement of an electron beam across the face of an electrostatic cathode ray tube. This movement of the electron beam is known as a SWEEP. The voltage which causes this movement is known as SWEEP VOLTAGE and the circuit which produces this voltage is the SWEEP GENERATOR, or TIME-BASE GENERATOR. Most common types of time-base generators develop the sawtooth waveform by using some type of switching action with either the charge or discharge of an RC or RL circuit.

Sawtooth Wave

A sawtooth wave can be generated by using an RC network. Possibly the simplest sawtooth generator is that which is shown in figure 3-38, view (A). Assume that at T0 (view (B)), S1 is placed in position P. At the instant the switch closes, the applied voltage (E_a) appears at R. C begins to charge to E_a through R. If S1 remains closed long enough, C will fully charge to E_a. You should remember from NEETS, Module 2, *Alternating Current and Transformers*, that a capacitor takes 5 time constants (5TC) to fully charge. As the capacitor charges to the applied voltage, the rate of charge follows an exponential curve. If a linear voltage is desired, the full charge time of the capacitor cannot be used because the exponential curve becomes nonlinear during the first time constant.

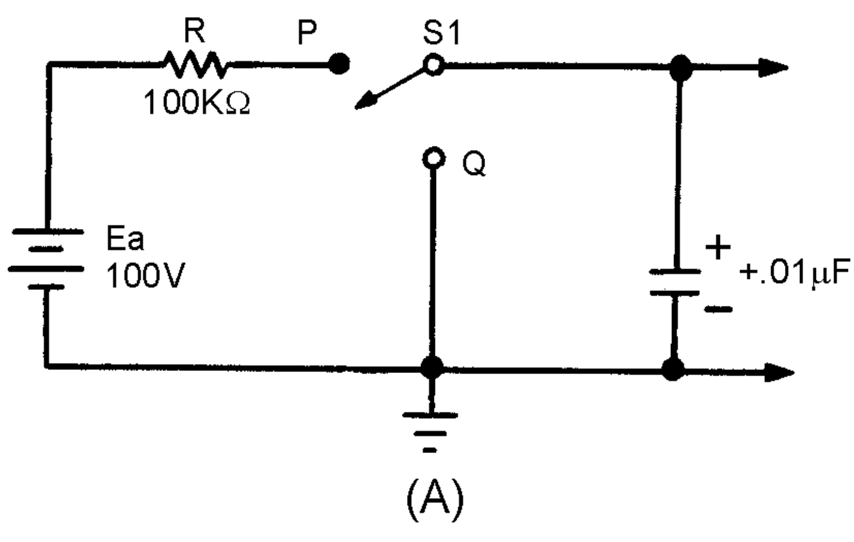


Figure 3-38A.—Series RC circuit.

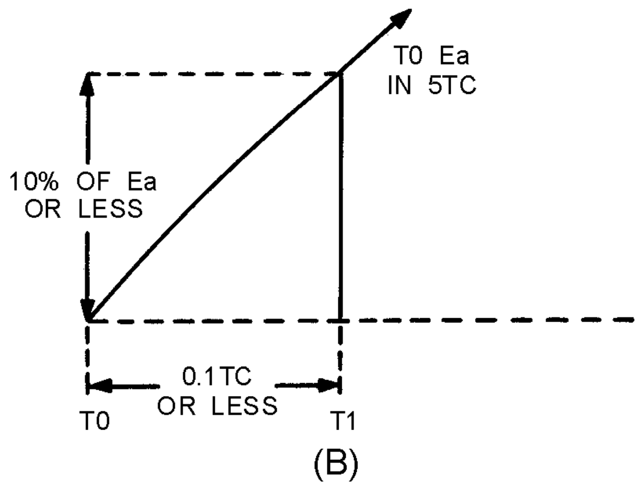


Figure 3-38B.—Series RC circuit.

However, during the first 10 percent of the first time constant, the rate of voltage change across the capacitor is almost constant (linear). Suppose that S1 is placed in position P at T0, and C is allowed to charge for 0.1 time constant. This is shown as T0 to T1 in view (B). Notice that the rate of voltage change across C is nearly constant between T0 and T1. Now, assume that at T1 the switch is moved from position P to position Q. This shorts the capacitor, and it discharges very rapidly. If the switch is placed back in position P, the capacitor will start charging again.

By selecting the sizes of R and C, you can have a time constant of any value you desire. Further, by controlling the time S1 remains closed, you can generate a sawtooth of any duration. Figure 3-39 is the Universal Time Constant Chart. Notice in the chart that if 1 time constant is 1,000 microseconds, S1 (figure 3-38, view (A)) can be closed no longer than 100 microseconds to obtain a reasonable linear sawtooth. In this example, C1 will charge to nearly 10 volts in 0.1 time constant.

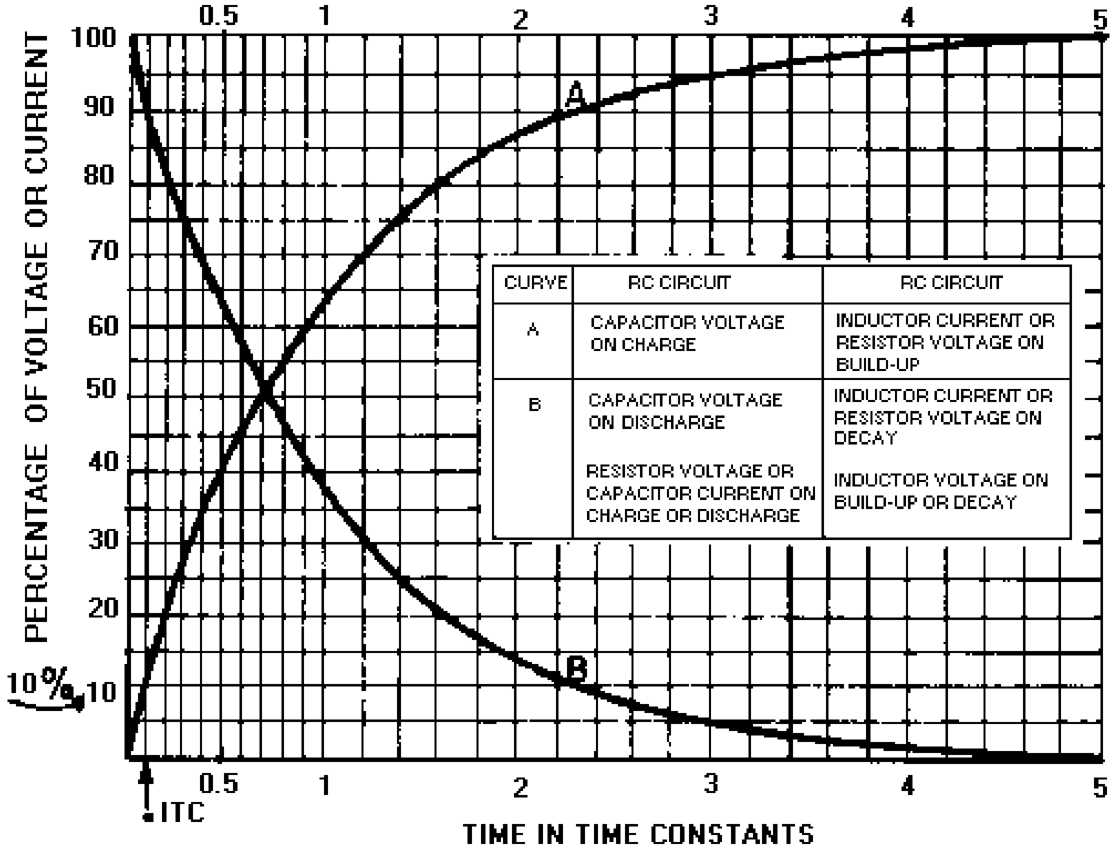


Figure 3-39.—Universal Time Constant Chart.

The dimensions of the sawtooth waveform used in oscilloscopes need to be discussed before going any further. Figure 3-40 shows a sawtooth waveform with the various dimensions labeled. The duration of the rise of voltage (T0 to T1) is known as the SWEEP TIME or ELECTRICAL LENGTH. The electron beam of an oscilloscope moves across the face of the cathode ray tube during this sweep time. The amount of voltage rise per unit of time is referred to as the SLOPE of the waveform. The time from T1 to T2 is the capacitor discharge time and is known as FALL TIME or FLYBACK TIME. This discharge time is known as flyback time because during this period the electron beam returns, or "flys" back, from the end of a scanning line to begin the next line.

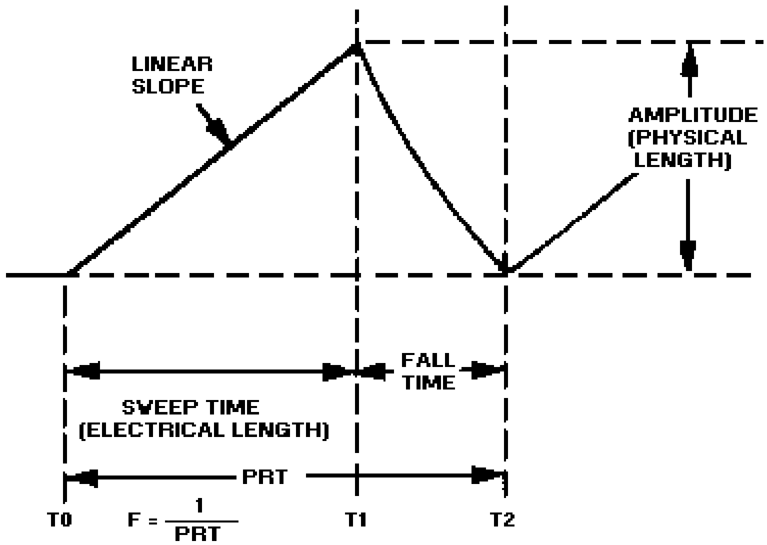


Figure 3-40.—Sawtooth waveform.

The amplitude of the rise of voltage is known as the PHYSICAL LENGTH. It is called physical length because the greater the peak voltage, the greater physical distance the beam will move. For example, the amount of voltage needed to move an electron beam 4 inches is twice the amount needed to move the beam 2 inches across the face of a given crt.

The voltage rise between T0 to T1 is the LINEAR SLOPE of the wave. The linearity of the rise of voltage is determined by the amount of time the capacitor is allowed to charge. If the charge time is kept short (10 percent or less of 1TC), the linearity is reasonably good.

As stated in the discussion of time-base generators, the waveform produced from any sawtooth generator must be linear. A LINEAR SAWTOOTH is one that has an equal change in voltage for an equal change in time. Referring to the Universal Time Constant Chart in figure 3-39, you can see that the most desirable part of the charge curve is the first one-tenth (0.1) of the first TC.

Figure 3-41, view (A), is a transistor sawtooth generator. In this figure R1 is a forward-biasing resistor for Q1, C1 is a coupling capacitor, and Q1 is serving as a switch for the RC network consisting of R2 and C2. With forward bias applied to Q1, the generator conducts at saturation, and its collector voltage (the output) is near 0 volts as indicated by the waveform in view (B). The charge felt by C1 is nearly 0. A negative gate is applied to the base of Q1 to cut off Q1 and allow C2 to charge. The length of time that the gate is negative determines how long Q1 will remain cut off and, in turn, how long C2 will be allowed to charge. The length of time that C2 is allowed to charge is referred to as the electrical length of the sawtooth that is produced.

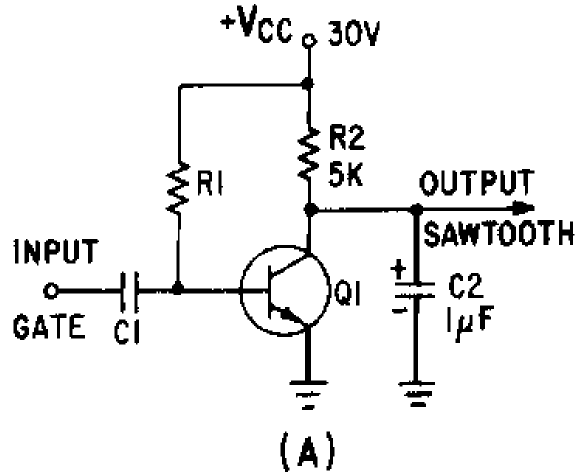


Figure 3-41A.—Transistor sawtooth generator.

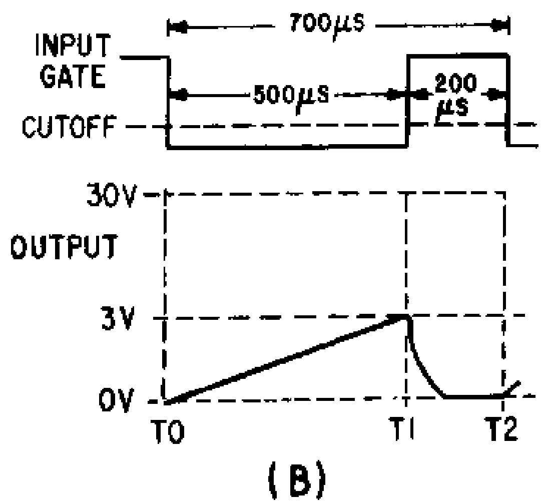


Figure 3-41B.—Transistor sawtooth generator.

The amplitude of the sawtooth that is produced is limited by the value of V_{CC} that is used in the circuit. For example, if the voltage is 30 volts, and the capacitor (C2) is allowed to charge to 10 percent of 30 volts, then the amplitude of the sawtooth will be 3 volts (see figure 3-41, view (B)). If V_{CC} is increased to 40 volts, C2 will charge to 10 percent of 40 volts and the output will increase in amplitude to 4 volts. Changing the value of V_{CC} in the circuit changes the amplitude of the sawtooth waveform that is produced; amplitude determines the physical length. Since the number of time constants used in the circuit has not been changed, linearity does not change with a change in V_{CC}.

The linear slope that is produced by the circuit is dependent on two variables; (1) the time constant of the RC circuit and (2) the gate length of the gate applied to the circuit. The circuit will produce a linear sawtooth waveshape if the components selected are such that only one-tenth of 1 TC or less is used. The GATE LENGTH is the amount of time that the gate is applied to the circuit and controls the time that the capacitor is allowed to charge. The value of R2 and C2 determines the time for 1 time constant

(TC = RC). To determine the number of time constants (or the fraction of 1TC) used, divide the time for 1 time constant into the time that the capacitor is allowed to charge:

$$\text{number of time constants} = \frac{\text{gate length}}{\text{TC}}$$

In figure 3-41, view (B), gate length is 500 microseconds and TC is the product of R2 (5 kilohms) and C2 (1 microfarad). The number of time constants is computed as follows:

$$\begin{aligned}\text{number of time constants} & \\ &= \frac{500 \times 10^{-6} \text{ seconds}}{(5 \times 10^{3})(1 \times 10^{-6} \text{ farads})} \\ &= \frac{500 \times 10^{-6}}{5 \times 10^{-3}} \\ &= 100 \times 10^{-3} \\ &= 0.1 \text{ TC}\end{aligned}$$

Therefore, 0.1TC is the length of time required to produce a linear rise in the sawtooth waveform.

The formula:

$$\text{number of time constants} = \frac{\text{gate length}}{\text{TC}}$$

shows that an increase in gate length increases the number of time constants. An increase in the number of time constants decreases linearity. The reason is that C2 now charges to a greater percentage of the applied voltage, and a portion of the charge curve is being used that is less linear. The waveform in figure 3-42, view (A), shows an increase in amplitude (physical length), an increase in the time that C2 is allowed to charge (electrical length), and a decrease in linearity. If a smaller percentage of V_{CC} is used, the gate length is decreased. As shown in view (B), this decreased gate length results in an increase in linearity, a decrease in the time that C2 is allowed to charge (electrical length), and a decrease in amplitude (physical length).

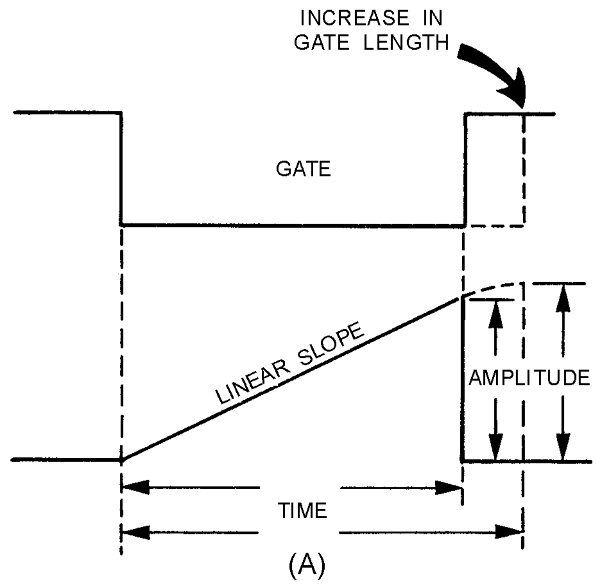


Figure 3-42A.—Relationship of gate to linearity.

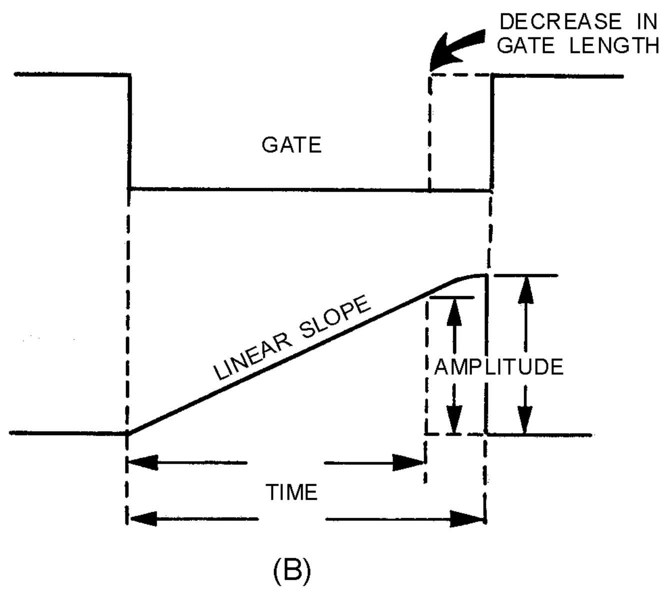


Figure 3-42B.—Relationship of gate to linearity.

Changing the value of R and C in the circuit affects linearity since they control the time for 1 time constant. For example, if the value of C2 is increased in the circuit, as shown in figure 3-43, view (A), the time for 1 time constant increases and the number of time constants then decreases. With a decrease in the number of time constants, linearity increases. The reason is that a smaller percentage of V_{CC} is used, and the circuit is operating in a more linear portion of the charge curve. Increasing the value of the TC (C2 or R2) decreases the amplitude of the sawtooth (physical length) because C2 now charges to a smaller percentage V_{CC} for a given time. The electrical length remains the same because the length of time that C2 is allowed to charge has not been changed.

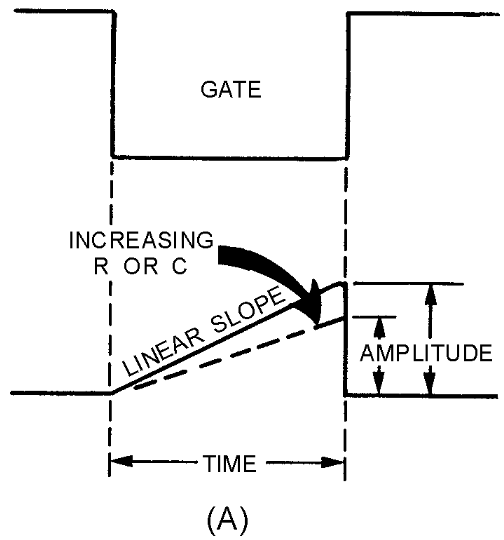


Figure 3-43A.—Relationship of R and C to linearity.

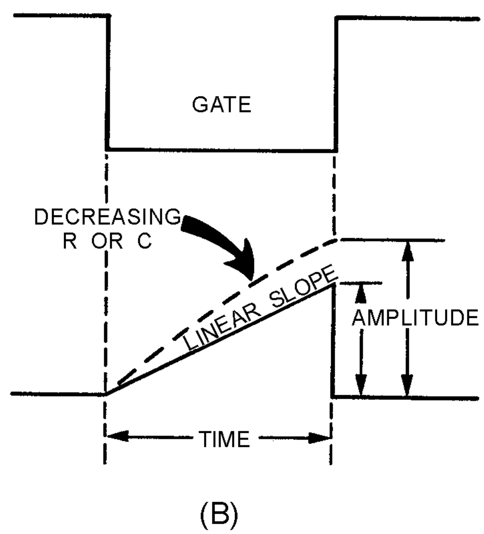


Figure 3-43B.—Relationship of R and C to linearity.

Decreasing the value of the TC (R2 or C2), as shown in figure 3-43, view (B), results in an increase in the number of time constants and therefore causes linearity to decrease. Anytime the number of time constants increases, the percentage of charge increases (see the Universal Time Constant Chart, figure 3-39), and amplitude (physical length) increases. Without an increase in gate length, the time that C2 is allowed to charge through R2 remains the same; therefore, electrical length remains the same. Linearity is affected by gate length, the value of R, and the value of C; but is not affected by changing the value of V_{CC}. Increasing the gate length decreases linearity, and decreasing gate length increases linearity. Increasing R or C in the circuit increases linearity, and decreasing R or C in the circuit decreases linearity.

The entire time of the sawtooth, from the time at which the capacitor begins charging (T0 in figure 3-41, view (B)) to the time when it starts charging again (T2), is known as the prt of the wave. The pulse repetition frequency of the sawtooth wave is:

$$prf = \frac{1}{prt}$$

UNIJUNCTION SAWTOOTH GENERATOR.—So far, you have learned in this chapter that a switch and an RC network can generate a sawtooth waveform. When using a unijunction transistor as the switch, a simple sawtooth generator looks like the circuit in figure 3-44, view (A); the output waveshapes are shown in view (B). You may want to review unijunction transistors in NEETS, Module 7, *Introduction to Solid-State Devices and Power Supplies*, chapter 3, before continuing.

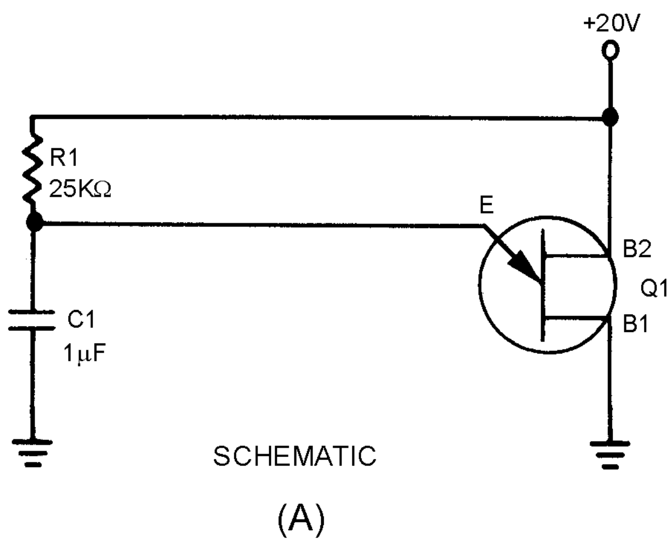


Figure 3-44A.—Unijunction sawtooth generator. SCHEMATIC.

When the 20 volts is applied across B2 and B1, the n-type bar acts as a voltage- divider. A voltage of 12.8 volts appears at a point near the emitter. At the first instant, C1 has no voltage across it, so the output of the circuit, which is taken across the capacitor (C1), is equal to 0 volts. (The voltage across C1 is also the voltage that is applied to the emitter of the unijunction.) The unijunction is now reverse biased. After T0, C1 begins to charge toward 20 volts.

At T1, the voltage across the capacitor (the voltage on the emitter) has reached approximately 12.8 volts. This is the peak point for the unijunction, and it now becomes forward biased. With the emitter forward biased, the impedance between the emitter and B1 is just a few ohms. This is similar to placing a short across the capacitor. The capacitor discharges very rapidly through the low resistance of B1 to E.

As C1 discharges, the voltage from the emitter to B1 also decreases. Q1 will continue to be forward biased as long as the voltage across C1 is larger than the valley point of the unijunction.

At T2 the 3-volt valley point of the unijunction has been reached. The emitter now becomes reverse biased and the impedance from the emitter to B1 returns to a high value. Immediately after T2, Q1 is reverse biased and the capacitor has a charge of approximately 3 volts. C1 now starts to charge toward 20 volts as it did originally (just after T0). This is shown from T2 to T3 in figure 3-44, view (B).

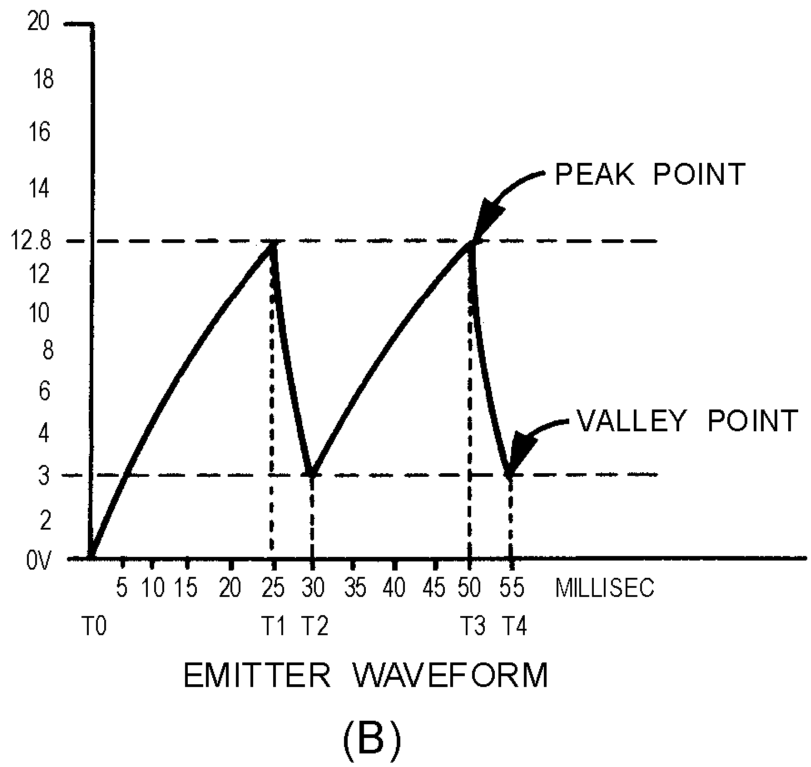


Figure 3-44B.—Unijunction sawtooth generator. EMITTER WAVEFORM.

The circuit operation from now on is just a continuous repetition of the actions between T2 and T4. The capacitor charges until the emitter becomes forward biased, the unijunction conducts and C1 discharges, and Q1 becomes reverse biased and C1 again starts charging.

Now, let's determine the linearity, electrical length, and amplitude of the output waveform. First, the linearity: To charge the circuit to the full 20 volts will take 5 time constants. In the circuit shown in figure 3-44, view (B), C1 is allowed to charge from T2 to T3. To find the percentage of charge, use the equation:

$$\text{percent of charge} = \frac{E_{peak} - E_{valley}}{E_a - E_{valley}} \times 100$$

$$= \frac{12.8 - 3}{20.0 - 3} \times 100$$

$$= \frac{9.9}{17} \times 100$$

$$= 57 \text{ percent}$$

This works out to be about 57 percent and is far beyond the 10 percent required for a linear sweep voltage. The linearity is very poor in this example.

The electrical length (sweep time), which is measured from T2 to T3, can be found by multiplying RC times the number of time constants. Refer to the Universal Time Constant Chart (figure 3-39) again to find that 57 percent is 0.83TC. By multiplying 0.83 times R1C1, you will find that the electrical length is approximately 21 milliseconds:

$$\begin{aligned}\text{electrical length} &= RC \times \text{number of TC}\\ &= 25\,k\Omega \times 1\mu F \times .83TC\\ &= 20.75 \times 10^{-3}\ \text{seconds}\\ &= 21\ \text{milliseconds}\end{aligned}$$

The physical length (amplitude) is determined by subtracting the valley point from the peak point. This is 9.8 volts in the example (12.8 volts – 3 volts).

For a sweep generator that produces a more linear output sawtooth waveform, refer to the circuit in figure 3-45, view (A). R1 and C1 form the RC time constant. Notice that the capacitor charges toward 35 volts (V_E) in this circuit.

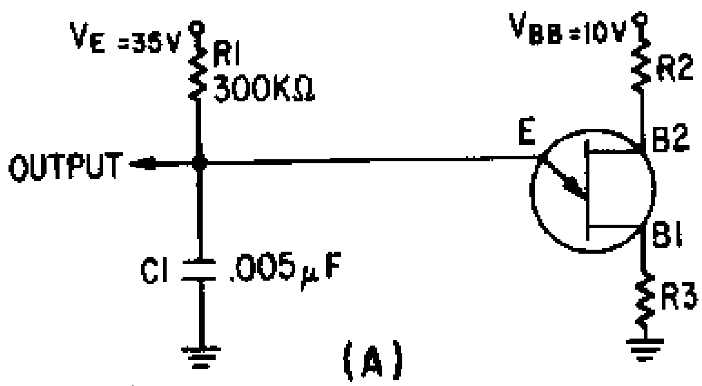


Figure 3-45A.—Improved unijunction sawtooth generator.

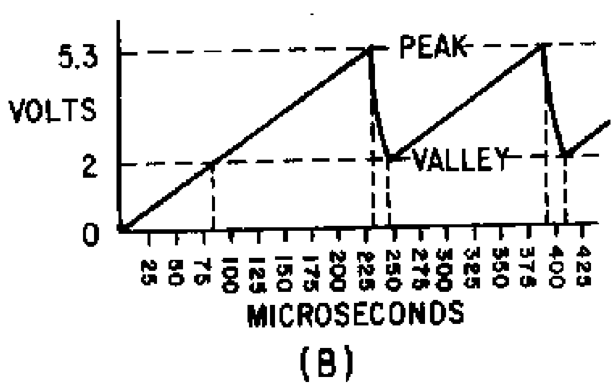


Figure 3-45B.—Improved unijunction sawtooth generator.

The output waveform is shown in figure 3-45, view (B). With a lower voltage applied from B1 to B2, the peak and valley points are closer together. Calculating the percentage of charge:

$$\text{percent of charge} = \frac{E_p - E_v}{E_a - E_v} \times 100$$

$$= \frac{5.3 - 2}{35 - 2} \times 100$$

$$= \frac{3.3}{33} \times 100$$

$$= 10 \text{ percent}$$

The linearity in this case is good. Using the Universal Time Constant Chart, a 10-percent charge amounts to 0.1 time constant. The electrical length is, again, RC times the number of time constants. With R1 at 300 kilohms and C1 at .005 microfarads, the time constant is 1,500 microseconds. One-tenth of a time constant is equal to 150 microseconds; so the electrical length is 150 microseconds. Prt is the electrical length plus the fall or flyback time. If C1 discharges from 5.3 volts to 2 volts in 15 microseconds, then the prt is 150 + 15, or 165 microseconds. The prf is about 6 kilohertz

$$(\text{prf} = \frac{1}{\text{prt}})$$

Some unijunction circuits are triggered to obtain a very stable prf. One method is to apply triggers to B2, as shown in figure 3-46. Negative triggers applied to B2 reduce the inter-base voltage enough to cause a forward bias condition in the emitter circuit. This cuts off the sweep and allows C1 to discharge through the B1-to-emitter circuit. Then, C1 recharges until the next trigger arrives and C1 discharges. Circuit operation and parameters are figured in the same manner as in the previous sawtooth circuits.

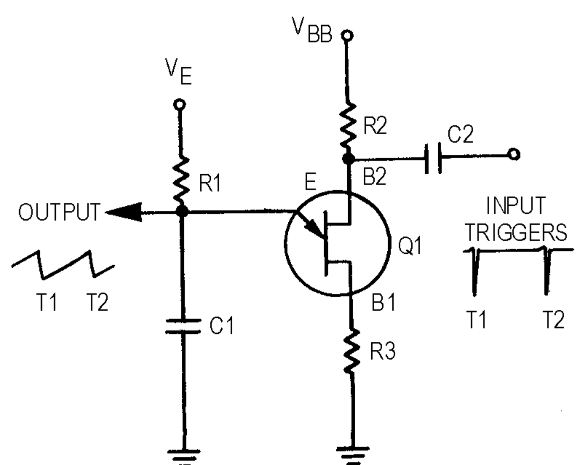


Figure 3-46.—Synchronized sawtooth generator.

TRANSISTOR SAWTOOTH GENERATOR.—The next sawtooth generator uses a conventional pnp transistor, as shown in figure 3-47, view (A). This generator also uses an RC network, and the transistor provides the switching action.

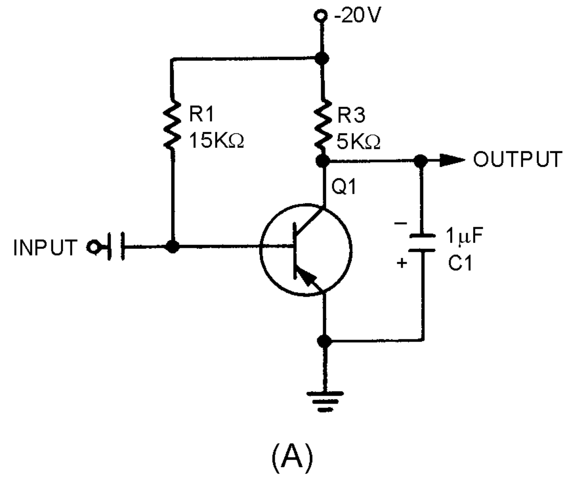


Figure 3-47A.—Transistor sawtooth generator (pnp).

The waveforms for the circuit are shown in views (B) and (C). With no input signals, Q1 is biased near saturation by R1. The voltage across C1 is very low (-2.5 volts) because load resistor R3 drops most of the applied voltage. The transistor must be cut off to allow C1 to charge. To cut off Q1, a positive rectangular wave is used.

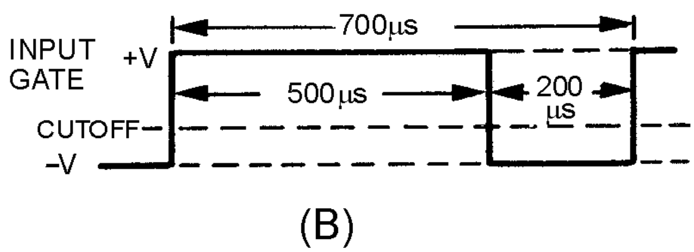


Figure 3-47B.—Transistor sawtooth generator (pnp).

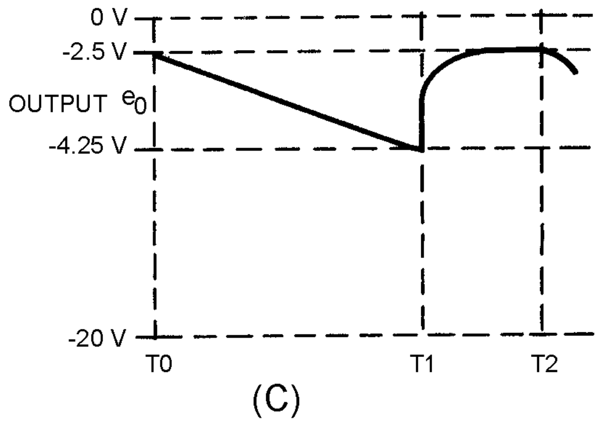


Figure 3-47C.—Transistor sawtooth generator (pnp).

Since Q1 is a pnp transistor, a positive voltage must be used to drive it to cutoff. Figure 3-47, view (B), shows a rectangular wave input 500 microseconds long on the positive alternation. At T0, the positive gate applied to the base of Q1 cuts off Q1. This effectively removes the transistor from the circuit (opens the switch), and C1 charges through R3 toward 20 volts. Starting with a charge of −2.5 volts at T0, C1 charges (T0 to T1) for 500 microseconds to -4.25 volts at T1. Let's determine the percent of charge:

$$\text{percent of charge} = \frac{E_C\,\text{max} - E_C\,\text{min}}{V_{CC} - E_C\,\text{min}} \times 100$$

$$= \frac{4.25 - 2.5}{20 - 2.5} \times 100$$

$$= \frac{1.75}{17.5} \times 100 = 10\%$$

This allows nearly a linear rise of voltage across C1.

Increasing the value of R3 or C1 increases the time constant. The capacitor will not charge to as high a voltage in the same period of time. Decreasing the width of the gate and maintaining the same time constant also prevents the capacitor from charging as much. With less charge on the capacitor, and the same voltage applied, linearity has been improved. Decreasing R3 or C1 or increasing gate width decreases linearity. Changing the applied voltage will change the charge on the capacitor. The percentage of charge remains constant; however, it does not affect linearity.

At T1, the positive alternation of the input gate ends, and Q1 returns to a forward-bias condition. A transistor that is near saturation has very low resistance, so C1 discharges rapidly between T1 and T2, as shown in figure 3-47, view (C). The capacitor discharges in less than 200 microseconds, the length of the negative alternation of the gate. The negative gate is made longer than the discharge time of the capacitor to ensure that the circuit has returned to its original condition.

From T1 to T2, the capacitor discharges and the circuit returns to its original condition, ready for another positive gate to arrive. The next positive gate arrives at T2 and the actions repeats.

The amplitude of the output sawtooth wave is equal to 1.75 volts (4.25 volts minus 2.5 volts). The electrical length is the same as the positive alternation of the input gate, or 500 microseconds. The prt is 700 microseconds (500 + 200) and the prf is 1/prt or 1,428 hertz.

Trapezoidal Sweep Generator

Normally, oscilloscopes and synchroscopes use ELECTROSTATIC DEFLECTION and, as the name implies, electrostatic fields move the electron beam. The need here is for a sawtooth voltage waveform.

Another method of electron beam deflection is ELECTROMAGNETIC DEFLECTION. Currents through a coil produce electromagnetic fields which position the beam of electrons. The electromagnetic system requires a sawtooth of current which increases at a linear rate. Because of the inherent characteristics of a coil, a sawtooth voltage does not cause a linear increase of current. A linear increase of current requires a TRAPEZOIDAL voltage waveform applied to a coil. This section discusses the generation of a trapezoidal wave.

Figure 3-48 shows a trapezoidal wave. The wave consists of a sharp, almost instantaneous jump in voltage followed by a linear rise to some peak value. The initial change in voltage at T0 is called a JUMP or STEP. The jump is followed by a linear sawtooth voltage rise. The time from the jump to the peak amplitude is the sum of the jump voltage and the sawtooth peak; where the peak value occurs is the electrical length. The peak voltage amplitude is the sum of the jump voltage and the sawtooth peak voltage. The waveshape can be considered a combination of a rectangular wave and a sawtooth wave.

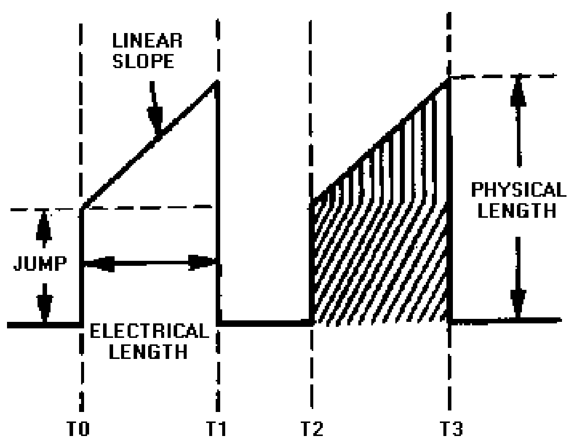


Figure 3-48.—Trapezoidal waveform.

The inductance and resistance of a coil form a series RL circuit. The voltage drop across this inductance and resistance must be added to obtain the voltage waveform required to produce a linear rise in current. A linear rise of current produces a linear rise of voltage across the resistance of the coil and a constant voltage drop across the inductance of the coil.

Assume figure 3-49, view (A), represents deflection coils. If we apply a voltage waveshape to the circuit, which will provide a square wave across inductor L, and a sawtooth across resistor R, then a linear current rise will result.

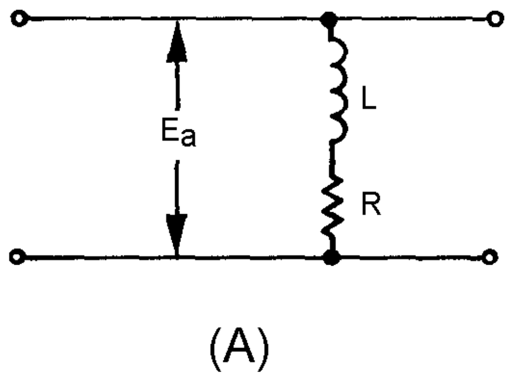


Figure 3-49A.—Series LR circuit.

View (B) of figure 3-49 shows the waveforms when E_a is a square wave. Recall that the inductor acts as an open circuit at this first instant. Current now starts to flow and develops a voltage across the resistor. With a square wave applied, the voltage across the inductor starts to drop as soon as any voltage appears across the resistor. This is due to the fact that the voltage across the inductor and resistor must add up to the applied voltage.

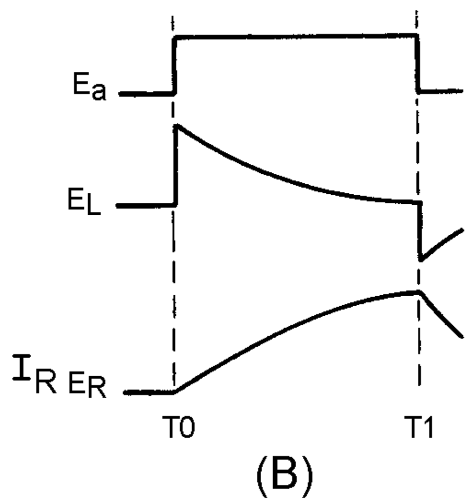


Figure 3-49B.—Series LR circuit.

With E_a being a trapezoidal voltage, as shown in figure 3-49, view (C), the instant current flows, a voltage appears across the resistor, and the applied voltage increases. With an increasing applied voltage, the inductor voltage remains constant (E_L) at the jump level and circuit current (I_R) rises at a linear rate from the jump voltage point. Notice that if you add the inductor voltage (E_L) and resistor voltage (E_R) at any point between times T0 and T1, the sum is the applied voltage (E_a). The key fact here is that a trapezoidal voltage must be applied to a sweep coil to cause a linear rise of current. The linear rise of current will cause a uniform, changing magnetic field which, in turn, will cause an electron beam to move at a constant rate across a crt.

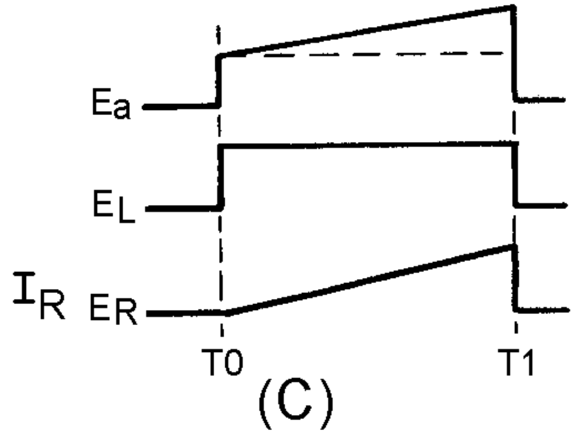


Figure 3-49C.—Series LR circuit.

There are many ways to generate a trapezoidal waveshape. For example, the rectangular part could be generated in one circuit, the sawtooth portion in another, and the two combined waveforms in still a third circuit. A far easier, and less complex, way is to use an RC circuit in combination with a transistor to generate the trapezoidal waveshape in one stage.

Figure 3-50, view (A), shows the schematic diagram of a trapezoidal generator. The waveshapes for the circuit are shown in view (B). R1 provides forward bias for Q1 and, without an input gate, Q1 conducts very hard (near saturation), C1 couples the input gate signal to the base of Q1. R2, R3, and C2 form the RC network which forms the trapezoidal wave. The output is taken across R3 and C2.

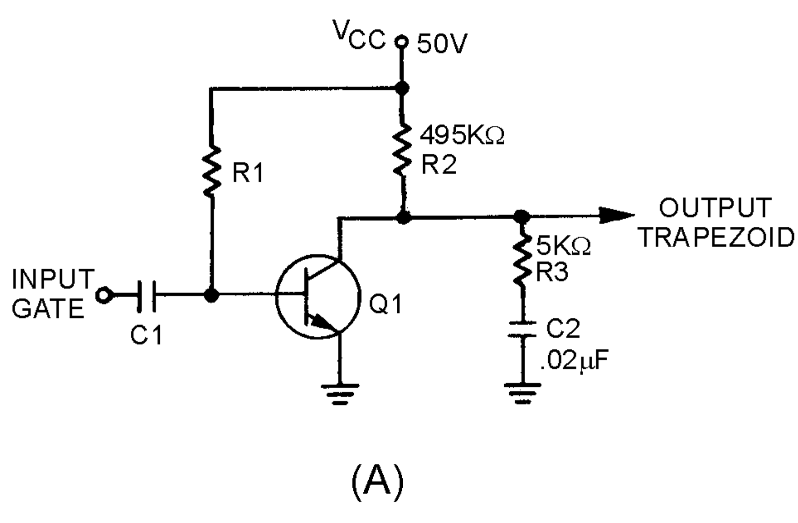


Figure 3-50A.—Trapezoidal waveform generator.

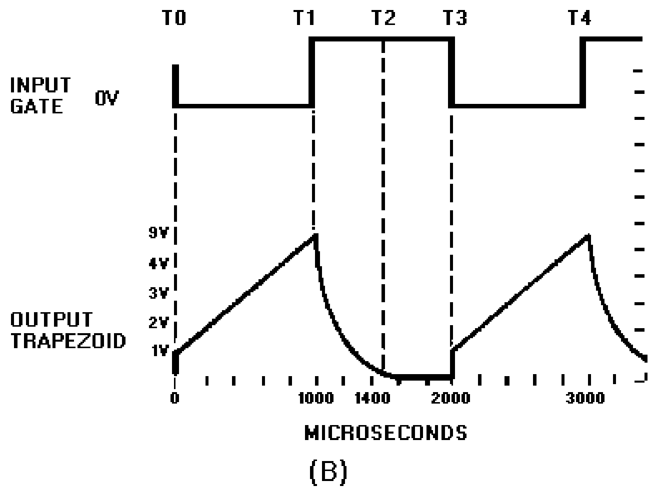


Figure 3-50B.—Trapezoidal waveform generator.

With Q1 conducting very hard, collector voltage is near 0 volts prior to the gate being applied. The voltage across R2 is about 50 volts. This means no current flows across R3, and C2 has no charge.

At T0, the negative alternation of the input gate is applied to the base of Q1, driving it into cutoff. At this time the transistor is effectively removed from the circuit. The circuit is now a series-RC network with 50 volts applied. At the instant Q1 cuts off, 50 volts will appear across the combination of R2 and R3 (the capacitor being a short at the first instant). The 50 volts will divide proportionally, according to the size of the two resistors. R2 then will have 49.5 volts and R3 will have 0.5 volt. The 0.5 volt across R3 (jump resistor) is the amplitude of the jump voltage. Since the output is taken across R3 and C2 in series, the output "jumps" to 0.5 volt.

Observe how a trapezoidal generator differs from a sawtooth generator. If the output were taken across the capacitor alone, the output voltage would be 0 at the first instant. But splitting the R of the RC network so that the output is taken across the capacitor and part of the total resistance produces the jump voltage.

Refer again to figure 3-50, view (A) and view (B). From T0 to T1, C2 begins charging toward 50 volts through R2 and R3. The time constant for this circuit is 10 milliseconds. If the input gate is 1,000 microseconds, the capacitor can charge for only 10 percent of 1TC, and the sawtooth part of the trapezoidal wave will be linear.

At T1, the input gate ends and Q1 begins to conduct heavily. C2 discharges through R3 and Q1. The time required to discharge C2 is primarily determined by the values of R3 and C2. The minimum discharge time (in this circuit) is 500 microseconds ($5K\Omega \times .02\mu F \times 5$). At T2, the capacitor has discharged back to 0 volts and the circuit is quiescent. It remains in this condition until T3 when another gate is applied to the transistor.

The amplitude of the jump voltage was calculated to be 0.5 volt. The sawtooth portion of the wave is linear because the time, T0 to T1, is only 10 percent of the total charge time. The amplitude of the trapezoidal wave is approximately 5 volts. The electrical length is the same as the input gate length, or 1,000 microseconds. Linearity is affected in the same manner as in the sawtooth generator. Increasing R2 or C2, or decreasing gate width, will improve linearity. Changing the applied voltage will increase output amplitude, but will not affect linearity.

Linearity of the trapezoidal waveform, produced by the circuit in figure 3-50, view (A) and view (B) depends on two factors, gate length and the time constant of the RC circuit. Recall that these are the same factors that controlled linearity in the sawtooth generator. The formula developed earlier still remains true and enables us to determine what effect these factors have on linearity.

$$\text{number of time constants} = \frac{\text{gate length}}{\text{TC}}$$

An increase in gate length results in an increase in the number of time constants and an increase in the percentage of charge that the capacitor will take on during this time interval. As stated earlier, if the number of time constants were to exceed 0.1, linearity would decrease. The reason for a decrease in linearity is that a greater percentage of V_{CC} is used. The Universal Time Constant Chart (figure 3-39) shows that the charge line begins to curve. A decrease in gate length has the opposite effect on linearity in that it causes linearity to increase. The reason for this increase is that a smaller number of time constants are used and, in turn, a smaller percentage of the applied V_{CC} is used.

Changing the value of resistance or capacitance in the circuit also affects linearity. If the value of C2 or R3 is increased, the time is increased for 1 time constant. An increase in the time for 1TC results in a decrease in the number of time constants required for good linearity. As stated earlier, a decrease in the number of time constants results in an increase in linearity (less than 0.1TC). In addition to an increase in jump voltage (larger value of R3) and a decrease in the amplitude (physical length) of the sawtooth produced by the circuit, electrical length remains the same because the length of the gate was not changed.

R2 has a similar effect on linearity because it is in series with R3. As an example, decreasing the value of R2 results in a decrease in linearity. The equation

$$\text{number of time constants} = \frac{\text{gate length}}{\text{TC}}$$

illustrates that by decreasing R (TC = RC), TC decreases and an increase in the number of time constants causes a decrease in linearity. Other effects are an increase in jump voltage and an increase in the amplitude (physical length) of the sawtooth.

Changing the value of V_{CC} does not affect linearity. Linearity is dependent on gate length, R, and C. V_{CC} does affect the amplitude of the waveform and the value of jump voltage that is obtained.

Q11. For an RC circuit to produce a linear output across the capacitor, the voltage across the capacitor may not exceed what percent of the applied voltage?

Q12. Increasing gate length in a sawtooth generator does what to linearity?

Q13. In a sawtooth generator, why is the transistor turned on for a longer time than the discharge time of the RC network?

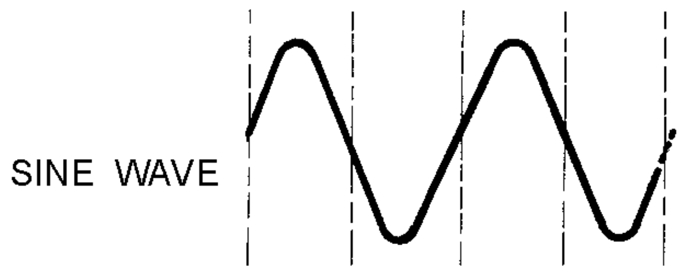
Q14. What is added to a sawtooth generator to produce a trapezoidal wave?

SUMMARY

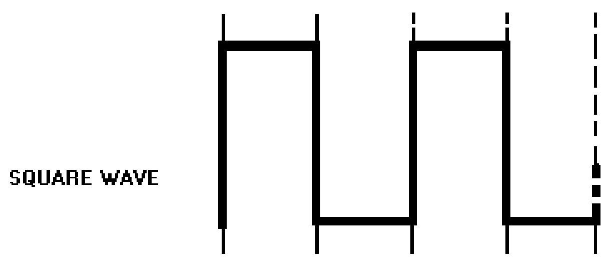
This chapter has presented information on waveforms and wave generators. The information that follows summarizes the important points of this chapter.

A waveform which undergoes a pattern of changes, returns to its original pattern, and repeats that same pattern of changes is called a PERIODIC waveform.

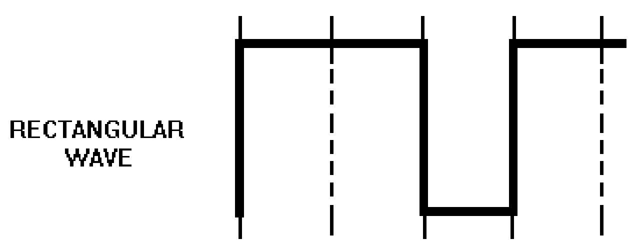
Each completed pattern of a waveform is called a CYCLE.



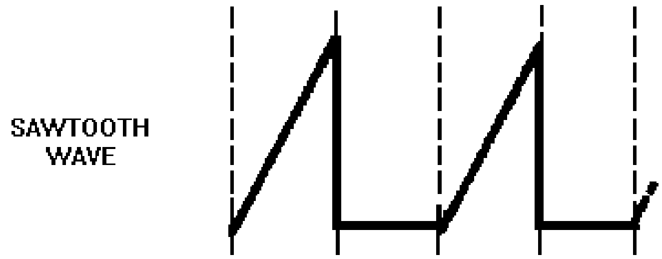
A **SQUARE WAVE** is identified by, two alternations equal in time that are square in appearance. One alternation is called a PULSE. The time for one complete cycle is called the PULSE REPETITION TIME (prt). The number of times in one second that the cycle repeats itself is called PULSE REPETITION RATE (prr) or PULSE REPETITION FREQUENCY (prf). The length of the pulse measured in the figure (T0 to T1) is referred to as the PULSE WIDTH (pw). The left side of the pulse is referred to as the LEADING EDGE and the right side as the TRAILING EDGE.



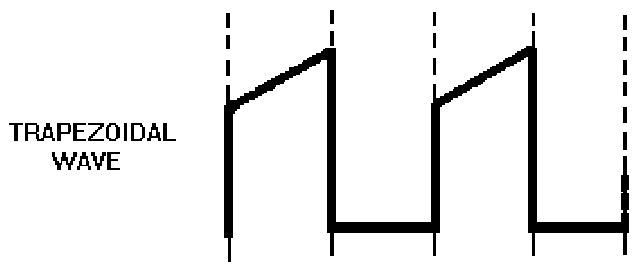
A **RECTANGULAR WAVE** has two alternations that are unequal in time.



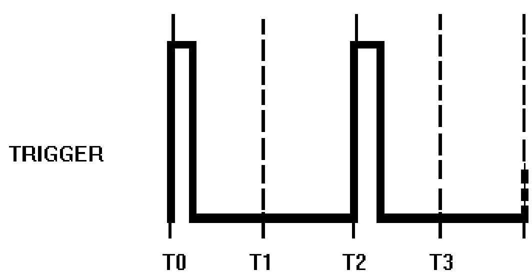
A **SAWTOOTH WAVE** has a linear increase in voltage followed by a rapid decrease of voltage at the end of the waveform.



A **TRAPEZOIDAL WAVE** looks like a sawtooth wave sitting on top of a square wave. The leading edge is called the JUMP voltage.

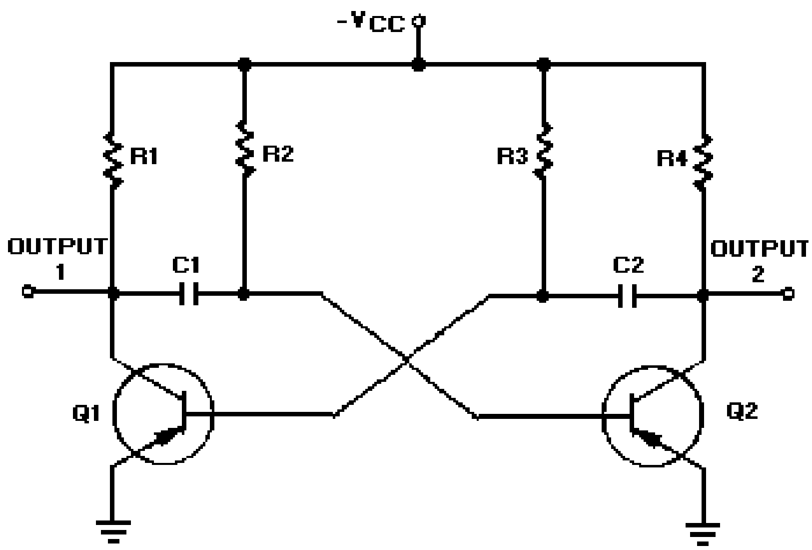


A **TRIGGER** is a very narrow pulse used to turn on or off another circuit.

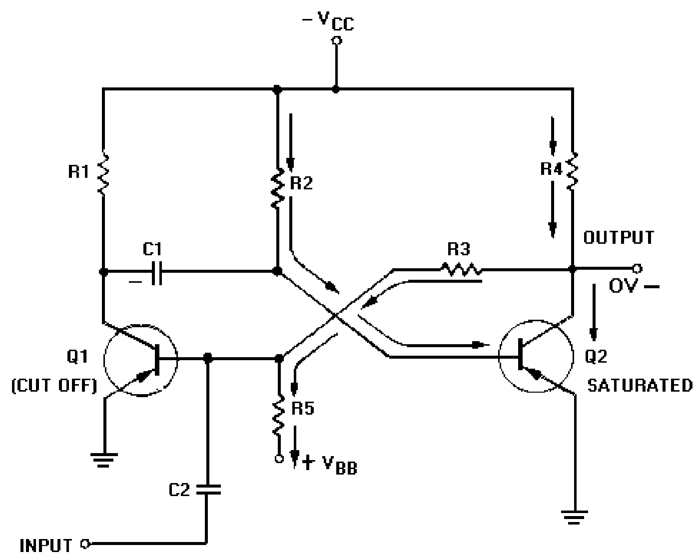


A **MULTIVIBRATOR** is used to generate a square or rectangular wave. A multivibrator is basically two amplifiers with regenerative feedback.

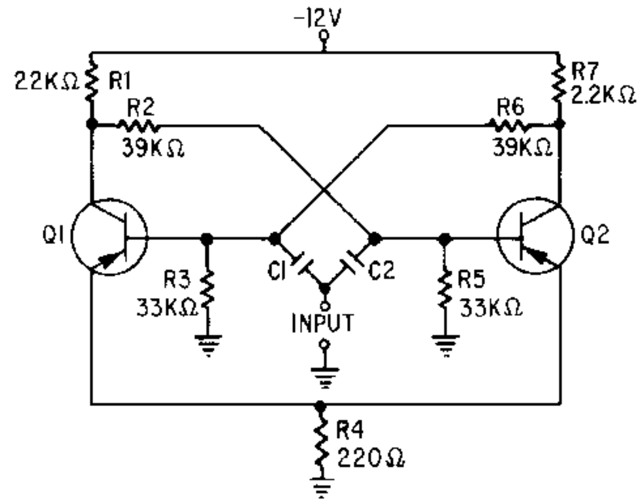
The **ASTABLE MULTIVIBRATOR** has no stable state. The transistors alternately switch from cutoff to saturation at a frequency determined by the RC time constants of the coupling circuits.



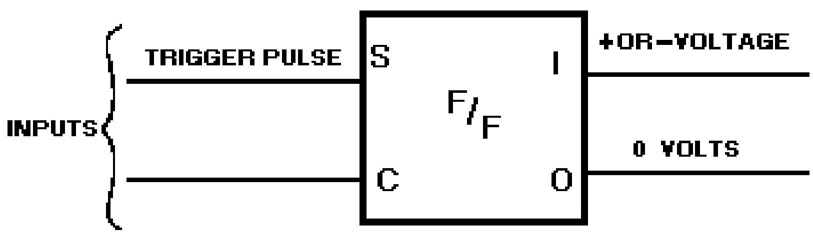
The **MONOSTABLE MULTIVIBRATOR** has one stable state. One transistor conducts while the other is cut off. An external trigger must be applied to change this condition.



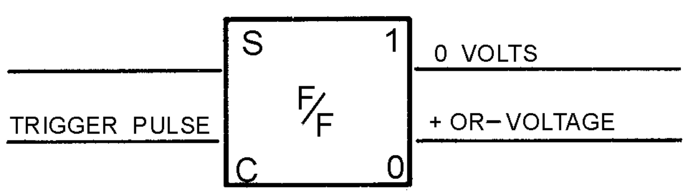
The **BISTABLE MULTIVIBRATOR** has two steady states. It remains in one of the stable states until a trigger is applied. It then switches to the other stable state until another trigger is applied.



The bistable multivibrator is also known as a FLIP-FLOP. The two inputs are SET and CLEAR. The two outputs are "1" and "0." A trigger pulse on the set input will cause the "1" output (negative or positive voltage depending on the type transistor used). At the same time the "0" output will equal 0 volts. This is the SET state.

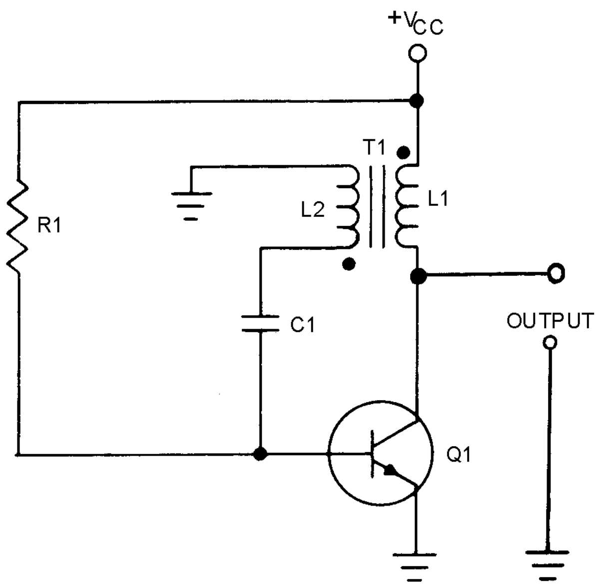


A **CLEAR STATE** of a flip-flop exists when the "1" output measures low voltage (or 0 volts) and the "0" output is high voltage. The flip-flop will flop to the CLEAR state only upon application of a trigger pulse to the CLEAR (C) input.

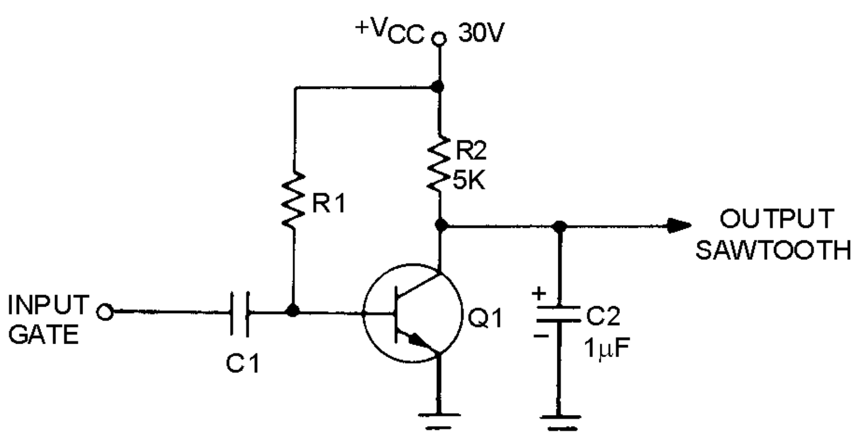


There is a third lead on some flip-flops. This lead is the TOGGLE (T) input. Every time a trigger pulse is applied to the (T) input, the flip-flop will change states.

BLOCKING OSCILLATORS are used in applications which require a narrow pulse with sharp leading and trailing edges. They are used as TRIGGER GENERATORS or FREQUENCY DIVIDERS.

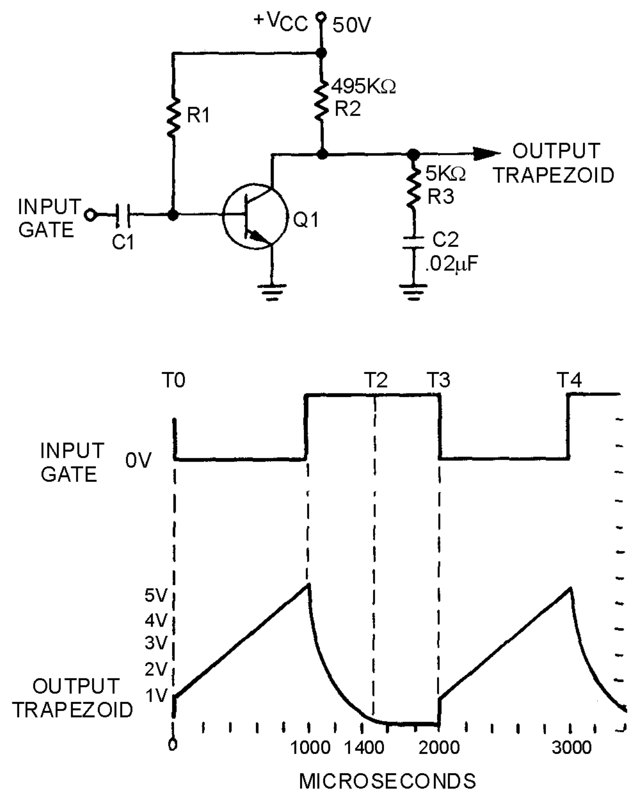


A **SAWTOOTH GENERATOR** voltage waveform has a linear change in voltage and a fast recovery time. The linear change in voltage is generated by taking the output from a capacitor. The sawtooth voltage waveform is used to provide electrostatic deflection in oscilloscopes.



(A)

A **TRAPEZOIDAL GENERATOR** voltage waveform is used to provide, a linear increase in current through a coil. A trapezoidal wave begins with a step or jump voltage, then a sawtooth wave. A trapezoidal wave of voltage is used in electromagnetic deflection display devices.



ANSWERS TO QUESTIONS Q1. THROUGH Q14.

A1. Multivibrator.

A2. Astable.

A3. Monostable.

A4. Bistable.

A5. RC coupling networks.

A6. One-shot.

A7. Two.

A8. Two.

A9. SET state.

A10. Transformer.

A11. Ten percent.

A12. Decreases linearity.

A13. To allow the capacitor time to discharge.

A14. A resistor.

CHAPTER 4

WAVE SHAPING

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Explain the operation of series-limiter circuits.
2. Explain the operation of parallel-limiter circuits.
3. Describe the operation of a dual-diode limiter circuit.
4. Explain the operation of clamper circuits.
5. Explain the composition of nonsinusoidal waves.
6. Explain how RC and RL circuits are used as integrators.
7. Explain how RC and RL circuits are used as differentiators.
8. Explain the operation of a counting circuit.
9. Explain the operation of a step-by-step counter used as a frequency divider.

LIMITERS

As a technician, you will be confronted with many different types of LIMITING circuits. A LIMITER is defined as a device which limits some part of a waveform from exceeding a specified value. Limiting circuits are used primarily for wave shaping and circuit-protection applications.

A limiter is little more than the half-wave rectifier you studied in *NEETS, Module 6, Introduction to Electronic Emission, Tubes, and Power Supplies*. By using a diode, a resistor, and sometimes a dc bias voltage, you can build a limiter that will eliminate the positive or negative alternations of an input waveform. Such a circuit can also limit a portion of the alternations to a specific voltage level. In this chapter you will be introduced to five types of limiters: SERIES-POSITIVE, SERIES-NEGATIVE, PARALLEL-POSITIVE, PARALLEL-NEGATIVE, and DUAL-DIODE LIMITERS. Both series- and parallel-positive and negative limiters use biasing to obtain certain wave shapes. They will be discussed in this chapter.

The diode in these circuits is the voltage-limiting component. Its polarity and location, with respect to ground, are the factors that determine circuit action. In series limiters, the diode is in series with the output. In parallel limiters, the diode is in parallel with the output.

SERIES LIMITERS

You should remember, from NEETS, Module 7, *Introduction to Solid-State Devices and Power Supplies*, that a diode will conduct when the anode voltage is positive with respect to the cathode voltage. The diode will not conduct when the anode is negative in respect to the cathode. Keeping these two simple facts in mind as you study limiters will help you understand their operation. Your knowledge of voltage divider action from NEETS, Module 1, *Introduction to Matter, Energy, and Direct Current* will also help you understand limiters.

In a SERIES LIMITER, a diode is connected in series with the output, as shown in view (A) of figure 4-1. The input signal is applied across the diode and resistor and the output is taken across the resistor. The series-limiter circuit can limit either the positive or negative alternation, depending on the polarity of the diode connection with respect to ground. The circuit shown in figure 4-1, view (B), is a SERIES-POSITIVE LIMITER. Reversing D1 would change the circuit to a SERIES-NEGATIVE LIMITER.

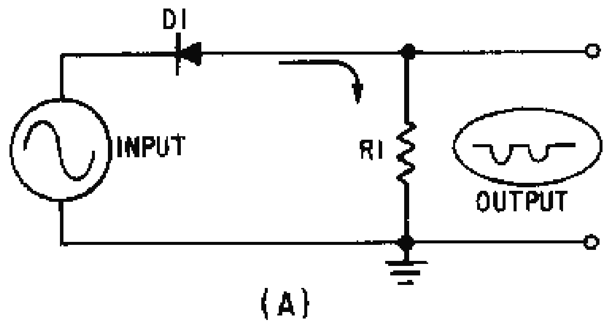


Figure 4-1A.—Series-positive limiter.

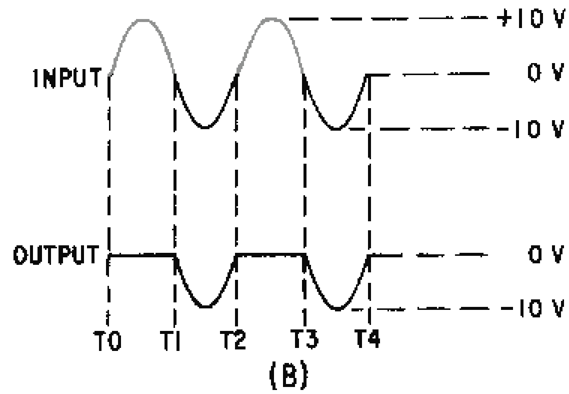


Figure 4-1B.—Series-positive limiter.

Series-Positive Limiter

Let's look at the series-positive limiter and its outputs in figure 4-1. Diode D1 is in series with the output and the output is taken across resistor R1. The input must be negative with respect to the anode of the diode to make the diode conduct. When the positive alternation of the input signal (T0 to T1) is applied to the circuit, the cathode is positive with respect to the anode. The diode is reverse biased and will not conduct. Since no current can flow, no output is developed across the resistor during the positive alternation of the input signal.

During the negative half cycle of the input signal (T1 to T2), the cathode is negative with respect to the anode. This causes D1 to be forward biased. Current flows through R1 and an output is developed.

The output during each negative alternation of the input is approximately the same as the input (−10 volts) because most of the voltage is developed across the resistor.

Ideally, the output wave shape should be exactly the same as the input wave shape with only the limited portion removed. When the diode is reverse biased, the circuit has a small amount of reverse current flow, as shown just above the 0-volt reference line in figure 4-2. During the limiting portion of the input signal, the diode resistance should be high compared to the resistor. During the time the diode is conducting, the resistance of the diode should be small as compared to that of the resistor. In other words, the diode should have a very high front-to-back ratio (forward resistance compared to reverse resistance). This relationship can be better understood if you study the effects that a front-to-back resistance ratio has on circuit output.

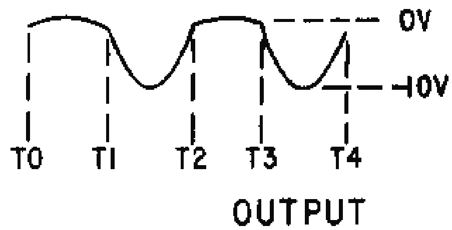


Figure 4-2.—Actual output of a series-positive limiter.

The following formula can be used to determine the output amplitude of the signal:

$$E_{out} = \frac{R}{R + R_{ac}} \cdot E_{in}$$

Where:

E_{out} = amplitude voltage

R = value of R_1

R_{ac} = value of ac resistance of the diode (under forward - and reverse - biased conditions)

E_{in} = input of signal amplitude

Let's use the formula to compare the front-to-back ratio of the diode in the forward- and reverse-biased conditions.

Given:

$R_1 = 1{,}000$ ohms

$R_{ac} = 1$ ohm (forward - biased condition)

$R_{ac} = 100{,}000$ ohms (reversed biased condition)

$E_{in} = 10$ volts

FORWARD BIAS

$$E_{out} = \frac{R}{R + R_{ac}} \cdot E_{in}$$

$$E_{out} = \frac{1{,}000}{1{,}000 + 1} \cdot 10 volts$$

$$E_{out} = \frac{1{,}000}{1{,}001} \cdot 10 volts$$

$$E_{out} = 0.999 \cdot 10 volts$$

$$E_{out} = 0.09 volt$$

REVERSE BIAS

$$E_{out} = \frac{R}{R + R_{ac}} \cdot E_{in}$$

$$E_{out} = \frac{1{,}000}{1{,}000 + 100{,}000} \cdot 10 \text{ volts}$$

$$E_{out} = \frac{1{,}000}{101{,}000} \cdot 10 \text{ volts}$$

$$E_{out} = 0.09 \cdot 10 \text{ volts}$$

$$E_{out} = 0.09 \text{ volt}$$

You can readily see that the formula comparison of the forward- and reverse-bias resistance conditions shows that a small amount of reverse current will flow during the limited portion of the input waveform. This small amount of reverse current will develop as the small positive voltage (0.09 volt) shown in figure 4-2 (T0 to T1 and T2 to T3). The actual amount of voltage developed will depend on the type of diode used. For the remainder of this chapter, we will use only idealized waveforms and disregard this small voltage.

SERIES-POSITIVE LIMITER WITH BIAS.—In the series-positive limiter (figure 4-1, view (A)), the reference point at the bottom of resistor R1 is ground, or 0 volts. By placing a dc potential at point (1) in figure 4-3 (views (A) and (B)), you can change the reference point. The reference point changes by the amount of dc potential that is supplied by the battery. The battery can either aid or oppose the flow of current in the series-limiter circuit. POSITIVE BIAS (aiding) is shown in view (A) and NEGATIVE BIAS (opposing) is shown in view (B).

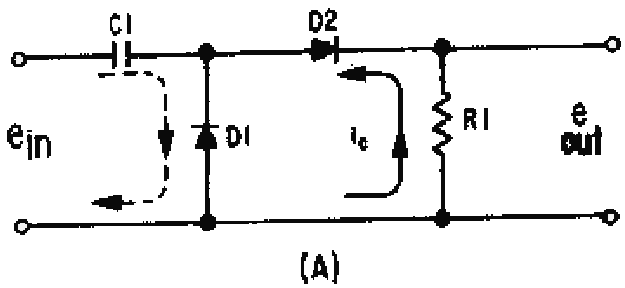


Figure 4-3A.—Positive and negative bias. POSITIVE BIAS.

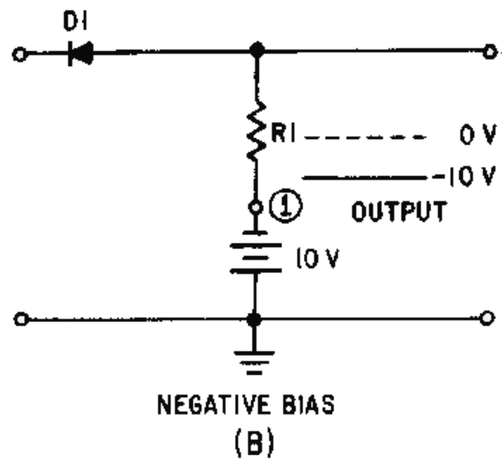


Figure 4-3B.—Positive and negative bias. NEGATIVE BIAS.

When the dc aids forward bias, as in view (A), the diode conducts even with no signal applied. An input signal sufficiently positive to overcome the dc bias potential is required to reverse bias and cut off the diode.

Let's look at a series-positive limiter with positive bias as shown in figure 4-4, views (A) and (B). The diode will conduct until the input signal exceeds +5 at T1 on the positive alternation of the input signal. When the positive alternation exceeds +5 volts, the diode becomes reverse biased and limits the positive alternation of the output signal to +5 volts. This is because there is no current flow through resistor R1 and battery voltage is felt at point (B). The diode will remain reverse biased until the positive alternation of the input signal decreases to just under +5 volts at T2. At this time, the diode again becomes forward biased and conducts. The diode will remain forward biased from T2 to T3. During this period the negative alternation of the input is passed through the diode without being limited. From T3 to T4 the diode is again reverse biased and the output is again limited.

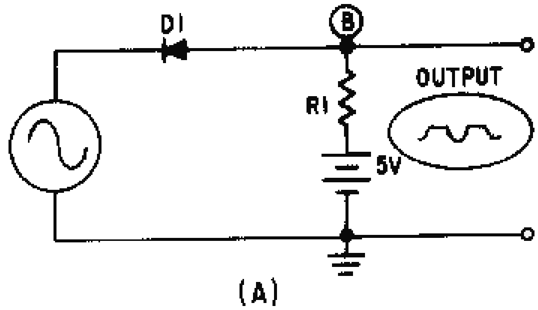


Figure 4-4A.—Series-positive limiter with positive bias.

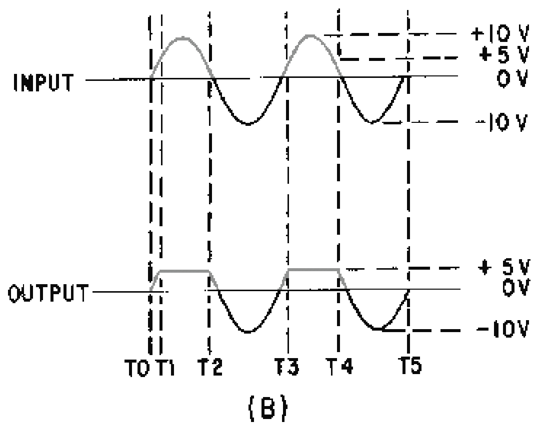


Figure 4-4B.—Series-positive limiter with positive bias.

Now let's look at what takes place when reverse bias is aided, as shown in figure 4-5, view (A). The diode is negatively biased with -5 volts from the battery. In view (B), compare the output to the input signal applied. From T0 to T1 the diode is reverse biased and limiting takes place. The output is at -5 volts (battery voltage) during this period. As the negative alternation increases toward -10 volts (T1), the cathode of the diode becomes more negative than the anode and is forward biased. From T1 to T2 the input signal is passed to the output. The diode remains forward biased until the negative alternation has decreased to −5 volts at T2. At T2 the cathode of the diode becomes more positive than the anode, and the diode is again reverse biased and remains so until T3.

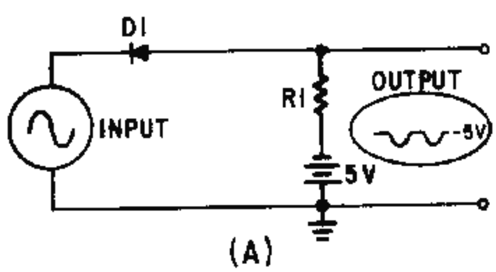


Figure 4-5A.—Series-positive limiter with negative bias.

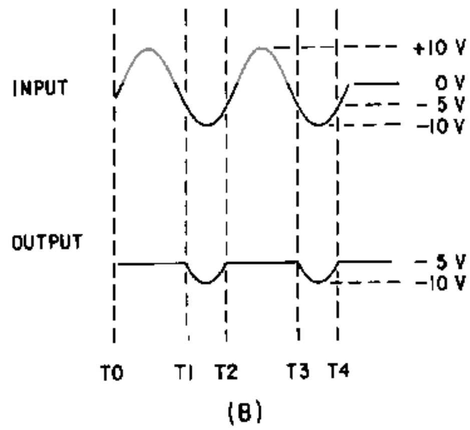


Figure 4-5B.—Series-positive limiter with negative bias.

Series-Negative Limiter

In view (A) of figure 4-6, the SERIES-NEGATIVE LIMITER limits the negative portion of the waveform, as shown in view (B). Let's consider the input signal and determine how the output is produced. During T0 to T1 (view (B)), the anode is more positive than the cathode and the diode conducts. Current flows up through the resistor and the diode, and a positive voltage is developed at the output. The voltage across the resistor is essentially the same as the voltage applied to the circuit.

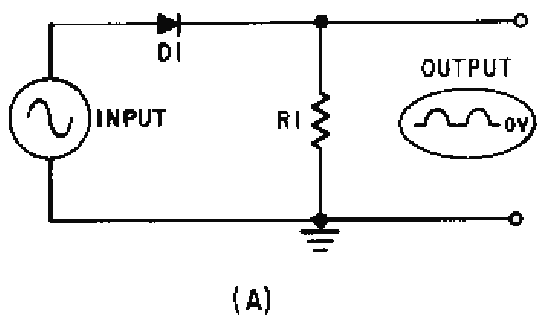


Figure 4-6A.—Series-negative limiter.

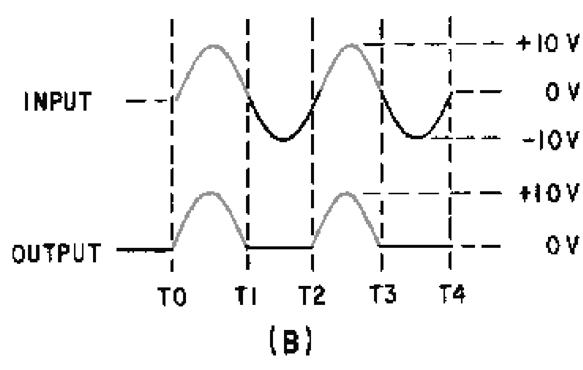


Figure 4-6B.—Series-negative limiter.

During T1 to T2 the anode is negative with respect to the cathode and the diode does not conduct. This portion of the output is limited because no current flows through the resistor.

As you can see, the only difference between series-positive and series-negative limiters is that the diode is reversed in the negative limiters.

SERIES-NEGATIVE LIMITER WITH BIAS.—View (A) of figure 4-7 shows a series-negative limiter with negative bias. The diode is forward biased and conducts with no input signal. In view (B) it will continue to conduct as the input signal swings first positive and then negative (but only to −5 volts) from T0 through T1. At T1 the input becomes negative with respect to the −5 volt battery bias. The diode becomes reverse biased and is cutoff until T2 when the anode again becomes positive with respect to the battery voltage (−5 volts) on the cathode. No voltage is developed in the output by R1 (no current flow) and the output is held at −5 volts from T1 to T2. With negative bias applied to a series-negative limiter, only a portion of the negative signal is limited.

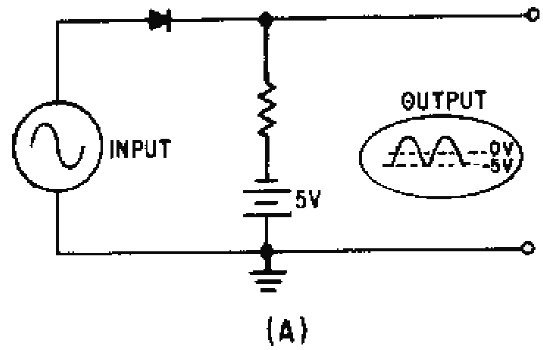


Figure 4-7A.—Series-negative limiter with negative bias.

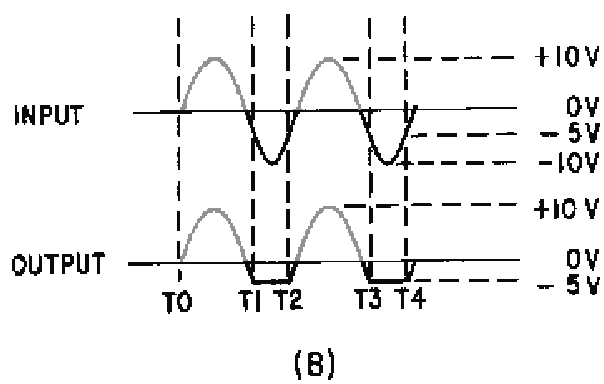


Figure 4-7B.—Series-negative limiter with negative bias.

Now let's look at a series-negative limiter with positive bias, as shown in figure 4-8, view (A). Here we will remove all of the negative alternation and part of the positive alternation of the input signal. We have given a full explanation of the series-positive limiter, series-positive limiter with bias, series-negative limiter, and series-negative limiter with negative bias; therefore, you should have little difficulty understanding what is happening in the circuit in the figure.

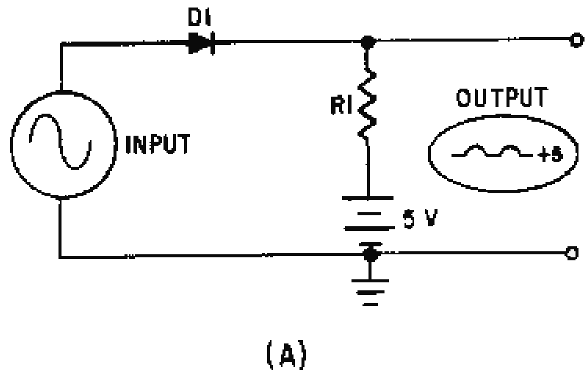


Figure 4-8A.—Series-negative limiter with positive bins.

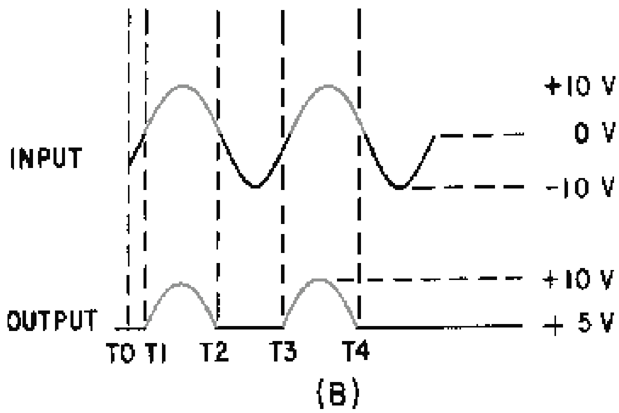


Figure 4-8B.—Series-negative limiter with positive bins.

The series-negative limiter with positive bias is different in only one aspect from the series-positive limiter with bias (figure 4-5) discussed earlier. The difference is that the diode is reversed and the output is of the opposite polarity.

Q1. Which portion of a sine-wave input is retained in the output of a series-positive limiter?

Q2. Which portion of a sine-wave input is retained in the output of a series-negative limiter?

Q3. How can a series-positive limiter be modified to limit unwanted negative portions of the input signal?

PARALLEL LIMITERS

A PARALLEL-LIMITER circuit uses the same diode theory and voltage divider action as series limiters. A resistor and diode are connected in series with the input signal and the output signal is developed across the diode. The output is in parallel with the diode, hence the circuit name, parallel limiter. The parallel limiter can limit either the positive or negative alternation of the input signal.

Recall that in the series limiter the output was developed while the diode was conducting. In the parallel limiter the output will develop when the diode is cut off. You should not try to memorize the outputs of these circuits; rather, you should study their actions and be able to figure them out.

Parallel-Positive Limiter.

The schematic diagram shown in figure 4-9, view (A), is a PARALLEL-POSITIVE LIMITER. The diode is in parallel with the output and only the positive half cycle of the input is limited. When the positive alternation of the input signal is applied to the circuit (T0 to T1), the diode is forward biased and conducts. This action may be seen in view (B). As current flows up through the diode and the resistor, a voltage is dropped across each. Since R1 is much larger than the forward resistance of D1, most of the input signal is developed across R1. This leaves only a very small voltage across the diode (output). The positive alternation of the input signal has been limited.

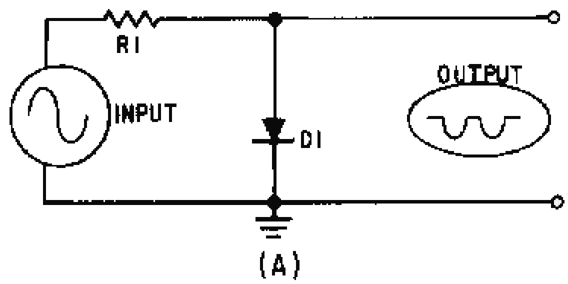


Figure 4-9A.—Parallel-positive limiter.

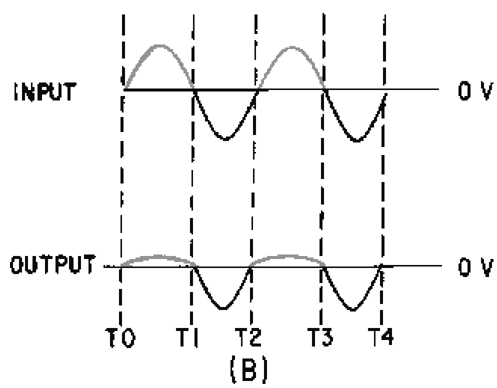


Figure 4-9B.—Parallel-positive limiter.

From T1 to T2 the diode is reverse biased and acts as an extremely high resistance. The negative alternation of the input signal appears across the diode at approximately the same amplitude as the input. The negative alternation of the input is not limited.

As with the series limiter, the parallel limiter should provide maximum output voltage for the unlimited part of the signal. The reverse-bias resistance of the diode must be very large compared to the series resistor. To determine the output amplitude, use the following formula:

$$E_{out} = \frac{R}{R + R_{ac}} \cdot E_{in}$$

Where:

E_{out} = amplitude voltage

R = value of R_1

R_{ac} = value of ac resistance of the diode (under forward - and reverse - biased conditions)

E_{in} = input of signal amplitude

PARALLEL-POSITIVE LIMITER WITH BIAS.—Figure 4-10, view (A), shows the schematic diagram of a PARALLEL-POSITIVE LIMITER WITH NEGATIVE BIAS. The diode is forward biased and conducts without an input signal. D1 is essentially a short circuit. The voltage at the output terminals is -4 volts.

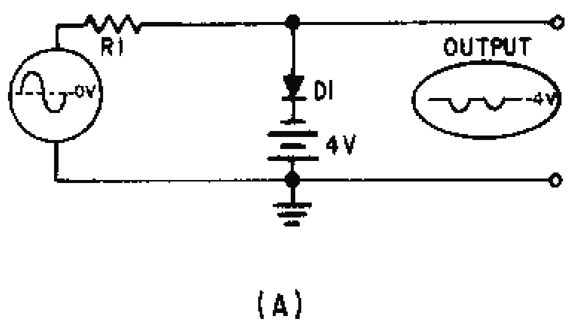


Figure 4-10A.—Parallel limiter with negative bias.

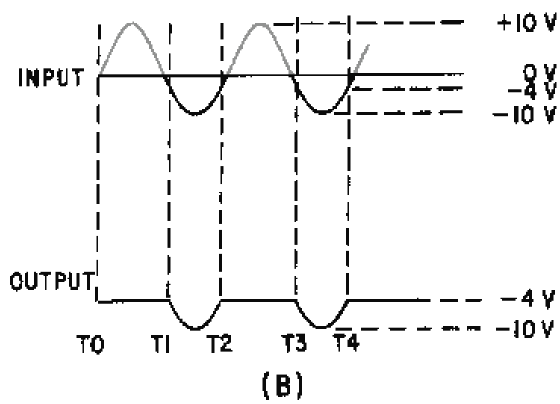


Figure 4-10B.—Parallel limiter with negative bias.

As the positive alternation of the input signal is applied to the circuit, the diode remains forward biased and limits the entire positive alternation, as shown in view (B). As the signal goes in a negative direction Oust before T1), the diode remains forward biased (limiting is still present) until the input signal exceeds –4 volts (T1). D1 becomes reverse biased as the anode becomes more negative than the cathode. While the input signal is more negative than the –4 volts of the bias battery (T1 to T2), the diode is reverse biased and remains cut off. The output follows the input signal from T1 to T2. At all other times during that cycle, the diode is forward biased and limiting occurs. This circuit is called a parallel-positive limiter with negative bias because the positive output is limited and the bias in the circuit is negative with reference to ground. Limiting takes place at all points more positive than –4 volts.

The circuit shown in figure 4-11, view (A), is a PARALLEL-POSITIVE LIMITER WITH POSITIVE BIAS. The positive terminal of the battery is connected to the cathode of the diode. This causes the diode to be reverse biased at all times except when the input signal is more positive than the bias voltage (T1 to T2), as shown in view (B).

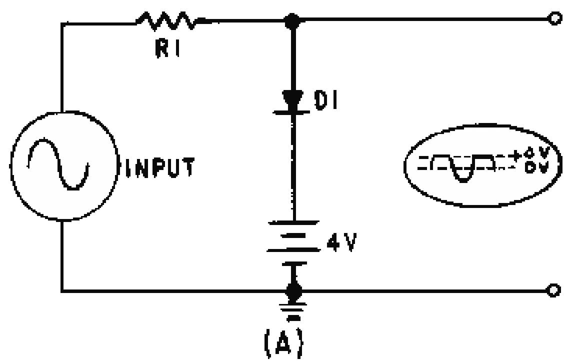


Figure 4-11A.—Parallel-positive Limiter with positive bias.

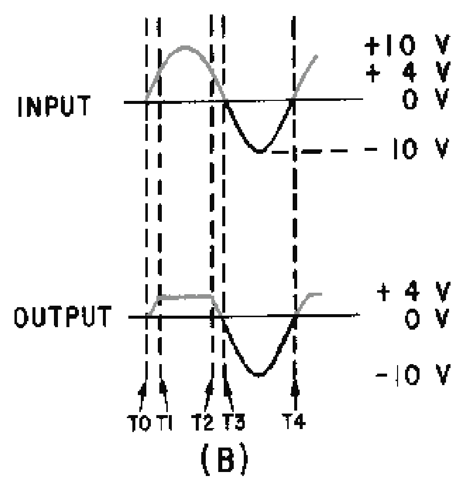


Figure 4-11B.—Parallel-positive Limiter with positive bias.

As the positive alternation of the input signal is applied (T0), the output voltage follows the input signal. From T1 to T2 the input signal is more positive than + 4 volts. The diode is forward biased and conducts. At this time the output voltage equals the bias voltage and limiting takes place. From T2 to T4 of the input signal, the diode is reverse biased and does not conduct. The output signal follows the input signal and no limiting takes place.

This circuit is called a parallel-positive limiter with positive bias because limiting takes place in the positive alternation and positive bias is used on the diode.

Parallel-Negative Limiter

A PARALLEL-NEGATIVE LIMITER is shown in view (A) of figure 4-12. Notice the similarity of the parallel-negative limiter and the parallel-positive limiter shown in view (A) of figure 4-9. From T0 to T1 of the input signal, the diode is reverse biased and does not conduct, as shown in view (B) of figure 4-12. The output signal follows the input signal and the positive alternation is not limited.

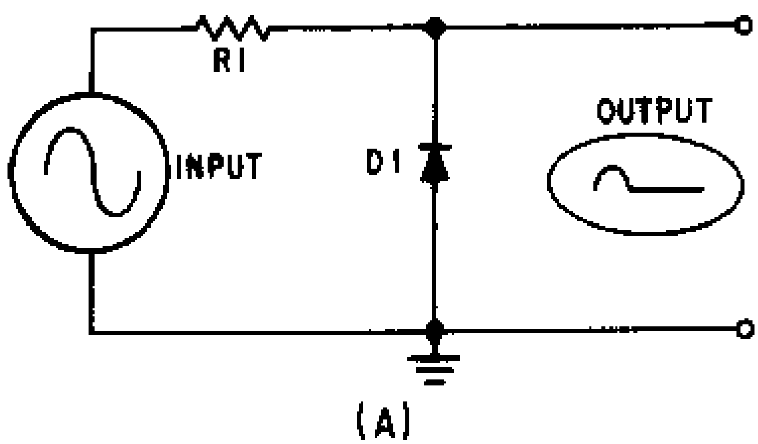


Figure 4-12A.—Parallel-negative limiter.

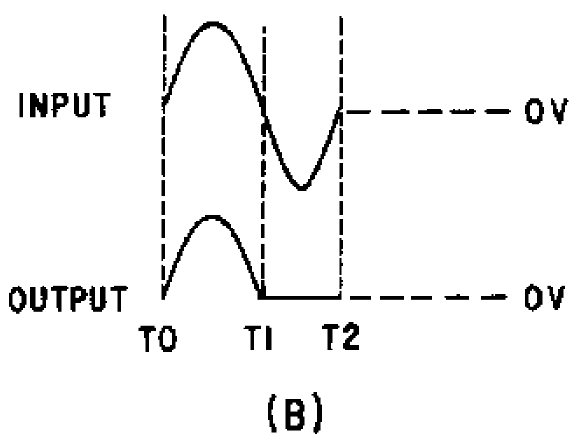


Figure 4-12B.—Parallel-negative limiter.

During the negative alternation of the input signal (T1 to T2), the diode is forward biased and conducts. The relatively low forward bias of D1 develops a very small voltage and, therefore, limits the output to nearly 0 volts. A voltage is developed across the resistor as current flows through the resistor and diode.

PARALLEL-NEGATIVE LIMITER WITH BIAS.—The circuit shown in figure 4-13, view (A), is a parallel-negative limiter with negative bias. With no input, the battery maintains D1 in a reverse-bias condition. D1 cannot conduct until its cathode is more negative than its anode. D1 acts as an open until the input signal dips below −4 volts at T2 in view (B). At T2 the input signal becomes negative enough to forward bias the diode, D1 conducts and acts like a short, and the output is limited to the −4 volts from the battery from T2 to T3. Between T3 and T4 the diode is again reverse biased. The output signal follows the input signal and no limiting occurs.

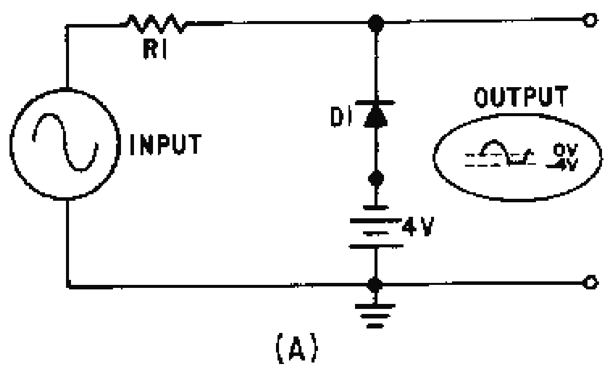


Figure 4-13A.—Parallel-negative limiter with negative bias.

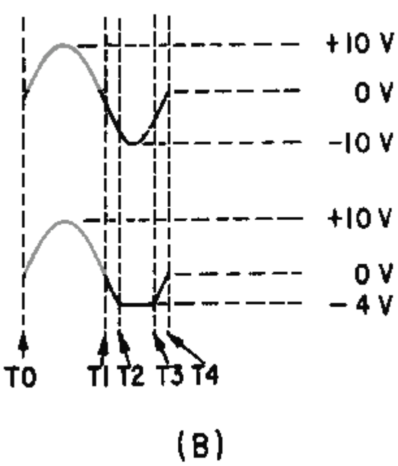


Figure 4-13B.—Parallel-negative limiter with negative bias.

Figure 4-14, view (A), shows a parallel-negative limiter with positive bias. The operation is similar to those circuits already explained. Limiting occurs when the diode conducts. No limiting occurs when the diode is reverse biased. In this circuit, the bias battery provides forward bias to the diode without an input signal. The output is at +4 volts, except where the input goes above +4 volts (T1 to T2), as shown in view (B). The parts of the signal more negative than +4 volts are limited.

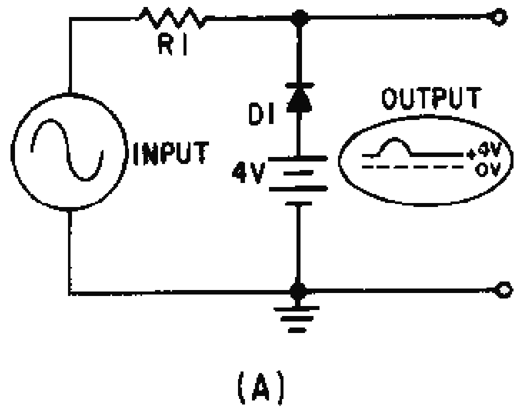


Figure 4-14A.—Parallel-negative limiter with positive bias.

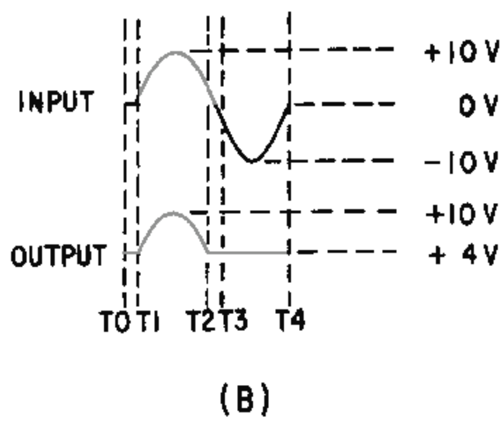


Figure 4-14B.—Parallel-negative limiter with positive bias.

Q4. What component is in parallel with the output in a parallel limiter?

Q5. What is the condition of the diode in a series limiter when an output is developed? In a parallel limiter?

DUAL-DIODE LIMITER

The last type of limiter to be discussed in this chapter is the DUAL-DIODE LIMITER, shown in figure 4-15, view (A). This limiter combines a parallel-negative limiter with negative bias (D1 and B1) and a parallel-positive limiter with positive bias (D2 and B2). Parts of both the positive and negative alternations are removed in this circuit. Each battery aids the reverse bias of the diode in its circuit; the circuit has no current flow with no input signal. When the input signal is below the value of the biasing batteries, both D1 and D2 are reverse biased. With D1 and D2 reverse biased, the output follows the input. When the input signal becomes more positive than +20 volts (view (B)), D2 conducts and limits the output to +20 volts. When the input signal becomes more negative than –20 volts, D1 conducts and limits the output to this, value. When neither diode conducts, the output follows the input waveform.

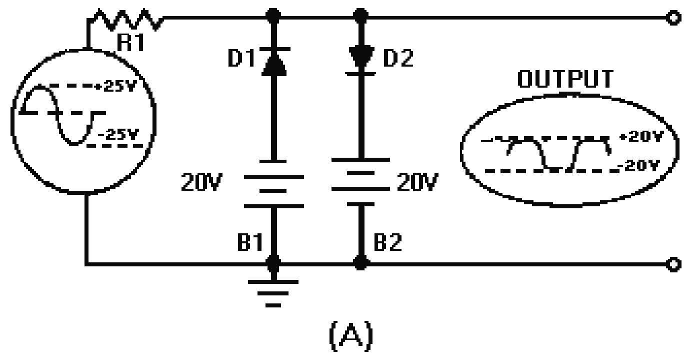


Figure 4-15A.—Dual-diode limiter.

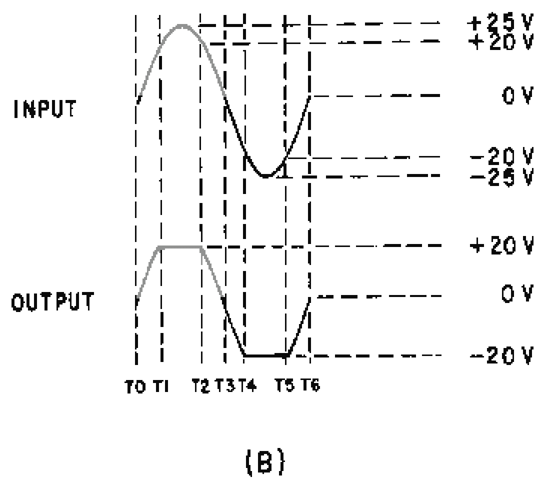


Figure 4-15B.—Dual-diode limiter.

CLAMPERS

Certain applications in electronics require that the upper or lower extremity of a wave be fixed at a specific value. In such applications, a CLAMPING (or CLAMPER) circuit is used. A clamping circuit clamps or restrains either the upper or lower extremity of a waveform to a fixed dc potential. This circuit is also known as a DIRECT-CURRENT RESTORER or a BASE-LINE STABILIZER. Such circuits are used in test equipment, radar systems, electronic countermeasure systems, and sonar systems. Depending upon the equipment, you could find negative or positive clampers with or without bias. Figure 4-16, views (A) through (E), illustrates some examples of waveforms created by clampers. However, before we discuss clampers, we will review some relevant points about series RC circuits.

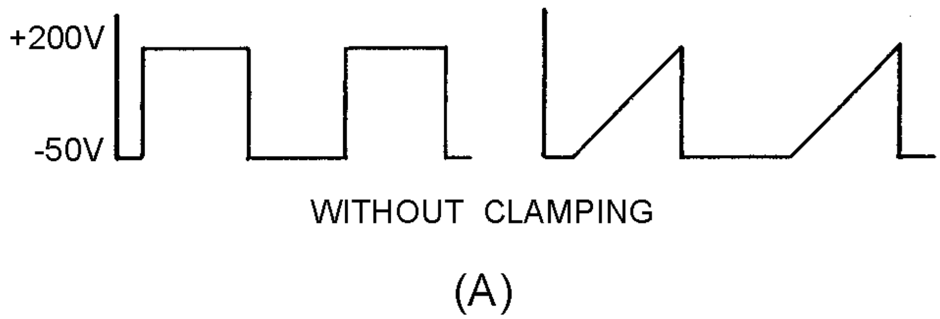


Figure 4-16A.—Clamping waveforms. WITHOUT CLAMPING.

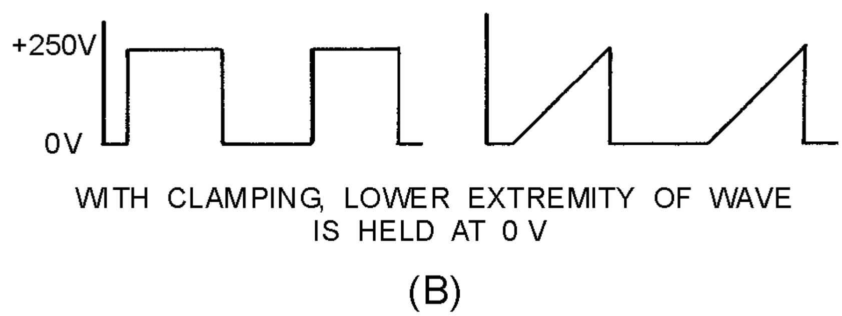


Figure 4-16B.—Clamping waveforms. WITH CLAMPING, LOWER EXTREMITY OF WAVE IS HELD AT 0V.

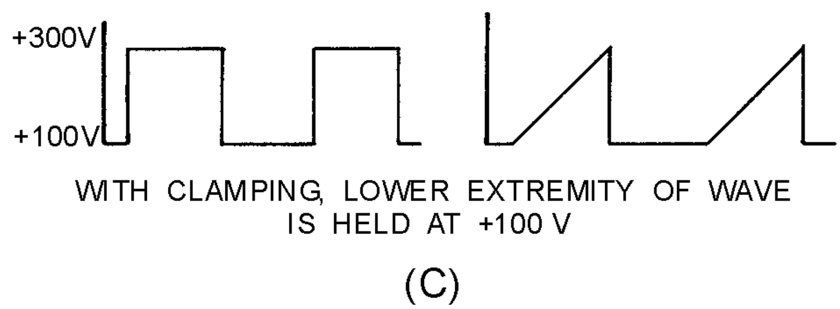


Figure 4-16C.—Clamping waveforms. WITH CLAMPING, LOWER EXTREMITY OF WAVE IS HELD AT +100 V.

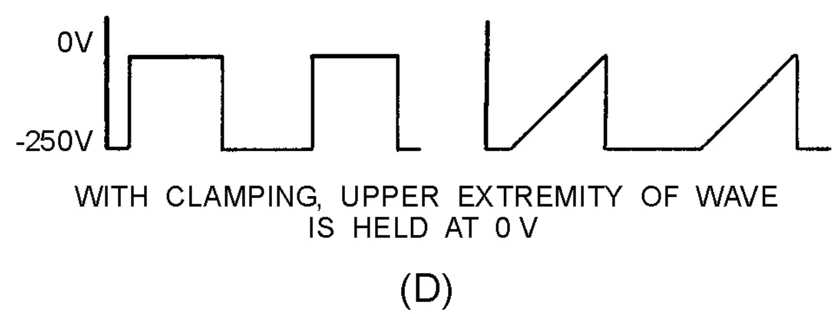


Figure 4-16D.—Clamping waveforms. WITH CLAMPING, UPPER EXTREMITY OF WAVE IS HELD AT 0V.

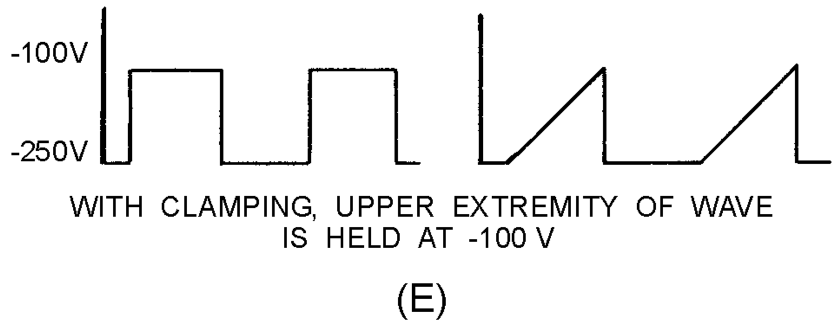


Figure 4-16E.—Clamping waveforms. WITH CLAMPING, UPPER EXTREMITY OF WAVE IS HELD AT –100 V.

SERIES RC CIRCUITS

Series RC circuits are widely used for coupling signals from one stage to another. If the time constant of the coupling circuit is comparatively long, the shape of the output waveform will be almost identical to that of the input. However, the output dc reference level may be different from that of the input. Figure 4-17, view (A), shows a typical RC coupling circuit in which the output reference level has been changed to 0 volts. In this circuit, the values of R1 and C1 are chosen so that the capacitor will charge (during T0 to T1) to 20 percent of the applied voltage, as shown in view (B). With this in mind, let's consider the operation of the circuit.

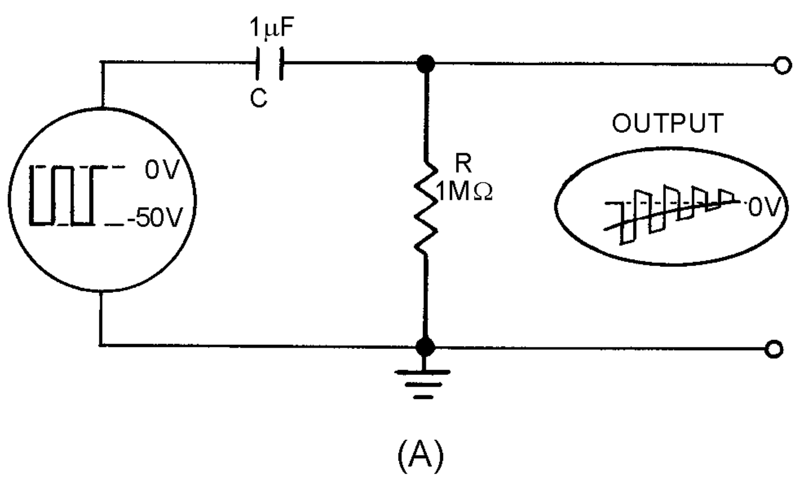


Figure 4-17A.—RC coupling.

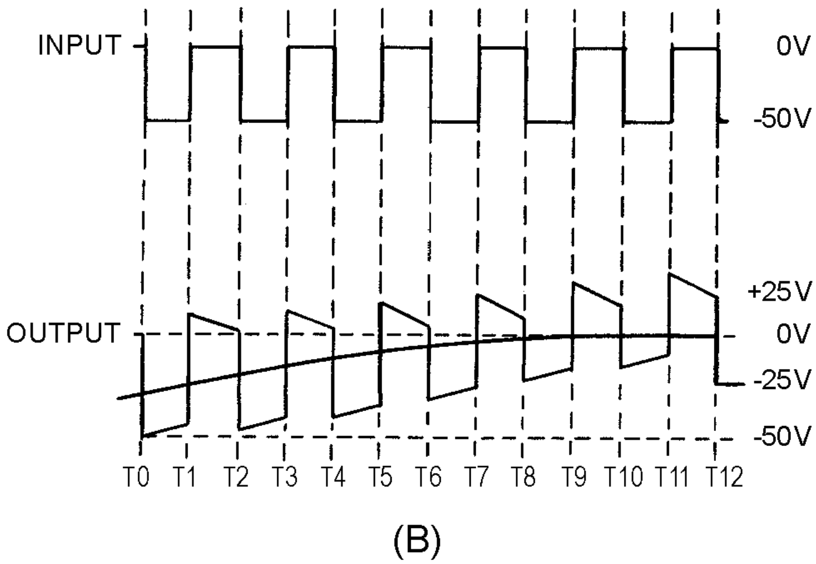


Figure 4-17B.—RC coupling.

At T0 the input voltage is −50 volts and the capacitor begins charging. At the first instant the voltage across C is 0 and the voltage across R is −50 volts. As C charges, its voltage increases. The voltage across R, which is the output voltage, begins to drop as the voltage across C increases. At T1 the capacitor has charged to 20 percent of the −50 volts input, or −10 volts. Because the input voltage is now 0 volts, the capacitor must discharge. It discharges through the low impedance of the signal source and through R, developing +10 volts across R at the first instant. C discharges 20 percent of the original 10-volt charge from T1 to T2. Thus, C discharges to +8 volts and the output voltage also drops to 8 volts.

At T2 the input signal becomes −50 volts again. This −50 volts is in series opposition to the 8-volt charge on the capacitor. Thus, the voltage across R totals −42 volts (−50 plus +8 volts). Notice that this value of voltage (−42 volts) is smaller in amplitude than the amplitude of the output voltage which occurred at TO (−50 volts). Capacitor C now charges from +8 to +16 volts. If we were to continue to follow the operation of the circuit, we would find that the output wave shape would become exactly distributed around the 0-volt reference point. At that time the circuit operation would have reached a stable operating point. Note that the output wave shape has the same amplitude and approximately the same shape as the input wave shape, but now "rides" equally above and below 0 volts. Clampers use this RC time so that the input and output waveforms will be almost identical, as shown from T11 to T12.

POSITIVE-DIODE CLAMPERS

Figure 4-18, view (A), illustrates the circuit of a positive-diode clamper. Resistor R1 provides a discharge path for C1. This resistance is large in value so that the discharge time of C1 will be long compared to the input pulse width. The diode provides a fast charge path for C1. After C1 becomes charged it acts as a voltage source. The input wave shape shown in view (B) is a square wave and varies between +25 volts and −25 volts. Compare each portion of the input wave shape with the corresponding output wave shape. Keep Kirchhoff's law in mind: The algebraic sum of the voltage drops around a closed loop is 0 at any instant.

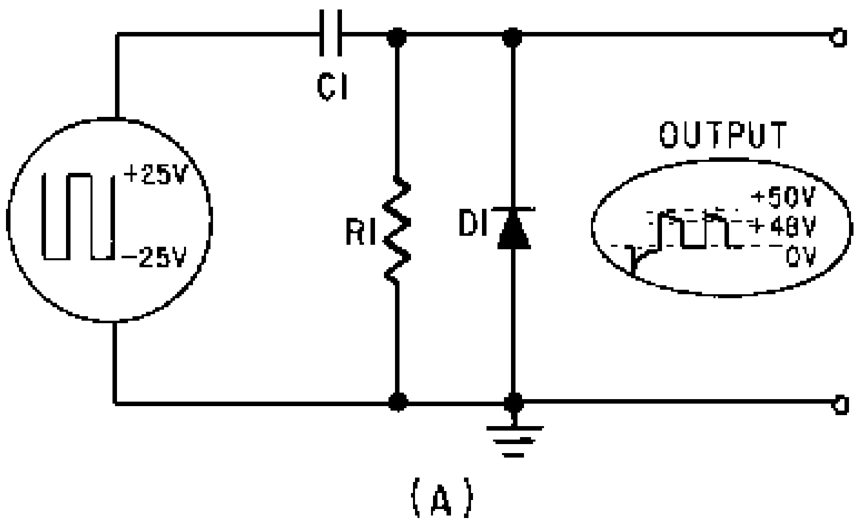


Figure 4-18A.—Positive damper and waveform.

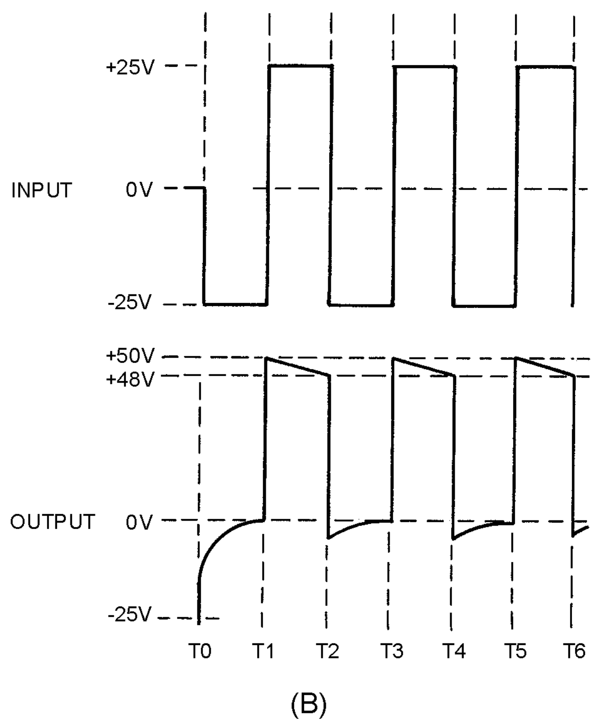


Figure 4-18B.—Positive damper and waveform.

At T0 the −25 volt input signal appears across R1 and D1 (the capacitor is a short at the first instant). The initial voltage across R1 and D1 causes a voltage spike in the output. Because the charge time of C1 through D1 is almost instantaneous, the duration of the pulse is so short that it has only a negligible effect on the output. The −25 volts across D1 makes the cathode negative with respect to the anode and the diode conducts heavily. C1 quickly charges through the small resistance of D1. As the voltage across C1 increases, the output voltage decreases at the same rate. The voltage across C1 reaches -25 volts and the output is at 0 volts.

At T1 the +25 volts already across the capacitor and the +25 volts from the input signal are in series and aid each other (SERIES AIDING). Thus, +50 volts appears across R1 and D1. At this time, the cathode of D1 is positive with respect to the anode, and the diode does not conduct. From T1 to T2, C1 discharges to approximately +23 volts (because of the large values of R and C) and the output voltage drops from +50 volts to +48 volts.

At T2 the input signal changes from +25 volts to −25 volts. The input is now SERIES OPPOSING with the +23 volts across C1. This leaves an output voltage of −2 volts (−25 plus +23 volts). The cathode of D1 is negative with respect to the anode and D1 conducts. From T2 to T3, C1 quickly charges through D1 from +23 volts to +25 volts; the output voltage changes from −2 volts to 0 volts.

At T3 the input signal and capacitor voltage are again series aiding. Thus, the output voltage felt across R1 and D1 is again +50 volts. During T3 and T4, C1 discharges 2 volts through R1. Notice that circuit operation from T3 to T4 is the same as it was from T1 to T2. The circuit operation for each square-wave cycle repeats the operation which occurred from T2 to T4.

Compare the input wave shape of figure 4-18, view (B), with the output wave shape. Note the following important points: (1) The peak-to-peak amplitude of the input wave shape has not been changed by the clamper circuit; (2) the shape of the output wave shape has not been significantly changed from that of the input by the action of the clamper circuit; and (3) the output wave shape is now all above 0 volts whereas the input wave shape is both above and below 0 volts. Thus, the lower part of the input wave shape has been clamped to a dc potential of 0 volts in the output. This circuit is referred to as a positive clamper since all of the output wave shape is above 0 volts and the bottom is clamped at 0 volts.

The positive clamper circuit is self-adjusting. This means that the bottom of the output waveform remains clamped at 0 volts during changes in input signal amplitude. Also, the output wave shape retains the form and peak-to-peak amplitude (50 volts in this case) of the input wave shape. When the input amplitude becomes greater, the charge of the capacitor becomes greater and the output amplitude becomes larger. When the input amplitude decreases, the capacitor does not charge as high as before and clamping occurs at a lower output voltage. The capacitor charge, therefore, changes with signal strength.

The size of R1 and C1 has a direct effect upon the operation of the clamper. Because of the small resistance of the diode, the capacitor charge time is short. If either R1 or C1 is made smaller, the capacitor discharges faster ($TC = R \cdot C$).

The ability of a smaller value capacitor to quickly discharge to a lower voltage is an *advantage* when the amplitude of the input wave shape is suddenly reduced. However, for normal clamper operation, quick discharge time is a *disadvantage*. This is because one objective of clamping is to keep the output wave shape the same as the input wave shape. If the small capacitor allows a relatively large amount of the voltage to discharge with each cycle, then distortion occurs in the output wave shape. A larger portion of the wave shape then appears on the wrong side of the reference line.

Increasing the value of the resistor increases the discharge time (again, $TC = R \cdot C$). This increased value causes the capacitor to discharge more slowly and produces an output wave shape which is a better reproduction of the input wave shape. A *disadvantage* of increasing the resistance value is that the larger resistance increases the discharge time of the capacitor and slows the self-adjustment rate of the circuit, particularly in case a sudden decrease in input amplitude should occur. The larger resistance has no effect on self-adjustment with a sudden *rise* in input amplitude. This is because the capacitor charges through the small resistance of the conducting diode.

Circuits often incorporate a compromise between a short RC time constant (for self-adjustment purposes) and a long RC time constant for less distortion. A point to observe is that the reverse resistance of the diode sometimes replaces the, physical resistor in the discharge path of the capacitor.

Positive-Diode Clamper With Bias

Biased clamping circuits operate in exactly the same manner as unbiased clampers, with one exception. That exception is the addition of a dc bias voltage in series with the diode and resistor. The size and polarity of this bias voltage determines the output clamping reference.

View (A) of figure 4-19 illustrates the circuit of a positive clamper with positive bias. It can be identified as a positive clamper because the cathode of the diode is connected to the capacitor. Positive bias can be observed by noting that the negative side of the battery is connected to ground. The purposes and actions of the capacitor, resistor, and diode are the same as in the unbiased clamper circuit just discussed.

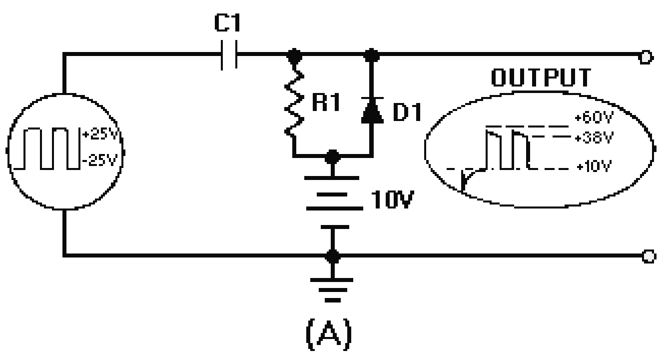


Figure 4-19A.—Positive clamper with positive bias.

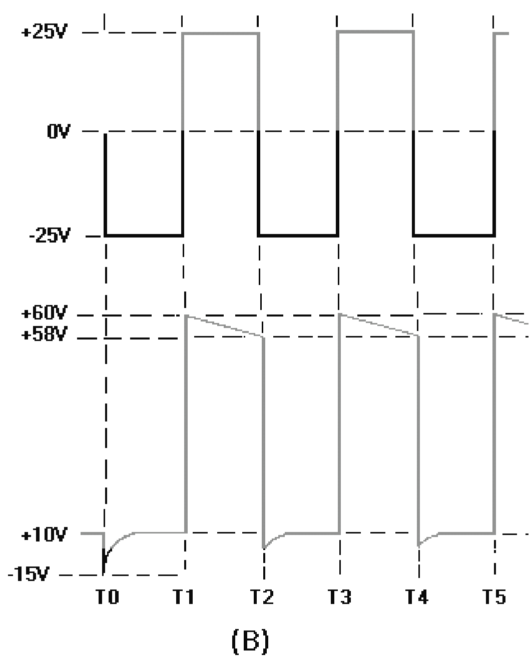


Figure 4-19B.—Positive clamper with positive bias.

With no input, D1 is forward biased and the +10 V battery is the output. C1 will charge to +10 V and hold this charge until the first pulse is applied. The battery establishes the dc reference level at +10 volts. The input wave shape at the top of view (B) is a square wave which alternates between +25 and −25 volts. The output wave shape is shown at the bottom half of view (B).

Here, as with previous circuits, let's apply Kirchhoff's voltage law to determine circuit operation. With no input signal, the output is just the +10 volts supplied by the battery.

At time T0 the −25 volt signal applied to the circuit is instantly felt across R1 and D1. The −25 volt input signal forward biases D1, and C1 quickly charges to 35 volts. This leaves +10 volts across the output terminals for much of the period from T0 to T1. The polarity of the charged capacitor is, from the left to the right, minus to plus.

At T1 the 35 volts across the capacitor is series aiding with the +25 volt input signal. At this point (T1) the output voltage becomes +60 volts; the voltage across R1 and D1 is +50 volts, and the battery is +10 volts. The cathode of D1 is positive with respect to the anode and the diode does not conduct. From T1 to T2, C1 discharges only slightly through the large resistance of R1. Assume that, because of the size of R1 and C1, the capacitor discharges just 2 volts (from +35 volts to +33 volts) during this period. Thus, the output voltage drops from +60 volts to +58 volts.

At T2 the −25 volt input signal and the +33 volts across C1 are series opposing. This makes the voltage across the output terminals +8 volts. The cathode of the diode is 2 volts negative with respect to its anode and D1 conducts. Again, since the forward-biased diode is essentially a short, C1 quickly charges from +33 volts to +35 volts. During most of the time from T2 to T3, then, we find the output voltage is +10 volts.

At T3 the +25 volts of the input signal is series aiding with the +35 volts across C1. Again the output voltage is +60 volts. Observe that at T3 the conditions in the circuit are the same as they were at T1. Therefore, the circuit operation from T3 to T4 is the same as it was from T1 to T2. Circuit operation continues as a duplication of the operations which occurred from T1 to T3.

By comparing the input and output wave shapes, you should note the following: (1) The peak-to-peak amplitude of the input wave shape has not been changed in the output (for all practical purposes) by the action of the clamper circuit; (2) the shape of the input wave has not been changed; (3) the output wave shape is now clamped above +10 volts. Remember that this clamping level (+10 volts) is determined by the bias battery.

Positive-Diode Clamper With Negative Bias

View (A) of figure 4-20 is a positive clamper with negative bias. Observe that with no input signal, the capacitor charges through R1 to the bias battery voltage; the output voltage equals −10 volts. The circuit has negative bias because the positive side of the battery is grounded. The output waveform is shown in view (B). Study the figure and waveforms carefully and note the following important points. Once again the peak-to-peak amplitude and shape of the output wave are, for all practical purposes, the same as the input wave. The lower extremity of the output wave is clamped to −10 volts, the value of the battery. Let's look at the circuit operation. The capacitor is initially charged to −10 volts with no input signal, and diode D1 does not conduct.

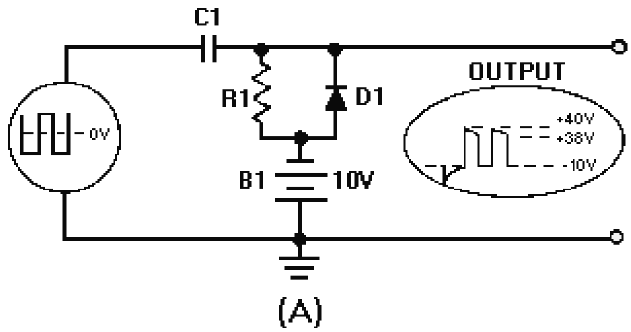


Figure 4-20A.—Positive clamper with negative bias.

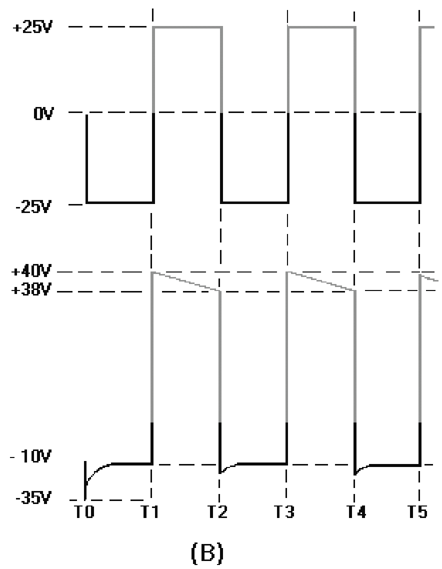
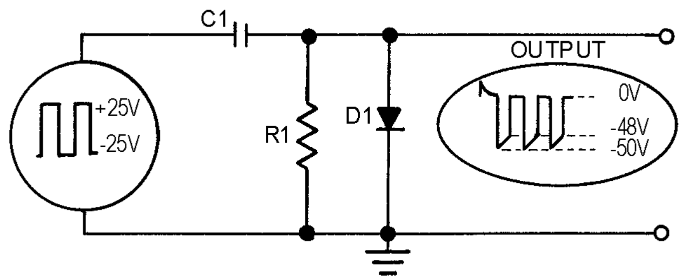


Figure 4-20B.—Positive clamper with negative bias.

The −25 volt input signal provides forward bias for D1. The capacitor charges to +15 volts and retains most of its charge because its discharge through R1 is negligible. The +25 volt input signal is series aiding the capacitor voltage and develops +40 volts between the output terminals. When the input voltage is −25 volts, D1 conducts and the output voltage is −10 volts (−25 volts plus +15 volts). In this way the output reference is clamped at −10 volts. Changing the size of the battery changes the clamping reference level to the new voltage.

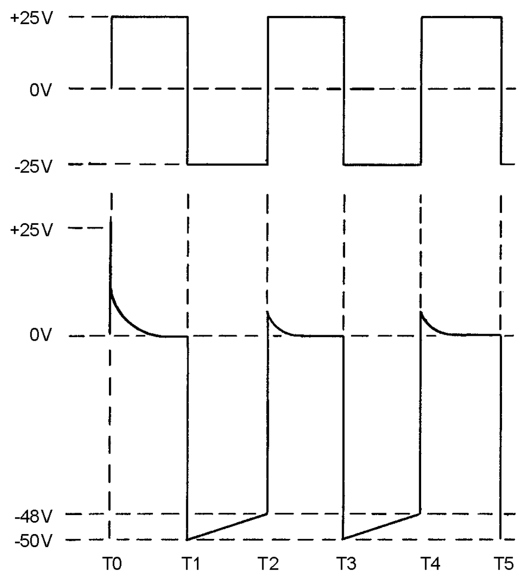
NEGATIVE-DIODE CLAMPERS

Figure 4-21, view (A), illustrates the circuit of a negative-diode clamper. Compare this with the positive-diode clamper in view (A) of figure 4-18. Note that the diode is reversed with reference to ground. Like the positive clamper, resistor R1 provides a discharge path for C1; the resistance must be a large value for C1 to have a long discharge time. The low resistance of the diode provides a fast charge path for C1. Once C1 becomes charged, it acts as a source of voltage which will help determine the maximum and minimum voltage levels of the output wave shape. The input wave shape shown in view (B) is a square wave which varies between +25 and −25 volts. The output wave shapes are shown in the bottom half of view (B). You will find that the operation of the negative clamper is similar to that of the positive clamper, except for the reversal of polarities.



(A)

Figure 4-21A.—Negative clamper and waveform.



(B)

Figure 4-21B.—Negative clamper and waveform.

At T0 the +25 volt input signal applied to the circuit appears across R1 and D1. This makes the anode of D1 positive with respect to the cathode and it conducts heavily. Diode resistance is very small causing C1 to charge quickly. As the voltage across C1 increases, the output voltage decreases. The voltage across C1 reaches 25 volts quickly; during most of T0 to T1, the output voltage is 0.

At T1 the voltage across the capacitor and the input voltage are series aiding and result in −50 volts appearing at the output. At this time the diode is reverse biased and does not conduct. Because of the size of R and C, the capacitor discharges only 2 volts to approximately 23 volts from T1 to T2. Using Kirchhoff's voltage law to determine voltage in the circuit, we find that the output voltage decreases from −50 to −48 volts.

At T2 the +25 volt input signal and the 23 volts across C1 are series opposing. The output voltage is +2 volts. The anode of D1 is positive with respect to the cathode and D1 will conduct. From T2 to T3, C1 charges quickly from 23 to 25 volts through D1. At the same time, the output voltage falls from +2 to 0 volts.

At T3 the input and capacitor voltages are series aiding and the total output voltage is −50 volts. From T3 to T4, D1 is reverse biased and C discharges through R. The circuit operation is now the same as it was from T1 to T2. The circuit operation for the following square-wave cycles duplicates the operation which occurred from T1 to T3.

As was the case with the positive clamper, the amplitude and wave, shape of the output is almost identical to that of the input. However, note that the upper extremity of the output wave shape is clamped to 0 volts; that is, the output wave shape, for all practical purposes, lies entirely below the 0-volt reference level.

Negative-Diode Clamper With Negative Bias

View (A) of figure 4-22 is the circuit of a negative clamper with negative bias. Again, with no input signal the capacitor charges to the battery voltage and the output is negative because the positive side of the battery is ground. The bottom of view (B) shows the output of the circuit. Study the figure carefully, and note the following important points. The peak-to-peak amplitude and shape of the output wave, for all practical purposes, are the same as that of the input wave. The output wave is clamped to −10 volts which is the value of the battery. Since this is a negative clamper, the upper extremity of the waveform touches the −10 volt reference line (and the rest of it lies below this voltage level).

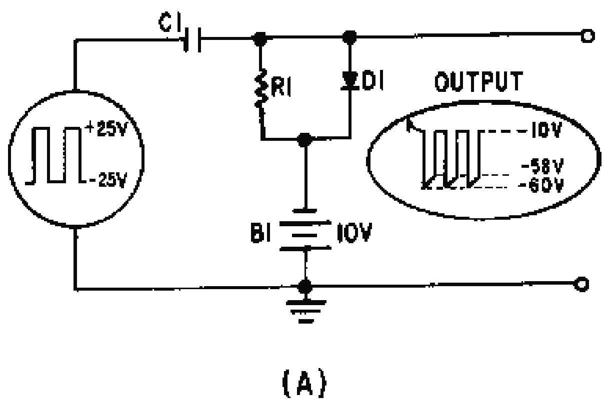


Figure 4-22A.—Negative damper with negative bias.

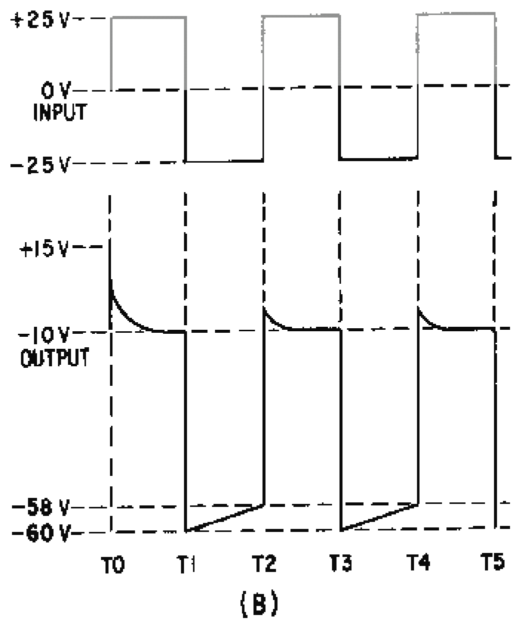


Figure 4-22B.—Negative damper with negative bias.

Let's review the important points of circuit operation. The capacitor is initially charged to –10 volts with no input signal. Applying Kirchhoff's law we find that the +25 volt input signal and the 10-volt battery are series opposing. This series opposing forward biases D1 and the capacitor charges to –35 volts. The output voltage is equal to the sum of the capacitor voltage and the input voltage. Thus, the output voltage is –10 volts and the wave shape is clamped to –10 volts. With a –25 volt input, the charge maintained across C1 and the input are series aiding and provide a –60 volt output. C1 will discharge just before the next cycle begins and the input becomes positive. The +25 volt input signal and the approximately –23 volt charge remaining on C1 will forward bias D1 and the output will be clamped to the battery voltage. C1 will quickly charge to the input signal level. Thus, the output voltage varies between –10 and –60 volts and the wave shape is clamped to –10 volts.

Negative Clamper With Positive Bias

View (A) of figure 4-23 illustrates the circuit of a negative clamper with positive bias. With no input signal the capacitor charges to the battery voltage and the output is positive because the negative side of the battery is grounded. The output is illustrated in the bottom half of view (B). Study the figure carefully and note the following important points. The peak-to-peak amplitude and shape of the output waveform, for all practical purposes, are the same as that of the input. The output wave is clamped to +10 volts, the value of the battery. Since this is a negative clamper (cathode to ground), the top of the output wave touches the +10 volt reference line.

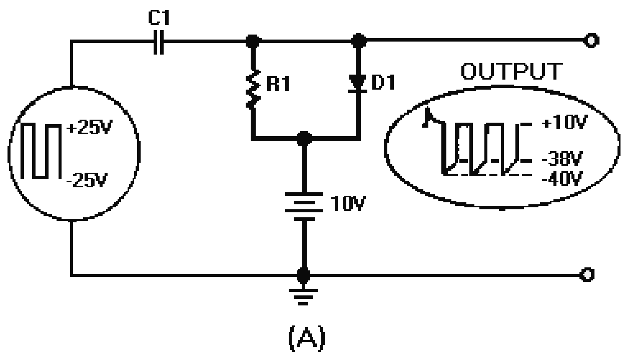


Figure 4-23A.—Negative clamper with positive bias.

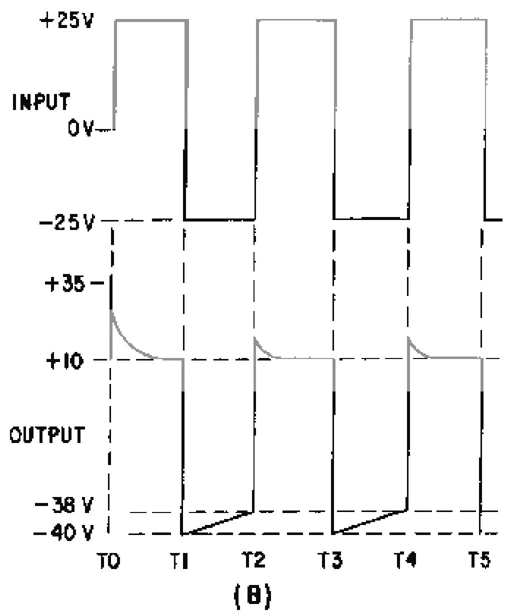


Figure 4-23B.—Negative clamper with positive bias.

Let's go over a summary of the circuit operation. With no input signal the capacitor charges to 10 volts. The +25 volt input signal forward biases D1. With the 10-volt battery and the input in series, the capacitor charges to −15 volts. The capacitor remains charged, for all practical purposes, since its discharge through R1 (very large) is almost negligible. The output voltage is equal to the algebraic sum of the capacitor voltage and the input voltage. The +25 volt input signal added to the −15 volt capacitor charge provides a +10 volt output. With a −25 volt input at T1, D1 is reverse-biased and the charge across C 1 adds to the input voltage to provide a −40 volt output. From T1 to T2, the capacitor loses only a small portion of its charge. At T2 the input signal is +25 volts and the input returns to +10 volts. The wave shape is negatively clamped to +10 volts by the battery.

We can say, then, that positive clamping sets the wave shape above (negative peak on) the reference level, and negative clamping places the wave shape below (positive peak on) the reference level.

Q6. What is the relative length of the time constant for the diode-capacitor combination in a damper (long or short)?

Q7. What is the relative length of the discharge time constant with respect to the charge time constant of a damper (long or short)?

Q8. A positive damper clamps which extremity of the output signal to 0 volts?

Q9. To which polarity does a positive damper with positive bins clamp the most negative extremity of the output waveform (positive or negative)?

Q10. What type damper (with bias) clamps the most negative extremity of the output waveform to a negative potential?

Q11. A negative damper damps which extremity of the output waveshape to 0 volts?

Q12. A negative damper with negative bias clamps the most positive extremity of the output wave shape to what polarity (positive or negative)?

Q13. What type of bias (positive or negative) is added to a negative damper for the most positive extremity of the wave shape to be clamped above 0 volts?

Q14. What would be the output of a negative clamper with a bias potential of –5 volts and an input voltage swing from +50 to –50 volts?

COMMON-BASE TRANSISTOR CLAMPER

The common-base transistor clamper is similar to the dual diode limiter in figure 4-15, except for the addition of a transistor. In the previous clampers, we have clamped the output signal to a reference. In the transistor common-base clamper, we want to clamp the amplitude of the input to no more than nor less than certain values in the output. Also, we do not want phase inversion in the output. View (A) of figure 4-24 shows such a circuit. The transistor does not amplify the input and the output is not inverted. However, the two diode circuits serve to clamp the output between –2 volts and –8 volts, no matter what the varying input positive and negative extremes.

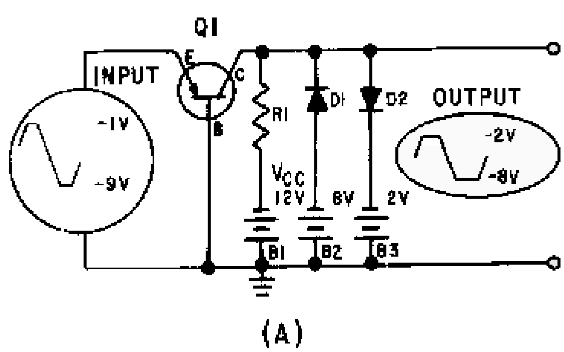


Figure 4-24A.—Common-base configuration clamper.

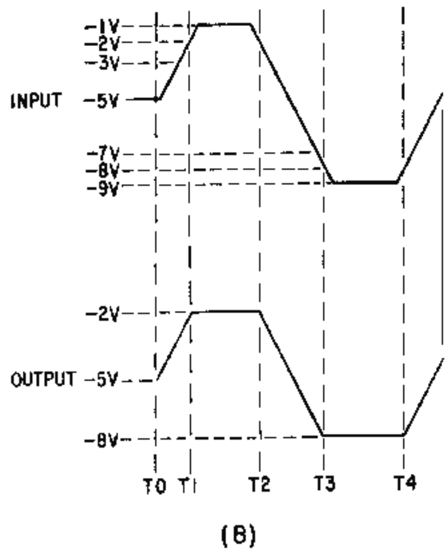


Figure 4-24B.—Common-base configuration clamper.

Look at view (B) as we continue. The input signal is a square-wave pulse type signal that varies in amplitude. Without an input signal, Q1 conducts and provides current through R1. This develops the output (collector to ground) potential which is assumed to be approximately −5 volts ($V_{CC} - E_{R1}$) for this discussion.

From T0 to T1 the output follows the input because of the increasing emitter-base forward bias. However, at T1 the collector voltage reaches −2 volts and D2 is forward biased. D2 conducts and limits the output to −2 volts (the value of B3). D2 conducts until T2 when the input decreases below −2 volts. At this time, D2 cuts off and the output again follows the input because of the decreasing forward bias on Q1. At T3 the input reaches −8 volts and forward biases D1. D1 conducts and any further increase (beyond −8 volts) of the input has no effect on the output. When the input returns to a value more positive than −8 volts, D1 cuts off and the output again follows the input. This circuit action is the same for all inputs. The output remains the same as the input except that both positive and negative extremes are clamped at −2 and −8 volts, respectively.

SHAPING CIRCUITS

Timing circuits and circuits which require a particular shape or "spike" of voltage, may use SHAPING circuits. Shaping circuits can be used to cause wave shapes, such as square waves, sawtooth waves, and trapezoidal waves, to change their shape. Shaping circuits may be either series RC or series RL circuits. The time constant is controlled in respect to the duration of the applied waveform. Notice that the wave shapes mentioned did not include the sine wave. These RC or RL shaping circuits do not change the shape of a pure sine wave.

The series RC and RL circuits electrically perform the mathematical operations of INTEGRATION and DIFFERENTIATION. Therefore, the circuits used to perform these operations are called INTEGRATORS and DIFFERENTIATORS. These names are applied to these circuits even though they do not always completely perform the operations of mathematical integration and differentiation.

COMPOSITION OF NONSINUSOIDAL WAVES

Pure sine waves are basic wave shapes from which other wave shapes can be constructed. Any waveform that is not a pure sine wave consists of two or more sine waves. Adding the correct frequencies at the proper phase and amplitude will form square waves, sawtooth waves, and other nonsinusoidal waveforms.

A waveform other than a sine wave is called a COMPLEX WAVE. You will see that a complex wave consists of a fundamental frequency plus one or more HARMONIC frequencies. The shape of a nonsinusoidal waveform is dependent upon the type of harmonics present as part of the waveform, their relative amplitudes, and their relative phase relationships. In general, the steeper the sides of a waveform, that is, the more rapid its rise and fall, the more harmonics it contains.

The sine wave which has the lowest frequency in the complex periodic wave is referred to as the FUNDAMENTAL FREQUENCY. The type and number of harmonics included in the waveform are dependent upon the shape of the waveform. Harmonics have two classifications — EVEN numbered and ODD numbered. Harmonics are always a whole number of times higher than the fundamental frequency and are designated by an integer (whole number). For example, the frequency twice as high as the fundamental frequency is the SECOND HARMONIC (or the first even harmonic).

View (A) of figure 4-25 compares a square wave with sine waves. Sine wave K is the same frequency as the square wave (its fundamental frequency). If another sine wave (L) of smaller amplitude but three times the frequency (referred to as the third harmonic) is added to sine wave K, curve M is produced. The addition of these two waveforms is accomplished by adding the instantaneous values of both sine waves algebraically. Curve M is called the resultant. Notice that curve M begins to assume the shape of a square wave. Curve M is shown again in view (B).

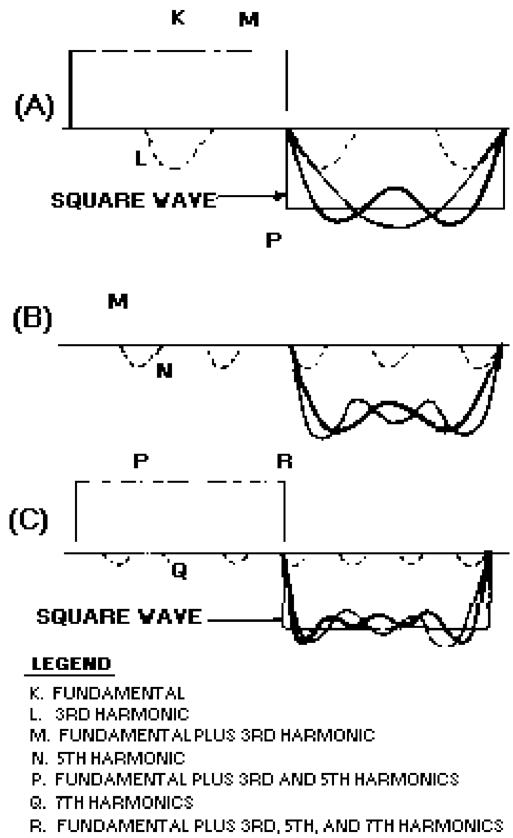


Figure 4-25.—Harmonic composition of a square wave.

As shown in view (B), when the fifth harmonic (curve N with its decreased amplitude) is added, the sides of the new resultant (curve P) are steeper than before. In view (C), the addition of the seventh harmonic (curve Q), which is of even smaller amplitude, makes the sides of the composite waveform (R) still steeper. The addition of more odd harmonics will bring the composite waveform nearer the shape of the perfect square wave. A perfect square wave is, therefore, composed of an infinite number of odd harmonics. In the composition of square waves, all the odd harmonics cross the reference line in phase with the fundamental.

A sawtooth wave, shown in figure 4-26, is made up of both even and odd harmonics. Notice that each higher harmonic is added in phase as it crosses the 0 reference in view (A), view (B), view (C), and view (D). The resultant, shown in view (D), closely resembles a sawtooth waveform.

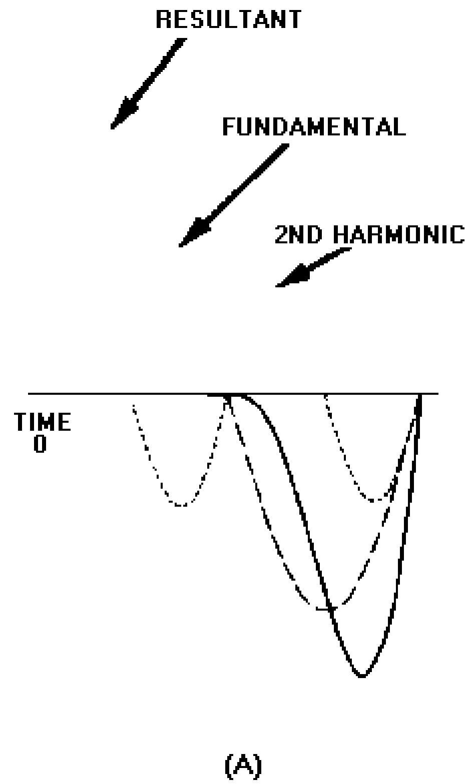


Figure 4-26A.—Composition of a sawtooth wave.

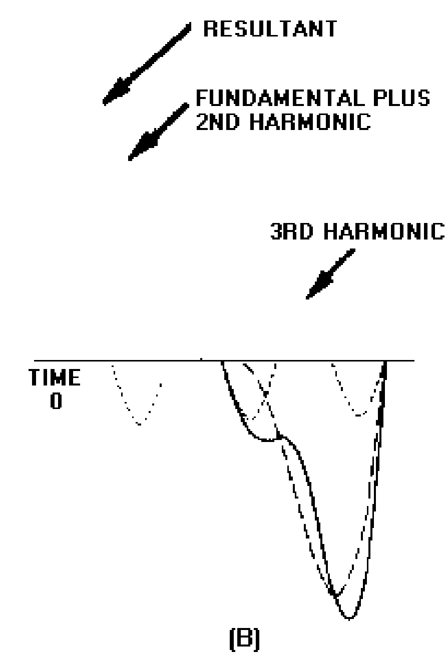


Figure 4-26B.—Composition of a sawtooth wave.

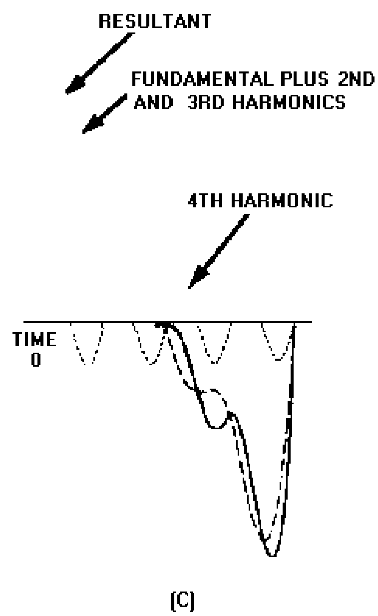


Figure 4-26C.—Composition of a sawtooth wave.

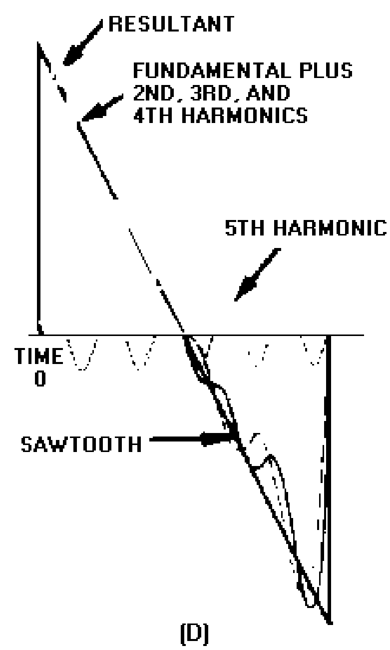


Figure 4-26D.—Composition of a sawtooth wave.

Figure 4-27 shows the composition of a peaked wave. Notice how the addition of each odd harmonic makes the peak of the resultant higher and the sides steeper. The phase relationship between the harmonics of the peaked wave is different from the phase relationship of the harmonics in the composition of the square wave. In the composition of the square wave, all the odd harmonics cross the reference line in phase with the fundamental. In the peaked wave, harmonics such as the third, seventh, and so forth, cross the reference line 180 degrees out of phase with the fundamental; the fifth, ninth, and so forth, cross the reference line in phase with the fundamental.

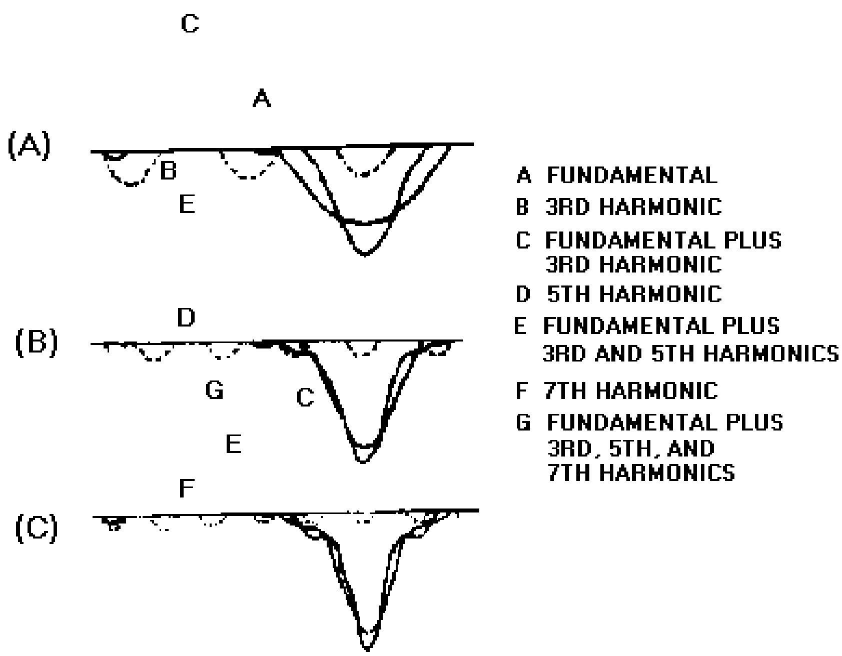


Figure 4-27.—Composition of a peaked wave.

Q15. What is the harmonic composition of a square wave?

Q16. What is the peaked wave composed of?

Q17. What is the fundamental difference between the phase relationship of the harmonics of the square wave as compared to the harmonics of a peaked wave?

Nonsinusoidal Voltages Applied to an RC Circuit

The harmonic content of a square wave must be complete to produce a pure square wave. If the harmonics of the square wave are not of the proper phase and amplitude relationships, the square wave will not be pure. The term PURE, as applied to square waves, means that the waveform must be perfectly square.

Figure 4-28 shows a pure square wave that is applied to a series-resistive circuit. If the values of the two resistors are equal, the voltage developed across each resistor will be equal; that is, from one pure square-wave input, two pure square waves of a lower amplitude will be produced. The value of the resistors does not affect the phase or amplitude relationships of the harmonics contained within the square waves. This is true because the same opposition is offered by the resistors to all the harmonics presented. However, if the same square wave is applied to a series RC circuit, as shown in figure 4-29, the circuit action is not the same.

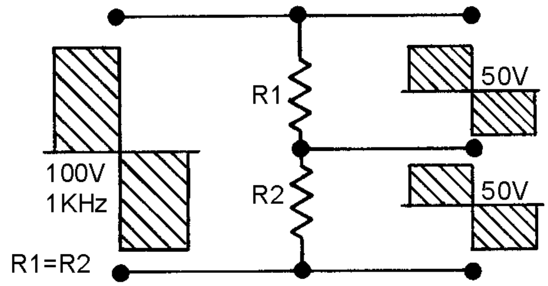


Figure 4-28.—Square wave applied to a resistive circuit.

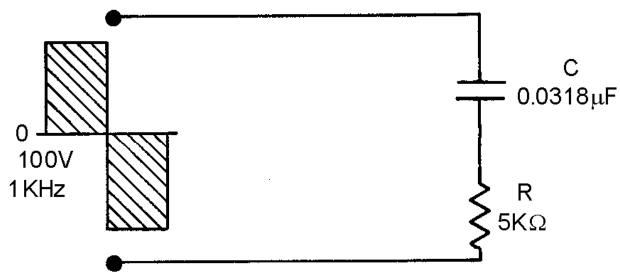


Figure 4-29.—Square wave applied to an RC circuit.

RC INTEGRATORS

The RC INTEGRATOR is used as a waveshaping network in communications, radar, and computers. The harmonic content of the square wave is made up of odd multiples of the fundamental frequency. Therefore SIGNIFICANT HARMONICS (those that have an effect on the circuit) as high as 50 or 60 times the fundamental frequency will be present in the wave. The capacitor will offer a reactance (X_C) of a different magnitude to each of the harmonics

$$X_C = \frac{1}{2\pi fC}$$

This means that the voltage drop across the capacitor for each harmonic frequency present will not be the same. To low frequencies, the capacitor will offer a large opposition, providing a large voltage drop across the capacitor. To high frequencies, the reactance of the capacitor will be extremely small, causing a small voltage drop across the capacitor. This is no different than was the case for low- and high-pass filters (discriminators) presented in chapter 1. If the voltage component of the harmonic is not developed across the reactance of the capacitor, it will be developed across the resistor, if we observe Kirchhoff's voltage law. The harmonic amplitude and phase relationship across the capacitor is not the same as that of the original frequency input; therefore, a perfect square wave will not be produced across the capacitor. You should remember that the reactance offered to each harmonic frequency will cause a change in both the amplitude and phase of each of the individual harmonic frequencies with respect to the current reference. The amount of phase and amplitude change taking place across the capacitor depends on the X_C of the capacitor. The value of the resistance offered by the resistor must also be considered here; it is part of the ratio of the voltage development across the network.

The circuit in figure 4-30 will help show the relationships of R and X_C more clearly. The square wave applied to the circuit is 100 volts peak at a frequency of 1 kilohertz. The odd harmonics will be 3 kilohertz, 5 kilohertz, 7 kilohertz, etc. Table 4-1 shows the values of X_C and R offered to several harmonics and indicates the approximate value of the cutoff frequency ($X_C = R$). The table clearly shows that the cutoff frequency lies between the fifth and seventh harmonics. Between these two values, the capacitive reactance will equal the resistance. Therefore, for all harmonic frequencies above the fifth, the majority of the output voltage will not be developed across the output capacitor. Rather, most of the output will be developed across R. The absence of the higher order harmonics will cause the leading edge of the waveform developed across the capacitor to be rounded. An example of this effect is shown in figure 4-31. If the value of the capacitance is increased, the reactances to each harmonic frequency will be further decreased. This means that even fewer harmonics will be developed across the capacitor.

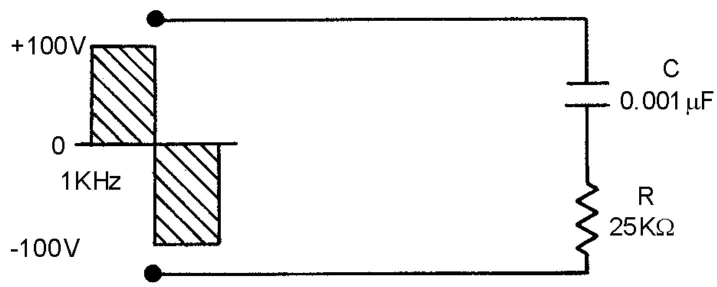


Figure 4-30.—Partial integration circuit.

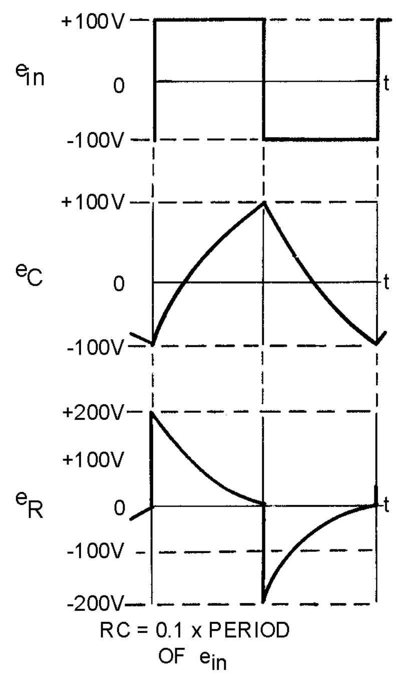


Figure 4-31.—Partial integration.

Table 4-1.—Resistive and reactive values

HARMONIC	X_C	R
FUNDAMENTAL	159 K Ω	25 K Ω
3rd	53 K Ω	25 K Ω
5th	31.8 K Ω	25 K Ω
7th	22.7 K Ω	25 K Ω
9th	17.7 K Ω	25 K Ω
11th	14.5KΩ	25 KΩ

The harmonics not effectively developed across the capacitor must be developed across the resistor to satisfy Kirchhoff's voltage law. Note the pattern of the voltage waveforms across the resistor and capacitor. If the waveforms across both the resistor and the capacitor were added graphically, the resultant would be an exact duplication of the input square wave.

When the capacitance is increased sufficiently, full integration of the input signal takes place in the output across the capacitor. An example of complete integration is shown in figure 4-32 (waveform e_C). This effect can be caused by significantly decreasing the value of capacitive reactance. The same effect would take place by increasing the value of the resistance. Integration takes place in an RC circuit when the output is taken across the capacitor.

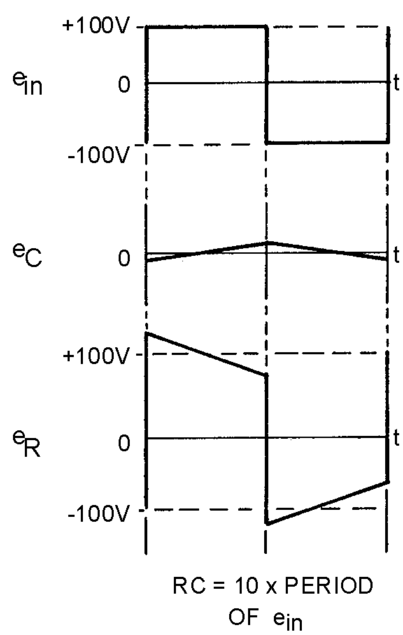


Figure 4-32.—Integration.

The amount of integration is dependent upon the values of R and C. The amount of integration may also be dependent upon the time constant of the circuit. The time constant of the circuit should be at least 10 TIMES GREATER than the time duration of the input pulse for integration to occur. The value of 10 is only an approximation. When the time constant of the circuit is 10 or more times the value of the duration of the input pulse, the circuit is said to possess a long time constant. When the time constant is long, the capacitor does not have the ability to charge instantly to the value of the applied voltage. Therefore, the result is the long, sloping, integrated waveform.

Q18. What are the requirements for an integration circuit?

Q19. Can a pure sine wave be integrated? Why?

RL INTEGRATORS

The RL circuit may also be used as an integrating circuit. An integrated waveform may be obtained from the series RL circuit by taking the output across the resistor. The characteristics of the inductor are such that at the first instant of time in which voltage is applied, current flow through the inductor is minimum and the voltage developed across it is maximum. Therefore, the value of the voltage drop across the series resistor at that first instant must be 0 volts because there is no current flow through it. As time passes, current begins to flow through the circuit and voltage develops across the resistor. Since the circuit has a long time constant, the voltage across the resistor does NOT respond to the rapid changes in voltage of the input square wave. Therefore, the conditions for integration in an RL circuit are a long time constant with the output taken across the resistor. These conditions are shown in figure 4-33.

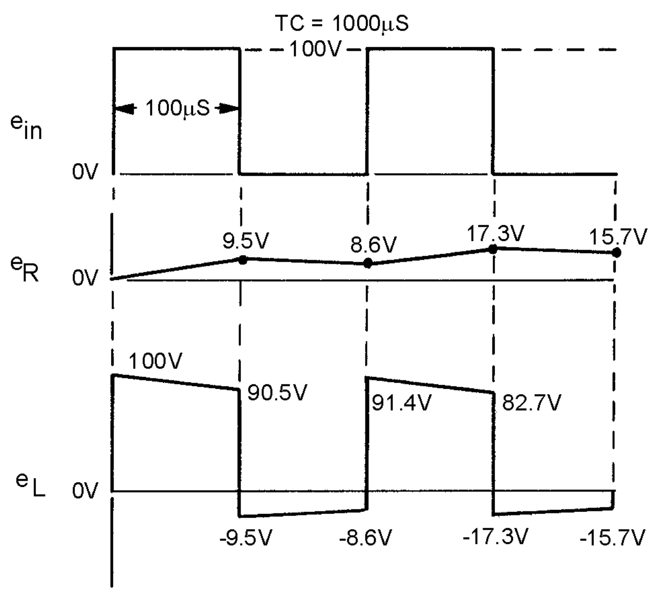


Figure 4-33.—RL integrator waveform.

Q20. What characteristic of an RL circuit allows it to act as an integrator?

INTEGRATOR WAVEFORM ANALYSIS

If either an RC or RL circuit has a time constant 10 times greater than the duration of the input pulse, the circuits are capable of integration. Let's compute and graph the actual waveform that would result from a long time constant (10 times the pulse duration), a short time constant (1/10 of the pulse duration), and a medium time constant (that time constant between the long and the short). To accurately plot values for the capacitor output voltage, we will use the Universal Time Constant Chart shown in figure 4-34.

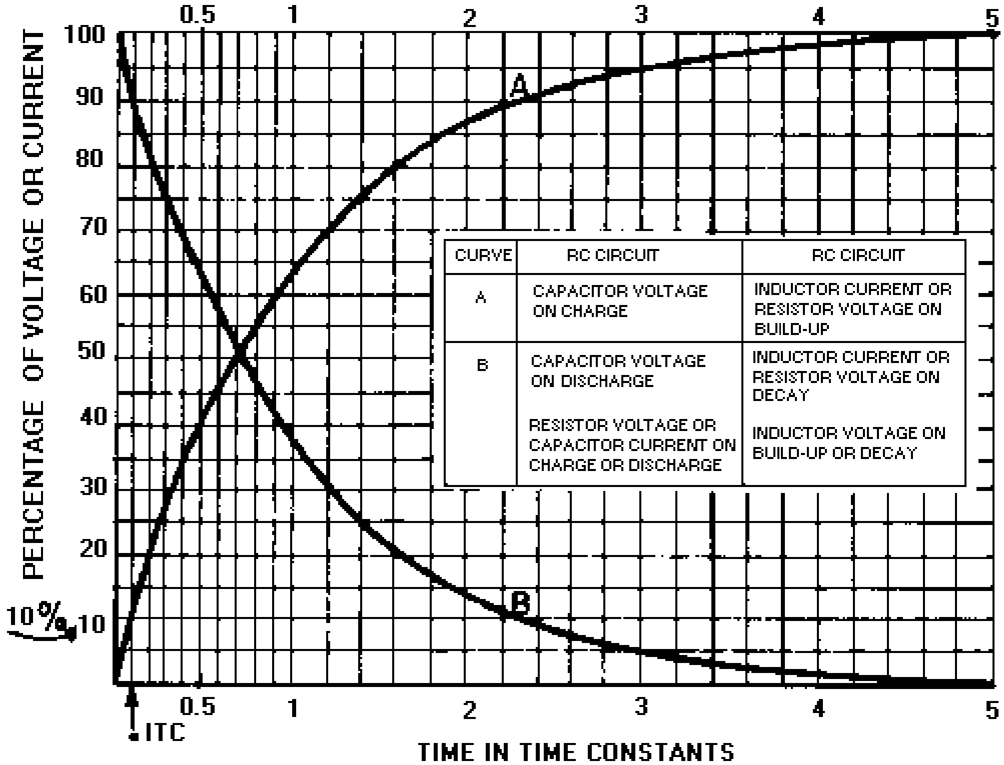


Figure 4-34.—Universal Time Constant Chart.

You already know that capacitor charge follows the shape of the curve shown in figure 4-34. This curve may be used to determine the amount of voltage across either component in the series RC circuit. As long as the time constant or a fractional part of the time constant is known, the voltage across either component may be determined.

Short Time-Constant Integrator

In figure 4-35, a 100-microsecond pulse at an amplitude of 100 volts is applied to the circuit. The circuit is composed of the, 0.01μF capacitor and the variable resistor, R. The square wave applied is a pure square wave. The resistance of the variable resistor is set at a value of 1,000 ohms. The time constant of the circuit is given by the equation:

$$TC = RC$$

Substituting values:

$$T = 1{,}000 \cdot 0.01\mu F$$

$$T = (1 \times 10^{3}) \cdot (1 \times^{-8})$$

$T = 1 \times 10^{-5}$ vor 10 microseconds

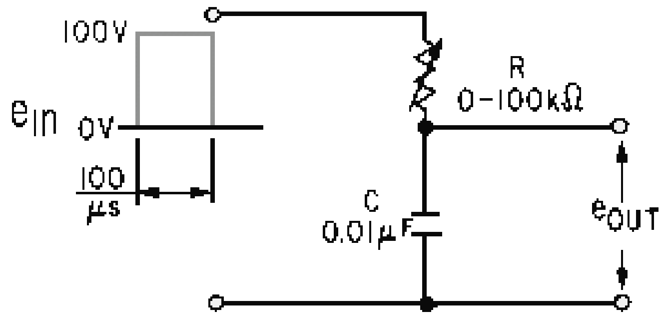


Figure 4-35.—RC integrator circuit.

Since the time constant of the circuit is 10 microseconds and the pulse duration is 100 microseconds, the time constant is short (1/10 of the pulse duration). The capacitor is charged exponentially through the resistor. In 5 time constants, the capacitor will be, for all practical purposes, completely charged. At the first time constant, the capacitor is charged to 63.2 volts, at the second 86.5 volts, at the third 95 volts, at the fourth 98 volts, and finally at the end of the fifth time constant (50 microseconds), the capacitor is fully charged. This is shown in figure 4-36.

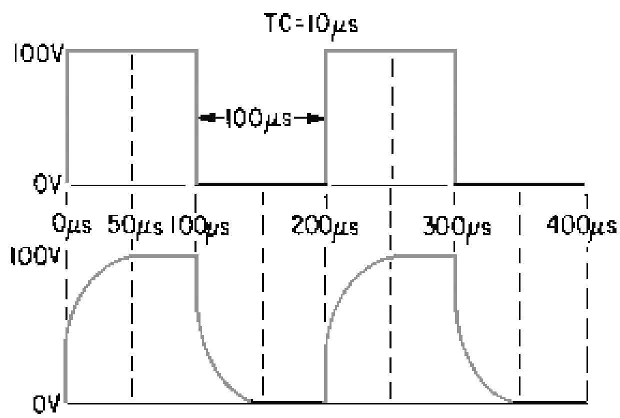


Figure 4-36.—Square wave applied to a short time-constant integrator.

Notice that the leading edge of the square wave taken across the capacitor is rounded. If the time constant were made extremely short, the rounded edge would become square.

Medium Time-Constant Integrator

The time constant, in figure 4-36 can be changed by increasing the value of the variable resistor (figure 4-35) to 10,000 ohms. The time constant will then be equal to 100 microseconds.

This time constant is known as a medium time constant. Its value lies between the extreme ranges of the short and long time constants. In this case, its value happens to be exactly equal to the duration of the input pulse, 100 microseconds. The output waveform, after several time constants, is shown in figure

4-37. The long, sloping rise and fall of voltage is caused by the inability of the capacitor to charge and discharge rapidly through the 10,000-ohm series resistance.

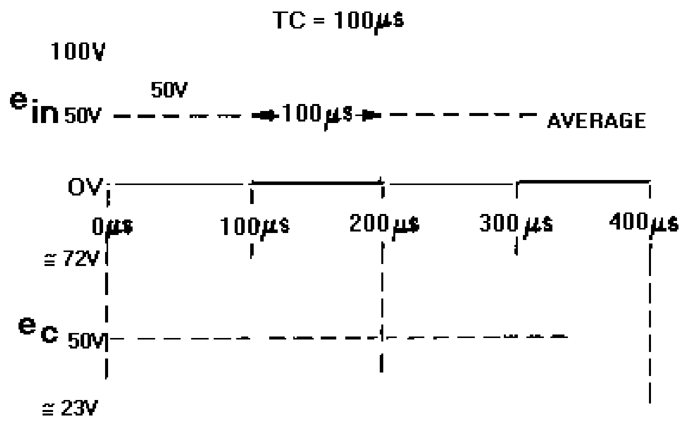


Figure 4-37.—Medium time-constant integrator.

At the first instant of time, 100 volts is applied to the medium time-constant circuit. In this circuit, 1TC is exactly equal to the duration of the input pulse. After 1TC the capacitor has charged to 63.2 percent of the input voltage (100 volts). Therefore, at the end of 1TC (100 microseconds), the voltage across the capacitor is equal to 63.2 volts. However, as soon as 100 microseconds has elapsed, and the initial charge on the capacitor has risen to 63.2 volts, the input voltage suddenly drops to 0. It remains there for 100 microseconds. The capacitor will now discharge for 100 microseconds. Since the discharge time is 100 microseconds (1TC), the capacitor will discharge 63.2 percent of its total 63.2-volt charge, a value of 23.3 volts. During the next 100 microseconds, the input voltage will increase from 0 to 100 volts instantaneously. The capacitor will again charge for 100 microseconds (1TC). The voltage available for this charge is the difference between the voltage applied and the charge on the capacitor (100 - 23.3 volts), or 76.7 volts. Since the capacitor will only be able to charge for 1TC, it will charge to 63.2 percent of the 76.7 volts, or 48.4 volts. The total charge on the capacitor at the end of 300 microseconds will be 23.3 + 48.4 volts, or 71.7 volts.

Notice that the capacitor voltage at the end of 300 microseconds is greater than the capacitor voltage at the end of 100 microseconds. The voltage at the end of 100 microseconds is 63.2 volts, and the capacitor voltage at the end of 300 microseconds is 71.7 volts, an increase of 8.5 volts.

The output waveform in this graph (e_C) is the waveform that will be produced after many cycles of input signal to the integrator. The capacitor will charge and discharge in a step-by-step manner until it finally charges and discharges above and below a 50-volt level. The 50-volt level is controlled by the maximum amplitude of the symmetrical input pulse, the average value of which is 50 volts.

Long Time-Constant Integrator

If the resistance in the circuit of figure 4-35 is increased to 100,000 ohms, the time constant of the circuit will be 1,000 microseconds. This time constant is 10 times the pulse duration of the input pulse. It is, therefore, a long time-constant circuit.

The shape of the output waveform across the capacitor is shown in figure 4-38. The shape of the output waveform is characterized by a long, sloping rise and fall of capacitor voltage.

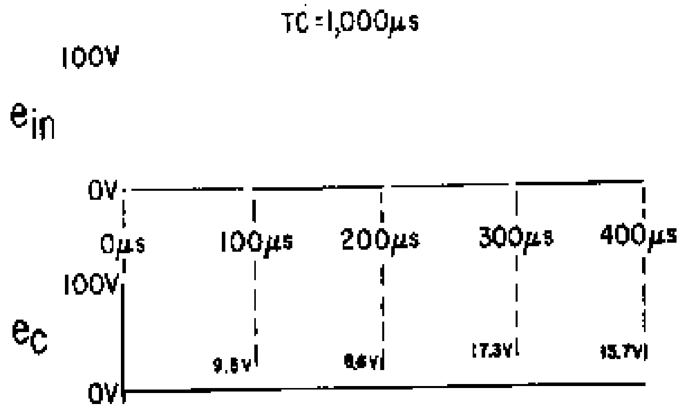


Figure 4-38.—Square wave applied to a long time-constant integrator.

At the first instant of time, 100 volts is applied to the long time-constant circuit. The value of charge on the capacitor at the end of the first 100 microseconds of the input signal can be found by using the Universal Time Constant Chart (figure 4-34). Assume that a line is projected up from the point on the base line corresponding to 0.1TC. The line will intersect the curve at a point that is the percentage of voltage across the capacitor at the end of the first 100 microseconds. Since the applied voltage is 100 volts, the charge on the capacitor at the end of the first 100 microseconds will be approximately 9.5 volts. At the end of the first 100 microseconds, the input signal will fall suddenly to 0 and the capacitor will begin to discharge. It will be able to discharge for 100 microseconds. Therefore, the capacitor will discharge 9.5 percent of its accumulated 9.5 volts ($.095 \times 9.5 = 0.90$ volt). The discharge of the 0.90 volt will result in a remaining charge on the capacitor of 8.6 volts. At the end of 200 microseconds, the input signal will again suddenly rise to a value of 100 volts. The capacitor will be able to charge to 9.5 percent of the difference ($100 - 8.6 = 91.4$ volts). This may also be figured as a value of 8.7 volts plus the initial 8.6 volts. This results in a total charge on the capacitor (at the end of the first 300 microseconds) of $8.7 + 8.6 = 17.3$ volts.

Notice that the capacitor voltage at the end of the first 300 microseconds is greater than the capacitor voltage at the end of the first 100 microseconds. The voltage at the end of the first 100 microseconds is 9.5 volts; the capacitor voltage at the end of the first 300 microseconds is 17.3 volts, an increase of 7.8 volts.

The capacitor charges and discharges in this step-by-step manner until, finally, the capacitor charges and discharges above and below a 50-volt level. The 50-volt level is controlled by the maximum amplitude of the square-wave input pulse, the average value of which is 50 volts.

Q21. What is the numerical difference (in terms of the time constant) between a long and a short time-constant circuit?

Q22. What would happen to the integrator output if the capacitor were made extremely large (all other factors remaining the same)?

DIFFERENTIATORS

DIFFERENTIATION is the direct opposite of integration. In the RC integrator, the output is taken from the capacitor. In the differentiator, the output is taken across the resistor. Likewise, this means that when the RL circuit is used as a differentiator, the differentiated output is taken across the inductor.

An application of Kirchhoff's law shows the relationship between the waveforms across the resistor and capacitor in a series network. Since the sum of the voltage drops in a closed loop must equal the total applied voltage, the graphical sum of the voltage waveforms in a closed loop must equal the applied waveform. Figure 4-39 shows a differentiator circuit with the output taken across a variable resistor.

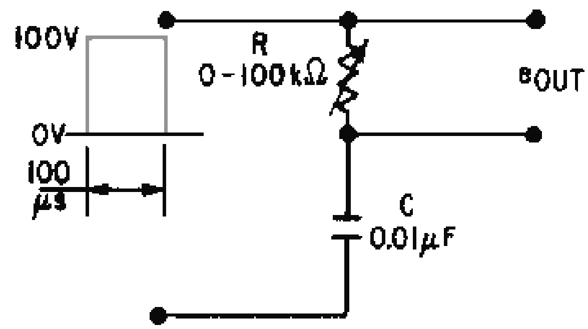


Figure 4-39.—RC circuit as a differentiator.

Short Time-Constant Differentiator

With the variable resistor set at 1,000 ohms and the capacitor value of 0.01 microfarad, the time constant of the circuit is 10 microseconds. Since the input waveform has a duration of 100 microseconds, the circuit is a short time-constant circuit.

At the first instant of time in the short time-constant circuit, the voltage across the capacitor is 0. Current flows through the resistor and causes a maximum voltage to be developed across it. This is shown at the first instant of time in the graph of figure 4-40.

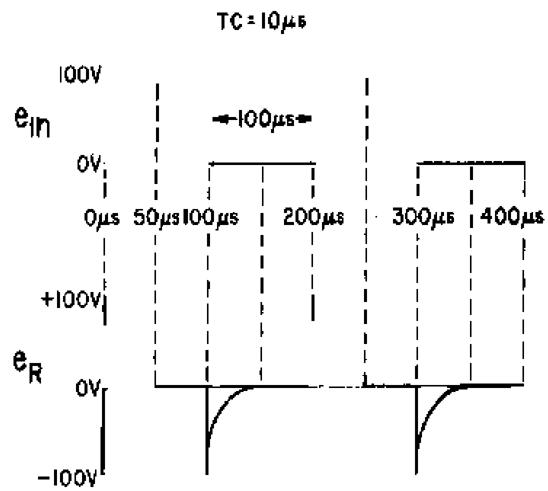


Figure 4-40.—Square wave applied to a short time-constant differentiator.

As the capacitor begins accumulating a charge, the voltage developed across the resistor will begin to decrease. At the end of the first time constant, the voltage developed across the resistor will have decreased by a value equal to 63.2 percent of the applied voltage. Since 100 volts is applied, the voltage across the resistor after 1TC will be equal to 36.8 volts. After the second time constant, the voltage across the resistor will be down to 13.5 volts. At the end of the third time constant, e_R will be 5 volts and at the end of the fourth time constant, 2 volts. At the end of the fifth time constant, the voltage across the resistor will be very close to 0 volts. Since the time constant is equal to 10 microseconds, it will take a total of 50 microseconds to completely charge the capacitor and stop current flow in the circuit.

As shown in figure 4-40 the slope of the charge curve will be very sharp. The voltage across the resistor will remain at 0 volts until the end of 100 microseconds. At that time, the applied voltage suddenly drops to 0, and the capacitor will now discharge through the resistor. At this time, the discharge current will be maximum causing a large discharge voltage to develop across the resistor. This is shown as the negative spike in figure 4-40. Since the current flow from the capacitor, which now acts like a source, is decreasing exponentially, the voltage across the resistor will also decrease. The resistor voltage will decrease exponentially to 0 volts in 5 time constants. All of this discharge action will take a total of 50 microseconds. The discharge curve is also shown in figure 4-40. At the end of 200 microseconds, the action begins again. The output waveform taken across the resistor in this short time-constant circuit is an example of differentiation. With the square wave applied, positive and negative spikes are produced in the output. These spikes approximate the rate of change of the input square wave.

Medium Time-Constant Differentiator

The output across the resistor in an RC circuit of a medium time constant is shown in figure 4-41. The value of the variable resistor has been increased to a value of 10,000 ohms. This means that the time constant of the circuit is equal to the duration of the input pulse or 100 microseconds. For clarity, the voltage waveforms developed across both the resistor and the capacitor are shown. As before, the sum of the voltages across the resistor and capacitor must be equal to the applied voltage of 100 volts.

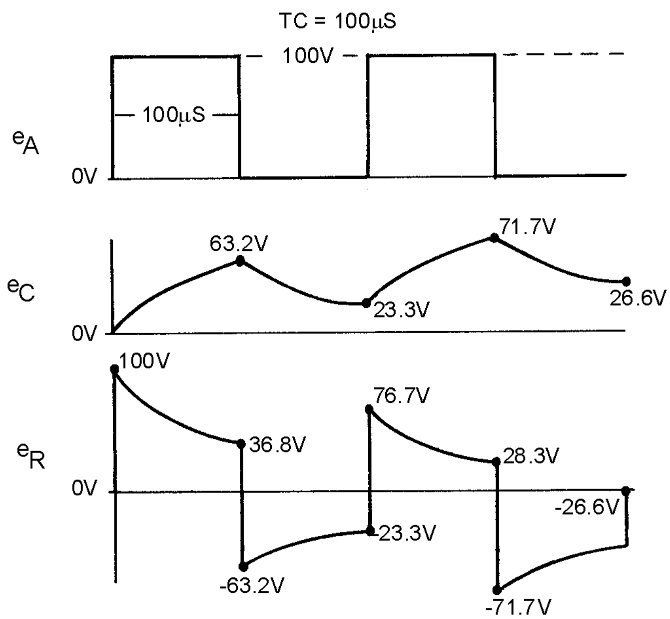


Figure 4-41.—Voltage outputs in a medium time-constant differentiator.

At the first instant of time, a pulse of 100 volts in amplitude with a duration of 100 microseconds is applied. Since the capacitor cannot respond quickly to the change in voltage, all of the applied voltage is felt across the resistor. Figure 4-41 shows the voltage across the resistor (e_R) to be 100 volts and the voltage across the capacitor (e_C) to be 0 volts. As time progresses, the capacitor charges. As the capacitor voltage increases, the resistor voltage decreases. Since the time that the capacitor is permitted to charge is 100 microseconds (equal to 1TC in this circuit), the capacitor will charge to 63.2 percent of the applied voltage at the end of 1TC, or 63.2 volts. Because Kirchhoff's law must be followed at all times, the voltage across the resistor must be equal to the difference between the applied voltage and the charge on the capacitor (100 – 63.2 volts), or 36.8 volts.

At the end of the first 100 microseconds, the input voltage suddenly drops to 0 volts. The charge on the capacitor (–63.2 volts) becomes the source and the entire voltage is developed across the resistor for the first instant.

The capacitor discharges during the next 100 microseconds. The voltage across the resistor decreases at the same rate as the capacitor voltage and total voltage is maintained at 0. This exponential decrease in resistor voltage is shown during the second 100 microseconds in figure 4-41. The capacitor will discharge 63.2 percent of its charge to a value of 23.3 volts at the end of the second 100 microseconds. The resistor voltage will rise in the positive direction to a value of –23.3 volts to maintain the total voltage at 0 volts.

At the end of 200 microseconds, the input voltage again rises suddenly to 100 volts. Since the capacitor cannot respond to the 100-volt increase instantaneously, the 100-volt change takes place across the resistor. The voltage across the resistor suddenly rises from –23.3 volts to +76.7 volts. The capacitor will now begin to charge for 100 microseconds. The voltage will decrease across the resistor. This charge and discharge action will continue for many cycles. Finally, the voltage across the capacitor will rise and fall by equal amounts both above and below about a 50-volt level. The resistor voltage will also rise and fall by equal amounts to about a 0-volt level.

Long Time-Constant Differentiator

If the time constant for the circuit in figure 4-39 is increased to make it a long time-constant circuit, the differentiator output will appear more like the input. The time constant for the circuit can be changed by either increasing the value of capacitance or resistance. In this circuit, the time constant will be increased by increasing the value of resistance from 10,000 ohms to 100,000 ohms. Increasing the value of resistance will result in a time constant of 1,000 microseconds. The time constant is 10 times the duration of the input pulse. The output of this long time-constant circuit is shown in figure 4-42.

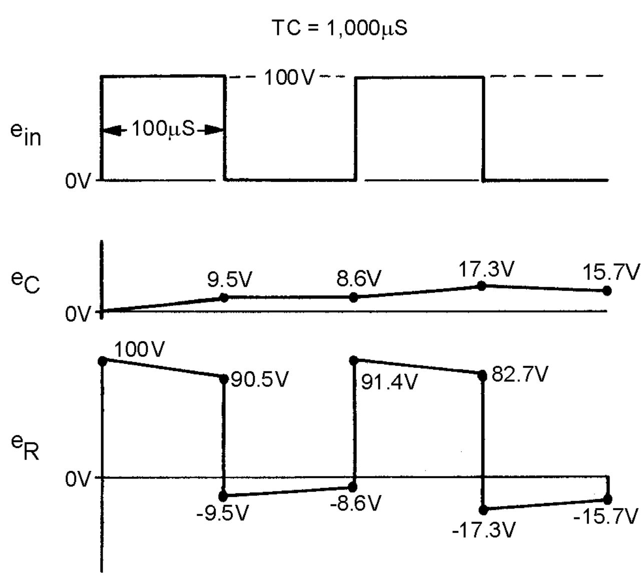


Figure 4-42.—Voltage outputs in a long time-constant differentiator.

At the first instant of time, a pulse of 100-volts amplitude with a duration of 100 microseconds is applied. Since the capacitor cannot respond instantaneously to a change in voltage, all of the applied voltage is felt across the resistor. As time progresses, the capacitor will charge and the voltage across the resistor will be reduced. Since the time that the capacitor is permitted to charge is 100 microseconds, the capacitor will charge for only 1/10 of 1TC or to 9.5 percent of the applied voltage. The voltage across the resistor must be equal to the difference between the applied voltage and the charge on the capacitor (100 – 9.5 volts), or 90.5 volts.

At the end of the first 100 microseconds of input, the applied voltage suddenly drops to 0 volts, a change of 100 volts. Since the capacitor is not able to respond to so rapid a voltage change, it becomes the source of 9.5 volts. This causes a –9.5 voltage to be felt across the resistor in the first instant of time. The sum of the voltage across the two components is now 0 volts.

During the next 100 microseconds, the capacitor discharges. The total circuit voltage is maintained at 0 by the voltage across the resistor decreasing at exactly the same rate as the capacitor discharge. This exponential decrease in resistor voltage is shown during the second 100 microseconds of operation. The capacitor will now discharge 9.5 percent of its charge to a value of 8.6 volts. At the end of the second 100 microseconds, the resistor voltage will rise in a positive direction to a value of –8.6 volts to maintain the total circuit voltage at 0 volts.

At the end of 200 microseconds, the input voltage again suddenly rises to 100 volts. Since the capacitor cannot respond to the 100-volt change instantaneously, the 100-volt change takes place across the resistor. This step-by-step action will continue until the circuit stabilizes. After many cycles have passed, the capacitor voltage varies by equal amounts above and below the 50-volt level. The resistor voltage varies by equal amounts both above and below a 0-volt level.

The RC networks which have been discussed in this chapter may also be used as coupling networks. When an RC circuit is used as a coupling circuit, the output is taken from across the resistor. Normally, a long time-constant circuit is used. This, of course, will cause an integrated wave shape across the capacitor if the applied signal is nonsinusoidal. However, in a coupling circuit, the signal across the resistor should closely resemble the input signal and will if the time constant is sufficiently long. By referring to the diagram in figure 4-42, you can see that the voltage across the resistor closely resembles the input signal. Consider what would happen if a pure sine wave were applied to a long time-constant RC circuit (R is much greater than X_C). A large percentage of the applied voltage would be developed across the resistor and only a small amount across the capacitor.

Q23. What is the difference between an RC and an RL differentiator in terms of where the output is developed?

COUNTERS

A counting circuit receives uniform pulses representing units to be counted. It provides a voltage that is proportional to the frequency of the units.

With slight modification, the counting circuit can be used with a blocking oscillator to produce trigger pulses which are a submultiple of the frequency of the pulses applied. In this case the circuit acts as a frequency divider.

The pulses applied to the counting circuit must be of the same time duration if accurate frequency division is to be made. Counting circuits are generally preceded by shaping circuits and limiting circuits (both discussed in this chapter) to ensure uniformity of amplitude and pulse width. Under those conditions, the pulse repetition frequency is the only variable and frequency variations may be measured.

Q24. Name a common application of counting circuits.

Positive Counters

The POSITIVE-DIODE COUNTER circuit is used in timing or counting circuits in which the number of input pulses are represented by the output voltage. The output may indicate frequency, count the rpm of a shaft, or register a number of operations. The counter establishes a direct relationship between the input frequency and the average dc output voltage. As the input frequency increases, the output voltage also increases; conversely, as the input frequency decreases, the output voltage decreases. In effect, the positive counter counts the number of positive input pulses by producing an average dc output voltage proportional to the repetition frequency of the input signal. For accurate counting, the pulse repetition frequency must be the only variable parameter in the input signal. Therefore, careful shaping and limiting of the input signal is essential for you to ensure that the pulses are of uniform width and that the amplitude is constant. When properly filtered and smoothed, the dc output voltage of the counter may be used to operate a direct reading indicator.

Solid-state and electron-tube counters operate in manners similar to each other. The basic solid-state (diode) counter circuit is shown in view (A) of figure 4-43. Capacitor C1 is the input coupling capacitor. Resistor R1 is the load resistor across which the output voltage is developed. For the purpose of circuit discussion, assume that the input pulses (shown in view (B)) are of constant amplitude and time duration and that only the pulse repetition frequency changes. At time T0, the positive-going input pulse is applied to C1 and causes the anode of D2 to become positive. D2 conducts and current i_c flows through R1 and D2 to charge C1. Current i_c, develops an output voltage across R1, shown as e_{out}.

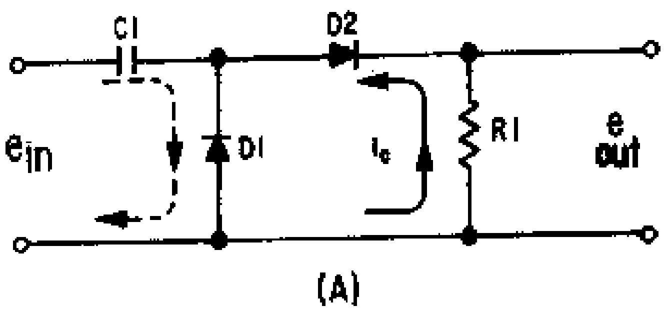


Figure 4-43A.—Positive-diode counter and waveform.

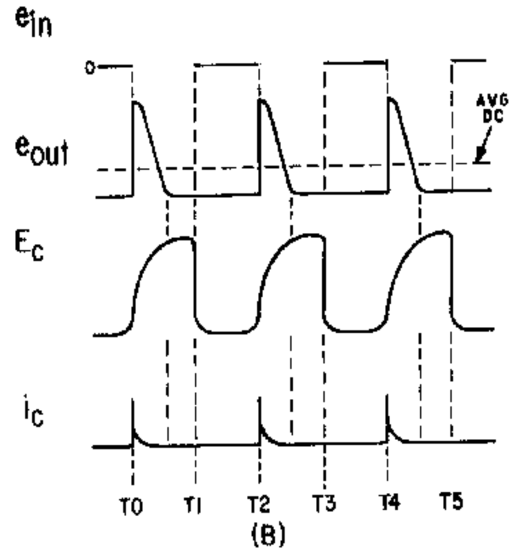


Figure 4-43B.—Positive-diode counter and waveform.

The initial heavy flow of current produces a large voltage across R1 which tapers off exponentially as C1 charges. The charge on C1 is determined by the time constant of R1 and the conducting resistance of the diode times the capacitance of C1. For ease of explanation, assume that C1 is charged to the peak value before T1.

At T1 the input signal reverses polarity and becomes negative-going. Although the charge on capacitor C1 cannot change instantly, the applied negative voltage is equal to or greater than the charge on C1. This causes the anode of D2 to become negative and conduction ceases. When D2 stops conducting e_{out} is at 0. C1 quickly discharges through D1 since its cathode is now negative with respect to ground. Between T1 and T2 the input pulse is again at the 0-volt level and D2 remains in a nonconducting state. Since the very short time constant provided by the conduction resistance of D1 and C1 is so much less than the long time constant offered by D2 and R1 during the conduction period, C1 is always completely discharged between pulses. Thus, for each input pulse, a precise level of charge is deposited on C1. For each charge of C1 an identical output pulse is produced by the flow of ic through R1. Since this current flow always occurs in the direction indicated by the solid arrow, the dc output voltage is positive.

At T2 the input signal again becomes positive and the cycle repeats. The time duration between pulses is the interval represented by the period between T1 and T2 or between T3 and T4. If the input-pulse frequency is reduced, these time periods become longer. On the other hand, if the frequency is increased, these time intervals become shorter. With shorter periods, more pulses occur in a given length of time and a higher average dc output voltage is produced; with longer periods, fewer pulses occur and a lower average dc output voltage is produced. Thus, the dc output is directly proportional to the repetition frequency of the input pulses. If the current and voltage are sufficiently large, a direct-reading meter can be used to indicate the count. If they are not large enough to actuate a meter directly, a dc amplifier may be added. In the latter case, a pi-type filter network is inserted at the output of R1 to absorb the instantaneous pulse variations and produce a smooth direct current for amplification.

From the preceding discussion, you should see that the voltage across the output varies in direct proportion to the input pulse repetition rate. Hence, if the repetition rate of the incoming pulses increases, the voltage across R1 also increases. For the circuit to function as a frequency counter, some method must be employed to use this frequency-to-voltage relationship to operate an indicator. The block diagram in view (A) of figure 4-44 represents one simple circuit which may be used to perform this function. In this circuit, the basic counter is fed into a low-pass filter and an amplifier with a meter that is calibrated in units of frequency.

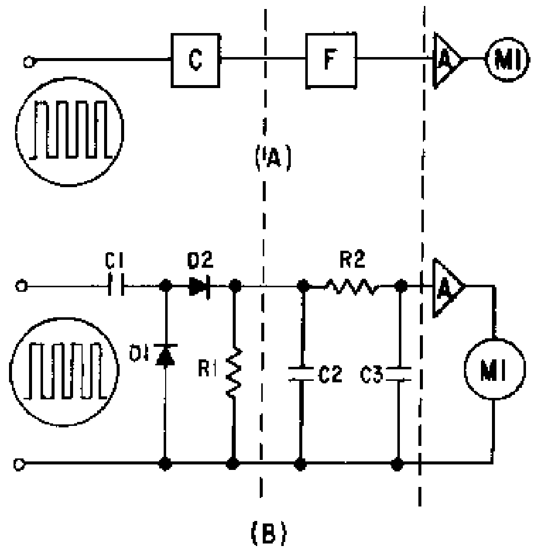


Figure 4-44.—Basic frequency counter.

A typical schematic diagram is shown in view (B). The positive pulses from the counter are filtered by C2, R2, and C3. The positive dc voltage from the filter is applied to the input of amplifier A. This voltage increases with frequency; as a consequence, the current through the device increases. Since emitter or cathode current flows through M1, an increase in amplifier current causes an increase in meter deflection. The meter may be calibrated in units of time, frequency, revolutions per minute, or any function based upon the relationship of output voltage to input frequency.

Q25. What establishes the value of the current that flows in the output of figure 4-43?

Q26. What is the purpose of D1 in figure 4-43?

Negative Counters

Reversing the connections of diodes D1 and D2 in the positive-counter circuit (view (A) of figure 4-43) will cause the circuit to respond to negative pulses and become a negative-counter circuit. Diode D2 conducts during the time the negative pulse is applied and current flows in the opposite direction through R1, as was indicated by the arrow. At the end of the negative pulse, D1 conducts and discharges C1. The current through R1 increases with an increase in pulse frequency as before. However, if the voltage developed across R1 is applied to the same control circuit, as shown in view (A) of figure 4-44, the increase in current will be in a negative direction and the amplifier will conduct less. Thus, the effect is opposite to that of the positive counter.

Step-by-Step (Step) Counters

The STEP-BY-STEP (STEP) COUNTER is used as a voltage multiplier when a stepped voltage must be provided to any device which requires such an input. The step counter provides an output which increases in one-step increments for each cycle of the input. At some predetermined level, the output voltage reaches a point which causes a circuit, such as a blocking oscillator, to be triggered.

A schematic diagram of a positive step counter is shown in view (A) of figure 4-45. For step counting, the load resistor of the positive-counting circuit is replaced by capacitor C2. This capacitor is relatively large in comparison to C1. Each time D2 conducts, the charge on C2 increases as shown in view (B). The steps are not the same height each time. They decrease exponentially with time as the voltage across C2 approaches the input voltage.

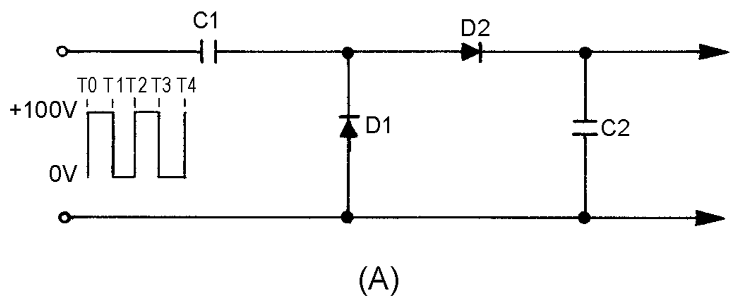


Figure 4-45A.—Basic step counter and waveforms.

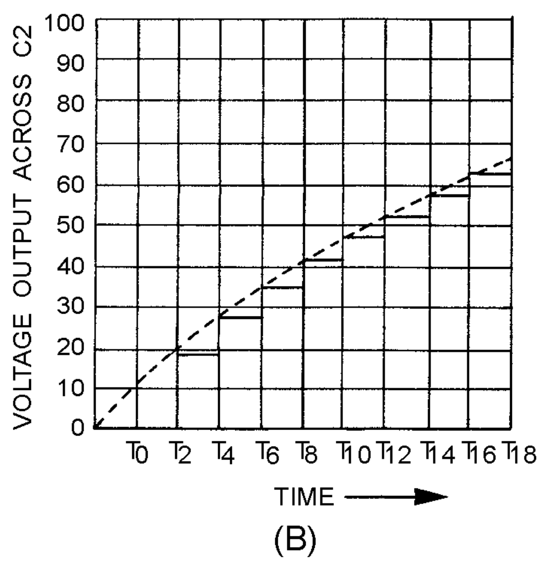


Figure 4-45B.—Basic step counter and waveforms.

As long as C2 has no discharge path, the voltage across its terminals increases with each successive step until it is equal in amplitude to the applied pulse. The voltage across C2 could be applied to a blocking-oscillator circuit to cause the oscillator to pulse after a certain amount of voltage is applied to it.

The circuit in figure 4-46, (view A) and (view B), may be used as a frequency divider. When used in this manner, Q1 is used as a single-swing blocking oscillator that is triggered when the voltage across C2 becomes great enough to forward bias Q1. At other times, the transistor is cut off by the bias voltage developed in the section of R2 that is between the ground and the slide.

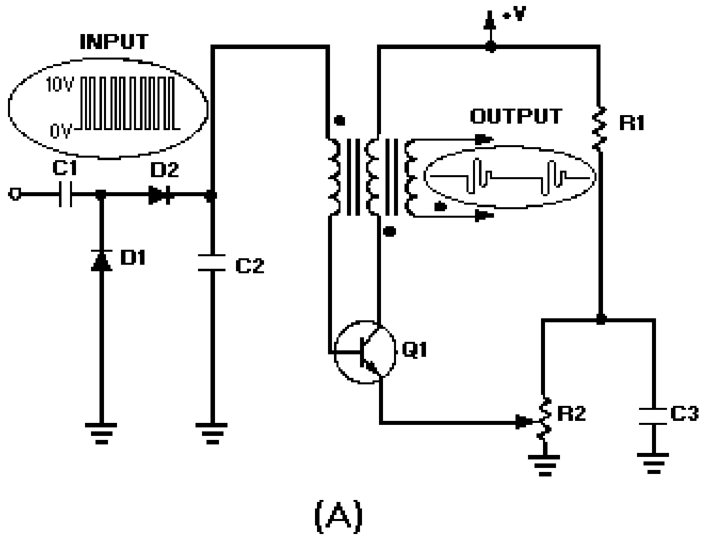


Figure 4-46A.—Step counter as a frequency divider and waveforms.

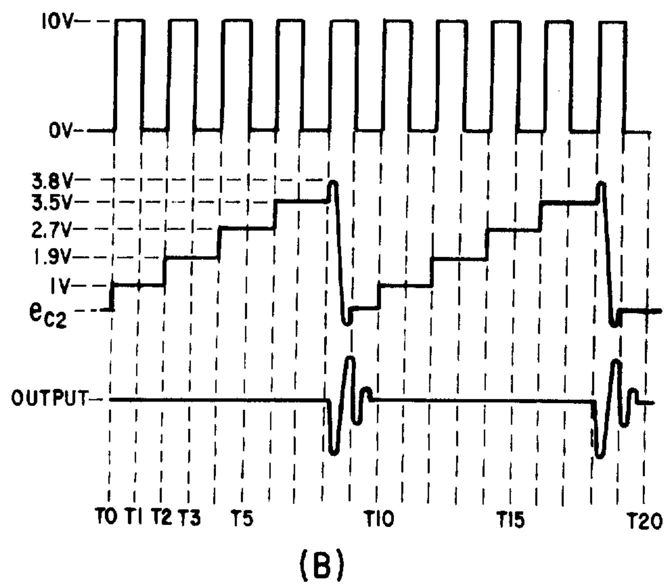


Figure 4-46B.—Step counter as a frequency divider and waveforms.

The action of the counter can best be understood by referring back to figure 4-45. Assume C2 is 10 times larger than C1 and the peak voltage is 10 volts. C1 will assume 9/10 of the positive input voltage at T0, while C2 will assume only 1/10, or 1 volt in this example. At T1 the input will drop in a negative direction and D2 will be cut off. The cathode of D1 will become more negative than its anode and conduct, discharging C1. The charge on C2 will remain at 1 volt because it has no discharge path. At T2 the second pulse will be applied. The 1-volt charge on C2 will oppose the 10 volts of the second pulse, and the applied voltage for the capacitors to charge will be 9 volts. C2 will again charge 10 percent, or 0.9 volt. This is in addition to the initial charge of volt. At the end of the second pulse, the voltage on C2 will be 1.9 volts. At T3 the third pulse will be 10 volts, but 1.9 volts will oppose it. Therefore, the applied voltage will be 10 – 1.9 volts, or 8.1 volts. C2 will charge to 10 percent of 8.1 volts, or .81 volt. The voltage on C2 will become 1 + .9 + .81, or 2.71 volts. Successive input pulses will raise C2 by 10 percent of the remaining voltage toward 10 volts until the blocking oscillator works. If the oscillator bias is set so that Q1 begins conduction at 3.8 volts, this will continue until 3.8 volts is exceeded. Since the fourth step is 3.5 volts and the fifth is 4.1 volts, the 3.8-volt level is crossed at the fifth step. If the oscillator goes through 1 cycle of operation every fifth step and C2 is discharged at this point, this circuit would be a 5-to-1 divider.

The circuit can be made to divide by 3, 4, or some other value by setting the bias at a different level. For example, if the bias is set at 2.9 volts, conduction will occur at the fourth step, making it a 4-to-1 divider.

The counting stability of the step counter is dependent upon the exponential charging rate of capacitor C2. As C2 increases to higher steps, the voltage increments are less and less. If the ratio becomes too great, the higher steps become almost indiscernible. For this reason, accuracy decreases as the ratio increases. When you desire to count by a large number, 24 for example, a 6-to-1 counter and a 4-to-1 counter are connected in cascade (series). A more stable method of counting 24 would be to use a 2:1, 3:1, 4:1 counter connected in cascade. Most step counters operate on a ratio of 5 to 1 or less.

Q27. What is the difference between a positive counter and a step counter?

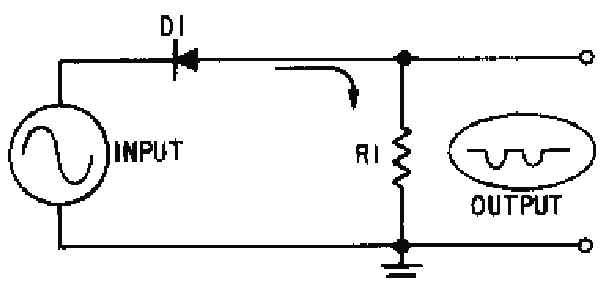
SUMMARY

This chapter has presented information on wave shaping. The information that follows summarizes the important points of this chapter.

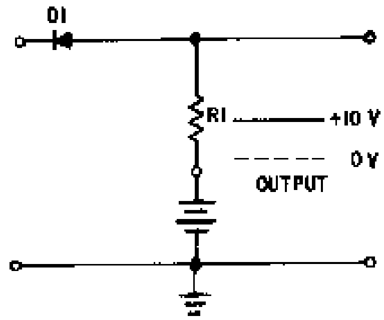
A **LIMITER** is a device which limits or prevents some part of a waveform from exceeding a specified value.

In a **SERIES LIMITER**, the diode is in series with the output. It can limit either the negative or positive alternation of the input signal.

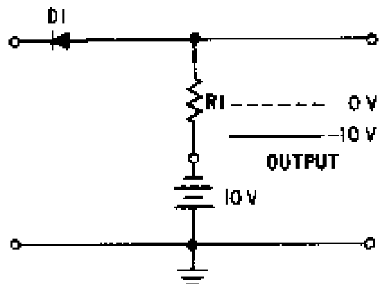
In a **SERIES-POSITIVE LIMITER**, the diode is in series with the output which is taken across the resistor. It removes the positive alternation of the input signal.



In a **SERIES-POSITIVE LIMITER WITH BIAS,** the bias potential will either aid or oppose the flow of current. When aiding forward bias, only a portion of the positive input pulse is removed. When the bias aids the reverse bias, all of the positive and a portion of the negative pulse is removed.

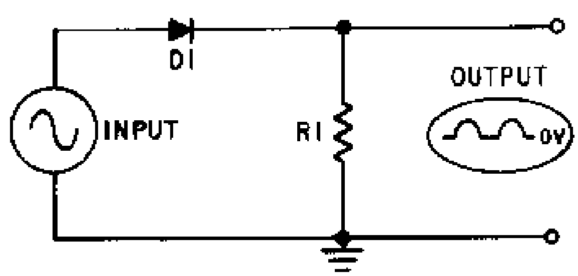


POSITIVE BIAS

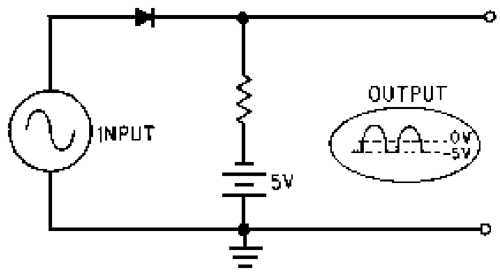


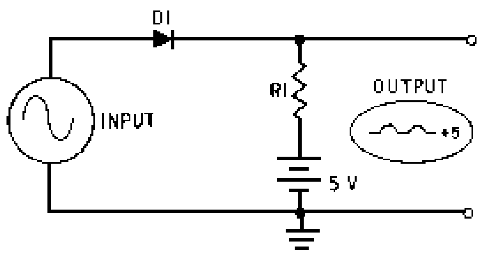
NEGATIVE BIAS

The **SERIES-NEGATIVE LIMITER** limits the negative portion of the input pulse. The difference between a series-negative limiter and a series-positive limiter is that the diode is reversed in the negative limiter.



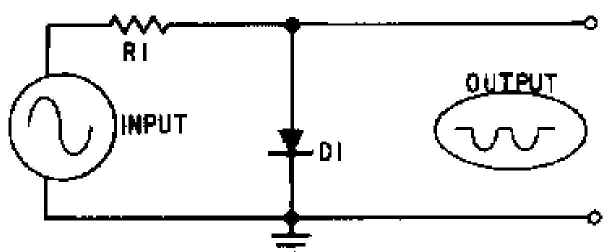
A **SERIES-NEGATIVE LIMITER** with bias is the same as the series-positive limiter with bias, but the outputs are opposite. When bias aids forward bias, only a portion of the negative input is removed. When bias aids reverse bias, all of the negative and a portion of the positive input is removed.



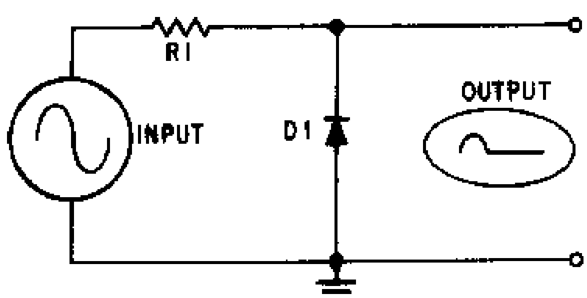


In a **PARALLEL LIMITER**, a resistor and diode are connected in series with the input signal. The output is taken across the diode.

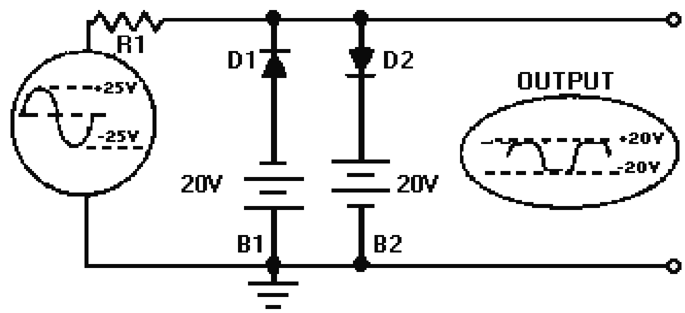
In the **PARALLEL-POSITIVE LIMITER**, the positive portion of the input signal is limited when the diode conducts.



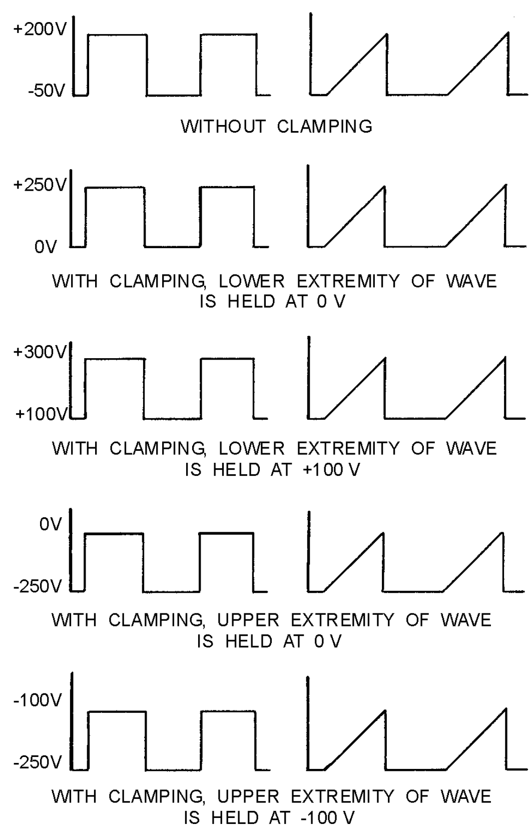
The **PARALLEL-NEGATIVE LIMITER** diode is reversed from that of the parallel positive limiter to limit only a portion of the negative input signal.



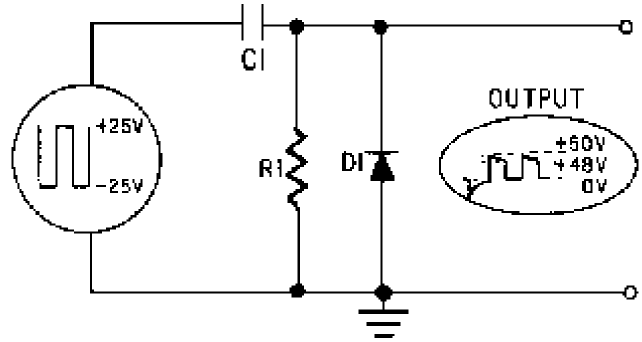
The **DUAL-DIODE LIMITER** combines the parallel negative limiter with negative bias (reverse bias) and the parallel positive limiter with positive bias (reverse bias). It will remove parts of the positive and negative input signal.



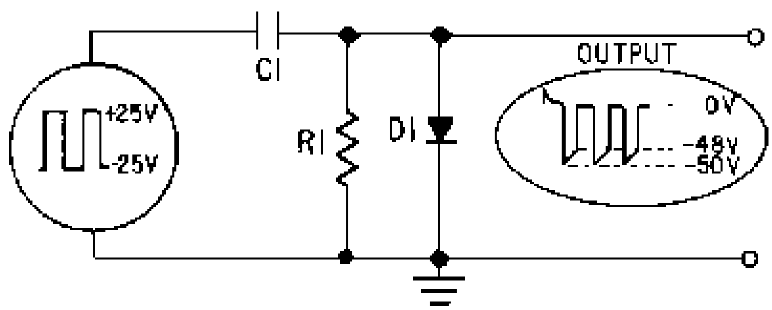
A **CLAMPING CIRCUIT** effectively clamps or ties down the upper or lower extremity of a waveform to a fixed dc potential. Clamping does not change the amplitude or shape of the input waveform.



A **POSITIVE CLAMPER** will clamp the lower extremity of the input waveform to a dc potential of 0 volts.



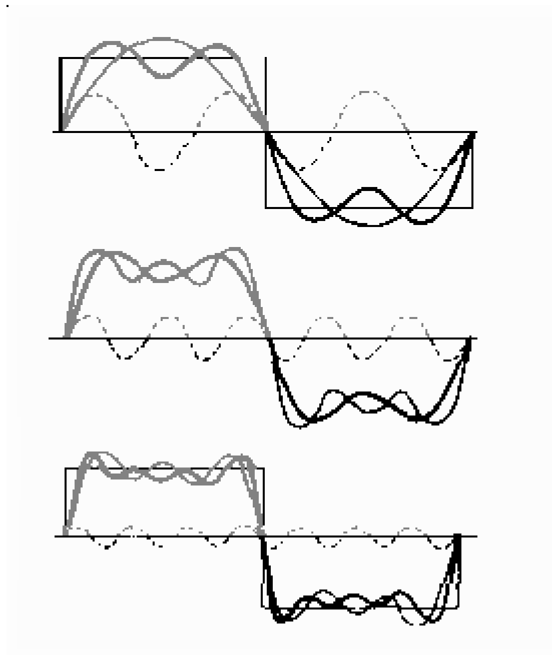
A **NEGATIVE CLAMPER** will clamp the upper extremity of the input waveform to a dc potential of 0 volts.



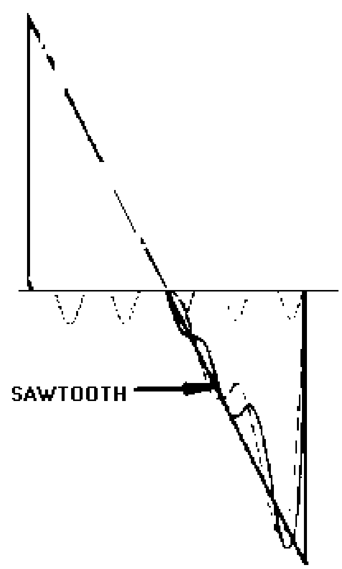
A **COMMON-BASE TRANSISTOR CLAMPER** clamps the collector voltage to a reference level.

A waveform other than a sine wave is called a COMPLEX WAVE.

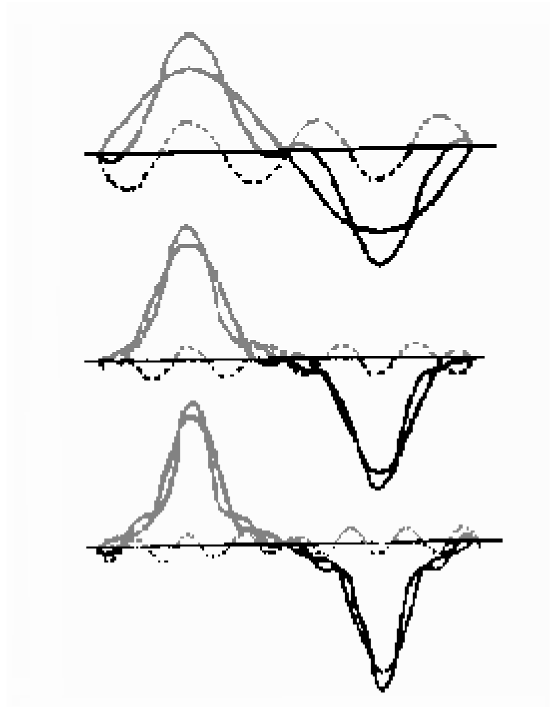
If the odd harmonics of a sine wave are added algebraically, the result is a square wave. A PERFECT SQUARE WAVE is composed of an infinite number of odd harmonics in phase with the fundamental wave.



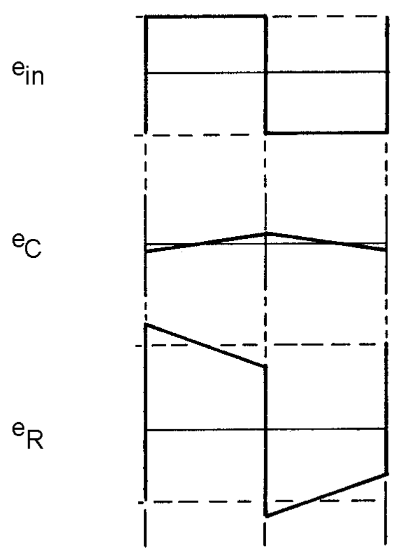
A **SAWTOOTH WAVE** is made up of different harmonics, both odd and even.



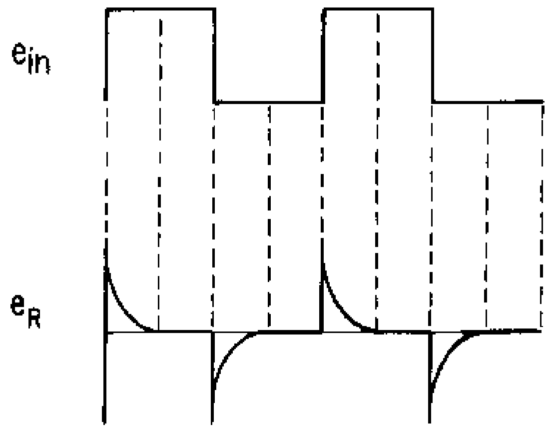
A **PEAKED WAVE** is made up of odd harmonics that are in phase and out of phase with the fundamental.



INTEGRATION takes place in an RC circuit with the output taken across the capacitor. The amount of integration is dependent upon the time constant of the circuit. Full integration takes place when the time constant of the RC circuit is at least 10 times greater than the duration of the input pulse. An RL circuit is also used as an integration circuit. The output is taken across the resistor and the time constant of the circuit is 10 times greater than the input pulse.



DIFFERENTIATION is the opposite of integration. In the differentiator, the output is taken across the resistor. Full differentiation takes place when the time constant of the circuit is 1/10 that of the input pulse.



A **COUNTING CIRCUIT** receives uniform pulses, representing units to be counted, and produces a voltage output proportional to its frequency.

ANSWERS TO QUESTION Q1. THROUGH Q27.

A1. Negative.

A2. Positive.

A3. Biasing.

A4. The diode.

A5. Conducting, cutoff.

A6. Short time constant.

A7. Long time constant.

A8. Most negative.

A9. Positive potential.

A10. Positive clamper with negative bias.

A11. Most positive.

A12. Negative potential.

A13. Positive bias.

A14. -5 volts.

A15. It is composed of an infinite number of odd harmonics in phase with the fundamental.

A16. It is composed of odd harmonics some of which are out phase with the fundamental.

A17. All the odd harmonics are in phase with the fundamental in the square wave. This is not true of the odd harmonics in the peaked wave.

A18. The time constant is long and the output is taken across the capacitor in an RC circuit.

A19. A pure sine wave cannot be integrated; it contains no harmonics.

A20. The ability of the inductor to oppose a change in current.

A21. The time-constant value of a long time constant-circuit is 10 times the value of the input pulse duration. The short time-constant circuit has a time constant of 1/10 of the pulse duration.

A22. A more complete integration of the waveform would result from the long time constant.

A23. In an RC circuit the output is taken across the resistor. In the RL circuit the output is taken across the inductor.

A24. Frequency counters or frequency dividers.

A25. The frequency of the voltage input.

A26. To provide a quick discharge path for C1.

A27. The load resistor in a positive counter is replaced by a capacitor in a step counter.

APPENDIX I

GLOSSARY

AMPLITUDE STABILITY—Amplitude stability refers to the ability of the oscillator to maintain a constant amplitude in the output waveform.

ASTABLE MULTIVIBRATOR—A multivibrator which has no stable state. Also called free-running because it alternates between two different output voltage levels during the time it is on. The frequency is determined by the RC time constant of the coupling circuit.

ATTENUATION—The ability of a filter circuit to reduce the amplitude of unwanted frequencies to a level below that of the desired output frequency.

BANDPASS FILTER—A filter which allows a narrow band of frequencies to pass through the circuit. Rejects or attenuates frequencies which are either higher or lower than the desired band of frequencies.

BAND-REJECT FILTER—Rejects the passage of current for a small band of frequencies. Allows current to flow at frequencies either above or below this band.

BANDWIDTH—The range of frequencies included between upper and lower frequencies.

BISTABLE MULTIVIBRATOR—A multivibrator that has two stable states. It remains in one of the states until a trigger is applied. It then flips to the other stable state and remains there until another trigger is applied. Also referred to as a flip-flop.

BUFFER AMPLIFIER—An amplifier which isolates one circuit from another. It decreases the loading effect on an oscillator by reducing the interaction between the load and the oscillator.

CAPACITIVE REACTANCE—The opposition, expressed in ohms, offered to the flow of an alternating current by capacitance. The symbol for capacitive reactance is Xc.

CLAMPER—A circuit in which either the upper or lower extremity of a waveform is fixed at a desired value.

COMPLEX WAVE—A waveform other than a sine wave.

COUNTER—A circuit which counts input pulses.

CRYSTAL OVEN—Closed oven maintained at a constant temperature in which a crystal and its holder are enclosed to reduce frequency drift.

DAMPED WAVE—A sinusoidal wave in which the amplitude steadily decreases with time. Often associated with energy loss.

FILTER CIRCUIT—Network of resistors, inductors, and/or capacitors which offers opposition to certain frequencies.

FLYWHEEL EFFECT—The ability of a resonant circuit to operate continuously because of stored energy or energy pulses.

FREQUENCY CUTOFF—The frequency at which the filter circuit changes from an action of rejecting the unwanted frequencies to an action of passing the desired frequencies. Conversely, the point at which the filter circuit changes from an action in which it passes the desired frequencies to an action in which it rejects the undesired frequencies.

FREQUENCY STABILITY—Refers to the ability of an oscillator to accurately maintain its operating frequency.

HALF-POWER POINT—Point on either side of resonance curve at which the power is approximately 70 percent of the maximum value.

HARMONIC—Integral multiples of a fundamental frequency. For example, the harmonics of 60 hertz are 120 hertz, 180 hertz, 240 hertz, and so forth.

HIGH-PASS FILTER—A filter that passes a majority of the high frequencies on to the next circuit and rejects, or attenuates, the lower frequencies. Also called a low-frequency discriminator.

INDUCTIVE REACTANCE—The opposition to the flow of an alternating current (expressed in ohms) caused by the inductance of a circuit. The symbol for inductive reactance is X_L.

IMPEDANCE—Total opposition to alternating current flow. Impedance may consist of any combination of resistance, inductive reactance, and capacitive reactance. The symbol for impedance is Z.

LIMITER—A device which prevents (limits) a waveform from exceeding a specified value.

LOWER-FREQUENCY CUTOFF—The lowest frequency a circuit will pass.

LOW-PASS FILTER—A filter that passes a majority of the low frequencies on to the next circuit and rejects, or attenuates, the higher frequencies. Also called a high-frequency discriminator.

MULTIVIBRATOR—A form of relaxation oscillator which comprises two stages that are coupled so that the input of one is derived from the output of the other.

MONOSTABLE MULTIVIBRATOR—A multivibrator which has one steady state. A signal (trigger) must be applied to cause change of states.

NATURAL FREQUENCY—See Resonance Frequency.

NEGATIVE CLAMPER—The upper extremity of the output waveshape is clamped to a dc potential of 0 volts.

OSCILLATOR—An oscillator is a nonrotating device which produces alternating current. The frequency is determined by the characteristics of the device.

PARALLEL LIMITER—A resistor and diode connected in series with the input signal. The output is taken across the diode.

PARALLEL-NEGATIVE LIMITER—A resistor and diode connected in series with the input signal. The output is taken across the diode and the negative alternation is eliminated.

PARALLEL-POSITIVE LIMITER—A resistor and diode connected in series with the input signal. The output is taken across the diode and the positive alternation of the input signal is eliminated.

PARALLEL-RESONANT CIRCUIT—A resonant circuit in which the source voltage is connected across a parallel circuit (formed by a capacitor and an inductor) to furnish a high impedance to the frequency at which the circuit is resonant. Often referred to as a tank circuit.

PERIODIC WAVE—A waveform that undergoes a pattern of changes, returns to its original pattern, and then repeats the same pattern of changes. Examples are square waves, rectangular waves, and sawtooth waves.

POSITIVE CLAMPER—The lower extremity of the output waveshape is clamped to a dc potential of 0 volts.

PULSE—Signal characterized by a rapid rise and decay from an initial level.

PULSE OSCILLATOR—A sine-wave oscillator that is turned on and off at specific times. Also known as a ringing oscillator.

PULSE-REPETITION FREQUENCY (PRF)—The number of times in 1 second that a waveform repeats itself.

PULSE-REPETITION RATE (PRR)—Same as Pulse-Repetition Frequency (prf).

Q—Figure of merit (efficiency) of a circuit or coil. Ratio of inductive reactance to resistance.

QUIESCENT STATE—Time during which a tube or transistor of an electrical circuit is not performing its active function.

RC CONSTANT—Time constant of a resistor-capacitor circuit. Equal in seconds to the resistance value multiplied by capacitance value.

RC DIFFERENTIATOR—An RC circuit in which the output is taken from the resistor.

RC INTEGRATOR—An RC circuit in which the output is taken from the capacitor.

RC NETWORK—A circuit containing resistances and capacitances arranged in a particular manner to perform a specific function.

RC OSCILLATOR—An oscillator in which the frequency is determined by resistive and capacitive elements.

REGENERATIVE FEEDBACK—The process by which a portion of the output power of an amplifying device is fed back to reinforce the input.

RESONANCE—The condition in a circuit containing inductance and capacitance in which the inductive reactance is equal and opposite to the capacitive reactance. This condition occurs at only one frequency and the circuit in that condition is said to be in resonance. The resonant frequency can be changed by varying the values of either the capacitance or inductance.

RESONANT CIRCUIT—A circuit that contains both inductance and capacitance and is resonant at one frequency ($X_L = X_C$).

RESONANT FREQUENCY—That frequency in a given resonant circuit at which the inductive and capacitive values are equal and cancel each other.

RL DIFFERENTIATOR—An RL circuit in which the output is taken from the inductor.

RL INTEGRATOR—An RL circuit in which the output is taken from the resistor.

SELECTIVITY—The ability of a circuit to discriminate between frequencies.

SERIES-FED OSCILLATOR—An oscillator in which dc power is supplied to the amplifier through the tank circuit or a portion of the tank circuit.

SERIES LIMITER—The diode is connected in series with the output and the output is taken across the resistor. Either the positive or negative alternation of the input wave is eliminated.

SERIES-NEGATIVE LIMITER—The diode is connected in series with the output and the output is taken across the resistor. Eliminates the negative alternation of the input wave.

SERIES-PARALLEL CIRCUIT—A circuit in which two or more parallel or series combinations are in series with each other.

SERIES-POSITIVE LIMITER—The diode is connected in series with the output and the output is taken across a resistor. Eliminates the positive alternation of the input wave.

SERIES-RESONANT CIRCUIT—A resonant circuit in which the source voltage is connected in series with a capacitor and an inductor (also in series) to furnish a low impedance at the frequency at which the circuit is resonant.

SHAPING CIRCUIT—A circuit which alters the shapes of input waveforms.

STEP-BY-STEP-COUNTER—A counter which provides an output for each cycle of the input in one-step increments.

SHUNT-FED OSCILLATOR—An oscillator which receives its dc power for the transistor or tube through a path both separate from and parallel to the tank circuit.

TANK CIRCUIT—A tuned circuit used to temporarily store energy. Also referred to as a parallel-resonant circuit.

TICKLER COIL—Small coil connected in series with the collector or plate circuit of a transistor or tube and inductively coupled to a base or grid-circuit coil to establish feedback or regeneration.

TIME CONSTANT—Time required for an exponential quantity to change by an amount equal to .632 times the total change that can occur.

TRIGGER—Short pulse, either positive or negative, which can be used to cause an electrical function to take place.

TUNED CIRCUIT—Circuit consisting of inductance and capacitance which can be adjusted for resonance at a desired frequency.

UPPER-FREQUENCY CUTOFF—The highest frequency a circuit can pass.

APPENDIX II

SQUARE AND SQUARE ROOTS

N	N^2	$\sqrt{N}$	N	N^2	$\sqrt{N}$	N	N^2	$\sqrt{N}$
1	1	1.000	41	1681	6.4031	81	6561	9.0000
2	4	1.414	42	1764	6.4807	82	6724	9.0554
3	9	1.732	43	1849	6.5574	83	6889	9.1104
4	16	2.000	44	1936	6.6332	84	7056	9.1652
5	25	2.236	45	2025	6.7082	85	7225	9.2195
6	36	2.449	46	2116	6.7823	86	7396	9.2736
7	49	2.646	47	2209	6.8557	87	7569	9.3274
8	64	2.828	48	2304	6.9282	88	7744	9.3808
9	81	3.000	49	2401	7.0000	89	7921	9.4340
10	100	3.162	50	2500	7.0711	90	8100	9.4868
11	121	3.3166	51	2601	7.1414	91	8281	9.5394
12	144	3.4641	52	2704	7.2111	92	8464	9.5917
13	169	3.6056	53	2809	7.2801	93	8649	9.6437
14	196	3.7417	54	2916	7.3485	94	8836	9.6954
15	225	3.8730	55	3025	7.4162	95	9025	9.7468
16	256	4.0000	56	3136	7.4833	96	9216	9.7980
17	289	4.1231	57	3249	7.5498	97	9409	9.8489
18	324	4.2426	58	3364	7.6158	98	9604	9.8995
19	361	4.3589	59	3481	7.6811	99	9801	9.9499
20	400	4.4721	60	3600	7.7460	100	10000	10.0000
21	441	4.5826	61	3721	7.8102	101	10201	10.0499
22	484	4.6904	62	3844	7.8740	102	10404	10.0995
23	529	4.7958	63	3969	7.9373	103	10609	10.1489
24	576	4.8990	64	4096	8.0000	104	10816	10.1980
25	625	5.0000	65	4225	8.0623	105	11025	10.2470
26	676	5.0990	66	4356	8.1240	106	11236	10.2956
27	729	5.1962	67	4489	8.1854	107	11449	10.3441
28	784	5.2915	68	4624	8.2462	108	11664	10.3923
29	841	5.3852	69	4761	8.3066	109	11881	10.4403
30	900	5.4772	70	4900	8.3666	110	12100	10.4881
31	961	5.5678	71	5041	8.4261	111	12321	10.5357
32	1024	5.6569	72	5184	8.4853	112	12544	10.5830
33	1089	5.7447	73	5329	8.5440	113	12769	10.6301
34	1156	5.8310	74	5476	8.6023	114	12996	10.6771
35	1225	5.9161	75	5625	8.6603	115	13225	10.7238
36	1296	6.0000	76	5776	8.7178	116	13456	10.7703
37	1369	6.0828	77	5929	8.7750	117	13689	10.8167
38	1444	6.1644	78	6084	8.8318	118	13924	10.8628
39	1521	6.2450	79	6241	8.8882	119	14161	10.9087
40	1600	6.3246	80	6400	8.9443	120	14400	10.9545

For numbers up to 120. For larger numbers divide into factors smaller than 120.

Examples: $\sqrt{225}$ and $\sqrt{16200}$

$225 = 5 \times 45$

$\sqrt{225} = \sqrt{5} \times \sqrt{45}$

$\sqrt{225} = 2.236 \times 6.7082$

$\sqrt{225} = 15$

$16200 = 100 \times 81 \times 2$

$\sqrt{16200} = \sqrt{100} \times \sqrt{81} \times \sqrt{2}$

$\sqrt{16200} = 10 \times 9 \times 1.414$

$\sqrt{16200} = 127.26$

MODULE 9 INDEX

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Tuned Circuits," pages 1-1 through 1-52.

1-1. An inductor presents which of the following types of electrical opposition to ac current flow?

1. Reactance
2. Resistance
3. Inductance
4. Capacitance

1-2. Which of the following electrical characteristics determines the magnitude of inductive reactance?

1. Resistance
2. Frequency only
3. Inductance only
4. Frequency and inductance

1-3. Which of the following values represents an inductive reactance?

1. X_C = 2,220 ohms
2. X_L = 220 ohms
3. L = 22 millihenries
4. C = 22 microfarads

1-4. What formula is used to calculate inductive reactance?

1. $X_L = 2\pi fC$
2. $X_L = 2\pi fL$
3. $X_L = \frac{1}{2\pi fL}$
4. $X_L = \frac{1}{2\pi fC}$

1-5. In an ac circuit, how does inductive reactance respond to an increase in applied frequency?

1. Inductive reactance increases
2. Inductive reactance decreases
3. Inductive reactance remains the same

1-6. What term describes the opposition to ac that causes current to lead voltage?

1. Resistance
2. Conductance
3. Inductive reactance
4. Capacitive reactance

1-7. In an ac circuit, how does capacitive reactance respond to an increase in applied frequency?

1. Capacitive reactance increases
2. Capacitive reactance decreases
3. Capacitive reactance remains the same

1-8. In an ac circuit, what is the term that describes the TOTAL opposition to current flow?

1. Impedance
2. Inductance
3. Resistance
4. Capacitance

1-9. In an ac circuit that contains an inductive reactance of 7,250 ohms and a capacitive reactance of 9,775 ohms, what is the resultant reactance?

1. - 2,525 ohms
2. -10,250 ohms
3. 2,525 ohms
4. 10,250 ohms

1-10. When an ac circuit is at resonance, what is the relationship between X_L and X_C

1. X_L is equal to X_C
2. X_L is less than X_C
3. X_L is greater than X_C

1-11. What formula is used to calculate resonant frequency?

1. $f_r = \frac{1}{2\pi fL}$
2. $f_r = \frac{1}{2\pi fC}$
3. $f_r = \frac{1}{2\pi\sqrt{LR}}$
4. $f_r = \frac{1}{2\pi\sqrt{LC}}$

1-12. In a tank circuit, how does the resonant frequency of the circuit respond to an increase in (a) capacitance and (b) inductance?

1. (a) Increases (b) increases
2. (a) Increases (b) decreases
3. (a) Decreases (b) decreases
4. (a) Decreases (b) increases

1-13. In a resonant circuit, what is the phase angle between voltage and current?

1. 0 degrees
2. 90 degrees
3. 180 degrees
4. 270 degrees

1-14. In a resonant circuit, how does resistance change, if at all, in response to an increase in frequency?

1. Increases
2. Decreases
3. Remains the same

1-15. In a series-LC circuit, which of the following component characteristics describes circuit action (a) below the resonant frequency and (b) above the resonant frequency?

1. (a) Inductive (b) Capacitive
2. (a) Inductive (b) Resistive
3. (a) Capacitive (b) Inductive
4. (a) Capacitive (b) Resistive

THIS SPACE LEFT BLANK INTENTIONALLY.

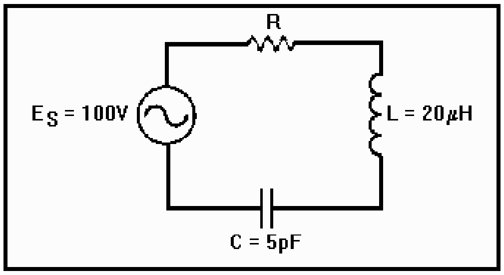


Figure 1A.—Series-resonant circuit.

IN ANSWERING QUESTIONS 1-16 THROUGH 1-21, REFER TO FIGURE 1A.

1-16. What is the resonant frequency for the circuit?

1. 1.592 MHz
2. 15.92 MHz
3. 159.2 MHz
4. 1,592 MHz

1-17. What is the value of inductive reactance?

1. 1.97 ohms
2. 97 ohms
3. 199.7 ohms
4. 1,997 ohms

1-18. If the resonant frequency is 7.96 MHz, what is the value of capacitive reactance?

1. 500 ohms
2. 1,000 ohms
3. 2,000 ohms
4. 4,000 ohms

IN ANSWERING QUESTIONS 1-19 THROUGH 1-21, ASSUME THE SOURCE FREQUENCY IN FIGURE 1A IS ABOVE THE RESONANT FREQUENCY. SELECT THE ANSWERS THAT DESCRIBE HOW AN ABOVE-RESONANCE FREQUENCY WILL CAUSE THE CIRCUIT CHARACTERISTICS IN THE QUESTIONS TO RESPOND WHEN COMPARED TO THEIR VALUES AT RESONANCE.

1-19. Impedance.

1. Increases
2. Decreases
3. Remains the same

1-20. Current.

1. Increases
2. Decreases
3. Remains the same

1-21. Voltage drops across the reactances.

1. Increases
2. Decreases
3. Remains the same

THIS SPACE LEFT BLANK INTENTIONALLY.

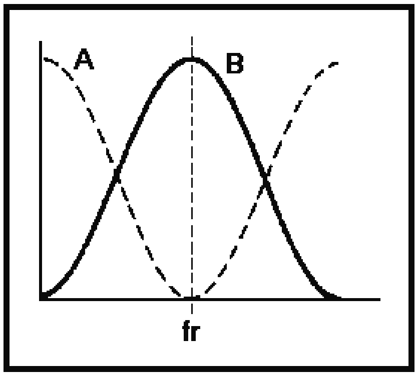


Figure 1B.—Series-resonant circuit curves.

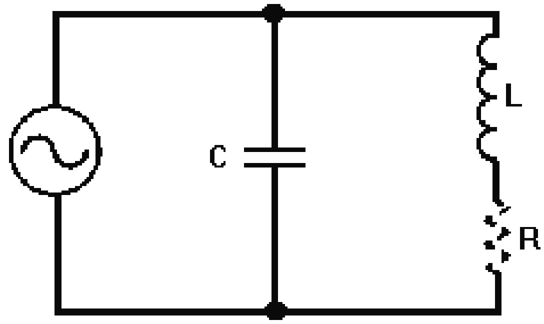


Figure 1C.—Parallel-resonant circuit.

IN ANSWERING QUESTIONS 1-22 AND 1-23, REFER TO FIGURE 1B.

1-22. Response curve B for a series-resonant circuit represents which of the following circuit characteristics?

1. Power
2. Voltage
3. Current
4. Impedance

1-23. At resonance, which of the following series-resonant circuit values is at a maximum value?

1. Circuit current
2. Voltage across L
3. Voltage across C
4. Circuit impedance

1-24. In a series-resonant circuit operating at f_r, what term describes the impedance of the circuit?

1. Resistive
2. Inductive only
3. Capacitive only
4. Capacitive-inductive

IN ANSWERING QUESTIONS 1-25 THROUGH 1-27, REFER TO FIGURE 1C.

1-25. In the parallel-resonant circuit, what is the phase relationship between the current in the inductor and the current in the capacitor?

1. Inductor current is in phase with capacitor current
2. Inductor current is 45 degrees out of phase with capacitor current
3. Inductor current is 90 degrees out of phase with capacitor current
4. Inductor current is 180 degrees out of phase with capacitor current

1-26. In the parallel-resonant circuit, what is the phase relationship between voltage in the inductor and the voltage in the capacitor.

1. Inductor voltage is in phase with capacitor voltage
2. Inductor voltage is 45 degrees out of phase with capacitor voltage
3. Inductor voltage is 90 degrees out of phase with capacitor voltage
4. Inductor voltage is 180 degrees out of phase with capacitor voltage

1-27. In the parallel-resonant circuit, which of the following circuit conditions is NOT normal?

1. X_C equals X_L
2. I_C equals I_L
3. I_{line} is minimum
4. I_{line} is maximum

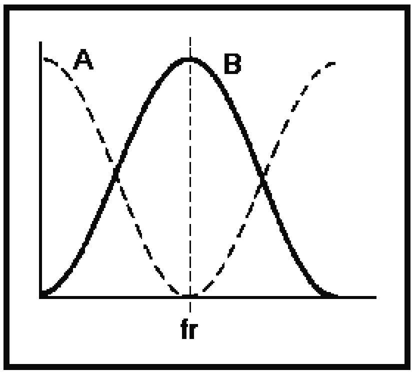


Figure 1D.—Parallel-resonant circuit curves.

IN ANSWERING QUESTIONS 1-28 AND 1-29, REFER TO FIGURE 1D.

1-28. In the figure, what does response curve A represent?

1. Current
2. Impedance
3. Reactance
4. Resistance

1-29. What does response curve B represent?

1. Power
2. Impedance
3. Reactance
4. Resistance

1-30. As a parallel-resonant circuit approaches resonance, which of the following circuit actions takes place?

1. Impedance decreases
2. Oscillating current increases
3. Inductance increases
4. Capacitance decreases

1-31. When a parallel-resonant circuit operates BELOW resonance, which of the following component characteristics describes circuit action?

1. Inductive
2. Capacitive
3. Resistive

1-32. When a parallel-resonant circuit operates ABOVE resonance, which of the following component characteristics describes circuit actions?

1. Inductive
2. Capacitive
3. Resistive

1-33. In a parallel-resonant circuit, which of the following circuit conditions is observed?

1. Oscillating current is less than line current
2. Oscillating current is greater than line current
3. Line current is maximum
4. Impedance is minimum

1-34. What is the level of impedance offered at resonance in (a) a series-resonant circuit and (b) a parallel-resonant circuit?

1. (a) High (b) high
2. (a) High (b) low
3. (a) Low (b) low
4. (a) Low (b) high

1-35. The ability of a resonant circuit to separate currents of desired frequencies from those of undesired frequencies makes them useful in which of the following circuit applications?

1. Filters
2. Counters
3. Amplifiers
4. Voltage dividers

1-36. The Q of a circuit is a measure of circuit

1. quality
2. permeance
3. conductance
4. inductive reactance

1-37. Which of the following circuit values has the greatest effect on the figure of merit of the circuit?

1. Reactance
2. Inductance
3. Resistance
4. Capacitance

1-38. What formula is used to figure the Q of a coil?

1. $Q = \frac{R}{Z}$
2. $Q = \frac{X_L}{Z}$
3. $Q = \frac{R}{X_L}$
4. $Q = \frac{X_L}{R}$

1-39. On which of the following coil characteristics is the Q of a coil dependent?

1. Size
2. Length
3. Material
4. All of the above

THIS SPACE LEFT BLANK INTENTIONALLY.

1-40. A series-resonant circuit, which of the following conditions results in a voltage gain?

1. $\frac{X_L}{R}$ increasing
2. $\frac{X_L}{R}$ decreasing
3. $\frac{R}{X_L}$ increasing
4. $\frac{R}{X_L}$ decreasing

1-41. In a parallel-resonant circuit, Q is used to figure which of the following circuit values?

1. Voltage gain
2. Voltage loss
3. Circulating tank current
4. Circulating line current

1-42. To determine bandwidth, you would use which of the following mathematical expressions?

1. $BW = \frac{f_r}{Q}$
2. $BW = \frac{f_r \times X_L}{f_r}$
3. $BW = \frac{R \times X_L}{f_r}$
4. $BW = \frac{R}{f_r \times X_L}$

1-43. To calculate (figure) the half-power points of a resonant circuit, which of the following mathematical expressions should you use?

1. $.707 \times I_{min}$
2. $.707 \times I_{max}$
3. $.637 \times I_{min}$
4. $.637 \times I_{max}$

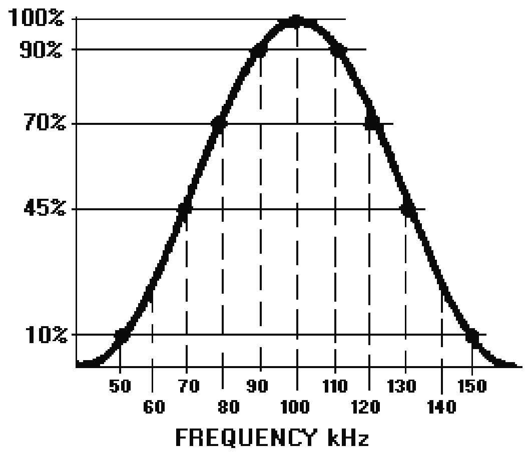


Figure 1E.—Response curve.

IN ANSWERING QUESTIONS 1-44 THROUGH 1-46, REFER TO FIGURE 1E.

1-44. In the response curve, what is the resonant frequency?

1. 50 kHz
2. 70 kHz
3. 100 kHz
4. 140 kHz

1-45. What is the bandwidth?

1. 10 kHz
2. 20 kHz
3. 30 kHz
4. 40 kHz

1-46. If the Q of the circuit represented by the response curve is 100, what is the bandwidth?

1. 1 kHz
2. 10 kHz
3. 20 kHz
4. 30 khz

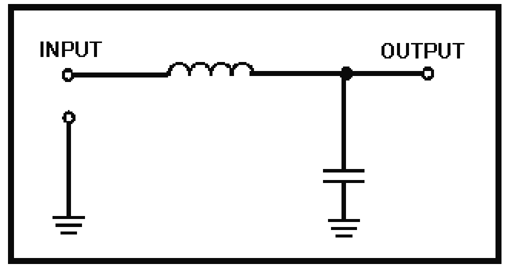


Figure 1F.—Filter circuit.

IN ANSWERING QUESTION 1-47, REFER TO FIGURE 1F.

1-47. If the applied frequency to the circuit is increased, what is the response of (a) X_C and (b) X_L

1. (a) X_C increases (b) X_L increases
2. (a) X_C increases (b) X_L decreases
3. (a) X_C decreases (b) X_L decreases
4. (a) X_C decreases (b) X_L increases

THIS SPACE LEFT BLANK INTENTIONALLY.

TO ANSWER QUESTIONS 1-48 THROUGH 1-50, SELECT FROM COLUMN B THE CIRCUIT WHICH DESCRIBES THE CIRCUIT OPERATION IN COLUMN A. CHOICES IN COLUMN B MAY BE USED ONCE, MORE THAN ONCE, OR NOT AT ALL.

	A. CIRCUIT OPERATION	B. CIRCUIT
1-48.	Passes the majority of current below a specific frequency and opposes current above that frequency.	1. Band-pass filter 2. High-pass filter 3. Low-pass filter 4. Band-reject filter
1-49.	Passes the majority of current above a specific frequency and opposes current below that frequency.	
1-50.	Passes a narrow band of frequencies and opposes all others.	

1-51. The action of a filter circuit that reduces the amplitude of unwanted frequencies below the amplitude of the desired frequency is known as

1. attenuation
2. amplification
3. discrimination
4. impedance matching

1-52. The frequency beyond which a filter circuit no longer passes current is referred to as the

1. filter frequency
2. cutoff frequency
3. resonant frequency
4. response frequency

THIS SPACE LEFT BLANK INTENTIONALLY.

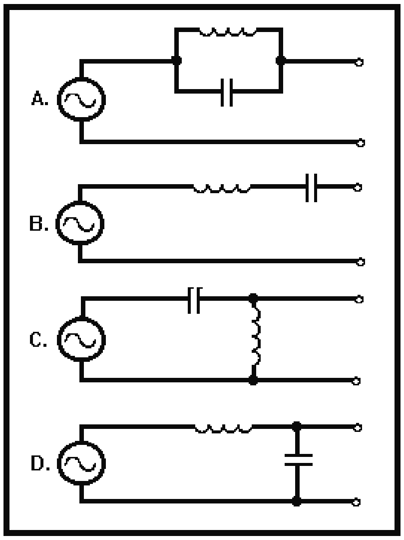


Figure 1G.—Filter circuits.

TO ANSWER QUESTIONS 1-53 AND 1-54, SELECT FROM FIGURE 1G THE CIRCUIT DIAGRAM WHICH MATCHES THE CIRCUIT NAME IN EACH QUESTION. CHOICES IN THE FIGURE MAY BE USED ONCE, MORE THAN ONCE, OR NOT AT ALL.

1-53. Band-reject filter.

1. A
2. B
3. C
4. D

1-54. High-pass filter.

1. A
2. B
3. C
4. D

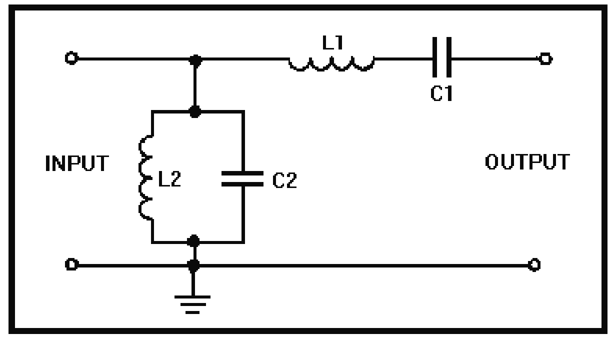


Figure 1H.—Filter circuit.

IN ANSWERING QUESTIONS 1-55 AND 1-56, REFER TO THE CIRCUIT IN FIGURE 1H.

1-55. L1 and C1 in the circuit offer what type of opposition to (a) frequencies near resonance and (b) all other frequencies?

1. (a) Minimum (b) minimum
2. (a) Minimum (b) maximum
3. (a) Maximum (b) maximum
4. (a) Maximum (b) minimum

1-56. In the type of filter circuit in the figure, what is/are the "cutoff point(s)?"

1. Upper frequency limit only
2. Lower frequency limit only
3. Both upper and lower frequency limits

1-57. In a series-resonant circuit that is operating at resonance, what is the amplitude of the applied voltage compared to (a) inductor voltage and (b) capacitor voltage?

1. (a) Lower (b) lower
2. (a) Lower (b) higher
3. (a) Higher (b) higher
4. (a) Higher (b) lower

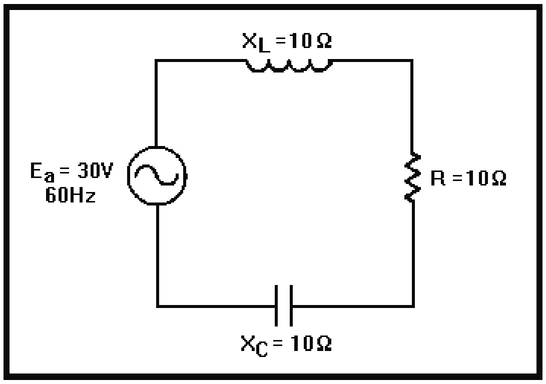


Figure 1I.—Series-RCL circuit at resonance.

IN ANSWERING QUESTIONS 1-58 THROUGH 1-60, REFER TO FIGURE 1I.

1-58. With the circuit in the figure at resonance, what is the circuit current?

1. 1 ampere
2. 2 amperes
3. 3 amperes
4. 0.5 ampere

1-59. If E_a were increased to 60 volts at the resonant frequency, what would be the voltage drop across the capacitor?

1. 10 volts
2. 20 volts
3. 30 volts
4. 60 volts

1-60. If the circuit is at resonance, what is circuit impedance?

1. 10 ohms
2. 20 ohms
3. 30 ohms
4. 40 ohms

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Oscillators," pages 2-1 through 2-38.

2-1. A sinusoidal oscillator can be regarded as which of the following types of amplifiers?

1. One that produces a trapezoidal wave
2. One that produces a sine wave
3. One that produces a square wave
4. One that produces a sawtooth wave

2-2. Wave generators are classified according to the

1. input wave shape
2. output wave shape
3. current in the output
4. voltage in the output

2-3. An IDEAL sinusoidal oscillator would produce which of the following outputs?

1. A square wave of constant frequency and amplitude
2. A square wave of varying frequency and amplitude
3. A sine wave of constant frequency and amplitude
4. A sine wave of varying frequency and constant amplitude

2-4. What three circuits are most commonly used as frequency determining devices?

1. Class C amplifier, class B amplifier, and class A amplifier
2. Crystal-controlled oscillator, RC oscillator, and LC oscillator
3. Common-emitter amplifier, common-base amplifier, and common-collector amplifier
4. Transformer coupler, RC coupler, and direct coupler

2-5. Which of the following circuits is NOT a relaxation oscillator?

1. A multivibrator
2. A sawtooth generator
3. A blocking oscillator
4. A sinusoidal oscillator

2-6. Which of the following definitions describes the basic oscillator?

1. A nonrotating device producing alternating current
2. A rotating device producing alternating current
3. A nonrotating device producing direct current
4. A rotating device producing direct current

2-7. Amplitude stability in an oscillator is the ability to

1. produce an increased amplitude in the output
2. produce a variable amplitude in the output
3. maintain a constant frequency in the output
4. maintain a constant amplitude in the output

2-8. Frequency stability in an oscillator refer to its ability to

1. maintain a constant operating frequency
2. maintain a variable operating amplitude
3. maintain a constant amplitude
4. vary operating frequency

2-9. What is the purpose of a buffer amplifier?

1. To provide a direct connection between the oscillator and the load
2. To amplify the output signal of the oscillator
3. To remove frequency distortion from the oscillator
4. To prevent load variations from affecting the oscillator

2-10. Why is class A bias used in oscillators?

1. To develop low power
2. To develop maximum power
3. To maintain low distortion
4. To maintain high efficiency

2-11. When a group of RC networks is used for regenerative feedback, which of the following waveform actions takes place in each successive stage?

1. Waveform is rectified
2. Amplitude is decreased
3. Amplitude is increased
4. Amplitude is held constant

2-12. When RC networks are connected in cascade (series), what amount of phase shift should you see?

1. The sum of the phase shifts of each RC network
2. The difference between the phase shifts of each RC network
3. The product of the phase shifts of each RC network
4. The square of the phase shifts of each RC network

2-13. Which of the following terms describes the gradual amplitude reduction in an oscillator?

1. Damping
2. Phase shift
3. Regeneration
4. Flywheel effect

2-14. Which of the following formulas can be used to figure frequency in an LC tank circuit?

1. $f_r = \frac{1}{2\pi LC}$
2. $f_r = \frac{1}{2\pi\sqrt{LC}}$
3. $f_r = \frac{1}{2\pi\sqrt{X_C}}$
4. $f_r = \frac{1}{2\pi\sqrt{X_C L_C}}$

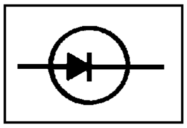
2-15. Which of the following actions best describes the piezoelectric effect?

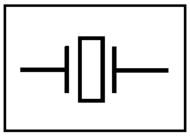
1. Produces an dc output voltage for a given ac input voltage
2. Produces an output voltage for a given mechanical input
3. Produces a mechanical output for a given input voltage
4. Both 2 and 3 above

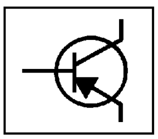
2-16. The piezoelectric effect is the property of a crystal which produces which of the following electrical characteristics?

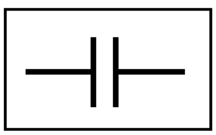
1. Resistance
2. Inductance
3. Capacitance
4. Each of the above

2-17. What is the schematic symbol for a crystal?

1. 

2. 

3. 

4. 

2-18. What electrical characteristic makes the frequency stability of a crystal better than that of an LC tank circuit?

1. Higher Q
2. Higher inductance
3. Higher resistance
4. Higher capacitance

2-19. How is feedback described?

1. Control of a circuit output signal by the input signal
2. Control of a circuit input signal by the output of the previous circuit
3. Transfer of a portion of the output circuit energy to control the input of the circuit
4. Transfer of a portion of the input circuit energy to control the output circuit

2-20. Which of the following terms describes the types of feedback?

1. Degenerative and regenerative
2. Negative and positive
3. Both 1 and 2 above
4. Bypassed and unbypassed

2-21. What type of feedback aids an input signal?

1. Positive
2. Negative
3. Bypassed
4. Degenerative

2-22. What type of feedback opposes an input signal?

1. Positive
2. Unbypassed
3. Degenerative
4. Regenerative

2-23. What type of feedback is used to sustain oscillations?

1. Bypassed
2. Negative
3. Degenerative
4. Regenerative

2-24. What oscillator uses a tickler coil for feedback?

1. Hartley
2. Colpitts
3. Armstrong
4. RC phase-shift

2-25. What oscillator uses a tapped coil for feedback?

1. Hartley
2. Colpitts
3. Armstrong
4. RC phase-shift

2-26. What oscillator uses split capacitors for feedback?

1. Hartley
2. Colpitts
3. Armstrong
4. RC phase-shift

TO ANSWER QUESTIONS 2-27 THROUGH 2-29, SELECT THE CONFIGURATIONS IN COLUMN B THAT MATCH THE AMPLIFIER CHARACTERISTICS IN COLUMN A. CHOICES IN COLUMN B MAY BE USED ONCE, MORE THAN ONCE OR NOT AT ALL.

A. CHARACTERISTICS	B. CONFIGURATIONS
2-27. Voltage gain is less than unity	1. Common-base
	2. Common-gate
	3. Common-emitter
2-28. Low power gain	4. Common-collector
2-29. Feedback signal requires phase shift	

2-30. Which of the following statements best describes tank current in a series-fed oscillator?

1. The dc path is through the tank circuit
2. The dc path does not go through the tank circuit
3. The ac path is through the tank circuit
4. The ac path does not go through the tank circuit

2-31. In a shunt-fed, tuned-collector Armstrong oscillator, what blocks the dc component from the tank circuit?

1. A resistor
2. A capacitor
3. An inductor
4. A transistor

2-32. Which of the following circuit arrangements aid in the frequency stability of an oscillator?

1. A regulated power supply
2. A common bias source for the emitter and collector
3. Both 1 and 2 above
4. Separate bias sources

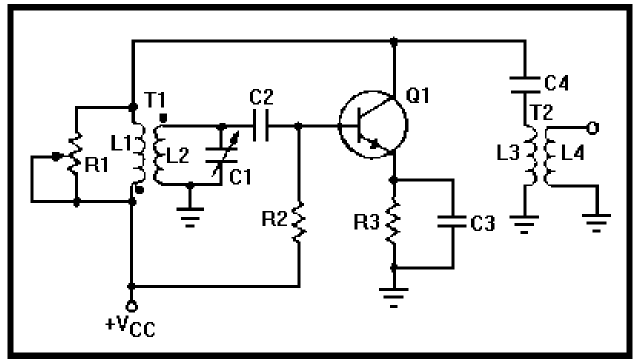


Figure 2A.—Tuned-base Armstrong oscillator.

IN ANSWERING QUESTIONS 2-33 THROUGH 2-37, REFER TO FIGURE 2A.

2-33. The frequency of the output signal of the oscillator is determined by what components?

1. R1 and L1
2. L2 and C1
3. L3 and C4
4. R3 and C3

2-34. Forward bias for the amplifier is developed by what component?

1. R1
2. R2
3. R3
4. L1

2-35. The resonant frequency is tuned to the desired value by what component?

1. C1
2. C2
3. L3
4. L1

2-36. What is the maximum degree of phase shift provided between the base and collector of Q1?

1. 0 degrees
2. 90 degrees
3. 120 degrees
4. 180 degrees

2-37. Temperature stability of the oscillator is improved by what component?

1. R1
2. R2
3. R3
4. C4

2-38. What feature in a Hartley oscillator differs from an Armstrong oscillator?

1. Tickler coil
2. Split inductor
3. Split coupling
4. Split capacitance

THIS SPACE LEFT BLANK INTENTIONALLY.

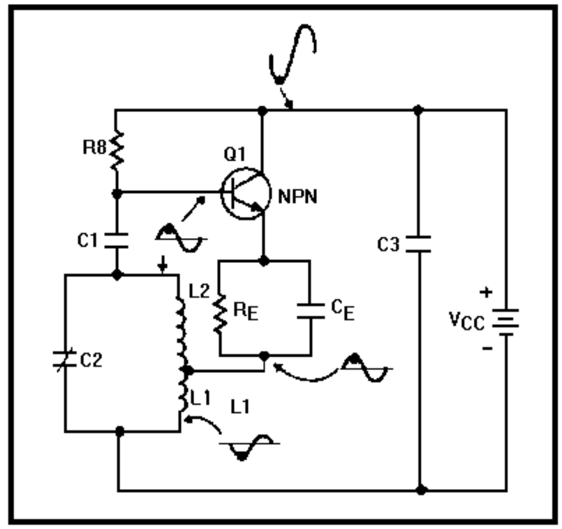


Figure 2B.—Series-fed, tuned-base Hartley oscillator.

IN ANSWERING QUESTIONS 2-39 THROUGH 2-42, REFER TO FIGURE 2B.

2-39. What components are part of the frequency-determining device of this oscillator?

1. C1, L1, and L2
2. C2, L1, and L2
3. C3, L1, and L2
4. C_E, R_E, and R_B

2-40. What circuit component prevents thermal runaway?

1. L1
2. C_E
3. R_B
4. R_E

2-41. The low resistance of L2 could place a short across the emitter-to-base junction network of Q1 and R_E. What component in the circuit prevents this from happening?

1. C1
2. C2
3. C3
4. C_E

2-42. When a positive signal is coupled to the base of Q1, what happens to (a) collector current and (b) emitter current?

1. (a) Increases (b) increases
2. (a) Increases (b) decreases
3. (a) Decreases (b) decreases
4. (a) Decreases (b) increases

2-43. A tuned-base Hartley oscillator is described as "shunt fed" when

1. ac flows through the tank circuit
2. dc flows through the tank circuit
3. ac does not flow through the tank circuit
4. dc does not flow through the tank circuit

2-44. Which of the following advantages does the Colpitts oscillator have over the Armstrong and Hartley oscillators?

1. Easier to tune
2. Wider frequency range
3. Better frequency stability
4. All of the above

THIS SPACE LEFT BLANK INTENTIONALLY.

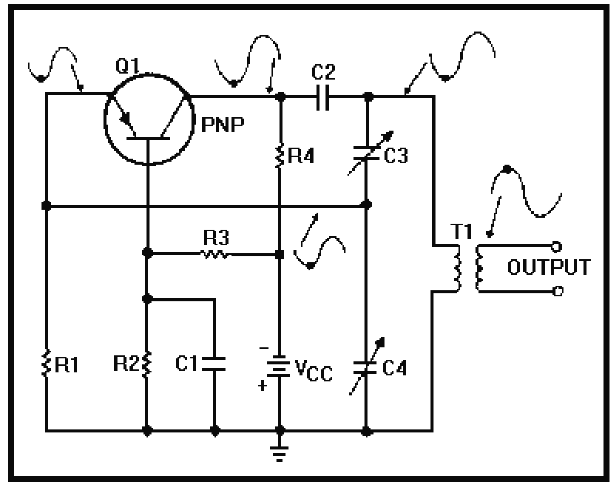


Figure 2C.—Oscillator.

IN ANSWERING QUESTIONS 2-45 THROUGH 2-47, REFER TO FIGURE 2C.

2-45. What type of oscillator is shown in the figure?

1. Common-base Hartley
2. Common-base Colpitts
3. Common-emitter Colpitts
4. Common-collector Hartley

2-46. What component is the collector load resistor?

1. R1
2. R2
3. R3
4. R4

2-47. What resistors provide the base bias?

1. R1, R2
2. R2, R3
3. R3, R4
4. R2, R4

2-48. What class of biasing does the RC oscillator use?

1. A
2. B
3. C
4. AB

2-49. In an RC network, (a) what type of impedance is presented and (b) does the current lead or lag?

1. (a) Inductive (b) leads
2. (a) Inductive (b) lags
3. (a) Capacitive (b) lags
4. (a) Capacitive (b) leads

2-50. In the phase-shift oscillator, a phase shift of 180 degrees for regenerative feedback is provided by what minimum number of RC networks?

1. One
2. Two
3. Three
4. Four

2-51. What determines the phase angle of an RC network?

1. Input voltage
2. Output voltage
3. Values of resistance and inductance
4. Values of resistance and capacitance

THIS SPACE LEFT BLANK INTENTIONALLY.

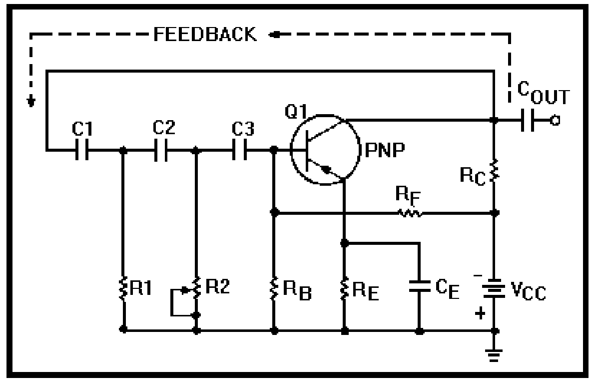


Figure 2D.—Phase-shift oscillator.

IN ANSWERING QUESTIONS 2-52 THROUGH 2-56, REFER TO FIGURE 2D.

2-52. What is the maximum amount of phase shift provided by Q1 in the figure?

1. 0 degrees
2. 60 degrees
3. 90 degrees
4. 180 degrees

2-53. What type of feedback is provided through the RC networks to the base of Q1?

1. Neutral
2. Negative
3. Regenerative
4. Degenerative

2-54. At any other than the desired frequency, what type of feedback is provided by the circuit?

1. Neutral
2. Positive
3. Regenerative
4. Degenerative

2-55. What components make up the frequency-determining device?

1. C1, C2, C_E, R1, R2, R_B
2. C2, C3, C_E, R2, R_B, R_E
3. C1, C2, C3, R1, R2, R_B
4. C_{out}, C1, C2, R1, R2, R_E

2-56. What is the maximum amount of phase shift provided by the C3-R_B network?

1. 90 degrees
2. 80 degrees
3. 70 degrees
4. 60 degrees

2-57. Which of the following is the correct formula for the resonant frequency of a phase-shift oscillator?

1. $f_r = \frac{1}{2\pi\sqrt{LC}}$
2. $f_r = \frac{1}{2\pi RC\sqrt{2n}}$
3. $f_r = \frac{1}{2\pi\sqrt{fL}}$
4. $f_r = \frac{1}{2\pi\sqrt{RC}}$

2-58. Which of the following oscillators is used to provide a highly stable output at a very precise frequency?

1. Crystal
2. Hartley
3. Colpitts
4. Armstrong

2-59. The frequency of a crystal-controlled oscillator is determined by which of the following physical actions?

1. Type of cut
2. Accuracy of cut
3. Thickness of grinding
4. All of the above

2-60. Why is the crystal in a crystal-controlled oscillator often installed in a temperature-controlled oven?

1. To increase frequency without changing the crystal
2. To decrease frequency without changing the crystal
3. To provide better amplitude stability
4. To provide better frequency stability

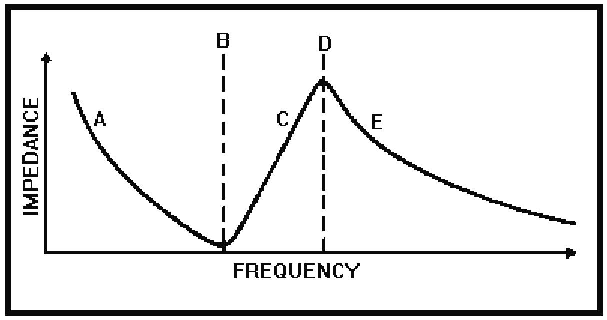


Figure 2E.—Frequency response of a crystal.

IN ANSWERING QUESTIONS 2-61 THROUGH 2-64, REFER TO FIGURE 2E. THE LETTERS A, B, C, D, AND E ARE POINTS ON THE FREQUENCY-RESPONSE CURVE FROM WHICH YOU SHOULD SELECT ANSWERS TO THE QUESTIONS.

2-61. At what point on the curve does a crystal act as a series-tuned circuit?

1. A
2. B
3. C
4. D

2-62. At what point does the crystal act inductively?

1. A
2. B
3. C
4. D

2-63. Below series resonance, a crystal acts capacitively at what point on the curve?

1. A
2. B
3. C
4. E

2-64. At what point does the crystal act purely as a parallel-resonant circuit?

1. B
2. C
3. D
4. E

2-65. How is the Q of a crystal determined?

1. Type of cut used
2. Type of holder used
3. Accuracy of the grinding
4. All of the above

2-66. An oscillator that is turned ON for a specific period of time, then is turned OFF and remains OFF until required at a later time, is which of following types?

1. LC
2. Pierce
3. Pulsed
4. Crystal

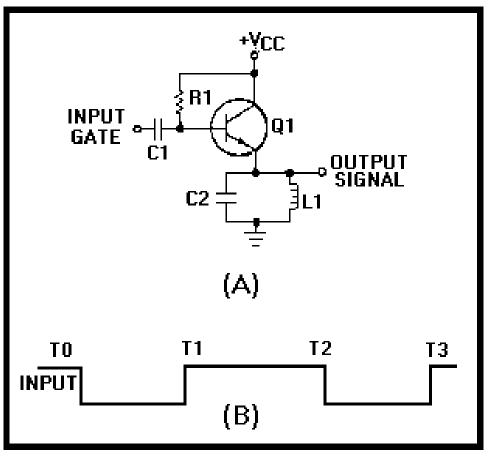


Figure 2F.—Oscillator circuit.

IN ANSWERING QUESTIONS 2-67 THROUGH 2-70, REFER TO FIGURE 2F.

2-67. What circuit is shown in the figure?

1. Pierce oscillator
2. Pulsed oscillator
3. Colpitts oscillator
4. Armstrong oscillator

2-68. Sine waves are generated in the emitter circuit of Q1 during which of the following time periods of the input gate?

1. T0 to T1 and T1 to T2
2. T0 to T1 and T2 to T3
3. T1 to T2 and T3 to T4
4. T1 to T3 and T0 to T4

2-69. The frequencies in the output are determined by what two circuit parameters?

1. Input gate time and the time the circuit is turned OFF
2. Output gate time and the time the circuit is turned ON
3. Input gate time and the resonant frequency of the tank circuit
4. Output gate time and the resonant frequency of the tank circuit,

2-70. If the resonant frequency of the tank circuit were 5 megahertz and transistor Q1 were cut off for 500 microseconds, what maximum number of cycles of the tank frequency would be present in each pulse of the output?

1. 500 cycles
2. 1,500 cycles
3. 2,500 cycles
4. 3,500 cycles

2-71. What is the fourth harmonic of a 2-megahertz signal?

1. 6 megahertz
2. 2 megahertz
3. 8 megahertz
4. 4 megahertz

2-72. What is the highest multiplication factor normally used in frequency multipliers?

1. One
2. Two
3. Three
4. Four

2-73. As the multiplication factor in a frequency multiplier circuit is increased, what happens to the output signal (a) amplitude and (b) frequency?

1. (a) Increases (b) increases
2. (a) Increases (b) decreases
3. (a) Decreases (b) decreases
4. (a) Decreases (b) increases

2-74. In a buffer amplifier, what is the impedance in the (a) input and (b) output?

1. (a) Low (b) low
2. (a) Low (b) high
3. (a) High (b) high
4. (a) High (b) low

ASSIGNMENT 3

Textbook assignment: Chapter 3, "Waveforms and Wave Generators," pages 3-1 through 3-56.

3-1. A waveform that repeats the same pattern of changes is a/an

1. periodic wave
2. sporadic wave
3. aperiodic wave
4. transverse wave

3-2. Which of the following waveforms is sinusoidal?

1. Sine wave
2. Square wave
3. Sawtooth wave
4. Rectangular wave

3-3. The time required to complete one full cycle of a square wave is referred to as the

1. pulse-repetition rate
2. pulse-repetition time
3. pulse-repetition cycle
4. pulse-repetition frequency

3-4. What term(s) describes the number of times in one second that a square wave repeats itself?

1. The pulse-repetition frequency (prf)
2. The pulse-repetition rate (prr)
3. Both 1 and 2 above
4. The pulse-repetition time (prt)

3-5. A square wave with a prf of 1,250 hertz has a prt of

1. 8 microseconds
2. 80 microseconds
3. 800 microseconds
4. 8,000 microseconds

THIS SPACE LEFT BLANK INTENTIONALLY.

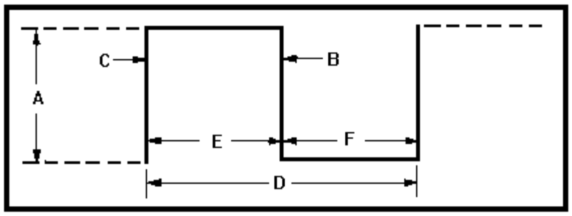


Figure 3A.—Square wave.

IN ANSWERING QUESTIONS 3-6 THROUGH 3-10, SELECT THE CORRESPONDING LETTER IN FIGURE 3A WHICH DESCRIBES THE PORTION OF THE WAVEFORM IN EACH OF THE FOLLOWING QUESTIONS.

3-6. The leading edge of the pulse.

1. C
2. D
3. E
4. F

3-7. The trailing edge of the pulse.

1. A
2. B
3. C
4. D

3-8. The positive alternation.

1. C
2. D
3. E
4. F

3-9. The amplitude of the pulse.

1. A
2. B
3. C
4. D

3-10. The pulse-repetition time of the pulse.

1. C
2. D
3. E
4. F

3-11. What type of waveform is used to furnish a linear rise in current for electromagnetic cathode ray tubes?

1. Square wave
2. Sawtooth wave
3. Trapezoidal wave
4. Rectangular wave

3-12. Which of the following multivibrators must have a signal applied (triggered) to change states?

1. Astable
2. Bistable
3. Monostable
4. Both 2 and 3 above

3-13. Which of the following multivibrators is also called a free running multivibrator?

1. Astable
2. Bistable
3. Monostable
4. Both 2 and 3 above

3-14. Which of the following waveforms could be the output of the astable multivibrator?

1. Sine wave
2. Sawtooth wave
3. Rectangular wave
4. Trapezoidal wave

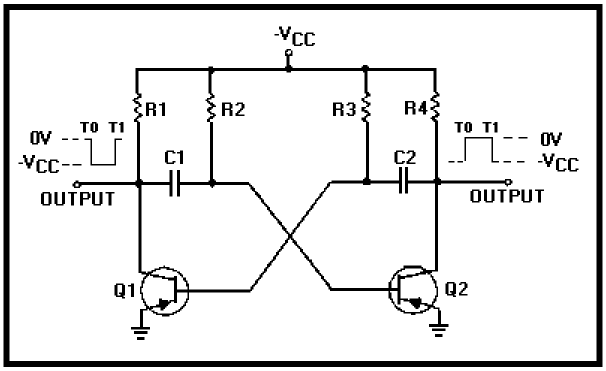


Figure 3B.—Multivibrator.

IN ANSWERING QUESTIONS 3-15 AND 3-16, REFER TO FIGURE 3B.

3-15. The time necessary for Q2 in the circuit to become saturated is controlled by what RC network?

1. R1, C1
2. R2, C1
3. R3, C2
4. R4, C2

3-16. Which of the following conditions exist in the outputs of the figure from T0 to T1?

1. Q1 saturated, Q2 saturated
2. Q1 saturated, Q2 cutoff
3. Q1 cutoff, Q2 cutoff
4. Q1 cutoff, Q2 saturated

3-17. What multivibrator is a square or rectangular-wave generator with only one stable condition?

1. Astable
2. Bistable
3. Monostable
4. Eccles-Jordan

3-18. What is the primary use for the monostable multivibrator circuit?

1. Filter
2. Amplifier
3. Oscillator
4. Pulse stretcher

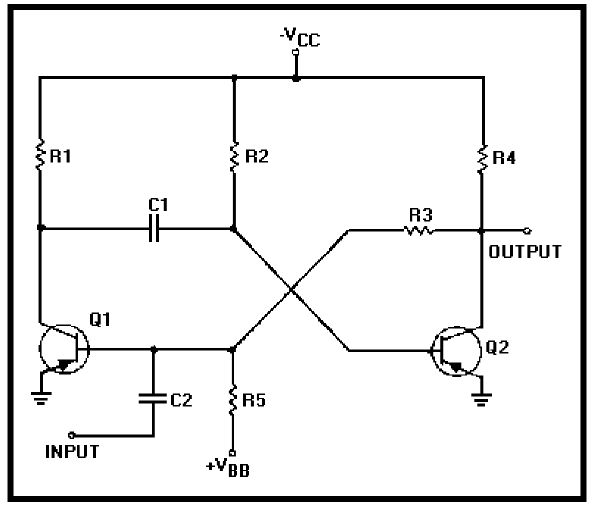


Figure 3C.—Monostable multivibrator circuit.

IN ANSWERING QUESTIONS 3-19 AND 3-20, REFER TO FIGURE 3C.

3-19. In the stable state of the circuit, what will be the condition of (a) Q1 and (b) Q2?

1. (a) Cutoff (b) cutoff
2. (a) Cutoff (b) saturated
3. (a) Saturated (b) saturated
4. (a) Saturated (b) cutoff

3-20. What is the discharge path for C1?

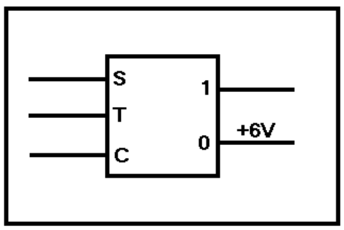
1. C1, Q2, ground, $-V_{CC}$, Q1, and C1
2. C1, Q2, ground, $+V_{BB}$, R5, Q1, and C1
3. C1, Q2, R4, R2, $+V_{BB}$, $-V_{CC}$, R2, and C1
4. C1, Q1, ground, $-V_{CC}$, R2, and C1

3-21. In a bistable multivibrator, what minimum number of triggers is required to produce one gate?

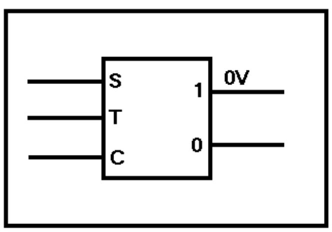
1. One
2. Two
3. Three
4. Four

3-22. Which of the flip flops shown below is in the SET state?

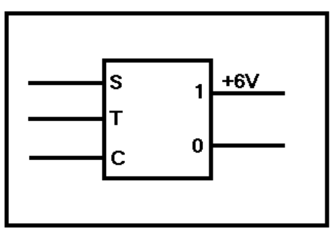
1.



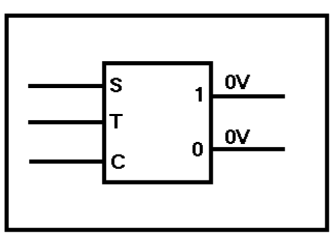
2.



3.

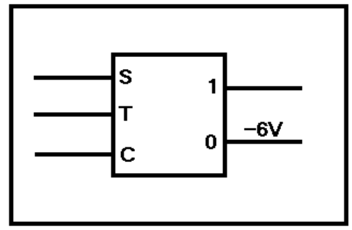


4.

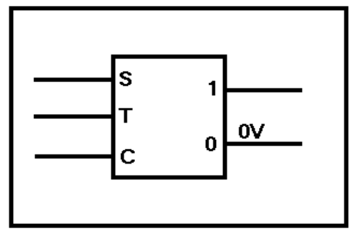


3-23. Which of the flip flops shown below is in the CLEAR state?

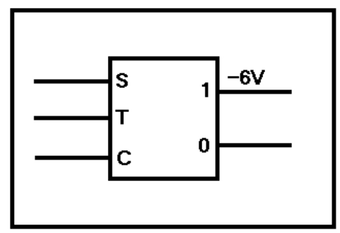
1.



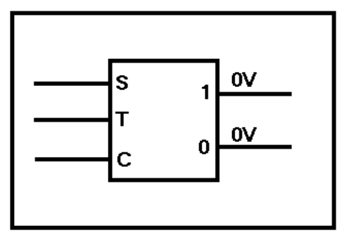
2.



3.



4.



3-24. Which of the following pulses is used to change states in a flip flop?

1. A trigger pulse
2. A clipping pulse
3. A modulating pulse
4. An interference pulse

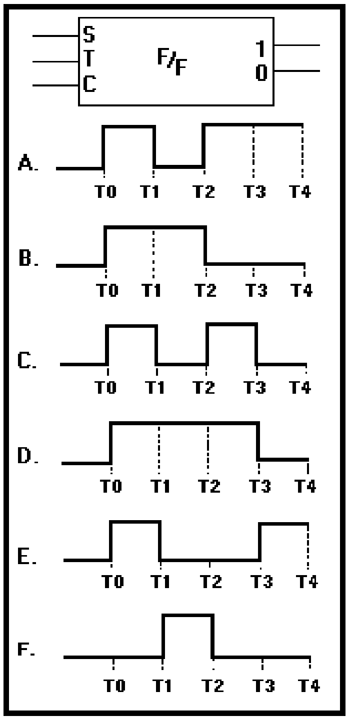
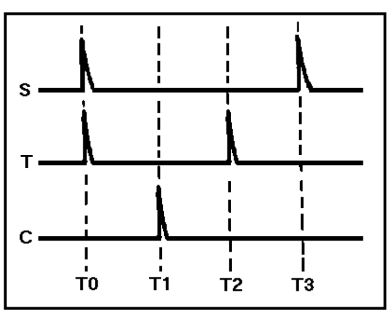


Figure 3D.—Flip-flop output 1 waveforms.

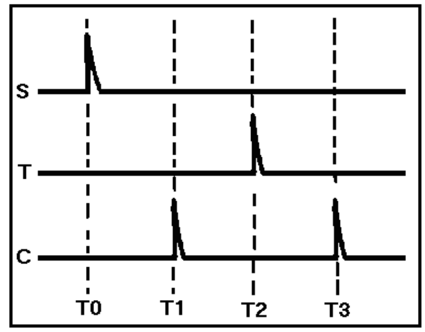
IN ANSWERING QUESTIONS 3-25 THROUGH 3-29, REFER TO FIGURE 3D. ASSUME THE FLIP-FLOP IS INITIALLY IN THE CLEAR STATE. SELECT THE WAVEFORM AT THE "1" OUTPUT IN THE FIGURE THAT WILL RESULT FROM THE INPUT PULSES SHOWN IN THE QUESTIONS. SOME CHOICES MAY BE USED ONCE, MORE THAN ONCE, OR NOT AT ALL.

3-25.



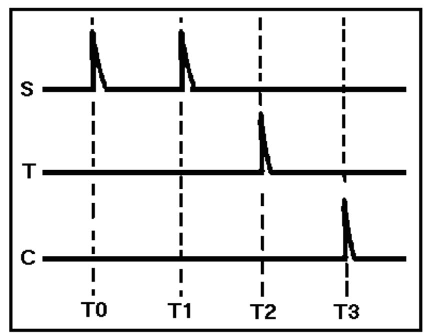
1. A
2. B
3. C
4. D

3-26.



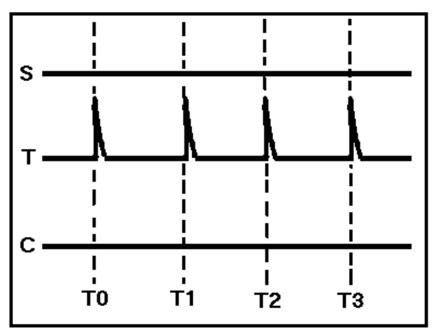
1. A
2. B
3. C
4. D

3-27.



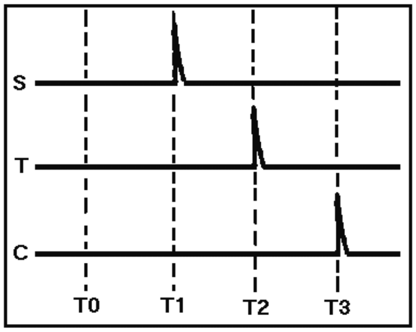
1. A
2. B
3. D
4. E

3-28.



1. B
2. C
3. D
4. F

3-29.



1. A
2. B
3. E
4. F

3-30. The toggle input on a flip flop is used to cause which of the following circuit actions?

1. SET the flip flop
2. CLEAR the flip flop
3. Both 1 and 2 above
4. Sample the condition of the flip flop

3-31. The blocking oscillator is NOT suitable for which of the following circuit applications?

1. Counter circuit
2. Frequency divider
3. Switching circuit
4. Sine-wave generator

THIS SPACE LEFT BLANK INTENTIONALLY.

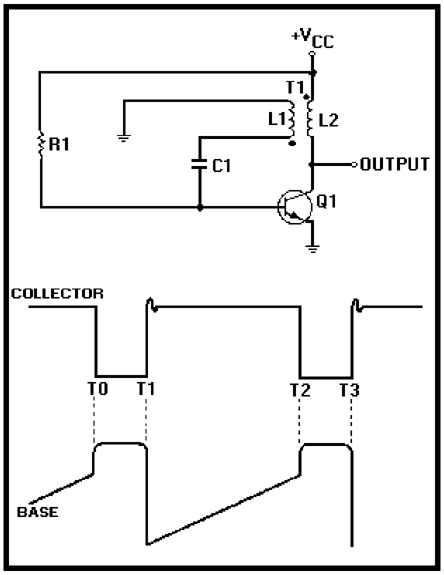


Figure 3E.—Blocking oscillator.

IN ANSWERING QUESTIONS 3-32 THROUGH 3-37, REFER TO FIGURE 3E.

3-32. What is indicated by the dots at each end of T1 in the figure?

1. 0-degree phase shift
2. 90-degree phase shift
3. 160-degree phase shift
4. 180-degree phase shift

3-33. Regenerative feedback to the base of Q1 is provided by what circuit component(s)?

1. L1 only
2. L2 only
3. L1 and L2 only
4. L1, L2, and C1

3-34. What circuit action is taking place from T0 to T1?

1. C1 is charging
2. C1 is discharging
3. L1 is discharging
4. I_C is decreasing

3-35. THIS QUESTION HAS BEEN DELETED.

3-36. During what total time period is Q1 blocked?

1. T0 to T1
2. T1 to T2
3. T0 to T2
4. T2 to T3

3-37. In the blocking oscillator, which of the following circuit actions is the primary cause of parasitic oscillations?

1. Collapse of the magnetic field of L1
2. Expansion of the magnetic field of L1
3. Inductive coupling between L1 and L2
4. C1 discharging through Q1

3-38. Which of the following circuit actions is a result of critical damping?

1. Rapid transient response without overshoot
2. Rapid transient response with overshoot
3. Slow transient response without overshoot
4. Slow transient response with overshoot

3-39. What type of damping is caused by (a) high resistance and (b) low resistance?

1. (a) Overdamping
   (b) Underdamping
2. (a) Underdamping
   (b) Overdamping
3. (a) Critical damping
   (b) Underdamping
4. (a) Overdamping
   (b) Critical damping

3-40. Applying synchronizing triggers with a frequency that is SLIGHTLY higher than the free-running frequency will cause a synchronized blocking oscillator to

1. divide in frequency
2. double in frequency
3. lock in at the higher frequency
4. stay locked at the free-running frequency

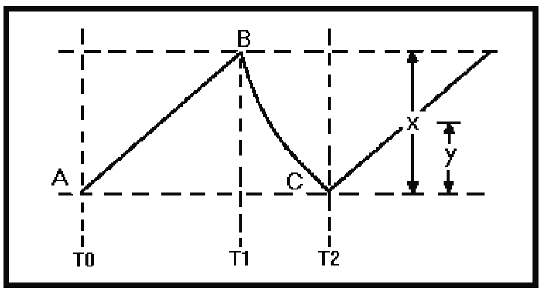


Figure 3F.—Sawtooth waveform.

IN ANSWERING QUESTIONS 3-41 THROUGH 3-46, REFER TO FIGURE 3F AND MATCH THE WAVEFORM POINTS (OR TIME REFERENCES) TO THE TERMS IN THE QUESTIONS. SOME CHOICES MAY BE USED ONCE, MORE THAN ONCE, OR NOT AT ALL.

3-41. Linear slope.

1. A to B
2. B to C
3. T0 to T2
4. T1 to T2

3-42. Physical length.

1. A to B
2. B to C
3. X
4. Y

3-43. Sweep time.

1. A to B
2. B to C
3. T0 to T1
4. T1 to T2

3-44. Fall time.

1. A to C
2. T0 to T1
3. T1 to T2
4. T0 to T2

3-45. Electrical length.

1. B to C
2. T0 to T1
3. T1 to T2
4. T0 to T2

3-46. Amplitude.

1. A to B
2. B to C
3. X
4. Y

3-47. The linearity of the rise voltage in a sawtooth wave is determined by which of the following circuit timing actions?

1. The time the capacitor is allowed to charge
2. The time it takes the capacitor to fully charge
3. The time the capacitor is allowed to discharge
4. The time it takes the capacitor to fully discharge

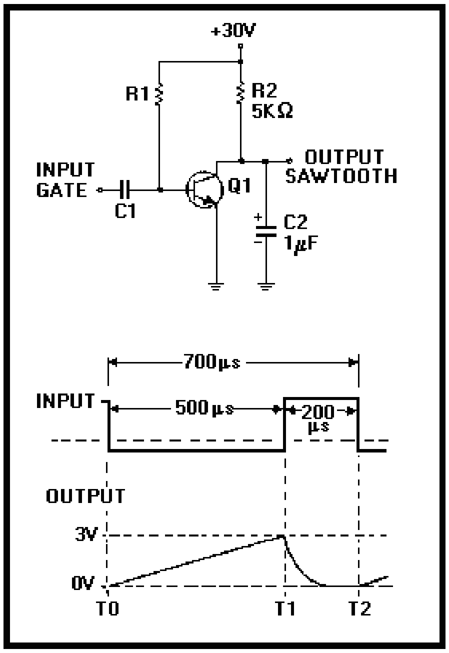


Figure 3G.—Transistor sawtooth generator.

IN ANSWERING QUESTIONS 3-48 THROUGH 3-53, REFER TO FIGURE 3G.

3-48. What component in the circuit develops the output sawtooth waveform?

1. R1
2. R2
3. C1
4. C2

3-49. What is the purpose of Q1?

1. Acts as a switch
2. Allows C2 to charge
3. Inverts the negative gate
4. Serves as a common-collector amplifier

3-50. What is the maximum length of time C2 is allowed to charge?

1. 200 microseconds
2. 500 microseconds
3. 700 microseconds
4. 900 microseconds

3-51. If V_{CC} were increased to 40 volts, which of the following parameters in the output sawtooth wave would increase?

1. Fall time
2. Amplitude
3. Sweep time
4. Linearity

3-52. What would be the effect on C2 if the negative gate length were increased?

1. Charge to V_{CC}
2. Discharge to V_{CC}
3. Charge to a larger percentage of V_{CC}
4. Charge to a smaller percentage of V_{CC}

3-53. What is the prf of the circuit?

1. 1,428 pulses per second
2. 1,450 pulses per second
3. 1,470 pulses per second
4. 1,482 pulses per second

3-54. In a sawtooth generator, a change in which of the following parameters will NOT affect the linearity of the output?

1. V_{CC}
2. Resistance
3. Capacitance
4. Gate length

THIS SPACE LEFT BLANK INTENTIONALLY.

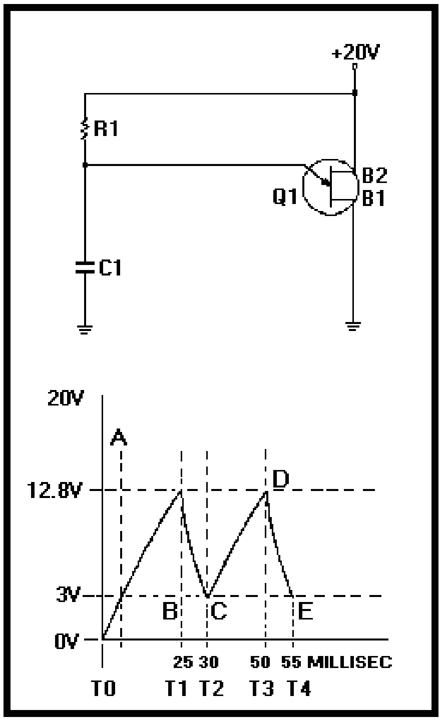


Figure 3H.—Unijunction sawtooth generator.

IN ANSWERING QUESTIONS 3-55 THROUGH 3-59, REFER TO FIGURE 3H.

3-55. The output in the circuit is taken across what component.

1. R1
2. B2
3. R1
4. C1

3-56. What is the discharge path for C1?

1. C1, R1, V_{CC}, ground, and C1
2. C1, emitter B2, V_{CC}, ground, and C1
3. C1, B1, emitter, and C1
4. C1, B1, B2, V_{CC}, ground, and C1

3-57. What part of the waveform is sweep time?

1. A to D
2. B to C
3. C to D
4. C to E

3-58. What is the action of C1 when the emitter-to-B1 junction is (a) reverse biased and (b) forward biased?

1. (a) Charging (b) Discharging
2. (a) Discharging (b) Charging
3. (a) Charging (b) Charging
4. (a) Discharging (b) Discharging

3-59. To obtain a more stable output frequency, you could modify the circuit by applying (a) what type of triggers to (b) what circuit element.

1. (a) Positive (b) B1
2. (a) Positive (b) B2
3. (a) Negative (b) B2
4. (a) Negative (b) B1

THIS SPACE LEFT BLANK INTENTIONALLY.

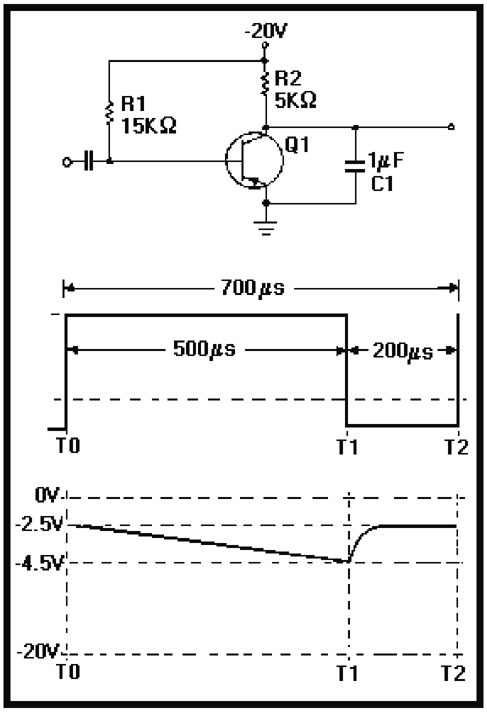


Figure 3I.—Transistor sawtooth generator.

IN ANSWERING QUESTIONS 3-60 THROUGH 3-64, REFER TO FIGURE 3I.

3-60. With no input voltage applied to the circuit, what voltage should you read at the collector?

1. -20 volts
2. -17.5 volts
3. -2.5 volts
4. 0 volts

3-61. To what voltage is C1 allowed to charge?

1. 0 volts
2. -2.5 volts
3. -4.5 volts
4. -20 volts

3-62. To determine the percent of charge on C1, what formula is used?

1. $$\frac{E_C\,max - V_{CC}}{E_C\,min - E_C\,max} \times 100$$
2. $$\frac{E_C\,max - V_{CC}}{E_C\,min + E_C\,max} \times 100$$
3. $$\frac{E_C\,max + E_C\,min}{V_{CC} - E_C\,min} \times 100$$
4. $$\frac{E_C\,max - E_C\,min}{V_{CC} - E_C\,min} \times 100$$

3-63. Which of the following actions will improve the linearity of the sawtooth?

1. Increasing the value of C1
2. Increasing the value of R1
3. Increasing the gate length
4. Each of the above

3-64. What is the maximum amplitude of the output sawtooth signal?

1. 1.75 volts
2. 2.0 volts
3. 2.5 volts
4. 4.5 volts

3-65. If applied to a coil, what voltage waveform will cause a linear rise in current?

1. Square wave
2. Sawtooth wave
3. Rectangular wave
4. Trapezoidal wave

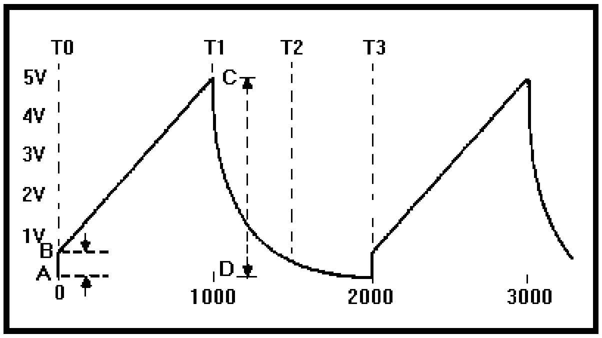


Figure 3J.—Trapezoidal waveform.

IN ANSWERING QUESTIONS 3-66 THROUGH 3-70, REFER TO FIGURE 3J AND MATCH THE WAVEFORM POINTS (OR TIME REFERENCES) TO THE TERMS IN THE QUESTIONS. SOME CHOICES MAY BE USED ONCE, MORE THAN ONCE, OR NOT AT ALL.

3-66. Pulse-repetition time (prt).

1. T1 to T2
2. T0 to T3
3. A to B
4. B to C

3-67. Physical length.

1. T0 to T3
2. A to B
3. B to C
4. C to D

3-68. Electrical length.

1. T0 to T1
2. T1 to T2
3. T0 to T3
4. A to B

3-69. Jump voltage.

1. T0 to T1
2. T1 to T2
3. T0 to T3
4. A to B

3-70. Linear slope.

1. C to D
2. B to C
3. A to B
4. T0 to T3

THIS SPACE LEFT BLANK INTENTIONALLY.

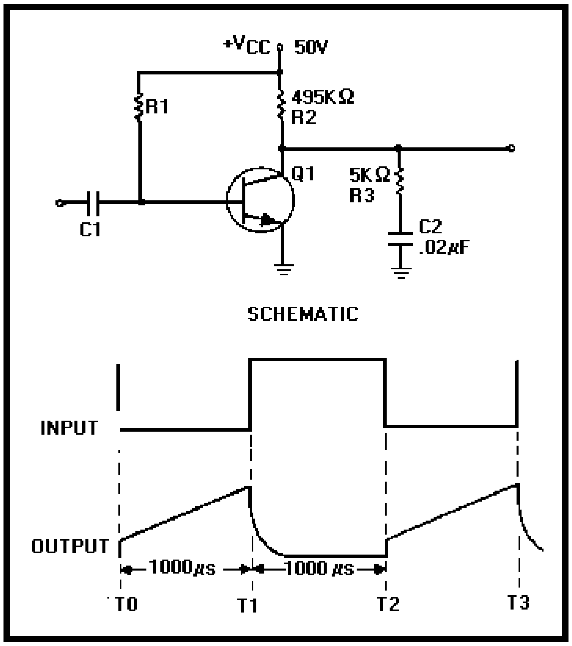


Figure 3K.—Trapezoidal-wave generator with input and output waveforms.

IN ANSWERING QUESTIONS 3-71 THROUGH 3-74, REFER TO FIGURE 3K.

3-71. The amplitude of the jump voltage is approximately

$$\frac{(a)}{0.5V,\ 5V}$$

and the amplitude of the trapezoidal wave is

$$\frac{(b)}{0.5V,\ 5V}.$$

1. (a) 0.5 volts (b) 5 volts
2. (a) 0.5 volts (b) 0.5 volts
3. (a) 5 volts (b) 0.5 volts
4. (a) 5 volts (b) 5 volts

3-72. Which of the following components and/or value(s) determines the amplitude of the jump voltage?

1. V_{CC} only
2. R2 and V_{CC} only
3. R3 and V_{CC} only
4. R2, R3, and V_{CC}

3-73. What is the minimum discharge time for C2?

1. 50 microseconds
2. 500 microseconds
3. 1,000 microseconds
4. 2,000 microseconds

3-74. Increasing which of the following values will NOT affect linearity of the circuit?

1. Resistance of R2
2. Capacitance of C2
3. Gate width
4. V_{CC}

ASSIGNMENT 4

Textbook assignment: Chapter 4, “Wave Shaping,” pages 4-1 through 4-61.

4-1. A wave-shaping circuit which restricts some portion of a waveform from exceeding a specified value is known as a/an

1. divider
2. clamper
3. limiter
4. oscillator

4-2. Limiting circuits are used in which of the following circuit applications?

1. Counting
2. Amplification
3. Wave generation
4. Circuit protection

4-3. In a series limiter, the diode is connected in

(a)
parallel, series

With the

(b)
input, output

1. (a) Parallel (b) output
2. (a) Parallel (b) input
3. (a) Series (b) input
4. (a) Series (b) output

4-4. In a series limiter, a voltage is developed across the output resistor when

1. the anode of the diode is negative with respect to the cathode
2. the anode of the diode is positive with respect to the cathode
3. the cathode of the diode is positive with respect to the anode
4. no current flows through the diode

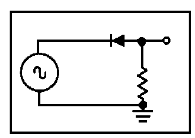
4-5. The diode in a series-positive limiter is (a) forward biased by what portion of the input signal and (b) reverse biased by what portion of the input signal?

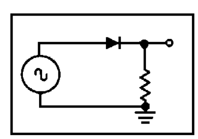
1. (a) Positive (b) positive
2. (a) Positive (b) negative
3. (a) Negative (b) negative
4. (a) Negative (b) positive

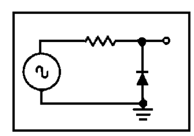
4-6. How does the value of diode resistance compare to that of the resistor (a) during the limiting portion of the input and (b) during the nonlimiting portion?

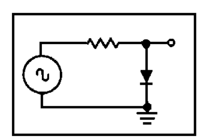
1. (a) High (b) high
2. (a) High (b) low
3. (a) Low (b) low
4. (a) Low (b) high

4-7. Which of the following circuits is a series-positive limiter?

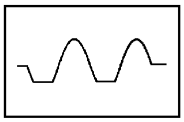
1. 

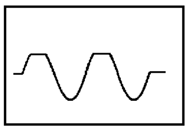
2. 

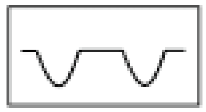
3. 

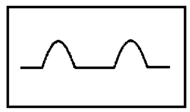
4. 

4-8. If a sine wave is applied to the input of a series-positive limiter, which of the following waveforms describes the output?

1. 

2. 

3. 

4. 

4-9. The amplitude of the output of a series-diode limiter is figured using which of the following formulas?

1. $$E_{out} = \frac{R + R_{ac}}{R} \times E_{in}$$

2. $$E_{out} = \frac{R}{R + R_{ac}} \times E_{in}$$

3. $$E_{out} = \frac{R_{ac}}{R_{ac} + R} \times E_{in}$$

4. $$E_{out} = \frac{R_{ac} + R}{R_{ac}} \times E_{in}$$

IN ANSWERING QUESTIONS 4-10 AND 4-11, ASSUME YOU ARE FIGURING OUTPUT AMPLITUDES FOR SERIES-POSITIVE LIMITERS AND CIRCUIT CONDITIONS ARE AS FOLLOWS:

E_{in} = 30 vac
R1 = 20,000Ω
R_{ac} = 100 Ω (forward bias)
R_{ac} = 150,000Ω (reverse bias)

4-10. With forward bias, what is the output amplitude?

1. 28.95 volts
2. 29 volts
3. 29.85 volts
4. 29.95 volts

4-11. With reverse bias, what is the output amplitude?

1. .175 volt
2. 1.75 volts
3. 3.53 volts
4. 3.75 volts

4-12. In a series-positive limiter, where is the input signal applied?

1. Directly to the anode of the diode
2. Directly to the cathode of the diode
3. To the anode of the diode through a series input resistor
4. To the cathode of the diode through the output resistor

4-13. With a sine wave applied, which of the following circuits limits only a portion of the positive input signal?

1. Series-negative limiter with negative bias
2. Series-negative limiter with positive bias
3. Series-positive limiter with negative bias
4. Series-positive limiter with positive bias

4-14. With a sine-wave input, which of the following types of series limiter allows only a portion of the negative input to be developed in the output?

1. Series-positive limiter without bias
2. Series-positive limiter with negative bias
3. Series-positive limiter with positive bias
4. Series-negative limiter without bias

4-15. In a series-negative limiter, how is the diode biased (a) by the positive half of the input sine wave and (b) by the negative half?

1. (a) Reverse (b) forward
2. (a) Reverse (b) reverse
3. (a) Forward (b) reverse
4. (a) Forward (b) forward

THIS SPACE LEFT BLANK INTENTIONALLY.

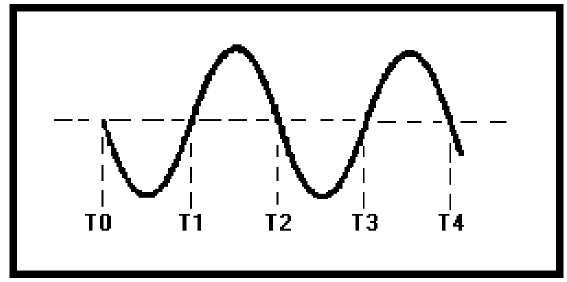
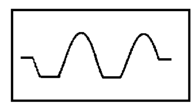
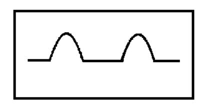


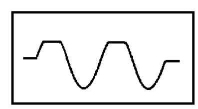
Figure 4A.—Sine-wave input.

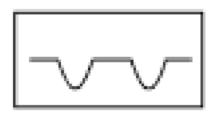
IN ANSWERING QUESTION 4-16, REFER TO FIGURE 4A.

4-16. If the input waveform shown in the figure is applied to a series-negative limiter, which of the following waveforms will be the output?

1. 

3. 

2. 

4. 

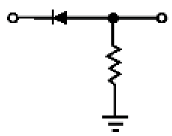
4-17. In a series-negative limiter with positive bias, which of the following output circuit actions takes place?

1. Both positive half cycles are eliminated
2. Both negative half cycles are eliminated
3. All of the positive half cycles and a portion of the negative half cycles are eliminated
4. All of the negative half cycles and a portion of the positive half cycles are eliminated

IN QUESTIONS 4-18 THROUGH 4-22, MATCH THE LIMITER CIRCUIT IN COLUMN A TO THE CIRCUIT DESCRIPTION IN COLUMN B. CHOICES MAY BE USED ONCE, MORE THAN ONCE, OR NOT AT ALL.

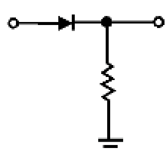
A. LIMITER CIRCUIT	B. CIRCUIT DESCRIPTION
	A. Series-positive limiter with no bias
	B. Series-negative limiter with no bias
	C. Series-positive limiter with negative bias
	D. Series-positive limiter with positive bias
	E. Series-negative limiter with negative bias
	F. Series-negative limiter with positive bias

4-18.



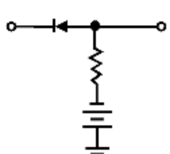
1. A
2. B
3. C
4. D

4-19.



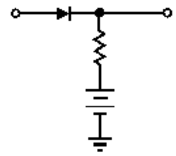
1. A
2. B
3. E
4. F

4-20.



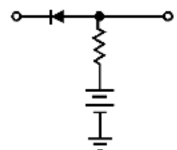
1. C
2. D
3. E
4. F

4-21.



1. C
2. D
3. E
4. F

4-22.



1. C
2. D
3. E
4. F

4-23. In a parallel-diode limiter, (a) how is the output taken and (b) under what diode condition is it developed?

1. (a) Across the resistor
   (b) When the diode is cut off
2. (a) Across the resistor
   (b) When the diode is conducting
3. (a) Across the diode
   (b) When the diode is conducting
4. (a) Across the diode
   (b) When the diode is cut off

4-24. In a parallel-positive limiter, where is the input sine wave applied?

1. At the anode of the diode
2. At the cathode of the diode
3. Through a series resistor to the anode of the diode
4. Through a series resistor to the cathode of the diode

4-25. Which of the following formulas is used to figure the output amplitude of a parallel-diode limiter?

1. $$E_{out} = \frac{R_{ac} + R}{R_{ac}} \times E_{in}$$

2. $$E_{out} = \frac{R}{R_{ac} + R} \times E_{in}$$

3. $$E_{out} = \frac{R_{ac}}{R_{ac} + R} \times E_{in}$$

4. $$E_{out} = \frac{R_{ac} + R}{R_{ac} - R} \times E_{in}$$

4-26. THIS QUESTION HAS BEEN DELETED.

4-27. THIS QUESTION HAS BEEN DELETED.

4-28. In a parallel-negative limiter, how is the diode biased (a) by the positive half of the input sine wave and (b) by the negative half?

1. (a) Forward (b) forward
2. (a) Forward (b) reverse
3. (a) Reverse (b) reverse
4. (a) Reverse (b) forward

4-29. If a sine wave is applied to a parallel-negative limiter with positive bias, what is the polarity of the output voltage?

1. Positive at all times
2. Negative at all times
3. Positive during the positive portion of the input cycle and negative during the negative portion
4. Negative during the positive portion of the input cycle and positive during the negative portion

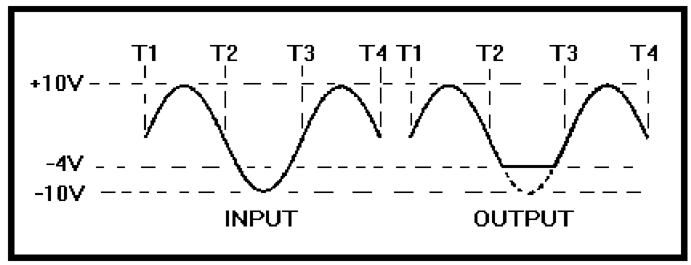


Figure 4B.—Parallel-limiter input and output waveforms.

IN ANSWERING QUESTION 4-30, REFER TO FIGURE 4B.

4-30. If the diode in the parallel limiter were reversed, what portions of the input waveform would be limited?

1. Positive peaks only
2. Negative peaks only
3. All hut the positive peaks would be limited
4. All but the negative peaks would be limited

4-31. Which of the following circuits would be used to fix the upper or lower extremity of a waveform at a specific value?

1. Clamper
2. Limiter
3. Counter
4. Amplifier

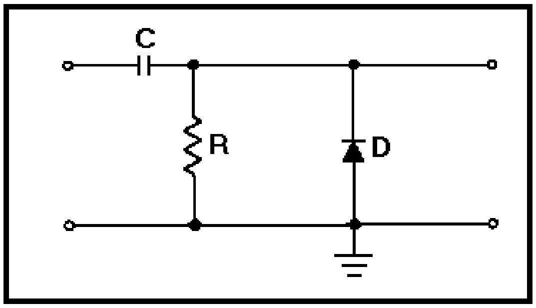


Figure 4C.—Positive clamper.

IN ANSWERING QUESTIONS 4-32 AND 4-33, REFER TO FIGURE 4C.

4-32. When a negative input signal is present, what components provide(s) the charge path for the capacitor?

1. Diode only
2. Resistor only
3. Both the resistor and diode

4-33. How does the length of time required for the capacitor to charge compare to the time for it to discharge?

1. Charge time is longer than discharge time
2. Charge time is shorter than discharge time
3. Charge and discharge times are the same

4-34. In a positive clamper with positive bias, (a) what extremity of the waveform is clamped and (b) to what potential is it clamped?

1. (a) Upper (b) positive
2. (a) Upper (b) negative
3. (a) Lower (b) positive
4. (a) Lower (b) negative

4-35. In a positive clamper with negative bias, (a) what extremity of the waveform is clamped and (b) to what potential is it clamped?

1. (a) Upper (b) positive
2. (a) Upper (b) negative
3. (a) Lower (b) negative
4. (a) Lower (b) positive

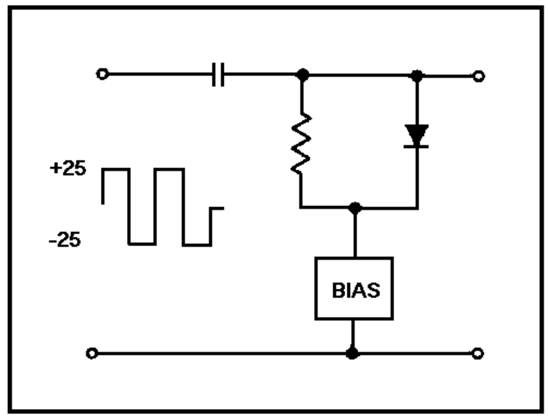


Figure 4D.—Clamper with bias.

IN ANSWERING QUESTIONS 4-36 AND 4-37, REFER TO FIGURE 4D.

4-36. With +10 volts of bias in the circuit, what is the maximum negative output voltage?

1. -15 volts
2. -25 volts
3. -40 volts
4. -50 volts

4-37. With -15 volts of bias, to what voltage level is the output clamped?

1. +75 volts
2. +45 volts
3. -15 volts
4. -75 volts

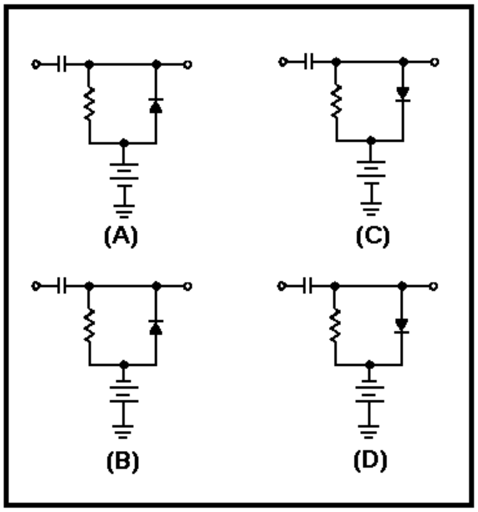


Figure 4E.—Clamper circuits.

IN ANSWERING QUESTIONS 4-38 THROUGH 4-40, SELECT THE CIRCUIT IN FIGURE 4E THAT MATCHES THE CIRCUIT NAMED IN THE QUESTIONS. CHOICES MAY BE USED ONCE, MORE THAN ONCE, OR NOT AT ALL.

4-38. Positive clamper with negative bias.

1. A
2. B
3. C
4. D

4-39. Negative clamper with negative bias.

1. A
2. B
3. C
4. D

4-40. Positive clamper with positive bias.

1. A
2. B
3. C
4. D

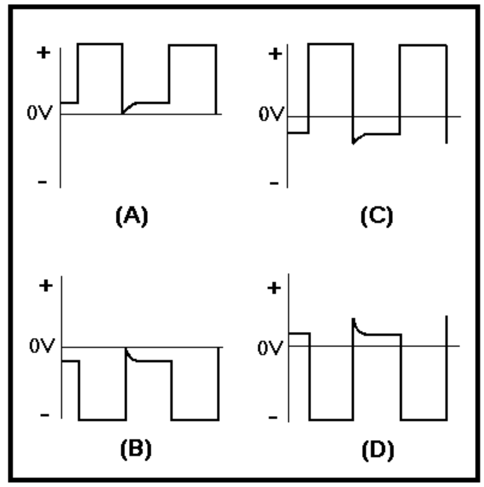


Figure 4F—Clamper circuit outputs.

IN ANSWERING QUESTIONS 4-41 THROUGH 4-43, SELECT THE OUTPUT IN FIGURE 4F WHICH IS PRODUCED BY THE CIRCUITS IN THE QUESTIONS. CHOICES MAY BE USED ONCE, MORE THAN ONCE, OR NOT AT ALL.

4-41. Positive clamper with positive bias.

1. A
2. B
3. C
4. D

4-42. Negative clamper with positive bias.

1. A
2. B
3. C
4. D

4-43. Negative clamper with negative bias.

1. A
2. B
3. C
4. D

4-44. Which of the following waves is NOT a complex wave?

1. Sine wave
2. Square wave
3. Rectangular wave
4. Trapezoidal wave

4-45. What is the harmonic content of a square wave?

1. A combination of odd harmonics only
2. A combination of even harmonics only
3. Both even and odd harmonic combinations

4-46. What is the harmonic composition within a sawtooth wave?

1. Odd harmonics only
2. Even harmonics only
3. Both even and odd harmonics

4-47. With a square wave applied to a resistive network, the circuit values of what components will NOT affect either the phases or amplitudes of the harmonics within the square wave?

1. Inductors
2. Resistors
3. Capacitors
4. Transformers

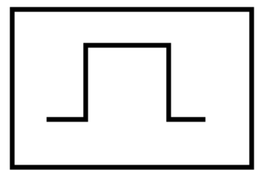
4-48. The time constant for full integration in an RC circuit should be what minimum number of times greater than the input-pulse duration?

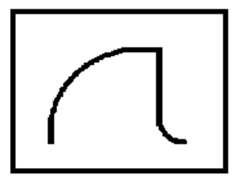
1. 1
2. 5
3. 10
4. 20

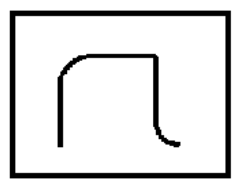
4-49. Integration in a circuit takes place when the output is taken (a) across the capacitor in what type of circuit and (b) across the resistor in what type of circuit?

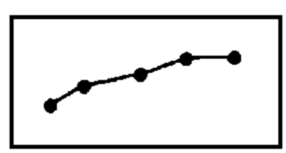
1. (a) Resistive only
   (b) Resistive only
2. (a) Resistive only
   (b) Resistive-inductive
3. (a) Resistive only
   (b) Resistive-capacitive
4. (a) Resistive-capacitive
   (b) Resistive-inductive

4-50. In an RC integrator, which of the following waveforms has the longest time constant?

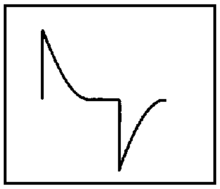
1. 

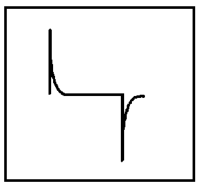
2. 

3. 

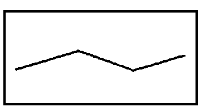
4. 

4-51. In an RC differentiator, which of the following waveforms has the shortest time constant?

1. 

2. 

3. 

4. 

4-52. In a short time-constant integrator circuit, the maximum amplitude of the input pulse is 100 volts and the time constant of the circuit is 1/10 the duration of the input pulse. At the end of three time constants, what is the maximum voltage across the capacitor?

1. 36.8 volts
2. 63.2 volts
3. 86.5 volts
4. 95 volts

4-53. In a medium time-constant circuit, the maximum amplitude of the input pulse is 100 volts and the pulse length is one time constant. At the end of two time constants, what is the maximum voltage across the capacitor?

1. 23.3 volts
2. 48.4 volts
3. 71.7 volts
4. 100 volts

4-54. In a short time-constant differentiator circuit, the maximum amplitude of the input pulse is 100 volts and the time constant of the circuit is 1/10 that of the input pulse. At the end of four time constants, what is the maximum voltage across the resistor?

1. 5 volts
2. 2 volts
3. 23.3 volts
4. 48.4 volts

4-55. In an RC differentiator circuit, the time constant for the circuit and the input pulse are equal. At the end of one time constant, to what maximum percentage of the applied voltage is the capacitor charged?

1. 5 percent
2. 13.5 percent
3. 36.8 percent
4. 63.2 percent

4-56. In an RC network that is used as a coupling circuit, (a) across what component is the output normally taken and (b) what is the relative length of the time constant?

1. (a) Capacitor (b) short
2. (a) Capacitor (b) long
3. (a) Resistor (b) long
4. (a) Resistor (b) short

4-57. In a positive-diode counter that provides accurate counting, what is the only variable allowed in the input signal?

1. Pulse width
2. Pulse duration
3. Pulse amplitude
4. Pulse-repetition frequency

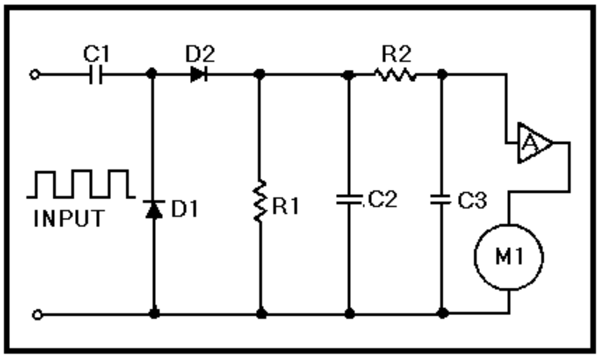


Figure 4G.—Basic frequency counter.

IN ANSWERING QUESTIONS 4-58 THROUGH 4-60, REFER TO FIGURE 4G.

4-58. An input signal to the counter will cause (a) what capacitor to charge through R1 and D2 and (b) what capacitor to discharge through D1?

1. (a) C1 (b) C1
2. (a) C1 (b) C2
3. (a) C2 (b) C1
4. (a) C2 (b) C2

4-59. What components produce the smooth dc output-voltage level?

1. C1, D1, and R1
2. C1, D2, and R1
3. C1, C2, and R1
4. C2, C3, and R2

4-60. When the input frequency is increased, what is the effect on (a) the input interval, (b) the number of pulses per given time and (c) the dc output voltage?

1. (a) Longer (b) less (c) higher
2. (a) Longer (b) more (c) lower
3. (a) Shorter (b) less (c) higher
4. (a) Shorter (b) more (c) higher

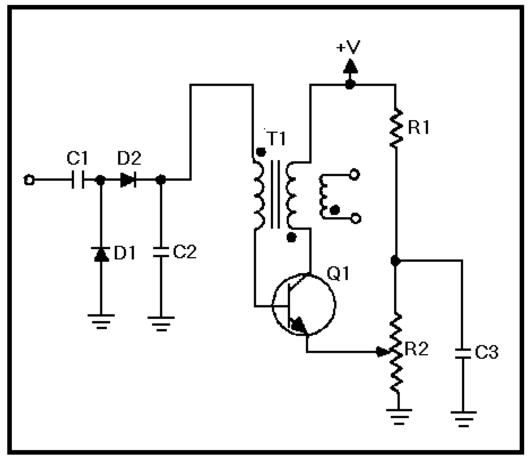


Figure 4H.—Step-counter frequency divider.

IN ANSWERING QUESTIONS 4-61 AND 4-62, REFER TO FIGURE 4H.

4-61. An output pulse will occur when the charge on C2

1. builds a magnetic field in the base winding
2. drops below the cutoff-bias level of Q1
3. exceeds the bias level of Q1
4. reaches the source potential of the voltage applied

4-62. What component develops the bias for Q1?

1. R1
2. R2
3. C2
4. T1

NONRESIDENT
TRAINING
COURSE

Navy Electricity And Electronics Training Series

Module 10—Introduction to Wave Propagation, Transmission Lines, and Antennas

NAVEDTRA 14182

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 10 of a series.

History of the course:

Sep 1998: Original edition released.
May 2003: Administrative update released. Entered administrative updates. Technical content was *not revised.*

NAVSUP Logistics Tracking Number
0504-LP-026-8350

TABLE OF CONTENTS

ASSIGNMENT QUESTIONS follow Index.

CHAPTER 1

WAVE PROPAGATION

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the NRTC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completion of this chapter, you should be able to:

1. State what wave motion is, define the terms reflection, refraction, and diffraction, and describe the Doppler effect.
2. State what sound waves are and define a propagating medium.
3. List and define terms as applied to sound waves, such as cycle, frequency, wavelength, and velocity.
4. List the three requirements for sound.
5. Define pitch, intensity, loudness, and quality and their application to sound waves.
6. State the acoustical effects that echoes, reverberation, resonance, and noise have on sound waves.
7. Define light waves and list their characteristics.
8. List the various colors of light and define the terms reflection, refraction, diffusion, and absorption as applied to light waves.
9. State the difference between sound waves and light waves.
10. State the electromagnetic wave theory and list the components of the electromagnetic wave.

INTRODUCTION TO WAVE PROPAGATION

Of the many technical subjects that naval personnel are expected to know, probably the one least susceptible to change is the theory of wave propagation. The basic principles that enable waves to be propagated (transmitted) through space are the same today as they were 70 years ago. One would think, then, that a thorough understanding of these principles is a relatively simple task. For the electrical engineer or the individual with a natural curiosity for the unknown, it is indeed a simple task. Most technicians, however, tend to view wave propagation as something complex and confusing, and would just as soon see this chapter completely disappear from training manuals. This attitude undoubtedly stems from the fact that wave propagation is an invisible force that cannot be detected by the sense of sight or touch. Understanding wave propagation requires the use of the imagination to visualize the associated concepts and how they are used in practical application. This manual was developed to help you visualize and understand those concepts. Through ample use of illustrations and a step-by-step transition from the simple to the complex, we will help you develop a better understanding of wave propagation. In this chapter, we will discuss propagation theory on an introductory level, without going into the technical details that concern the engineer. However, you must still use thought and imagination to understand the new ideas and concepts as they are presented.

To understand radio wave propagation, you must first learn what wave propagation is and some of the basic physics or properties that affect propagation. Many of these properties are common everyday occurrences, with which you are already familiar.

WHAT IS PROPAGATION?

Early man was quick to recognize the need to communicate beyond the range of the human voice. To satisfy this need, he developed alternate methods of communication, such as hand gestures, beating on a hollow log, and smoke signals. Although these methods were effective, they were still greatly limited in range. Eventually, the range limitations were overcome by the development of courier and postal systems; but there was then a problem of speed. For centuries the time required for the delivery of a message depended on the speed of a horse.

During the latter part of the 19th century, both distance and time limitations were largely overcome. The invention of the telegraph made possible instantaneous communication over long wires. Then a short time later, man discovered how to transmit messages in the form of RADIO WAVES.

As you will learn in this chapter, radio waves are propagated. PROPAGATION means "movement through a medium." This is most easily illustrated by light rays. When a light is turned on in a darkened room, light rays travel from the light bulb throughout the room. When a flashlight is turned on, light rays also radiate from its bulb, but are focused into a narrow beam. You can use these examples to picture how radio waves propagate. Like the light in the room, radio waves may spread out in all directions. They can also be focused (concentrated) like the flashlight, depending upon the need. Radio waves are a form of radiant energy, similar to light and heat. Although they can neither be seen nor felt, their presence can be detected through the use of sensitive measuring devices. The speed at which both forms of waves travel is the same; they both travel at the speed of light.

You may wonder why you can see light but not radio waves, which consist of the same form of energy as light. The reason is that you can only "see" what your eyes can detect. Your eyes can detect radiant energy only within a fixed range of frequencies. Since the frequencies of radio waves are below the frequencies your eyes can detect, you cannot see radio waves.

The theory of wave propagation that we discuss in this module applies to Navy electronic equipment, such as radar, navigation, detection, and communication equipment. We will not discuss these individual systems in this module, but we will explain them in future modules.

Q1. What is propagation?

PRINCIPLES OF WAVE MOTION

All things on the earth—on the land, or in the water—are showered continually with waves of energy. Some of these waves stimulate our senses and can be seen, felt, or heard. For instance, we can see light, hear sound, and feel heat. However, there are some waves that do not stimulate our senses. For example, radio waves, such as those received by our portable radio or television sets, cannot be seen, heard, or felt. A device must be used to convert radio waves into light (TV pictures) and sound (audio) for us to sense them.

A WAVE can be defined as a DISTURBANCE (sound, light, radio waves) that moves through a MEDIUM (air, water, vacuum). To help you understand what is meant by "a disturbance which moves through a medium," picture the following illustration. You are standing in the middle of a wheat field. As the wind blows across the field toward you, you can see the wheat stalks bending and rising as the force of the wind moves into and across them. The wheat appears to be moving toward you, but it isn't. Instead, the stalks are actually moving back and forth. We can then say that the "medium" in this illustration is the wheat and the "disturbance" is the wind moving the stalks of wheat.

WAVE MOTION can be defined as a recurring disturbance advancing through space with or without the use of a physical medium. Wave motion, therefore, is a means of moving or transferring energy from one point to another point. For example, when sound waves strike a microphone, sound energy is converted into electrical energy. When light waves strike a phototransistor or radio waves strike an antenna, they are likewise converted into electrical energy. Therefore, sound, light, and radio waves are all forms of energy that are moved by wave motion. We will discuss sound waves, light waves, and radio waves later.

Q2. How is a wave defined as it applies to wave propagation?

Q3. What is wave motion?

Q4. What are some examples of wave motion?

WAVE MOTION IN WATER

A type of wave motion familiar to almost everyone is the movement of waves in water. We will explain these waves first to help you understand wave motion and the terms used to describe it.

Basic wave motion can be shown by dropping a stone into a pool of water (see figure 1-1). As the stone enters the water, a surface disturbance is created, resulting in an expanding series of circular waves. Figure 1-2 is a diagram of this action. View A shows the falling stone just an instant before it strikes the water. View B shows the action taking place at the instant the stone strikes the surface, pushing the water that is around it upward and outward. In view C, the stone has sunk deeper into the water, which has closed violently over it causing some spray, while the leading wave has moved outward. An instant later, the stone has sunk out of sight, leaving the water disturbed as shown in view D. Here the leading wave has continued to move outward and is followed by a series of waves gradually diminishing in amplitude. Meanwhile, the disturbance at the original point of contact has gradually subsided.

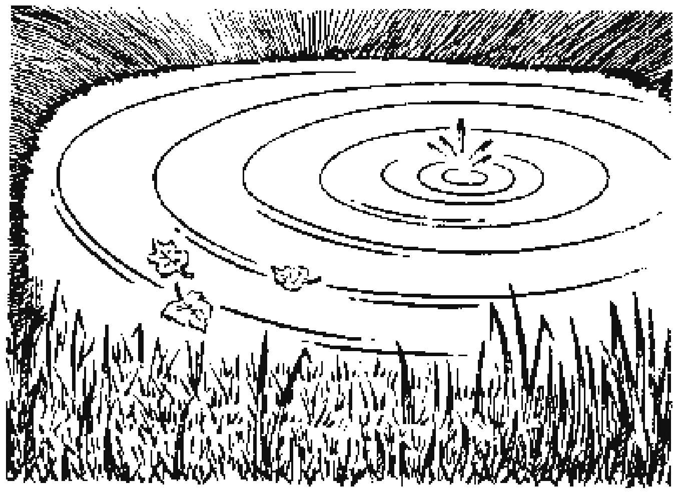


Figure 1-1.—Formation of waves in water.

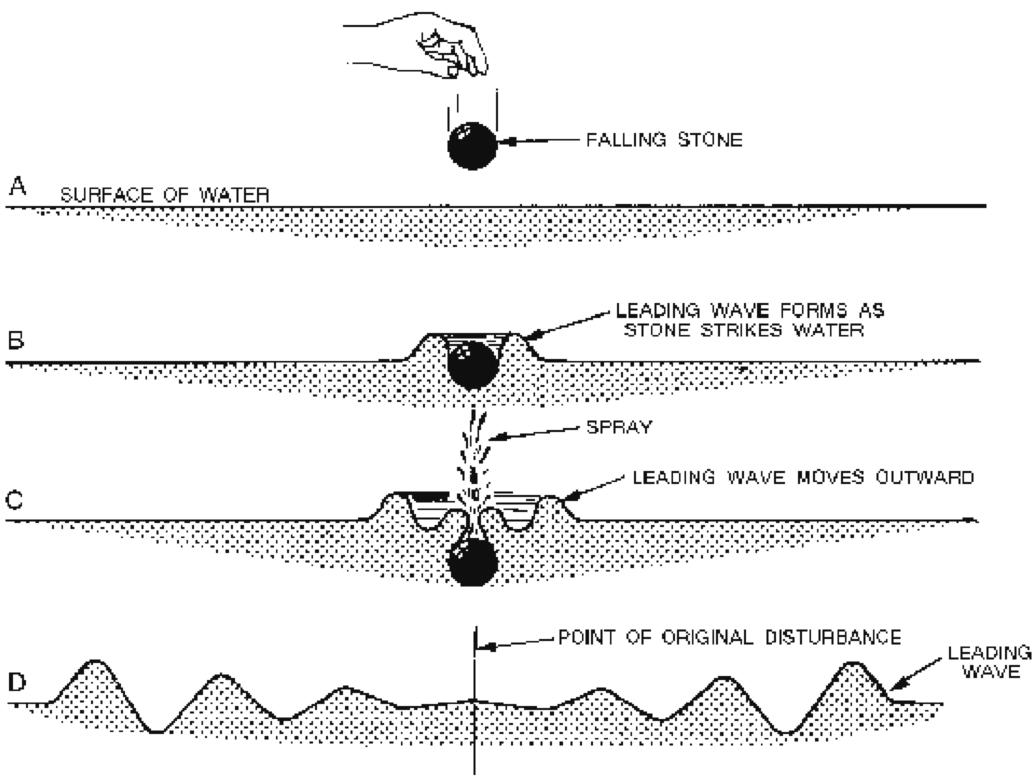


Figure 1-2.—How a falling stone creates wave motion to the surface of water.

In this example, the water is not actually being moved outward by the motion of the waves, but up and down as the waves move outward. The up and down motion is transverse, or at right angles, to the outward motion of the waves. This type of wave motion is called TRANSVERSE WAVE MOTION.

Q5. What type of wave motion is represented by the motion of water?

TRANSVERSE WAVES

To explain transverse waves, we will again use our example of water waves. Figure 1-3 is a cross section diagram of waves viewed from the side. Notice that the waves are a succession of crests and troughs. The wavelength (one 360 degree cycle) is the distance from the crest of one wave to the crest of the next, or between any two similar points on adjacent waves. The amplitude of a transverse wave is half the distance measured vertically from the crest to the trough. Water waves are known as transverse waves because the motion of the water is up and down, or at right angles to the direction in which the waves are traveling. You can see this by observing a cork bobbing up and down on water as the waves pass by; the cork moves very little in a sideways direction. In figure 1-4, the small arrows show the up-and-down direction the cork moves as the transverse wave is set in motion. The direction the wave travels is shown by the large arrow. Radio waves, light waves, and heat waves are examples of transverse waves.

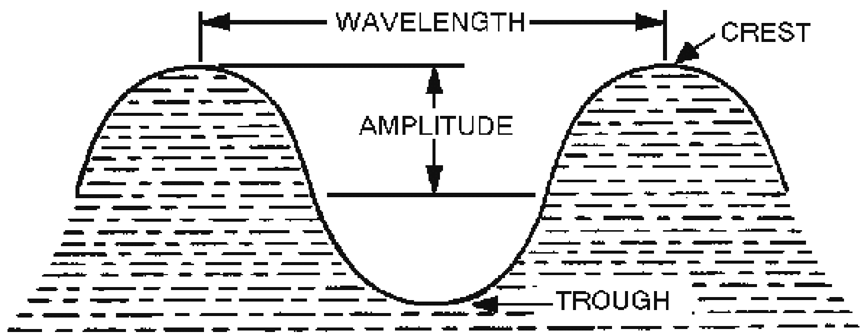


Figure 1-3.—Elements of a wave.

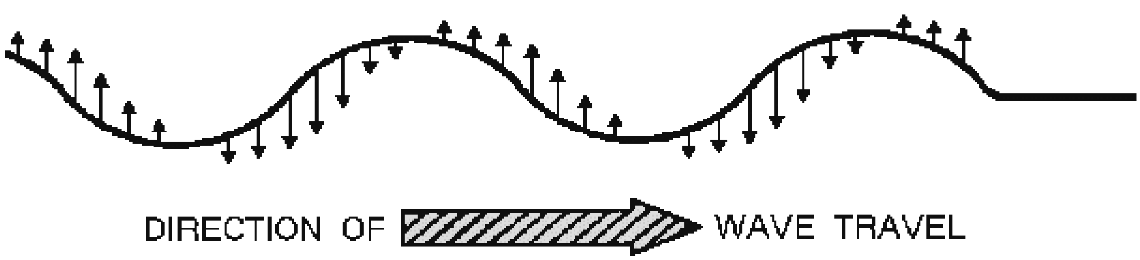


Figure 1-4.—Transverse wave.

LONGITUDINAL WAVES

In the previous discussion, we listed radio waves, light waves, and heat waves as examples of transverse waves, but we did not mention sound waves. Why? Simply because sound waves are LONGITUDINAL WAVES. Unlike transverse waves, which travel at right angles to the direction of propagation, sound waves travel back and forth in the same direction as the wave motion. Therefore, longitudinal waves are waves in which the disturbance takes place in the direction of propagation. Longitudinal waves are sometimes called COMPRESSION WAVES.

Waves that make up sound, such as those set up in the air by a vibrating tuning fork, are longitudinal waves. In figure 1-5, the tuning fork, when struck, sets up vibrations. As the tine moves in an outward direction, the air immediately in front of it is compressed (made more dense) so that its momentary pressure is raised above that at other points in the surrounding medium (air). Because air is elastic, the disturbance is transmitted in an outward direction as a COMPRESSION WAVE. When the tine returns and moves in the inward direction, the air in front of the tine is rarefied (made less dense or expanded) so that its pressure is lowered below that of the other points in the surrounding air. The rarefied wave is propagated from the tuning fork and follows the compressed wave through the medium (air).

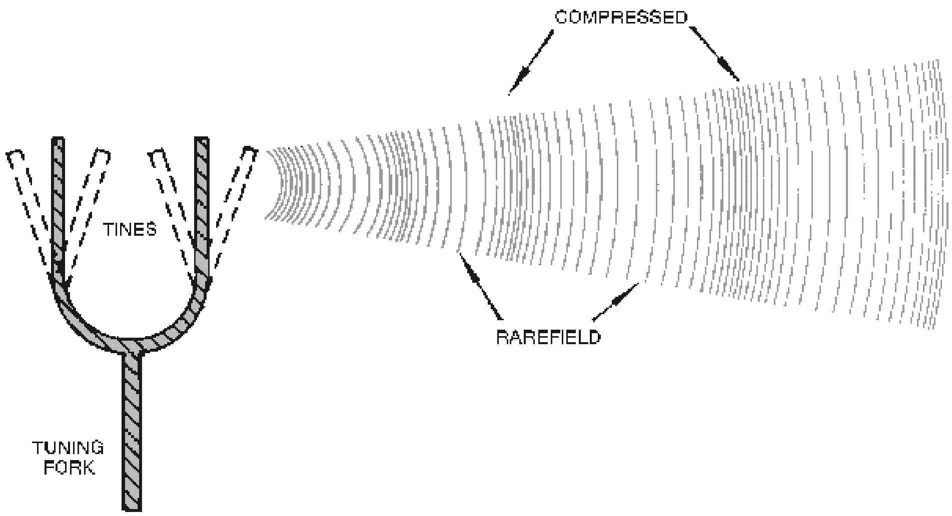


Figure 1-5.—Sound propagation by a tuning fork.

Q6. What are some examples of transverse waves?

Q7. What example of a longitudinal wave was given in the text?

MEDIUM

We have used the term *medium* in describing the motion of waves. Since *medium* is a term that is used frequently in discussing propagation, it needs to be defined so you will understand what a medium is and its application to propagation.

A MEDIUM is the vehicle through which the wave travels from one point to the next. The vehicle that carries a wave can be just about anything. An example of a medium, already mentioned, is air. Air, as defined by the dictionary, is the mixture of invisible, odorless, tasteless gases that surrounds the earth (the atmosphere). Air is made up of molecules of various gases (and impurities). We will call these molecules of air *particles of air* or simply *particles*.

Figure 1-6 will help you to understand how waves travel through air. The object producing the waves is called the SOURCE—a bell in this illustration. The object responding to the waves is called a DETECTOR or RECEIVER—in this case, the human ear. The medium is air, which is the means of conveying the waves from the source to the detector. The source, detector, and medium are all necessary for wave motion and wave propagation (except for electromagnetic waves which require no medium). The waves shown in figure 1-6 are sound waves. As the bell is rung, the particles of air around the bell are compressed and then expanded. This compression and expansion of particles of air set up a wave motion in the air. As the waves are produced, they carry energy from particle to particle through the medium (air) to the detector (ear).

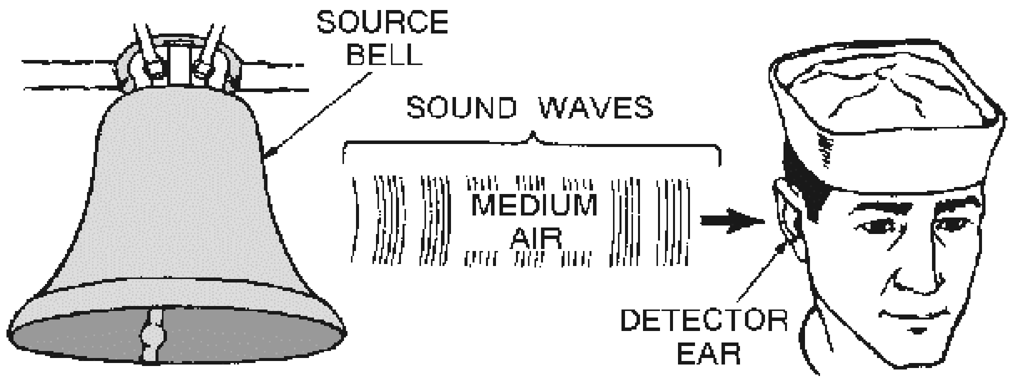


Figure 1-6.—The three elements of sound.

Q8. What are the three requirements for a wave to be propagated?

TERMS USED IN WAVE MOTION

There are a number of special terms concerning waves that you should know. Many of the terms, such as CYCLE, WAVELENGTH, AMPLITUDE, and FREQUENCY were introduced in previous *NEETS* modules. We will now discuss these terms in detail as they pertain to wave propagation. Before we begin our discussion, however, note that in the figure, wave 1 and wave 2 have equal frequency and wavelength but different amplitudes. The REFERENCE LINE (also known as REST POSITION or POINT OF ZERO DISPLACEMENT) is the position that a particle of matter would have if it were not disturbed by wave motion. For example, in the case of the water wave, the reference line is the level of the water when no wave motion is present. With this in mind, let's go on to our discussion of the four terms, as shown in figure 1-7.

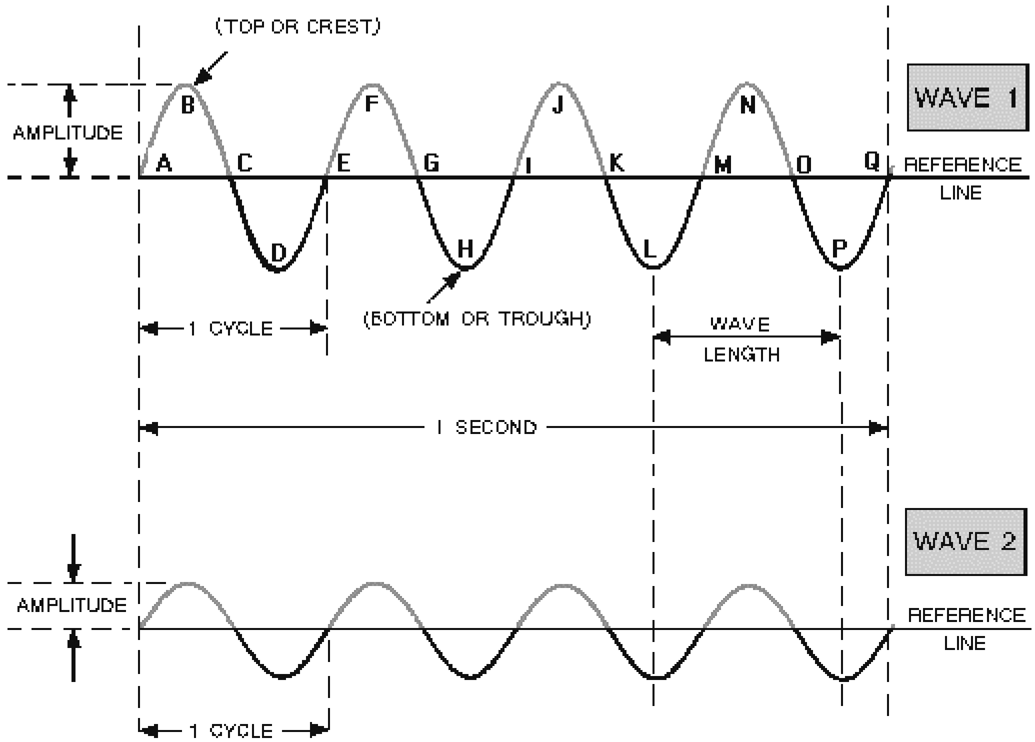


Figure 1-7.—Comparison of waves with different amplitudes.

Cycle

Refer to wave 1 in figure 1-7. Notice how similar it is to the sine wave you have already studied. All transverse waves appear as sine waves when viewed from the side. In figure 1-7, wave 1 has four complete cycles. Points ABCDE comprise one complete cycle having a maximum value above and a maximum value below the reference line. The portion above the reference line (between points A and C) is called a POSITIVE ALTERNATION and the portion below the reference line (between points C and E) is known as a NEGATIVE ALTERNATION. The combination of one complete positive and one complete negative alternation represents one cycle of the wave. At point E, the wave begins to repeat itself with a second cycle completed at point I, a third at point M, etc. The peak of the positive alternation (maximum value above the line) is sometimes referred to as the TOP or CREST, and the peak of the negative alternation (maximum value below the line) is sometimes called the BOTTOM or TROUGH, as depicted in the figure. Therefore, one cycle has one crest and one trough.

Wavelength

A WAVELENGTH is the distance in space occupied by one cycle of a radio wave at any given instant. If the wave could be frozen in place and measured, the wavelength would be the distance from the leading edge of one cycle to the corresponding point on the next cycle. Wavelengths vary from a few hundredths of an inch at extremely high frequencies to many miles at extremely low frequencies; however, common practice is to express wavelengths in meters. Therefore, in figure 1-7 (wave 1), the distance between A and E, or B and F, etc., is one wavelength. The Greek letter lambda (λ) is used to signify wavelength. Why lambda and not "l" or "L"? This is because "L" is used conventionally as the symbol for inductance, and "l" is used for dimensional length; therefore, λ; is used to indicate the length of waves.

Amplitude

Two waves may have the same wavelength, but the crest of one may rise higher above the reference line than the crest of the other. Compare wave 1 and wave 2 of figure 1-7 again. The height of a wave crest above the reference line is called the AMPLITUDE of the wave. The amplitude of a wave gives a relative indication of the amount of energy the wave transmits. A continuous series of waves, such as A through Q, having the same amplitude and wavelength, is called a train of waves or WAVE TRAIN.

Frequency and Time

Time is an important factor in wave studies. When a wave train passes through a medium, a certain number of individual waves pass a given point in a specific unit of time. For example, if a cork on a water wave rises and falls once every second, the wave makes one complete up-and-down vibration every second. The number of vibrations, or cycles, of a wave train in a unit of time is called the FREQUENCY of the wave train and is measured in HERTZ. If 5 waves pass a point in one second, the frequency of the wave train is 5 cycles per second. In figure 1-7, the frequency of both wave 1 and wave 2 is four cycles per second (cycles per second is abbreviated as cps).

In 1967, in honor of the German physicist Heinrich Hertz, the term HERTZ was designated for use in lieu of the term "cycle per second" when referring to the frequency of radio waves. It may seem confusing that in one place the term "cycle" is used to designate the positive and negative alternations of a wave, but in another instance the term "hertz" is used to designate what appears to be the same thing. The key is the time factor. The term cycle refers to any sequence of events, such as the positive and negative alternations, comprising one cycle of electrical current. The term hertz refers to the number of occurrences that take place in one second.

Q9. What is a cycle?

Q10. What is wavelength (λ)?

CHARACTERISTICS OF WAVE MOTION

The two types of wave motion, transverse and longitudinal, have many of the same characteristics, such as frequency, amplitude, and wavelength. Another important characteristic that these two types of wave motion share is VELOCITY. Velocity of propagation is the rate at which the disturbance travels through the medium, or the velocity with which the crest of the wave moves along. The velocity of the wave depends both on the type of wave (light, sound, or radio) and type of medium (air, water, or metal). If longitudinal waves are plotted as a graph, they appear as transverse waves. This fact is illustrated in figure 1-8.

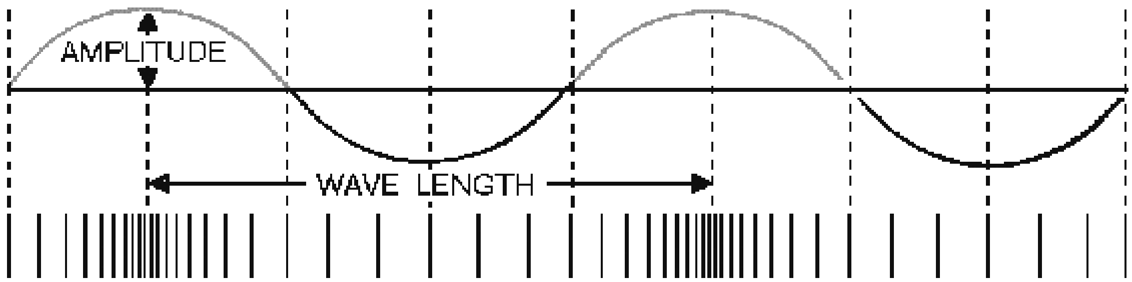


Figure 1-8.—Longitudinal wave represented graphically by a transverse wave.

The frequency of a longitudinal wave, like that of a transverse wave, is the number of complete cycles the wave makes during a specific unit of time. The higher the frequency, the greater is the number of compressions and expansions per unit of time.

In the two types of wave motion described in the preceding discussion, the following quantities are of interest:

a. The PERIOD, which is the time (T) in which one complete vibratory cycle of events occurs,

b. The FREQUENCY OF VIBRATION (f), which is the number of cycles taking place in one second, and

c. The WAVELENGTH, which is the distance the disturbance travels during one period of vibration.

Now, consider the following concept. If a vibrating object makes a certain number of vibrations per second, then 1 second divided by the number of vibrations is equal to the period of time of 1 vibration. In other words, the period, or time, of 1 vibration is the reciprocal of the frequency; thus,

$$\text{time (T) of one vibration} = \frac{1}{\text{frequency (f)}}$$

or

$$T = \frac{1}{f}$$

If you know the velocity of a wave, you can determine the wavelength by dividing the velocity by the frequency. As an equation:

Where:

$$\lambda = \frac{v}{f}$$

λ = wavelength
v = velocity of propagation
f = frequency of vibration

When you use the above equation, be careful to express velocity and wavelength in the proper units of length. For example, in the English system, if the velocity (expressed in feet per second) is divided by the frequency (expressed in cycles per second, or Hz), the wavelength is given in feet per cycle. If the metric system is used and the velocity (expressed in meters per second) is divided by the frequency (expressed in cycles per second), the wavelength is given in meters per cycle. Be sure to express both the wavelength and the frequency in the same units. (Feet per cycle and meters per cycle are normally abbreviated as feet or meters because one wavelength indicates one cycle.) Because this equation holds true for both transverse and longitudinal waves, it is used in the study of both electromagnetic waves and sound waves.

Consider the following example. Two cycles of a wave pass a fixed point every second, and the velocity of the wave train is 4 feet per second. What is the wavelength? The formula for determining wavelength is as follows:

$$\lambda = \frac{v}{f}$$

Where:

λ = wavelength in feet
v = velocity in feet per second
f = frequency in Hz

Given:

v = 4 feet per second
f = 2 Hz

Solution:

$$\lambda = \frac{v}{f}$$

$$\lambda = \frac{4 \text{ feet per second}}{2 \text{ Hz}}$$

$$\lambda = 2 \text{ feet}$$

NOTE: In problems of this kind, be sure NOT to confuse wave velocity with frequency. FREQUENCY is the number of cycles per unit of time (Hz). WAVE VELOCITY is the speed with which a wave train passes a fixed point.

Here is another problem. If a wave has a velocity of 1,100 feet per second and a wavelength of 30 feet, what is the frequency of the wave?

By transposing the general equation:

By transposing the general equation:

$$f = \frac{v}{\lambda}$$

We have the equation:

$$\lambda = \frac{v}{f}$$

Given:

v = 1,100 feet per second
λ = 30 feet

Solution:

$$f = \frac{1{,}100 \text{ feet per second}}{30 \text{ feet}}$$

$$f = 36.67\text{Hz}$$

To find the velocity, rewrite the equation as:

$$v = \lambda f$$

Let's work one more problem, this time using the metric system.

Suppose the wavelength is 0.4 meters and the frequency is 12 kHz. What is the velocity?

Use the formula:

velocity = wavelength × frequency ($v = \lambda f$)

Given:

wavelength (λ) = 0.4 meters
frequency (f) = 12kHz

Solution:

$$v = \lambda \times f$$
$$v = 0.4 \text{ meters} \times 12{,}000\text{Hz}$$
$$v = 4{,}800 \text{ meters per second}$$

Other important characteristics of wave motion are reflection, refraction, diffraction, and the Doppler effect. Big words, but the concept of each is easy to see. For ease of understanding, we will explain the first two characteristics using light waves, and the last two characteristics using sound waves. You should keep in mind that all waves react in a similar manner.

Within mediums, such as air, solids, or gases, a wave travels in a straight line. When the wave leaves the boundary of one medium and enters the boundary of a different medium, the wave changes direction. For our purposes in this module, a boundary is an imaginary line that separates one medium from another.

When a wave passes through one medium and encounters a medium having different characteristics, three things can occur to the wave: (1) Some of the energy can be reflected back into the initial medium; (2) some of the energy can be transmitted into the second medium where it may continue at a different velocity; or (3) some of the energy can be absorbed by the medium. In some cases, all three processes (reflection, transmission, and absorption) may occur to some degree.

Reflection

REFLECTION WAVES are simply waves that are neither transmitted nor absorbed, but are reflected from the surface of the medium they encounter. If a wave is directed against a reflecting surface, such as a mirror, it will reflect or "bounce" from the mirror. Refer to figure 1-9. A wave directed *toward* the surface of the mirror is called the INCIDENT wave. When the wave bounces off of the mirror, it becomes known as the REFLECTED wave. An imaginary line perpendicular to the mirror at the point at which the incident wave strikes the mirror's surface is called the NORMAL, or perpendicular. The angle between the incident wave and the normal is called the ANGLE OF INCIDENCE. The angle between the reflected wave and the normal is called the ANGLE OF REFLECTION.

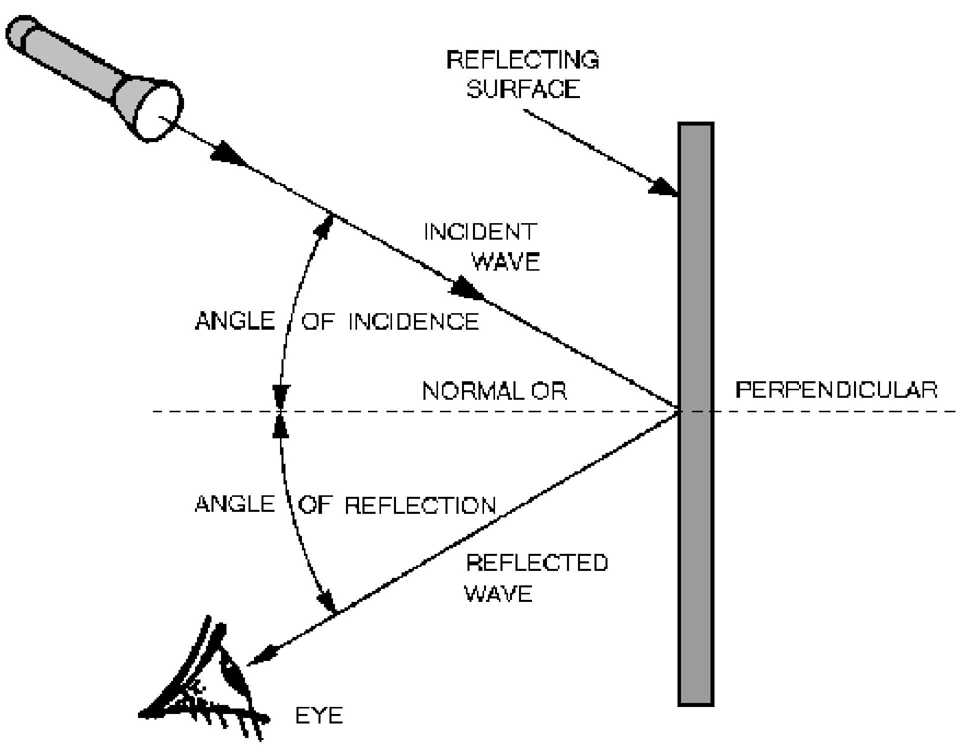


Figure 1-9.—Reflection of a wave.

If the reflecting surface is smooth and polished, the angle between the incident ray and the normal will be the same as the angle between the reflected ray and the normal. This conforms to the law of reflection which states: The angle of incidence is equal to the angle of reflection.

The amount of incident wave energy reflected from a given surface depends on the nature of the surface and the angle at which the wave strikes the surface. As the angle of incidence increases, the amount of wave energy reflected increases. The reflected energy is the greatest when the wave is nearly parallel to the reflecting surface. When the incident wave is perpendicular to the surface, more of the energy is transmitted into the substance and less is reflected. At any incident angle, a mirror reflects almost all of the wave energy, while a dull, black surface reflects very little.

Q11. What is the law of reflection?

Q12. When a wave is reflected from a surface, energy is transferred. When is the transfer of energy greatest?

Q13. When is the transfer of energy minimum?

Refraction

When a wave passes from one medium into another medium that has a different velocity of propagation, a change in the direction of the wave will occur. This changing of direction as the wave enters the second medium is called REFRACTION. As in the discussion of reflection, the wave striking the boundary (surface) is called the INCIDENT WAVE, and the imaginary line perpendicular to the boundary is called the NORMAL. The angle between the incident wave and the normal is called the ANGLE OF INCIDENCE. As the wave passes through the boundary, it is bent either toward or away from the normal. The angle between the normal and the path of the wave through the second medium is the ANGLE OF REFRACTION.

A light wave passing through a block of glass is shown in figure 1-10. The wave moves from point A to B at a constant speed. This is the incident wave. As the wave penetrates the glass boundary at point B, the velocity of the wave is slowed down. This causes the wave to bend toward the normal. The wave then takes the path from point B to C through the glass and becomes BOTH the refracted wave from the top surface and the incident wave to the lower surface. As the wave passes from the glass to the air (the second boundary), it is again refracted, this time away from the normal and takes the path from point C to D. As the wave passes through the last boundary, its velocity increases to the original velocity. As figure 1-10 shows, refracted waves can bend toward or away from the normal. This bending depends on the velocity of the wave through each medium. The broken line between points B and E is the path that the wave would travel if the two mediums (air and glass) had the same density.

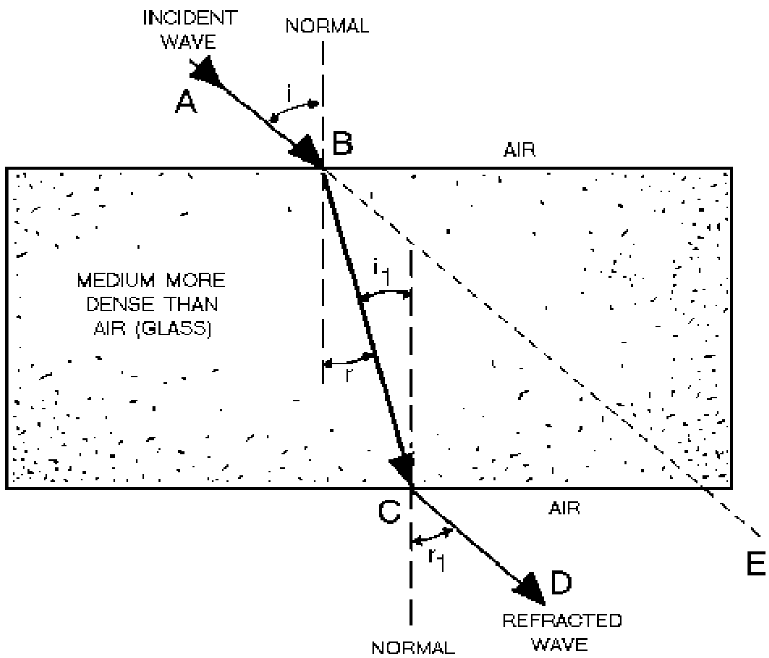


Figure 1-10.—Refraction of a wave.

To summarize what figure 1-10 shows:

1. If the wave passes from a less dense medium to a more dense medium, it is bent toward the normal, and the angle of refraction (r) is less than the angle of incidence (i).

2. If the wave passes from a more dense to a less dense medium, it is bent away from the normal, and the angle of refraction (r_1) is greater than the angle of incidence (i_1).

You can more easily understand refraction by looking at figure 1-11. There is a plowed field in the middle of a parade ground. Think of the incident wave as a company of recruits marching four abreast at an angle across the parade ground to the plowed field, then crossing the plowed field and coming out on the other side onto the parade ground again. As the recruits march diagonally across the parade ground and begin to cross the boundary onto the plowed field, the front line is slowed down. Because the recruits arrive at the boundary at different times, they will begin to slow down at different times (number 1 slows down first and number 4 slows down last in each line). The net effect is a bending action. When the recruits leave the plowed field and reenter the parade ground, the reverse action takes place.

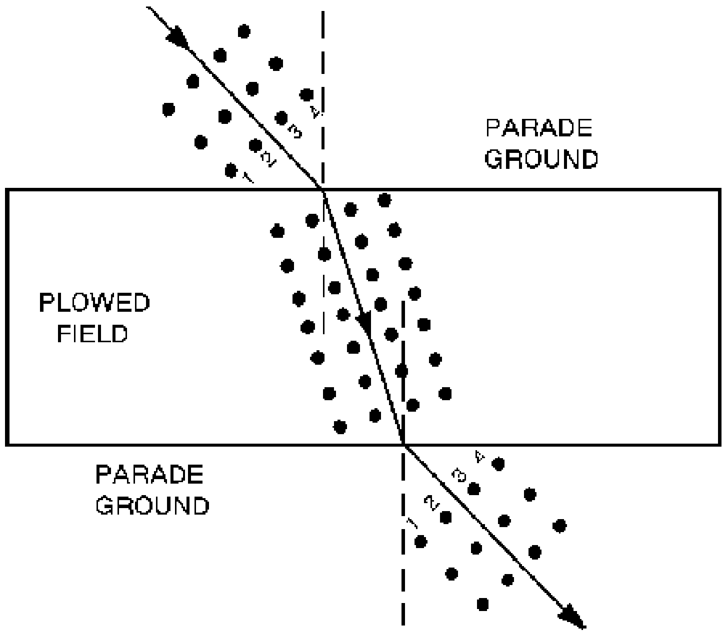


Figure 1-11.—Analogy of refraction.

Q14. A refracted wave occurs when a wave passes from one medium into another medium. What determines the angle of refraction?

Diffraction

DIFFRACTION is the bending of the wave path when the waves meet an obstruction. The amount of diffraction depends on the wavelength of the wave. Higher frequency waves are rarely diffracted in the normal world that surrounds us. Since light waves are high frequency waves, you will rarely see light diffracted. You can, however, observe diffraction in sound waves by listening to music. Suppose you are outdoors listening to a band. If you step behind a solid obstruction, such as a brick wall, you will hear mostly low notes. This is because the higher notes, having short wave lengths, undergo little or no diffraction and pass by or over the wall without wrapping around the wall and reaching your ears. The low notes, having longer wavelengths, wrap around the wall and reach your ears. This leads to the general statement that lower frequency waves tend to diffract more than higher frequency waves. Broadcast band (AM band) radio waves (lower frequency waves) often travel over a mountain to the opposite side from their source because of diffraction, while higher frequency TV and FM signals from the same source tend to be stopped by the mountain.

Doppler Effect

The last, but equally important, characteristic of a wave that we will discuss is the Doppler effect. The DOPPLER EFFECT is the apparent change in frequency or pitch when a sound source moves either toward or away from the listener, or when the listener moves either toward or away from the sound source. This principle, discovered by the Austrian physicist Christian Doppler, applies to all wave motion.

The apparent change in frequency between the source of a wave and the receiver of the wave is because of relative motion between the source and the receiver. To understand the Doppler effect, first assume that the frequency of a sound from a source is held constant. The wavelength of the sound will also remain constant. If both the source and the receiver of the sound remain stationary, the receiver will hear the same frequency sound produced by the source. This is because the receiver is receiving the same number of waves per second that the source is producing. Now, if either the source or the receiver or both move toward the other, the receiver will perceive a higher frequency sound. This is because the receiver will receive a greater number of sound waves per second and interpret the greater number of waves as a higher frequency sound. Conversely, if the source and the receiver are moving apart, the receiver will receive a smaller number of sound waves per second and will perceive a lower frequency sound. In both cases, the frequency of the sound produced by the source will have remained constant.

For example, the frequency of the whistle on a fast-moving train sounds increasingly higher in pitch as the train is approaching than when the train is departing. Although the whistle is generating sound waves of a constant frequency, and though they travel through the air at the same velocity in all directions, the distance between the approaching train and the listener is decreasing. As a result, each wave has less distance to travel to reach the observer than the wave preceding it. Thus, the waves arrive with decreasing intervals of time between them.

These apparent changes in frequency, called the Doppler effect, affect the operation of equipment used to detect and measure wave energy. In dealing with electromagnetic wave propagation, the Doppler principle is used in equipment such as radar, target detection, weapons control, navigation, and sonar.

Q15. The apparent change in frequency or pitch because of motion is explained by what effect?

SOUND WAVES

The study of sound is important because of the role sound plays in the depth finding equipment (fathometer) and underwater detection equipment (sonar) used by the Navy.

As you know, sound travels through a medium by wave motion. Although sound waves and the electromagnetic waves used in the propagation of radio and radar differ, both types of waves have many of the same characteristics. Studying the principles of sound-wave motion will help you understand the actions of both sound waves and the more complex radio and radar electromagnetic waves. The major differences among sound waves, heat waves, and light waves are (1) their frequencies; (2) their types; the mediums through which they travel; and the velocities at which they travel.

SOUND—WHAT IS IT?

The word SOUND is used in everyday speech to signify a variety of things. One definition of sound is the sensation of hearing. Another definition refers to a stimulus that is capable of producing the sensation of hearing. A third definition limits sound to what is actually heard by the human ear.

In the study of physics, sound is defined as *a range of compression-wave frequencies to which the human ear is sensitive*. For the purpose of this chapter, however, we need to broaden the definition of sound to include compression waves that are not always audible to the human ear. To distinguish frequencies in the audible range from those outside that range, the words SONIC, ULTRASONIC, and INFRASONIC are used. Sounds capable of being heard by the human ear are called SONICS. The normal hearing range extends from about 20 to 20,000 hertz. However, to establish a standard sonic range, the Navy has set an arbitrary upper limit for sonics at 10,000 hertz and a lower limit at 15 hertz. Even though the average person can hear sounds above 10,000 hertz, it is standard practice to refer to sounds above that frequency as *ultrasonic*. Sounds between 15 hertz and 10,000 hertz are called *sonic*, while sounds below 15 hertz are known as *infrasonic* (formerly referred to as subsonic) sounds.

Q16. What term describes sounds capable of being heard by the human ear?

Q17. Are all sounds audible to the human ear? Why?

REQUIREMENTS FOR SOUND

Recall that sound waves are compression waves. The existence of compression waves depends on the transfer of energy. To produce vibrations that become sounds, a mechanical device (the source) must first receive an input of energy. Next, the device must be in contact with a medium that will receive the sound energy and carry it to a receiver. If the device is not in contact with a medium, the energy will not be transferred to a receiver, and there will be no sound.

Thus, three basic elements for transmission and reception of sound must be present before a sound can be produced. They are (1) the source (or transmitter), (2) a medium for carrying the sound (air, water, metal, etc.), and (3) the detector (or receiver).

A simple experiment provides convincing evidence that a medium must be present if sound is to be transferred. In figure 1-12, an electric bell is suspended by rubber bands in a bell jar from which the air can be removed. An external switch is connected from a battery to the bell so the bell may be rung intermittently. As the air is pumped out, the sound from the bell becomes weaker and weaker. If a perfect vacuum could be obtained, and if no sound were conducted out of the jar by the rubber bands, the sound from the bell would be completely inaudible. In other words, sound cannot be transmitted through a vacuum. When the air is admitted again, the sound is as loud as it was at the beginning. This experiment shows that when air is in contact with the vibrating bell, it carries energy to the walls of the jar, which in turn are set in vibration. Thus, the energy passes into the air outside of the jar and then on to the ear of the observer. This experiment illustrates that sound cannot exist in empty space (or a vacuum).

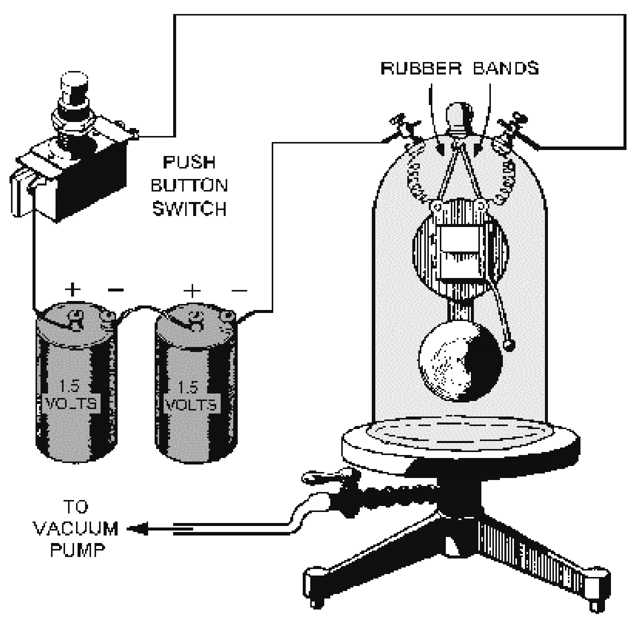


Figure 1-12.—No air, no sound.

Any object that moves rapidly back and forth, or vibrates, and thus disturbs the medium around it may be considered a source for sound. Bells, speakers, and stringed instruments are familiar sound sources.

The material through which sound waves travel is called the *medium*. The density of the medium determines the ease, distance, and speed of sound transmission. The higher the density of the medium, the slower sound travels through it.

The detector acts as the receiver of the sound wave. Because it does not surround the source of the sound wave, the detector absorbs only part of the energy from the wave and sometimes requires an amplifier to boost the weak signal.

As an illustration of what happens if one of these three elements is not present, let's refer to our experiment in which a bell was placed in a jar containing a vacuum. You could see the bell being struck, but you could hear no sound because there was no medium to transmit sound from the bell to you. Now let's look at another example in which the third element, the detector, is missing. You see a source (such as an explosion) apparently producing a sound, and you know the medium (air) is present, but you are too far away to hear the noise. Thus, as far as you are concerned, there is no detector and, therefore, no sound. We must assume, then, that sound can exist only when a source transmits sound through a medium, which passes it to a detector. Therefore, in the absence of any one of the basic elements (source, medium, detector) there can be NO sound.

Q18. Sound waves transmitted from a source are sometimes weak when they reach the detector. What instrument is needed to boost the weak signal?

TERMS USED IN SOUND WAVES

Sound waves vary in length according to their frequency. A sound having a long wavelength is heard at a <u>low pitch</u> (low frequency); one with a short wavelength is heard at a <u>high pitch</u> (high frequency). A complete wavelength is called a cycle. The distance from one point on a wave to the corresponding point on the next wave is a <u>wavelength</u>. The number of cycles per second (hertz) is the frequency of the sound. The <u>frequency</u> of a sound wave is also the number of <u>vibrations</u> per second produced by the sound source.

Q19. What are the three basic requirements for sound?

CHARACTERISTICS OF SOUND

Sound waves travel at great distances in a very short time, but as the distance increases the waves tend to spread out. As the sound waves spread out, their energy simultaneously spreads through an increasingly larger area. Thus, the wave energy becomes weaker as the distance from the source is increased.

Sounds may be broadly classified into two general groups. One group is NOISE, which includes sounds such as the pounding of a hammer or the slamming of a door. The other group is musical sounds, or TONES. The distinction between noise and tone is based on the regularity of the vibrations, the degree of damping, and the ability of the ear to recognize components having a musical sequence. You can best understand the physical difference between these kinds of sound by comparing the waveshape of a musical note, depicted in view A of figure 1-13, with the waveshape of noise, shown in view B. You can see by the comparison of the two waveshapes, that noise makes a very irregular and haphazard curve and a musical note makes a uniform and regular curve.

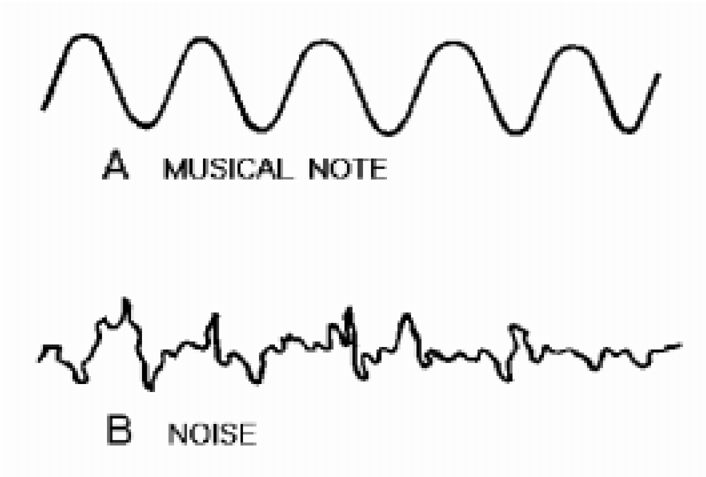


Figure 1-13.—Musical sound versus noise.

Sound has three basic characteristics: pitch, intensity, and quality. Each of these three characteristics is associated with one of the properties of the source or the type of waves which it produces. The pitch depends upon the frequency of the waves; the intensity depends upon the amplitude of the waves; and the quality depends upon the form of the waves. With the proper combination of these characteristics, the tone is pleasant to the ear. With the wrong combination, the sound quality turns into noise.

The Pitch of Sound

The term PITCH is used to describe the frequency of a sound. An object that vibrates many times per second produces a sound with a high pitch, as with a police whistle. The slow vibrations of the heavier strings of a violin cause a low-pitched sound. Thus, the frequency of the wave determines pitch. When the frequency is low, sound waves are long; when it is high, the waves are short. A sound can be so high in frequency that the waves reaching the ear cannot be heard. Likewise, some frequencies are so low that the eardrums do not convert them into sound. The range of sound that the human ear can detect varies with each individual.

The Intensity of Sound

The intensity of sound, at a given distance, depends upon the amplitude of the waves. Thus, a tuning fork gives out more energy in the form of sound when struck hard than when struck gently. You should remember that when a tuning fork is struck, the sound is omnidirectional (heard in all directions), because the sound waves spread out in all directions, as shown in figure 1-14. You can see from the figure that as the distance between the waves and the sound source increases, the energy in each wave spreads over a greater area; hence, the intensity of the sound decreases. The speaking tubes sometimes used aboard a ship prevent the sound waves from spreading in all directions by concentrating them in one desired direction (unidirectional), producing greater intensity. Therefore, the sound is heard almost at its original intensity at the opposite end of the speaking tube. The unidirectional megaphone and the directional loudspeaker also prevent sound waves from spreading in all directions.

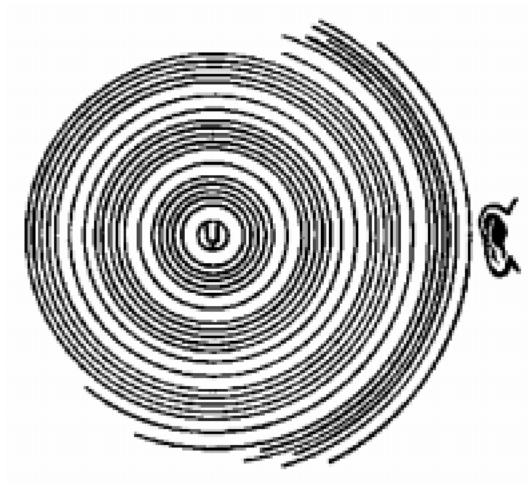


Figure 1-14.—Sound waves spread in all directions.

Sound intensity and loudness are often mistakenly interpreted as having the same meaning. Although they are related, they are not the same. Sound INTENSITY is a measure of the sound energy of a wave. LOUDNESS, on the other hand, is the sensation the intensity (and sometimes frequency) the sound wave produces on the ear. Increasing the intensity causes an increase in loudness but not in a direct proportion. For instance, doubling the loudness of a sound requires about a tenfold increase in the intensity of the sound.

Sound Quality

Most sounds, including musical notes, are not pure tones. They are a mixture of different frequencies (tones). A tuning fork, when struck, produces a pure tone of a specific frequency. This pure tone is produced by regular vibrations of the source (tines of the tuning fork). On the other hand, scraping your fingernails across a blackboard only creates noise, because the vibrations are irregular. Each individual pipe of a pipe organ is similar to a tuning fork, and each pipe produces a tone of a specific frequency. But sounding two or more pipes at the same time produces a complex waveform. A tone that closely imitates any of the vowel sounds can be produced by selecting the proper pipes and sounding them at the same time. Figure 1-15 illustrates the combining of two pure tones to make a COMPLEX WAVE.

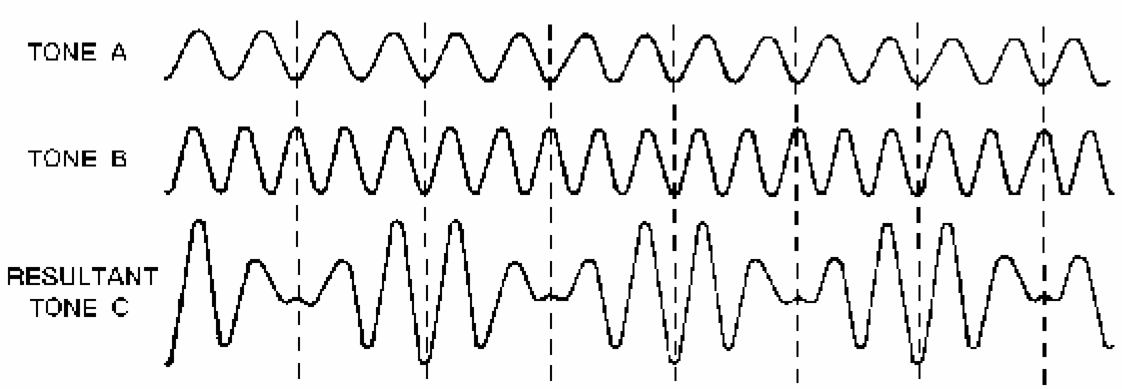


Figure 1-15.—Combination of tones.

The QUALITY of a sound depends on the complexity of its sound waves, such as the waves shown in tone C of figure 1-15. Almost all sounds (musical and vocal included) have complicated (complex) waveforms. Tone A is a simple wave of a specific frequency that can be produced by a tuning fork, piano, organ, or other musical instrument. Tone B is also a simple wave but at a different frequency. When the two tones are sounded together, the complex waveform in tone C is produced. Note that tone C has the same frequency as tone A with an increase in amplitude. The human ear could easily distinguish between tone A and tone C because of the quality. Therefore, we can say that quality distinguishes tones of like pitch and loudness when sounded on different types of musical instruments. It also distinguishes the voices of different persons.

Q20. What are the two general groups of sound?

Q21. What are the three basic characteristics of sound?

Q22. What is the normal audible range of the human ear?

Q23. What is intensity as it pertains to sound?

Q24. What characteristic of sound enables a person to distinguish one musical instrument from another, if they are all playing the same note?

ELASTICITY AND DENSITY AND VELOCITY OF TRANSMISSION

Sound waves travel through any medium to a velocity that is controlled by the medium. Varying the frequency and intensity of the sound waves will not affect the speed of propagation. The ELASTICITY and DENSITY of a medium are the two basic physical properties that govern the velocity of sound through the medium.

Elasticity is the ability of a strained body to recover its shape after deformation, as from a vibration or compression. The measure of elasticity of a body is the force it exerts to return to its original shape.

The *density* of a medium or substance is the mass per unit volume of the medium or substance. Raising the temperature of the medium (which decreases its density) has the effect of increasing the velocity of sound through the medium.

The velocity of sound in an elastic medium is expressed by the formula:

$$v = \sqrt{\frac{E}{d}}$$

Even though solids such as steel and glass are far more dense than air, their elasticity’s are so much greater that the velocities of sound in them are 15 times greater than the velocity of sound in air. Using elasticity as a rough indication of the speed of sound in a given medium, we can state as a general rule that *sound travels faster in harder materials* (such as steel), *slower in liquids, and slowest in gases*. Density has the opposite effect on the velocity of sound, that is, with other factors constant, a denser material (such as lead) passes sound slower.

At a given temperature and atmospheric pressure, all sound waves travel in air at the same speed. Thus the velocity that sound will travel through air at 32º F (0º C) is 1,087 feet per second. But for practical purposes, the speed of sound in air may be considered as 1,100 feet per second. Table 1-1 gives a comparison of the velocity of sound in various mediums.

Table 1-1.—Comparison of Velocity of Sound in Various Mediums

MEDIUM	TEMPERATURE ºF	ºC	VELOCITY (FT/SEC)
AIR	32	0	1,087
AIR	68	20	1,127
ALUMINUM	68	20	16,700
CARBON DIOXIDE	32	0	856
FRESH WATER	32	0	4,629
FRESH WATER	68	20	4,805
HYDROGEN	32	0	4,219
LEAD	32	20	4,030
SALT WATER	32	0	4,800
SALT WATER	68	20	4,953
STEEL	32	0	16,410
STEEL	68	20	16,850

Q25. How does density and temperature affect the velocity of sound?

ACOUSTICS

The science of sound is called ACOUSTICS. This subject could fill volumes of technical books, but we will only scratch the surface in this chapter. We will present important points that you will need for a better understanding of sound waves.

Acoustics, like sound, relates to the sense of hearing. It also deals with the production, control, transmission, reception, and the effects of sound. For the present, we are concerned only with the last relationship—the effects of sound. These same effects will be used throughout your study of wave propagation.

Echo

An ECHO is the reflection of the original sound wave as it bounces off a distant surface. Just as a rubber ball bounces back when it is thrown against a hard surface, sound waves also bounce off most surfaces. As you have learned from the study of the law of conservation of energy, a rubber ball never bounces back with as much energy as the initial bounce. Similarly, a reflected sound wave is not as loud as the original sound wave. In both cases, some of the energy is absorbed by the reflecting surface. Only a portion of the original sound is reflected, and only a portion of the reflected sound returns to the listener. For this reason, an echo is never as loud as the original sound.

Sound reflections (echoes) have many applications in the Navy. The most important of these applications can be found in the use of depth finding equipment (the fathometer) and sonar. The fathometer sends sound-wave pulses from the bottom of the ship and receives echoes from the ocean floor to indicate the depth of the ocean beneath the ship. The sonar transmits a pulse of sound energy and receives the echo to indicate range and bearing of objects or targets in the ocean depths.

Refraction

When sound waves traveling at different velocities pass obliquely (at an angle) from one medium into another, the waves are refracted; that is, their line of travel is bent. Refraction occurs gradually when one part of a sound wave is traveling faster than the other parts. For example, the wind a few feet above the surface of the earth has a greater velocity than that near the surface because friction retards the lower layers (see figure 1-16). The velocity of the wind is added to the velocity of the sound through the air. The result is that the upper portion of the sound wave moves faster than the lower portion and causes a gradual change in the direction of travel of the wave. Refraction causes sound to travel farther with the wind than against it.

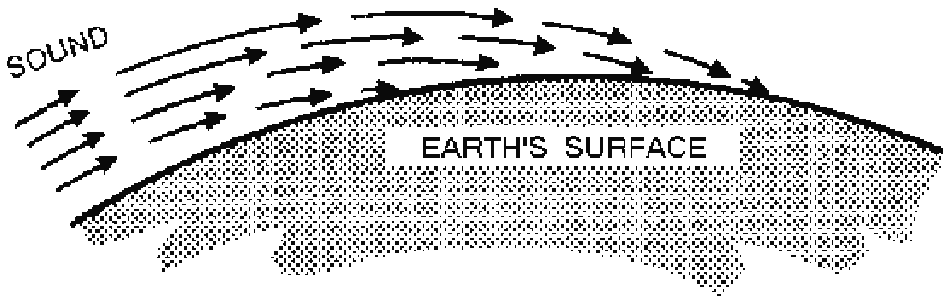


Figure 1-16.—Refraction of sound.

Reverberation

In empty rooms or other confined spaces, sound may be reflected several times to cause what is known as reverberation. REVERBERATION is the multiple reflections of sound waves. Reverberations seem to prolong the time during which a sound is heard. Examples of this often occur in nature. For instance, the discharge of lightning causes a sharp, quick sound. By the time this sound has reached the ears of a distant observer, it is usually drawn out into a prolonged roar by reverberations that we call thunder. A similar case often arises with underwater sound equipment. Reverberations from nearby points may continue for such a long time that they interfere with the returning echoes from targets.

Interference

Any disturbance, man-made or natural, that causes an undesirable response or the degradation of a wave is referred to as INTERFERENCE.

Two sound waves moving simultaneously through the same medium will advance independently, each producing a disturbance as if the other were not present. If the two waves have the same frequency—in phase with each other—and are moving in the same direction, they are additive and are said to interfere constructively. If the two waves have the same frequency and are moving in the same direction, but out of phase with each other, they are subtractive and are said to interfere destructively. If these two subtractive waves have equal amplitudes, the waves cancel each other. This addition or subtraction of waves is often called interference.

Resonance

At some time during your life you probably observed someone putting his or her head into an empty barrel or other cavity and making noises varying in pitch. When that person's voice reached a certain pitch, the tone produced seemed much louder than the others. The reason for this phenomenon is that at that a certain pitch the frequency of vibrations of the voice matched the resonant (or natural) frequency of the cavity. The resonant frequency of a cavity is the frequency at which the cavity body will begin to vibrate and create sound waves. When the resonant frequency of the cavity was reached, the sound of the voice was reinforced by the sound waves created by the cavity, resulting in a louder tone.

This phenomenon occurs whenever the frequency of vibrations is the same as the natural frequency of a cavity, and is called RESONANCE.

Noise

The most complex sound wave that can be produced is noise. Noise has no tonal quality. It distracts and distorts the sound quality that was intended to be heard. NOISE is generally an unwanted disturbance caused by spurious waves originating from man-made or natural sources, such as a jet breaking the sound barrier, or thunder.

Q26. What term is used in describing the science of sound?

Q27. A sound wave that is reflected back toward the source is known as what type of sound?

Q28. What is the term for multiple reflections of sound waves?

Q29. A cavity that vibrates at its natural frequency produces a louder sound than at other frequencies. What term is used to describe this phenomenon?

Q30. What do we call a disturbance that distracts or distorts the quality of sound?

LIGHT WAVES

Technicians maintain equipment that use frequencies from one end of the electromagnetic spectrum to the other—from low-frequency radio waves to high-frequency X-rays and cosmic rays. Visible light is a small but very important part of this electromagnetic spectrum.

Most of the important terms that pertain to the behavior of waves, such as reflection, refraction, diffraction, etc., were discussed earlier in this chapter. We will now discuss how these terms are used in understanding light and light waves. The relationship between light and light waves (rays) is the same as sound and sound waves.

Light is a form of energy. It can be produced by various means (mechanical, electrical, chemical, etc.). We can see objects because the light rays they give off or reflect reach our eyes. If the object is the source of light energy, it is called luminous. If the object is not the source of light but reflects light, it is called an illuminated body.

PROPAGATION OF LIGHT

The exact nature of light is not fully understood, although scientists have been studying the subject for many centuries. Some experiments seem to show that light is composed of tiny particles, and some suggest that it is made up of waves.

One theory after another attracted the approval and acceptance of physicists. Today, some scientific phenomena can be explained only by the wave theory and others only by the particle theory. Physicists, constantly searching for some new discovery that would bring these two theories into agreement, gradually have come to accept a theory that combines the principles of the two theories.

According to the view now generally accepted, light is a form of electromagnetic radiation; that is, light and similar forms of radiation are made up of moving electric and magnetic fields. These two fields will be explained thoroughly later in this chapter.

ELECTROMAGNETIC THEORY OF LIGHT

James Clark Maxwell, a brilliant Scottish scientist Of the middle l9th century, showed, by constructing an oscillating electrical circuit, that electromagnetic waves could move through empty space. Light eventually was proved to be electromagnetic.

Current light theory says that light is made up of very small packets of electromagnetic energy called PHOTONS (the smallest unit of radiant energy). These photons move at a constant speed in the medium through which they travel. Photons move at a faster speed through a vacuum than they do in the atmosphere, and at a slower speed through water than air.

The electromagnetic energy of light is a form of electromagnetic radiation. Light and similar forms of radiation are made up of moving electric and magnetic forces and move as waves. Electromagnetic waves move in a manner similar to the waves produced by the pebble dropped in the pool of water discussed earlier in this chapter. The transverse waves of light from a light source spread out in expanding circles much like the waves in the pool. However, the waves in the pool are very slow and clumsy in comparison with light, which travels approximately 186,000 miles per second.

Light radiates from its source in all directions until absorbed or diverted by some substance (fig. 1-17). The lines drawn from the light source (a light bulb in this instance) to any point on one of these waves indicate the direction in which the waves are moving. These lines, called radii of the spheres, are formed by the waves and are called LIGHT RAYS.

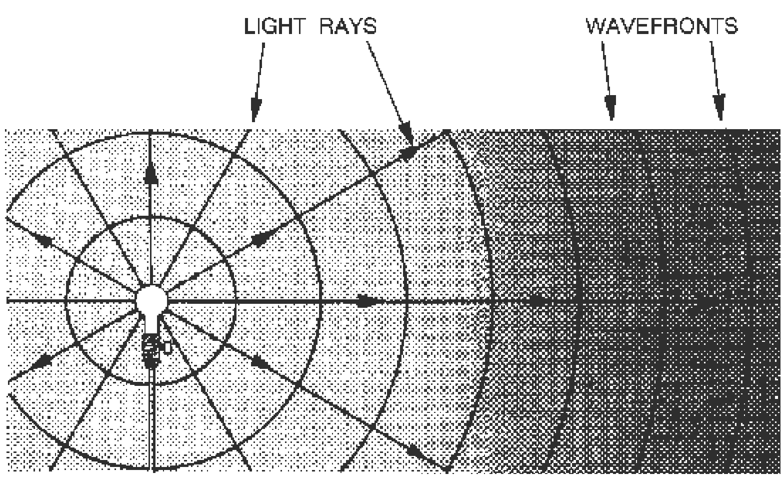


Figure 1-17.—Waves and radii from a nearby light source.

Although single rays of light do not exist, light "rays" as used in illustrations are a convenient method used to show the direction in which light is traveling at any point.

A large volume of light is called a beam; a narrow beam is called a pencil; and the smallest portion of a pencil is called a light ray. A ray of light, can be illustrated as a straight line. This straight line drawn from a light source will represent an infinite number of rays radiating in all directions from the source.

Q31. What are three means of producing light?

Q32. What is the smallest unit of radiant energy?

FREQUENCIES AND WAVELENGTHS

Compared to sound waves, the frequency of light waves is very high and the wavelength is very short. To measure these wavelengths more conveniently, a special unit of measure called an ANGSTROM UNIT, or more usually, an ANGSTROM (Å) was devised. Another common unit used to measure these waves is the millimicron (mµ), which is one millionth of a millimeter. One mF equals ten angstroms. One angstrom equals 1055^{-10}m.

Q33. What unit is used to measure the different wavelengths of light?

FREQUENCIES AND COLOR

For our discussion of light wave waves, we will use the millimicron measurement. The wavelength of a light determines the color of the light. Figure 1-18 indicates that light with a wavelength of 700 millimicrons is red, and that light with a wavelength of 500 millimicrons is blue-green. This illustration shows approximate wavelengths of the different colors in the visible spectrum. In actual fact, the color of light depends on its frequency, not its wavelength. However, light is measured in wavelengths.

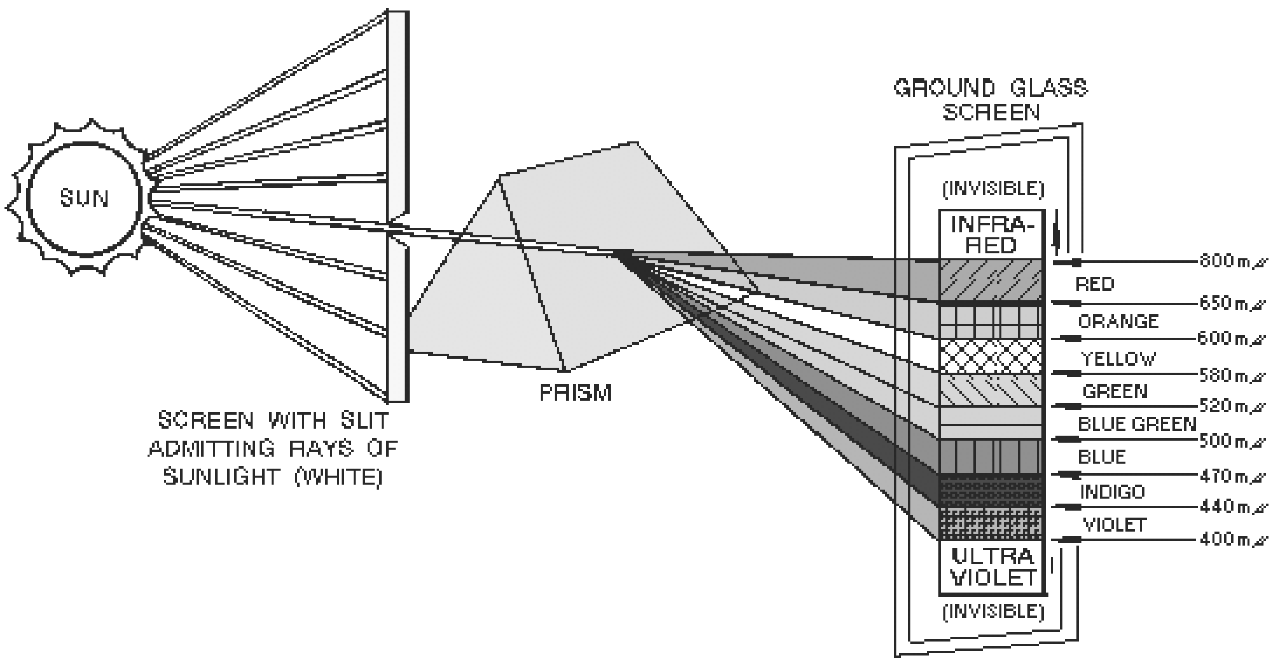


Figure 1-18.—Use of a prism to split white light into different colors.

When the wavelength of 700 millimicrons is measured in a medium such as air, it produces the color red, but the same wave measured in a different medium will have a different wavelength. When red light which has been traveling in air enters glass, it loses speed. Its wavelength becomes shorter or compressed, but it continues to be red. This illustrates that the color of light depends on frequency and not on wavelength. The color scale in figure 1-18 is based on the wavelengths in air.

When a beam of white light (sunlight) is passed through a PRISM, as shown in figure 1-18, it is refracted and dispersed (the phenomenon is known as DISPERSION) into its component wavelengths. Each of these wavelengths causes a different reaction of the eye, which sees the various colors that compose the visible spectrum. The visible spectrum is recorded as a mixture of red, orange, yellow, green, blue, indigo, and violet. White light results when the PRIMARIES (red, green, and blue) are mixed together in overlapping beams of light. (NOTE: These are not the primaries used in mixing pigments, such as in paint.) Furthermore, the COMPLEMENTARY or SECONDARY colors (magenta, yellow, and cyan) may be shown with equal ease by mixing any two of the primary colors in overlapping beams of light. Thus, red and green light mixed in equal intensities will make yellow light; green and blue will produce cyan (blue-green light); and blue and red correctly mixed will produce magenta (a purplish red light).

LIGHT AND COLOR

All objects absorb some of the light that falls on them. An object appears to be a certain color because it absorbs all of the light waves except those whose frequency corresponds to that particular color. Those waves are reflected from the surface, strike your eye, and cause you to see the particular color. The color of an object therefore depends on the frequency of the electromagnetic wave reflected.

LUMINOUS BODIES

Certain bodies, such as the sun, a gas flame, and an electric light filament, are visible because they are light sources. They are called SELF-LUMINOUS bodies. Objects other than self-luminous bodies become visible only when they are in the presence of light from luminous bodies.

Most NONLUMINOUS bodies are visible because they diffuse or reflect the light that falls on them. A good example of a nonluminous diffusing body is the moon, which shines only because the sunlight falling onto its surface is diffused.

Black objects do not diffuse or reflect light. They are visible only when outlined against a background of light from a luminous or diffusing body.

PROPERTIES OF LIGHT

When light waves, which travel in straight lines, encounter any substance, they are either transmitted, refracted, reflected, or absorbed. This is illustrated in figure 1-19. When light strikes a substance, some absorption and some reflection always take place. No substance completely transmits, reflects, or absorbs all of the light rays that reach its surface. Substances that transmit almost all the light waves that fall upon them are said to be TRANSPARENT. A transparent substance is one through which you can see clearly. Clear glass is transparent because it transmits light rays without diffusing them (view A of figure 1-20). There is no known perfectly transparent substance, but many substances are nearly so. Substances through which some light rays can pass but through which objects cannot be seen clearly because the rays are diffused are called TRANSLUCENT (view B of figure 1-20). The frosted glass of a light bulb and a piece of oiled paper are examples of translucent materials. Substances that do not transmit any light rays are called OPAQUE (view C of figure 1-20). Opaque substances can either reflect or absorb all of the light rays that fall upon them.

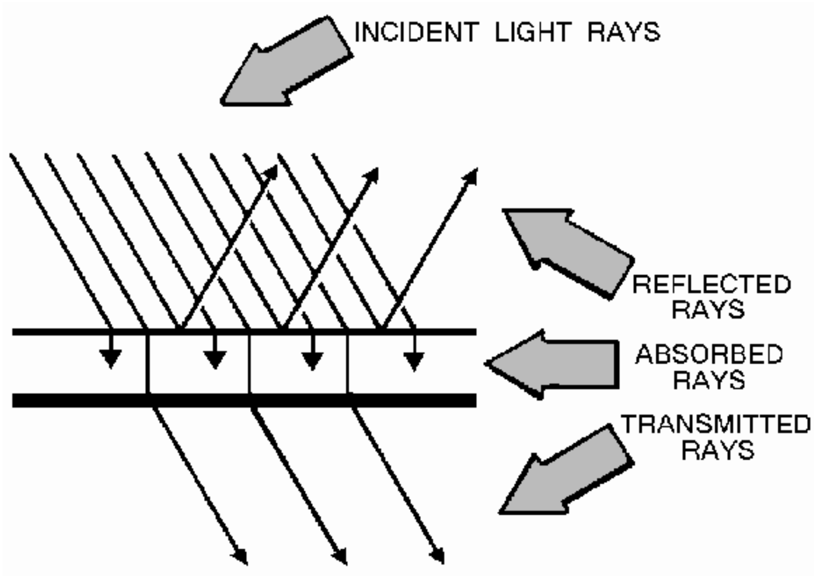


Figure 1-19.—Light waves reflected, absorbed, and transmitted.

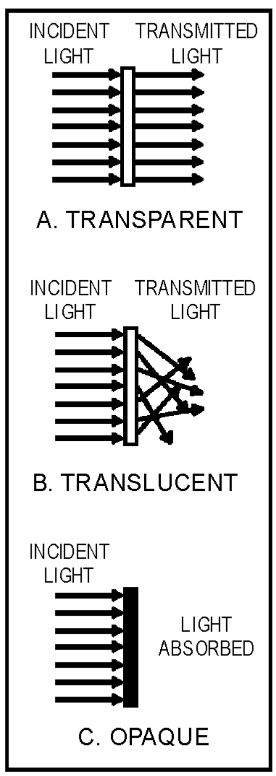


Figure 1-20.—Transparent, translucent, and opaque substances.

Q34. What are the three primary colors of light?

Q35. What are the three secondary colors of light?

Q36. White light falls upon a dull, rough, dark-brown object. Will the light primarily be reflected, diffused, or absorbed by the object?

Q37. What color will be emitted by a dull, rough, black object when white light falls upon it?

Q38. A substance that transmits light but through which an object cannot be seen clearly is known as what kind of substance?

Speed of Light

You probably have heard people say, "quick as lightning" or "fast as light" to describe rapid motion; nevertheless, it is difficult to realize how fast light actually travels. Not until recent years have scientists been able to measure accurately the speed of light.

Prior to the middle 17th century, scientists thought that light required no time at all to pass from the source to the observer. Then in 1675, Ole Roemer, a Danish astronomer, discovered that light travels approximately 186,000 miles per second in space. At this velocity, a light beam can circle the earth 7 1/2 times in one second.

The speed of light depends on the medium through which the light travels. In empty space, the speed is 186,000 (1.86×10^5) miles per second. It is almost the same in air. In water, it slows down to approximately 140,000 (1.4×10^5) miles per second. In glass, the speed of light is 124,000 (1.24×10^5) miles per second. In other words, the speed of light decreases as the density of the substance through which the light passes increases.

The velocity of light, which is the same as the velocity of other electromagnetic waves, is considered to be constant, at 186,000 miles per second. If expressed in meters, it is 300,000,000 meters per second.

Reflection of Light

Light waves obey the law of reflection in the same manner as other types of waves. Consider the straight path of a light ray admitted through a narrow slit into a darkened room. The straight path of the beam is made visible by illuminated dust particles suspended in the air. If the light beam is made to fall onto the surface of a mirror or other reflecting surface, however, the direction of the beam changes sharply. The light can be reflected in almost any direction depending on the angle at which the mirror is held.

As shown earlier in figure 1-9, if a light beam strikes a mirror, the angle at which the beam is reflected depends on the angle at which it strikes the mirror. The beam approaching the mirror is the INCIDENT or striking beam, and the beam leaving the mirror is the REFLECTED beam.

The term "reflected light" simply refers to light waves that are neither transmitted nor absorbed, but are thrown back from the surface of the medium they encounter.

You will see this application used in our discussion of radio waves (chapter 2) and antennas (chapter 4).

Q39. At what speed does light travel?

Refraction of Light

The change of direction that occurs when a ray of light passes from one transparent substance into another of different density is called refraction. Refraction is due to the fact that light travels at various speeds in different transparent substances. For example, water never appears as deep as it really is, and objects under water appear to be closer to the surface than they really are. A bending of the light rays causes these impressions.

Another example of refraction is the apparent bending of a spoon when it is immersed in a cup of water. The bending seems to take place at the surface of the water, or exactly at the point where there is a change of density. Obviously, the spoon does not bend from the pressure of the water. The light forming the image of the spoon is bent as it passes from the water (a medium of high density) to the air (a medium of comparatively low density).

Without refraction, light waves would pass in straight lines through transparent substances without any change of direction. Refer back to figure 1-10, which shows refraction of a wave. As you can see, all rays striking the glass at any angle other than perpendicular are refracted. However, the perpendicular ray, which enters the glass normal to the surface, continues through the glass and into the air in a straight line no refraction takes place.

Diffusion of Light

When light is reflected from a mirror, the angle of reflection of each ray equals the angle of incidence. When light is reflected from a piece of plain white paper, however, the reflected beam is scattered, or DIFFUSED, as shown in figure 1-21. Because the surface of the paper is not smooth, the reflected light is broken up into many light beams that are reflected in all directions.

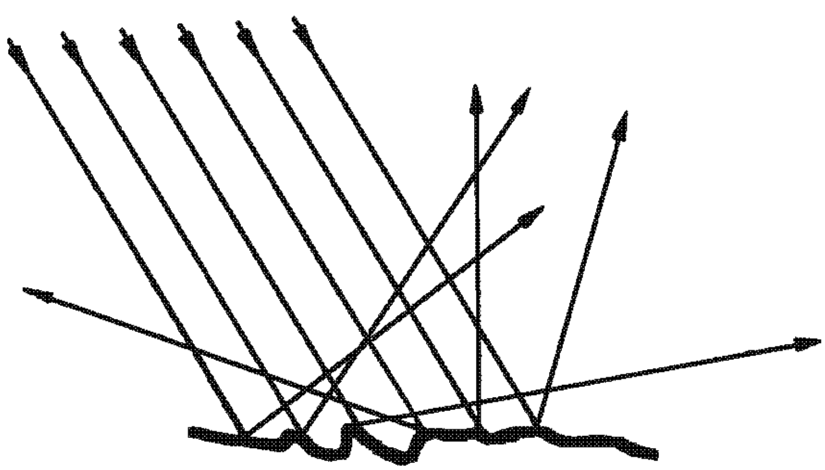


Figure 1-21.—Diffusion of light.

Absorption of Light

You have just seen that a light beam is reflected and diffused when it falls onto a piece of white paper. If a light beam falls onto a piece of black paper, the black paper absorbs most of the light rays and very little light is reflected from the paper. If the surface on which the light beam falls is perfectly black, there is no reflection; that is, the light is totally absorbed. No matter what kind of surface light falls on, however, some of the light is absorbed.

Q40. A light wave enters a sheet of glass at a perfect right angle to the surface. Is the majority of the wave reflected, refracted, transmitted, or absorbed?

Q41. When light strikes a piece of white paper, the light is reflected in all directions. What do we call this scattering of light?

COMPARISON OF LIGHT WAVES WITH SOUND WAVES

There are two main differences between sound waves and light waves. The first difference is in velocity. Sound waves travel through air at the speed of approximately 1,100 feet per second; light waves travel through air and empty space at a speed of approximately 186,000 miles per second. The second difference is that sound is composed of longitudinal waves (alternate compressions and expansions of matter) and light is composed of transverse waves in an electromagnetic field.

Although both are forms of wave motion, sound requires a solid, liquid, or gaseous medium; whereas light travels through empty space. The denser the medium, the greater the speed of sound. The opposite is true of light. Light travels approximately one-third slower in water than in air. Sound travels through all substances, but light cannot pass through opaque materials.

Frequency affects both sound and light. A certain range of sound frequencies produces sensations that you can hear. A slow vibration (low frequency) in sound gives the sensation of a low note. A more rapid sound vibration (higher frequency) produces a higher note. Likewise, a certain range of light frequencies produces sensations that you can see. Violet light is produced at the high-frequency end of the light spectrum, while red light is produced at the low-frequency end of the light spectrum. A change in frequency of sound waves causes an audible sensation—a difference in pitch. A change in the frequency of a light wave causes a visual sensation—a difference in color.

For a comparison of light waves with sound waves, see table 1-2.

Table 1-2.—Comparison of Light Waves and Sound Waves

	SOUND WAVES	LIGHT WAVES
VELOCITY IN AIR	APPROXIMATELY 1,100 FEET PER SECOND	APPROXIMATELY 186,000 MILES PER SECOND
FORM	A FORM OF WAVE MOTION	A FORM OF WAVE MOTION
WAVE COMPOSITION	LONGITUDINAL	TRANSVERSE
TRANSMITTING MEDIUM	ALL SUBSTANCES	EMPTY SPACE AND ALL SUBSTANCES EXCEPT OPAQUE MATERIALS
RELATION OF TRANSMITTING MEDIUM VELOCITY TO VELOCITY	THE DENSER THE MEDIUM, THE GREATER THE SPEED	THE DENSER THE MEDIUM, THE SLOWER THE SPEED
SENSATIONS PRODUCED	HEARING	SEEING
VARIATIONS IN SENSATIONS PRODUCED	A LOW FREQUENCY CAUSES A LOW NOTE; A HIGH FREQUENCY, A HIGH NOTE	A LOW FREQUENCY CAUSES RED LIGHT; A HIGH FREQUENCY, VIOLET LIGHT

Q42. What three examples of electromagnetic energy are mentioned in the text?

Q43. What is the main difference between the bulk of the electromagnetic spectrum and the visual spectrum?

ELECTROMAGNETIC SPECTRUM

Light is one kind of electromagnetic energy. There are many other types, including heat energy and radio energy. The only difference between the various types of electromagnetic energy is the frequency of their waves (rate of vibration). The term SPECTRUM is used to designate the entire range of electromagnetic waves arranged in order of their frequencies. The VISIBLE SPECTRUM contains only those waves which stimulate the sense of sight. You, as a technician, might be expected to maintain equipment that uses electromagnetic waves within, above, and below the visible spectrum.

There are neither sharp dividing lines nor gaps in the ELECTROMAGNETIC SPECTRUM. Figure 1-22 illustrates how portions of the electromagnetic spectrum overlap. Notice that only a small portion of the electromagnetic spectrum contains visible waves, or light, which can be seen by the human eye.

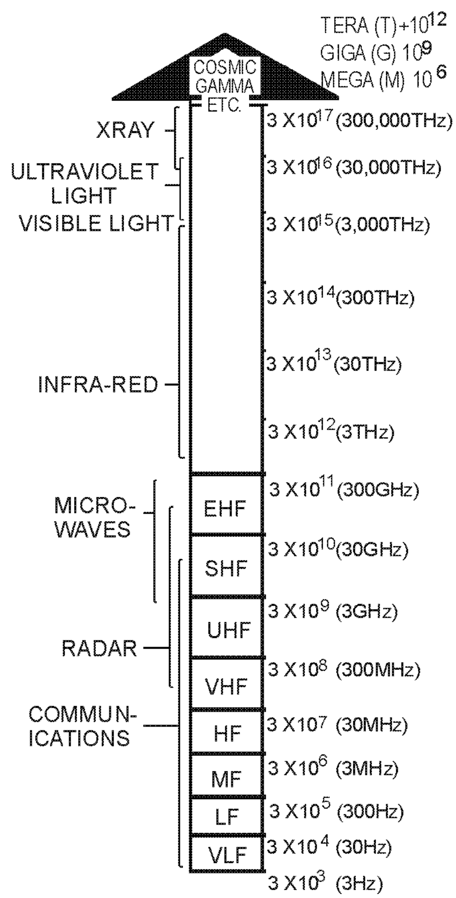


Figure 1-22.—Electromagnetic spectrum.

ELECTROMAGNETIC WAVES

In general, the same principles and properties of light waves apply to the communications electromagnetic waves you are about to study. The electromagnetic field is used to transfer energy (as communications) from point to point. We will introduce the basic ANTENNA as the propagation source of these electromagnetic waves.

THE BASIC ANTENNA

The study of antennas and electromagnetic wave propagation is essential to a complete understanding of radio communication, radar, loran, and other electronic systems. Figure 1-23 shows a simple radio communication system. In the illustration, the transmitter is an electronic device that generates radio-frequency energy. The energy travels through a transmission line (we will discuss this in chapter 3) to an antenna. The antenna converts the energy into radio waves that radiate into space from the antenna at the speed of light. The radio waves travel through the atmosphere or space until they are either reflected by an object or absorbed. If another antenna is placed in the path of the radio waves, it absorbs part of the waves and converts them to energy. This energy travels through another transmission line and is fed to a receiver. From this example, you can see that the requirements for a simple communications system are (1) transmitting equipment, (2) transmission line, (3) transmitting antenna, (4) medium, (5) receiving antenna, and (6) receiving equipment.

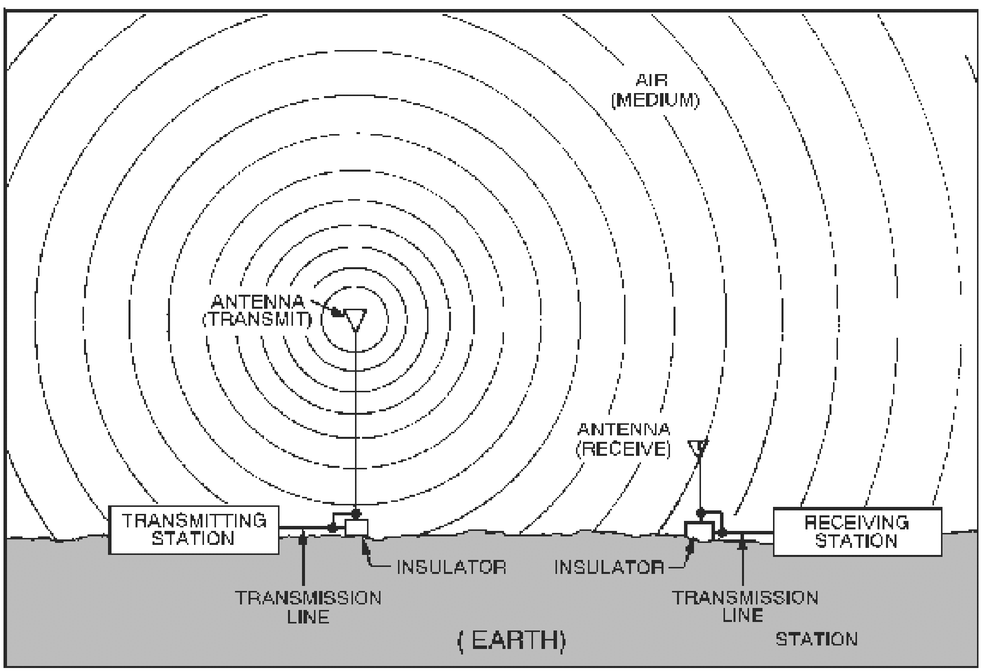


Figure 1-23.—Simple radio communication system.

An antenna is a conductor or a set of conductors used either to radiate electromagnetic energy into space or to collect this energy from space. Figure 1-24 shows an antenna. View A is a drawing of an actual antenna; view B is a cut-away view of the antenna; and view C is a simplified diagram of the antenna.

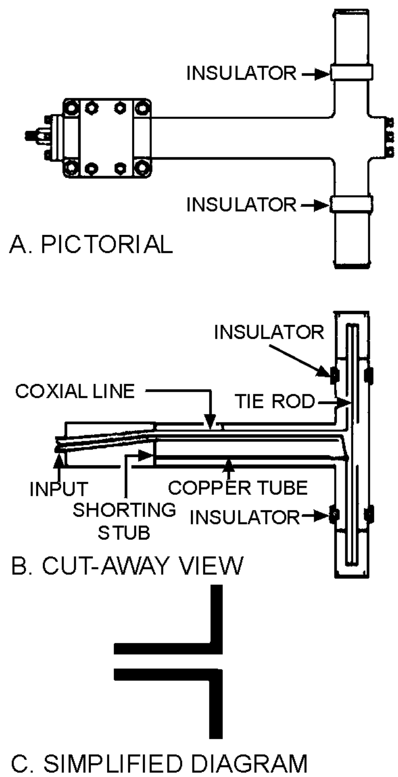


Figure 1-24.—Antenna.

COMPONENTS OF THE ELECTROMAGNETIC WAVE

An electromagnetic wave consists of two primary components—an ELECTRIC FIELD and a MAGNETIC FIELD. The electric field results from the force of voltage, and the magnetic field results from the flow of current.

Although electromagnetic fields that are radiated are commonly considered to be waves, under certain circumstances their behavior makes them appear to have some of the properties of particles. In general, however, it is easier to picture electromagnetic radiation in space as horizontal and vertical lines of force oriented at right angles to each other. These lines of force are made up of an electric field (E) and a magnetic field (H), which together make up the electromagnetic field in space.

The electric and magnetic fields radiated from an antenna form the electromagnetic field. This field is responsible for the transmission and reception of electromagnetic energy through free space. An antenna, however, is also part of the electrical circuit of a transmitter or a receiver and is equivalent to a circuit containing inductance, capacitance, and resistance. Therefore, the antenna can be expected to display definite voltage and current relationships with respect to a given input. A current through the antenna produces a magnetic field, and a charge on the antenna produces an electric field. These two fields combine to form the INDUCTION field. To help you gain a better understanding of antenna theory, we must review some basic electrical concepts. We will review voltage and its electric field, current and its magnetic field, and their relationship to propagation of electrical energy.

Q44. What are the two components (fields) that make up the electromagnetic wave?

Q45. What do we call a conductor (or set of conductors) that radiates electromagnetic energy into space?

Electric Field

Around every electrically charged object is a force field that can be detected and measured. This force field can cause electric charges to move in the field. When an object is charged electrically, there is either a greater or a smaller concentration of electrons than normal. Thus, a difference of potential exists between a charged object and an uncharged object. An electric field is, therefore, associated with a difference of potential, or a voltage.

This invisible field of force is commonly represented by lines that are drawn to show the paths along which the force acts. The lines representing the electric field are drawn in the direction that a single positive charge would normally move under the influence of that field. A large electric force is shown by a large concentration of lines; a weak force is indicated by a few lines.

When a capacitor is connected across a source of voltage, such as a battery, it is charged by a particular amount, depending on the voltage and the value of capacitance. (See figure 1-25.) Because of the emf (electromotive force) of the battery, negative charges flow to the lower plate, leaving the upper plate positively charged. Along with the growth of charge, the electric field is also building up. The flux lines are directed from the positive to the negative charges and at right angles to the plates. When the capacitor is fully charged, the voltage of the capacitor is equal to the voltage of the source and opposite in polarity. The charged capacitor stores the energy in the form of an electric field. It can be said, therefore, that an electric field indicates voltage.

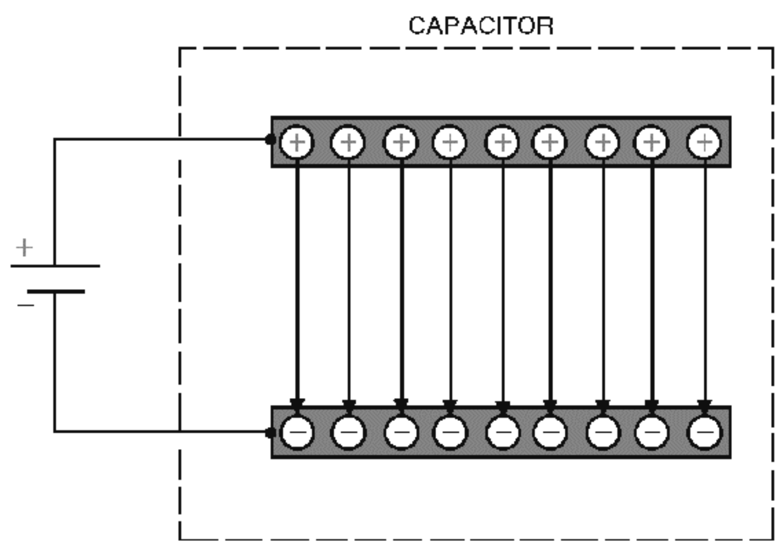


Figure 1-25.—Electric fields between plates.

If the two plates of the capacitor are spread farther apart, the electric field must curve to meet the plates at right angles (fig. 1-26). The straight lines in view A of figure 1-26 become arcs in view B, and approximately semicircles in view C, where the plates are in a straight line. Instead of flat metal plates, as in the capacitor, the two elements can take the form of metal rods or wires and form the basic antenna.

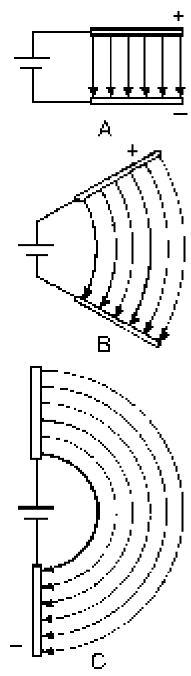


Figure 1-26.—Electric fields between plates at different angles.

In figure 1-27, two rods replace the plates of the capacitor, and the battery is replaced by an ac source generating a 60-hertz signal. On the positive alternation of the 60-hertz generator, the electric field extends from the positively charged rod to the negatively charged rod, as shown. On the negative alternation, the charge is reversed. The previous explanation of electrons moving from one plate to the other of the capacitor in figure 1-25 can also be applied to the rods in figure 1-27.

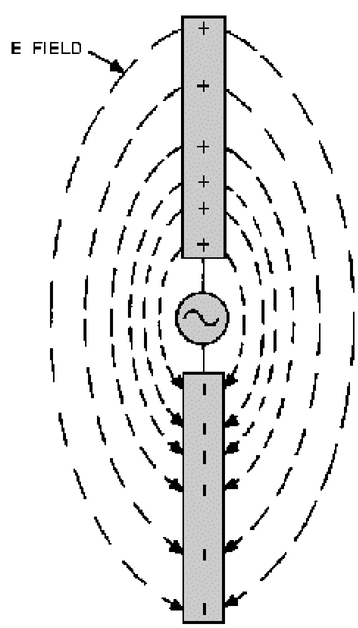


Figure 1-27.—Electric fields between elements.

The polarity of charges and the direction of the electric fields will reverse polarity and direction periodically at the frequency of the voltage source. The electric field will build up from zero to maximum in one direction and then collapse back to zero. Next, the field will build up to maximum in the opposite direction and then collapse back to zero. This complete reversal occurs during a single cycle of the source voltage. The HALF-WAVE DIPOLE ANTENNA (two separate rods in line as illustrated in figure 1-27) is the fundamental element normally used as a starting point of reference in any discussion concerning the radiation of electromagnetic energy into space. If rf energy from the ac generator (or transmitter) is supplied to the element of an antenna, the voltage across the antenna lags the current by 90 degrees. The antenna acts as if it were a capacitor.

Magnetic Field

When current flows through a conductor, a magnetic field is set up in the area surrounding the conductor. In fact, any moving electrical charge will create a magnetic field. The magnetic field is a region in space where a magnetic force can be detected and measured. There are two other fields involved—an INDUCTION FIELD, which exists close to the conductor carrying the current, and the RADIATION FIELD, which becomes detached from the current-carrying rod and travels through space.

To represent the magnetic field, lines of force are again used to illustrate the energy. Magnetic lines are not drawn between the rods, nor between high- and low-potential points, as the E lines that were discussed earlier. Magnetic lines are created by the flow of current rather than the force of voltage. The magnetic lines of force, therefore, are drawn at right angles to the direction of current flow.

The magnetic fields that are set up around two parallel rods, as shown in figure 1-28 view A, are in maximum opposition. Rod 1 contains a current flowing from the generator, while rod 2 contains a current flowing toward the generator. As a result, the direction of the magnetic field surrounding rod 1 is opposite the direction of the magnetic field surrounding rod 2. This will cause cancellation of part or all of both magnetic fields with a resultant decrease in radiation of the electromagnetic energy. View B illustrates the fact that if the far ends of rods 1 and 2 are separated from each other while the rods are still connected to the generator at the near ends, more space, and consequently less opposition, will occur between the magnetic fields of the two rods. View C illustrates the fact that placing the rods in line makes the currents through both rods flow in the same direction. Therefore, the two magnetic fields are in the same direction; thus, maximum electromagnetic radiation into space can be obtained.

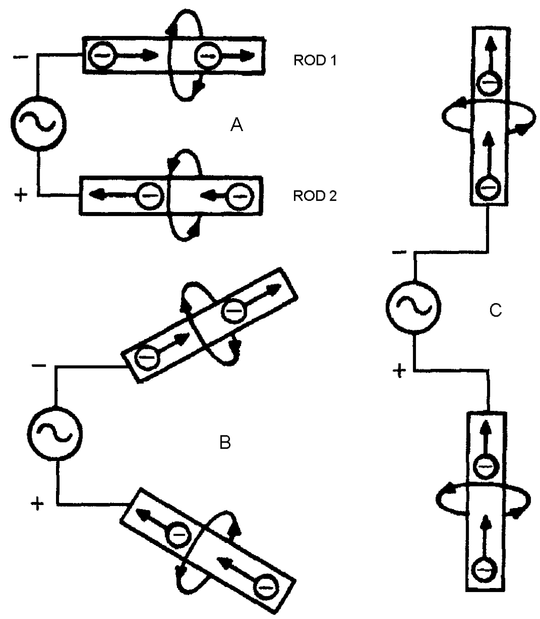


Figure 1-28.—Magnetic fields around elements.

Magnetic lines of force are indicated by the letter H and are called H lines. The direction of the magnetic lines may be determined by use of the left-hand rule for a conductor: If you grasp the conductor in your left hand with the thumb extended in the direction of the current flow, your fingers will point in the direction of the magnetic lines of force. In view C of figure 1-28, the direction of current flow is upward along both halves of the elements (conductors). The lines of magnetic force (flux) form concentric loops that are perpendicular to the direction of current flow. The arrowheads on the loops indicate the direction of the field. The left-hand rule is used to determine the direction of the magnetic field and is illustrated in figure 1-29. If the thumb of the left hand is extended in the direction of current flow and the fingers clenched, then the rough circles formed by the fingers indicate the direction of the magnetic field.

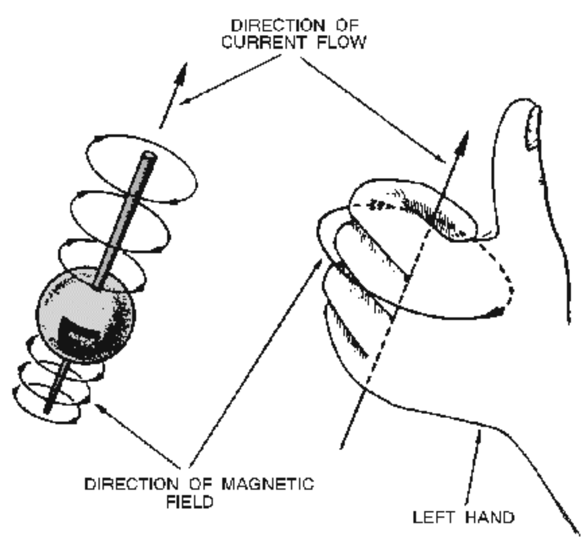


Figure 1-29.—Left-hand rule for conducting elements.

Q46. What do we call the field that is created between two rods when a voltage is applied to them?

Q47. When current flows through a conductor, a field is created around the conductor. What do we call this field?

Combined Electric and Magnetic Fields

The generator, shown in figure 1-30, provides the voltage, which creates an electric field, and current, which creates a magnetic field. This source voltage and current build up to maximum values in one direction during one half-cycle, and then build up to maximum values in the other direction during the next half-cycle. Both the electric and magnetic fields alternate from minimum through maximum values in synchronization with the changing voltage and current. The electric and magnetic fields reach their maximum intensity a quarter-cycle apart. These fields form the induction field. Since the current and voltage that produce these E and H fields are 90 degrees out of phase, the fields will also be 90 degrees out of phase.

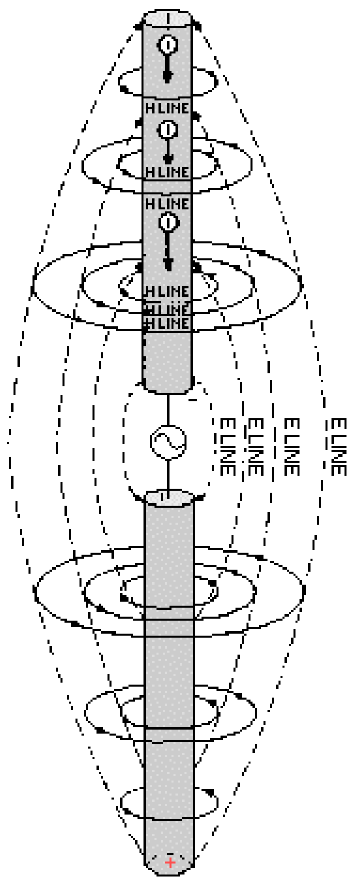


Figure 1-30.—Relationship of E-lines, and current flow.

Q48. An induction field is created around a conductor when current flows through it. What do we call the field that detaches itself from the conductor and travels through space?

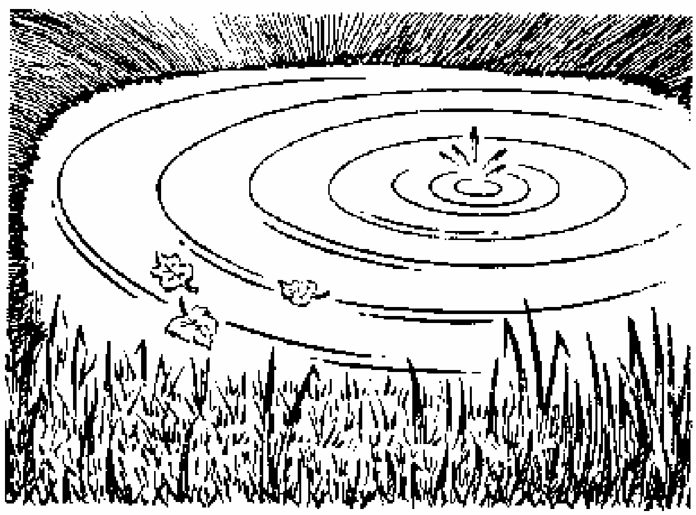
SUMMARY

Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas you have learned. You should have a thorough understanding of these principles before moving on to chapter 2.

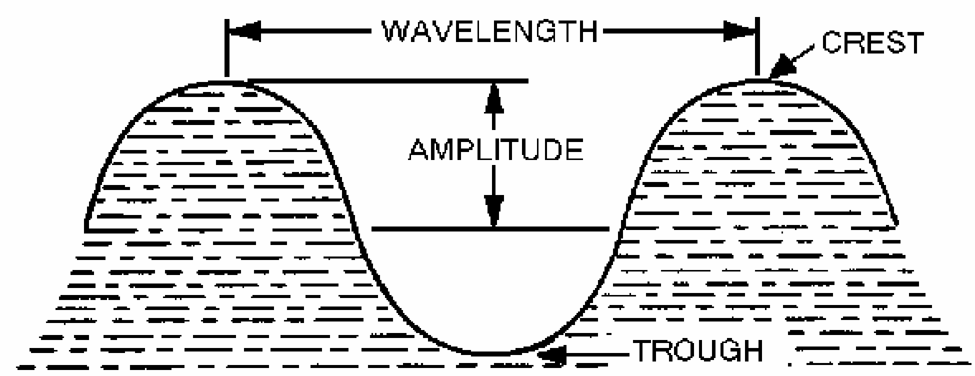
WAVE PROPAGATION is an invisible force that enables man to communicate over long distances. Wave transmission can take many forms, such as **LIGHT**, **SOUND**, and **RADIO**.

LIGHT is a form of wave motion that can be seen. Heat cannot normally be seen, but can be felt. Radio waves cannot be seen or felt.

WAVE MOTION can be seen in action by throwing a pebble into a pool of still water. The ripples that move toward the edge of the pool demonstrate the **PROPAGATION** theory.

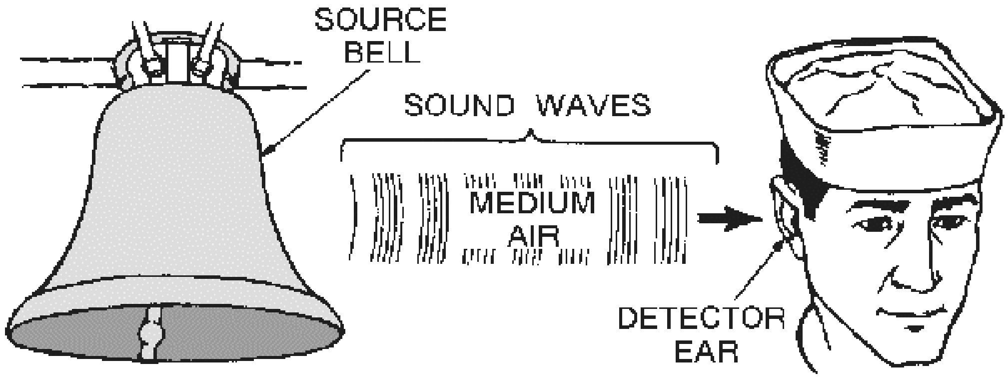


The **TRANSVERSE WAVE** is a type of wave motion. Radio, light, and heat waves are examples of transverse waves.



The **LONGITUDINAL WAVE** is another type of wave motion. The sound wave is the only example of a longitudinal wave given in this text.

SOURCE, MEDIUM, AND DETECTOR (RECEIVER) are the three requirements for all wave motion.



A **SOURCE** can be anything that emits or expends energy (waves).

The **MEDIUM** is the vehicle for carrying waves from one point to another. Water, air, metal, empty space, etc., are examples of a medium. Empty space is considered a medium for electro-magnetic waves but not a medium for sound waves.

The **SOUND DETECTOR** absorbs the waves emitted by the source. The human ear is an example of a detector.

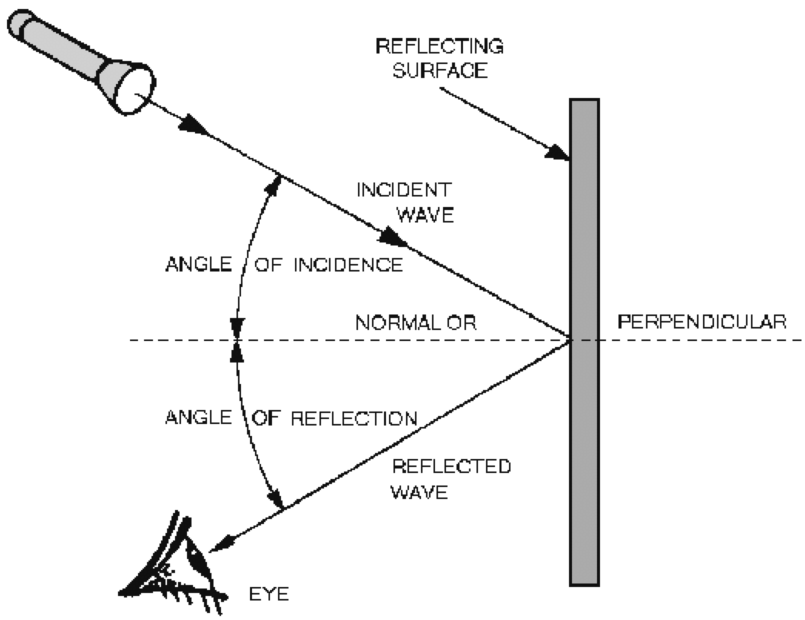
HERTZ, which is abbreviated **Hz**, is used in lieu of "cycle per second" when referring to radio frequencies.

VELOCITY OF PROPAGATION is the speed (or rate) at which the crest of a wave moves through a medium. Velocity can be calculated by using the formula:

$$V = \lambda f$$

Where v is velocity of propagation and is expressed in feet (meters) per second, λ is the wavelength in feet (meters), and f is the frequency in hertz.

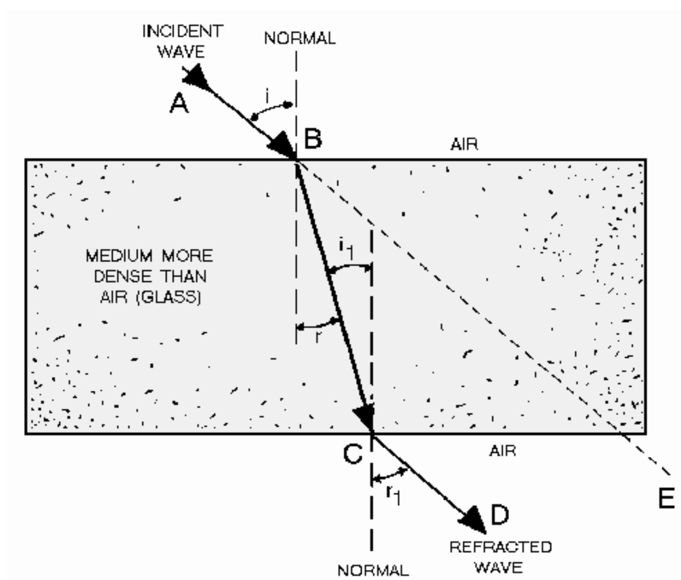
REFLECTION occurs when a wave strikes an object and bounces back (toward the source). The wave that moves from the source to the object is called the **INCIDENT WAVE**, and the wave that moves away from the object is called the **REFLECTED WAVE**.



The **LAW OF REFLECTION** states:

<u>The angle of incidence is equal to the angle of reflection</u>.

REFRACTION occurs when a wave traveling through two different mediums passes through the **BOUNDARY** of the mediums and bends toward or away from the **NORMAL**.



DIFFRACTION can account for the ability of the AM radio waves (due to their low frequency) to travel over a mountain, while FM and TV signals (due to their higher frequencies) are blocked.

The **DOPPLER EFFECT** is the apparent change in frequency of a source as it moves toward or away from a detector. It can affect the operation of equipment used to detect and measure wave energy.

SOUND can be audible to the human ear or it can be outside the hearing range.

NOISE AND TONES are the two general groups that broadly classify ALL sounds.

A MUSICAL NOTE

B NOISE

PITCH, INTENSITY, AND QUALITY are the three basic characteristics of sound. Pitch describes the frequency of sound. Intensity describes how much energy is transmitted. Quality enables us to distinguish one sound from another.

The **DENSITY** of a **MEDIUM**, **TEMPERATURE**, and **ATMOSPHERIC PRESSURE** affect the velocity of sound. If temperature, density, or pressure increases, the velocity of sound increases and vice versa.

ACOUSTICS is the science of sound and relates to the sense of hearing.

ECHO is an example of reflection. Sound echoes are used in sonar and depth finders to determine or measure the range of an object or the depth of the ocean bottom.

REVERBERATION is the multiple reflections of sound waves. The prolonged roar of thunder is caused by reverberations. With underwater sound equipment, reverberations of nearby objects may interfere with returning echoes from actual targets.

INTERFERENCE occurs when two waves move simultaneously through a medium. They can interfere constructively, destructively, or produce a resultant of zero.

RESONANCE occurs when an objects vibrates (or resonates) at its natural frequency. When different frequencies are produced inside a cavity, the sound from the cavity sounds louder at its resonant frequency than at all other frequencies.

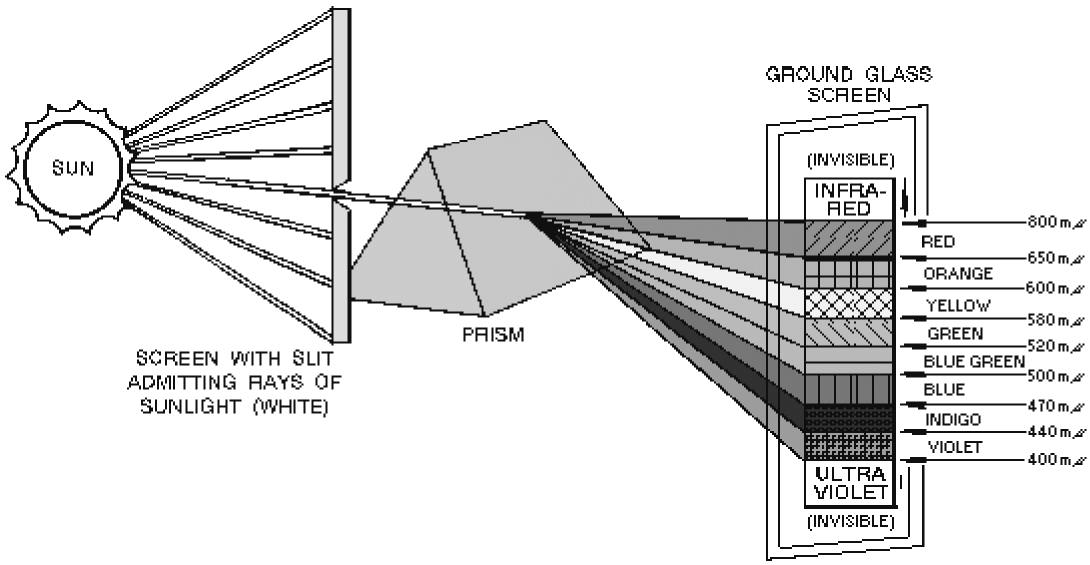
NOISE is any disturbance that distracts from or distorts the quality of sound.

A **PHOTON** is the smallest unit of radiant energy that makes up light waves and radio waves.

ANGSTROM (Å) units are used for measuring the wavelength of light. One angstrom = 1055^{-10} m.

The **VISIBLE SPECTRUM** contains all the colors between infrared and ultraviolet. **INFRA-RED** and **ULTRA-VIOLET** are invisible to the human eye.

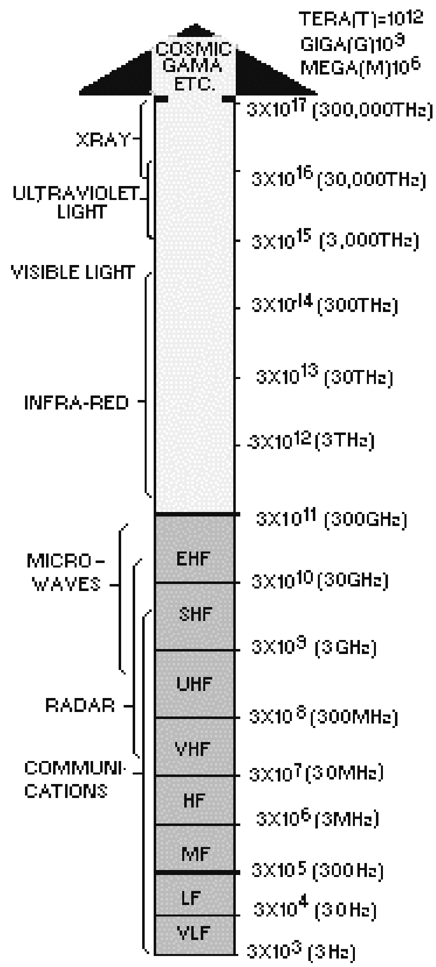
The **PRIMARY COLORS** of light are red, green, and blue. These primaries can be mixed to make any color between red and violet. If the three colors are mixed equally, they produce white light.



The **COMPLEMENTARY COLORS** of light are magenta, yellow, and cyan. They are produced by mixing any two of the primary colors together in overlapping beams.

The **SPEED OF LIGHT** in empty space is considered to be 186,000 miles per second (or 300,000,000 meters per second). This speed varies in different mediums, but the constant of 186,000 miles per second is always used as the speed of light.

The **ELECTROMAGNETIC SPECTRUM** is the complete range of electromagnetic frequencies from 3 kHz to beyond 300,000 THz. Audio frequencies (15 Hz-20 kHz) are not electromagnetic energy and are not included in the electromagnetic spectrum.



The **ELECTROMAGNETIC FIELD** consists of an **ELECTRIC FIELD** and a **MAGNETIC FIELD**. These fields are responsible for the transmission and reception of electromagnetic energy through free space.

ANSWERS TO QUESTIONS Q1. THROUGH Q48.

A1. Propagation means spreading out.

A2. A wave is a disturbance which moves through a medium.

A3. A means of transferring energy from one place to another.

A4. Sound waves, light waves, radio waves, heat waves, water waves.

A5. Transverse waves.

A6. Radio waves, light waves, and heat waves.

A7. A sound wave.

A8. A source, medium, and detector (receiver).

A9. A sequence of events, such as the positive and negative alternation of electrical current.

A10. The space occupied by one cycle of a radio wave at any given instant.

A11. The law of reflection states: The angle of incidence is equal to the angle of reflection.

A12. When the incident wave is nearly parallel with the surface.

A13. When the incident wave is perpendicular to the surface. Also a dull (or black) surface reflects very little regardless of the angle.

A14. The density of the two mediums, and the velocity of the waves.

A15. The Doppler effect.

A16. Sonics.

A17. No. The average human ear cannot hear all sounds in the infrasonic and ultrasonic regions.

A18. An amplifier.

A19. A source, medium, and detector (receiver).

A20. Noise and tones.

A21. Pitch, intensity, and quality.

A22. 20 Hz to 20 kHz.

A23. The amount of energy transmitted from a source.

A24. Quality.

A25. Velocity increases as density decreases and temperature increases.

A26. Acoustics.

A27. Echo.

A28. Reverberation.

A29. Resonance.

A30. Noise.

A31. Mechanical, electrical, and chemical.

A32. A photon.

A33. Angstrom unit.

A34. Red, green and blue.

A35. Magenta, yellow and cyan.

A36. Reflected or absorbed.

A37. None, all colors would be absorbed.

A38. Translucent.

A39. 186,000 miles per second or 300,000,000 meters per second.

A40. Transmitted.

A41. Diffused.

A42. Light waves, heat waves, and radio waves.

A43. The visible spectrum can be seen.

A44. Electric field and magnetic field.

A45. An antenna.

A46. Electric field.

A47. Magnetic field.

A48. Radiation field.

CHAPTER 2

RADIO WAVE PROPAGATION

LEARNING OBJECTIVES

Upon completion of this unit, you should be able to:

1. State what the electromagnetic field is and what components make up the electromagnetic field.
2. State the difference between the induction field and the radiation field.
3. State what radio waves are.
4. List the components of a radio wave and define the terms cycle, frequency, harmonics, period, wavelength, and velocity as applied to radio wave propagation.
5. Compute the wavelength of radio waves.
6. State how radio waves are polarized, vertically and horizontally.
7. State what reflection, refraction, and diffraction are as applied to radio waves.
8. State what influence the Earth's atmosphere has on radio waves and list the different layers of the Earth's atmosphere.
9. Identify a ground wave, a sky wave, and state the effects of the ionosphere on the sky wave.
10. Identify the structure of the ionosphere.
11. Define density of layer, frequency, angle of incidence, skip distance, and skip zone.
12. Describe propagation paths.
13. Describe fading, multipath fading, and selective fading. Describe propagation paths.
14. State how transmission losses affect radio wave propagation.
15. State how electromagnetic interference, man-made/natural interference, and ionospheric disturbances affect radio wave propagation. State how transmission losses affect radio wave propagation.
16. Identify variations in the ionosphere.
17. Identify the maximum, optimum, and lowest usable frequencies of radio waves.
18. State what temperature inversion is, how frequency predictions are made, and how weather affects frequency.
19. State what tropospheric scatter is and how it affects radio wave propagation.

ELECTROMAGNETIC FIELDS

The way energy is propagated into free space is a source of great dispute among people concerned with it. Although many theories have been proposed, the following theory adequately explains the phenomena and has been widely accepted. There are two basic fields associated with every antenna; an INDUCTION FIELD and a RADIATION FIELD. The field associated with the energy stored in the antenna is the induction field. This field is said to provide no part in the transmission of electromagnetic energy through free space. However, without the presence of the induction field, there would be no energy radiated.

INDUCTION FIELD

Figure 2-1, a low-frequency generator connected to an antenna, will help you understand how the induction field is produced. Let's follow the generator through one cycle of operation.

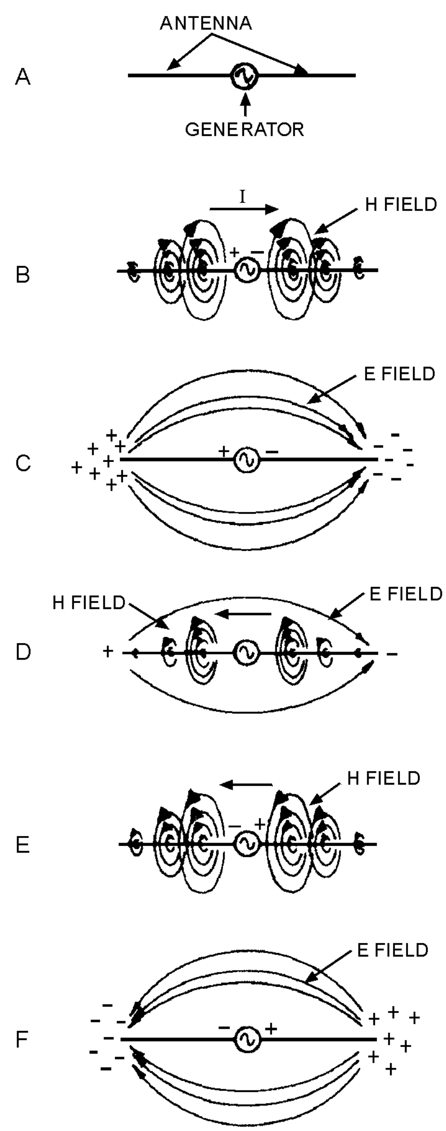


Figure 2-1.—Induction field about an antenna.

Initially, you can consider that the generator output is zero and that no fields exist about the antenna, as shown in view A. Now assume that the generator produces a slight potential and has the instantaneous polarity shown in view B. Because of this slight potential, the antenna capacitance acts as a short, allowing a large flow of current (I) through the antenna in the direction shown. This current flow, in turn, produces a large magnetic field about the antenna. Since the flow of current at each end of the antenna is minimum, the corresponding magnetic fields at each end of the antenna are also minimum. As time passes, charges, which oppose antenna current and produce an electrostatic field (E field), collect at each end of the antenna. Eventually, the antenna capacitance becomes fully charged and stops current flow through the antenna. Under this condition, the electrostatic field is maximum, and the magnetic field (H field) is fully collapsed, as shown in view C.

As the generator potential decreases back to zero, the potential of the antenna begins to discharge. During the discharging process, the electrostatic field collapses and the direction of current flow reverses, as shown in view D. When the current again begins to flow, an associated magnetic field is generated. Eventually, the electrostatic field completely collapses, the generator potential reverses, and current is maximum, as shown in view E. As charges collect at each end of the antenna, an electrostatic field is produced and current flow decreases. This causes the magnetic field to begin collapsing. The collapsing magnetic field produces more current flow, a greater accumulation of charge, and a greater electrostatic field. The antenna gradually reaches the condition shown in view F, where current is zero and the collected charges are maximum.

As the generator potential again decreases toward zero, the antenna begins to discharge and the electrostatic field begins to collapse. When the generator potential reaches zero, discharge current is maximum and the associated magnetic field is maximum. A brief time later, generator potential reverses, and the condition shown in view B recurs.

NOTE: The electric field (E field) and the electrostatic field (E field) are the same. They will be used interchangeably throughout this text.

The graph shown in figure 2-2 shows the relationship between the magnetic (H) field and the electric (E) field plotted against time. Note that the two fields are 90 degrees out of phase with each other. If you compare the graph in figure 2-2 with figure 2-1, you will notice that the two fields around the antenna are displaced 90 degrees from each other in space. (The H field exists in a plane perpendicular to the antenna. The E field exists in a plane parallel with the antenna, as shown in figure 2-1.)

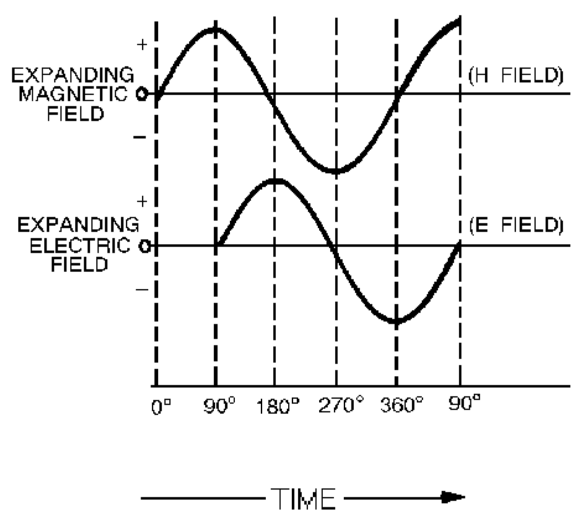


Figure 2-2.—Phase relationship of induction field components.

All the energy supplied to the induction field is returned to the antenna by the collapsing E and H fields. No energy from the induction field is radiated from the antenna. Therefore, the induction field is considered a local field and plays no part in the transmission of electromagnetic energy. The induction field represents only the stored energy in the antenna and is responsible only for the resonant effects that the antenna reflects to the generator.

RADIATION FIELDS

The E and H fields that are set up in the transfer of energy through space are known collectively as the radiation field. This radiation field is responsible for electromagnetic radiation from the antenna. The radiation field decreases as the distance from the antenna is increased. Because the decrease is linear, the radiation field reaches great distances from the antenna.

Let's look at a half-wave antenna to illustrate how this radiation actually takes place. Simply stated, a half-wave antenna is one that has an electrical length equal to half the wavelength of the signal being transmitted. Assume, for example, that a transmitter is operating at 30 megahertz. If a half-wave antenna is used with the transmitter, the antenna's electrical length would have to be at least 16 feet long. (The formula used to compute the electrical length of an antenna will be explained in chapter 4.) When power is delivered to the half-wave antenna, both an induction field and a radiation field are set up by the fluctuating energy. At the antenna, the intensities of these fields are proportional to the amount of power delivered to the antenna from a source such as a transmitter. At a short distance from the antenna and beyond, only the radiation field exists. This radiation field is made up of an electric component and a magnetic component at right angles to each other in space and varying together in intensity.

With a high-frequency generator (a transmitter) connected to the antenna, the induction field is produced as described in the previous section. However, the generator potential reverses before the electrostatic field has had time to collapse completely. The reversed generator potential neutralizes the remaining antenna charges, leaving a resultant E field in space.

Figure 2-3 is a simple picture of an E field detaching itself from an antenna. (The H field will not be considered, although it is present.) In view A the voltage is maximum and the electric field has maximum intensity. The lines of force begin at the end of the antenna that is positively charged and extend to the end of the antenna that is negatively charged. Note that the outer E lines are stretched away from the inner lines. This is because of the repelling force that takes place between lines of force in the same direction. As the voltage drops (view B), the separated charges come together, and the ends of the lines move toward the center of the antenna. But, since lines of force in the same direction repel each other, the centers of the lines are still being held out.

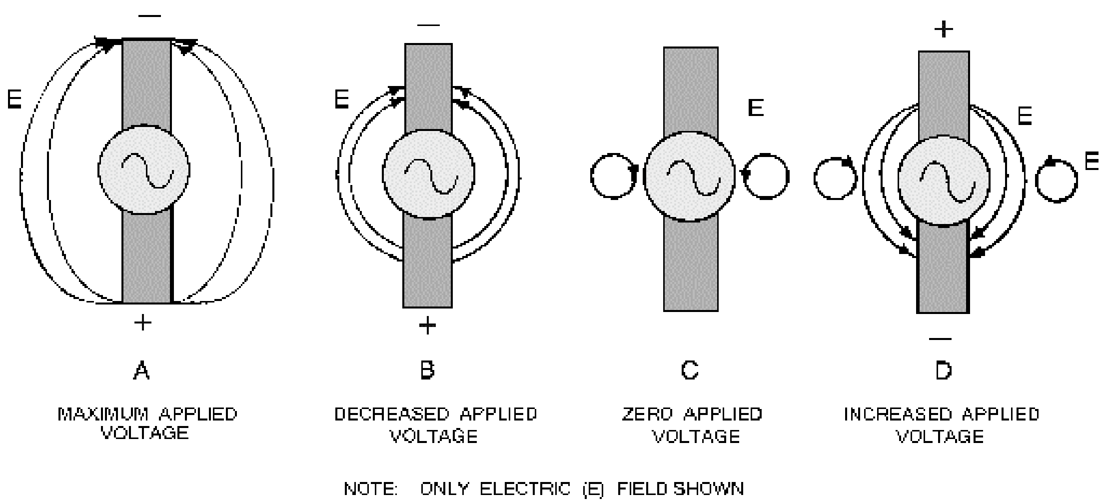


Figure 2-3.—Radiation from an antenna.

As the voltage approaches zero (view B), some of the lines collapse back into the antenna. At the same time, the ends of other lines begin to come together to form a complete loop. Notice the direction of these lines of force next to the antenna in view C. At this point the voltage on the antenna is zero. As the charge starts to build up in the opposite direction (view D), electric lines of force again begin at the positive end of the antenna and stretch to the negative end of the antenna. These lines of force, being in the same direction as the sides of the closed loops next to the antenna, repel the closed loops and force them out into space at the speed of light. As these loops travel through space, they generate a magnetic field in phase with them.

Since each successive E field is generated with a polarity that is opposite the preceding E field (that is, the lines of force are opposite), an oscillating electric field is produced along the path of travel. When an electric field oscillates, a magnetic field having an intensity that varies directly with that of the E field is produced. The variations in magnetic field intensity, in turn, produce another E field. Thus, the two varying fields sustain each other, resulting in electromagnetic wave propagation.

During this radiation process, the E and H fields are in phase in time but physically displaced 90 degrees in space. Thus, the varying magnetic field produces a varying electric field; and the varying electric field, in turn, sustains the varying magnetic field. Each field supports the other, and neither can be propagated by itself. Figure 2-4 shows a comparison between the induction field and the radiation field.

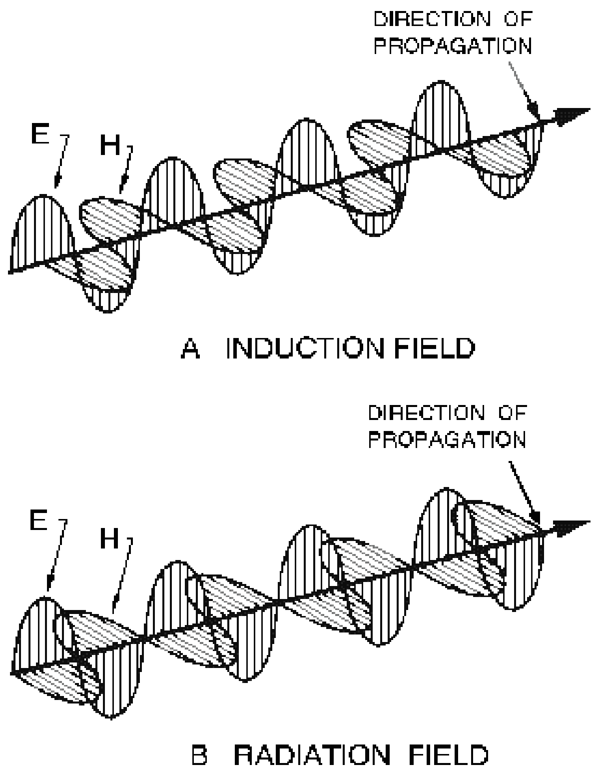


Figure 2-4.—E and H components of induction and radiation fields.

Q1. Which two composite fields (composed of E and H fields) are associated with every antenna?

Q2. What composite field (composed of E and H fields) is found stored in the antenna?

Q3. What composite field (composed of E and H fields) is propagated into free space?

RADIO WAVES

An energy wave generated by a transmitter is called a RADIO WAVE. The radio wave radiated into space by the transmitting antenna is a very complex form of energy containing both electric and magnetic fields. Because of this combination of fields, radio waves are also referred to as ELECTROMAGNETIC RADIATION.

This discussion will explain the Earth's atmosphere and its effect on radio waves. All the principles of wave motion that were discussed in chapter 1 also apply to radio waves.

NOTE: The term *radio wave* is not limited to communications equipment alone. The term applies to all equipment that generate signals in the form of electromagnetic energy.

COMPONENTS OF RADIO WAVES

The basic shape of the wave generated by a transmitter is that of a sine wave. The wave radiated out into space, however, may or may not retain the characteristics of the sine wave.

A sine wave can be one cycle or many cycles. Recall from chapter 1 that the number of cycles of a sine wave that are completed in 1 second is known as the *frequency* of the sine wave. For example, 60 cycles of ordinary house current occur each second, so house current is said to have a frequency of 60 cycles per second or 60 hertz.

The frequencies falling between 3000 hertz (3 kHz) and 300,000,000,000 hertz (300 GHz) are called RADIO FREQUENCIES (abbreviated rf) since they are commonly used in radio communications. This part of the radio frequency spectrum is divided into bands, each band being 10 times higher in frequency than the one immediately below it. This arrangement serves as a convenient way to remember the range of each band. The rf bands are shown in table 2-1. The usable radio-frequency range is roughly 10 kilohertz to 100 gigahertz.

Table 2-1.—Radio Frequency Bands

DESCRIPTION	ABBREVIATION	FREQUENCY
Very low	VLF	3 to 30 KHz
Low	LF	30 to 300 KHz
Medium	MF	300 to 3000 KHz
High	HF	3 to 30 MHz
Very high	VHF	30 to 300 MHz
Ultrahigh	UHF	300 to 3000 MHz
Super high	SHF	3 to 30 GHz
Extremely high	EHF	30 to 300 GHz

Any frequency that is a whole number multiple of a smaller basic frequency is known as a HARMONIC of that basic frequency. The basic frequency itself is called the first harmonic or, more commonly, the FUNDAMENTAL FREQUENCY. A frequency that is twice as great as the fundamental frequency is called the second harmonic; a frequency three times as great is the third harmonic; and so on. For example:

First harmonic (Fundamental frequency)	3000 kHz
Second harmonic	6000 kHz
Third harmonic	9000 kHz

The PERIOD of a radio wave is simply the amount of time required for the completion of one full cycle. If a sine wave has a frequency of 2 hertz, each cycle has a duration, or period, of one-half second. If the frequency is 10 hertz, the period of each cycle is one-tenth of a second. Since the frequency of a radio wave is the number of cycles that are completed in one second, you should be able to see that as the frequency of a radio wave increases, its period decreases.

A wavelength is the space occupied by one full cycle of a radio wave at any given instant. Wavelengths are expressed in meters (1 meter is equal to 3.28 feet). You need to have a good understanding of frequency and wavelength to be able to select the proper antenna(s) for use in successful communications. The relationship between frequency, wavelength, and antennas will be discussed in chapter 4 of this module.

The velocity (or speed) of a radio wave radiated into free space by a transmitting antenna is equal to the speed of light—186,000 miles per second or 300,000,000 meters per second. Because of various factors, such as barometric pressure, humidity, molecular content, etc., radio waves travel inside the Earth's atmosphere at a speed slightly less than the speed of light. Normally, in discussions of the velocity of radio waves, the velocity referred to is the speed at which radio waves travel in free space.

The frequency of a radio wave has nothing to do with its velocity. A 5-megahertz wave travels through space at the same velocity as a 10-megahertz wave. However, the velocity of radio waves is an important factor in making wavelength-to-frequency conversions, the subject of our next discussion.

Q4. What is the term used to describe the basic frequency of a radio wave?

Q5. What is the term used to describe a whole number multiple of the basic frequency of a radio wave?

WAVELENGTH-TO-FREQUENCY CONVERSIONS

Radio waves are often referred to by their wavelength in meters rather than by frequency. For example, most people have heard commercial radio stations make announcements similar to the following: "Station WXYZ operating on 240 meters..." To tune receiving equipment that is calibrated by frequency to such a station, you must first convert the designated wavelength to its equivalent frequency.

As discussed earlier, a radio wave travels 300,000,000 meters a second (speed of light); therefore, a radio wave of 1 hertz would have traveled a distance (or wavelength) of 300,000,000 meters. Obviously then, if the frequency of the wave is increased to 2 hertz, the wavelength will be cut in half to 150,000,000 meters. This illustrates the principle that the HIGHER THE FREQUENCY, the SHORTER THE WAVELENGTH.

Wavelength-to-frequency conversions of radio waves are really quite simple because wavelength and frequency are reciprocals: Either one divided into the velocity of a radio wave yields the other. Remember, the formula for wavelength is:

$$\lambda = \frac{v}{f} \quad \text{or} \quad f = \frac{v}{\lambda}$$

Where:

λ = wavelength in meters

v = velocity of radio wave (speed of light)

f = frequency of radio wave (in Hz, kHz or Mhz)

The wavelength in meters divided into 300,000,000 yields the <u>frequency of a radio wave in hertz</u>. Likewise, the wavelength divided into 300,000 yields the <u>frequency of a radio wave in kilohertz</u>, and the wavelength divided into 300 yields the frequency in megahertz.

Now, let us apply the formula to determine the frequency to which the receiving equipment must be tuned to receive station WXYZ operating on 240 meters. Radio wave frequencies are normally expressed in kilohertz or megahertz.

To find the frequency in hertz, use the formula:

$$f = \frac{v}{\lambda}$$

Given:

v = 300,000,000 meters per second

λ = 240 meters

Solution:

$$f = \frac{300{,}000{,}000 \text{ meters per second}}{240 \text{ meters}}$$

$$f = 1{,}250{,}000 \text{ Hz}$$

To find the frequency in kilohertz, use the formula:

$$f_{[kHz]} = \frac{300{,}000}{\lambda}$$

Given:

λ = 240 meters

Solution:

$$f_{[kHz]} = \frac{300{,}000}{240 \text{ meters}}$$

$$f = 1250\text{kHz}$$

To find the frequency in megahertz, use the formula:

$$f_{[MHz]} = \frac{300}{\lambda}$$

Given:

λ = 240 meters

Solution:

$$f_{[MHz]} = \frac{300}{240 \text{ meters}}$$

$$f = 1.25\text{MHz}$$

Q6. It is known that WWV operates on a frequency of 10 megahertz. What is the wavelength of WWV?

Q7. A station is known to operate at 60-meters. What is the frequency of the unknown station?

POLARIZATION

For maximum absorption of energy from the electromagnetic fields, the receiving antenna must be located in the plane of polarization. This places the conductor of the antenna at right angles to the magnetic lines of force moving through the antenna and parallel to the electric lines, causing maximum induction.

Normally, the plane of polarization of a radio wave is the plane in which the E field propagates with respect to the Earth. If the E field component of the radiated wave travels in a plane perpendicular to the Earth's surface (vertical), the radiation is said to be VERTICALLY POLARIZED, as shown in figure 2-5, view A. If the E field propagates in a plane parallel to the Earth's surface (horizontal), the radiation is said to be HORIZONTALLY POLARIZED, as shown in view B.

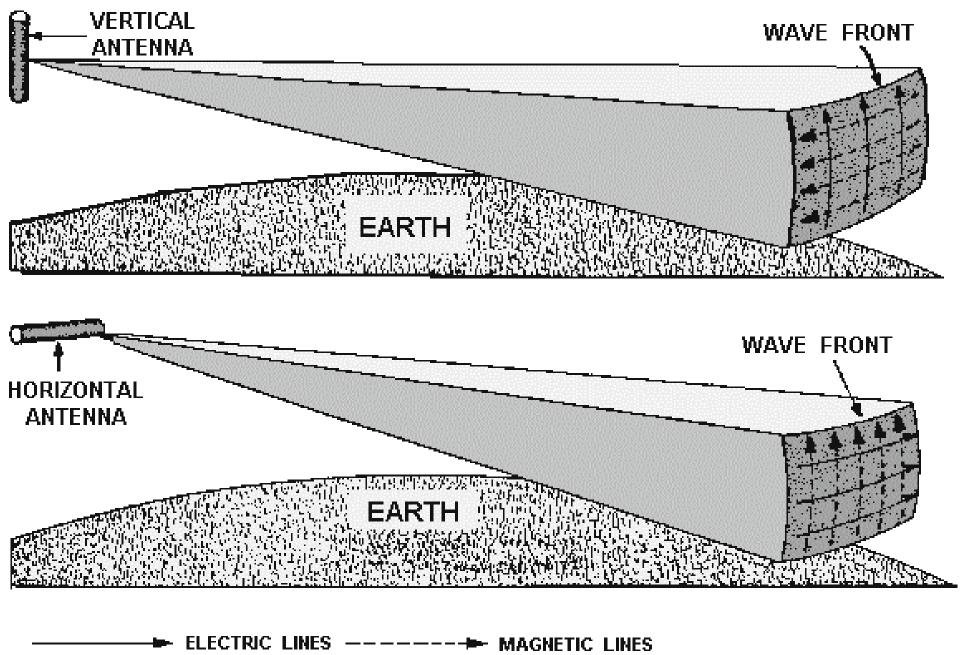


Figure 2-5.—Vertical and horizontal polarization.

The position of the antenna in space is important because it affects the polarization of the electromagnetic wave. When the transmitting antenna is close to the ground, vertically polarized waves cause a greater signal strength along the Earth's surface. On the other hand, antennas high above the ground should be horizontally polarized to get the greatest possible signal strength to the Earth's surface. Vertically and horizontally polarized antennas will be discussed in more detail in chapter 4.

The radiated energy from an antenna is in the form of an expanding sphere. Any small section of this sphere is perpendicular to the direction the energy travels and is called a WAVEFRONT. All energy on a wavefront is in phase. Usually all points on the wavefront are at equal distances from the antenna. The farther the wavefront is from the antenna, the less spherical the wave appears. At a considerable distance the wavefront can be considered as a plane surface at a right angle to the direction of propagation.

If you know the directions of the E and H components, you can use the "right-hand rule" (see figure 2-6) to determine the direction of wave propagation. This rule states that if the thumb, forefinger, and middle finger of the right hand are extended so they are mutually perpendicular, the middle finger will point in the direction of wave propagation if the thumb points in the direction of the E field and the forefinger points in the direction of the H field. Since both the E and H fields reverse directions simultaneously, propagation of a particular wavefront is always in the same direction (away from the antenna).

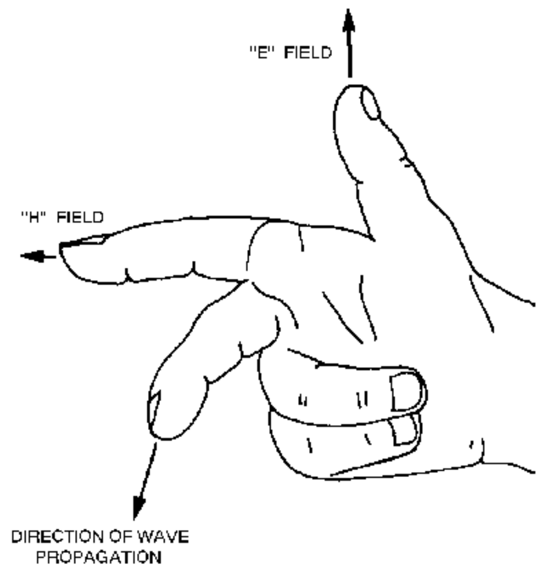


Figure 2-6.—Right-hand rule for propagation.

Q8. If a transmitting antenna is placed close to the ground, how should the antenna be polarized to give the greatest signal strength?

Q9. In the right-hand rule for propagation, the thumb points in the direction of the E field and the forefinger points in the direction of the H field. In what direction does the middle finger point?

ATMOSPHERIC PROPAGATION

Within the atmosphere, radio waves can be reflected, refracted, and diffracted like light and heat waves.

Reflection

Radio waves may be reflected from various substances or objects they meet during travel between the transmitting and receiving sites. The amount of reflection depends on the reflecting material. Smooth metal surfaces of good electrical conductivity are efficient reflectors of radio waves. The surface of the Earth itself is a fairly good reflector. The radio wave is not reflected from a single point on the reflector but rather from an area on its surface. The size of the area required for reflection to take place depends on the wavelength of the radio wave and the angle at which the wave strikes the reflecting substance.

When radio waves are reflected from flat surfaces, a phase shift in the alternations of the wave occurs. Figure 2-7 shows two radio waves being reflected from the Earth's surface. Notice that the positive and negative alternations of radio waves (A) and (B) are in phase with each other in their paths toward the Earth's surface. After reflection takes place, however, the waves are approximately 180 degrees out of phase from their initial relationship. The amount of phase shift that occurs is not constant.

It depends on the polarization of the wave and the angle at which the wave strikes the reflecting surface. Radio waves that keep their phase relationships after reflection normally produce a stronger signal at the receiving site. Those that are received out of phase produce a weak or fading signal. The shifting in the phase relationships of reflected radio waves is one of the major reasons for fading. Fading will be discussed in more detail later in this chapter.

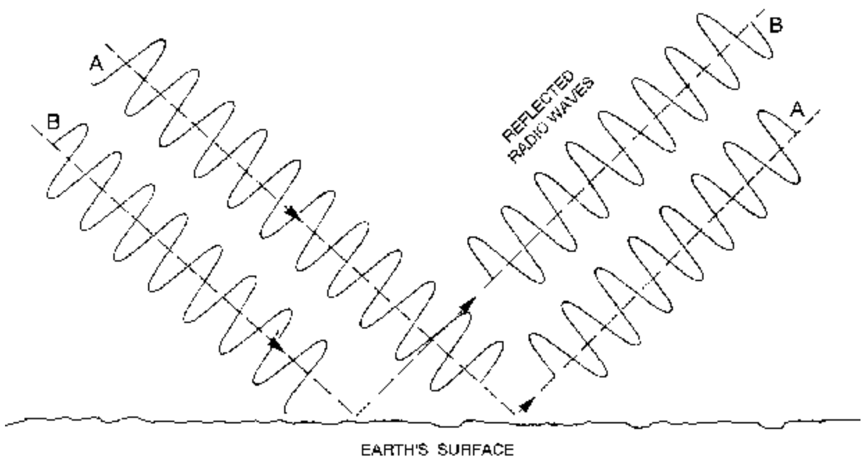


Figure 2-7.—Phase shift of reflected radio waves.

Refraction

Another phenomenon common to most radio waves is the bending of the waves as they move from one medium into another in which the velocity of propagation is different. This bending of the waves is called refraction. For example, suppose you are driving down a smoothly paved road at a constant speed and suddenly one wheel goes off onto the soft shoulder. The car tends to veer off to one side. The change of medium, from hard surface to soft shoulder, causes a change in speed or velocity. The tendency is for the car to change direction. This same principle applies to radio waves as changes occur in the medium through which they are passing. As an example, the radio wave shown in figure 2-8 is traveling through the Earth's atmosphere at a constant speed. As the wave enters the dense layer of electrically charged ions, the part of the wave that enters the new medium first travels faster than the parts of the wave that have not yet entered the new medium. This abrupt increase in velocity of the upper part of the wave causes the wave to bend back toward the Earth. This bending, or change of direction, is always toward the medium that has the lower velocity of propagation.

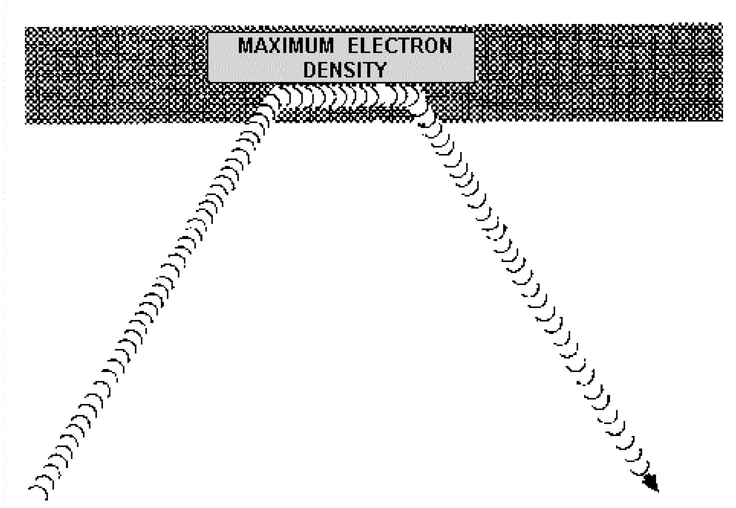


Figure 2-8.—Radio wave refraction.

Radio waves passing through the atmosphere are affected by certain factors, such as temperature, pressure, humidity, and density. These factors can cause the radio waves to be refracted. This effect will be discussed in greater detail later in this chapter.

Diffraction

A radio wave that meets an obstacle has a natural tendency to bend around the obstacle as illustrated in figure 2-9. The bending, called diffraction, results in a change of direction of part of the wave energy from the normal line-of-sight path. This change makes it possible to receive energy around the edges of an obstacle as shown in view A or at some distances below the highest point of an obstruction, as shown in view B. Although diffracted rf energy usually is weak, it can still be detected by a suitable receiver. The principal effect of diffraction extends the radio range beyond the visible horizon. In certain cases, by using high power and very low frequencies, radio waves can be made to encircle the Earth by diffraction.

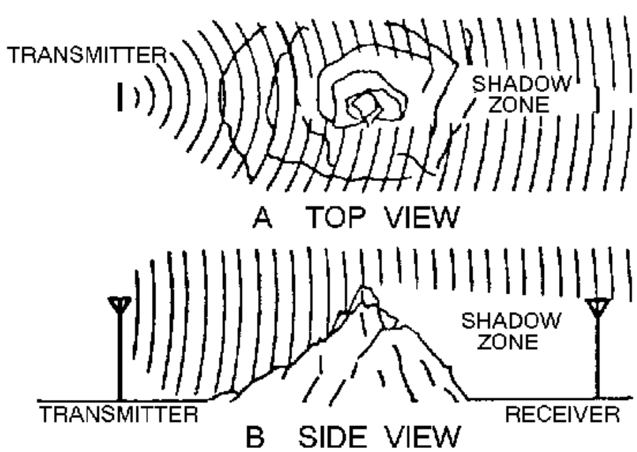


Figure 2-9.—Diffraction around an object.

Q10. What is one of the major reasons for the fading of radio waves which have been reflected from a surface?

THE EFFECT OF THE EARTH'S ATMOSPHERE ON RADIO WAVES

This discussion of electromagnetic wave propagation is concerned mainly with the properties and effects of the medium located between the transmitting antenna and the receiving antenna. While radio waves traveling in free space have little outside influence affecting them, radio waves traveling within the Earth's atmosphere are affected by varying conditions. The influence exerted on radio waves by the Earth's atmosphere adds many new factors to complicate what at first seems to be a relatively simple problem. These complications are because of a lack of uniformity within the Earth's atmosphere. Atmospheric conditions vary with changes in height, geographical location, and even with changes in time (day, night, season, year). A knowledge of the composition of the Earth's atmosphere is extremely important for understanding wave propagation.

The Earth's atmosphere is divided into three separate regions, or layers. They are the TROPOSPHERE, the STRATOSPHERE, and the IONOSPHERE. The layers of the atmosphere are illustrated in figure 2-10.

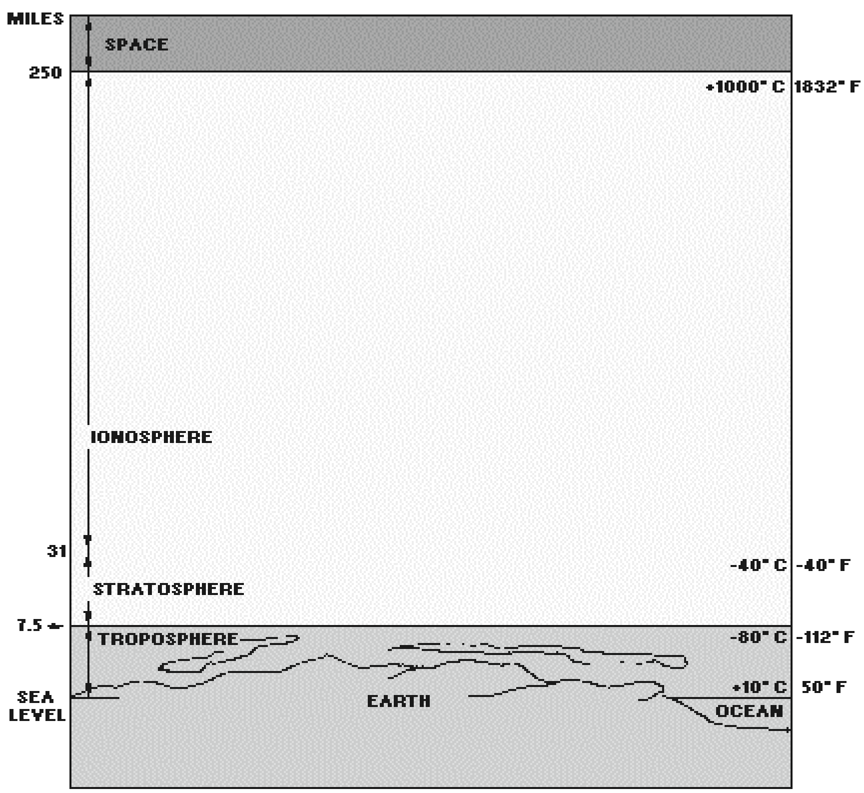


Figure 2-10.—Layers of the earth's atmosphere.

TROPOSPHERE

The troposphere is the portion of the Earth's atmosphere that extends from the surface of the Earth to a height of about 3.7 miles (6 km) at the North Pole or the South Pole and 11.2 miles (18 km) at the equator. Virtually all weather phenomena take place in the troposphere. The temperature in this region decreases rapidly with altitude, clouds form, and there may be much turbulence because of variations in temperature, density, and pressure. These conditions have a great effect on the propagation of radio waves, which will be explained later in this chapter.

STRATOSPHERE

The stratosphere is located between the troposphere and the ionosphere. The temperature throughout this region is considered to be almost constant and there is little water vapor present. The stratosphere has relatively little effect on radio waves because it is a relatively calm region with little or no temperature changes.

IONOSPHERE

The ionosphere extends upward from about 31.1 miles (50 km) to a height of about 250 miles (402 km). It contains four cloud-like layers of electrically charged ions, which enable radio waves to be propagated to great distances around the Earth. This is the most important region of the atmosphere for long distance point-to-point communications. This region will be discussed in detail a little later in this chapter.

Q11. What are the three layers of the atmosphere?

Q12. Which layer of the atmosphere has relatively little effect on radio waves?

RADIO WAVE TRANSMISSION

There are two principal ways in which electromagnetic (radio) energy travels from a transmitting antenna to a receiving antenna. One way is by GROUND WAVES and the other is by SKY WAVES. Ground waves are radio waves that travel near the surface of the Earth (surface and space waves). Sky waves are radio waves that are reflected back to Earth from the ionosphere. (See figure 2-11.)

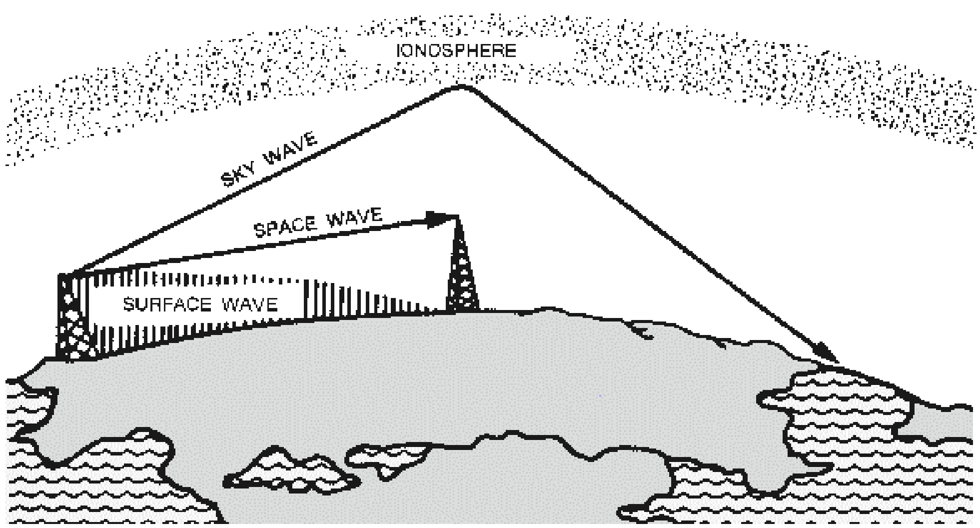


Figure 2-11.—Ground waves and sky waves.

Ground Waves

The ground wave is actually composed of two separate component waves. These are known as the SURFACE WAVE and the SPACE WAVE (fig. 2-11). The determining factor in whether a ground wave component is classified as a space wave or a surface wave is simple. A surface wave travels along the surface of the Earth. A space wave travels over the surface.

SURFACE WAVE.—The surface wave reaches the receiving site by traveling along the surface of the ground as shown in figure 2-12. A surface wave can follow the contours of the Earth because of the process of diffraction. When a surface wave meets an object and the dimensions of the object do not exceed its wavelength, the wave tends to curve or bend around the object. The smaller the object, the more pronounced the diffractive action will be.

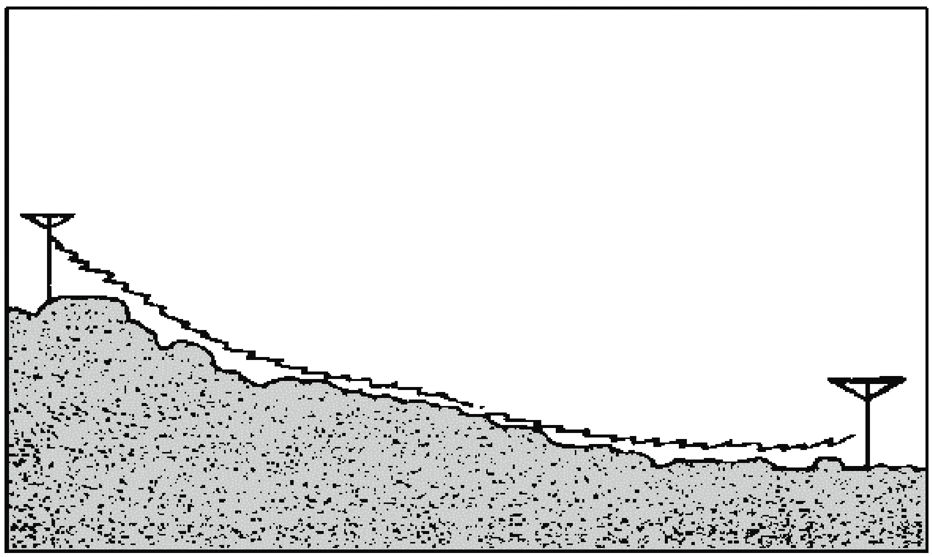


Figure 2-12.—Surface wave propagation.

As a surface wave passes over the ground, the wave induces a voltage in the Earth. The induced voltage takes energy away from the surface wave, thereby weakening, or attenuating, the wave as it moves away from the transmitting antenna. To reduce the attenuation, the amount of induced voltage must be reduced. This is done by using vertically polarized waves that minimize the extent to which the electric field of the wave is in contact with the Earth. When a surface wave is horizontally polarized, the electric field of the wave is parallel with the surface of the Earth and, therefore, is constantly in contact with it. The wave is then completely attenuated within a short distance from the transmitting site. On the other hand, when the surface wave is vertically polarized, the electric field is vertical to the Earth and merely dips into and out of the Earth's surface. For this reason, vertical polarization is vastly superior to horizontal polarization for surface wave propagation.

The attenuation that a surface wave undergoes because of induced voltage also depends on the electrical properties of the terrain over which the wave travels. The best type of surface is one that has good electrical conductivity. The better the conductivity, the less the attenuation. Table 2-2 gives the relative conductivity of various surfaces of the Earth.

Table 2-2.—Surface Conductivity

SURFACE	RELATIVE CONDUCTIVITY
Sea water	Good
Flat, loamy soil	Fair
Large bodies of fresh water	Fair
Rocky terrain	Poor
Desert	Poor
Jungle	Unusable

Another major factor in the attenuation of surface waves is frequency. Recall from earlier discussions on wavelength that the higher the frequency of a radio wave, the shorter its wavelength will be. These high frequencies, with their shorter wavelengths, are not normally diffracted but are absorbed by the Earth at points relatively close to the transmitting site. You can assume, therefore, that as the frequency of a surface wave is increased, the more rapidly the surface wave will be absorbed, or attenuated, by the Earth. Because of this loss by attenuation, the surface wave is impractical for long-distance transmissions at frequencies above 2 megahertz. On the other hand, when the frequency of a surface wave is low enough to have a very long wavelength, the Earth appears to be very small, and diffraction is sufficient for propagation well beyond the horizon. In fact, by lowering the transmitting frequency into the very low frequency (vlf) range and using very high-powered transmitters, the surface wave can be propagated great distances. The Navy's extremely high-powered vlf transmitters are actually capable of transmitting surface wave signals around the Earth and can provide coverage to naval units operating anywhere at sea.

SPACE WAVE.—The space wave follows two distinct paths from the transmitting antenna to the receiving antenna—one through the air directly to the receiving antenna, the other reflected from the ground to the receiving antenna. This is illustrated in figure 2-13. The primary path of the space wave is directly from the transmitting antenna to the receiving antenna. So, the receiving antenna must be located within the radio horizon of the transmitting antenna. Because space waves are refracted slightly, even when propagated through the troposphere, the radio horizon is actually about one-third farther than the line-of-sight or natural horizon.

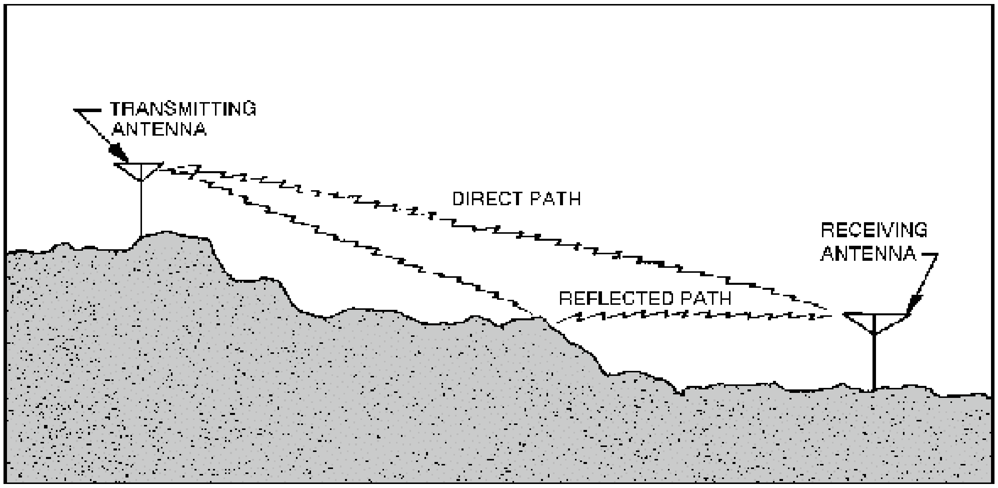


Figure 2-13.—Space wave propagation.

Although space waves suffer little ground attenuation, they nevertheless are susceptible to fading. This is because space waves actually follow two paths of different lengths (direct path and ground reflected path) to the receiving site and, therefore, may arrive in or out of phase. If these two component waves are received in phase, the result is a reinforced or stronger signal. Likewise, if they are received out of phase, they tend to cancel one another, which results in a weak or fading signal.

Q13. What is the determining factor in classifying whether a radio wave is a ground wave or a space wave?

Q14. What is the best type of surface or terrain to use for radio wave transmission?

Q15. What is the primary difference between the radio horizon and the natural horizon?

Q16. What three factors must be considered in the transmission of a surface wave to reduce attenuation?

Sky Wave

The sky wave, often called the ionospheric wave, is radiated in an upward direction and returned to Earth at some distant location because of refraction from the ionosphere. This form of propagation is relatively unaffected by the Earth's surface and can propagate signals over great distances. Usually the high frequency (hf) band is used for sky wave propagation. The following in-depth study of the ionosphere and its effect on sky waves will help you to better understand the nature of sky wave propagation.

STRUCTURE OF THE IONOSPHERE

As we stated earlier, the ionosphere is the region of the atmosphere that extends from about 30 miles above the surface of the Earth to about 250 miles. It is appropriately named the ionosphere because it consists of several layers of electrically charged gas atoms called ions. The ions are formed by a process called ionization.

Ionization

Ionization occurs when high energy ultraviolet light waves from the sun enter the ionospheric region of the atmosphere, strike a gas atom, and literally knock an electron free from its parent atom. A normal atom is electrically neutral since it contains both a positive proton in its nucleus and a negative orbiting electron. When the negative electron is knocked free from the atom, the atom becomes positively charged (called a positive ion) and remains in space along with the free electron, which is negatively charged. This process of upsetting electrical neutrality is known as IONIZATION.

The free negative electrons subsequently absorb part of the ultraviolet energy, which initially freed them from their atoms. As the ultraviolet light wave continues to produce positive ions and negative electrons, its intensity decreases because of the absorption of energy by the free electrons, and an ionized layer is formed. The rate at which ionization occurs depends on the density of atoms in the atmosphere and the intensity of the ultraviolet light wave, which varies with the activity of the sun.

Since the atmosphere is bombarded by ultraviolet light waves of different frequencies, several ionized layers are formed at different altitudes. Lower frequency ultraviolet waves penetrate the atmosphere the least; therefore, they produce ionized layers at the higher altitudes. Conversely, ultraviolet waves of higher frequencies penetrate deeper and produce layers at the lower altitudes.

An important factor in determining the density of ionized layers is the elevation angle of the sun, which changes frequently. For this reason, the height and thickness of the ionized layers vary, depending on the time of day and even the season of the year.

Recombination

Recall that the process of ionization involves ultraviolet light waves knocking electrons free from their atoms. A reverse process called RECOMBINATION occurs when the free electrons and positive ions collide with each other. Since these collisions are inevitable, the positive ions return to their original neutral atom state.

The recombination process also depends on the time of day. Between the hours of early morning and late afternoon, the rate of ionization exceeds the rate of recombination. During this period, the ionized layers reach their greatest density and exert maximum influence on radio waves. During the late afternoon and early evening hours, however, the rate of recombination exceeds the rate of ionization, and the density of the ionized layers begins to decrease. Throughout the night, density continues to decrease, reaching a low point just before sunrise.

Four Distinct Layers

The ionosphere is composed of three layers designated D, E, and F, from lowest level to highest level as shown in figure 2-14. The F layer is further divided into two layers designated F1 (the lower layer) and F2 (the higher layer). The presence or absence of these layers in the ionosphere and their height above the Earth varies with the position of the sun. At high noon, radiation in the ionosphere directly above a given point is greatest. At night it is minimum. When the radiation is removed, many of the particles that were ionized recombine. The time interval between these conditions finds the position and number of the ionized layers within the ionosphere changing. Since the position of the sun varies daily, monthly, and yearly, with respect to a specified point on Earth, the exact position and number of layers present are extremely difficult to determine. However, the following general statements can be made:

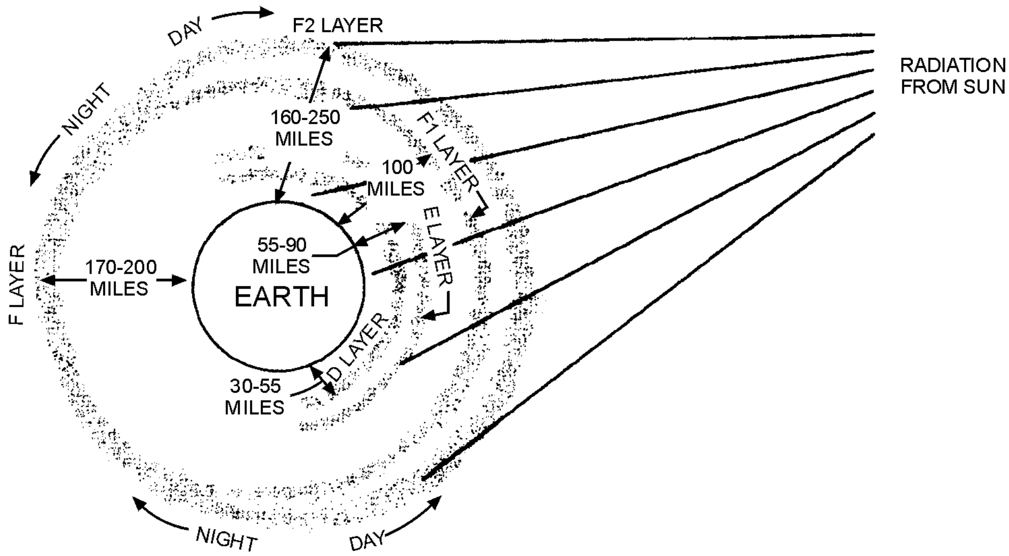


Figure 2-14.—Layers of the ionosphere.

a. The D layer ranges from about 30 to 55 miles. Ionization in the D layer is low because it is the lowest region of the ionosphere. This layer has the ability to refract signals of low frequencies. High frequencies pass right through it and are attenuated. After sunset, the D layer disappears because of the rapid recombination of ions.

b. The E layer limits are from about 55 to 90 miles. This layer is also known as the Kennelly-Heaviside layer, because these two men were the first to propose its existence. The rate of ionic recombination in this layer is rather rapid after sunset and the layer is almost gone by midnight. This layer has the ability to refract signals as high as 20 megahertz. For this reason, it is valuable for communications in ranges up to about 1500 miles.

c. The F layer exists from about 90 to 240 miles. During the daylight hours, the F layer separates into two layers, the F1 and F2 layers. The ionization level in these layers is quite high and varies widely during the day. At noon, this portion of the atmosphere is closest to the sun and the degree of ionization is maximum. Since the atmosphere is rarefied at these heights, recombination occurs slowly after sunset. Therefore, a fairly constant ionized layer is always present. The F layers are responsible for high-frequency, long distance transmission.

Q17. What causes ionization to occur in the ionosphere?

Q18. How are the four distinct layers of the ionosphere designated?

Q19. What is the height of the individual layers of the ionosphere?

REFRACTION IN THE IONOSPHERE

When a radio wave is transmitted into an ionized layer, refraction, or bending of the wave, occurs. As we discussed earlier, refraction is caused by an abrupt change in the velocity of the upper part of a radio wave as it strikes or enters a new medium. The amount of refraction that occurs depends on three main factors: (1) the density of ionization of the layer, (2) the frequency of the radio wave, and (3) the angle at which the wave enters the layer.

Density of Layer

Figure 2-15 illustrates the relationship between radio waves and ionization density. Each ionized layer has a central region of relatively dense ionization, which tapers off in intensity both above and below the maximum region. As a radio wave enters a region of INCREASING ionization, the increase in velocity of the upper part of the wave causes it to be bent back TOWARD the Earth. While the wave is in the highly dense center portion of the layer, however, refraction occurs more slowly because the density of ionization is almost uniform. As the wave enters into the upper part of the layer of DECREASING ionization, the velocity of the upper part of the wave decreases, and the wave is bent AWAY from the Earth.

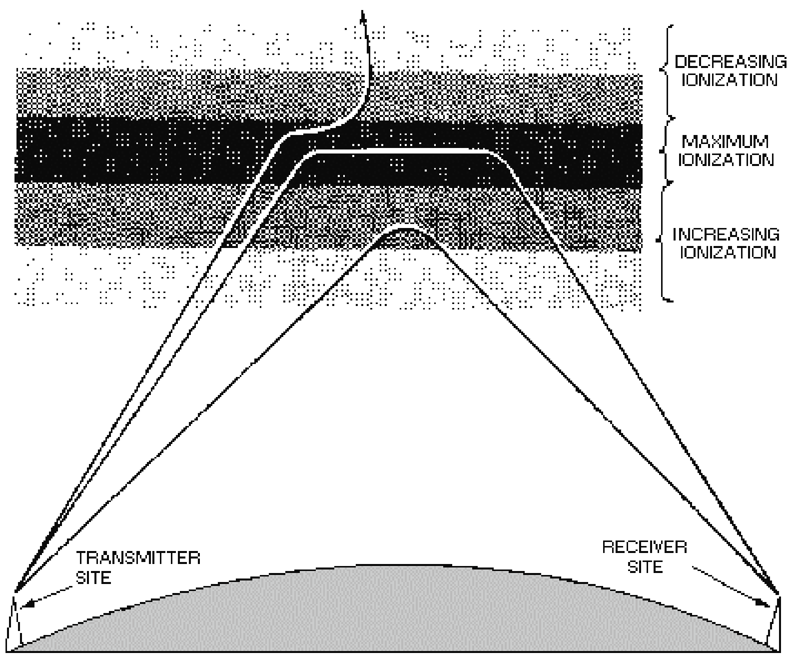


Figure 2-15.—Effects of ionospheric density on radio waves.

If a wave strikes a thin, very highly ionized layer, the wave may be bent back so rapidly that it will appear to have been reflected instead of refracted back to Earth. To reflect a radio wave, the highly ionized layer must be approximately no thicker than one wavelength of the radio wave. Since the ionized layers are often several miles thick, ionospheric reflection is more likely to occur at long wavelengths (low frequencies).

Frequency

For any given time, each ionospheric layer has a maximum frequency at which radio waves can be transmitted vertically and refracted back to Earth. This frequency is known as the CRITICAL FREQUENCY. It is a term that you will hear frequently in any discussion of radio wave propagation. Radio waves transmitted at frequencies higher than the critical frequency of a given layer will pass through the layer and be lost in space; but if these same waves enter an upper layer with a higher critical frequency, they will be refracted back to Earth. Radio waves of frequencies lower than the critical frequency will also be refracted back to Earth unless they are absorbed or have been refracted from a lower layer. The lower the frequency of a radio wave, the more rapidly the wave is refracted by a given degree of ionization. Figure 2-16 shows three separate waves of different frequencies entering an ionospheric layer at the same angle. Notice that the 5-megahertz wave is refracted quite sharply. The 20-megahertz wave is refracted less sharply and returned to Earth at a greater distance. The 100-megahertz wave is obviously greater than the critical frequency for that ionized layer and, therefore, is not refracted but is passed into space.

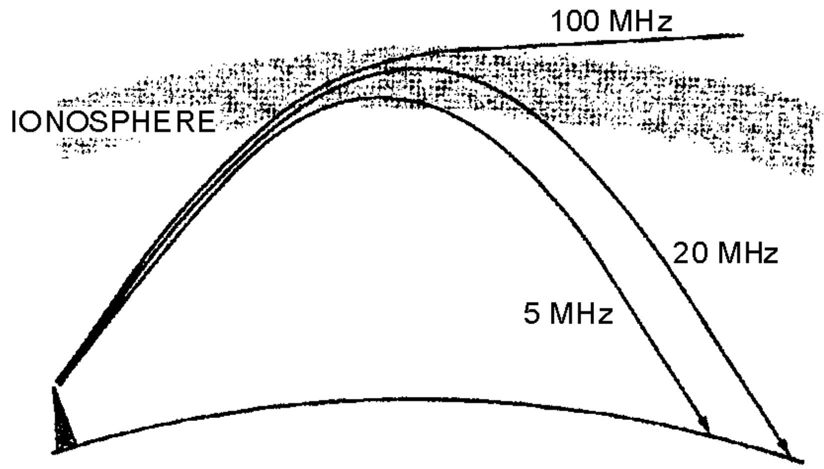


Figure 2-16.—Frequency versus refraction and distance.

Angle of Incidence

The rate at which a wave of a given frequency is refracted by an ionized layer depends on the angle at which the wave enters the layer. Figure 2-17 shows three radio waves of the same frequency entering a layer at different angles. The angle at which wave A strikes the layer is too nearly vertical for the wave to be refracted to Earth. As the wave enters the layer, it is bent slightly but passes through the layer and is lost. When the wave is reduced to an angle that is less than vertical (wave B), it strikes the layer and is refracted back to Earth. The angle made by wave B is called the CRITICAL ANGLE for that particular frequency. Any wave that leaves the antenna at an angle greater than the critical angle will penetrate the ionospheric layer for that frequency and then be lost in space. Wave C strikes the ionosphere at the smallest angle at which the wave can be refracted and still return to Earth. At any smaller angle, the wave will be refracted but will not return to Earth.

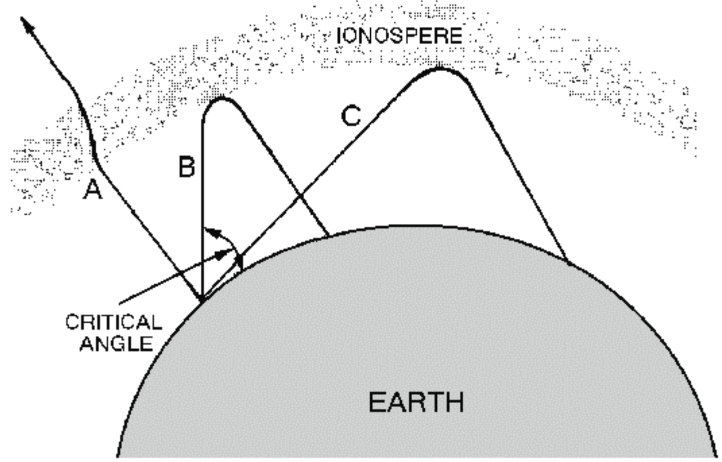


Figure 2-17.—Different incident angles of radio waves.

As the frequency of the radio wave is increased, the critical angle must be reduced for refraction to occur. This is illustrated in figure 2-18. The 2-megahertz wave strikes the layer at the critical angle for that frequency and is refracted back to Earth. Although the 5-megahertz wave (broken line) strikes the ionosphere at a lesser angle, it nevertheless penetrates the layer and is lost. As the angle is lowered from the vertical, however, a critical angle for the 5-megahertz wave is reached, and the wave is then refracted to Earth.

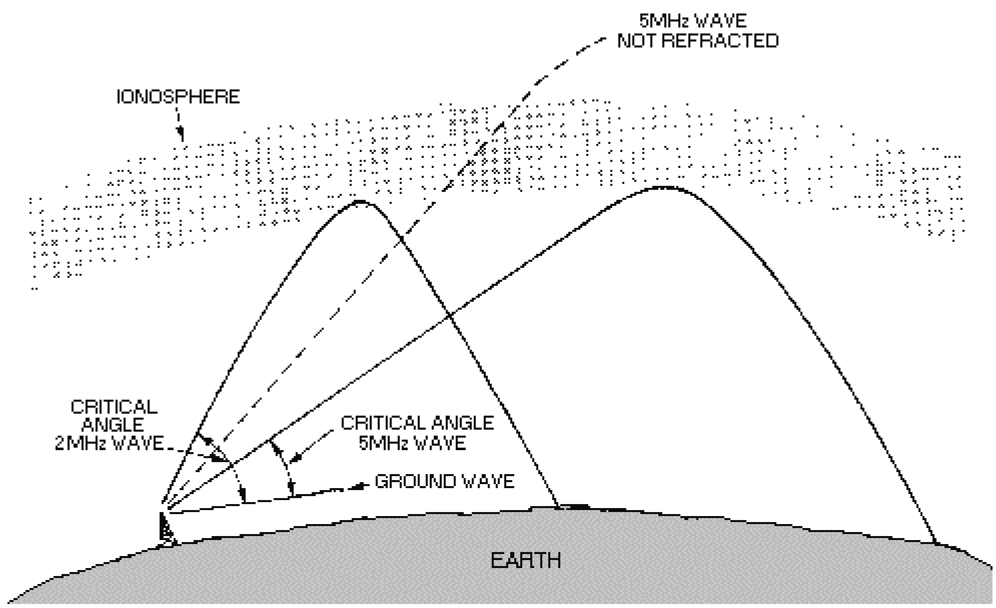


Figure 2-18.—Effects of frequency on the critical angle.

Q20. What factor determines whether a radio wave is reflected or refracted by the ionosphere?

Q21. There is a maximum frequency at which vertically transmitted radio waves can be refracted back to Earth. What is this maximum frequency called?

Q22. What three main factors determine the amount of refraction in the ionosphere?

Skip Distance/Skip Zone

In figure 2-19, note the relationship between the sky wave skip distance, the skip zone, and the ground wave coverage. The SKIP DISTANCE is the distance from the transmitter to the point where the sky wave is first returned to Earth. The size of the skip distance depends on the frequency of the wave, the angle of incidence, and the degree of ionization present.

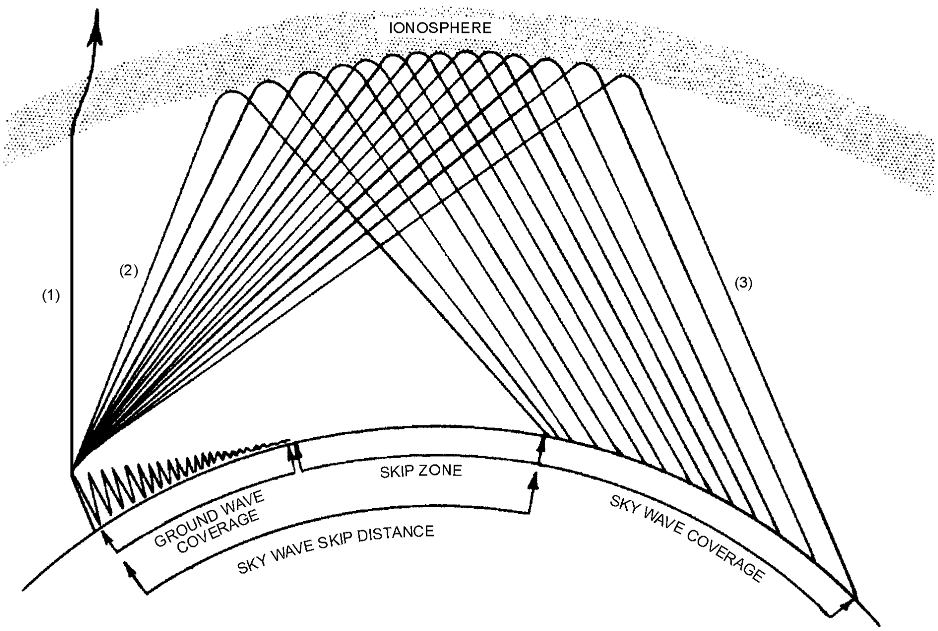


Figure 2-19.—Relationship between skip zone, skip distance, and ground wave.

The SKIP ZONE is a zone of silence between the point where the ground wave becomes too weak for reception and the point where the sky wave is first returned to Earth. The size of the skip zone depends on the extent of the ground wave coverage and the skip distance. When the ground wave coverage is great enough or the skip distance is short enough that no zone of silence occurs, there is no skip zone.

Occasionally, the first sky wave will return to Earth within the range of the ground wave. If the sky wave and ground wave are nearly of equal intensity, the sky wave alternately reinforces and cancels the ground wave, causing severe fading. This is caused by the phase difference between the two waves, a result of the longer path traveled by the sky wave.

PROPAGATION PATHS

The path that a refracted wave follows to the receiver depends on the angle at which the wave strikes the ionosphere. You should remember, however, that the rf energy radiated by a transmitting antenna spreads out with distance. The energy therefore strikes the ionosphere at many different angles rather than a single angle.

After the rf energy of a given frequency enters an ionospheric region, the paths that this energy might follow are many. It may reach the receiving antenna via two or more paths through a single layer. It may also, reach the receiving antenna over a path involving more than one layer, by multiple hops between the ionosphere and Earth, or by any combination of these paths.

Figure 2-20 shows how radio waves may reach a receiver via several paths through one layer. The various angles at which rf energy strikes the layer are represented by dark lines and designated as rays 1 through 6.

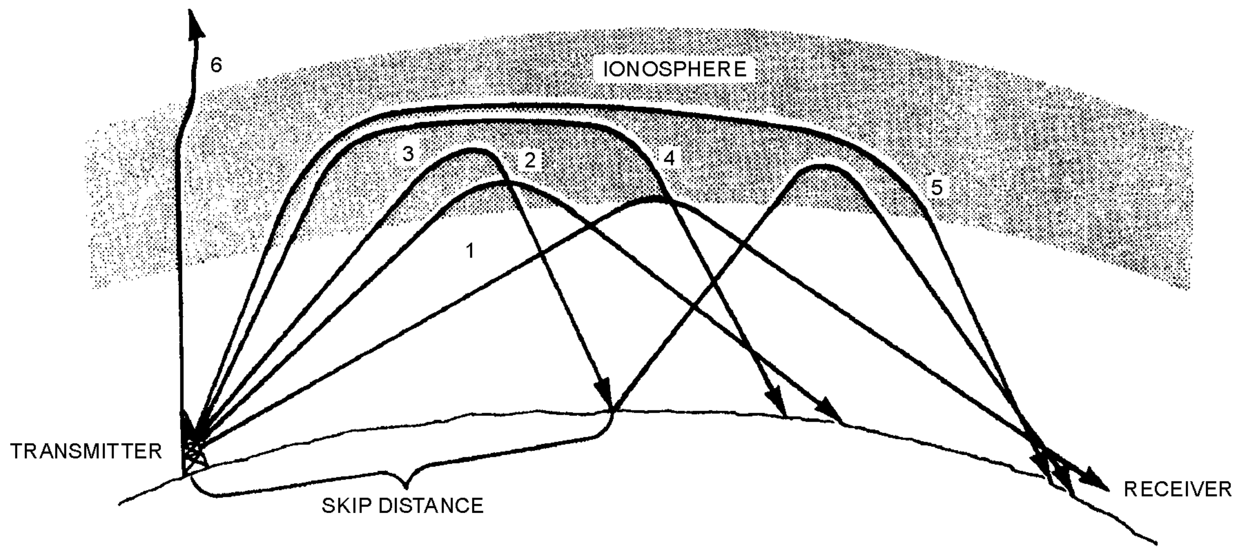


Figure 2-20.—Ray paths for a fixed frequency with varying angles of incidence.

When the angle is relatively low with respect to the horizon (ray 1), there is only slight penetration of the layer and the propagation path is long. When the angle of incidence is increased (rays 2 and 3), the rays penetrate deeper into the layer but the range of these rays decreases. When a certain angle is reached (ray 3), the penetration of the layer and rate of refraction are such that the ray is first returned to Earth at a minimal distance from the transmitter. Notice, however, that ray 3 still manages to reach the receiving site on its second refraction (called a hop) from the ionospheric layer.

As the angle is increased still more (rays 4 and 5), the rf energy penetrates the central area of maximum ionization of the layer. These rays are refracted rather slowly and are eventually returned to Earth at great distances. As the angle approaches vertical incidence (ray 6), the ray is not returned at all, but passes on through the layer.

ABSORPTION IN THE IONOSPHERE

Many factors affect a radio wave in its path between the transmitting and receiving sites. The factor that has the greatest adverse effect on radio waves is ABSORPTION. Absorption results in the loss of energy of a radio wave and has a pronounced effect on both the strength of received signals and the ability to communicate over long distances.

You learned earlier in the section on ground waves that surface waves suffer most of their absorption losses because of ground-induced voltage. Sky waves, on the other hand, suffer most of their absorption losses because of conditions in the ionosphere. Note that some absorption of sky waves may also occur at lower atmospheric levels because of the presence of water and water vapor. However, this becomes important only at frequencies above 10,000 megahertz.

Most ionospheric absorption occurs in the lower regions of the ionosphere where ionization density is greatest. As a radio wave passes into the ionosphere, it loses some of its energy to the free electrons and ions. If these high-energy free electrons and ions do not collide with gas molecules of low energy, most of the energy lost by the radio wave is reconverted into electromagnetic energy, and the wave continues to be propagated with little change in intensity. However, if the high-energy free electrons and ions do collide with other particles, much of this energy is lost, resulting in absorption of the energy from the wave. Since absorption of energy depends on collision of the particles, the greater the density of the ionized layer, the greater the probability of collisions; therefore, the greater the absorption. The highly dense D and E layers provide the greatest absorption of radio waves.

Because the amount of absorption of the sky wave depends on the density of the ionosphere, which varies with seasonal and daily conditions, it is impossible to express a fixed relationship between distance and signal strength for ionospheric propagation. Under certain conditions, the absorption of energy is so great that communicating over any distance beyond the line of sight is difficult.

FADING

The most troublesome and frustrating problem in receiving radio signals is variations in signal strength, most commonly known as FADING. There are several conditions that can produce fading. When a radio wave is refracted by the ionosphere or reflected from the Earth's surface, random changes in the polarization of the wave may occur. Vertically and horizontally mounted receiving antennas are designed to receive vertically and horizontally polarized waves, respectively. Therefore, changes in polarization cause changes in the received signal level because of the inability of the antenna to receive polarization changes.

Fading also results from absorption of the rf energy in the ionosphere. Absorption fading occurs for a longer period than other types of fading, since absorption takes place slowly.

Usually, however, fading on ionospheric circuits is mainly a result of multipath propagation.

Multipath Fading

MULTIPATH is simply a term used to describe the multiple paths a radio wave may follow between transmitter and receiver. Such propagation paths include the ground wave, ionospheric refraction, reradiation by the ionospheric layers, reflection from the Earth's surface or from more than one ionospheric layer, etc. Figure 2-21 shows a few of the paths that a signal can travel between two sites in a typical circuit. One path, XYZ, is the basic ground wave. Another path, XEA, refracts the wave at the E layer and passes it on to the receiver at A. Still another path, XFZFA, results from a greater angle of incidence and two refractions from the F layer. At point Z, the received signal is a combination of the ground wave and the sky wave. These two signals having traveled different paths arrive at point Z at different times. Thus, the arriving waves may or may not be in phase with each other. Radio waves that are received in phase reinforce each other and produce a stronger signal at the receiving site. Conversely, those that are received out of phase produce a weak or fading signal. Small alternations in the transmission path may change the phase relationship of the two signals, causing periodic fading. This condition occurs at point A. At this point, the double-hop F layer signal may be in or out of phase with the signal arriving from the E layer.

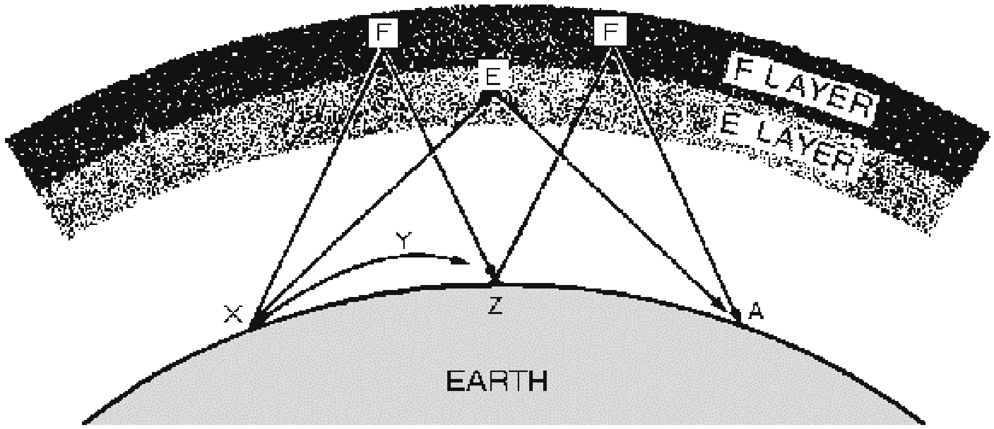


Figure 2-21.—Multipath transmission.

Multipath fading may be minimized by practices called SPACE DIVERSITY and FREQUENCY DIVERSITY. In space diversity, two or more receiving antennas are spaced some distance apart. Fading does not occur simultaneously at both antennas; therefore, enough output is almost always available from one of the antennas to provide a useful signal. In frequency diversity, two transmitters and two receivers are used, each pair tuned to a different frequency, with the same information being transmitted simultaneously over both frequencies. One of the two receivers will almost always provide a useful signal.

Selective Fading

Fading resulting from multipath propagation is variable with frequency since each frequency arrives at the receiving point via a different radio path. When a wide band of frequencies is transmitted simultaneously, each frequency will vary in the amount of fading. This variation is called SELECTIVE FADING. When selective fading occurs, all frequencies of the transmitted signal do not retain their original phases and relative amplitudes. This fading causes severe distortion of the signal and limits the total signal transmitted.

Q23. What is the skip zone of a radio wave?

Q24. Where does the greatest amount of ionospheric absorption occur in the ionosphere?

Q25. What is meant by the term "multipath"?

Q26. When a wide band of frequencies is transmitted simultaneously, each frequency will vary in the amount of fading. What is this variable fading called?

TRANSMISSION LOSSES

All radio waves propagated over ionospheric paths undergo energy losses before arriving at the receiving site. As we discussed earlier, absorption in the ionosphere and lower atmospheric levels account for a large part of these energy losses. There are two other types of losses that also significantly affect the ionospheric propagation of radio waves. These losses are known as ground reflection loss and free space loss. The combined effects of absorption, ground reflection loss, and free space loss account for most of the energy losses of radio transmissions propagated by the ionosphere.

Ground Reflection Loss

When propagation is accomplished via multihop refraction, rf energy is lost each time the radio wave is reflected from the Earth's surface. The amount of energy lost depends on the frequency of the wave, the angle of incidence, ground irregularities, and the electrical conductivity of the point of reflection.

Free space Loss

Normally, the major loss of energy is because of the spreading out of the wavefront as it travels away from the transmitter. As the distance increases, the area of the wavefront spreads out, much like the beam of a flashlight. This means the amount of energy contained within any unit of area on the wavefront will decrease as distance increases. By the time the energy arrives at the receiving antenna, the wavefront is so spread out that the receiving antenna extends into only a very small fraction of the wavefront. This is illustrated in figure 2-22.

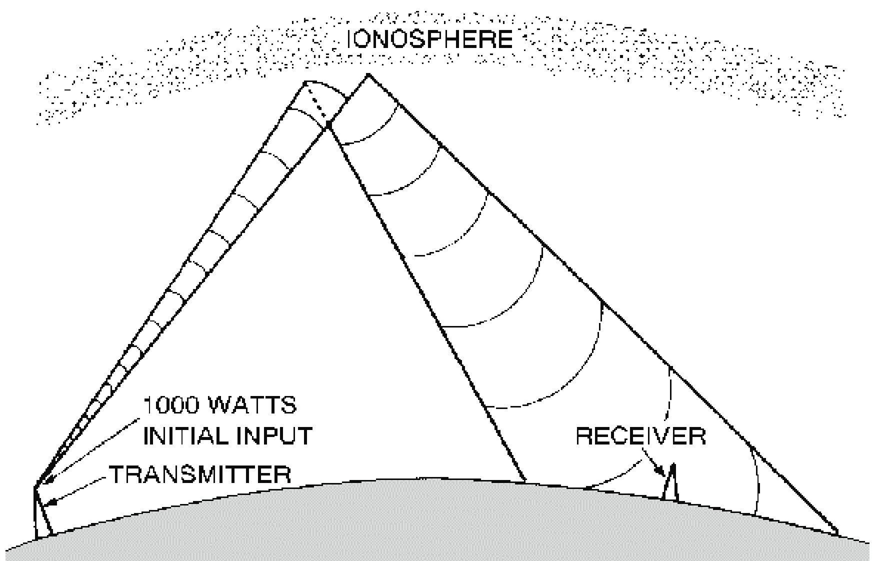


Figure 2-22.—Free space loss principle.

ELECTROMAGNETIC INTERFERENCE (EMI)

The transmission losses just discussed are not the only factors that interfere with communications. An additional factor that can interfere with radio communications is the presence of ELECTROMAGNETIC INTERFERENCE (EMI). This interference can result in annoying or impossible operating conditions. Sources of emi are both man-made and natural.

Man-Made Interference

Man-made interference may come from several sources. Some of these sources, such as oscillators, communications transmitters, and radio transmitters, may be specifically designed to generate radio frequency energy. Some electrical devices also generate radio frequency energy, although they are not specifically designed for this purpose. Examples are ignition systems, generators, motors, switches, relays, and voltage regulators. The intensity of man-made interference may vary throughout the day and drop off to a low level at night when many of these sources are not being used. Man-made interference may be a critical limiting factor at radio receiving sites located near industrial areas.

Natural Interference

Natural interference refers to the static that you often hear when listening to a radio. This interference is generated by natural phenomena, such as thunderstorms, snowstorms, cosmic sources, and the sun. The energy released by these sources is transmitted to the receiving site in roughly the same manner as radio waves. As a result, when ionospheric conditions are favorable for the long distance propagation of radio waves, they are likewise favorable for the propagation of natural interference. Natural interference is very erratic, particularly in the hf band, but generally will decrease as the operating frequency is increased and wider bandwidths are used. There is little natural interference above 30 megahertz.

Control of EMI

Electromagnetic interference can be reduced or eliminated by using various suppression techniques. The amount of emi that is produced by a radio transmitter can be controlled by cutting transmitting antennas to the correct frequency, limiting bandwidth, and using electronic filtering networks and metallic shielding.

Radiated emi during transmission can be controlled by the physical separation of the transmitting and receiving antennas, the use of directional antennas, and limiting antenna bandwidth.

Q27. What are the two main sources of emi with which radio waves must compete?

Q28. Thunderstorms, snowstorms, cosmic sources, the sun, etc., are a few examples of emi sources. What type of emi comes from these sources?

Q29. Motors, switches, voltage regulators, generators, etc., are a few examples of emi sources. What type of emi comes from these sources?

Q30. What are three ways of controlling the amount of transmitter-generated emi?

Q31. What are three ways of controlling radiated emi during transmission?

VARIATIONS IN THE IONOSPHERE

Because the existence of the ionosphere is directly related to radiations emitted from the sun, the movement of the Earth about the sun or changes in the sun's activity will result in variations in the ionosphere. These variations are of two general types: (1) those which are more or less regular and occur in cycles and, therefore, can be predicted in advance with reasonable accuracy, and (2) those which are irregular as a result of abnormal behavior of the sun and, therefore, cannot be predicted in advance. Both regular and irregular variations have important effects on radio wave propagation.

Regular Variations

The regular variations that affect the extent of ionization in the ionosphere can be divided into four main classes: daily, seasonal, 11-year, and 27-day variations.

DAILY.—Daily variations in the ionosphere are a result of the 24-hour rotation of the Earth about its axis. Daily variations of the different layers (fig. 2-14) are summarized as follows:

- The D layer reflects vlf waves; is important for long range vlf communications; refracts lf and mf waves for short range communications; absorbs hf waves; has little effect on vhf and above; and disappears at night.

- In the E layer, ionization depends on the angle of the sun. The E layer refracts hf waves during the day up to 20 megahertz to distances of about 1200 miles. Ionization is greatly reduced at night.

- Structure and density of the F region depend on the time of day and the angle of the sun. This region consists of one layer during the night and splits into two layers during daylight hours.

- Ionization density of the F1 layer depends on the angle of the sun. Its main effect is to absorb hf waves passing through to the F2 layer.

- The F2 layer is the most important layer for long distance hf communications. It is a very variable layer and its height and density change with time of day, season, and sunspot activity.

SEASONAL.—Seasonal variations are the result of the Earth revolving around the sun; the relative position of the sun moves from one hemisphere to the other with changes in seasons. Seasonal variations of the D, E, and F1 layers correspond to the highest angle of the sun; thus the ionization density of these layers is greatest during the summer. The F2 layer, however, does not follow this pattern; its ionization is greatest in winter and least in summer, the reverse of what might be expected. As a result, operating frequencies for F2 layer propagation are higher in the winter than in the summer.

ELEVEN-YEAR SUN SPOT CYCLE.—One of the most notable phenomena on the surface of the sun is the appearance and disappearance of dark, irregularly shaped areas known as SUNSPOTS. The exact nature of sunspots is not known, but scientists believe they are caused by violent eruptions on the sun and are characterized by unusually strong magnetic fields. These sunspots are responsible for variations in the ionization level of the ionosphere. Sunspots can, of course, occur unexpectedly, and the life span of individual sunspots is variable; however, a regular cycle of sunspot activity has also been observed. This cycle has both a minimum and maximum level of sunspot activity that occur approximately every 11 years.

During periods of maximum sunspot activity, the ionization density of all layers increases. Because of this, absorption in the D layer increases and the critical frequencies for the E, F1, and F2 layers are higher. At these times, higher operating frequencies must be used for long distance communications.

27-DAY SUNSPOT CYCLE.—The number of sunspots in existence at any one time is continually subject to change as some disappear and new ones emerge. As the sun rotates on its own axis, these sunspots are visible at 27-day intervals, the approximate period required for the sun to make one complete rotation.

The 27-day sunspot cycle causes variations in the ionization density of the layers on a day-to-day basis. The fluctuations in the F2 layer are greater than for any other layer. For this reason, precise predictions on a day-to-day basis of the critical frequency of the F2 layer are not possible. In calculating frequencies for long-distance communications, allowances for the fluctuations of the F2 layer must be made.

Irregular Variations

Irregular variations in ionospheric conditions also have an important effect on radio wave propagation. Because these variations are irregular and unpredictable, they can drastically affect communications capabilities without any warning.

The more common irregular variations are sporadic E, sudden ionospheric disturbances, and ionospheric storms.

SPORADIC E.—Irregular cloud-like patches of unusually high ionization, called sporadic E, often form at heights near the normal E layer. Exactly what causes this phenomenon is not known, nor can its occurrence be predicted. It is known to vary significantly with latitude, and in the northern latitudes, it appears to be closely related to the aurora borealis or northern lights.

At times the sporadic E is so thin that radio waves penetrate it easily and are returned to earth by the upper layers. At other times, it extends up to several hundred miles and is heavily ionized.

These characteristics may be either harmful or helpful to radio wave propagation. For example, sporadic E may blank out the use of higher, more favorable ionospheric layers or cause additional absorption of the radio wave at some frequencies. Also, it can cause additional multipath problems and delay the arrival times of the rays of rf energy.

On the other hand, the critical frequency of the sporadic E is very high and can be greater than double the critical frequency of the normal ionospheric layers. This condition may permit the long distance transmission of signals at unusually high frequencies. It may also permit short distance communications to locations that would normally be in the skip zone.

The sporadic E can form and disappear in a short time during either the day or night. However, it usually does not occur at the same time at all transmitting or receiving stations.

SUDDEN IONOSPHERIC DISTURBANCES.—The most startling of the ionospheric irregularities is known as a SUDDEN IONOSPHERIC DISTURBANCE (sid). These disturbances may occur without warning and may prevail for any length of time, from a few minutes to several hours. When sid occurs, long distance propagation of hf radio waves is almost totally "blanked out." The immediate effect is that radio operators listening on normal frequencies are inclined to believe their receivers have gone dead.

When sid has occurred, examination of the sun has revealed a bright solar eruption. All stations lying wholly, or in part, on the sunward side of the Earth are affected. The solar eruption produces an unusually intense burst of ultraviolet light, which is not absorbed by the F2, F1, and E layers, but instead causes a sudden abnormal increase in the ionization density of the D layer. As a result, frequencies above 1 or 2 megahertz are unable to penetrate the D layer and are usually completely absorbed by the layer.

IONOSPHERIC STORMS.—Ionospheric storms are disturbances in the Earth's magnetic field. They are associated, in a manner not fully understood, with both solar eruptions and the 27-day intervals, thus corresponding to the rotation of the sun.

Scientists believe that ionospheric storms result from particle radiation from the sun. Particles radiated from a solar eruption have a slower velocity than ultraviolet light waves produced by the eruption. This would account for the 18-hour or so time difference between a sid and an ionospheric storm. An ionospheric storm that is associated with sunspot activity may begin anytime from 2 days before an active sunspot crosses the central meridian of the sun until four days after it passes the central meridian. At times, however, active sunspots have crossed the central region of the sun without any ionospheric storms occurring. Conversely, ionospheric storms have occurred when there were no visible spots on the sun and no preceding sid. As you can see, some correlation between ionospheric storms, sid, and sunspot activity is possible, but there are no hard and fast rules. Ionospheric storms can occur suddenly without warning.

The most prominent effects of ionospheric storms are a turbulent ionosphere and very erratic sky wave propagation. Critical frequencies are lower than normal, particularly for the F2 layer. Ionospheric storms affect the higher F2 layer first, reducing its ion density. Lower layers are not appreciably affected by the storms unless the disturbance is great. The practical effect of ionospheric storms is that the range of frequencies that can be used for communications on a given circuit is much smaller than normal, and communications are possible only at the lower working frequencies.

Q32. What are the two general types of variations in the ionosphere?

Q33. What is the main difference between these two types of variations?

Q34. What are the four main classes of regular variation which affect the extent of ionization in the ionosphere?

Q35. What are the three more common types of irregular variations in the ionosphere?

FREQUENCY SELECTION CONSIDERATIONS

Up to this point, we have covered various factors that control the propagation of radio waves through the ionosphere, such as the structure of the ionosphere, the incidence angle of radio waves, operating frequencies, etc. There is a very good reason for studying radio wave propagation. You must have a thorough knowledge of radio wave propagation to exercise good judgment when you select transmitting and receiving antennas and operating frequencies. Selection of a suitable operating frequency (within the bounds of frequency allocations and availability) is of prime importance in maintaining reliable communications.

For successful communications between <u>any two specified locations</u> at <u>any given time of the day</u>, there is a <u>maximum</u> frequency, a <u>lowest</u> frequency, and an <u>optimum</u> frequency that can be used.

Maximum Usable Frequency

As we discussed earlier, the higher the frequency of a radio wave, the lower the rate of refraction by an ionized layer. Therefore, for a given angle of incidence and time of day, there is a maximum frequency that can be used for communications between two given locations. This frequency is known as the MAXIMUM USABLE FREQUENCY (muf).

Waves at frequencies above the muf are normally refracted so slowly that they return to Earth beyond the desired location, or pass on through the ionosphere and are lost. You should understand, however, that use of an established muf certainly does not guarantee successful communications between a transmitting site and a receiving site. Variations in the ionosphere may occur at any time and consequently raise or lower the predetermined muf. This is particularly true for radio waves being refracted by the highly variable F2 layer.

The muf is highest around noon when ultraviolet light waves from the sun are the most intense. It then drops rather sharply as recombination begins to take place.

Lowest Usable Frequency

As there is a maximum operating frequency that can be used for communications between two points, there is also a minimum operating frequency. This is known as the LOWEST USABLE FREQUENCY (luf).

As the frequency of a radio wave is lowered, the rate of refraction increases. So a wave whose frequency is below the established luf is refracted back to Earth at a shorter distance than desired, as shown in figure 2-23.

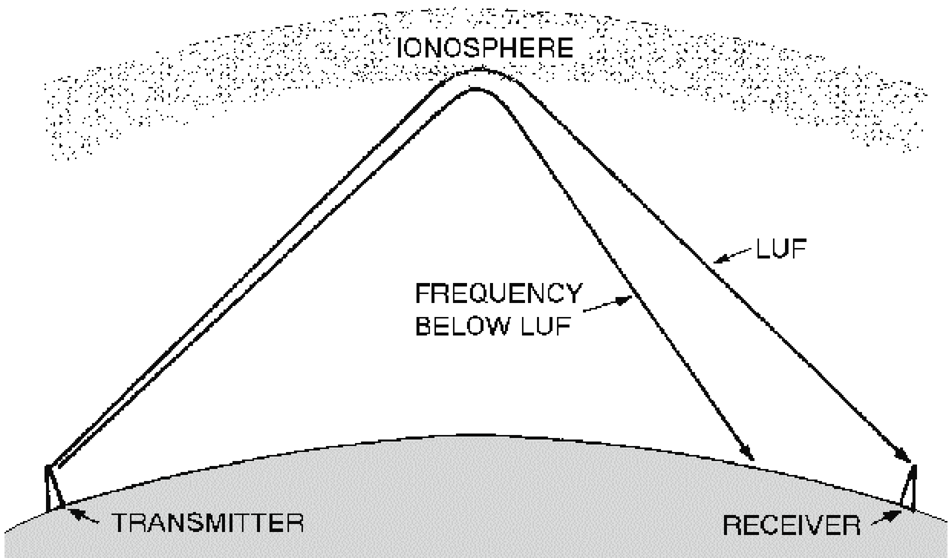


Figure 2-23.—Refraction of frequency below the lowest usable frequency (luf).

The transmission path that results from the rate of refraction is not the only factor that determines the luf. As a frequency is lowered, absorption of the radio wave increases. A wave whose frequency is too low is absorbed to such an extent that it is too weak for reception. Likewise, atmospheric noise is greater at lower frequencies; thus, a low-frequency radio wave may have an unacceptable signal-to-noise ratio.

For a given angle of incidence and set of ionospheric conditions, the luf for successful communications between two locations depends on the refraction properties of the ionosphere, absorption considerations, and the amount of atmospheric noise present.

Optimum Working Frequency

Neither the muf nor the luf is a practical operating frequency. While radio waves at the luf can be refracted back to Earth at the desired location, the signal-to-noise ratio is still much lower than at the higher frequencies, and the probability of multipath propagation is much greater. Operating at or near the muf can result in frequent signal fading and dropouts when ionospheric variations alter the length of the transmission path.

The most practical operating frequency is one that you can rely on with the least amount of problems. It should be high enough to avoid the problems of multipath, absorption, and noise encountered at the lower frequencies; but not so high as to result in the adverse effects of rapid changes in the ionosphere.

A frequency that meets the above criteria has been established and is known as the OPTIMUM WORKING FREQUENCY. It is abbreviated "fot" from the initial letters of the French words for optimum working frequency, "frequence optimum de travail." The fot is roughly about 85 percent of the muf but the actual percentage varies and may be either considerably more or less than 85 percent.

Q36. What do the letters muf, luf, and fot stand for?

Q37. When is muf at its highest and why?

Q38. What happens to the radio wave if the luf is too low?

Q39. What are some disadvantages of operating transmitters at or near the luf?

Q40. What are some disadvantages of operating a transmitter at or near the muf?

Q41. What is fot?

WEATHER VERSUS PROPAGATION

Weather is an additional factor that affects the propagation of radio waves. In this section, we will explain how and to what extent the various weather phenomena affect wave propagation.

Wind, air temperature, and water content of the atmosphere can combine in many ways. Certain combinations can cause radio signals to be heard hundreds of miles beyond the ordinary range of radio communications. Conversely, a different combination of factors can cause such attenuation of the signal that it may not be heard even over a normally satisfactory path. Unfortunately, there are no hard and fast rules on the effects of weather on radio transmissions since the weather is extremely complex and subject to frequent change. We will, therefore, limit our discussion on the effects of weather on radio waves to general terms.

PRECIPITATION ATTENUATION

Calculating the effect of weather on radio wave propagation would be comparatively simple if there were no water or water vapor in the atmosphere. However, some form of water (vapor, liquid, or solid) is always present and must be considered in all calculations. Before we begin discussing the specific effects that individual forms of precipitation (rain, snow, fog) have on radio waves, you should understand that attenuation because of precipitation is generally proportionate to the frequency and wavelength of the radio wave. For example, rain has a pronounced effect on waves at microwave frequencies. However, rain hardly affects waves with long wavelengths (hf range and below). You can assume, then, that as the wavelength becomes shorter with increases in frequency, precipitation has an increasingly important attenuation effect on radio waves. Conversely, you can assume that as the wavelength becomes longer with decreases in frequency, precipitation has little attenuation effect.

Rain

Attenuation because of raindrops is greater than attenuation because of other forms of precipitation. Attenuation may be caused by absorption, in which the raindrop, acting as a poor dielectric, absorbs power from the radio wave and dissipates the power by heat loss or by scattering (fig. 2-24). Raindrops cause greater attenuation by scattering than by absorption at frequencies above 100 megahertz. At frequencies above 6 gigahertz, attenuation by raindrop scatter is even greater.

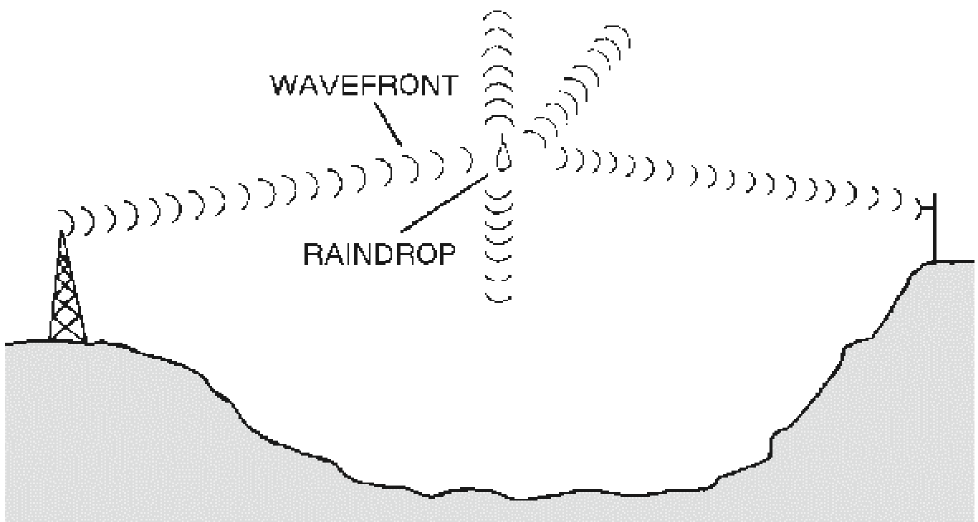


Figure 2-24.—Rf energy losses from scattering.

Fog

In the discussion of attenuation, fog may be considered as another form of rain. Since fog remains suspended in the atmosphere, the attenuation is determined by the quantity of water per unit volume and by the size of the droplets. Attenuation because of fog is of minor importance at frequencies lower than 2 gigahertz. However, fog can cause serious attenuation by absorption, at frequencies above 2 gigahertz.

Snow

The scattering effect because of snow is difficult to compute because of irregular sizes and shapes of the flakes. While information on the attenuating effect of snow is limited, scientists assume that attenuation from snow is less than from rain falling at an equal rate. This assumption is borne out by the fact that the density of rain is eight times the density of snow. As a result, rain falling at 1 inch per hour would have more water per cubic inch than snow falling at the same rate.

Hail

Attenuation by hail is determined by the size of the stones and their density. Attenuation of radio waves by scattering because of hailstones is considerably less than by rain.

TEMPERATURE INVERSION

Under normal atmospheric conditions, the warmest air is found near the surface of the Earth. The air gradually becomes cooler as altitude increases. At times, however, an unusual situation develops in which layers of warm air are formed above layers of cool air. This condition is known as TEMPERATURE INVERSION. These temperature inversions cause channels, or ducts, of cool air to be sandwiched between the surface of the Earth and a layer of warm air, or between two layers of warm air.

If a transmitting antenna extends into such a duct of cool air, or if the radio wave enters the duct at a very low angle of incidence, vhf and uhf transmissions may be propagated far beyond normal line-of-sight distances. When ducts are present as a result of temperature inversions, good reception of vhf and uhf television signals from a station located hundreds of miles away is not unusual. These long distances are possible because of the different densities and refractive qualities of warm and cool air. The sudden change in density when a radio wave enters the warm air above a duct causes the wave to be refracted back toward Earth. When the wave strikes the Earth or a warm layer below the duct, it is again reflected or refracted upward and proceeds on through the duct with a multiple-hop type of action. An example of the propagation of radio waves by ducting is shown in figure 2-25.

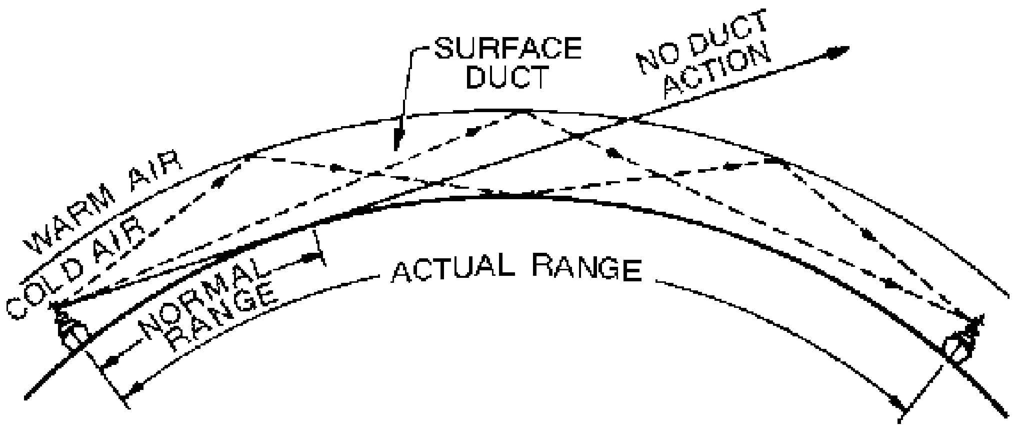


Figure 2-25.—Duct effect caused by temperature inversion.

Q42. How do raindrops affect radio waves?

Q43. How does fog affect radio waves at frequencies above 2 gigahertz?

Q44. How is the term "temperature inversion" used when referring to radio waves?

Q45. How does temperature inversion affect radio transmission?

TROPOSPHERIC PROPAGATION

As the lowest region of the Earth's atmosphere, the troposphere extends from the Earth's surface to a height of slightly over 7 miles. Virtually all weather phenomena occur in this region. Generally, the troposphere is characterized by a steady decrease in both temperature and pressure as height is increased. However, the many changes in weather phenomena cause variations in humidity and an uneven heating of the Earth's surface. As a result, the air in the troposphere is in constant motion. This motion causes small turbulences, or eddies, to be formed, as shown by the bouncing of aircraft entering turbulent areas of the atmosphere. These turbulences are most intense near the Earth's surface and gradually diminish with height. They have a refractive quality that permits the refracting or scattering of radio waves with short wavelengths. This scattering provides enhanced communications at higher frequencies.

Recall that in the relationship between frequency and wavelength, wavelength decreases as frequency increases and vice versa. Radio waves of frequencies below 30 megahertz normally have wavelengths longer than the size of weather turbulences. These radio waves are, therefore, affected very little by the turbulences. On the other hand, as the frequency increases into the vhf range and above, the wavelengths decrease in size, to the point that they become subject to tropospheric scattering. The usable frequency range for tropospheric scattering is from about 100 megahertz to 10 gigahertz.

TROPOSPHERIC SCATTERING

When a radio wave passing through the troposphere meets a turbulence, it makes an abrupt change in velocity. This causes a small amount of the energy to be scattered in a forward direction and returned to Earth at distances beyond the horizon. This phenomenon is repeated as the radio wave meets other turbulences in its path. The total received signal is an accumulation of the energy received from each of the turbulences.

This scattering mode of propagation enables vhf and uhf signals to be transmitted far beyond the normal line-of-sight. To better understand how these signals are transmitted over greater distances, you must first consider the propagation characteristics of the space wave used in vhf and uhf line-of-sight communications. When the space wave is transmitted, it undergoes very little attenuation within the line-of-sight horizon. When it reaches the horizon, the wave is diffracted and follows the Earth's curvature. Beyond the horizon, the rate of attenuation increases very rapidly and signals soon become very weak and unusable.

Tropospheric scattering, on the other hand, provides a usable signal at distances beyond the point where the diffracted space wave drops to an unusable level. This is because of the height at which scattering takes place. The turbulence that causes the scattering can be visualized as a relay station located above the horizon; it receives the transmitted energy and then reradiates it in a forward direction to some point beyond the line-of-sight distance. A high gain receiving antenna aimed toward this scattered energy can then capture it.

The magnitude of the received signal depends on the number of turbulences causing scatter in the desired direction and the gain of the receiving antenna. The scatter area used for tropospheric scatter is known as the *scatter volume*. The angle at which the receiving antenna must be aimed to capture the scattered energy is called the *scatter angle*. The scatter volume and scatter angle are shown in figure 2-26.

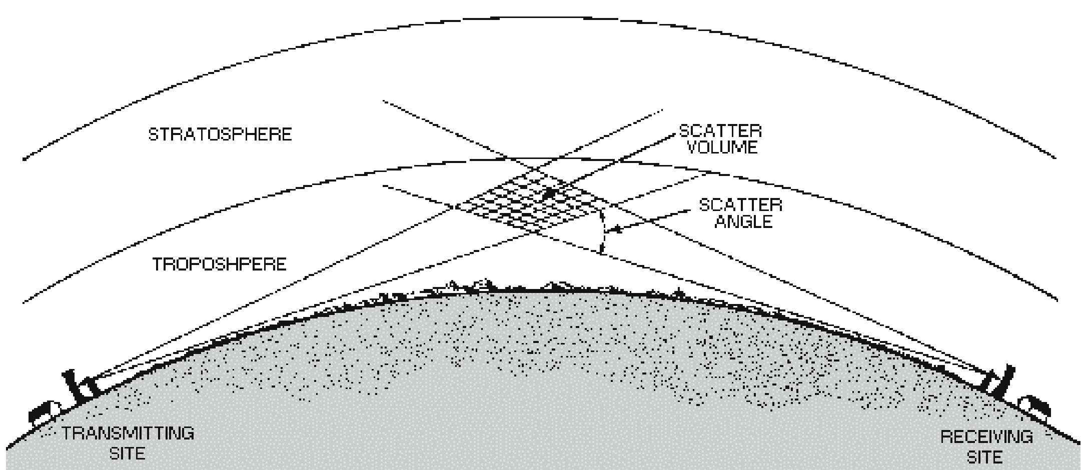


Figure 2-26.—Tropospheric scattering propagation.

The signal take-off angle (transmitting antenna's angle of radiation) determines the height of the scatter volume and the size of the scatter angle. A low signal take-off angle produces a low scatter volume, which in turn permits a receiving antenna that is aimed at a low angle to the scatter volume to capture the scattered energy.

As the signal take-off angle is increased, the height of the scatter volume is increased. When this occurs, the amount of received energy decreases. There are two reasons for this: (1) scatter angle increases as the height of the scatter volume is increased; (2) the amount of turbulence decreases with height. As the distance between the transmitting and receiving antennas is increased, the height of the scatter volume must also be increased. The received signal level, therefore, decreases as circuit distance is increased.

The tropospheric region that contributes most strongly to tropospheric scatter propagation lies near the midpoint between the transmitting and receiving antennas and just above the radio horizon of the antennas.

Since tropospheric scatter depends on turbulence in the atmosphere, changes in atmospheric conditions have an effect on the strength of the received signal. Both daily and seasonal variations in signal strength occur as a result of changes in the atmosphere. These variations are called *long-term fading*.

In addition to long-term fading, the tropospheric scatter signal often is characterized by very rapid fading because of multipath propagation. Since the turbulent condition is constantly changing, the path lengths and individual signal levels are also changing, resulting in a rapidly changing signal. Although the signal level of the received signal is constantly changing, the average signal level is stable; therefore, no complete fade out occurs.

Another characteristic of a tropospheric scatter signal is its relatively low power level. Since very little of the scattered energy is reradiated toward the receiver, the efficiency is very low and the signal level at the final receiver point is low. Initial input power must be high to compensate for the low efficiency in the scatter volume. This is accomplished by using high-power transmitters and high-gain antennas, which concentrate the transmitted power into a beam, thus increasing the intensity of energy of each turbulence in the volume. The receiver must also be very sensitive to detect the low-level signals.

APPLICATION OF TROPOSPHERIC SCATTERING

Tropospheric scatter propagation is used for point-to-point communications. A correctly designed tropospheric scatter circuit will provide highly reliable service for distances ranging from 50 miles to 500 miles. Tropospheric scatter systems may be particularly useful for communications to locations in rugged terrain that are difficult to reach with other methods of propagation. One reason for this is that the tropospheric scatter circuit is not affected by ionospheric and auroral disturbances.

Q46. In what layer of the atmosphere does virtually all weather phenomena occur?

Q47. Which radio frequency bands use the tropospheric scattering principle for propagation of radio waves?

Q48. Where is the tropospheric region that contributes most strongly to tropospheric scatter propagation?

SUMMARY

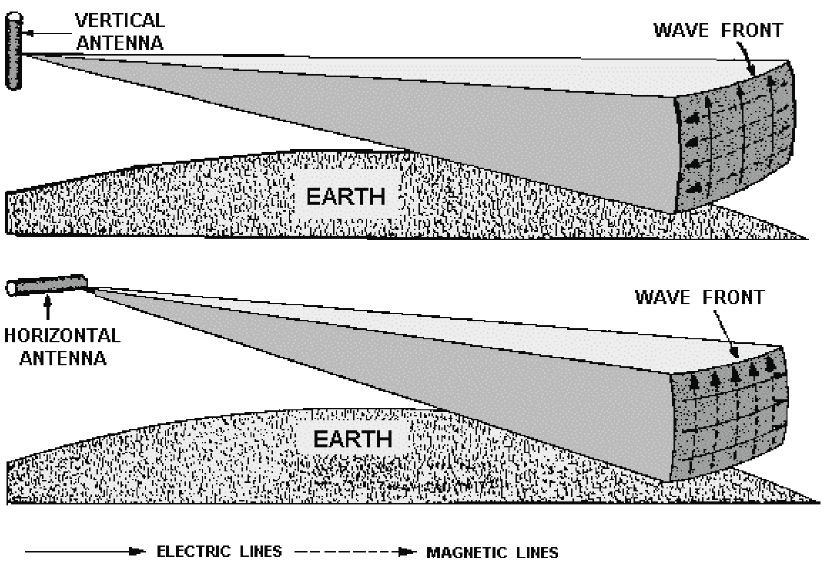
Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas that you have learned. You should have a thorough understanding of these principles before moving on to chapter 3.

The **INDUCTION FIELD** contains an E field and an H field and is localized near the antenna. The E and H fields of the induction field are 90 degrees out of phase with each other.

The **RADIATION FIELD** contains E and H fields that are propagated from the antenna into space in the form of electromagnetic waves. The E and H fields of the radiation field are in phase with each other.

A **HARMONIC FREQUENCY** is any frequency that is a whole number multiple of a smaller basic frequency. For example, a radio wave transmitted at a fundamental frequency of 3000 hertz can have a second harmonic of 6000 hertz, a third harmonic frequency of 9000 hertz, etc., transmitted at the same time.

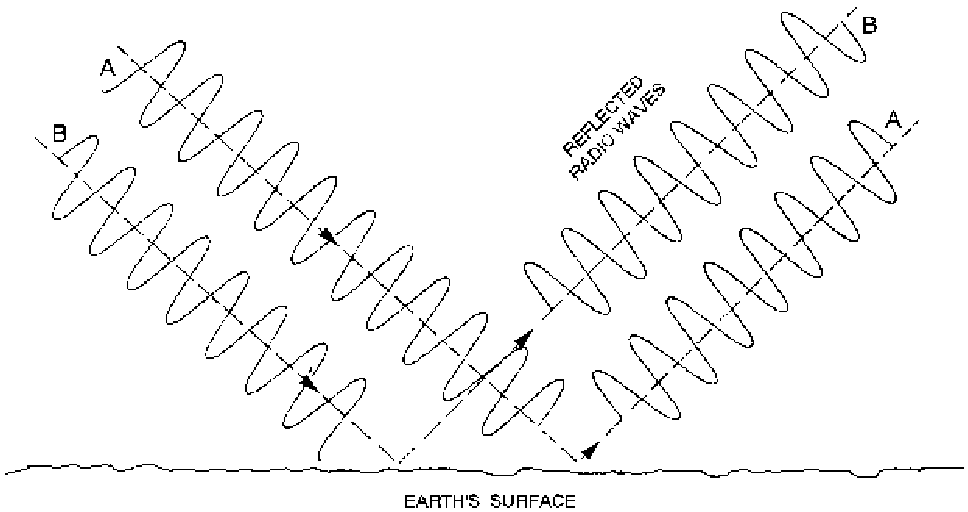
A **VERTICALLY POLARIZED** antenna transmits an electromagnetic wave with the E field perpendicular to the Earth's surface. A **HORIZONTALLY POLARIZED** antenna transmits a radio wave with the E field parallel to the Earth's surface.



A **WAVEFRONT** is a small section of an expanding sphere of radiated energy and is perpendicular to the direction of travel from the antenna.

RADIO WAVES are electromagnetic waves that can be reflected, refracted, and diffracted in the atmosphere like light and heat waves.

REFLECTED RADIO WAVES are waves that have been reflected from a surface and are 180 degrees out of phase with the initial wave.



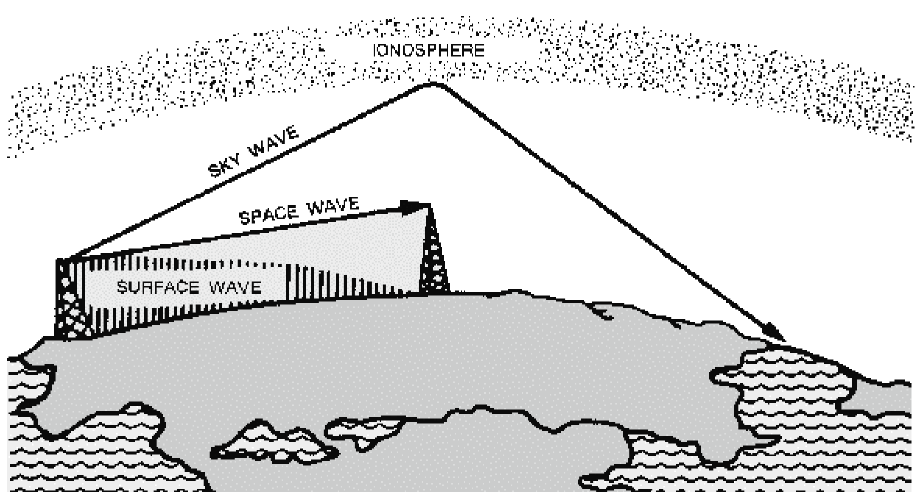
The Earth's atmosphere is divided into three separate layers: The **TROPOSPHERE**, **STRATOSPHERE**, and **IONOSPHERE**.

The **TROPOSPHERE** is the region of the atmosphere where virtually all weather phenomena take place. In this region, rf energy is greatly affected.

The **STRATOSPHERE** has a constant temperature and has little effect on radio waves.

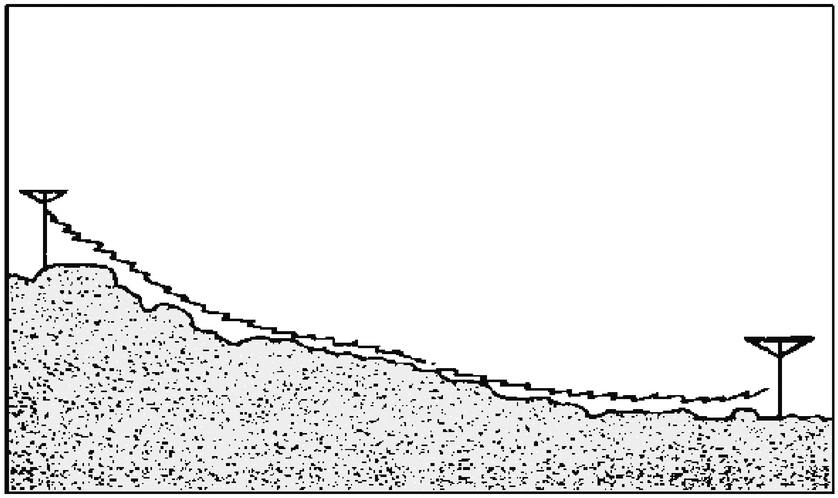
The **IONOSPHERE** contains four cloud-like layers of electrically charged ions which aid in long distance communications.

GROUND WAVES and **SKY WAVES** are the two basic types of radio waves that transmit energy from the transmitting antenna to the receiving antenna.

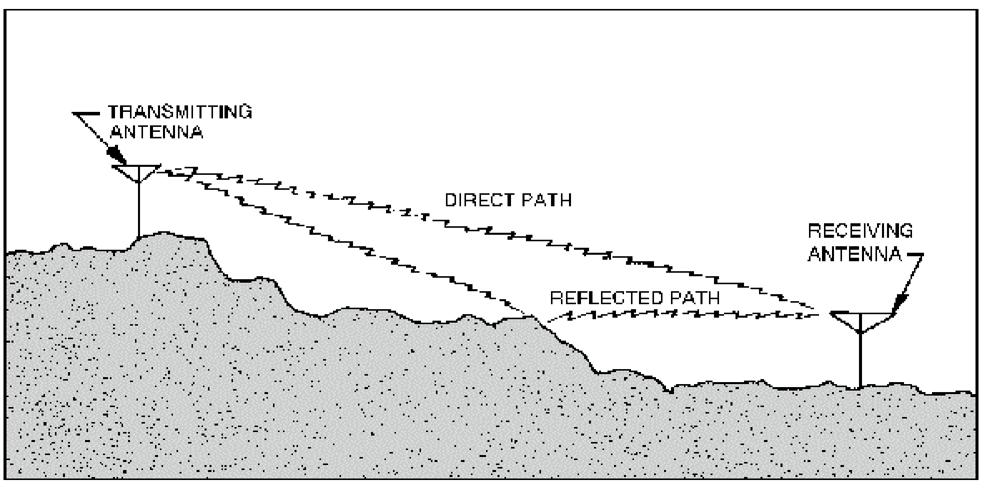


GROUND WAVES are composed of two separate component waves: the **SURFACE WAVE** and the **SPACE WAVE**.

SURFACE WAVES travel along the contour of the Earth by diffraction.



SPACE WAVES can travel through the air directly to the receiving antenna or can be reflected from the surface of the Earth.



SKY WAVES, often called ionospheric waves, are radiated in an upward direction and returned to Earth at some distant location because of refraction.

NATURAL HORIZON is the line-of-sight horizon.

RADIO HORIZON is one-third farther than the natural horizon.

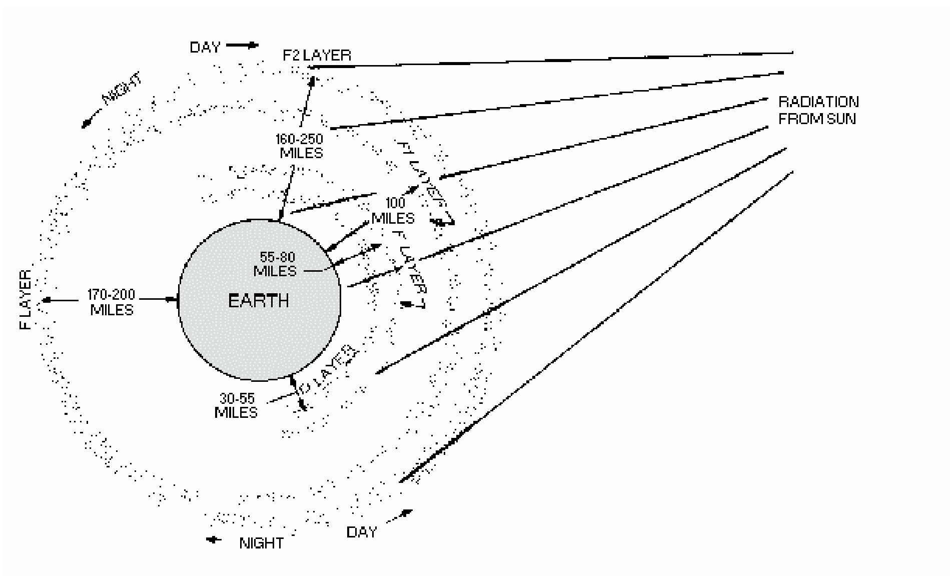
The **IONOSPHERE** consists of several layers of ions, formed by the process called ionization.

IONIZATION is the process of knocking electrons free from their parent atom, thus upsetting electrical neutrality.

RECOMBINATION is the opposite of ionization; that is, the free ions combine with positive ions, causing the positive ions to return to their original neutral atom state.

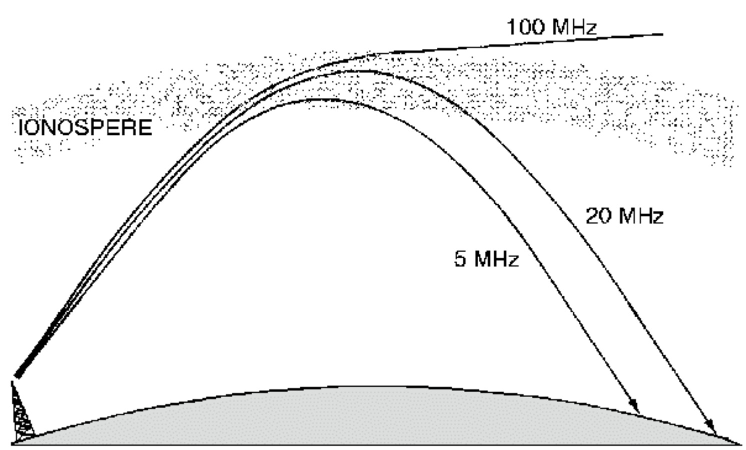
The **D LAYER** is the lowest region of the ionosphere and refracts signals of low frequencies back to Earth.

The **E LAYER** is present during the daylight hours; refracts signals as high as 20 megahertz back to Earth; and is used for communications up to 1500 miles.

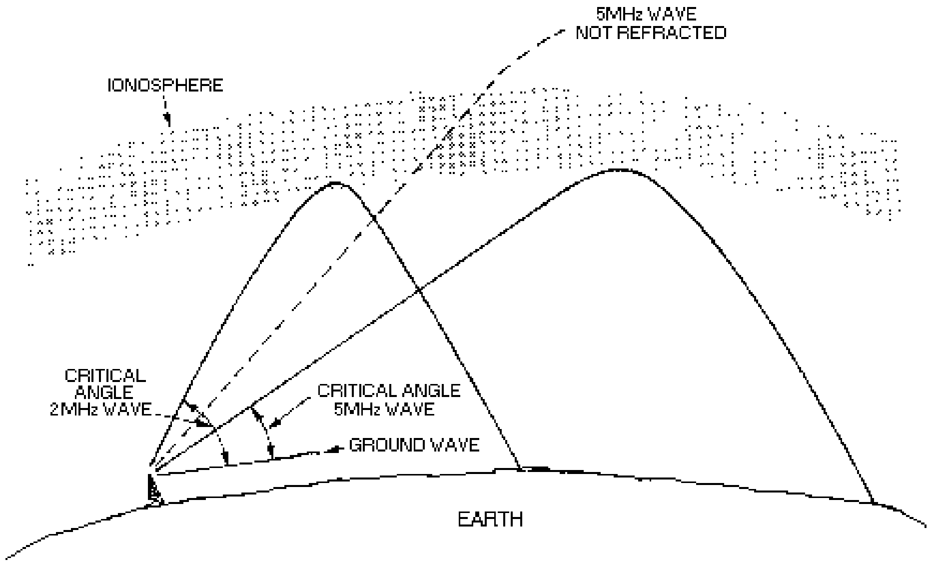


The **F LAYER** is divided into the F1 and F2 layers during the day but combine at night to form one layer. This layer is responsible for high-frequency, long-range transmission.

The **CRITICAL FREQUENCY** is the maximum frequency that a radio wave can be transmitted vertically and still be refracted back to Earth.

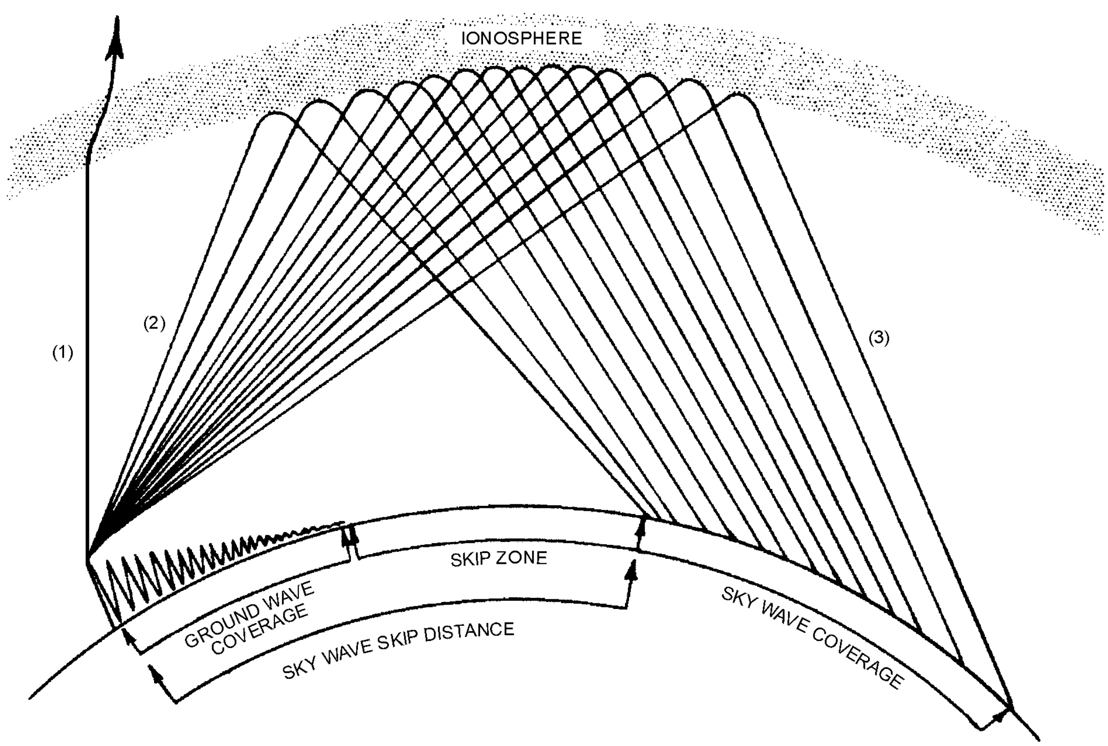


The **CRITICAL ANGLE** is the maximum and/or minimum angle that a radio wave can be transmitted and still be refracted back to Earth.



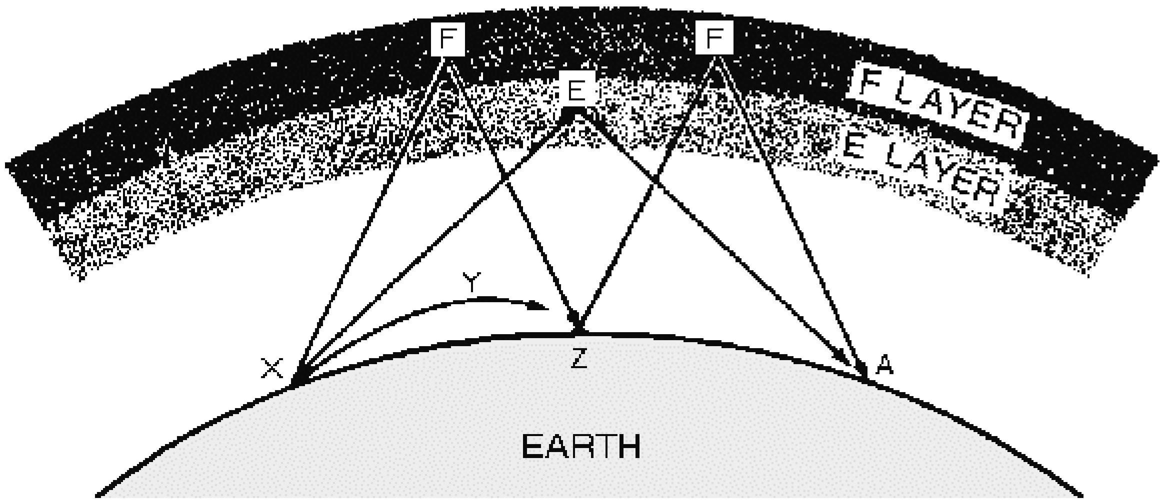
SKIP DISTANCE is the distance between the transmitter and the point where the sky wave first returns to Earth.

SKIP ZONE is the zone of silence between the point where the ground wave becomes too weak for reception and the point where the sky wave is first returned to Earth.



FADING is caused by variations in signal strength, such as absorption of the rf energy by the ionosphere.

MULTIPATH FADING occurs when a transmitted signal divides and takes more than one path to a receiver and some of the signals arrive out of phase, resulting in a weak or fading signal.

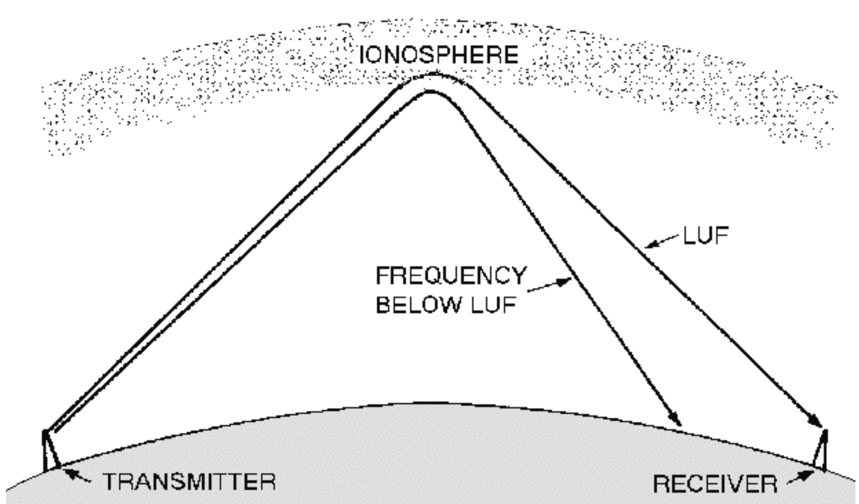


Some **TRANSMISSION LOSSES** that affect radio-wave propagation are ionospheric absorption, ground reflection, and free-space losses.

ELECTROMAGNETIC INTERFERENCE (emi), both natural and man-made, interfere with radio communications.

The **MAXIMUM USABLE FREQUENCY (muf)** is the highest frequency that can be used for communications between two locations at a given angle of incidence and time of day.

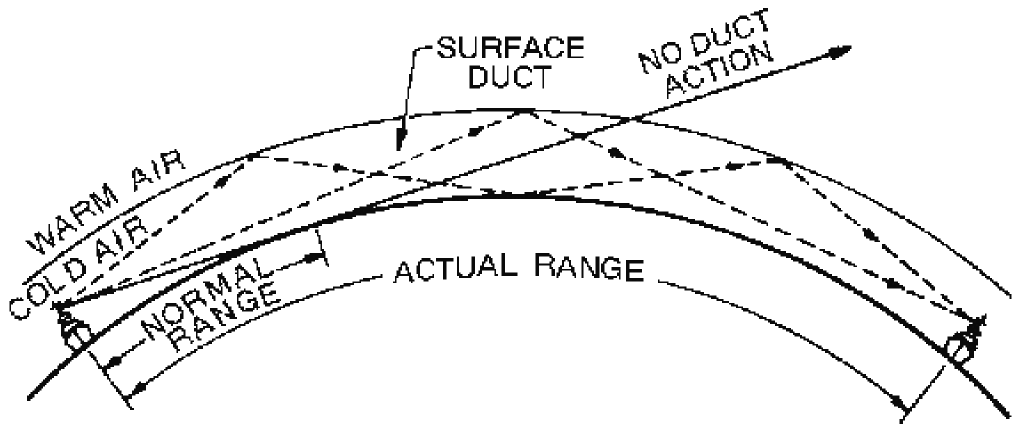
The **LOWEST USABLE FREQUENCY (luf)** is the lowest frequency that can be used for communications between two locations.



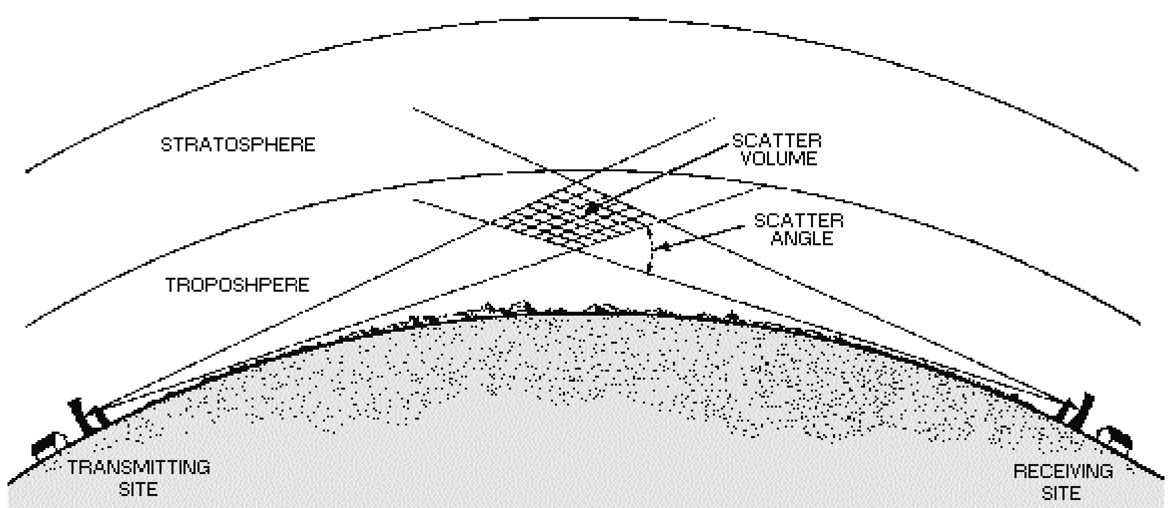
OPTIMUM WORKING FREQUENCY (fot) is the most practical operating frequency and the one that can be relied on to have the fewest problems.

PRECIPITATION ATTENUATION can be caused by rain, fog, snow, and hail; and can affect overall communications considerably.

TEMPERATURE INVERSION causes channels, or ducts, of cool air to form between layers of warm air, which can cause radio waves to travel far beyond the normal line-of-sight distances.



TROPOSPHERIC PROPAGATION uses the scattering principle to achieve beyond the line-of-sight radio communications within the troposphere.



ANSWERS TO QUESTIONS Q1. THROUGH Q48.

A1. *Induction field and radiation field.*

A2. *Induction field.*

A3. *Radiation field.*

A4. *Fundamental frequency.*

A5. *Harmonic frequency or harmonics.*

A6. *30 meters.*

A7. *5 megahertz.*

A8. *Vertically polarized.*

A9. *Direction of wave propagation.*

A10. *Shifting in the phase relationships of the wave.*

A11. *Troposphere, stratosphere, and ionosphere.*

A12. *Stratosphere.*

A13. *Whether the component of the wave is travelling along the surface or over the surface of the earth.*

A14. *Radio horizon is about 1/3 farther.*

A15. *Sea water.*

A16. *(a) electrical properties of the terrain (b) frequency (c) polarization of the antenna*

A17. *High energy ultraviolet light waves from the sun.*

A18. *D, E, F_1, and F_2 layers.*

A19. *D layer is 30-55 miles, E layer 55-90 miles, and F layers are 90-240 miles.*

A20. *Thickness of ionized layer.*

A21. *Critical frequency.*

A22. *(a) density of ionization of the layer (b) frequency (c) angle at which it enters the layer*

A23. *A zone of silence between the ground wave and sky wave where there is no reception.*

A24. *Where ionization density is greatest.*

A25. *A term used to describe the multiple pattern a radio wave may follow.*

A26. *Selective fading.*

A27. *Natural and man-made interference.*

A28. Natural.

A29. Man-made.

A30. (a) filtering and shielding of the transmitter (b) limiting bandwidth (c) cutting the antenna to the correct frequency

A31. (a) physical separation of the antenna (b) limiting bandwidth of the antenna (c) use of directional antennas

A32. Regular and irregular variations.

A33. Regular variations can be predicted but irregular variations are unpredictable.

A34. Daily, seasonal, 11-year, and 27-days variation.

A35. Sporadic E, sudden disturbances, and ionospheric storms.

A36. Muf is maximum usable frequency. Luf is lowest usable frequency. Fot is commonly known as optimum working frequency.

A37. Muf is highest around noon. Ultraviolet light waves from the sun are most intense.

A38. When luf is too low it is absorbed and is too weak for reception.

A39. Signal-to-noise ratio is low and the probability of multipath propagation is greater.

A40. Frequent signal fading and dropouts.

A41. Fot is the most practical operating frequency that can be relied on to avoid problems of multipath, absorbtion, and noise.

A42. They can cause attenuation by scattering.

A43. It can cause attenuation by absorbtion.

A44. It is a condition where layers of warm air are formed above layers of cool air.

A45. It can cause vhf and uhf transmission to be propagated far beyond normal line-of-sight distances.

A46. Troposphere.

A47. Vhf and above.

A48. Near the mid-point between the transmitting and receiving antennas, just above the radio horizon.

CHAPTER 3

PRINCIPLES OF TRANSMISSION LINES

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to:

1. State what a transmission line is and how transmission lines are used.
2. Explain the operating principles of transmission lines.
3. Describe the five types of transmission lines.
4. State the length of a transmission line.
5. Explain the theory of the transmission line.
6. Define the term LUMPED CONSTANTS in relation to a transmission line.
7. Define the term DISTRIBUTED CONSTANTS in relation to a transmission line.
8. Define LEAKAGE CURRENT.
9. Describe how the electromagnetic lines of force around a transmission line are affected by the distributed constants.
10. Define the term CHARACTERISTIC IMPEDANCE and explain how it affects the transfer of energy along a transmission line.
11. State how the energy transfer along a transmission line is affected by characteristic impedance and the infinite line.
12. Identify the cause of and describe the characteristics of reflections on a transmission line.
13. Define the term STANDING WAVES as applied to a transmission line.
14. Describe how standing waves are produced on a transmission line and identify the types of terminations.
15. Describe the types of standing-wave ratios.

INTRODUCTION TO TRANSMISSION LINES

A TRANSMISSION LINE is a device designed to guide electrical energy from one point to another. It is used, for example, to transfer the output rf energy of a transmitter to an antenna. This energy will not travel through normal electrical wire without great losses. Although the antenna can be connected directly to the transmitter, the antenna is usually located some distance away from the transmitter. On board ship, the transmitter is located inside a radio room and its associated antenna is mounted on a mast. A transmission line is used to connect the transmitter and the antenna.

The transmission line has a single purpose for both the transmitter and the antenna. This purpose is to transfer the energy output of the transmitter to the antenna with the least possible power loss. How well this is done depends on the special physical and electrical characteristics (impedance and resistance) of the transmission line.

TERMINOLOGY

All transmission lines have two ends (see figure 3-1). The end of a two-wire transmission line connected to a source is ordinarily called the INPUT END or the GENERATOR END. Other names given to this end are TRANSMITTER END, SENDING END, and SOURCE. The other end of the line is called the OUTPUT END or RECEIVING END. Other names given to the output end are LOAD END and SINK.

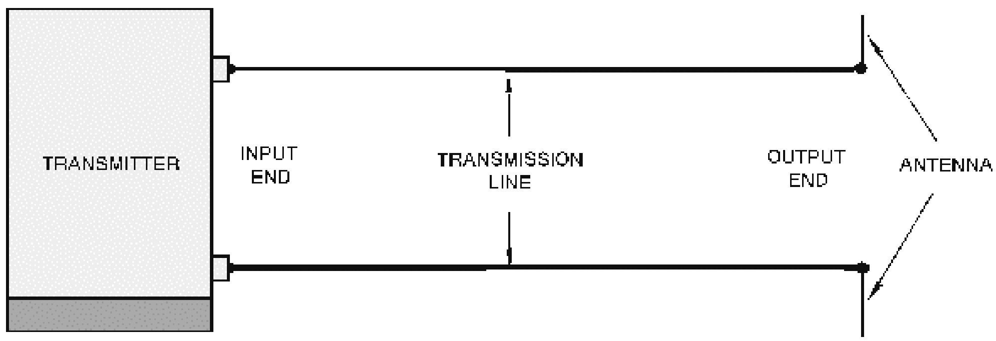


Figure 3-1.—Basic transmission line.

You can describe a transmission line in terms of its impedance. The ratio of voltage to current (E_{in}/I_{in}) at the input end is known as the INPUT IMPEDANCE (Z_{in}). This is the impedance presented to the transmitter by the transmission line and its load, the antenna. The ratio of voltage to current at the output (E_{out}/I_{out}) end is known as the OUTPUT IMPEDANCE (Z_{out}). This is the impedance presented to the load by the transmission line and its source. If an infinitely long transmission line could be used, the ratio of voltage to current at any point on that transmission line would be some particular value of impedance. This impedance is known as the CHARACTERISTIC IMPEDANCE.

Q1. What connecting link is used to transfer energy from a radio transmitter to its antenna located on the mast of a ship?

Q2. What term is used for the end of the transmission line that is connected to a transmitter?

Q3. What term is used for the end of the transmission line that is connected to an antenna?

TYPES OF TRANSMISSION MEDIUMS

The Navy uses many different types of TRANSMISSION MEDIUMS in its electronic applications. Each medium (line or wave guide) has a certain characteristic impedance value, current-carrying capacity, and physical shape and is designed to meet a particular requirement.

The five types of transmission mediums that we will discuss in this chapter include PARALLEL-LINE, TWISTED PAIR, SHIELDED PAIR, COAXIAL LINE, and WAVEGUIDES. The use of a particular line depends, among other things, on the applied frequency, the power-handling capabilities, and the type of installation.

NOTE: In the following paragraphs, we will mention LOSSES several times. We will discuss these losses more thoroughly under "LOSSES IN TRANSMISSION LINES."

Two-Wire Open Line

One type of parallel line is the TWO-WIRE OPEN LINE illustrated in figure 3-2. This line consists of two wires that are generally spaced from 2 to 6 inches apart by insulating spacers. This type of line is most often used for power lines, rural telephone lines, and telegraph lines. It is sometimes used as a transmission line between a transmitter and an antenna or between an antenna and a receiver. An advantage of this type of line is its simple construction. The principal disadvantages of this type of line are the high radiation losses and electrical noise pickup because of the lack of shielding. Radiation losses are produced by the changing fields created by the changing current in each conductor.

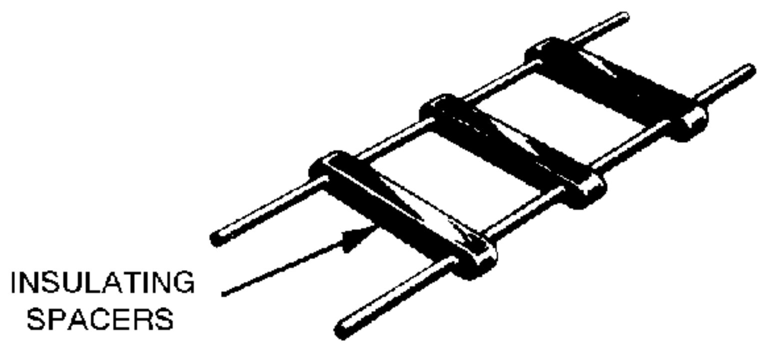


Figure 3-2.—Parallel two-wire line.

Another type of parallel line is the TWO-WIRE RIBBON (TWIN LEAD) illustrated in figure 3-3. This type of transmission line is commonly used to connect a television receiving antenna to a home television set. This line is essentially the same as the two-wire open line except that uniform spacing is assured by embedding the two wires in a low-loss dielectric, usually polyethylene. Since the wires are embedded in the thin ribbon of polyethylene, the dielectric space is partly air and partly polyethylene.

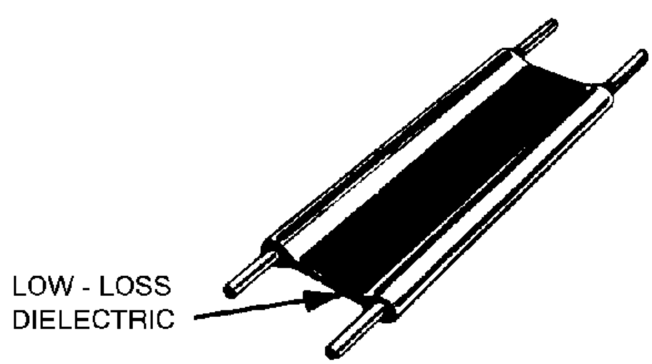


Figure 3-3.—Two-wire ribbon type line.

Twisted Pair

The TWISTED PAIR transmission line is illustrated in figure 3-4. As the name implies, the line consists of two insulated wires twisted together to form a flexible line without the use of spacers. It is not used for transmitting high frequency because of the high dielectric losses that occur in the rubber insulation. When the line is wet, the losses increase greatly.

Figure 3-4.—Twisted pair.

Shielded Pair

The SHIELDED PAIR, shown in figure 3-5, consists of parallel conductors separated from each other and surrounded by a solid dielectric. The conductors are contained within a braided copper tubing that acts as an electrical shield. The assembly is covered with a rubber or flexible composition coating that protects the line from moisture and mechanical damage. Outwardly, it looks much like the power cord of a washing machine or refrigerator.

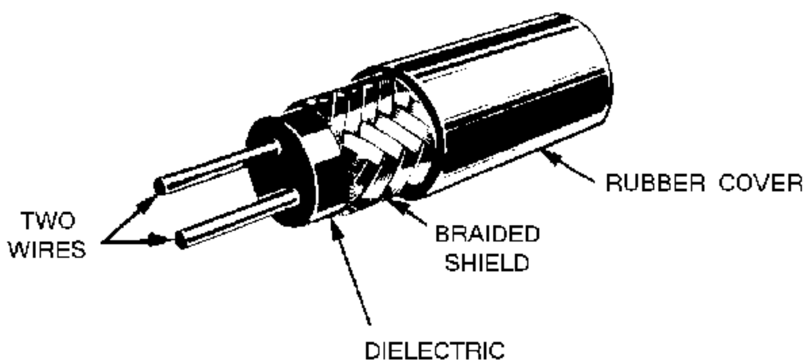


Figure 3-5.—Shielded pair.

The principal advantage of the shielded pair is that the conductors are balanced to ground; that is, the capacitance between the wires is uniform throughout the length of the line. This balance is due to the uniform spacing of the grounded shield that surrounds the wires along their entire length. The braided copper shield isolates the conductors from stray magnetic fields.

Coaxial Lines

There are two types of COAXIAL LINES, RIGID (AIR) COAXIAL LINE and FLEXIBLE (SOLID) COAXIAL LINE. The physical construction of both types is basically the same; that is, each contains two concentric conductors.

The rigid coaxial line consists of a central, insulated wire (inner conductor) mounted inside a tubular outer conductor. This line is shown in figure 3-6. In some applications, the inner conductor is also tubular. The inner conductor is insulated from the outer conductor by insulating spacers or beads at regular intervals. The spacers are made of Pyrex, polystyrene, or some other material that has good insulating characteristics and low dielectric losses at high frequencies.

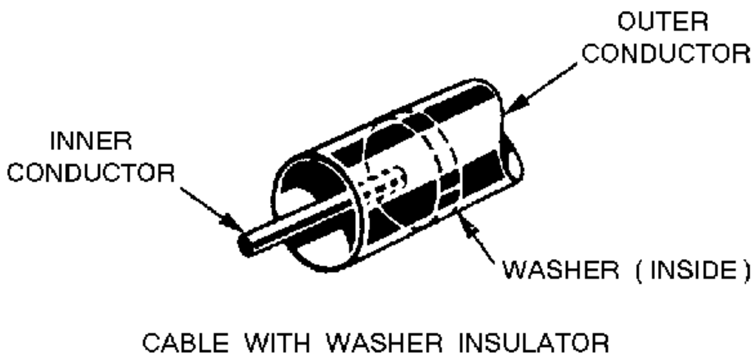


Figure 3-6.—Air coaxial line.

The chief advantage of the rigid line is its ability to minimize radiation losses. The electric and magnetic fields in a two-wire parallel line extend into space for relatively great distances and radiation losses occur. However, in a coaxial line no electric or magnetic fields extend outside of the outer conductor. The fields are confined to the space between the two conductors, resulting in a perfectly shielded coaxial line. Another advantage is that interference from other lines is reduced.

The rigid line has the following disadvantages: (1) it is expensive to construct; (2) it must be kept dry to prevent excessive leakage between the two conductors; and (3) although high-frequency losses are somewhat less than in previously mentioned lines, they are still excessive enough to limit the practical length of the line.

Leakage caused by the condensation of moisture is prevented in some rigid line applications by the use of an inert gas, such as nitrogen, helium, or argon. It is pumped into the dielectric space of the line at a pressure that can vary from 3 to 35 pounds per square inch. The inert gas is used to dry the line when it is first installed and pressure is maintained to ensure that no moisture enters the line.

Flexible coaxial lines (figure 3-7) are made with an inner conductor that consists of flexible wire insulated from the outer conductor by a solid, continuous insulating material. The outer conductor is made of metal braid, which gives the line flexibility. Early attempts at gaining flexibility involved using rubber insulators between the two conductors. However, the rubber insulators caused excessive losses at high frequencies.

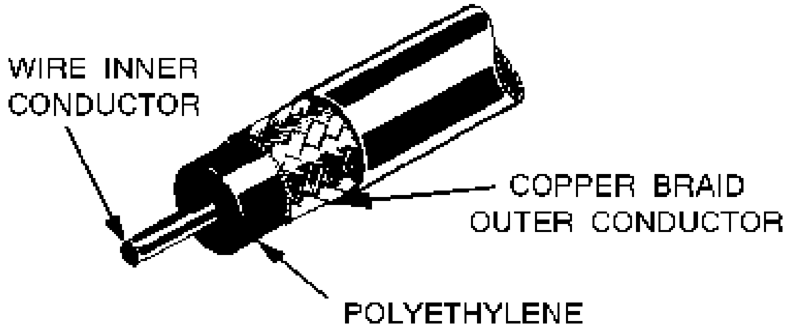


Figure 3-7.—Flexible coaxial line.

Because of the high-frequency losses associated with rubber insulators, polyethylene plastic was developed to replace rubber and eliminate these losses. Polyethylene plastic is a solid substance that remains flexible over a wide range of temperatures. It is unaffected by seawater, gasoline, oil, and most other liquids that may be found aboard ship. The use of polyethylene as an insulator results in greater high-frequency losses than the use of air as an insulator. However, these losses are still lower than the losses associated with most other solid dielectric materials.

Waveguides

The WAVEGUIDE is classified as a transmission line. However, the method by which it transmits energy down its length differs from the conventional methods. Waveguides are cylindrical, elliptical, or rectangular (cylindrical and rectangular shapes are shown in figure 3-8). The rectangular waveguide is used more frequently than the cylindrical waveguide.

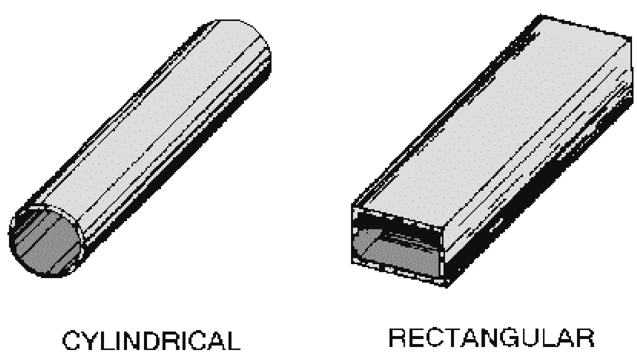


Figure 3-8.—Waveguides.

The term *waveguide* can be applied to all types of transmission lines in the sense that they are all used to guide energy from one point to another. However, usage has generally limited the term to mean a hollow metal tube or a dielectric transmission line. In this chapter, we use the term *waveguide* only to mean "hollow metal tube." It is interesting to note that the transmission of electromagnetic energy along a waveguide travels at a velocity somewhat slower than electromagnetic energy traveling through free space.

A waveguide may be classified according to its cross section (rectangular, elliptical, or circular), or according to the material used in its construction (metallic or dielectric). Dielectric waveguides are seldom used because the dielectric losses for all known dielectric materials are too great to transfer the electric and magnetic fields efficiently.

The installation of a complete waveguide transmission system is somewhat more difficult than the installation of other types of transmission lines. The radius of bends in the waveguide must measure greater than two wavelengths at the operating frequency of the equipment to avoid excessive attenuation. The cross section must remain uniform around the bend. These requirements hamper installation in confined spaces. If the waveguide is dented, or if solder is permitted to run inside the joints, the attenuation of the line is greatly increased. Dents and obstructions in the waveguide also reduce its breakdown voltage, thus limiting the waveguide's power-handling capability because of possible arc over. Great care must be exercised during installation; one or two carelessly made joints can seriously inhibit the advantage of using the waveguide.

We will not consider the waveguide operation in this module, since waveguide theory is discussed in *NEETS*, Module 11, *Microwave Principles*.

Q4. List the five types of transmission lines in use today.

Q5. Name two of the three described uses of a two-wire open line.

Q6. What are the two primary disadvantages of a two-wire open line?

Q7. What type of transmission line is often used to connect a television set to its antenna?

Q8. What is the primary advantage of the shielded pair?

Q9. What are the two types of coaxial lines in use today?

Q10. What is the chief advantage of the air coaxial line?

Q11. List the three disadvantages of the air coaxial line.

Q12. List the two common types of waveguides in use today.

LOSSES IN TRANSMISSION LINES

The discussion of transmission lines so far has not directly addressed LINE LOSSES; actually some line losses occur in all lines. Line losses may be any of three types—COPPER, DIELECTRIC, and RADIATION or INDUCTION LOSSES.

NOTE: Transmission lines are sometimes referred to as rf lines. In this text the terms are used interchangeably.

Copper Losses

One type of copper loss is I^2R LOSS. In rf lines the resistance of the conductors is never equal to zero. Whenever current flows through one of these conductors, some energy is dissipated in the form of heat. This heat loss is a POWER LOSS. With copper braid, which has a resistance higher than solid tubing, this power loss is higher.

Another type of copper loss is due to SKIN EFFECT. When dc flows through a conductor, the movement of electrons through the conductor's cross section is uniform. The situation is somewhat different when ac is applied. The expanding and collapsing fields about each electron encircle other electrons. This phenomenon, called SELF INDUCTION, retards the movement of the encircled electrons.

The flux density at the center is so great that electron movement at this point is reduced. As frequency is increased, the opposition to the flow of current in the center of the wire increases. Current in the center of the wire becomes smaller and most of the electron flow is on the wire surface. When the frequency applied is 100 megahertz or higher, the electron movement in the center is so small that the center of the wire could be removed without any noticeable effect on current. You should be able to see that the effective cross-sectional area decreases as the frequency increases. Since resistance is inversely proportional to the cross-sectional area, the resistance will increase as the frequency is increased. Also, since power loss increases as resistance increases, power losses increase with an increase in frequency because of skin effect.

Copper losses can be minimized and conductivity increased in an rf line by plating the line with silver. Since silver is a better conductor than copper, most of the current will flow through the silver layer. The tubing then serves primarily as a mechanical support.

Dielectric Losses

DIELECTRIC LOSSES result from the heating effect on the dielectric material between the conductors. Power from the source is used in heating the dielectric. The heat produced is dissipated into the surrounding medium. When there is no potential difference between two conductors, the atoms in the dielectric material between them are normal and the orbits of the electrons are circular. When there is a potential difference between two conductors, the orbits of the electrons change. The excessive negative charge on one conductor repels electrons on the dielectric toward the positive conductor and thus distorts the orbits of the electrons. A change in the path of electrons requires more energy, introducing a power loss.

The atomic structure of rubber is more difficult to distort than the structure of some other dielectric materials. The atoms of materials, such as polyethylene, distort easily. Therefore, polyethylene is often used as a dielectric because less power is consumed when its electron orbits are distorted.

Radiation and Induction Losses

RADIATION and INDUCTION LOSSES are similar in that both are caused by the fields surrounding the conductors. Induction losses occur when the electromagnetic field about a conductor cuts through any nearby metallic object and a current is induced in that object. As a result, power is dissipated in the object and is lost.

Radiation losses occur because some magnetic lines of force about a conductor do not return to the conductor when the cycle alternates. These lines of force are projected into space as radiation and this results in power losses. That is, power is supplied by the source, but is not available to the load.

Q13. What are the three types of line losses associated with transmission lines?

Q14. Losses caused by skin effect and the I^2R *(power) loss are classified as what type of loss?*

Q15. What types of losses cause the dielectric material between the conductors to be heated?

LENGTH OF A TRANSMISSION LINE

A transmission line is considered to be electrically short when its physical length is short compared to a quarter-wavelength ($1/4\lambda$) of the energy it is to carry.

NOTE: In this module, for ease of reading, the value of the wavelength will be spelled out in some cases, and in other cases, the numerical value will be used.

A transmission line is electrically long when its physical length is long compared to a quarter-wavelength of the energy it is to carry. You must understand that the terms "short" and "long" are relative ones. For example, a line that has a physical length of 3 meters (approximately 10 feet) is considered quite short electrically if it transmits a radio frequency of 30 kilohertz. On the other hand, the same transmission line is considered electrically long if it transmits a frequency of 30,000 megahertz.

To show the difference in physical and electrical lengths of the lines mentioned above, compute the wavelength of the two frequencies, taking the 30-kilohertz example first:

Given:

$$\lambda = \frac{v}{f}$$

Where:

λ = Wavelength
v = Velocity of rf in free space
f = Frequency of transmission
Hz = Cycles per second

$$\lambda = \frac{300 \times 10^6 \text{ meters / second}}{30 \times 10^3 \text{ cycles / second (Hz)}}$$

$$\lambda = 10 \times 10^3 \text{ meters/cycle}$$

λ = 10,000 meters, or approxiamtely 6 miles for complete wavelength

Now, computing the wavelength for the line carrying 30,000 megahertz:

$$\lambda = \frac{v}{f}$$

$$\lambda = \frac{300 \times 10^6 \text{ meters / second}}{30{,}000 \times 10^6 \text{ cycles / second (Hz)}}$$

$$\lambda = \frac{1}{100} \text{ meter / cycle}$$

λ =.01 meter, or approximately .03 foot for a complete wavelength

Thus, you can see that a 3-meter line is electrically very short for a frequency of 30 kilohertz. Also, the 3-meter line is electrically very long for a frequency of 30,000 megahertz.

When power is applied to a very short transmission line, practically all of it reaches the load at the output end of the line. This very short transmission line is usually considered to have practically no electrical properties of its own, except for a small amount of resistance.

However, the picture changes considerably when a long line is used. Since most transmission lines are electrically long (because of the distance from transmitter to antenna), the properties of such lines must be considered. Frequently, the voltage necessary to drive a current through a long line is considerably greater than the amount that can be accounted for by the impedance of the load in series with the resistance of the line.

TRANSMISSION LINE THEORY

The electrical characteristics of a two-wire transmission line depend primarily on the construction of the line. The two-wire line acts like a long capacitor. The change of its capacitive reactance is noticeable as the frequency applied to it is changed. Since the long conductors have a magnetic field about them when electrical energy is being passed through them, they also exhibit the properties of inductance. The values of inductance and capacitance presented depend on the various physical factors that we discussed earlier. For example, the type of line used, the dielectric in the line, and the length of the line must be considered. The effects of the inductive and capacitive reactances of the line depend on the frequency applied. Since no dielectric is perfect, electrons manage to move from one conductor to the other through the dielectric. Each type of two-wire transmission line also has a conductance value. This conductance value represents the value of the current flow that may be expected through the insulation. If the line is uniform (all values equal at each unit length), then one small section of the line may represent several feet. This illustration of a two-wire transmission line will be used throughout the discussion of transmission lines; but, keep in mind that the principles presented apply to all transmission lines. We will explain the theories using LUMPED CONSTANTS and DISTRIBUTED CONSTANTS to further simplify these principles.

LUMPED CONSTANTS

A transmission line has the properties of inductance, capacitance, and resistance just as the more conventional circuits have. Usually, however, the constants in conventional circuits are lumped into a single device or component. For example, a coil of wire has the property of inductance. When a certain amount of inductance is needed in a circuit, a coil of the proper dimensions is inserted. The inductance of the circuit is lumped into the one component. Two metal plates separated by a small space, can be used to supply the required capacitance for a circuit. In such a case, most of the capacitance of the circuit is lumped into this one component. Similarly, a fixed resistor can be used to supply a certain value of circuit resistance as a lumped sum. Ideally, a transmission line would also have its constants of inductance, capacitance, and resistance lumped together, as shown in figure 3-9. Unfortunately, this is not the case. Transmission line constants are *distributed,* as described below.

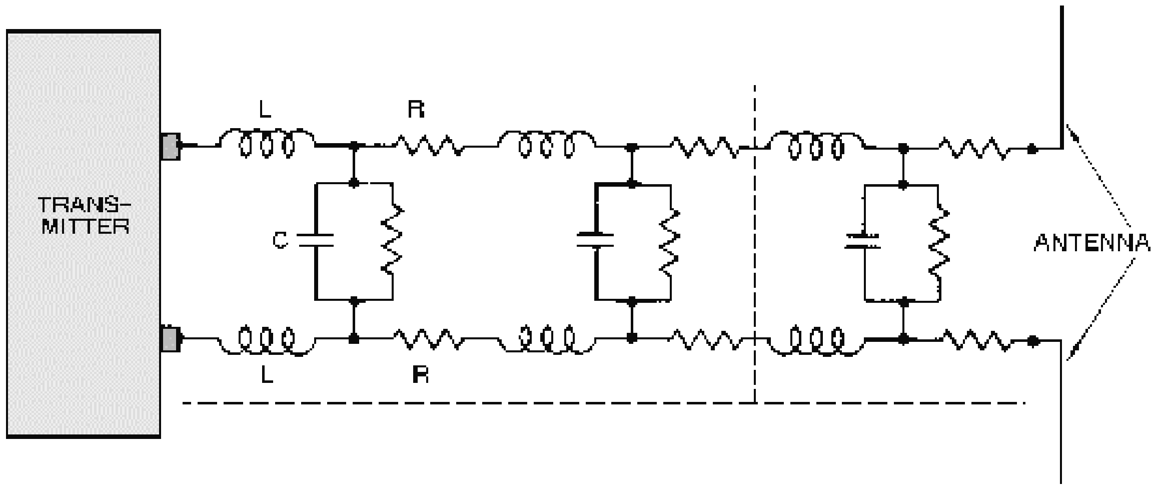


Figure 3-9.—Equivalent circuit of a two-wire transmission line.

DISTRIBUTED CONSTANTS

Transmission line constants, called *distributed constants*, are spread along the entire length of the transmission line and cannot be distinguished separately. The amount of inductance, capacitance, and resistance depends on the length of the line, the size of the conducting wires, the spacing between the wires, and the dielectric (air or insulating medium) between the wires. The following paragraphs will be useful to you as you study distributed constants on a transmission line.

Inductance of a Transmission Line

When current flows through a wire, magnetic lines of force are set up around the wire. As the current increases and decreases in amplitude, the field around the wire expands and collapses accordingly. The energy produced by the magnetic lines of force collapsing back into the wire tends to keep the current flowing in the same direction. This represents a certain amount of inductance, which is expressed in *microhenrys per unit length*. Figure 3-10 illustrates the inductance and magnetic fields of a transmission line.

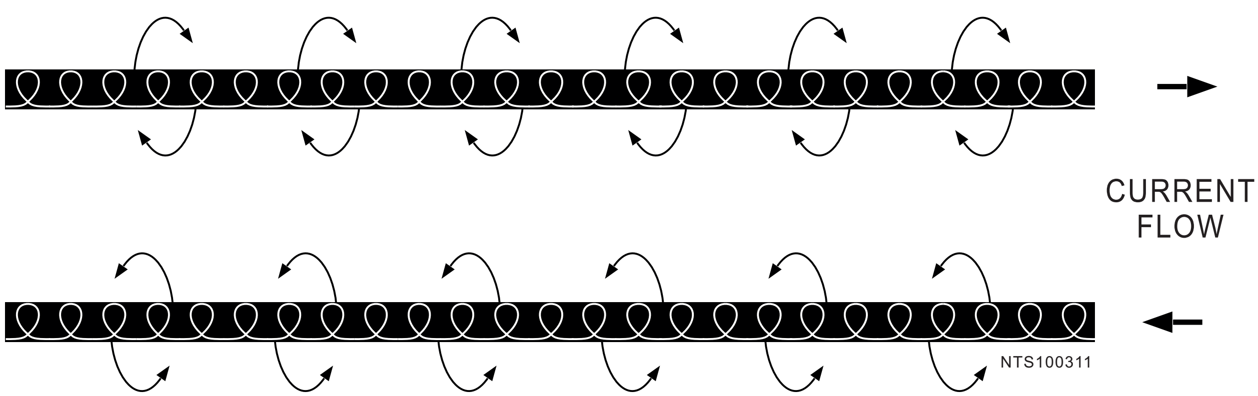


Figure 3-10.—Distributed inductance

Capacitance of a Transmission Line

Capacitance also exists between the transmission line wires, as illustrated in figure 3-11. Notice that the two parallel wires act as plates of a capacitor and that the air between them acts as a dielectric. The capacitance between the wires is usually expressed in *picofarads per unit length*. This electric field between the wires is similar to the field that exists between the two plates of a capacitor.

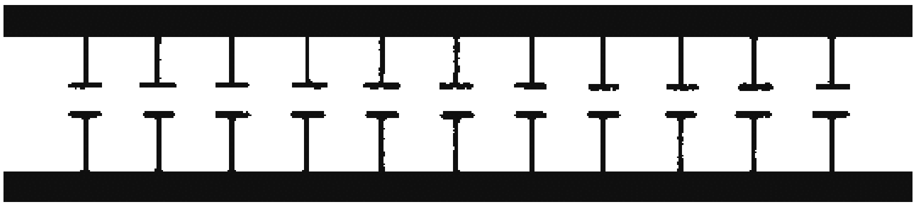


Figure 3-11.—Distributed capacitance.

Resistance of a Transmission Line

The transmission line shown in figure 3-12 has electrical resistance along its length. This resistance is usually expressed in *ohms per unit length* and is shown as existing continuously from one end of the line to the other.

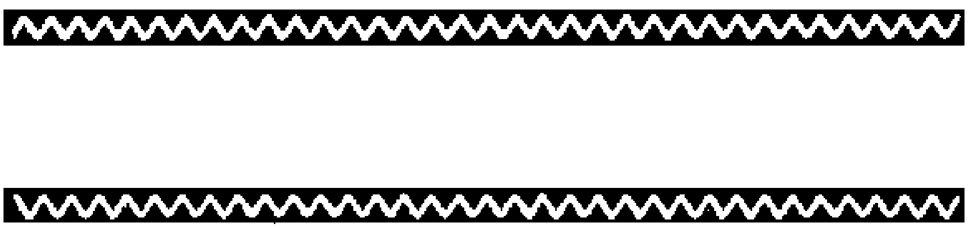


Figure 3-12.—Distributed resistance.

Q16. *What must the physical length of a transmission line be if it will be operated at 15,000,000 Hz? Use the formula:*

$$\lambda = \frac{v}{f}$$

Q17. *What are two of the three physical factors that determine the values of capacitance and inductance of a transmission line?*

Q18. *A transmission line is said to have distributed constants of inductance, capacitance, and resistance along the line. What units of measurement are used to express these constants?*

Leakage Current

Since any dielectric, even air, is not a perfect insulator, a small current known as LEAKAGE CURRENT flows between the two wires. In effect, the insulator acts as a resistor, permitting current to pass between the two wires. Figure 3-13 shows this leakage path as resistors in parallel connected between the two lines. This property is called CONDUCTANCE (G) and is the opposite of resistance.

Conductance in transmission lines is expressed as the reciprocal of resistance and is usually given in *micromhos per unit length*.

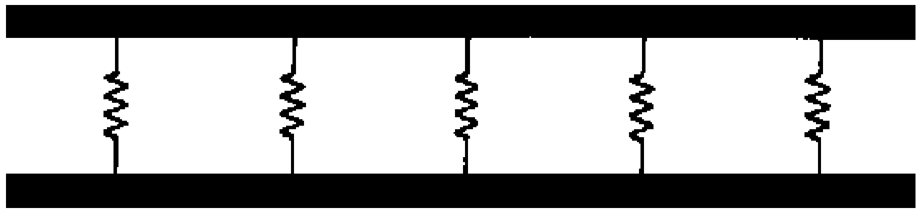


Figure 3-13.—Leakage in a transmission line.

ELECTROMAGNETIC FIELDS ABOUT A TRANSMISSION LINE

The distributed constants of resistance, inductance, and capacitance are basic properties common to all transmission lines and exist whether or not any current flow exists. As soon as current flow and voltage exist in a transmission line, another property becomes quite evident. This is the presence of an electromagnetic field, or lines of force, about the wires of the transmission line. The lines of force themselves are not visible; however, understanding the force that an electron experiences while in the field of these lines is very important to your understanding of energy transmission.

There are two kinds of fields; one is associated with voltage and the other with current. The field associated with voltage is called the ELECTRIC (E) FIELD. It exerts a force on any electric charge placed in it. The field associated with current is called a MAGNETIC (H) FIELD, because it tends to exert a force on any magnetic pole placed in it. Figure 3-14 illustrates the way in which the E fields and H fields tend to orient themselves between conductors of a typical two-wire transmission line. The illustration shows a cross section of the transmission lines. The E field is represented by solid lines and the H field by dotted lines. The arrows indicate the direction of the lines of force. Both fields normally exist together and are spoken of collectively as the electromagnetic field.

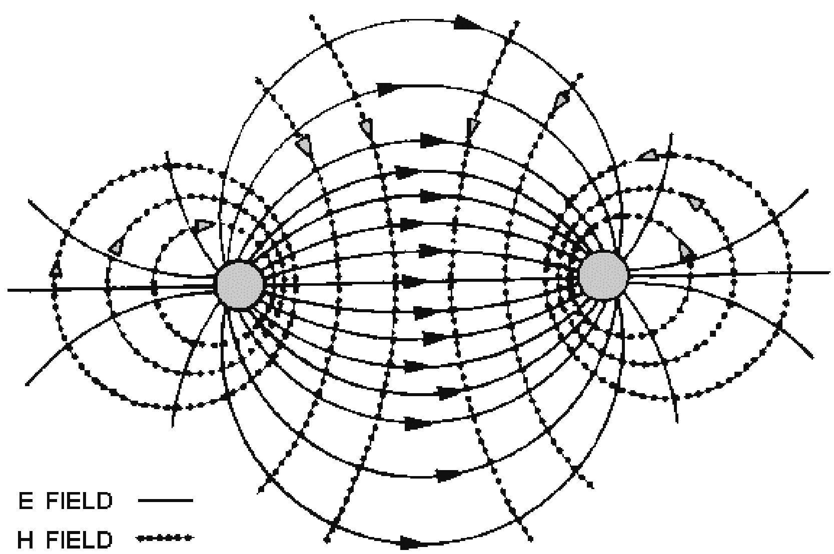


Figure 3-14.—Fields between conductors.

CHARACTERISTIC IMPEDANCE OF A TRANSMISSION LINE

You learned earlier that the maximum (and most efficient) transfer of electrical energy takes place when the source impedance is matched to the load impedance. This fact is very important in the study of transmission lines and antennas. If the characteristic impedance of the transmission line and the load impedance are equal, energy from the transmitter will travel down the transmission line to the antenna with no power loss caused by reflection.

Definition and Symbols

Every transmission line possesses a certain CHARACTERISTIC IMPEDANCE, usually designated as Z_0. Z_0 is the ratio of E to I at every point along the line. If a load equal to the characteristic impedance is placed at the output end of any length of line, the same impedance will appear at the input terminals of the line. The characteristic impedance is the only value of impedance for any given type and size of line that acts in this way. The characteristic impedance determines the amount of current that can flow when a given voltage is applied to an infinitely long line. Characteristic impedance is comparable to the resistance that determines the amount of current that flows in a dc circuit.

In a previous discussion, lumped and distributed constants were explained. Figure 3-15, view A, shows the properties of resistance, inductance, capacitance, and conductance combined in a short section of two-wire transmission line. The illustration shows the evenly distributed capacitance as a single lumped capacitor and the distributed conductance as a lumped leakage path. Lumped values may be used for transmission line calculations if the physical length of the line is very short compared to the wavelength of energy being transmitted. Figure 3-15, view B, shows all four properties lumped together and represented by their conventional symbols.

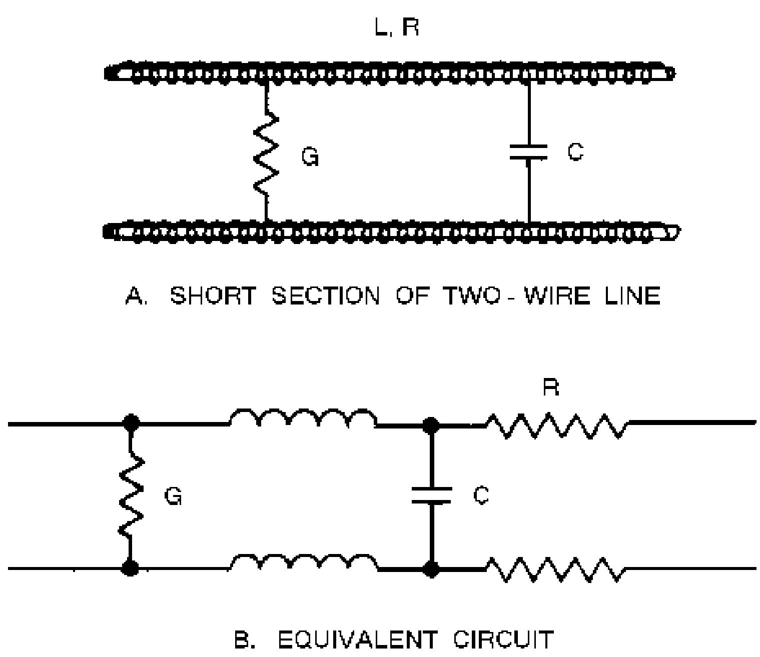


Figure 3-15.—Short section of two-wire transmission line and equivalent circuit.

Q19. Describe the leakage current in a transmission line and in what unit it is expressed.

Q20. All the power sent down a transmission line from a transmitter can be transferred to an antenna under what optimum conditions?

Q21. What symbol is used to designate the characteristic impedance of a line, and what two variables does it compare?

Characteristic Impedance and the Infinite Line

Several short sections, as shown in figure 3-15, can be combined to form a large transmission line, as shown in figure 3-16. Current will flow if voltage is applied across points K and L. In fact, any circuit, such as that represented in figure 3-16, view A, has a certain current flow for each value of applied voltage. The ratio of the voltage to the current is the impedance (Z).

Recall that:

$$Z = \frac{E}{I}$$

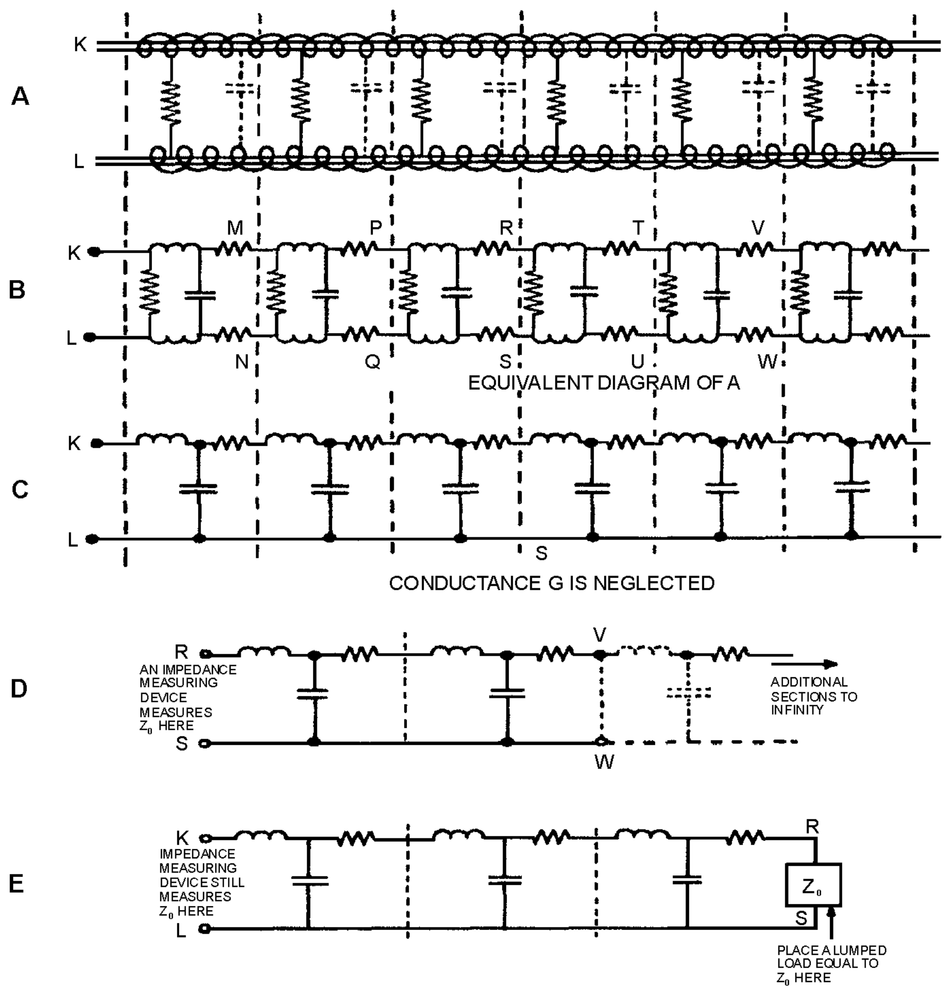


Figure 3-16.—Characteristic impedance.

The impedance presented to the input terminals of the transmission line is not merely the resistance of the wire in series with the impedance of the load. The effects of series inductance and shunt capacitance of the line itself may overshadow the resistance, and even the load, as far as the input terminals are concerned.

To find the input impedance of a transmission line, determine the impedance of a single section of line. The impedance between points K and L, in view B of figure 3-16, can be calculated by the use of series-parallel impedance formulas, provided the impedance across points M and N is known. But since this section is merely one small part of a longer line, another similar section is connected to points M and N. Again, the impedance across points K and L of the two sections can be calculated, provided the impedance of the third section is known. This process of adding one section to another can be repeated endlessly. The addition of each section produces an impedance across points K and L of a new and lower value. However, after many sections have been added, each successive added section has less and less effect on the impedance across points K and L. If sections are added to the line endlessly, the line is infinitely long, and a certain finite value of impedance across points K and L is finally reached.

In this discussion of transmission lines, the effect of conductance (G) is minor compared to that of inductance (L) and capacitance (C), and is frequently neglected. In figure 3-16, view C, G is omitted and the inductance and resistance of each line can be considered as one line.

Let us assume that the sections of view C continue to the right with an infinite number of sections. When an infinite number of sections extends to the right, the impedance appearing across K and L is Z_0. If the line is cut at R and S, an infinite number of sections still extends to the right since the line is endless in that direction. Therefore, the impedance now appearing across points R and S is also Z_0, as illustrated in view D. You can see that if only the first three sections are taken and a load impedance of Z_0 is connected across points R and S, the impedance across the input terminals K and L is still Z_0. The line continues to act as an infinite line. This is illustrated in view E.

Figure 3-17, view A, illustrates how the characteristic impedance of an infinite line can be calculated. Resistors are added in series parallel across terminals K and L in eight steps, and the resultant impedances are noted. In step 1 the impedance is infinite; in step 2 the impedance is 110 ohms. In step 3 the impedance becomes 62.1 ohms, a change of 47.9 ohms. In step 4 the impedance is 48.5 ohms, a change of only 13.6 ohms. The resultant changes in impedance from each additional increment become progressively smaller. Eventually, practically no change in impedance results from further additions to the line. The total impedance of the line at this point is said to be at its characteristic impedance; which, in this case, is 37 ohms. This means that an infinite line constructed as indicated in step 8 could be effectively replaced by a 37-ohm resistor. View B shows a 37-ohm resistor placed in the line at various points to replace the infinite line of step 8 in view A. There is no change in total impedance.

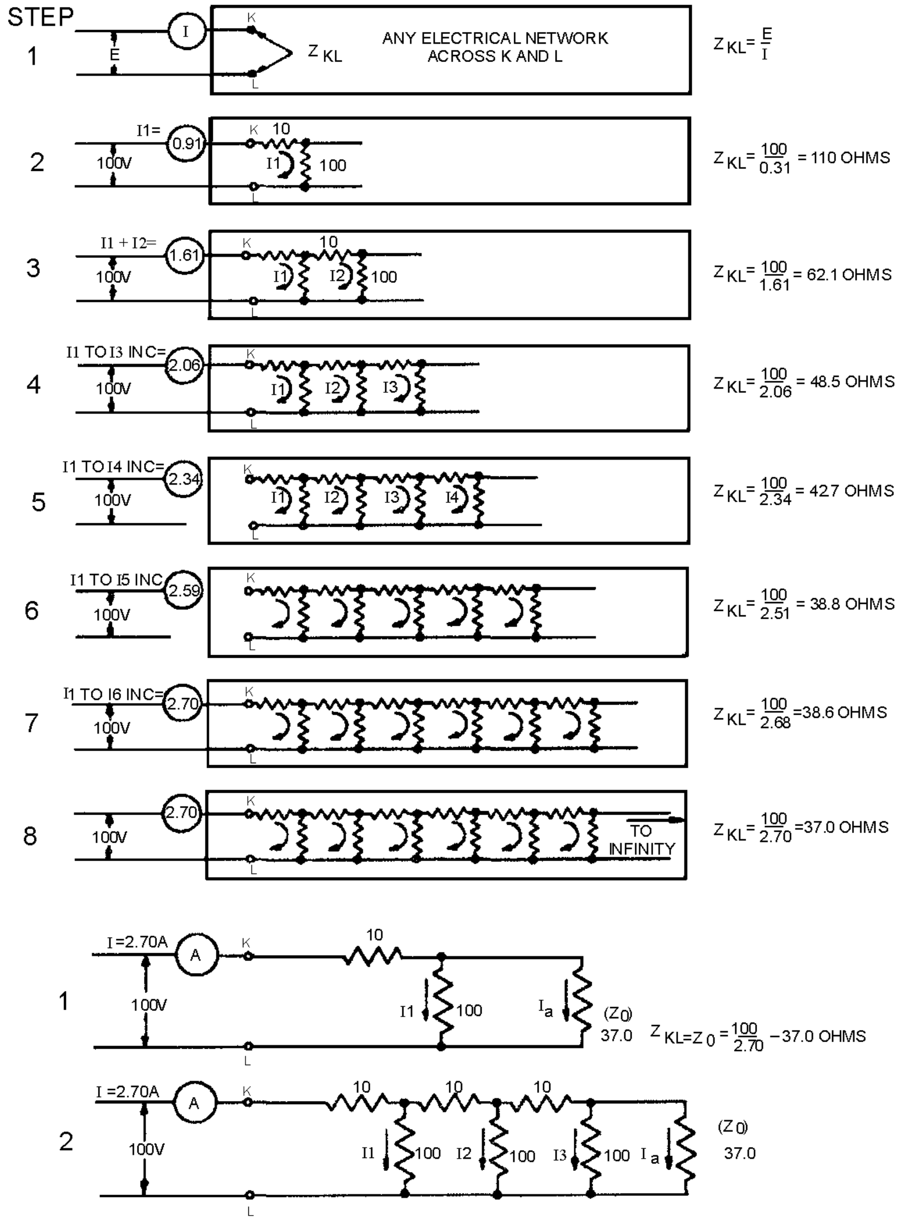


Figure 3-17.—Termination of a line.

In figure 3-17, resistors were used to show impedance characteristics for the sake of simplicity. Figuring the actual impedance of a line having reactance is very similar, with inductance taking the place of the series resistors and capacitance taking the place of the shunt resistors. The characteristic impedance of lines in actual use normally lies between 50 and 600 ohms.

When a transmission line is "short" compared to the length of the radio-frequency waves it carries, the opposition presented to the input terminals is determined primarily by the load impedance. A small amount of power is dissipated in overcoming the resistance of the line. However, when the line is "long" and the load is an incorrect impedance, the voltages necessary to drive a given amount of current through the line cannot be accounted for by considering just the impedance of the load in series with the impedance of the line. The line has properties other than resistance that affect input impedance. These properties are inductance in series with the line, capacitance across the line, resistance leakage paths across the line, and certain radiation losses.

Q22. What is the range of the characteristic impedance of lines used in actual practice?

VOLTAGE CHANGE ALONG A TRANSMISSION LINE

Let us summarize what we have just discussed. In an electric circuit, energy is stored in electric and magnetic fields. These fields must be brought to the load to transmit that energy. At the load, energy contained in the fields is converted to the desired form of energy.

Transmission of Energy

When the load is connected directly to the source of energy, or when the transmission line is short, problems concerning current and voltage can be solved by applying Ohm's law. When the transmission line becomes long enough so the time difference between a change occurring at the generator and the change appearing at the load becomes appreciable, analysis of the transmission line becomes important.

Dc Applied to a Transmission Line

In figure 3-18, a battery is connected through a relatively long two-wire transmission line to a load at the far end of the line. At the instant the switch is closed, neither current nor voltage exists on the line. When the switch is closed, point A becomes a positive potential, and point B becomes negative. These points of difference in potential move down the line. However, as the initial points of potential leave points A and B, they are followed by new points of difference in potential which the battery adds at A and B. This is merely saying that the battery maintains a constant potential difference between points A and B. A short time after the switch is closed, the initial points of difference in potential have reached points A' and B'; the wire sections from points A to A' and points B to B' are at the same potential as A and B, respectively. The points of charge are represented by plus (+) and minus (-) signs along the wires. The directions of the currents in the wires are represented by the arrowheads on the line, and the direction of travel is indicated by an arrow below the line. Conventional lines of force represent the electric field that exists between the opposite kinds of charge on the wire sections from A to A' and B to B'. Crosses (tails of arrows) indicate the magnetic field created by the electric field moving down the line. The moving electric field and the accompanying magnetic field constitute an electromagnetic wave that is moving from the generator (battery) toward the load. This wave travels at approximately the speed of light in free space. The energy reaching the load is equal to that developed at the battery (assuming there are no losses in the transmission line). If the load absorbs all of the energy, the current and voltage will be evenly distributed along the line.

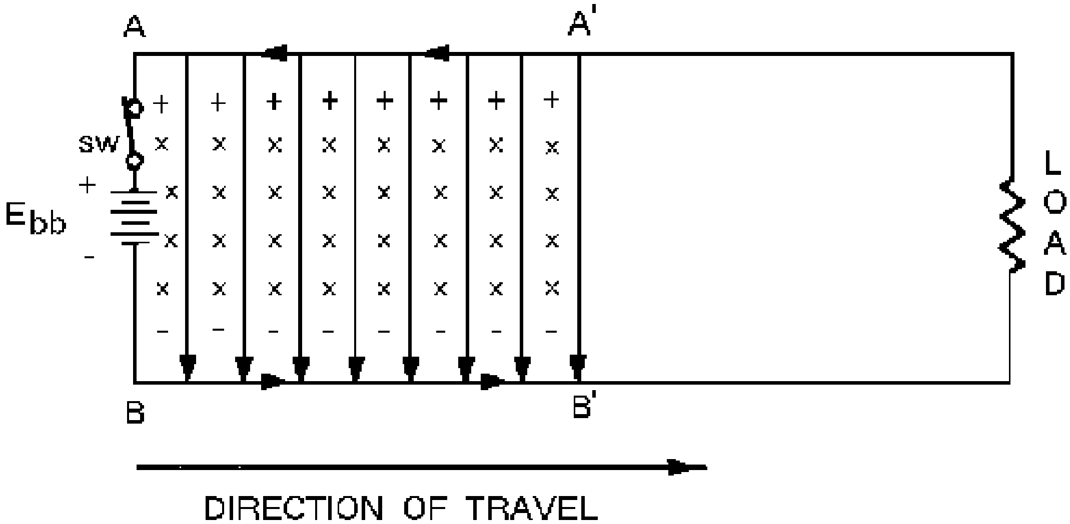


Figure 3-18.—Dc voltage applied to a line.

Ac Applied to a Transmission Line

When the battery of figure 3-18 is replaced by an ac generator (fig. 3-19), each successive instantaneous value of the generator voltage is propagated down the line at the speed of light. The action is similar to the wave created by the battery except that the applied voltage is sinusoidal instead of constant. Assume that the switch is closed at the moment the generator voltage is passing through zero and that the next half cycle makes point A positive. At the end of one cycle of generator voltage, the current and voltage distribution will be as shown in figure 3-19.

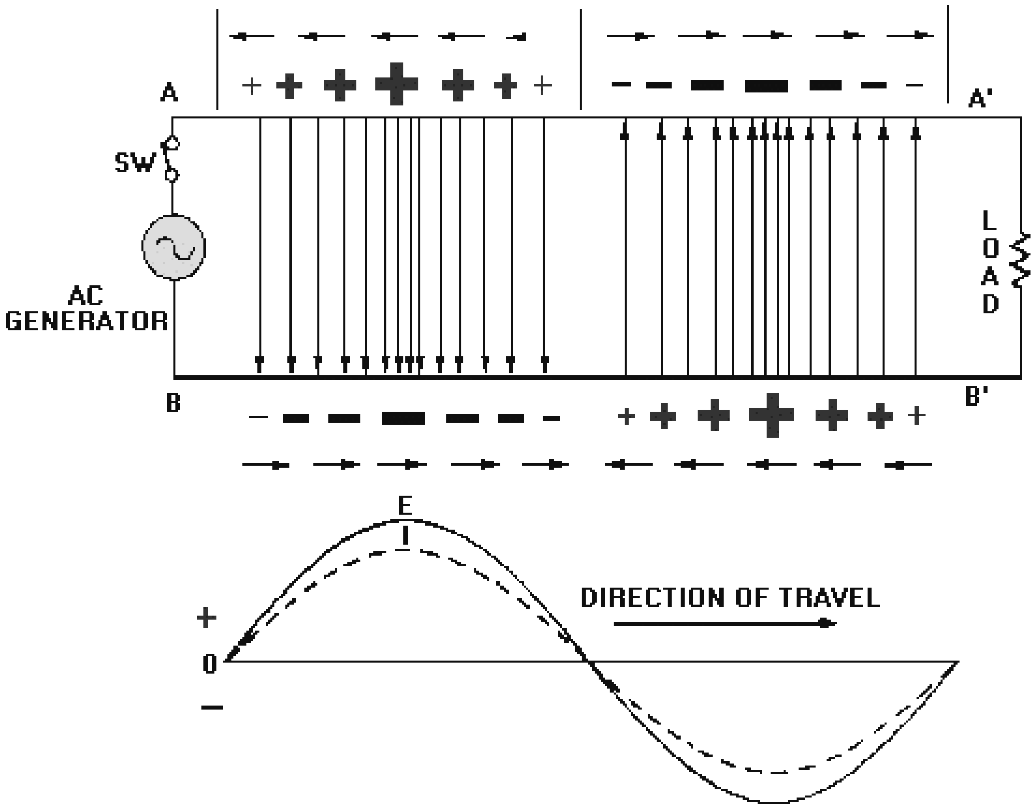


Figure 3-19.—Ac voltage applied to a line.

In this illustration the conventional lines of force represent the electric fields. For simplicity, the magnetic fields are not shown. Points of charge are indicated by plus (+) and minus (−) signs, the larger signs indicating points of higher amplitude of both voltage and current. Short arrows indicate direction of current (electron flow). The waveform drawn below the transmission line represents the voltage (E) and current (I) waves. The line is assumed to be infinite in length so there is no reflection. Thus, traveling sinusoidal voltage and current waves continually travel in phase from the generator toward the load, or far end of the line. Waves traveling from the generator to the load are called INCIDENT WAVES. Waves traveling from the load back to the generator are called REFLECTED WAVES and will be explained in later paragraphs.

Dc Applied to an Infinite Line

Figure 3-20 shows a battery connected to a circuit that is the equivalent of a transmission line. In this line the series resistance and shunt conductance are not shown. In the following discussion the line will be considered to have no losses.

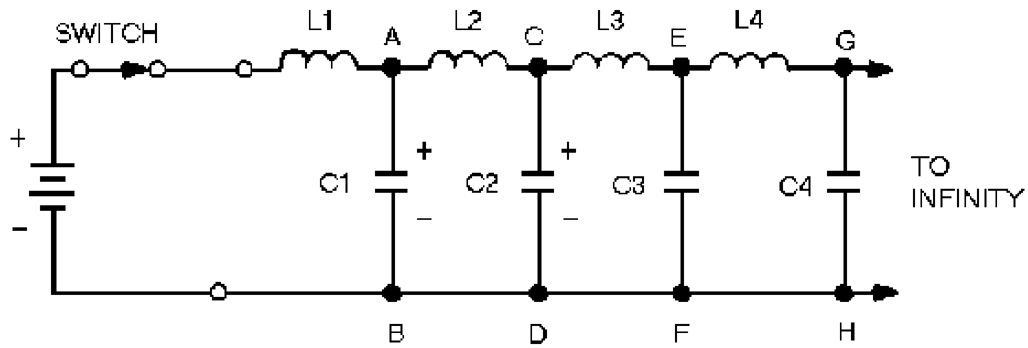


Figure 3-20.—Dc applied to an equivalent transmission line.

As the switch is closed, the battery voltage is applied to the input terminals of the line. Now, C1 has no charge and appears, effectively, as a short circuit across points A and B. The full battery voltage appears across inductor L1. Inductor L1 opposes the change of current (0 now) and limits the rate of charge of C1.

Capacitor C2 cannot begin to charge until after C1 has charged. No current can flow beyond points A and B until C1 has acquired some charge. As the voltage across C1 increases, current through L2 and C2 charges C2. This action continues down the line and charges each capacitor, in turn, to the battery voltage. Thus a voltage wave is traveling along the line. Beyond the wavefront, the line is uncharged. Since the line is infinitely long, there will always be more capacitors to be charged, and current will not stop flowing. Thus current will flow indefinitely in the line.

Notice that current flows to charge the capacitors along the line. The flow of current is not advanced along the line until a voltage is developed across each preceding capacitor. In this manner voltage and current move down the line together in phase.

Ac Applied to an Infinite Line

An rf line displays similar characteristics when an ac voltage is applied to its sending end or input terminals. In figure 3-21, view A, an ac voltage is applied to the line represented by the circuit shown.

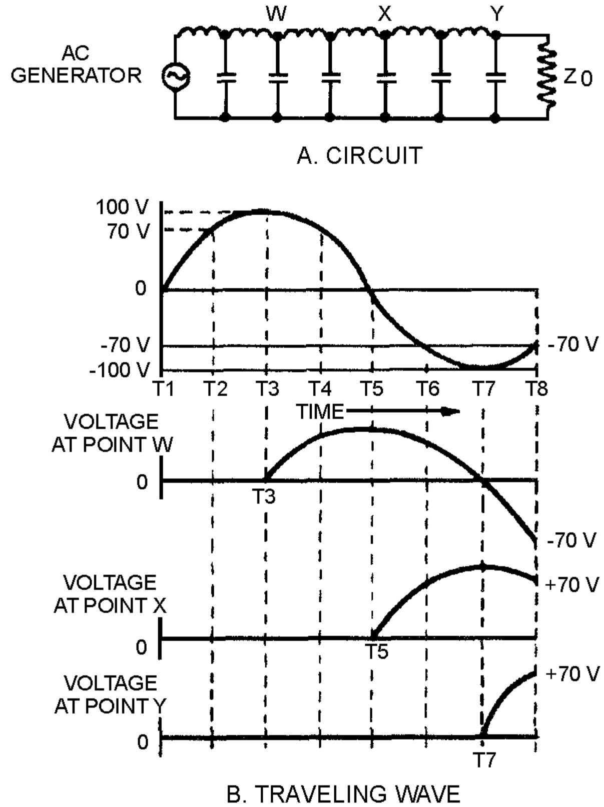


Figure 3-21.—Ac applied to an equivalent transmission line.

In view B the generator voltage starts from zero (T1) and produces the voltage shown. As soon as a small voltage change is produced, it starts its journey down the line while the generator continues to produce new voltages along a sine curve. At T2 the generator voltage is 70 volts. The voltages still move along the line until, at T3, the first small change arrives at point W, and the voltage at that point starts increasing. At T5, the same voltage arrives at point X on the line. Finally, at T7, the first small change arrives at the receiving end of the line. Meanwhile, all the changes in the sine wave produced by the generator pass each point in turn. The amount of time required for the changes to travel the length of the line is the same as that required for a dc voltage to travel the same distance.

At T7, the voltage at the various points on the line is as follows:

At the generator:	-100 V
At point W:	0 V
At point X:	+100 V
At point Y:	0 V

If these voltages are plotted along the length of the line, the resulting curve is like the one shown in figure 3-22, view A. Note that such a curve of instantaneous voltages resembles a sine wave. The changes in voltage that occur between T7 and T8 are as follows:

At the generator:	Rise from	-100 V to -70 V
At point W:	Drop from	0 V to -70 V
At point X:	Drop from	+100 V to +70 V
At point Y:	Rise from	0 V to + 70 V

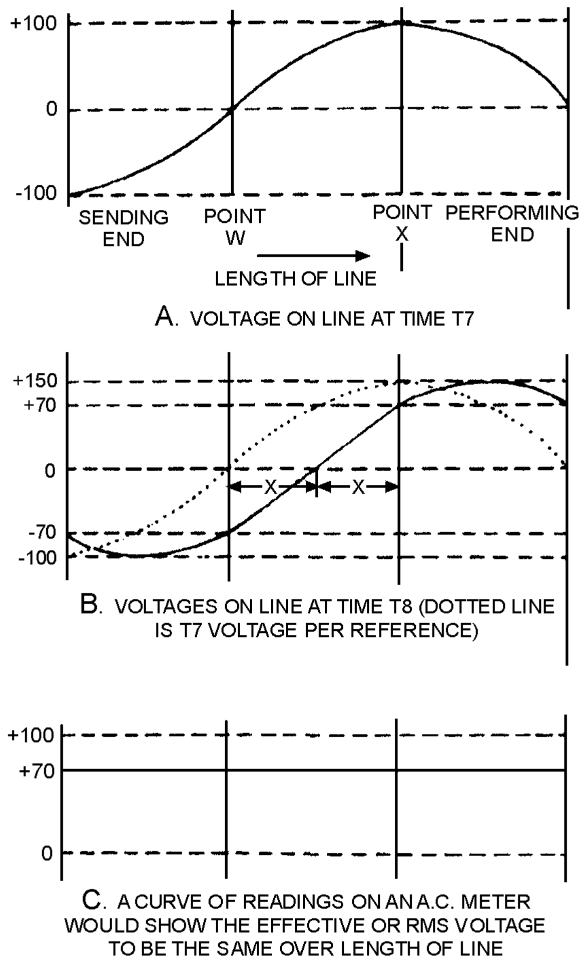


Figure 3-22.—Instantaneous voltages along a transmission line.

A plot of these new voltages produces the solid curve shown in figure 3-22, view B. For reference, the curve from T7 is drawn as a dotted line. The solid curve has exactly the same shape as the dotted curve, but has moved to the right by the distance X. Another plot at T9 would show a new curve similar to the one at T8, but moved to the right by the distance Y.

By analyzing the points along the graph just discussed, you should be able to see that the actions associated with voltage changes along an rf line are as follows:

1. All instantaneous voltages of the sine wave produced by the generator travel down the line in the order they are produced.

2. At any point, a sine wave can be obtained if all the instantaneous voltages passing the point are plotted. An oscilloscope can be used to plot these values of instantaneous voltages against time.

3. The instantaneous voltages (oscilloscope displays) are the same in all cases except that a phase difference exists in the displays seen at different points along the line. The phase changes continually with respect to the generator until the change is 360 degrees over a certain length of line.

4. All parts of a sine wave pass every point along the line. A plot of the readings of an ac meter (which reads the effective value of the voltage over a given time) taken at different points along the line shows that the voltage is constant at all points. This is shown in view C of figure 3-22.

5. Since the line is terminated with a resistance equal to Z_0, the energy arriving at the end of the line is absorbed by the resistance.

VELOCITY OF WAVE PROPAGATION

If a voltage is initially applied to the sending end of a line, that same voltage will appear later some distance from the sending end. This is true regardless of any change in voltage, whether the change is a jump from zero to some value or a drop from some value to zero. The voltage change will be conducted down the line at a constant rate.

Recall that the inductance of a line delays the charging of the line capacitance. The velocity of propagation is therefore related to the values of L and C. If the inductance and capacitance of the rf line are known, the time required for any waveform to travel the length of the line can be determined. To see how this works, observe the following relationship:

$$Q = IT$$

This formula shows that the total charge or quantity is equal to the current multiplied by the time the current flows. Also:

$$Q = CE$$

This formula shows that the total charge on a capacitor is equal to the capacitance multiplied by the voltage across the capacitor.

If the switch in figure 3-23 is closed for a given time, the quantity (Q) of electricity leaving the battery can be computed by using the equation Q = IT. The electricity leaves the battery and goes into the line, where a charge is built up on the capacitors. The amount of this charge is computed by using the equation Q = CE.

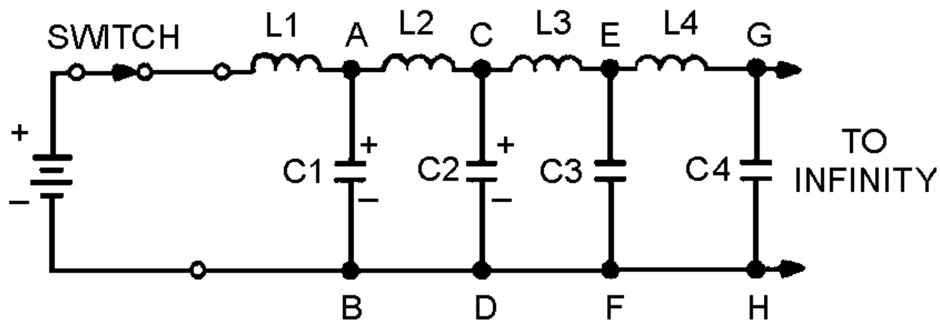


Figure 3-23.—Dc applied to an equivalent transmission line.

Since none of the charge is lost, the total charge leaving the battery during T is equal to the total charge on the line. Therefore:

$$Q = IT = CE$$

As each capacitor accumulates a charge equal to CE, the voltage across each inductor must change. As C1 in figure 3-23 charges to a voltage of E, point A rises to a potential of E volts while point B is still at zero volts. This makes E appear across L2. As C2 charges, point B rises to a potential of E volts as did point A. At this time, point B is at E volts and point C rises. Thus, we have a continuing action of voltage moving down the infinite line.

In an inductor, these circuit components are related, as shown in the formula

$$E = L\left(\frac{\Delta I}{\Delta T}\right).$$

This shows that the voltage across the inductor is directly proportional to inductance and the change in current, but inversely proportional to a change in time. Since current and time start from zero, the change in time (ΔT) and the change in current (ΔI) are equal to the final time (T) and final current (I). For this case the equation becomes:

$$ET = LI$$

If voltage E is applied for time (T) across the inductor (L), the final current (I) will flow. The following equations show how the three terms (T, L, and C) are related:

$$IT = CE$$
$$ET = LI$$

For convenience, you can find T in terms of L and C in the following manner. Multiply the left and right member of each equation as follows:

$$(IT)(ET) = (CE)(LI)$$

Then: $EIT^2 = LCEI$

Dividing by (EI): $T^2 = LC$

and $T = \sqrt{LC}$

This final equation is used for finding the time required for a voltage change to travel a unit length, since L and C are given in terms of unit length. The velocity of the waves may be found by:

$$V = \frac{D}{T} \quad \text{or} \quad V = \frac{D}{\sqrt{LC}}$$

Where: D is the physical length of a unit

This is the rate at which the wave travels over a unit length. The units of L and C are henrys and farads, respectively. T is in seconds per unit length and V is in unit lengths per second.

DETERMINING CHARACTERISTIC IMPEDANCE

As previously discussed, an infinite transmission line exhibits a definite input impedance. This impedance is the CHARACTERISTIC IMPEDANCE and is independent of line length. The exact value of this impedance is the ratio of the input voltage to the input current. If the line is infinite or is terminated in a resistance equal to the characteristic impedance, voltage and current waves traveling the line are in phase. To determine the characteristic impedance or voltage-to-current ratio, use the following procedure:

Divide the equation:

$ET = LI$ by $IT = CE$

$$\frac{ET}{IT} = \frac{LI}{CE}$$

Multiply by $\frac{E}{I}$:

$$\frac{E^2T}{I^2T} = \frac{LIE}{CEI}$$

Simplify:

$$\frac{E^2}{I^2} = \frac{L}{C}$$

Take the square root:

$$\frac{E}{I} = \sqrt{\frac{L}{C}} = Z_0 (\text{characteristic impedance})$$

Example:

A problem using this equation will illustrate how to determine the characteristics of a transmission line. Assume that the line shown in figure 3-23 is 1000 feet long. A 100-foot (approximately 30.5 meter) section is measured to determine L and C. The section is found to have an inductance of 0.25 millihenries and a capacitance of 1000 picofarads. Find the characteristic impedance of the line and the velocity of the wave on the line.

The characteristic impedance is:

$$Z_0 = \sqrt{LC}$$

$$Z_0 = \sqrt{\frac{0.25 \times 10^{-3}}{1000 \times 10^{-12}}}$$

$$Z_0 = \sqrt{0.25 \times 10^{6}}$$

$$Z_0 = 0.5 \times 10^{3}$$

$$Z_0 = 500\Omega$$

If any other unit length had been considered, the values of L and C would be different, but their ratio would remain the same as would the characteristic impedance.

The formula for T is:

$$T = \sqrt{LC}$$

$$T = \sqrt{0.25 \times 10^{-3} \times 1000 \times 10^{-12}}$$

$$T = \sqrt{.025 \times 10^{-12}}$$

$$T = 0.5 \times 10^{-6} \text{ second}$$

$$T = 0.5 \text{ microsecond}$$

The formula for the velocity of a wave is:

$$V = \frac{D}{T}$$

$$V = \frac{100 \text{ feet}}{0.5 \times 10^{-6} \text{ second}}$$

$$V = 200 \times 10^{6} \text{ feet/second}$$

$$V = 200{,}000{,}000 \text{ feet/second}$$

REFLECTIONS ON A TRANSMISSION LINE

Transmission line characteristics are based on an infinite line. A line cannot always be terminated in its characteristic impedance since it is sometimes operated as an OPEN-ENDED line and other times as a SHORT-CIRCUIT at the receiving end. If the line is open-ended, it has a terminating impedance that is infinitely large. If a line is not terminated in characteristic impedance, it is said to be finite.

When a line is not terminated in Z_0, the incident energy is not absorbed but is returned along the only path available—the transmission line. Thus, the behavior of a finite line may be quite different from that of the infinite line.

REFLECTION OF DC VOLTAGE FROM AN OPEN CIRCUIT

The equivalent circuit of an open-ended transmission line is shown in figure 3-24, view A. Again, losses are to be considered as negligible, and L is lumped in one branch. Assume that (1) the battery in this circuit has an internal impedance equal to the characteristic impedance of the transmission line ($Z_i = Z_0$); (2) the capacitors in the line are not charged before the battery is connected; and (3) since the line is open-ended, the terminating impedance is infinitely large.

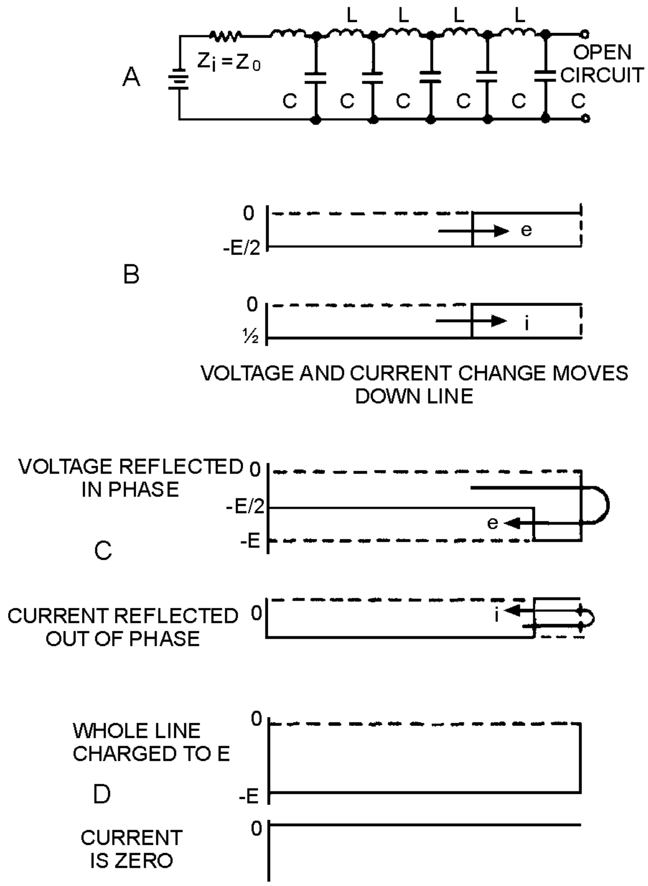


Figure 3-24.—Reflection from an open-ended line.

When the battery is connected to the sending end as shown, a negative voltage moves down the line. This voltage charges each capacitor, in turn, through the preceding inductor. Since Z_i equals Z_0, one-half the applied voltage will appear across the internal battery impedance, Z_i, and one-half across the impedance of the line, Z_0. Each capacitor is then charged to E/2 (view B). When the last capacitor in the line is charged, there is no voltage across the last inductor and current flow through the last inductor stops. With no current flow to maintain it, the magnetic field in the last inductor collapses and forces current to continue to flow in the same direction into the last capacitor. Because the direction of current has not changed, the capacitor charges in the same direction, thereby increasing the charge in the capacitor. Since the energy in the magnetic field equals the energy in the capacitor, the energy transfer to the capacitor doubles the voltage across the capacitor. The last capacitor is now charged to E volts and the current in the last inductor drops to zero.

At this point, the same process takes place with the next to the last inductor and capacitor. When the magnetic field about the inductor collapses, current continues to flow into the next to the last capacitor, charging it to E volts. This action continues backward down the line until the first capacitor has been fully charged to the applied voltage. This change of voltage, moving backward down the line, can be thought of in the following manner. The voltage, arriving at the end of the line, finds no place to go and returns to the sending end with the same polarity (view C). Such action is called REFLECTION.

When a reflection of voltage occurs on an open-ended line, the polarity is unchanged. The voltage change moves back to the source, charging each capacitor in turn until the first capacitor is charged to the source voltage and the action stops (view D). As each capacitor is charged, current in each inductor drops to zero, effectively reflecting the current with the opposite polarity (view C). Reflected current of opposite polarity cancels the original current at each point, and the current drops to zero at that point. When the last capacitor is charged, the current from the source stops flowing (view D).

Important facts to remember in the reflection of dc voltages in open-ended lines are:

- Voltage is reflected from an open end without change in polarity, amplitude, or shape.

- Current is reflected from an open end with opposite polarity and without change in amplitude or shape.

REFLECTION OF DC VOLTAGE FROM A SHORT CIRCUIT

A SHORT-CIRCUITED line affects voltage change differently from the way an open-circuited line affects it. The voltage across a perfect short circuit must be zero; therefore, no power can be absorbed in the short, and the energy is reflected toward the generator.

The initial circuit is shown in figure 3-25, view A. The initial voltage and current waves (view B) are the same as those given for an infinite line. In a short-circuited line the voltage change arrives at the last inductor in the same manner as the waves on an open-ended line. In this case, however, there is no capacitor to charge. The current through the final inductor produces a voltage with the polarity shown in view C. When the field collapses, the inductor acts as a battery and forces current through the capacitor in the opposite direction, causing it to discharge (view D). Since the amount of energy stored in the magnetic field is the same as that in the capacitor, the capacitor discharges to zero.

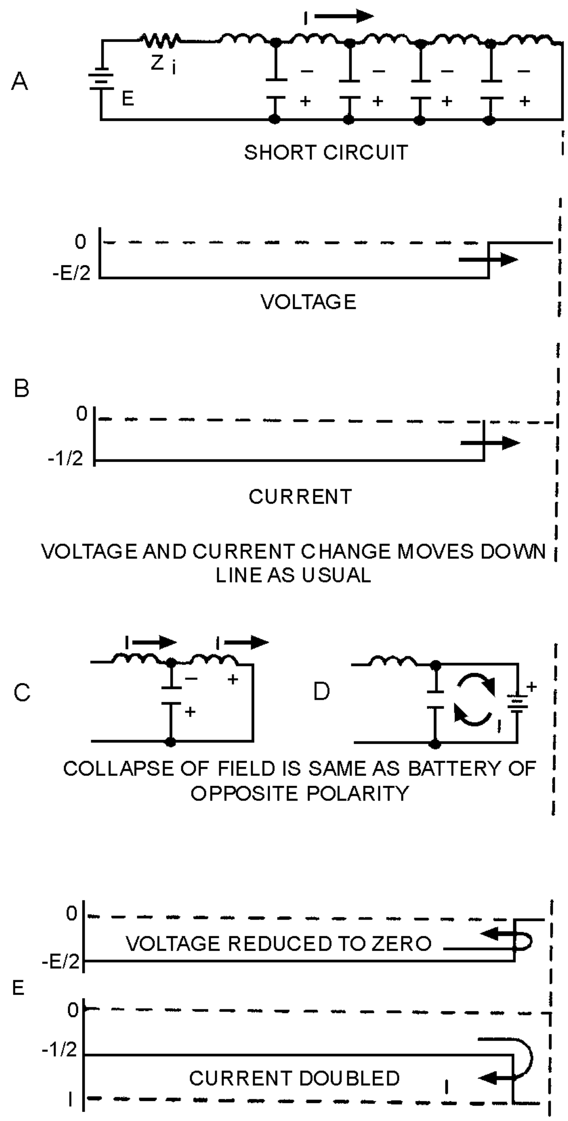


Figure 3-25.—Reflection from a short-circuited line.

Now there is no voltage to maintain the current through the next to the last inductor. Therefore, this inductor discharges the next to the last capacitor.

As each capacitor is discharged to zero, the next inductor effectively becomes a new source of voltage. The amplitude of each of these voltages is equal to E/2, but the polarity is the opposite of the battery at the input end of the line. The collapsing field around each inductor, in turn, produces a voltage that forces the current to continue flowing in the same direction, adding to the current from the source to make it 2I. This action continues until all the capacitors are discharged (view E).

Reflected waves from a short-circuited transmission line are characterized as follows:

- The reflected voltage has the opposite polarity but the same amplitude as the incident wave.
- The reflected current has the same polarity and the same amplitude as the incident current.

REFLECTION OF AC VOLTAGE FROM AN OPEN CIRCUIT

In most cases where rf lines are used, the voltages applied to the sending end are ac voltages. The action at the receiving end of the line is exactly the same for ac as for dc. In the open-ended line, shown in figure 3-26, view A, the generated ac voltage is distributed along the line, shown in view B. This voltage is distributed in such a way that as each instantaneous voltage arrives at the end, it is reflected with the same polarity and amplitude. When ac is used, this reflection is in phase. Each of the reflected voltages travels back along the line until it reaches the generator. If the generator impedance is the same as the line impedance, energy arriving at the generator is absorbed and not reflected again. Now two voltages are on the line.

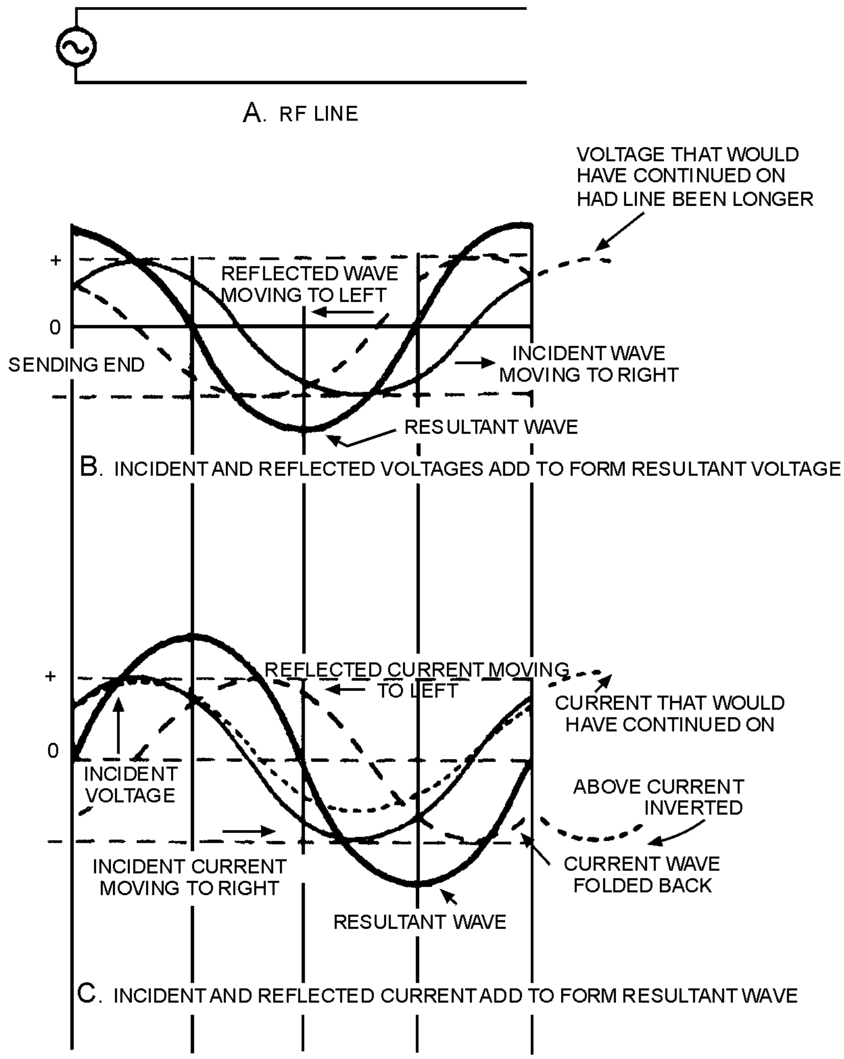


Figure 3-26.—Formation of standing waves.

View B shows how two waves of the same frequency and amplitude moving in opposite directions on the same conductor will combine to form a resultant wave. The small solid line is moving steadily from left to right and is the INCIDENT WAVE (from the source). The broken-line waveform is moving from right to left and is the REFLECTED WAVE. The resultant waveform, the heavy line, is found by algebraically adding instantaneous values of the two waveforms. The resultant waveform has an instantaneous peak amplitude that is equal to the sum of the peak amplitudes of the incident and reflected waves. Since most indicating instruments are unable to separate these voltages, they show the vector sum. An oscilloscope is usually used to study the instantaneous voltages on rf lines.

Since two waves of voltage are moving on the line, you need to know how to distinguish between the two. The voltages moving toward the receiving end are called INCIDENT VOLTAGES, and the whole waveshape is called the INCIDENT WAVE. The wave moving back to the sending end after reflection is called the REFLECTED WAVE. The resultant voltage curve (view B of figure 3-26) shows that the voltage is maximum at the end of the line, a condition that occurs across an open circuit.

Another step in investigating the open-circuited rf line is to see how the current waves act. The incident current wave is the solid line in figure 3-26, view C. The voltage is represented by the dotted line. The current is in phase with the voltage while traveling toward the receiving end. At the end of the line, the current is reflected in the opposite polarity; that is, it is shifted 180 degrees in phase, but its amplitude remains the same. The reflected wave of current is shown by dashed lines in view C. The heavy-line curve represents the sum of the two instantaneous currents and is the resultant wave. Notice that current is zero at the end of the line. This is reasonable, since there can be no current flow through an open circuit.

Views B and C of figure 3-26 show the voltage and current distribution along a transmission line at a point about 1/8 after a maximum voltage or current reaches the end of the line. Since the instantaneous values are continuously changing during the generation of a complete cycle, a large number of these pictures are required to show the many different relationships.

Figure 3-27 shows the incident and reflected waveshapes at several different times. The diagrams in the left column of figure 3-27 (representing *voltage*) show the incident wave and its reflection without change in polarity. In figure 3-27, waveform (1), the incident wave and the reflected wave are added algebraically to produce the resultant wave indicated by the heavy line. In waveform (2), a zero point preceding the negative-going cycle of the incident wave is at the end of the line. The reflected wave and incident wave are 180 degrees out of phase at all points. (The reflected wave is the positive cycle that just preceded the negative cycle now approaching the end of the line.) The resultant of the incident and reflected waves is zero at all points along the line. In waveform (3), the waves have moved $1/8\lambda$ along the line; the incident wave has moved 45 degrees to the right, and the reflected wave has moved 45 degrees to the left. The resultant voltage, shown by the heavy line, has a maximum negative at the end of the line and a maximum positive $1/2\lambda$ from the end of the line.

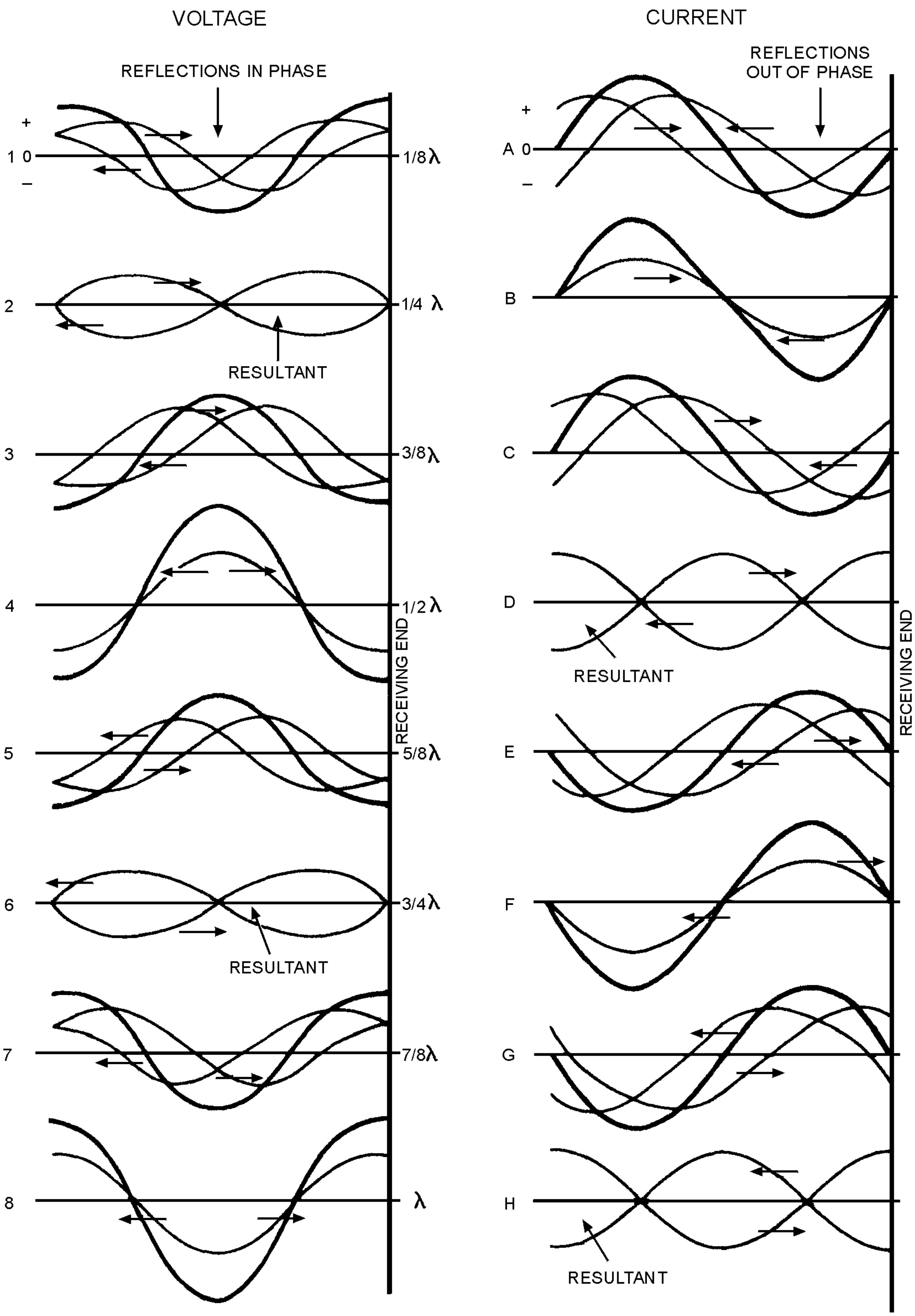


Figure 3-27.—Instantaneous values of incident and reflected waves on an open-ended line.

In waveform (4), the incident wave is at a maximum negative value at the end of the line. The wave has moved another 45 degrees to the right from the wave in the preceding illustration. The reflected wave has also moved 45 degrees, but to the left. The reflected wave is in phase with the incident wave. The resultant of these two waves, shown by the dark line, again has a negative maximum at the end of the line and a positive maximum 1/2λ from the end of the line. Notice that these maxima have a greater amplitude than those in waveform (3).

In waveform (5), the incident wave has moved another 45 degrees to the right and the reflected wave 45 degrees to the left. The resultant again is maximum negative at the end and positive maximum 1/2λ from the end. The maxima are lower than those in waveform (4). In waveform (6), the incident and reflected wave have moved another 1/8λ. The two waves again are 180 degrees out of phase, giving a resultant wave with no amplitude. The incident and reflected waves continue moving in opposite directions, adding to produce the resultant waveshapes shown in waveforms (7) and (8). Notice that the maximum voltage in each resultant wave is at the end and 1/2λ from the end.

Study each part of figure 3-27 carefully and you will get a clear picture of how the resultant waveforms of voltage are produced. You will also see that the resultant voltage wave on an open-ended line is always zero at 1/4λ and 3/4λ from the end of the transmission line. Since the zero and maximum points are always in the same place, the resultant of the incident and the reflected wave is called a STANDING WAVE of voltage.

The right-hand column in figure 3-27 shows the *current* waveshapes on the open-ended line. Since the current is reflected out of phase at an open end, the resultant waveshapes differ from those for voltage. The two out-of-phase components always cancel at the end of the transmission line, so the resultant is always zero at that point. If you check all the resultant waveshapes shown in the right-hand column of figure 3-27, you will see that a zero point always occurs at the end and at a point 1/2λ from the end. Maximum voltages occur 1/4λ and 3/4λ from the end.

When an ac meter is used to measure the voltages and currents along a line, the polarity is not indicated. If you plot all the current and voltage readings along the length of the line, you will get curves like the ones shown in figure 3-28. Notice that all are positive. These curves are the conventional method of showing current and voltage standing waves on rf lines.

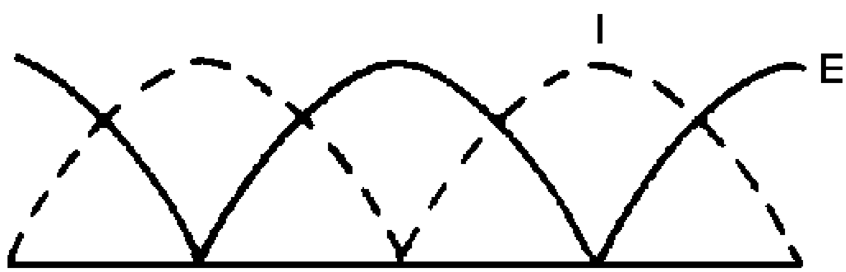


Figure 3-28.—Conventional picture of standing waves.

When an rf line is terminated in a short circuit, reflection is complete, but the effect on voltage and current differs from that in an open-ended line. Voltage is reflected in opposite phase, while current is reflected in phase. Again refer to the series of pictures shown in figure 3-27. However, this time the left column represents *current*, since it shows reflection in phase; and the right column of pictures now represents the *voltage* changes on the shorted line, since it shows reflection out of phase.

The composite diagram in figure 3-29 shows all resultant curves on a full-wavelength section of line over a complete cycle. Notice that the amplitude of the voltage varies between zero and maximum in both directions at the center and at both ends as well but, one-fourth of the distance from each end the voltage is always zero. The resultant waveshape is referred to as a standing wave of voltage. Standing waves, then, are caused by reflections, which occur only when the line is not terminated in its characteristic impedance.

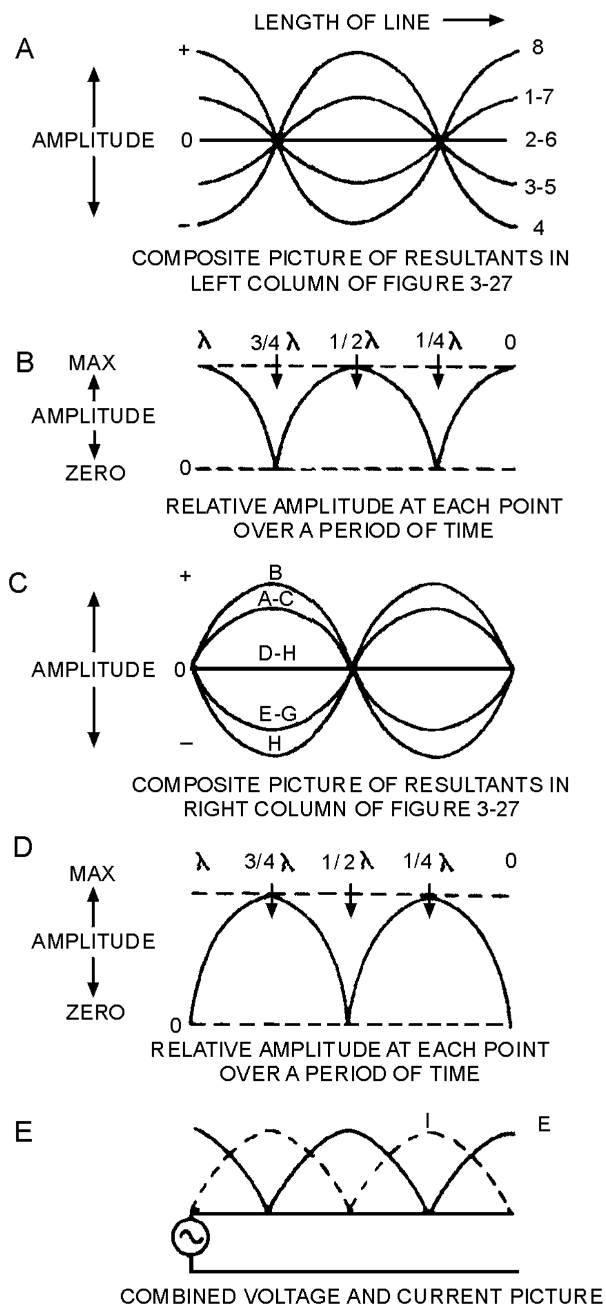


Figure 3-29.—Composite results of instantaneous waves.

The voltage at the center and the ends varies at a sinusoidal rate between the limits shown. At the one-fourth the three-fourths points, the voltage is always zero. A continuous series of diagrams such as these is difficult to see with conventional test equipment, which reads the effective or average voltage over several cycles. The curve of amplitude over the length of line for several cycles is shown in figure 3-29, view B. A meter will read zero at the points shown and will show a maximum voltage at the center, no matter how many cycles pass.

As shown in view D, the amplitude varies along the length of the line. In this case it is zero at the end and center but maximum at the one-fourth and three-fourths points. The entire diagram of the open-ended line conditions is shown in view E. The standing waves of voltage and current appear together. Observe that one is maximum when the other is minimum. The current and voltage standing waves are one-quarter cycle, or 90 degrees, out of phase with one another.

REFLECTION OF AC VOLTAGE FROM A SHORT CIRCUIT

Reflection is complete when an rf line is terminated in a short circuit, but the effect on voltage and current differs from the effect obtained in an open-ended line. Voltage is reflected in opposite phase, while current is reflected in phase. Again look at the series of diagrams in figure 3-27. The left column represents current, and the right column shows voltage changes on the shorted line. The standard representation of standing waves on a shorted line is shown in figure 3-30; the voltage is a solid line, and the current is a dashed line. The voltage is zero at the end and center (1/2λ) and maximum at the 1/4λ and 3/4λ points, while the current is maximum at the end and center and minimum at the 1/4λ and 3/4λ points.

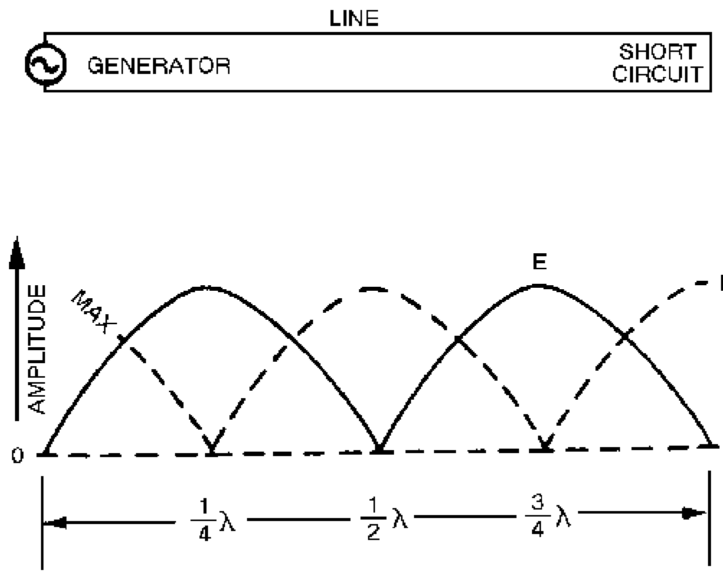


Figure 3-30.—Standing waves on a shorted line.

As we discussed voltage and current waves on transmission lines, we pointed out several differences between open and shorted lines. Basic differences also appear in the standing-wave patterns for open and shorted lines. You can see these differences by comparing figure 3-29, view E, and figure 3-30. Notice that the current and voltage standing waves are shifted 90 degrees with respect to the termination. At the open end of a line, voltage is maximum (zero if there are no losses in the line). At a short circuit, current is maximum and voltage is minimum.

Q23. Two types of waves are formed on a transmission line. What names are given to these waves?

Q24. In figure 3-27, which waveforms on the left have a resultant wave of zero, and what is indicated by these waves?

Q25. On an open-ended transmission line, the voltage is always zero at what distance from each end of the line?

TERMINATING A TRANSMISSION LINE

A transmission line is either NONRESONANT or RESONANT. First, let us define the terms nonresonant lines and resonant lines. A nonresonant line is a line that has no standing waves of current and voltage. A resonant line is a line that has standing waves of current and voltage.

Nonresonant Lines

A nonresonant line is either infinitely long or terminated in its characteristic impedance. Since no reflections occur, all the energy traveling down the line is absorbed by the load which terminates the line. Since no standing waves are present, this type of line is sometimes spoken of as a FLAT line. In addition, because the load impedance of such a line is equal to Z_0, no special tuning devices are required to effect a maximum power transfer; hence, the line is also called an UNTUNED line.

Resonant Lines

A resonant line has a finite length and is not terminated in its characteristic impedance. Therefore reflections of energy do occur. The load impedance is different from the Z_0 of the line; therefore, the input impedance may not be purely resistive but may have reactive components. Tuning devices are used to eliminate the reactance and to bring about maximum power transfer from the source to the line. Therefore, a resonant line is sometimes called a TUNED line. The line also may be used for a resonant or tuned circuit.

A resonant line is sometimes said to be resonant at an applied frequency. This means that at one frequency the line acts as a resonant circuit. It may act either as a high-resistive circuit (parallel resonant) or as a low-resistive circuit (series resonant). The line may be made to act in this manner by either open- or short-circuiting it at the output end and cutting it to some multiple of a quarter-wavelength.

At the points of voltage maxima and minima on a short-circuited or open-circuited line, the line impedance is resistive. On a short-circuited line, each point at an odd number of quarter-wavelengths from the receiving end has a high impedance (figure 3-31, view A). If the frequency of the applied voltage to the line is varied, this impedance decreases as the effective length of the line changes. This variation is exactly the same as the change in the impedance of a parallel-resonant circuit when the applied frequency is varied.

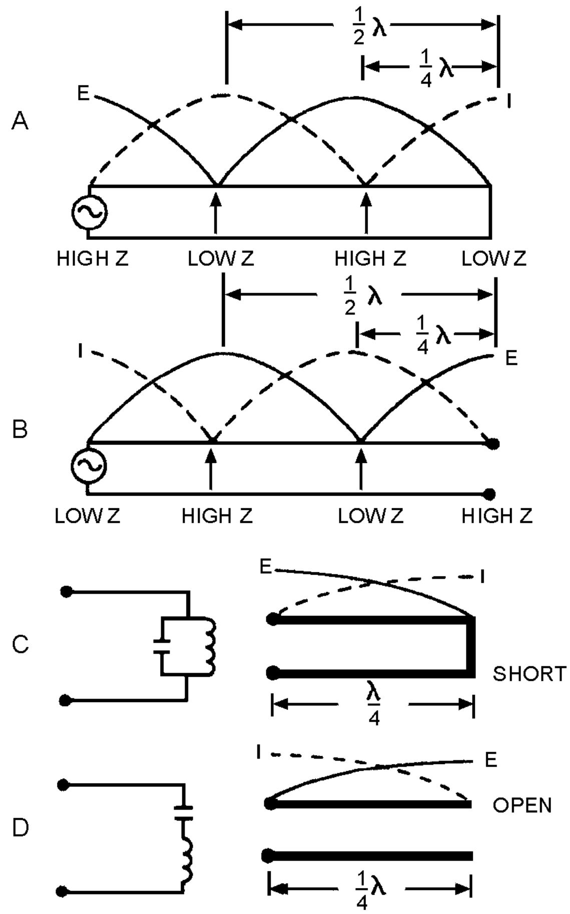


Figure 3-31.—Sending-end impedance of various lengths and terminations.

At all even numbered quarter-wavelength points from the short circuit, the impedance is extremely low. When the frequency of the voltage applied to the line is varied, the impedance at these points increases just as the impedance of a series-resonant circuit varies when the frequency applied to it is changed. The same is true for an open-ended line (figure 3-31, view B) except that the points of high and low impedance are reversed.

At this point let us review some of the characteristics of resonant circuits so we can see how resonant line sections may be used in place of LC circuits.

A PARALLEL-RESONANT circuit has the following characteristics:

- At resonance the impedance appears as a very high resistance. A loss-free circuit has infinite impedance (an open circuit). Other than at resonance, the impedance decreases rapidly.
- If the circuit is resonant at a point above the generator frequency (the generator frequency is too low), more current flows through the coil than through the capacitor. This happens because X_L decreases with a decrease in frequency but X_C increases.

A SERIES-RESONANT circuit has these characteristics:

- At resonance the impedance appears as a very low resistance. A loss-free circuit has zero impedance (a short circuit). Other than at resonance the impedance increases rapidly.

- If the circuit is resonant at a point above the generator frequency (the generator frequency is too low), then X_C is larger than X_L and the circuit acts capacitively.

- If the circuit is resonant at a point below the generator frequency (the generator frequency is too high), then X_L is larger than X_C and the circuit acts inductively.

Since the impedance a generator sees at the quarter-wave point in a shorted line is that of a parallel-resonant circuit, a shorted quarter-wave- length of line may be used as a parallel-resonant circuit (figure 3-31, view C). An open quarter-wavelength of line may be used as a series-resonant circuit (view D). The Q of such a resonant line is much greater than can be obtained with lumped capacitance and inductance.

Impedance for Various Lengths of Open Lines

In figure 3-32, the impedance (Z) the generator sees for various lengths of line is shown at the top. The curves above the letters of various heights show the relative value of the impedances presented to the generator for the various line lengths. The circuit symbols indicate the equivalent electrical circuits for the transmission lines at each particular length. The standing waves of voltage and current are shown on each length of line.

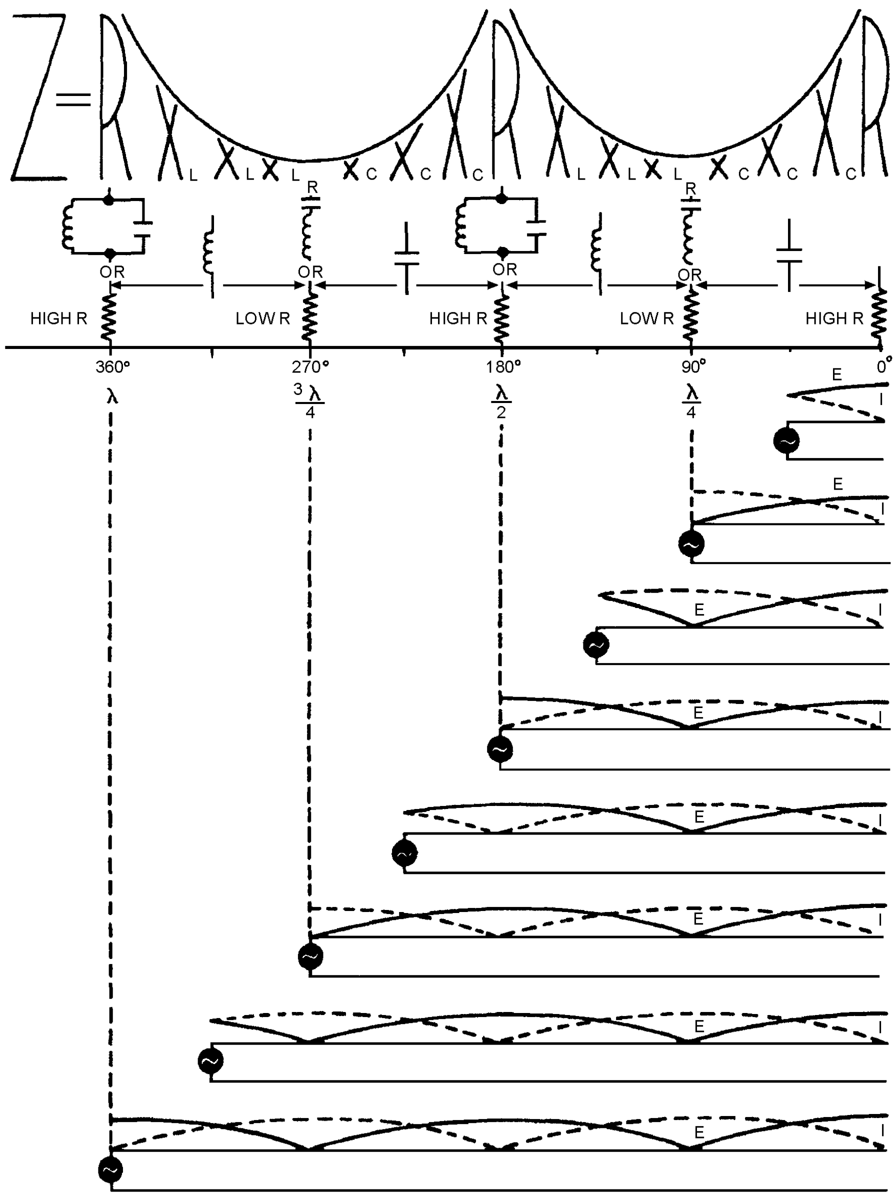


Figure 3-32.—Voltage, current, and impedance on open line.

At all odd quarter-wave points (1/4λ, 3/4λ, etc.), the voltage is minimum, the current is maximum, and the impedance is minimum. Thus, at all odd quarter-wave points, the open-ended transmission line acts as a series-resonant circuit. The impedance is equivalent to a very low resistance, prevented from being zero only by small circuit losses.

At all even quarter-wave points (1/2λ, 1λ, 3/2λ, etc.), the voltage is maximum, the current is minimum, and the impedance is maximum. Comparison of the line with an LC resonant circuit shows that at an even number of quarter-wavelengths, an open line acts as a parallel-resonant circuit. The impedance is therefore an extremely high resistance.

In addition, resonant open lines may also act as nearly pure capacitances or inductances. The illustration shows that an open line less than a quarter-wavelength long acts as a capacitance. Also, it acts as an inductance from 1/4 to 1/2 wavelength, as a capacitance from 1/2 to 3/4 wavelength, and as an inductance from 3/4 to 1 wavelength, etc. A number of open transmission lines, with their equivalent circuits, are shown in the illustration.

Impedance of Various Lengths of Shorted Lines

Follow figure 3-33 as we study the shorted line. At the odd quarter-wavelength points, the voltage is high, the current is low, and the impedance is high. Since these conditions are similar to those found in a parallel-resonant circuit, the shorted transmission line acts as a parallel-resonant circuit at these lengths.

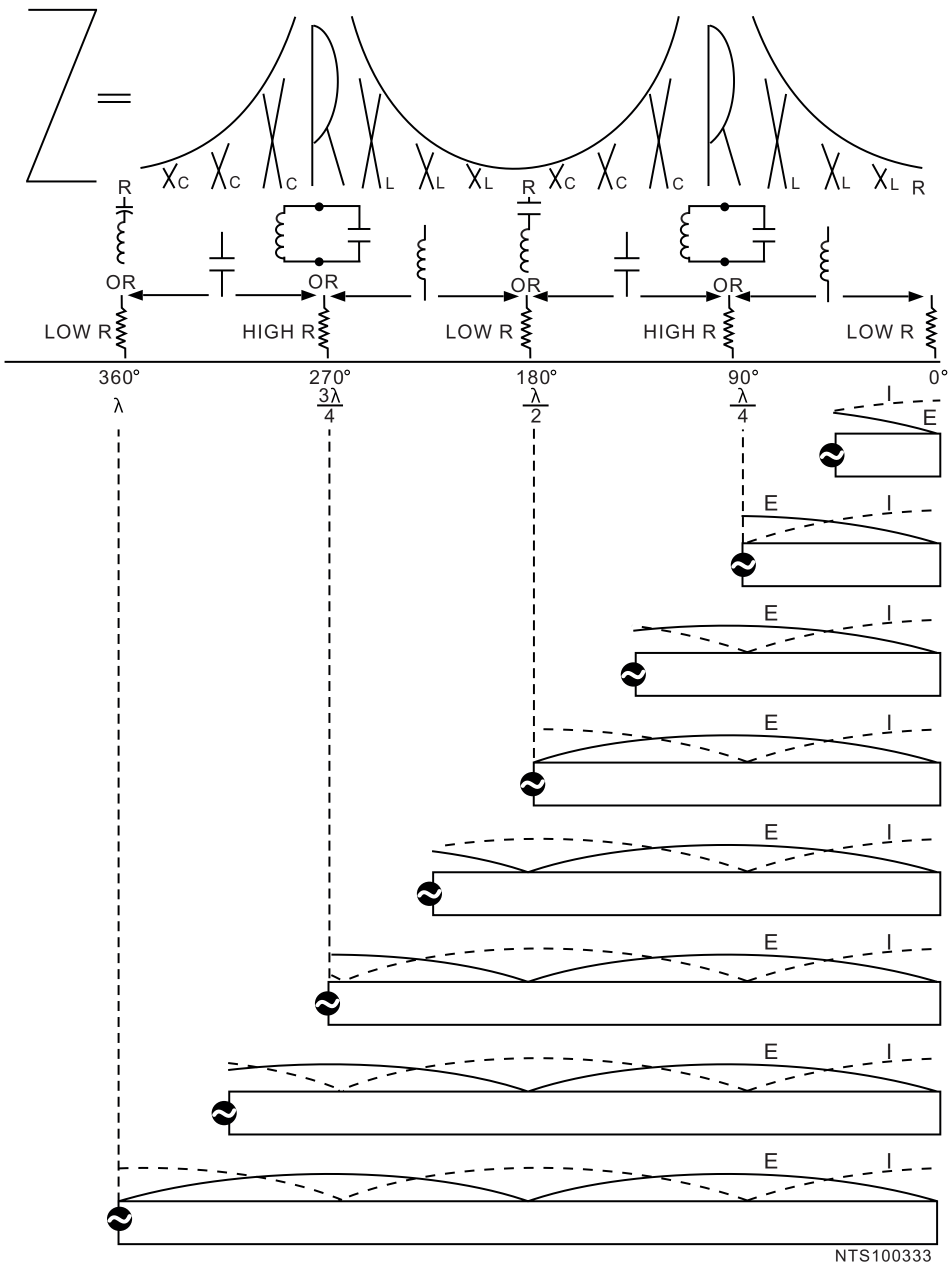


Figure 3-33.—Voltage, current, and impedance on shorted line.

At the even quarter-wave points voltage is minimum, current is maximum, and impedance is minimum. Since these characteristics are similar to those of a series-resonant LC circuit, a shorted transmission line whose length is an even number of quarter-wavelengths acts as a series-resonant circuit.

Resonant shorted lines, like open-end lines, also may act as pure capacitances or inductances. The illustration shows that a shorted line less than 1/4 wavelength long acts as an inductance. A shorted line with a length of from 1/4 to 1/2 wavelength acts as a capacitance. From 1/2 to 3/4 wavelength, the line acts as an inductance; and from 3/4 to 1 wavelength, it acts as a capacitance, and so on. The equivalent circuits of shorted lines of various lengths are shown in the illustration. Thus, properly chosen line segments may be used as parallel-resonant, series-resonant, inductive, or capacitive circuits.

STANDING WAVES ON A TRANSMISSION LINE

There is a large variety of terminations for rf lines. Each type of termination has a characteristic effect on the standing waves on the line. From the nature of the standing waves, you can determine the type of termination that produces the waves.

TERMINATION IN Z_0

Termination in Z_0 (characteristic impedance) will cause a constant reading on an ac meter when it is moved along the length of the line. As illustrated in figure 3-34, view A, the curve, provided there are no losses in the line, will be a straight line. If there are losses in the line, the amplitude of the voltage and current will diminish as they move down the line (view B). The losses are due to dc resistance in the line itself.

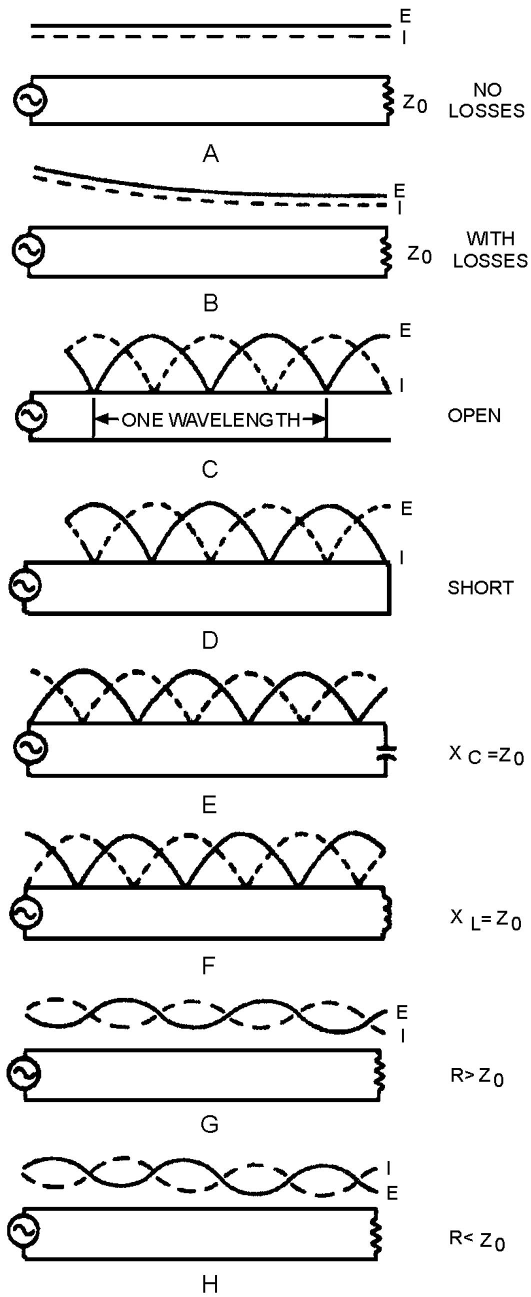


Figure 3-34.—Effects of various terminations on standing waves.

TERMINATION IN AN OPEN CIRCUIT

In an open-circuited rf line (figure 3-34, view C), the voltage is maximum at the end, but the current is minimum. The distance between two adjacent zero current points is $1/2\lambda$, and the distance between alternate zero current points is 1λ. The voltage is zero at a distance of $1/4\lambda$ from the end of the line. This is true at any frequency. A voltage peak occurs at the end of the line, at $1/2\lambda$ from the end, and at each $1/2\lambda$ thereafter.

TERMINATION IN A SHORT CIRCUIT

On the line terminated in a short circuit, shown in figure 3-34, view D, the voltage is zero at the end and maximum at 1/4λ from the end. The current is maximum at the end, zero at 1/4λ from the end, and alternately maximum and zero every 1/4λ thereafter.

TERMINATION IN CAPACITANCE

When a line is terminated in capacitance, the capacitor does not absorb energy, but returns all of the energy to the circuit. This means there is 100 percent reflection. The current and voltage relationships are somewhat more involved than in previous types of termination. For this explanation, assume that the capacitive reactance is equal to the Z_0 of the line. Current and voltage are in phase when they arrive at the end of the line, but in flowing through the capacitor and the characteristic impedance (Z_0) connected in series, they shift in phase relationship. Current and voltage arrive in phase and leave out of phase. This results in the standing-wave configuration shown in figure 3-34, view E. The standing wave of voltage is minimum at a distance of exactly 1/8λ from the end. If the capacitive reactance is greater than Z_0 (smaller capacitance), the termination looks more like an open circuit; the voltage minimum moves away from the end. If the capacitive reactance is smaller than Z_0, the minimum moves toward the end.

TERMINATION IN INDUCTANCE

When the line is terminated in an inductance, both the current and voltage shift in phase as they arrive at the end of the line. When X_L is equal to Z_0, the resulting standing waves are as shown in figure 3-34, view F. The current minimum is located 1/8λ from the end of the line. When the inductive reactance is increased, the standing waves appear closer to the end. When the inductive reactance is decreased, the standing waves move away from the end of the line.

TERMINATION IN A RESISTANCE NOT EQUAL TO THE CHARACTERISTIC IMPEDANCE (Z_0)

Whenever the termination is not equal to Z_0, reflections occur on the line. For example, if the terminating element contains resistance, it absorbs some energy, but if the resistive element does not equal the Z_0 of the line, some of the energy is reflected. The amount of voltage reflected may be found by using the equation:

$$E_R = E_i\left(\frac{R_L - Z_0}{R_L + Z_0}\right)$$

Where:

E_R = the reflected voltage

E_i = the incident voltage

R_R = the terminating resistance

Z_0= the characteristic impedance of the line

If you try different values of R_L in the preceding equation, you will find that the reflected voltage is equal to the incident voltage only when R_L equals 0 or is infinitely large. When R_L equals Z_0, no reflected voltage occurs. When R_L is greater than Z_0, E_R is positive, but less than E_i. As R_L increases and approaches an infinite value, E_R increases and approaches E_i in value. When R_L is smaller than Z_0, E_R has a negative value. This means that the reflected voltage is of opposite polarity to the incident wave at the termination of the line. As R_L approaches zero, E_R approaches E_i in value. The smaller the value of E_R, the smaller is the peak amplitude of the standing waves and the higher are the minimum values.

TERMINATION IN A RESISTANCE GREATER THAN Z_0

When R_L is greater than Z_0, the end of the line is somewhat like an open circuit; that is, standing waves appear on the line. The voltage maximum appears at the end of the line and also at half-wave intervals back from the end. The current is minimum (not zero) at the end of the line and maximum at the odd quarter-wave points. Since part of the power in the incident wave is consumed by the load resistance, the minimum voltage and current are less than for the standing waves on an open-ended line. Figure 3-34, view G, illustrates the standing waves for this condition.

TERMINATION IN A RESISTANCE LESS THAN Z_0

When R_L is less than Z_0, the termination appears as a short circuit. The standing waves are shown in figure 3-34, view H. Notice that the line terminates in a current LOOP (peak) and a voltage NODE (minimum). The values of the maximum and minimum voltage and current approach those for a shorted line as the value of R_L approaches zero.

A line does not have to be any particular length to produce standing waves; however, it cannot be an infinite line. Voltage and current must be reflected to produce standing waves. For reflection to occur, a line must not be terminated in its characteristic impedance. Reflection occurs on lines terminated in opens, shorts, capacitances, and inductances, because no energy is absorbed by the load. If the line is terminated in a resistance not equal to the characteristic impedance of the line, some energy will be absorbed and the rest will be reflected.

The voltage and current relationships for open-ended and shorted lines are opposite to each other, as shown in figure 3-34, views C and D. The points of maximum and minimum voltage and current are determined from the output end of the line, because reflection always begins at that end.

Q26. A nonresonant line is a line that has no standing waves of current and voltage on it and is considered to be flat. Why is this true?

Q27. On an open line, the voltage and impedance are maximum at what points on the line?

STANDING-WAVE RATIO

The measurement of standing waves on a transmission line yields information about equipment operating conditions. Maximum power is absorbed by the load when $Z_L = Z_0$. If a line has no standing waves, the termination for that line is correct and maximum power transfer takes place.

You have probably noticed that the variation of standing waves shows how near the rf line is to being terminated in Z_0. A wide variation in voltage along the length means a termination far from Z_0. A small variation means termination near Z_0. Therefore, the ratio of the maximum to the minimum is a measure of the perfection of the termination of a line. This ratio is called the STANDING-WAVE RATIO (swr) and is always expressed in whole numbers. For example, a ratio of 1:1 describes a line terminated in its characteristic impedance (Z_0).

Voltage Standing-Wave Ratio

The ratio of maximum voltage to minimum voltage on a line is called the VOLTAGE STANDING-WAVE RATIO (vswr). Therefore:

$$vswr = \frac{E_{max}}{E_{min}}$$

The vertical lines in the formula indicate that the enclosed quantities are absolute and that the two values are taken without regard to polarity. Depending on the nature of the standing waves, the numerical value of vswr ranges from a value of 1 ($Z_L = Z_0$, no standing waves) to an infinite value for theoretically complete reflection. Since there is always a small loss on a line, the minimum voltage is never zero and the vswr is always some finite value. However, if the vswr is to be a useful quantity, the power losses along the line must be small in comparison to the transmitted power.

Power Standing-Wave Ratio

The square of the voltage standing-wave ratio is called the POWER STANDING-WAVE RATIO (pswr). Therefore:

$$pswr = \frac{P_{max}}{P_{min}}$$

This ratio is useful because the instruments used to detect standing waves react to the square of the voltage. Since power is proportional to the square of the voltage, the ratio of the square of the maximum and minimum voltages is called the power standing-wave ratio. In a sense, the name is misleading because the power along a transmission line does not vary.

Current Standing-Wave Ratio

The ratio of maximum to minimum current along a transmission line is called CURRENT STANDING-WAVE RATIO (iswr). Therefore:

$$iswr = \frac{I_{max}}{I_{min}}$$

This ratio is the same as that for voltages. It can be used where measurements are made with loops that sample the magnetic field along a line. It gives the same results as vswr measurements.

Q28. At what point on an open-circuited rf line do voltage peaks occur?

Q29. What is the square of the voltage standing-wave ratio called?

Q30. What does vswr measure?

SUMMARY

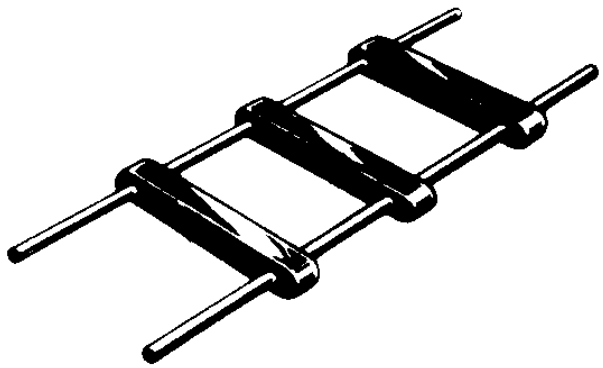
This chapter has presented information on the characteristics of transmission lines. The information that follows summarizes the important points of this chapter.

TRANSMISSION LINES are devices for guiding electrical energy from one point to another.

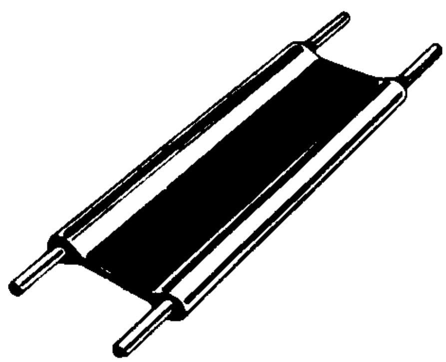
INPUT IMPEDANCE is the ratio of voltage to current at the input end of a transmission line.

OUTPUT IMPEDANCE is the ratio of voltage to current at the output end of the line.

TWO-WIRE OPEN LINES are parallel lines and have uses such as power lines, rural telephone lines, and telegraph lines. This type of line has high radiation losses and is subject to noise pickup.

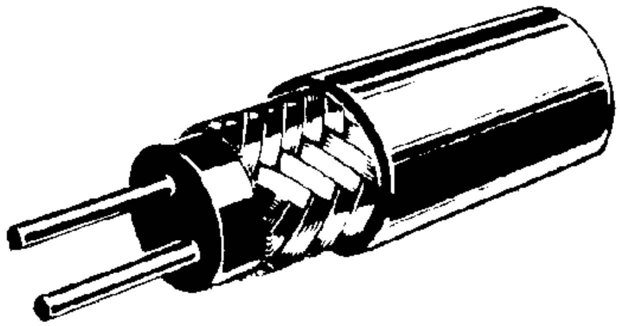


TWIN LEAD has parallel lines and is most often used to connect televisions to their antennas.

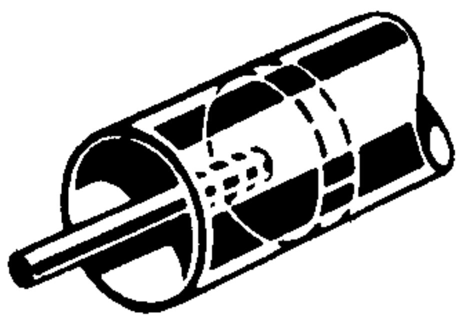


A **TWISTED PAIR** consists of two insulated wires twisted together. This line has high insulation loss.

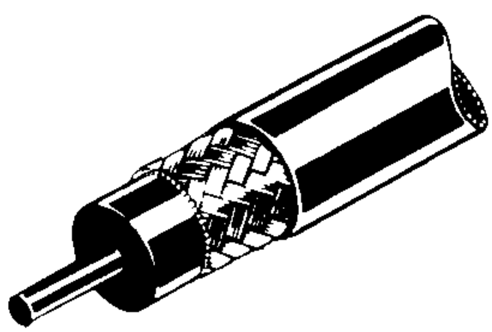
A **SHIELDED PAIR** has parallel conductors separated by a solid dielectric and surrounded by copper braided tubing. The conductors are balanced to ground.



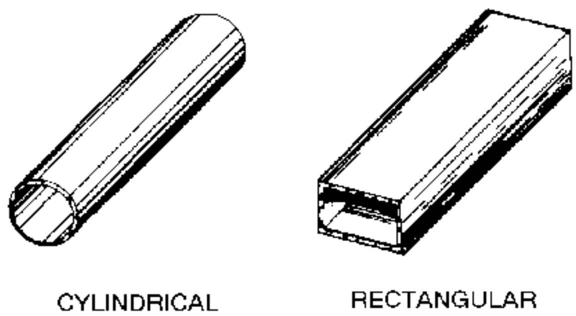
RIGID COAXIAL LINE contains two concentric conductors insulated from each other by spacers. Some rigid coaxial lines are pressurized with an inert gas to prevent moisture from entering. High-frequency losses are less than with other lines.



FLEXIBLE COAXIAL LINES consist of a flexible inner conductor and a concentric outer conductor of metal braid. The two are separated by a continuous insulating material.



WAVEGUIDES are hollow metal tubes used to transfer energy from one point to another. The energy travels slower in a waveguide than in free space.



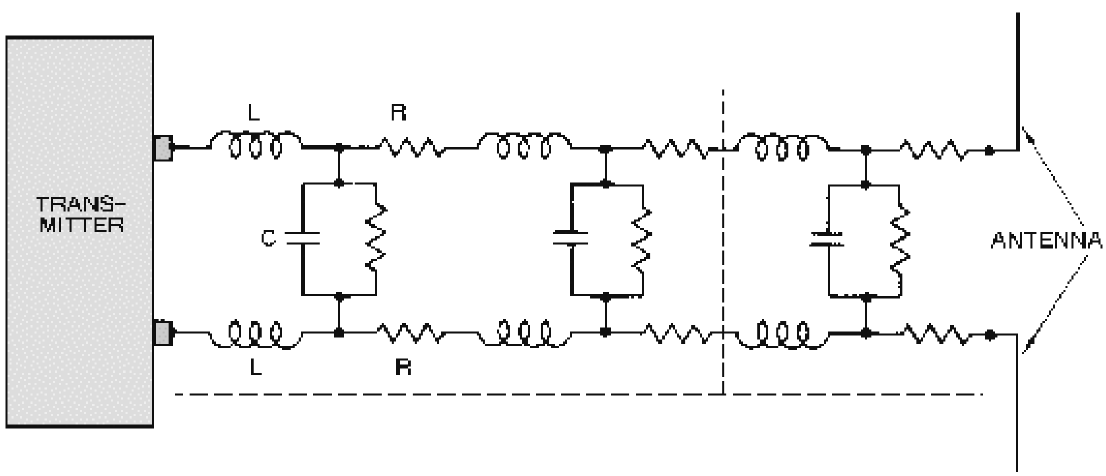
COPPER LOSSES can result from power (I^2R) loss, in the form of heat, or skin effect. These losses decrease the conductivity of a line.

DIELECTRIC LOSSES are caused by the heating of the dielectric material between conductors, taking power from the source.

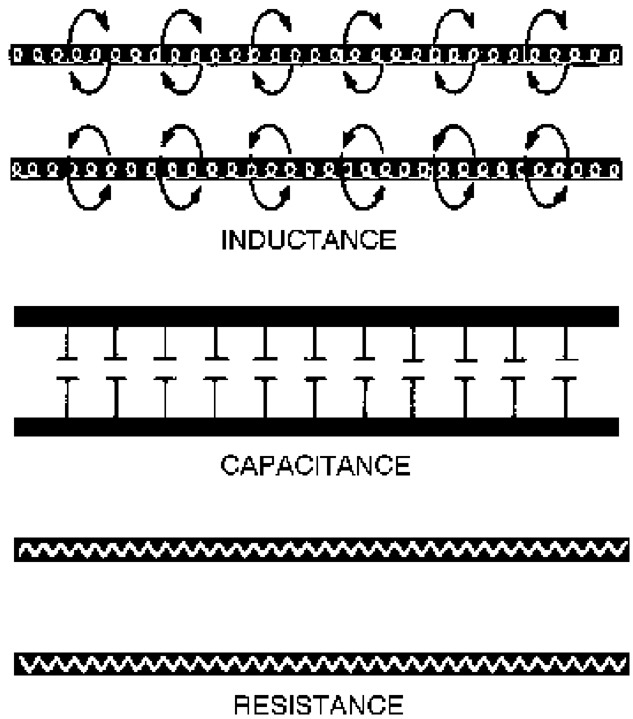
RADIATION and **INDUCTION LOSSES** are caused by part of the electromagnetic fields of a conductor being dissipated into space or nearby objects.

A transmission line is either electrically **LONG** or **SHORT** if its physical length is not equal to 1/4λ for the frequency it is to carry.

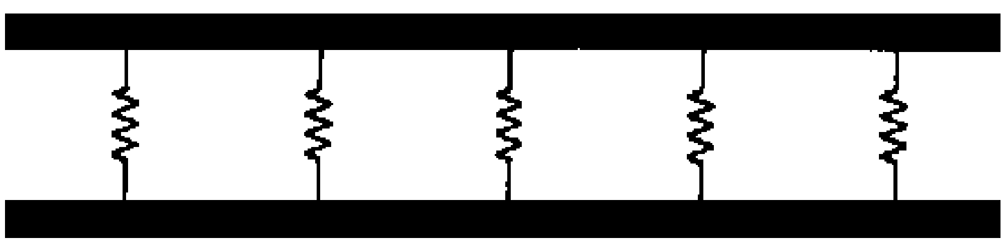
LUMPED CONSTANTS are theoretical properties (inductance, resistance, and capacitance) of a transmission line that are lumped into a single component.



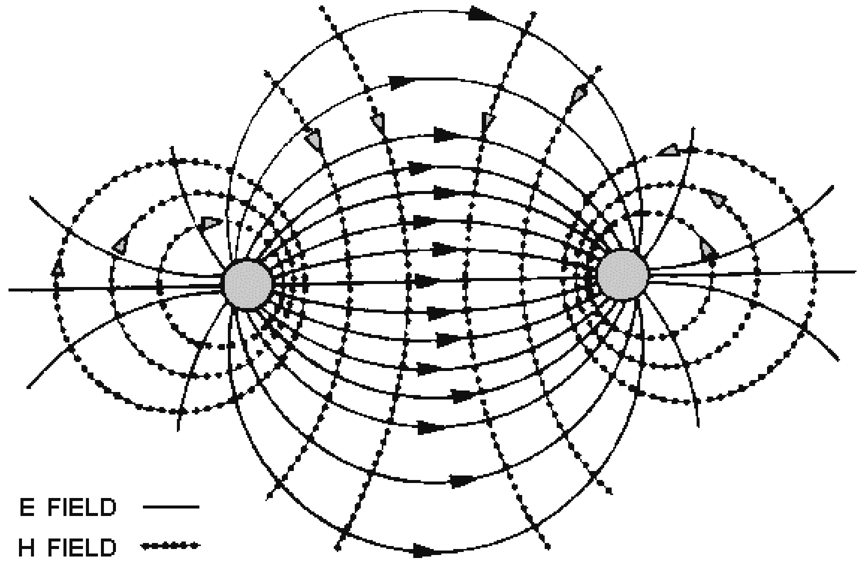
DISTRIBUTED CONSTANTS are constants of inductance, capacitance and resistance that are distributed along the transmission line.



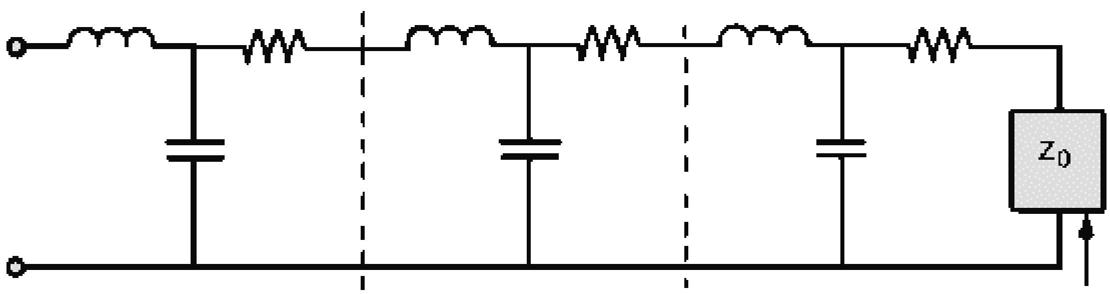
LEAKAGE CURRENT flows between the wires of a transmission line through the dielectric. The dielectric acts as a resistor.



An **ELECTROMAGNETIC FIELD** exists along transmission line when current flows through it.



CHARACTERISTIC IMPEDANCE, Z_0, is the ratio of E to I at every point along the line. For maximum transfer of electrical power, the characteristic impedance and load impedance must be matched.



The **VELOCITY** at which a wave travels over a given length of transmission line can be found by using the formula:

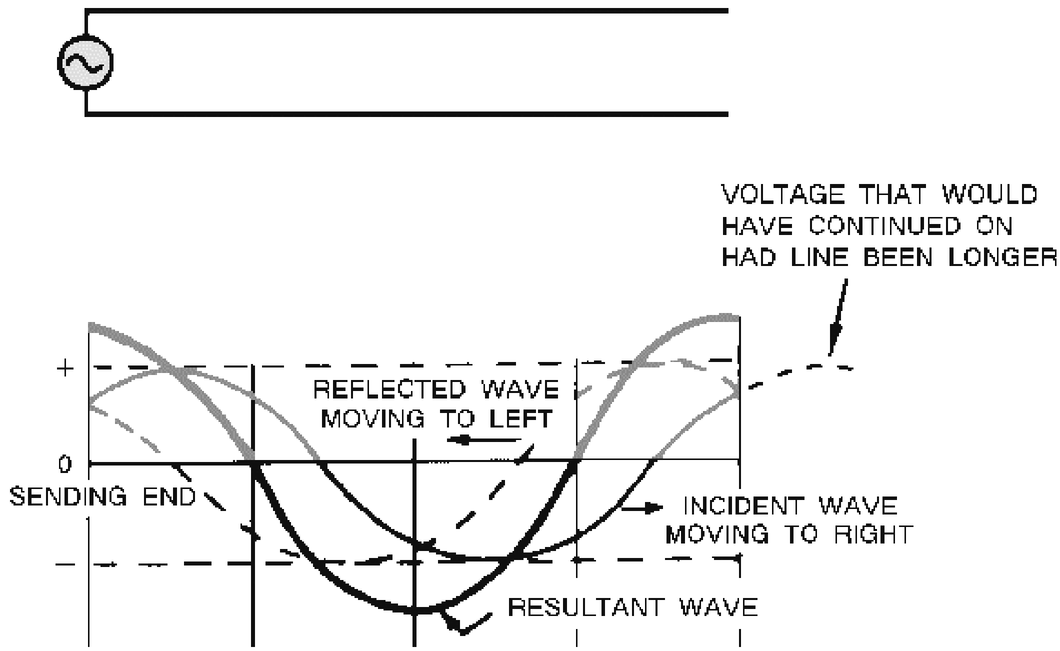
$$V = \frac{D}{\sqrt{LC}}$$

A transmission line that is not terminated in its characteristic impedance is said to be **FINITE**.

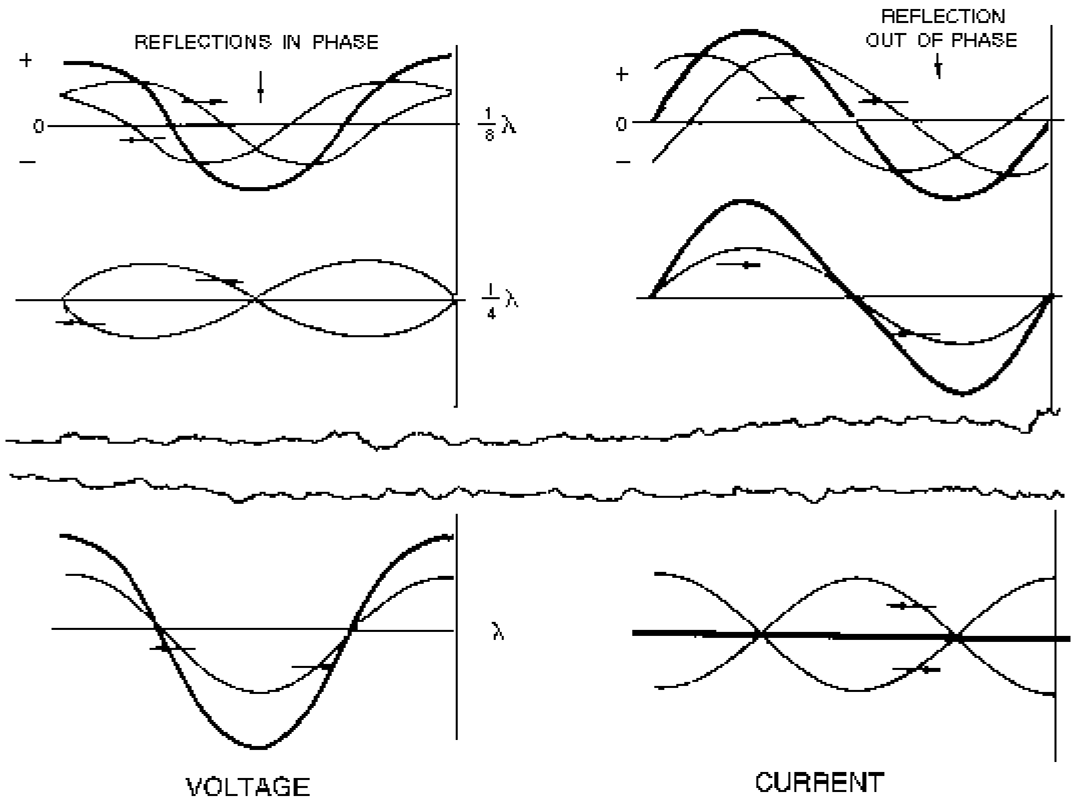
When dc is applied to an **OPEN-ENDED** line, the voltage is reflected back from the open end without any change in polarity, amplitude, or shape. Current is reflected back with the same amplitude and shape but with opposite polarity.

When dc is applied to a **SHORT-CIRCUITED** line, the current is reflected back with the same amplitude, and polarity. The voltage is reflected back with the same amplitude but with opposite polarity.

When ac is applied to an **OPEN-END** line, voltage is always reflected back in phase with the incident wave and current is reflected back out of phase.



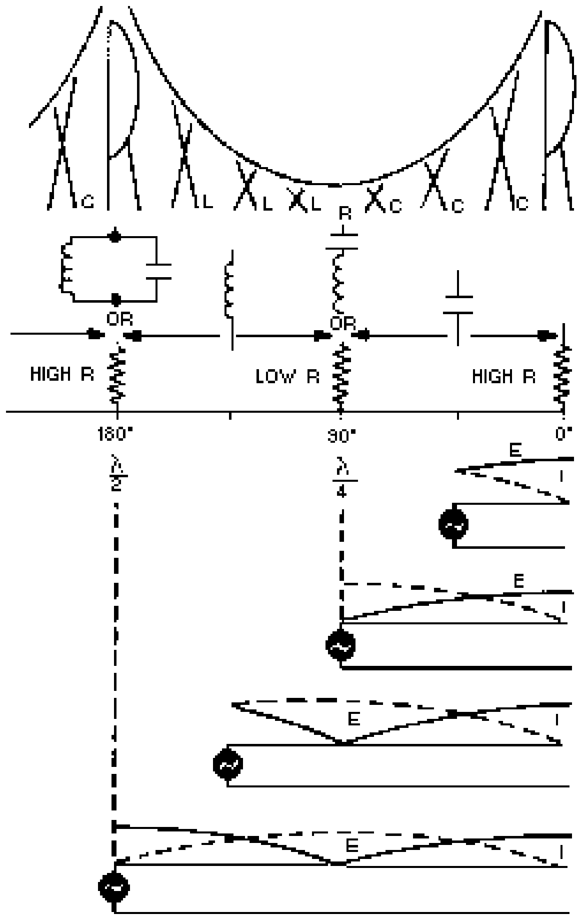
When ac is applied to a **SHORT-CIRCUITED** line, voltage is reflected in opposite phase, while current is reflected in phase.



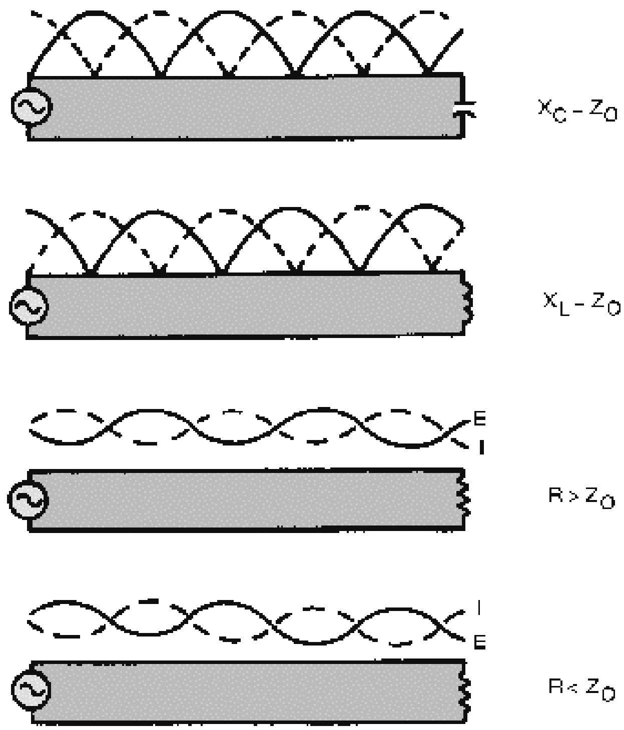
A **NONRESONANT** line has **NO STANDING WAVES** of current and voltage and is either infinitely long or terminated in its characteristic impedance.

A **RESONANT** line has **STANDING WAVES** of current and voltage and is of finite length and is NOT terminated in its characteristic impedance.

On an open-ended resonant line, and at all odd 1/4λ points, the voltage is minimum, the current is maximum, and the impedance is minimum. At all even 1/4λ points, the voltage is maximum, the current is minimum and the impedance is maximum.



There are a variety of **TERMINATIONS** for rf lines. Each termination has an effect on the standing waves on the line.



A transmission line can be terminated in its characteristic impedance as an open- or short-circuit, or in capacitance or inductance.

Whenever the termination on a transmission line is NOT EQUAL TO Z_0, there are reflections on the line. The amount of voltage reflected may be found by using the equation:

$$E_R = E_i\left(\frac{R_L - Z_0}{R_L + Z_0}\right)$$

When the termination on a transmission line EQUALS Z_0, there is NO reflected voltage.

The measurement of standing waves on a transmission line yields information about operating conditions. If there are NO standing waves, the termination for that line is correct and maximum power transfer takes place.

The **STANDING WAVE RATIO** is the measurement of maximum voltage (current) to minimum voltage (current) on a transmission line and measures the perfection of the termination of the line. A ratio of 1:1 describes a line terminated in its characteristic impedance.

ANSWERS TO QUESTIONS Q1. THROUGH Q30.

A1. *Transmission line.*

A2. *Input end, generator end, transmitter end, sending end, and source.*

A3. *Output end, receiving end, load end and sink.*

A4. *Parallel two-wire, twisted pair, shielded pair, coaxial line and waveguide.*

A5. *Power lines, rural telephone lines, and telegraph lines.*

A6. *High radiation losses and noise pickup.*

A7. *Twin lead.*

A8. *The conductors are balanced to ground.*

A9. *Air coaxial (rigid) and solid coaxial (flexible).*

A10. *The ability to minimize radiation losses.*

A11. *Expensive to construct, must be kept dry, and high frequency losses limit the practical length of the line.*

A12. *Cylindrical and rectangular.*

A13. *Copper, dielectric, and radiation.*

A14. *Copper losses.*

A15. *Dielectric losses.*

A16. $\lambda = 20$ *meters.*

A17. *(1) Type of line used, (2) dielectric in the line, and (3) length of line.*

A18. *Inductance is expressed in microhenrys per unit length, capacitance is expressed in picofarads per unit length, and resistance is expressed in ohms per unit length.*

A19. *The small amount of current that flows through the dielectric between two wires of a transmission line and is expressed in micromhos per unit length.*

A20. *When the characteristic impedance of the transmission line and the load impedance are equal.*

A21. Z_0 *and it is the ratio of E to I at every point along the line.*

A22. *Between 50 and 600 ohms.*

A23. *Incident waves from generator to load. Reflected waves from load back to generator.*

A24. *2 and 6 have zero resultant wave and they indicate that the incident and reflected waves are 180 degrees out of phase at all parts.*

A25. *One-fourth the distance from each end of the line.*

A26. The load impedance of such a line is equal to Z_0.

A27. Even quarter-wave points (1/2λ, 1λ, 3/2λ, etc.).

A28. At 1/2 wavelength from the end and at every 1/2 wavelength along the line.

A29. Power standing-wave ratio (pswr).

A30. The existence of voltage variations on a line.

CHAPTER 4

ANTENNAS

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. State the basic principles of antenna radiation and list the parts of an antenna.
2. Explain current and voltage distribution on an antenna.
3. Describe how electromagnetic energy is radiated from an antenna.
4. Explain polarization, gain, and radiation resistance characteristics of an antenna.
5. Describe the theory of operation of half- wave and quarter-wave antennas.
6. List the various array antennas.
7. Describe the directional array antennas presented and explain the basic operation of each.
8. Identify various special antennas presented, such as long-wire, V, rhombic, turnstile, ground-plane, and corner-reflector; describe the operation of each.
9. List safety precautions when working aloft and around antennas.

INTRODUCTION

If you had been around in the early days of electronics, you would have considered an ANTENNA (AERIAL) to be little more than a piece of wire strung between two trees or upright poles. In those days, technicians assumed that longer antennas automatically provided better reception than shorter antennas. They also believed that a mysterious MEDIUM filled all space, and that an antenna used this medium to send and receive its energy. These two assumptions have since been discarded. Modern antennas have evolved to the point that highly directional, specially designed antennas are used to relay worldwide communications in space through the use of satellites and Earth station antennas (fig. 4-1). Present transmission theories are based on the assumption that space itself is the only medium necessary to propagate (transmit) radio energy.

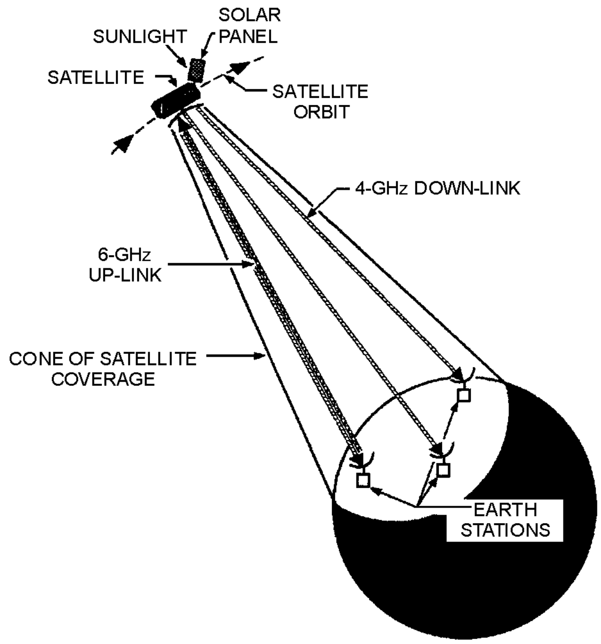


Figure 4-1.—Satellite/earth station communications system.

A tremendous amount of knowledge and information has been gained about the design of antennas and radio-wave propagation. Still, many old-time technicians will tell you that when it comes to designing the length of an antenna, the best procedure is to perform all calculations and try out the antenna. If it doesn't work right, use a cut-and-try method until it does. Fortunately, enough information has been collected over the last few decades that it is now possible to predict the behavior of antennas. This chapter will discuss and explain the basic design and operation of antennas.

PRINCIPLES OF ANTENNA RADIATION

After an rf signal has been generated in a transmitter, some means must be used to radiate this signal through space to a receiver. The device that does this job is the antenna. The transmitter signal energy is sent into space by a TRANSMITTING ANTENNA; the rf signal is then picked up from space by a RECEIVING ANTENNA.

The rf energy is transmitted into space in the form of an electromagnetic field. As the traveling electromagnetic field arrives at the receiving antenna, a voltage is induced into the antenna (a conductor). The rf voltages induced into the receiving antenna are then passed into the receiver and converted back into the transmitted rf information.

The design of the antenna system is very important in a transmitting station. The antenna must be able to radiate efficiently so the power supplied by the transmitter is not wasted. An efficient transmitting antenna must have exact dimensions. The dimensions are determined by the transmitting frequencies. The dimensions of the receiving antenna are not critical for relatively low radio frequencies. However, as the frequency of the signal being received increases, the design and installation of the receiving antenna become more critical. An example of this is a television receiving antenna. If you raise it a few more inches from the ground or give a slight turn in direction, you can change a snowy blur into a clear picture.

The conventional antenna is a conductor, or system of conductors, that radiates or intercepts electromagnetic wave energy. An ideal antenna has a definite length and a uniform diameter, and is completely isolated in space. However, this ideal antenna is not realistic. Many factors make the design of an antenna for a communications system a more complex problem than you would expect. These factors include the height of the radiator above the earth, the conductivity of the earth below it, and the shape and dimensions of the antenna. All of these factors affect the radiated-field pattern of the antenna in space. Another problem in antenna design is that the radiation pattern of the antenna must be directed between certain angles in a horizontal or vertical plane, or both.

Most practical transmitting antennas are divided into two basic classifications, HERTZ (half-wave) ANTENNAS and MARCONI (quarter-wave) ANTENNAS. Hertz antennas are generally installed some distance above the ground and are positioned to radiate either vertically or horizontally. Marconi antennas operate with one end grounded and are mounted perpendicular to the Earth or to a surface acting as a ground. Hertz antennas are generally used for frequencies above 2 megahertz. Marconi antennas are used for frequencies below 2 megahertz and may be used at higher frequencies in certain applications.

A complete antenna system consists of three parts: (1) The COUPLING DEVICE, (2) the FEEDER, and (3) the ANTENNA, as shown in figure 4-2. The coupling device (coupling coil) connects the transmitter to the feeder. The feeder is a transmission line that carries energy to the antenna. The antenna radiates this energy into space.

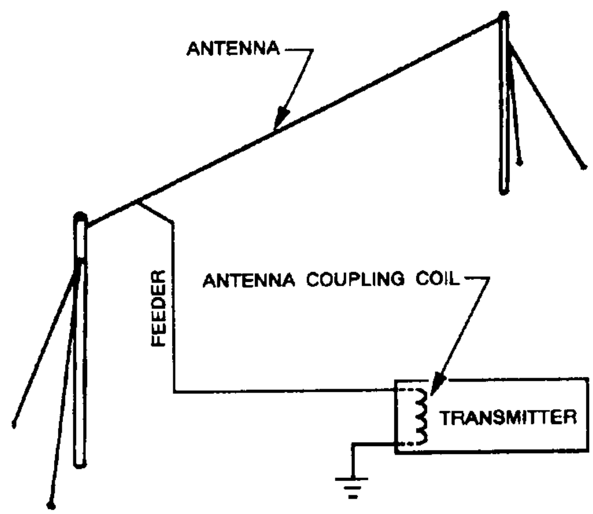


Figure 4-2.—Typical antenna system.

The factors that determine the type, size, and shape of the antenna are (1) the frequency of operation of the transmitter, (2) the amount of power to be radiated, and (3) the general direction of the receiving set. Typical antennas are shown in figure 4-3.

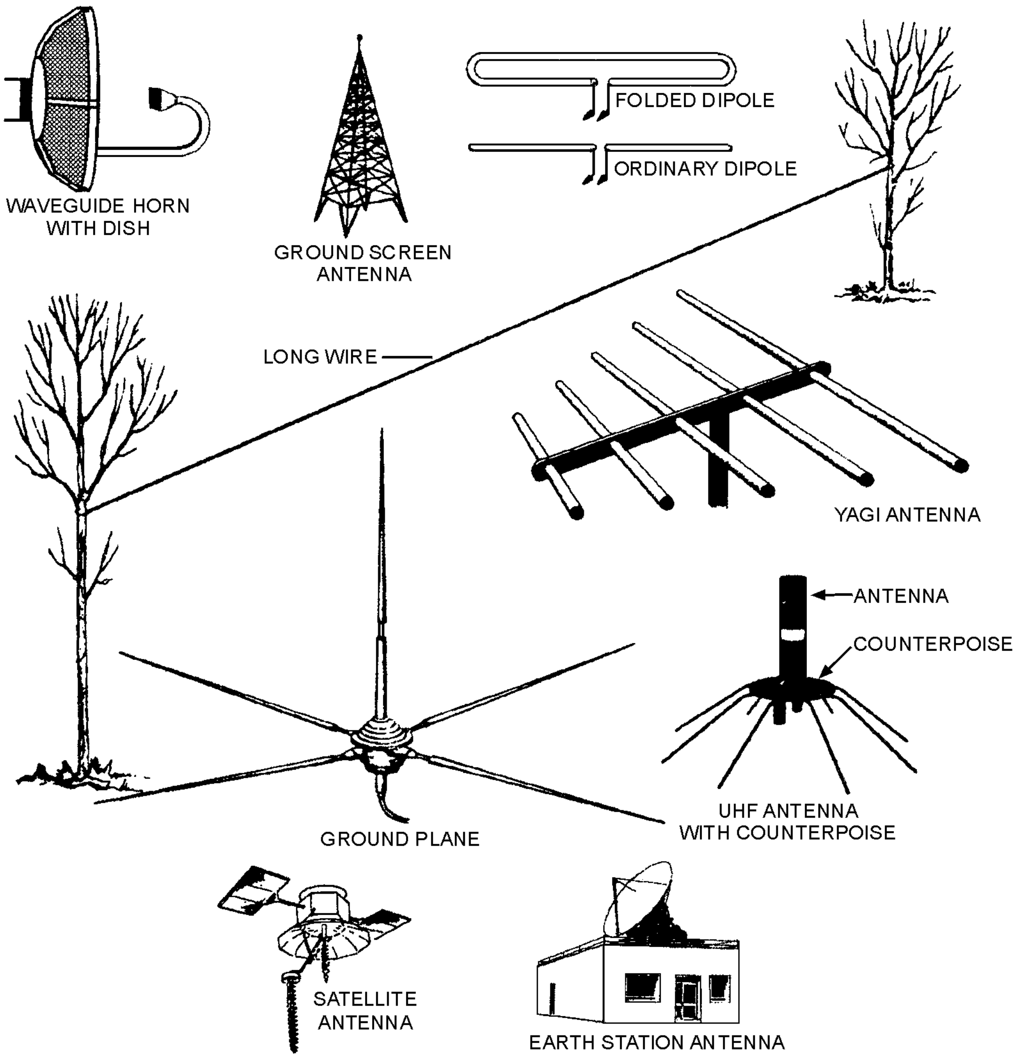


Figure 4-3.—Typical antennas.

CURRENT AND VOLTAGE DISTRIBUTION ON AN ANTENNA

A current flowing in a wire whose length is properly related to the rf produces an electro magnetic field. This field is radiated from the wire and is set free in space. We will discuss how these waves are set free later in this chapter. Remember, the principles of radiation of electromagnetic energy are based on two laws:

1. A MOVING ELECTRIC FIELD CREATES A MAGNETIC (H) FIELD.

2. A MOVING MAGNETIC FIELD CREATES AN ELECTRIC (E) FIELD.

In space, these two fields will be in phase and perpendicular to each other at any given time. Although a conductor is usually considered present when a moving electric or magnetic field is mentioned, the laws that govern these fields say nothing about a conductor. Therefore, these laws hold true whether a conductor is present or not.

Figure 4-4 shows the current and voltage distribution on a half-wave (Hertz) antenna. In view A, a piece of wire is cut in half and attached to the terminals of a high-frequency ac generator. The frequency of the generator is set so that each half of the wire is 1/4 wavelength of the output. The result is a common type of antenna known as a DIPOLE.

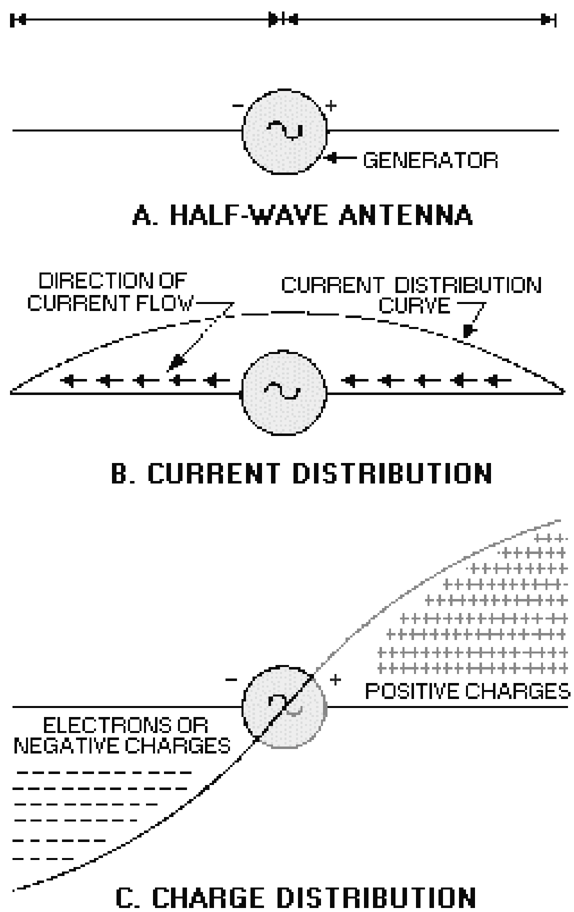


Figure 4-4.—Current and voltage distribution on an antenna.

At a given time the right side of the generator is positive and the left side negative. Remember that like charges repel. Because of this, electrons will flow away from the negative terminal as far as possible, but will be attracted to the positive terminal. View B shows the direction and distribution of electron flow. The distribution curve shows that most current flows in the center and none flows at the ends. The current distribution over the antenna will always be the same no matter how much or how little current is flowing. However, current at any given point on the antenna will vary directly with the amount of voltage developed by the generator.

One-quarter cycle after electrons have begun to flow, the generator will develop its maximum voltage and the current will decrease to 0. At that time the condition shown in view C will exist. No current will be flowing, but a maximum number of electrons will be at the left end of the line and a minimum number at the right end. The charge distribution view C along the wire will vary as the voltage of the generator varies. Therefore, you may draw the following conclusions:

1. A current flows in the antenna with an amplitude that varies with the generator voltage.
2. A sinusoidal distribution of charge exists on the antenna. Every 1/2 cycle, the charges reverse polarity.
3. The sinusoidal variation in charge magnitude lags the sinusoidal variation in current by 1/4 cycle.

Q1. What are the two basic classifications of antennas?

Q2. What are the three parts of a complete antenna system?

Q3. What three factors determine the type, size, and shape of an antenna?

RADIATION OF ELECTROMAGNETIC ENERGY

The electromagnetic radiation from an antenna is made up of two components, the E field and the H field. We discussed these fields in chapters 1 and 2. The two fields occur 90 degrees out of phase with each other. These fields add and produce a single electromagnetic field. The total energy in the radiated wave remains constant in space except for some absorption of energy by the Earth. However, as the wave advances, the energy spreads out over a greater area and, at any given point, decreases as the distance increases.

Various factors in the antenna circuit affect the radiation of these waves. In figure 4-5, for example, if an alternating current is applied at the A end of the length of wire from A to B, the wave will travel along the wire until it reaches the B end. Since the B end is free, an open circuit exists and the wave cannot travel farther. This is a point of high impedance. The wave bounces back (reflects) from this point of high impedance and travels toward the starting point, where it is again reflected. The energy of the wave would be gradually dissipated by the resistance of the wire of this back-and-forth motion (oscillation); however, each time it reaches the starting point, the wave is reinforced by an amount sufficient to replace the energy lost. This results in continuous oscillations of energy along the wire and a high voltage at the A end of the wire. These oscillations are applied to the antenna at a rate equal to the frequency of the rf voltage.

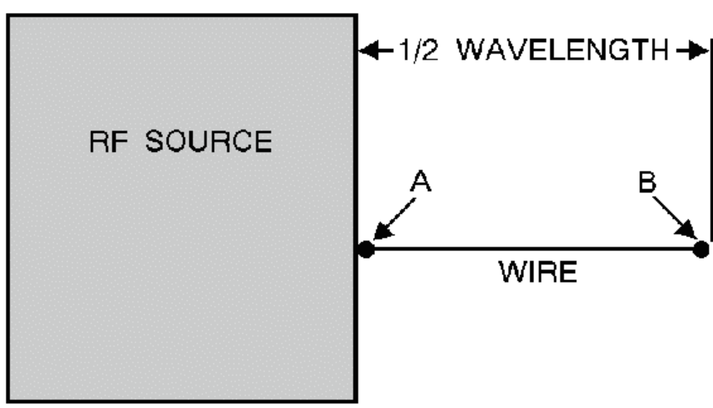


Figure 4-5.—Antenna and rf source.

These impulses must be properly timed to sustain oscillations in the antenna. The rate at which the waves travel along the wire is constant at approximately 300,000,000 meters per second. The length of the antenna must be such that a wave will travel from one end to the other and back again during the period of 1 cycle of the rf voltage. Remember, the distance a wave travels during the period of 1 cycle is known as the wavelength and is found by dividing the rate of travel by the frequency.

Look at the current and voltage (charge) distribution on the antenna in figure 4-6. A maximum movement of electrons is in the center of the antenna at all times; therefore, the center of the antenna is at a low impedance. This condition is called a STANDING WAVE of current. The points of high current and high voltage are known as current and voltage LOOPS. The points of minimum current and minimum voltage are known as current and voltage NODES. View A shows a current loop and current nodes. View B shows voltage loops and a voltage node. View C shows the resultant voltage and current loops and nodes. The presence of standing waves describes the condition of resonance in an antenna. At resonance the waves travel back and forth in the antenna reinforcing each other and the electromagnetic waves are transmitted into space at maximum radiation. When the antenna is not at resonance, the waves tend to cancel each other and lose energy in the form of heat.

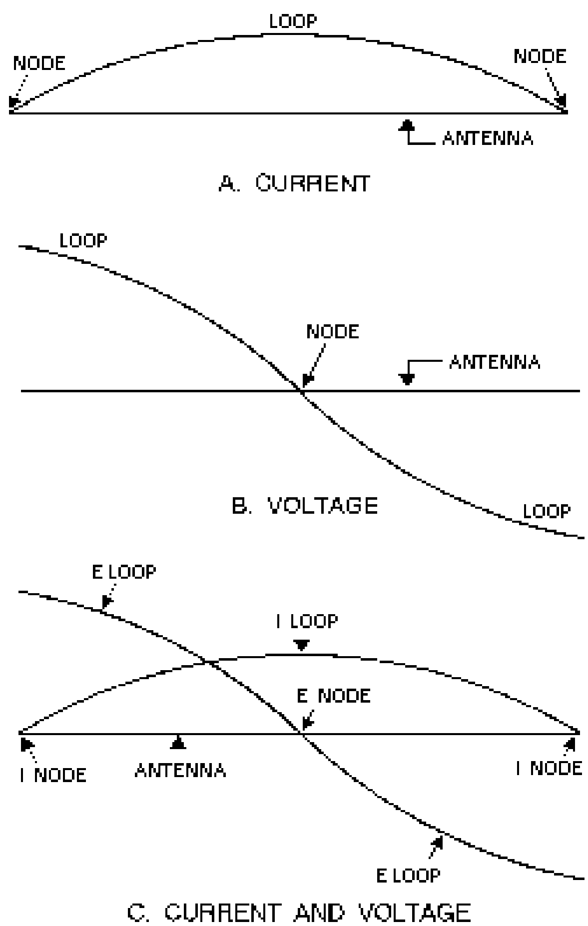


Figure 4-6.—Standing waves of voltage and current on an antenna.

Q4. If a wave travels exactly the length of an antenna from one end to the other and back during the period of 1 cycle, what is the length of the antenna?

Q5. What is the term used to identify the points of high current and high voltage on an antenna?

Q6. What is the term used to identify the points of minimum current and minimum voltage on an antenna?

ANTENNA CHARACTERISTICS

You can define an antenna as a conductor or group of conductors used either for radiating electromagnetic energy into space or for collecting it from space. Electrical energy from the transmitter is converted into electromagnetic energy by the antenna and radiated into space. On the receiving end, electromagnetic energy is converted into electrical energy by the antenna and is fed into the receiver.

Fortunately, separate antennas seldom are required for both transmitting and receiving rf energy. Any antenna can transfer energy from space to its input receiver with the same efficiency that it transfers energy from the transmitter into space. Of course, this is assuming that the same frequency is used in both cases. This property of interchangeability of the same antenna for transmitting and receiving is known as antenna RECIPROCITY. Antenna reciprocity is possible because antenna characteristics are essentially the same for sending and receiving electromagnetic energy.

RECIPROCITY OF ANTENNAS

In general, the various properties of an antenna apply equally, regardless of whether you use the antenna for transmitting or receiving. The more efficient a certain antenna is for transmitting, the more efficient it will be for receiving on the same frequency. Likewise, the directive properties of a given antenna also will be the same whether it is used for transmitting or receiving.

Assume, for example, that a certain antenna used with a transmitter radiates a maximum amount of energy at right angles to the axis of the antenna, as shown in figure 4-7, view A. Note the minimum amount of radiation along the axis of the antenna. Now, if this same antenna were used as a receiving antenna, as shown in view B, it would receive best in the same directions in which it produced maximum radiation; that is, at right angles to the axis of the antenna.

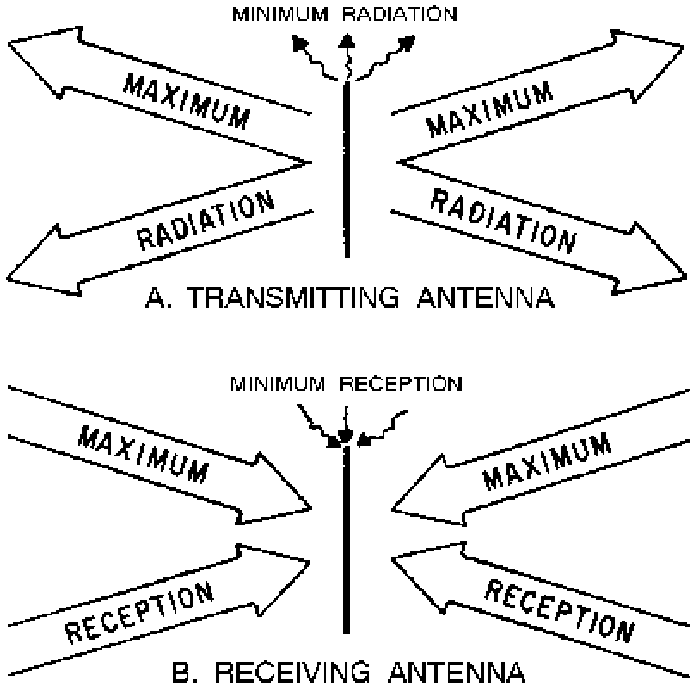


Figure 4-7.—Reciprocity of antennas.

ANTENNA GAIN

Another characteristic of a given antenna that remains the same whether the antenna is used for transmitting or receiving is GAIN. Some antennas are highly directional that is, more energy is propagated in certain directions than in others. The ratio between the amount of energy propagated in these directions compared to the energy that would be propagated if the antenna were not directional is known as its gain. When a transmitting antenna with a certain gain is used as a receiving antenna, it will also have the same gain for receiving.

POLARIZATION

Let's review polarization briefly. In chapter 2 you learned that the radiation field is composed of electric and magnetic lines of force. These lines of force are always at right angles to each other. Their intensities rise and fall together, reaching their maximums 90 degrees apart. The electric field determines the direction of polarization of the wave. In a vertically polarized wave, the electric lines of force lie in a vertical direction. In a horizontally polarized wave, the electric lines of force lie in a horizontal direction. Circular polarization has the electric lines of force rotating through 360 degrees with every cycle of rf energy.

The electric field was chosen as the reference field because the intensity of the wave is usually measured in terms of the electric field intensity (volts, millivolts, or microvolts per meter). When a single-wire antenna is used to extract energy from a passing radio wave, maximum pickup will result when the antenna is oriented in the same direction as the electric field. Thus a vertical antenna is used for the efficient reception of vertically polarized waves, and a horizontal antenna is used for the reception of horizontally polarized waves. In some cases the orientation of the electric field does not remain constant.

Instead, the field rotates as the wave travels through space. Under these conditions both horizontal and vertical components of the field exist and the wave is said to have an elliptical polarization.

Q7. The various properties of a transmitting antenna can apply equally to the same antenna when it is used as a receiving antenna. What term is used for this property?

Q8. The direction of what field is used to designate the polarization of a wave?

Q9. If a wave's electric lines of force rotate through 360 degrees with every cycle of rf energy, what is the polarization of this wave?

Polarization Requirements for Various Frequencies

Ground-wave transmission is widely used at medium and low frequencies. Horizontal polarization cannot be used at these frequencies because the electric lines of force are parallel to and touch the earth. Since the earth acts as a fairly good conductor at low frequencies, it would short out the horizontal electric lines of force and prevent the radio wave from traveling very far. Vertical electric lines of force, on the other hand, are bothered very little by the earth. Therefore vertical polarization is used for ground-wave transmission, allowing the radio wave to travel a considerable distance along the ground surface with minimum attenuation.

Sky-wave transmission is used at high frequencies. Either horizontal or vertical polarization can be used with sky-wave transmission because the sky wave arrives at the receiving antenna elliptically polarized. This is the result of the wave traveling obliquely through the Earth's magnetic field and striking the ionosphere. The radio wave is given a twisting motion as it strikes the ionosphere. Its orientation continues to change because of the unstable nature of the ionosphere. The relative amplitudes and phase differences between the horizontal and vertical components of the received wave also change. Therefore, the transmitting and receiving antennas can be mounted either horizontally or vertically.

Although either horizontally or vertically polarized antennas can be used for high frequencies, horizontally polarized antennas have certain advantages and are therefore preferred. One advantage is that vertically polarized interference signals, such as those produced by automobile ignition systems and electrical appliances, are minimized by horizontal polarization. Also, less absorption of radiated energy by buildings or wiring occurs when these antennas are used. Another advantage is that support structures for these antennas are of more convenient size than those for vertically polarized antennas.

For frequencies in the vhf or uhf range, either horizontal or vertical polarization is satisfactory. These radio waves travel directly from the transmitting antenna to the receiving antenna without entering the ionosphere. The original polarization produced at the transmitting antenna is maintained throughout the entire travel of the wave to the receiver. Therefore, if a horizontally polarized antenna is used for transmitting, a horizontally polarized antenna must be used for receiving. The requirements would be the same for a vertical transmitting and receiving antenna system.

For satellite communications, parallel frequencies can be used without interference by using polarized radiation. The system setup is shown in figure 4-8. One pair of satellite antennas is vertically polarized and another pair is horizontally polarized. Either vertically or horizontally polarized transmissions are received by the respective antenna and retransmitted in the same polarization. For example, transmissions may be made in the 3.7 to 3.74 GHz range on the vertical polarization path and in the 3.72 to 3.76 GHz range on the horizontal polarization path without adjacent frequency (co-channel) interference.

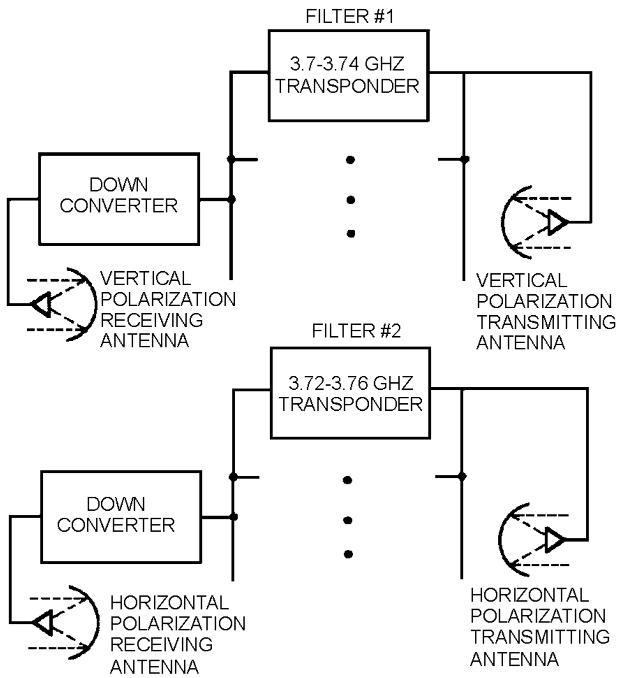


Figure 4-8.—Satellite transmissions using polarized radiation.

Advantages of Vertical Polarization

Simple vertical antennas can be used to provide OMNIDIRECTIONAL (all directions) communication. This is an advantage when communications must take place from a moving vehicle.

In some overland communications, such as in vehicular installations, antenna heights are limited to 3 meters (10 feet) or less. In such instances vertical polarization results in a stronger receiver signal than does horizontal polarization at frequencies up to about 50 megahertz. From approximately 50 to 100 megahertz, vertical polarization results in a slightly stronger signal than does horizontal polarization with antennas at the same height. Above 100 megahertz, the difference in signal strength is negligible.

For transmission over bodies of water, vertical polarization is much better than horizontal polarization for antennas at the lower heights. As the frequency increases, the minimum antenna height decreases. At 30 megahertz, vertical polarization is better for antenna heights below about 91 meters (300 feet); at 85 megahertz, antenna heights below 15 meters (50 feet); and still lower heights at the high frequencies. Therefore, at ordinary antenna mast heights of 12 meters (40 feet), vertical polarization is advantageous for frequencies less than about 100 megahertz.

Radiation is somewhat less affected by reflections from aircraft flying over the transmission path when vertical polarization is used instead of horizontal polarization. With horizontal polarization, such reflections cause variations in received signal strength. This factor is important in locations where aircraft traffic is heavy.

When vertical polarization is used, less interference is produced or picked up because of strong vhf and uhf broadcast transmissions (television and fm). This is because vhf and uhf transmissions use horizontal polarization. This factor is important when an antenna must be located in an urban area having several television and fm broadcast stations.

Advantages of Horizontal Polarization

A simple horizontal antenna is bi-directional. This characteristic is useful when you desire to minimize interference from certain directions. Horizontal antennas are less likely to pick up man-made interference, which ordinarily is vertically polarized.

When antennas are located near dense forests or among buildings, horizontally polarized waves suffer lower losses than vertically polarized waves, especially above 100 megahertz. Small changes in antenna locations do not cause large variations in the field intensity of horizontally polarized waves. When vertical polarization is used, a change of only a few meters in the antenna location may have a considerable effect on the received signal strength. This is the result of interference patterns that produce standing waves in space when spurious reflections from trees or buildings occur.

When simple antennas are used, the transmission line, which is usually vertical, is less affected by a horizontally mounted antenna. When the antenna is mounted at right angles to the transmission line and horizontal polarization is used, the line is kept out of the direct field of the antenna. As a result, the radiation pattern and electrical characteristics of the antenna are practically unaffected by the presence of the vertical transmission line.

Q10. What type of polarization should be used at medium and low frequencies?

Q11. What is an advantage of using horizontal polarization at high frequencies?

Q12. What type of polarization should be used if an antenna is mounted on a moving vehicle at frequencies below 50 megahertz?

RADIATION RESISTANCE

Radiated energy is the useful part of the transmitter's signal. However, it represents as much of a loss to the antenna as the energy lost in heating the antenna wire. In either case, the dissipated power is equal to I^2R. In the case of heat losses, the R is real resistance. In the case of radiation, R is an assumed resistance; if this resistance were actually present, it would dissipate the same amount of power that the antenna takes to radiate the energy. This assumed resistance is referred to as the RADIATION RESISTANCE.

Radiation resistance varies at different points on the antenna. This resistance is always measured at a current loop. For the antenna in free space, that is, entirely removed from any objects that might affect its operation, the radiation resistance is 73 ohms. A practical antenna located over a ground plane may have any value of radiation resistance from 0 to approximately 100 ohms. The exact value of radiation resistance depends on the height of the antenna above the ground. For most half-wave wire antennas, the radiation resistance is about 65 ohms. It will usually vary between 55 and 600 ohms for antennas constructed of rod or tubing. The actual value of radiation resistance, so long as it is 50 ohms or more, has little effect on the radiation efficiency of the antenna. This is because the ohmic resistance is about 1 ohm for conductors of large diameter. The ohmic resistance does not become important until the radiation resistance drops to a value less than 10 ohms. This may be the case when several antennas are coupled together.

RADIATION TYPES AND PATTERNS

The energy radiated from an antenna forms a field having a definite RADIATION PATTERN. A radiation pattern is a plot of the radiated energy from an antenna. This energy is measured at various angles at a constant distance from the antenna. The shape of this pattern depends on the type of antenna used. In this section, we will introduce the basic types of radiation (isotropic and anisotropic) and their radiation patterns.

Isotropic Radiation

Some antenna sources radiate energy equally in all directions. Radiation of this type is known as ISOTROPIC RADIATION. We all know the Sun radiates energy in all directions. The energy radiated from the Sun measured at any fixed distance and from any angle will be approximately the same. Assume that a measuring device is moved around the Sun and stopped at the points indicated in figure 4-9 to make a measurement of the amount of radiation. At any point around the circle, the distance from the measuring device to the Sun is the same. The measured radiation will also be the same. The Sun is therefore considered an isotropic radiator.

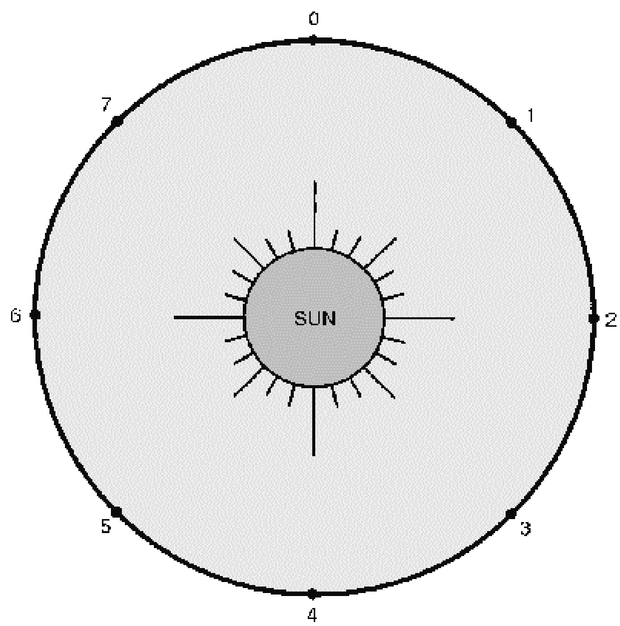


Figure 4-9.—Isotropic radiator.

To plot this pattern, we will assume that the radiation is measured on a scale of 0 to 10 units and that the measured amount of radiation is 7 units at all points. We will then plot our measurements on two different types of graphs, rectangular- and polar-coordinate graphs. The RECTANGULAR-COORDINATE GRAPH of the measured radiation, shown in view A of figure 4-10, is a straight line plotted against positions along the circle. View B shows the POLAR-COORDINATE GRAPH for the same isotropic source.

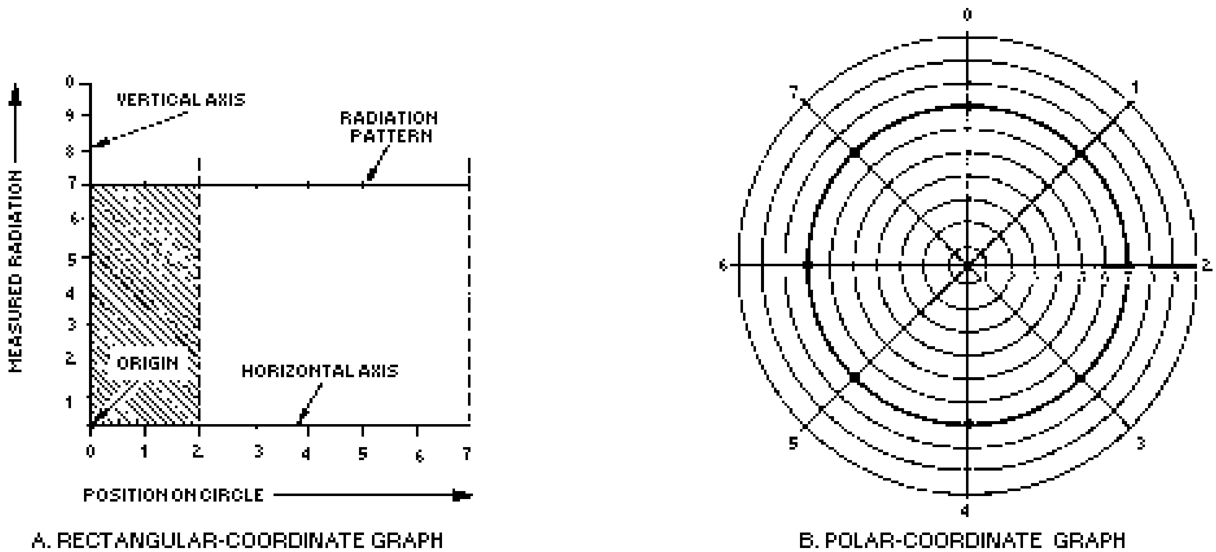


Figure 4-10.—Comparison of rectangular- and polar-coordinate graphs for an isotropic source.

In the rectangular-coordinate graph, points are located by projection from a pair of stationary, perpendicular axes. In the polar-coordinate graph, points are located by projection along a rotating axis (radius) to an intersection with one of several concentric, equally-spaced circles. The horizontal axis on the rectangular-coordinate graph corresponds to the circles on the polar-coordinate graph. The vertical axis on the rectangular-coordinate graph corresponds to the rotating axis (radius) on the polar-coordinate graph.

Rectangular-Coordinate Pattern

Look at view A of figure 4-10. The numbered positions around the circle are laid out on the HORIZONTAL AXIS of the graph from 0 to 7 units. The measured radiation is laid out on the VERTICAL AXIS of the graph from 0 to 10 units. The units on both axes are chosen so the pattern occupies a convenient part of the graph.

The horizontal and vertical axes are at a right angle to each other. The point where the axes cross each other is known as the ORIGIN. In this case, the origin is 0 on both axes. Now, assume that a radiation value of 7 units view B is measured at position 2. From position 2 on the horizontal axis, a dotted line is projected upwards that runs parallel to the vertical axis. From position 7 on the vertical axis, a line is projected to the right that runs parallel to the horizontal axis. The point where the two lines cross (INTERCEPT) represents a value of 7 radiation units at position 2. This is the only point on the graph that can represent this value.

As you can see from the figure, the lines used to plot the point form a rectangle. For this reason, this type of plot is called a *rectangular-coordinate graph*. A new rectangle is formed for each different point plotted. In this example, the points plotted lie in a straight line extending from 7 units on the vertical scale to the projection of position 7 on the horizontal scale. This is the characteristic pattern in rectangular coordinates of an isotropic source of radiation.

Polar-Coordinate Pattern

The polar-coordinate graph has proved to be of great use in studying radiation patterns. Compare views A and B of figure 4-10. Note the great difference in the shape of the radiation pattern when it is transferred from the rectangular-coordinate graph in view A to the polar-coordinate graph in view B. The scale of radiation values used in both graphs is identical, and the measurements taken are both the same. However, the shape of the pattern is drastically different.

Look at view B of figure 4-10 and assume that the center of the concentric circles is the Sun. Assume that a radius is drawn from the Sun (center of the circle) to position 0 of the circle. When you move to position 1, the radius moves to position 1; when you move to position 2, the radius also moves to position 2, and so on.

The positions where a measurement was taken are marked as 0 through 7 on the graph. Note how the position of the radius indicates the actual direction from the source at which the measurement was taken. This is a distinct advantage over the rectangular-coordinate graph in which the position is indicated along a straight-line axis and has no physical relation to the actual direction of measurement. Now that we have a way to indicate the *direction* of measurement, we must devise a way to indicate the *magnitude* of the radiation.

Notice that the rotating axis is always drawn from the center of the graph to some position on the edge of the graph. As the axis moves toward the edge of the graph, it passes through a set of equally-spaced, concentric circles. In this example view B, they are numbered successively from 1 to 10 from the center out. These circles are used to indicate the magnitude of the radiation.

The advantages of the polar-coordinate graph are immediately evident. The source, which is at the center of the observation circles, is also at the center of the graph. By looking at a polar-coordinate plot of a radiation pattern, you can immediately see the direction and strength of radiation put out by the source. Therefore, the polar-coordinate graph is more useful than the rectangular-coordinate graph in plotting radiation patterns.

Anisotropic Radiation

Most radiators emit (radiate) stronger radiation in one direction than in another. A radiator such as this is referred to as ANISOTROPIC. An example of an anisotropic radiator is an ordinary flashlight. The beam of the flashlight lights only a portion of the space surrounding it. If a circle is drawn with the flashlight as the center, as shown in view B of figure 4-11, the radiated light can be measured at different positions around the circle. Again, as with the isotropic radiator, all positions are the same distance from the center, but at different angles. However, in this illustration the radiated light is measured at 16 different positions on the circle.

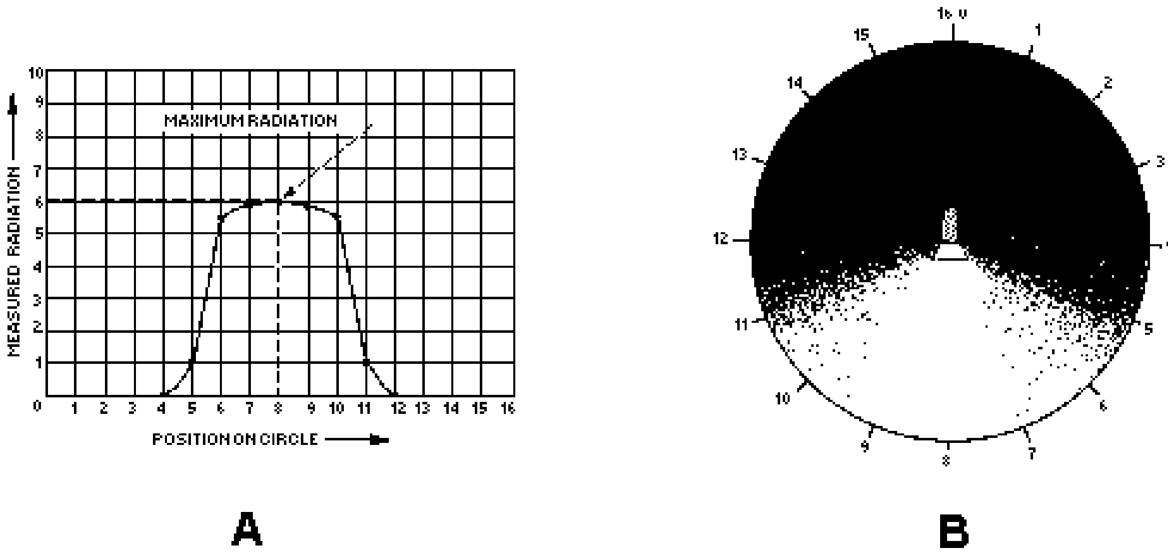


Figure 4-11.—Anisotropic radiator.

Directly behind the flashlight (position 0) the radiation measured is minimum. Accordingly, a 0 value is assigned to this position in the rectangular-coordinate graph (fig. 4-11, view A). This radiation remains at minimum until position 4 is reached. Between positions 4 and 6, the measuring device enters the flashlight beam. You can see this transition from darkness to brightness easily in view B. Radiation is fairly constant between positions 6 and 10. Maximum brightness occurs at position 8, which is directly in the path of the flashlight beam. From positions 10 to 12, the measuring device leaves the flashlight beam and the radiation measurement falls off sharply. At position 13 the radiation is again at 0 and stays at this value back to position 0.

Radiation from a light source and radiation from an antenna are both forms of electromagnetic waves. Therefore, the measurement of radiation of an antenna follows the same basic procedure as that just described for the Sun and the flashlight. Before proceeding further with the study of antenna patterns, you should be sure you understand the methods used to graph the measured radiation (magnitude of the radiation). Study the rectangular- and polar-coordinate systems of plotting presented in the following section.

Q13. What is the radiation resistance of a half-wave antenna in free space?

Q14. A radiating source that radiates energy stronger in one direction than another is known as what type of radiator?

Q15. A radiating source that radiates energy equally in all directions is known as what type of radiator?

Q16. A flashlight is an example of what type of radiator?

In figure 4-11, view A, the radiation pattern of the flashlight is graphed in rectangular coordinates. The illustration of the flashlight beam in view B clearly indicates the shape of the flashlight beam. This is not evident in the radiation pattern plotted on the rectangular-coordinate graph. Now look at figure 4-12. The radiation pattern shown in this figure looks very much like the actual flashlight beam. The pattern in figure 4-12 is plotted using the same values as those of figure 4-11, view A, but is drawn using polar coordinates.

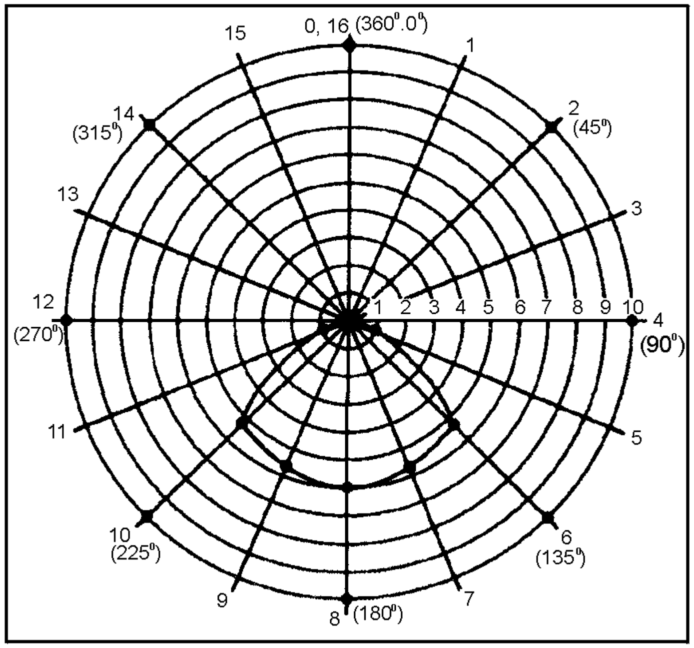


Figure 4-12.—Polar-coordinate graph for anisotropic radiator.

The positions marked off on the two polar-coordinate graphs in figures 4-10 and 4-12 were selected and numbered arbitrarily. However, a standard method allows the positions around a source to be marked off so that one radiation pattern can easily be compared with another. This method is based on the fact that a circle has a radius of 360 degrees. The radius extending vertically from the center (position 0 in figure 4-10) is designated 0 degrees. At position 4 the radius is at a right angle to the 0-degree radius. Accordingly, the radius at position 4 is marked 90 degrees, position 8 is 180 degrees, position 12 is 270 degrees, and position 16 is 360 degrees. The various radii drawn on the graph are all marked according to the angle each radius makes with the reference radius at 0 degrees.

The radiation pattern in figure 4-12 is obtained by using the same procedure that was used for (figure 4-10, view B). The radiation measured at positions 1, 2, 3, and 4 is 0. Position 5 measures approximately 1 unit. This is marked on the graph and the rotating radius moves to position 6. At this position a reading of 5.5 units is taken. As before, this point is marked on the graph. The procedure is repeated around the circle and a reading is obtained from positions 6 through 11. At position 12 no radiation is indicated, and this continues on to position 16.

The polar-coordinate graph now shows a definite area enclosed by the radiation pattern. This pattern indicates the general direction of radiation from the source. The enclosed area is called a LOBE. Outside of this area, minimum radiation is emitted in any direction. For example, at position 2 the radiation is 0. Such a point is called a NULL. In real situations, some radiation is usually transmitted in all directions. Therefore, a null is used to indicate directions of minimum radiation. The pattern of figure 4-12 shows one lobe and one continuous null.

ANTENNA LOADING

You will sometimes want to use one antenna system for transmitting and receiving on several different frequencies. Since the antenna must always be in resonance with the applied frequency, you may need to either physically or electrically lengthen or shorten the antenna.

Except for trailing-wire antennas used in aircraft installations (which may be lengthened or shortened), physically lengthening the antenna is not very practical. But you can achieve the same result by changing the electrical length of the antenna. To change the electrical length, you can insert either an inductor or a capacitor in series with the antenna. This is shown in figure 4-13, views A and B. Changing the electrical length by this method is known as LUMPED-IMPEDANCE TUNING, or LOADING. The electrical length of any antenna wire can be increased or decreased by loading. If the antenna is too short for the wavelength being used, it is resonant at a higher frequency than that at which it is being excited. Therefore, it offers a capacitive reactance at the excitation frequency. This capacitive reactance can be compensated for by introducing a lumped-inductive reactance, as shown in view A. Similarly, if the antenna is too long for the transmitting frequency, it offers an inductive reactance. Inductive reactance can be compensated for by introducing a lumped-capacitive reactance, as shown in view B. An antenna without loading is represented in view C.

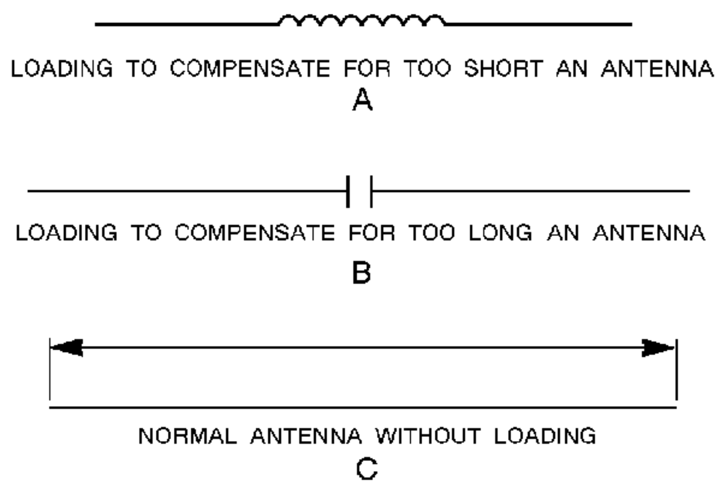


Figure 4-13.—Electrically equal antenna.

BASIC ANTENNAS

Before you look at the various types of antennas, consider the relationship between the wavelength at which the antenna is being operated and the actual length of the antenna. An antenna does not necessarily radiate or receive more energy when it is made longer. Specific dimensions must be used for efficient antenna operation.

Nearly all antennas have been developed from two basic types, the Hertz and the Marconi. The basic Hertz antenna is 1/2 wavelength long at the operating frequency and is insulated from ground. It is often called a DIPOLE or a DOUBLET. The basic Marconi antenna is 1/4 wavelength long and is either grounded at one end or connected to a network of wires called a COUNTERPOISE. The ground or counterpoise provides the equivalent of an additional 1/4 wavelength, which is required for the antenna to resonate.

HALF-WAVE ANTENNAS

A half-wave antenna (referred to as a dipole, Hertz, or doublet) consists of two lengths of wire rod, or tubing, each 1/4 wavelength long at a certain frequency. It is the basic unit from which many complex antennas are constructed. The half-wave antenna operates independently of ground; therefore, it may be installed far above the surface of the Earth or other absorbing bodies. For a dipole, the current is maximum at the center and minimum at the ends. Voltage is minimum at the center and maximum at the ends, as was shown in figure 4-6.

Radiation Patterns

In the following discussion, the term DIPOLE is used to mean the basic half-wave antenna. The term DOUBLET is used to indicate an antenna that is very short compared with the wavelength of the operating frequency. Physically, it has the same shape as the dipole.

RADIATION PATTERN OF A DOUBLET.—The doublet is the simplest form of a practical antenna. Its radiation pattern can be plotted like the radiation pattern of the flashlight (fig. 4-12). Figure 4-14 shows the development of vertical and horizontal patterns for a doublet. This in NOT a picture of the radiation, but three-dimensional views of the pattern itself. In three views the pattern resembles a doughnut. From the dimensions in these views, two types of polar-coordinate patterns can be drawn, horizontal and vertical. The HORIZONTAL PATTERN view A is derived from the solid pattern view C by slicing it horizontally. This produces view B, which is converted to the polar coordinates seen in view A. The horizontal pattern illustrates that the radiation is constant in any direction along the horizontal plane.

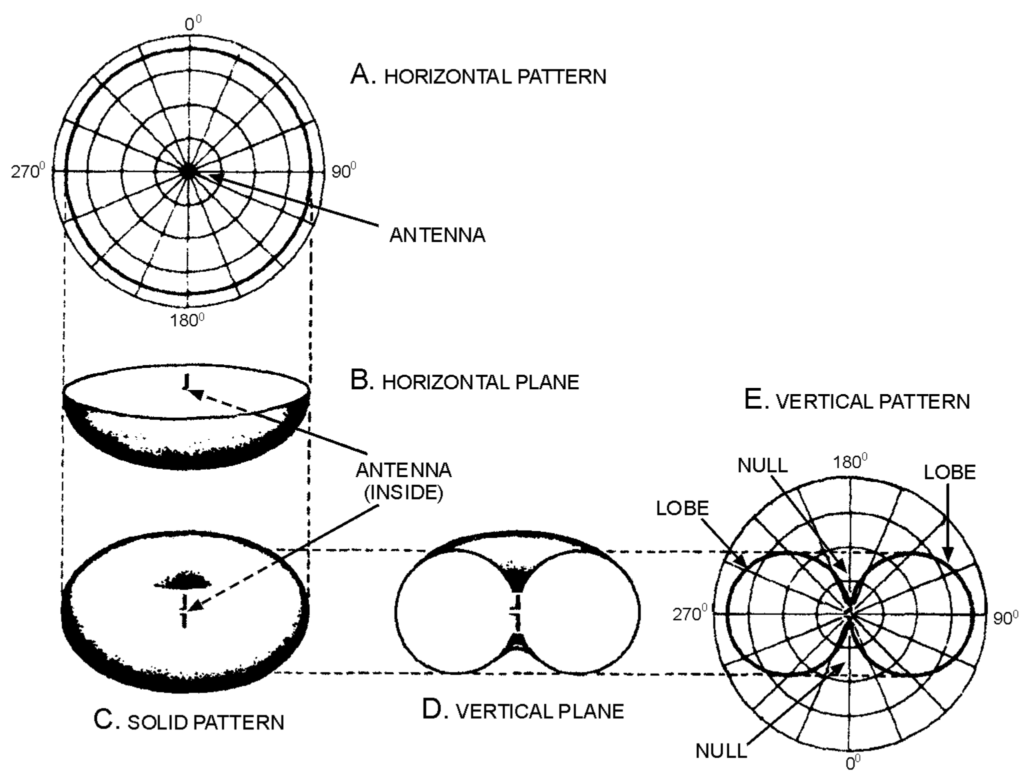


Figure 4-14.—Development of vertical and horizontal patterns.

A VERTICAL PATTERN view E is obtained from the drawing of the vertical plane view D of the radiation pattern view C. The radiation pattern view C is sliced in half along a vertical plane through the antenna. This produces the vertical plane pattern in view D. Note how the vertical plane in view D of the radiation pattern differs from the horizontal plane in view B. The vertical pattern view E exhibits two lobes and two nulls. The difference between the two patterns is caused by two facts: (1) no radiation is emitted from the ends of the doublet; and (2) maximum radiation comes from the doublet in a direction perpendicular to the antenna axis. This type of radiation pattern is both NONDIRECTIONAL (in a horizontal plane) and DIRECTIONAL (in a vertical plane).

From a practical viewpoint, the doublet antenna can be mounted either vertically or horizontally. The doublet shown in figure 4-14 is mounted vertically, and the radiated energy spreads out about the antenna in every direction in the horizontal plane. Since ordinarily the horizontal plane is the useful plane, this arrangement is termed NONDIRECTIONAL. The directional characteristics of the antenna in other planes is ignored. If the doublet were mounted horizontally, it would have the effect of turning the pattern on edge, reversing the patterns given in figure 4-14. The antenna would then be directional in the horizontal plane. The terms "directional" and "nondirectional" are used for convenience in describing specific radiation patterns. A complete description always involves a figure in three dimensions, as in the radiation pattern of figure 4-14.

Q17. What terms are often used to describe basic half-wave antennas?

Q18. If a basic half-wave antenna is mounted vertically, what type of radiation pattern will be produced?

Q19. In which plane will the half-wave antenna be operating if it is mounted horizontally?

RADIATION PATTERN OF A DIPOLE.—The radiation pattern of a dipole (fig. 4-15) is similar to that of the doublet (fig. 4-14). Increasing the length of the doublet to 1/2 wavelength has the effect of flattening out the radiation pattern. The radiation pattern in the horizontal plane of a dipole is a larger circle than that of the doublet. The vertical-radiation pattern lobes are no longer circular. They are flattened out and the radiation intensity is greater.

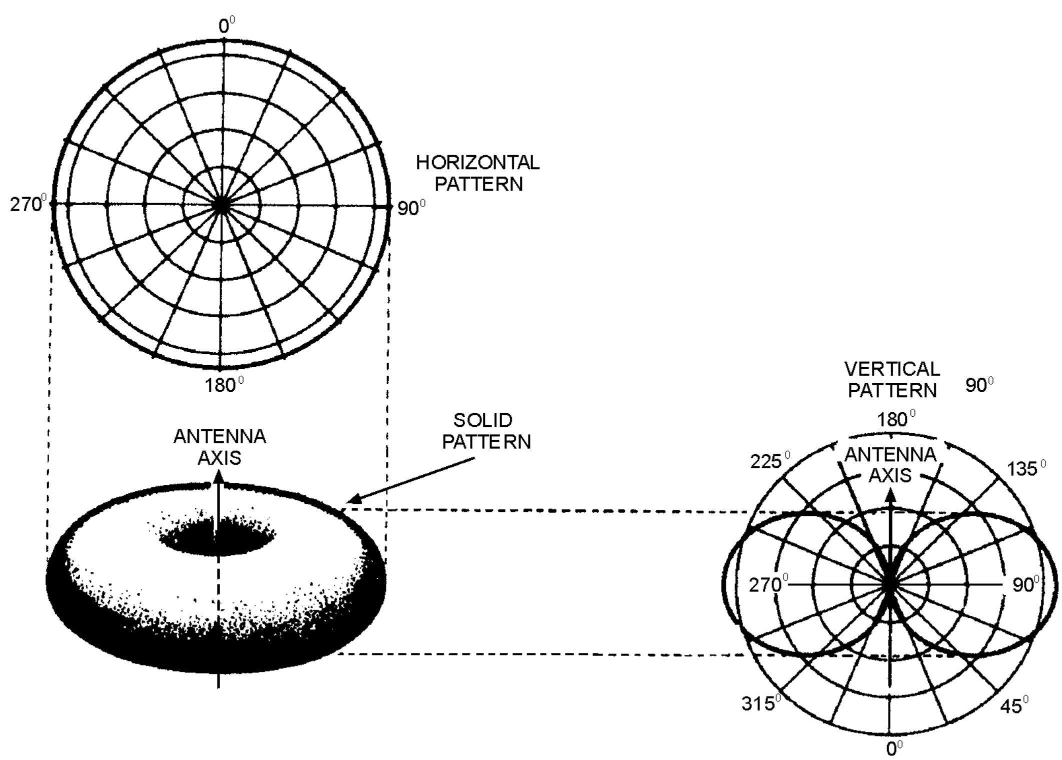


Figure 4-15.—Radiation pattern of a dipole.

Methods of Feeding Energy to an Antenna

Voltage and current distribution for the half-wave antenna (shown in figure 4-16) is the same as that for the antenna discussed earlier in this chapter. A point closely related to the voltage and current distribution on an antenna is the method of feeding the transmitter output to the antenna. The simplest method of feeding energy to the half-wave antenna is to connect one end through a capacitor to the final output stage of the transmitter. This method is often called the END-FEED or VOLTAGE-FEED method. In this method the antenna is fed at a point of high voltage (the end).

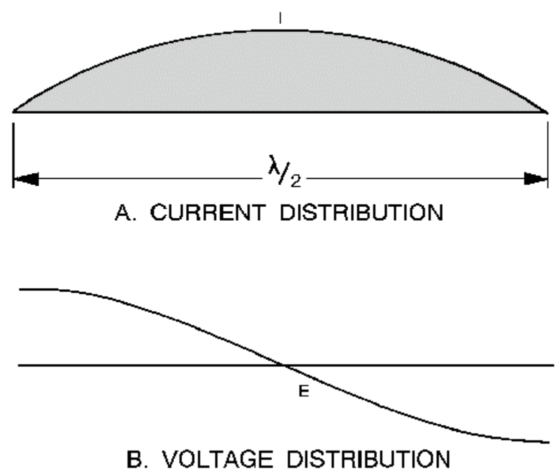


Figure 4-16.—Standing waves of current and voltage.

Energy may also be fed to the half-wave antenna by dividing the antenna at its center and connecting the transmission line from the final transmitter output stage to the two center ends of the halved antenna. Since the antenna is now being fed at the center (a point of low voltage and high current), this type of feed is known as the CENTER-FEED or CURRENT-FEED method. The point of feed is important in determining the type of transmission line to be used.

QUARTER-WAVE ANTENNAS

As you have studied in the previous sections, a 1/2 wavelength antenna is the shortest antenna that can be used in free space. If we cut a half-wave antenna in half and then ground one end, we will have a grounded quarter-wave antenna. This antenna will resonate at the same frequency as the ungrounded half-wave antenna. Such an antenna is referred to as a QUARTER-WAVE or Marconi antenna. Quarter-wave antennas are widely used in the military. Most mobile transmitting and receiving antennas (fig. 4-17) are quarter-wave antennas.

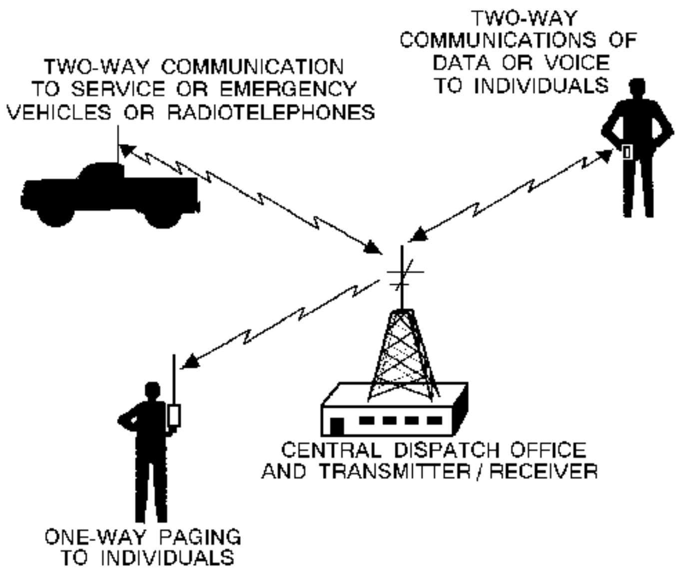


Figure 4-17.—Mobile antennas.

As stated above, a grounded quarter-wave antenna will resonate at the same frequency as an ungrounded half-wave antenna. This is because ground has high conductivity and acts as an electrical mirror image. This characteristic provides the missing half of the antenna, as shown in the bottom part of figure 4-18. In other words, the grounded quarter-wave antenna acts as if another quarter-wave were actually down in the earth.

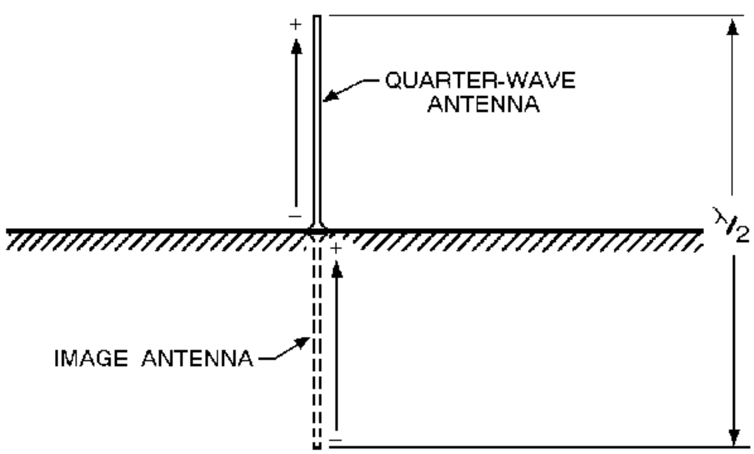


Figure 4-18.—Grounded quarter-wave antenna image.

Characteristics of Quarter-Wave Antennas

The grounded end of the quarter-wave antenna has a low input impedance and has low voltage and high current at the input end, as shown in figure 4-18. The ungrounded end has a high impedance, which causes high voltage and low current. The directional characteristics of a grounded quarter-wave antenna are the same as those of a half-wave antenna in free space.

As explained earlier, ground losses affect radiation patterns and cause high signal losses for some frequencies. Such losses may be greatly reduced if a perfectly conducting ground is provided in the vicinity of the antenna. This is the purpose of a GROUND SCREEN (figure 4-19, view A) and COUNTERPOISE view B.

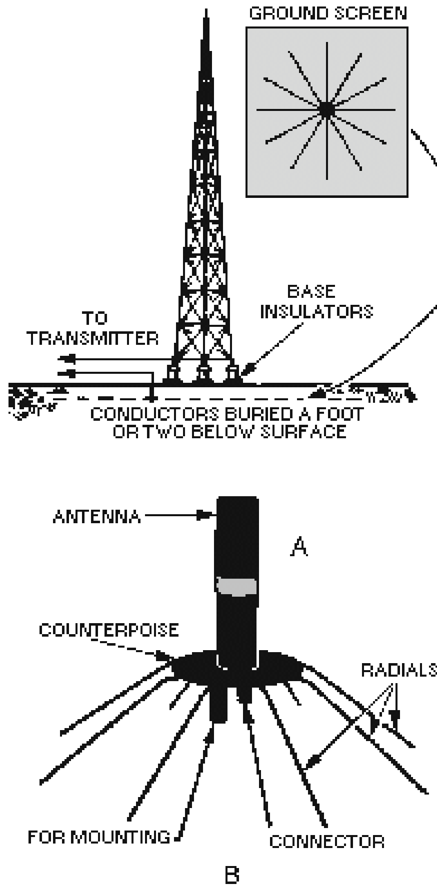


Figure 4-19.—Groundscreen and counterpoise.

The ground screen in view A is composed of a series of conductors buried 1 or 2 feet (0.3 to 0.6 meter) below the surface of the earth and arranged in a radial pattern. These conductors reduce losses in the ground in the immediate vicinity of the antenna. Such a radial system of conductors is usually 1/2 wavelength in diameter.

A counterpoise view B is used when easy access to the base of the antenna is necessary. It is also used when the earth is not a good conducting surface, such as ground that is sandy or solid rock. The counterpoise serves the same purpose as the ground screen but it is usually elevated above the earth. No specific dimensions are necessary in the construction of a counterpoise nor is the number of wires particularly critical. A practical counterpoise may be assembled from a large screen of chicken wire or some similar material. This screen may be placed on the ground, but better results are obtained if it is placed a few feet above the ground.

Q20. Since the radiation pattern of a dipole is similar to that of a doublet, what will happen to the pattern if the length of the doublet is increased?

Q21. What is the simplest method of feeding power to the half-wave antenna?

Q22. What is the radiation pattern of a quarter-wave antenna?

Q23. Describe the physical arrangement of a ground screen.

FOLDED DIPOLE

The use of parasitic elements and various stacking arrangements causes a reduction in the radiation resistance of a center-fed, half-wave antenna. Under these conditions obtaining a proper impedance match between the radiator and the transmission line is often difficult. A convenient method of overcoming these difficulties is to use a FOLDED DIPOLE in place of the center-fed radiator. (See views A and B of figure 4-20).

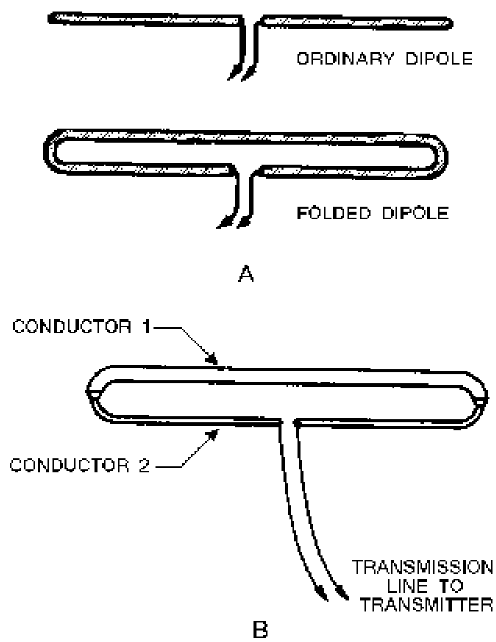


Figure 4-20.—Folded-dipole antennas.

A FOLDED DIPOLE is an ordinary half-wave antenna that has one or more additional conductors connected across its ends. Additional conductors are mounted parallel to the dipole elements at a distance equal to a very small fraction of a wavelength. Spacing of several inches is common.

The feed-point impedance can be further increased by using three or four properly spaced parallel conductors. Standard feed-line SPREADERS are used to maintain this spacing when required. In any folded dipole, the increase of impedance is the square of the number of conductors used in the radiator. Thus, a three-wire dipole has nine times (3^2) the feed-point impedance of a simple center-fed dipole. A second method of stepping up the impedance of a folded dipole is to use two conductors with different radii, as shown in view B.

The directional characteristics of a folded dipole are the same as those of a simple dipole. However, the reactance of a folded dipole varies much more slowly as the frequency is varied from resonance. Because of this the folded dipole can be used over a much wider frequency range than is possible with a simple dipole.

Q24. What is the difference in the amount of impedance between a three-wire dipole and a simple center-fed dipole?

Q25. Which has a wider frequency range, a simple dipole or a folded dipole?

ARRAY ANTENNAS

An array antenna is a special arrangement of basic antenna components involving new factors and concepts. Before you begin studying about arrays, you need to study some new terminology.

DEFINITION OF TERMS

An array antenna is made up of more than one ELEMENT, but the basic element is generally the dipole. Sometimes the basic element is made longer or shorter than a half-wave, but the deviation usually is not great.

A DRIVEN element is similar to the dipole you have been studying and is connected directly to the transmission line. It obtains its power directly from the transmitter or, as a receiving antenna, it delivers the received energy directly to the receiver. A PARASITIC ELEMENT is located near the driven element from which it gets its power. It is placed close enough to the driven element to permit coupling.

A parasitic element is sometimes placed so it will produce maximum radiation (during transmission) from its associated driver. When it operates to reinforce energy coming from the driver toward itself, the parasitic element is referred to as a DIRECTOR. If a parasitic element is placed so it causes maximum energy radiation in a direction away from itself and toward the driven element, that parasitic element is called a REFLECTOR.

If all of the elements in an array are driven, the array is referred to as a DRIVEN ARRAY (sometimes as a CONNECTED ARRAY). If one or more elements are parasitic, the entire system usually is considered to be a PARASITIC ARRAY.

MULTIELEMENT ARRAYS frequently are classified according to their directivity. A BIDIRECTIONAL ARRAY radiates in opposite directions along the line of maximum radiation. A UNIDIRECTIONAL ARRAY radiates in only one general direction.

Arrays can be described with respect to their radiation patterns and the types of elements of which they are made. However, you will find it useful to identify them by the physical placement of the elements and the direction of radiation with respect to these elements. Generally speaking, the term BROADSIDE ARRAY designates an array in which the direction of maximum radiation is perpendicular to the plane containing these elements. In actual practice, this term is confined to those arrays in which the elements themselves are also broadside, or parallel, with respect to each other.

A COLLINEAR ARRAY is one in which all the elements lie in a straight line with no radiation at the ends of the array. The direction of maximum radiation is perpendicular to the axis of the elements.

An END-FIRE ARRAY is one in which the principal direction of radiation is along the plane of the array and perpendicular to the elements. Radiation is from the end of the array, which is the reason this arrangement is referred to as an end-fire array.

Sometimes a system uses the characteristics of more than one of the three types mentioned. For instance, some of the elements may be collinear while others may be parallel. Such an arrangement is often referred to as a COMBINATION ARRAY or an ARRAY OF ARRAYS. Since maximum radiation occurs at right angles to the plane of the array, the term broadside array is also used.

The FRONT-TO-BACK RATIO is the ratio of the energy radiated in the principal direction compared to the energy radiated in the opposite direction for a given antenna.

PHASING

Various reflected and refracted components of the propagated wave create effects of reinforcement and cancellation. At certain distant points from the transmitter, some of the wave components meet in space. Reception at these points is either impaired or improved. If the different components arrive at a given point in the same phase, they add, making a stronger signal available. If they arrive out of phase, they cancel, reducing the signal strength.

Radiation Pattern

Effects similar to those described in the preceding paragraph can be produced at the transmitting point itself. Consider the antennas shown in figure 4-21, views A and B. View A shows an unobstructed view of the radiation pattern of a single dipole. In view B two dipoles, shown as points 1 and 2, are perpendicular to the plane of the page. They are spaced 1/4 wavelength apart at the operating frequency. The radiation pattern from either antenna 1 or 2, operating alone, would be uniform in all directions in this plane, as shown in view A. Suppose that current is being fed to both antennas from the same transmitter in such a way that the current fed to antenna 2 lags the current in antenna 1 by 90 degrees. Energy radiating from antenna 1 toward receiving location X will reach antenna 2 after 1/4 cycle of operation. The energy from both antennas will add, and propagation toward X will be strong.

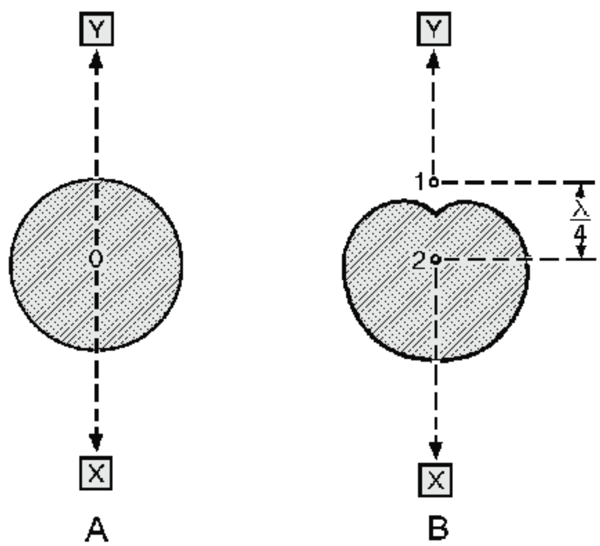


Figure 4-21.—Phasing of antenna in free space.

Radiation from antenna 2 toward receiving location Y will reach antenna 1 after 1/4 cycle. The energy in antenna 1 was 1/4 cycle behind that of antenna 2 to begin with; therefore, the radiation from antenna 1 toward receiving point Y will be exactly 180 degrees out of phase with that of antenna 2. As a result, the radiation fields will cancel and there will be no radiation toward Y.

At receiving points away from the line of radiation, phase differences occur between 0 and 180 degrees, producing varying amounts of energy in that direction. The overall effect is shown by the radiation pattern shown in view B. The physical phase relationship caused by the 1/4-wavelength spacing between the two elements, as well as the phase of the currents in the elements, has acted to change the radiation pattern of the individual antennas.

Stub Phasing

In the case just discussed, the currents fed to the two antennas from the same transmitter were 90 degrees out of phase. Sections of transmission line, called STUBS, are frequently used for this purpose. These stubs can be adjusted to produce any desired phase relationship between connected elements.

When two collinear half-wave elements are connected directly so their currents are in the same phase, the effect is the same as that of a full-wave antenna, as shown in figure 4-22, view A. The current in the first 1/2 wavelength is exactly 180 degrees out of phase with that in the second 1/2 wavelength. This is the opposite of the desired condition. In the illustration, arrows are used to indicate the direction of current flow in the antenna. (Using arrows is a convenient means of determining the phase on more complicated arrays.)

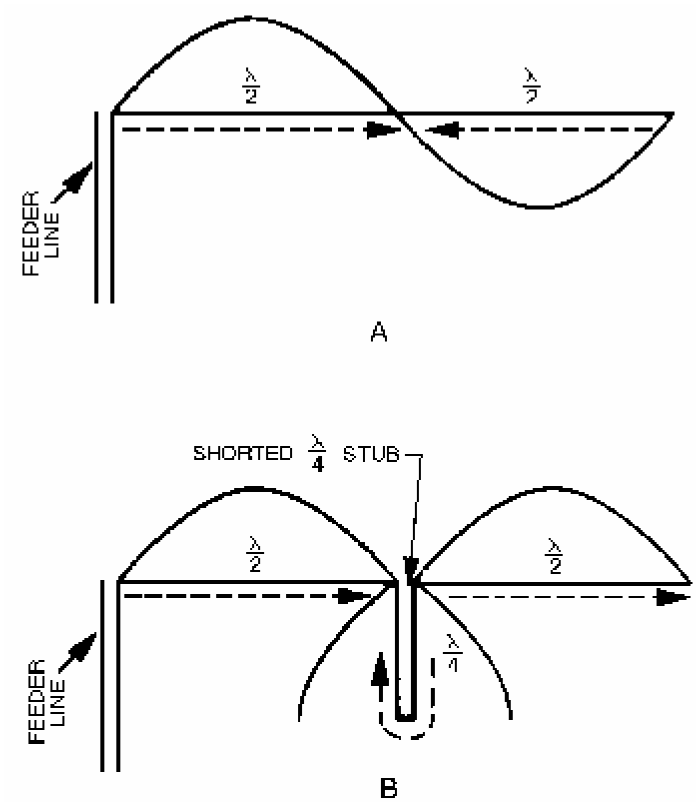


Figure 4-22.—Phasing of connected elements.

When the two elements are connected by a shorted 1/4-wavelength stub, as shown in view B, current travels down one side of the stub and up the other. It travels a distance of a 1/2 wavelength in the stub itself. As a result, the current moves through 1/2 cycle of change. When the current reaches the second element, it is in the desired phase. Since the current on one side of the stub is equal and opposite to the current on the other side, the fields produced here cancel and no radiation is transmitted from the stub itself.

DIRECTIVITY

The DIRECTIVITY of an antenna or an array can be determined by looking at its radiation pattern. In an array propagating a given amount of energy, more radiation takes place in certain directions than in others. The elements in the array can be altered in such a way that they change the pattern and distribute it more uniformly in all directions. The elements can be considered as a group of antennas fed from a common source and facing different directions. On the other hand, the elements could be arranged so that the radiation would be focused in a single direction. With no increase in power from the transmitter, the amount of radiation in a given direction would be greater. Since the input power has no increase, this increased directivity is achieved at the expense of gain in other directions.

Directivity and Interference

In many applications, sharp directivity is desirable although no need exists for added gain. Examine the physical disposition of the units shown in figure 4-23. Transmitters 1 and 2 are sending information to receivers 1 and 2, respectively, along the paths shown by the solid arrows. The distance between transmitter 1 and receiver 1 or between transmitter 2 and receiver 2 is short and does not require high-power transmission. The antennas of the transmitters propagate well in all directions. However, receiver 1 picks up some of the signals from transmitter 2, and receiver 2 picks up some of the signals from transmitter 1, as shown by the broken arrows. This effect is emphasized if the receiving antennas intercept energy equally well in all directions.

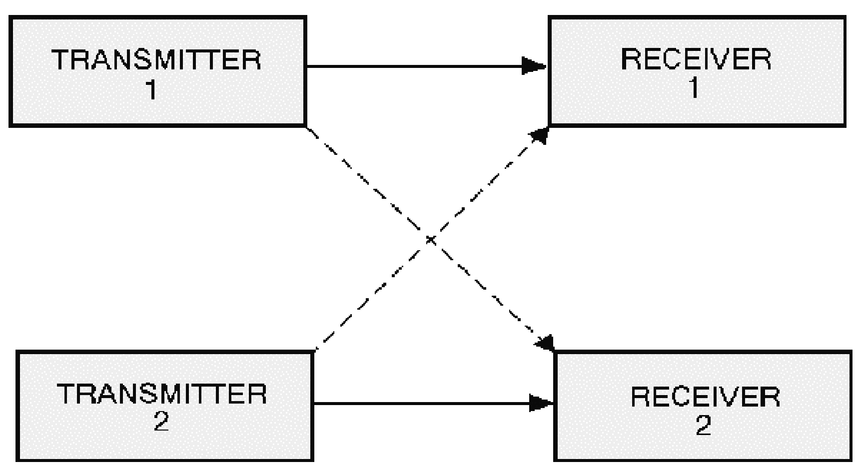


Figure 4-23.—Directivity and interference.

The use of highly directional arrays as radiators from the transmitters tends to solve the problem. The signals are beamed along the paths of the solid arrows and provide very low radiation along the paths of the broken arrows. Further improvement along these lines is obtained by the use of narrowly directed arrays as receiving antennas. The effect of this arrangement is to select the desired signal while discriminating against all other signals. This same approach can be used to overcome other types of radiated interference. In such cases, preventing radiation in certain directions is more important than producing greater gain in other directions.

Look at the differences between the field patterns of the single-element antenna and the array, as illustrated in figure 4-24. View A shows the relative field-strength pattern for a horizontally polarized single antenna. View B shows the horizontal-radiation pattern for an array. The antenna in view A radiates fairly efficiently in the desired direction toward receiving point X. It radiates equally as efficiently toward Y, although no radiation is desired in this direction. The antenna in view B radiates strongly to point X, but very little in the direction of point Y, which results in more satisfactory operation.

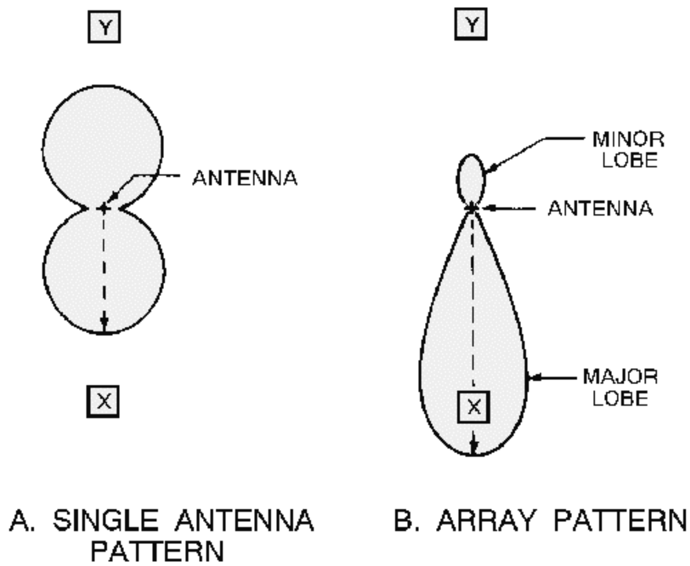


Figure 4-24.—Single antenna versus array.

Major and Minor Lobes

The pattern shown in figure 4-24, view B, has radiation concentrated in two lobes. The radiation intensity in one lobe is considerably stronger than in the other. The lobe toward point X is called a MAJOR LOBE; the other is a MINOR LOBE. Since the complex radiation patterns associated with arrays frequently contain several lobes of varying intensity, you should learn to use appropriate terminology. In general, major lobes are those in which the greatest amount of radiation occurs. Minor lobes are those in which the radiation intensity is least.

Q26. What is the purpose of antenna stubs?

Q27. What is the primary difference between the major and minor lobes of a radiation pattern?

DIRECTIONAL ARRAYS

You have already learned about radiation patterns and directivity of radiation. These topics are important to you because using an antenna with an improper radiation pattern or with the wrong directivity will decrease the overall performance of the system. In the following paragraphs, we discuss in more detail the various types of directional antenna arrays mentioned briefly in the "definition of terms" paragraph above.

Collinear Array

The pattern radiated by the collinear array is similar to that produced by a single dipole. The addition of the second radiator, however, tends to intensify the pattern. Compare the radiation pattern of the dipole (view A of figure 4-25) and the two-element antenna in view B. You will see that each pattern consists of two major lobes in opposite directions along the same axis, QQ1. There is little or no radiation along the

PP1 axis. QQ1 represents the line of maximum propagation. You can see that radiation is stronger with an added element. The pattern in view B is sharper, or more directive, than that in view A. This means that the gain along the line of maximum energy propagation is increased and the beam width is decreased. As more elements are added, the effect is heightened, as shown in view C. Unimportant minor lobes are generated as more elements are added.

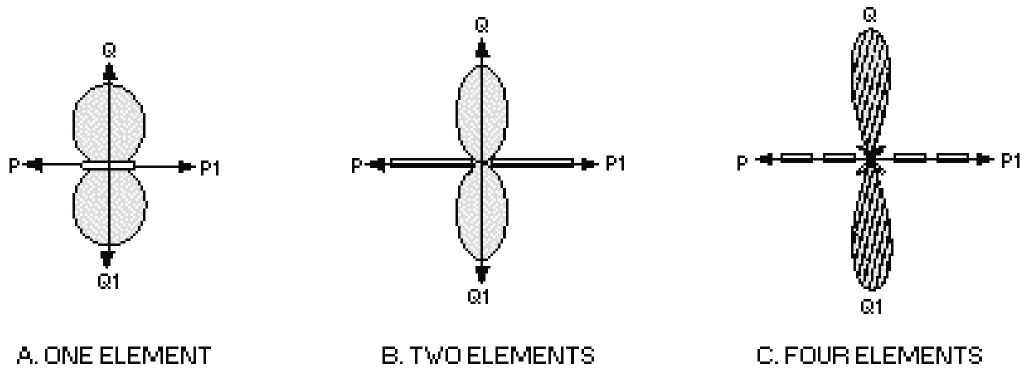


Figure 4-25.—Single half-wave antenna versus two half-wave antennas in phase.

More than four elements are seldom used because accumulated losses cause the elements farther from the point of feeding to have less current than the nearer ones. This introduces an unbalanced condition in the system and impairs its efficiency. Space limitations often are another reason for restricting the number of elements. Since this type of array is in a single line, rather than in a vertically stacked arrangement, the use of too many elements results in an antenna several wavelengths long.

RADIATION PATTERN.—The characteristic radiation pattern of a given array is obtained at the frequency or band of frequencies at which the system is resonant. The gain and directivity characteristics are lost when the antenna is not used at or near this frequency and the array tunes too sharply. A collinear antenna is more effective than an end-fire array when used off its tuned frequency. This feature is considered when transmission or reception is to be over a wide frequency band. When more than two elements are used, this advantage largely disappears.

LENGTH AND PHASING.—Although the 1/2 wavelength is the basis for the collinear element, you will find that greater lengths are often used. Effective arrays of this type have been constructed in which the elements are 0.7 and even 0.8 wavelength long. This type of array provides efficient operation at more than one frequency or over a wider frequency range. Whatever length is decided upon, all of the elements in a particular array should closely adhere to that length. If elements of different lengths are combined, current phasing and distribution are changed, throwing the system out of balance and seriously affecting the radiation pattern.

Q28. What is the maximum number of elements ordinarily used in a collinear array?

Q29. Why is the number of elements used in a collinear array limited?

Q30. How can the frequency range of a collinear array be increased?

Q31. How is directivity of a collinear array affected when the number of elements is increased?

SPACING.—The lower relative efficiency of collinear arrays of many elements, compared with other multi-element arrays, relates directly to spacing and mutual impedance effects. Mutual impedance is an important factor to be considered when any two elements are parallel and are spaced so that considerable coupling is between them. There is very little mutual impedance between collinear sections. Where impedance does exist, it is caused by the coupling between the ends of adjacent elements. Placing the ends of elements close together is frequently necessary because of construction problems, especially where long lengths of wire are involved.

The effects of spacing and the advantages of proper spacing can be demonstrated by some practical examples. A collinear array consisting of two half-wave elements with 1/4-wavelength spacing between centers has a gain of 1.8 dB. If the ends of these same dipoles are separated so that the distance from center to center is 3/4 wavelengths and they are driven from the same source, the gain increases to approximately 2.9 dB.

A three-dipole array with negligible spacing between elements gives a gain of 3.3 dB. In other words, when two elements are used with wider spacing, the gain obtained is approximately equal to the gain obtainable from three elements with close spacing. The spacing of this array permits simpler construction, since only two dipoles are used. It also allows the antenna to occupy less space. Construction problems usually dictate small-array spacing.

Broadside Arrays

A broadside array is shown in figure 4-26, view A. Physically, it looks somewhat like a ladder. When the array and the elements in it are polarized horizontally, it looks like an upright ladder. When the array is polarized vertically, it looks like a ladder lying on one side (view B). View C is an illustration of the radiation pattern of a broadside array. Horizontally polarized arrays using more than two elements are not common. This is because the requirement that the bottom of the array be a significant distance above the earth presents construction problems. Compared with collinear arrays, broadside arrays tune sharply, but lose efficiency rapidly when not operated on the frequencies for which they are designed.

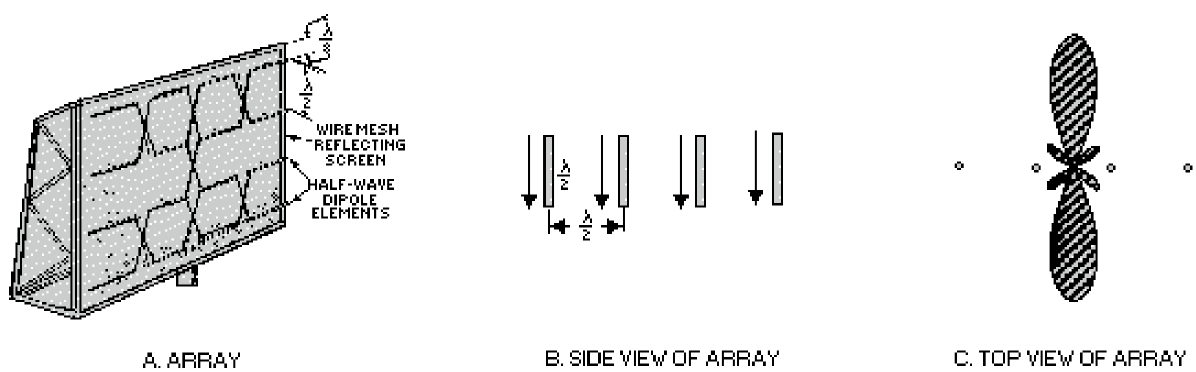


Figure 4-26.—Typical broadside array.

RADIATION PATTERN.—Figure 4-27 shows an end view of two parallel half-wave antennas (A and B) operating in the same phase and located 1/2 wavelength apart. At a point (P) far removed from the antennas, the antennas appear as a single point. Energy radiating toward P from antenna A starts out in phase with the energy radiating from antenna B in the same direction. Propagation from each antenna travels over the same distance to point P, arriving there in phase. The antennas reinforce each other in this direction, making a strong signal available at P. Field strength measured at P is greater than it would be if the total power supplied to both antennas had been fed to a single dipole. Radiation toward point P1 is built up in the same manner.

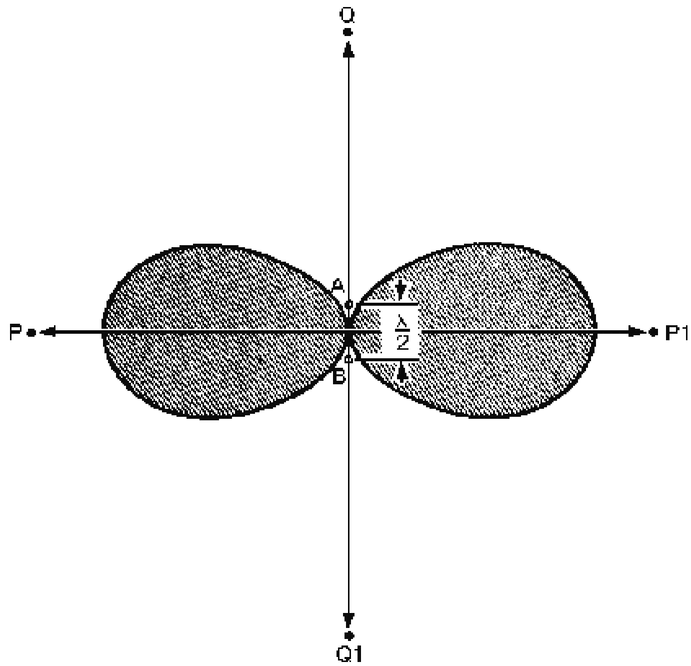


Figure 4-27.—Parallel elements in phase.

Next consider a wavefront traveling toward point Q from antenna B. By the time it reaches antenna A, 1/2 wavelength away, 1/2 cycle has elapsed. Therefore energy from antenna B meets the energy from antenna A 180 degrees out of phase. As a result, the energy moving toward point Q from the two sources cancels. In a like manner, radiation from antenna A traveling toward point Q1 meets and cancels the radiation in the same direction from antenna B. As a result, little propagation takes place in either direction along the QQ1 axis. Most of the energy is concentrated in both directions along the PP1 axis. When both antenna elements are fed from the same source, the result is the basic broadside array.

When more than two elements are used in a broadside arrangement, they are all parallel and in the same plane, as shown in figure 4-26, view B. Current phase, indicated by the arrows, must be the same for all elements. The radiation pattern shown in figure 4-26, view C, is always bi-directional. This pattern is sharper than the one shown in figure 4-27 because of the additional two elements. Directivity and gain depend on the number of elements and the spacing between them.

GAIN AND DIRECTIVITY.—The physical disposition of dipoles operated broadside to each other allows for much greater coupling between them than can occur between collinear elements. Moving the parallel antenna elements closer together or farther apart affects the actual impedance of the entire array and the overall radiation resistance as well. As the spacing between broadside elements increases, the effect on the radiation pattern is a sharpening of the major lobes. When the array consists of only two dipoles spaced exactly 1/2 wavelength apart, no minor lobes are generated at all. Increasing the distance between the elements beyond that point, however, tends to throw off the phase relationship between the original current in one element and the current induced in it by the other element. The result is that, although the major lobes are sharpened, minor lobes are introduced, even with two elements. These, however, are not large enough to be of concern.

If you add the same number of elements to both a broadside array and a collinear array, the gain of the broadside array will be greater. Reduced radiation resistance resulting from the efficient coupling between dipoles accounts for most of this gain. However, certain practical factors limit the number of elements that may be used. The construction problem increases with the number of elements, especially when they are polarized horizontally.

Q32. What is the primary cause of broadside arrays losing efficiency when not operating at their designed frequency?

Q33. When more than two elements are used in a broadside array, how are the elements arranged?

Q34. As the spacing between elements in a broadside array increases, what is the effect on the major lobes?

End-Fire Arrays

An end-fire array looks similar to a broadside array. The ladder-like appearance is characteristic of both (fig. 4-28, view A). The currents in the elements of the end-fire array, however, are usually 180 degrees out of phase with each other as indicated by the arrows. The construction of the end-fire array is like that of a ladder lying on its side (elements horizontal). The dipoles in an end-fire array are closer together (1/8-wavelength to 1/4 -wavelength spacing) than they are for a broadside array.

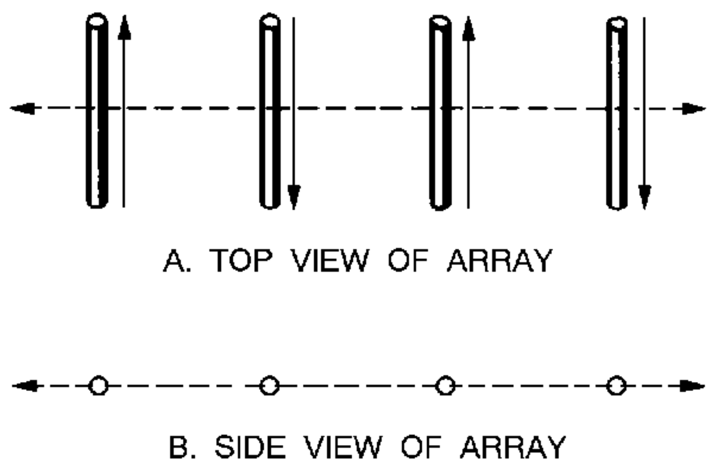


Figure 4-28.—Typical end-fire array.

Closer spacing between elements permits compactness of construction. For this reason an end-fire array is preferred to other arrays when high gain or sharp directivity is desired in a confined space. However, the close coupling creates certain disadvantages. Radiation resistance is extremely low, sometimes as low as 10 ohms, making antenna losses greater. The end-fire array is confined to a single frequency. With changes in climatic or atmospheric conditions, the danger of detuning exists.

RADIATION PATTERN.—The radiation pattern for a pair of parallel half-wave elements fed 180 degrees out of phase is shown in figure 4-29, view A. The elements shown are spaced 1/2 wavelength apart. In practice, smaller spacings are used. Radiation from elements L and M traveling toward point P begins 180 degrees out of phase. Moving the same distance over approximately parallel paths, the respective wavefronts from these elements remain 180 degrees out of phase. In other words, maximum cancellation takes place in the direction of P. The same condition is true for the opposite direction (toward P1). The P to P1 axis is the line of least radiation for the end-fire array.

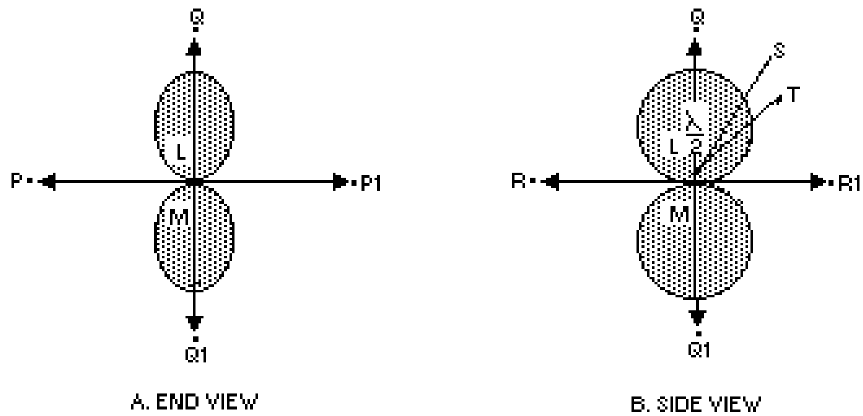


Figure 4-29.—Parallel elements 180 degrees out of phase.

Consider what happens along the QQ1 axis. Energy radiating from element M toward Q reaches element L in about 1/2 cycle (180 degrees) after it leaves its source. Since element L was fed 180 degrees out of phase with element M, the wavefronts are now in the same phase and are both moving toward Q reinforcing each other. Similar reinforcement occurs along the same axis toward Q1. This simultaneous movement towards Q and Q1 develops a bi-directional pattern. This is not always true in end-fire operation. Another application of the end-fire principle is one in which the elements are spaced 1/4 wavelength apart and phased 90 degrees from each other to produce a unidirectional pattern.

In figure 4-29, view A, elements A and B are perpendicular to the plane represented by the page; therefore, only the ends of the antennas appear. In view B the antennas are rotated a quarter of a circle in space around the QQ1 axis so that they are seen in the plane of the elements themselves. Therefore, the PP1 axis, now perpendicular to the page, is not seen as a line. The RR1 axis, now seen as a line, is perpendicular to the PP1 axis as well as to the QQ1 axis. The end-fire array is directional in this plane also, although not quite as sharply. The reason for the greater broadness of the lobes can be seen by following the path of energy radiating from the midpoint of element B toward point S in view B. This energy passes the A element at one end after traveling slightly more than the perpendicular distance between the dipoles. Energy, therefore, does not combine in exact phase toward point S. Although maximum radiation cannot take place in this direction, energy from the two sources combines closely enough in phase to produce considerable reinforcement. A similar situation exists for wavefronts traveling toward T. However, the wider angle from Q to T produces a greater phase difference and results in a decrease in the strength of the combined wave.

Directivity occurs from either one or both ends of the end-fire array, along the axis of the array, as shown by the broken arrows in figure 4-28, view A; hence, the term *end-fire* is used.

The major lobe or lobes occur along the axis of the array. The pattern is sharper in the plane that is at right angles to the plane containing the elements (figure 4-29, view A). If the elements are not exact half-wave dipoles, operation is not significantly affected. However, because of the required balance of phase relationships and critical feeding, the array must be symmetrical. Folded dipoles, such as the one shown in figure 4-20, view A, are used frequently because the impedance at their terminals is higher. This is an effective way of avoiding excessive antenna losses. Another expedient to reduce losses is the use of tubular elements of wide diameter.

GAIN AND DIRECTIVITY.—In end-fire arrays, directivity increases with the addition of more elements and with spacings approaching the optimum. The directive pattern for a two-element, bi-directional system is illustrated in figure 4-29. View A shows radiation along the array axis in a plane perpendicular to the dipoles, and view B shows radiation along the array axis in the plane of the elements. These patterns were developed with a 180-degree phase difference between the elements. Additional elements introduce small, minor lobes.

With a 90-degree phase difference in the energy fed to a pair of end-fire elements spaced approximately 1/4 wavelength apart, unidirectional radiation can be obtained. The pattern perpendicular to the plane of the two elements is shown in figure 4-30, view A. The pattern shown in view B, taken in the same plane, is for a six-element array with 90-degree phasing between adjacent elements. Since both patterns show relative gain only, the increase in gain produced by the six-element array is not evident. End-fire arrays are the only unidirectional arrays wholly made up of driven elements.

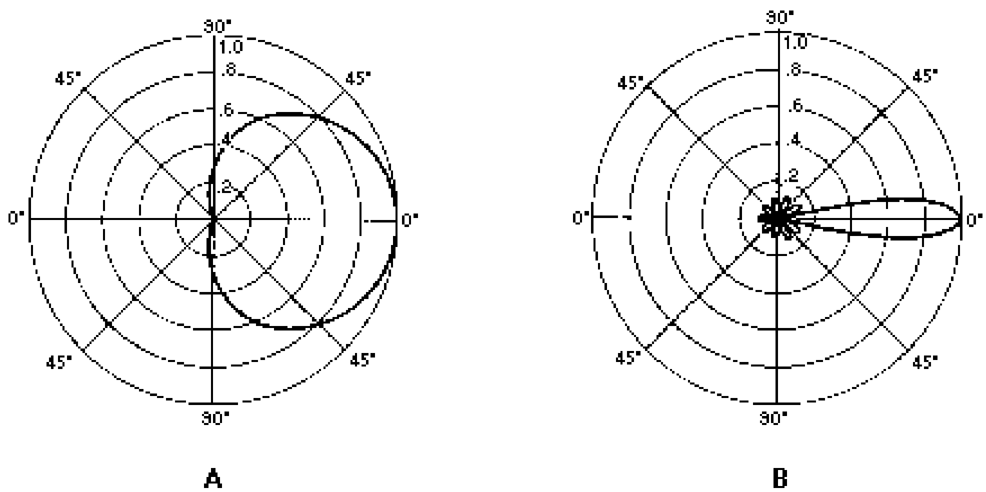


Figure 4-30.—Unidirectional end-fire arrays.

Q35. What are some disadvantages of the end-fire array?

Q36. Where does the major lobe in the end-fire array occur?

Q37. To maintain the required balance of phase relationships and critical feeding, how must the end-fire array be constructed?

Parasitic Arrays

If a small light bulb were placed in the center of a large room, the illumination would be very poor. However, if a reflector were placed behind the bulb, the space in front of the reflector would be brighter and the space behind the reflector would be dimmer. The light rays would be concentrated. Also, if a lens were placed in front of the bulb, the light would be even more concentrated and a very bright spot would appear on the wall in front of the lens. A flashlight is a practical combination of the small bulb, the reflector, and the lens. The energy from an antenna can be reflected and concentrated in a similar manner.

Although we do not usually discuss the gain of a flashlight, we can continue the comparison of an antenna and a flashlight to explain the meaning of antenna gain. Suppose the spot on the wall in front of the flashlight becomes 10 times brighter than it was when only the open bulb was used. The lens and reflector have then produced a 10-fold gain in light. For antennas, the simple half-wave antenna corresponds to the open bulb in the flashlight. Suppose an antenna system concentrates the radio waves so that at a particular point the field strength is 10 times more than it would be at the same distance from a half-wave antenna. The antenna system is then said to have a gain of 10.

Parasitic arrays represent another method of achieving high antenna gains. A parasitic array consists of one or more parasitic elements placed in parallel with each other and, in most cases, at the same line-of-sight level. The parasitic element is fed inductively by radiated energy coming from the driven element connected to the transmitter. It is in NO way connected directly to the driven element.

When the parasitic element is placed so that it radiates away from the driven element, the element is a director. When the parasitic element is placed so that it radiates toward the driven element, the parasitic element is a reflector.

The directivity pattern resulting from the action of parasitic elements depends on two factors. These are (1) the tuning, determined by the length of the parasitic element; and (2) the spacing between the parasitic and driven elements. To a lesser degree, it also depends on the diameter of the parasitic element, since diameter has an effect on tuning.

OPERATION.—When a parasitic element is placed a fraction of a wavelength away from the driven element and is of approximately resonant length, it will re-radiate the energy it intercepts. The parasitic element is effectively a tuned circuit coupled to the driven element, much as the two windings of a transformer are coupled together. The radiated energy from the driven element causes a voltage to be developed in the parasitic element, which, in turn, sets up a magnetic field. This magnetic field extends over to the driven element, which then has a voltage induced in it. The magnitude and phase of the induced voltage depend on the length of the parasitic element and the spacing between the elements. In actual practice the length and spacing are arranged so that the phase and magnitude of the induced voltage cause a unidirectional, horizontal-radiation pattern and an increase in gain.

In the parasitic array in figure 4-31, view A, the parasitic and driven elements are spaced 1/4 wavelength apart. The radiated signal coming from the driven element strikes the parasitic element after 1/4 cycle. The voltage developed in the parasitic element is 180 degrees out of phase with that of the driven element. This is because of the distance traveled (90 degrees) and because the induced current lags the inducing flux by 90 degrees (90 + 90 = 180 degrees). The magnetic field set up by the parasitic element induces a voltage in the driven element 1/4 cycle later because the spacing between the elements is 1/4 wavelength. This induced voltage is in phase with that in the driven element and causes an increase in radiation in the direction indicated in figure 4-31, view A. Since the direction of the radiated energy is stronger in the direction away from the parasitic element (toward the driven element), the parasitic element is called a reflector. The radiation pattern as it would appear if you were looking down on the antenna is shown in view B. The pattern as it would look if viewed from the ends of the elements is shown in view C.

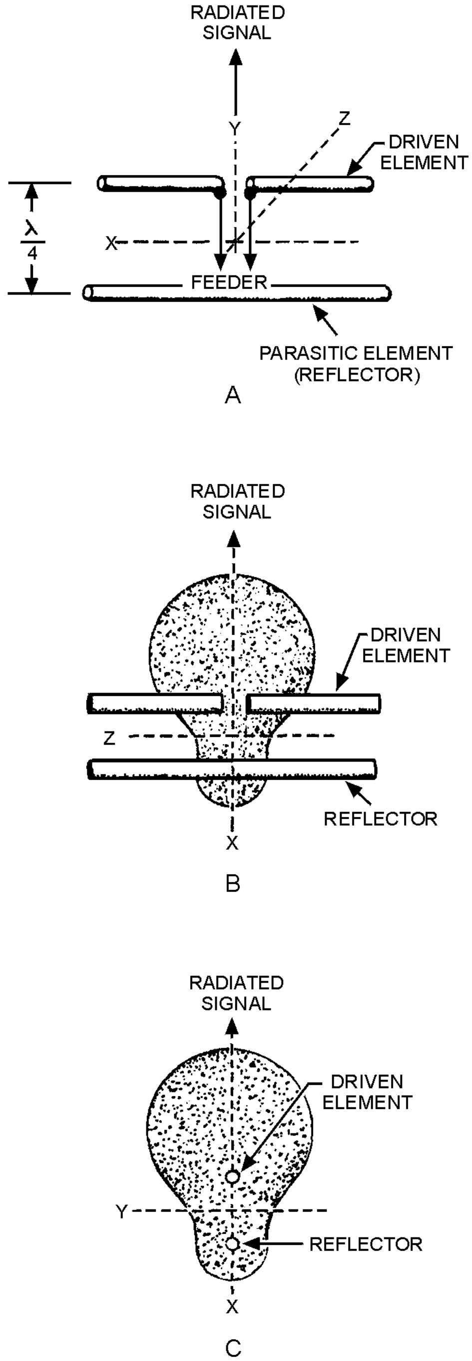


Figure 4-31.—Patterns obtained using a reflector with proper spacing.

Because the voltage induced in the reflector is 180 degrees out of phase with the signal produced at the driven element, a reduction in signal strength exists behind the reflector. Since the magnitude of an induced voltage never quite equals that of the inducing voltage, even in very closely coupled circuits, the energy behind the reflector (minor lobe) is not reduced to 0.

The spacing between the reflector and the driven element can be reduced to about 15 percent of a wavelength. The parasitic element must be made electrically inductive before it will act as a reflector. If this element is made about 5 percent longer than 1/2 wavelength, it will act as a reflector when the spacing is 15 percent of a wavelength.

Changing the spacing and length can change the radiation pattern so that maximum radiation is on the same side of the driven element as the parasitic element. In this instance the parasitic element is called a director.

Combining a reflector and a director with the driven element causes a decrease in back radiation and an increase in directivity. This combination results in the two main advantages of a parasitic array—unidirectivity and increased gain. If the parasitic array is rotated, it can pick up or transmit in different directions because of the reduction of transmitted energy in all but the desired direction. An antenna of this type is called a ROTARY ARRAY. Size for size, both the gain and directivity of parasitic arrays are greater than those of driven arrays. The disadvantage of parasitic arrays is that their adjustment is critical and they do not operate over a wide frequency range.

GAIN AND DIRECTIVITY.—Changing the spacing between either the director or the reflector and the driven element results in a change in the radiation pattern. More gain and directivity are obtained by changing the length of the parasitic elements.

The FRONT-TO-BACK RATIO of an array is the proportion of energy radiated in the principal direction of radiation to the energy radiated in the opposite direction. A high front-to-back ratio is desirable because this means that a minimum amount of energy is radiated in the undesired direction. Since completely suppressing all such radiation is impossible, an infinite ratio cannot be achieved. In actual practice, however, rather high values can be attained. Usually the length and spacing of the parasitic elements are adjusted so that a maximum front-to-back ratio is obtained, rather than maximum gain in the desired direction.

Q38. What two factors determine the directivity pattern of the parasitic array?

Q39. What two main advantages of a parasitic array can be obtained by combining a reflector and a director with the driven element?

Q40. The parasitic array can be rotated to receive or transmit in different directions. What is the name given to such an antenna?

Q41. What are the disadvantages of the parasitic array?

Multielement Parasitic Array

A MULTIELEMENT PARASITIC array is one that contains two or more parasitic elements with the driven element. If the array contains two parasitic elements (a reflector and a director) in addition to the driven element, it is usually known as a THREE-ELEMENT ARRAY. If three parasitic elements are used, the array is known as a FOUR-ELEMENT ARRAY, and so on. Generally speaking, if more parasitic elements are added to a three-element array, each added element is a director. The field behind a reflector is so small that additional reflectors would have little effect on the overall radiation pattern. In radar, from one to five directors are used.

CONSTRUCTION.—The parasitic elements of a multi-element parasitic array usually are positioned as shown in figure 4-32, views A and B. Proper spacings and lengths are determined experimentally. A folded dipole (view B) is often used as the driven element to obtain greater values of radiation resistance.

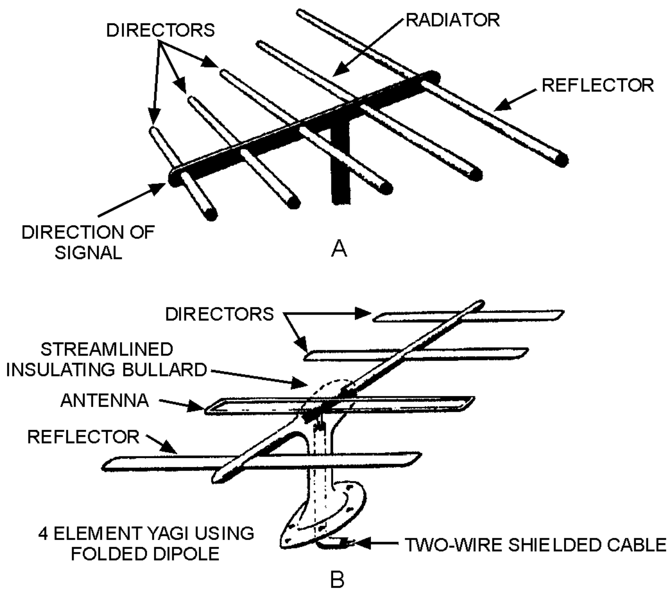


Figure 4-32.—Yagi antenna.

YAGI ANTENNAS.—An example of a multielement parasitic array is the YAGI ANTENNA (figure 4-32, views A and B). The spacings between the elements are not uniform. The radiation from the different elements arrives in phase in the forward direction, but out of phase by various amounts in the other directions.

The director and the reflector in the Yagi antenna are usually welded to a conducting rod or tube at their centers. This support does not interfere with the operation of the antenna. Since the driven element is center-fed, it is not welded to the supporting rod. The center impedance can be increased by using a folded dipole as the driven element.

The Yagi antenna shown in figure 4-32, view A, has three directors. In general, the greater number of parasitic elements used, the greater the gain. However, a greater number of such elements causes the array to have a narrower frequency response as well as a narrower beamwidth. Therefore, proper adjustment of the antenna is critical. The gain does not increase directly with the number of elements used. For example, a three-element Yagi array has a relative power gain of 5 dB. Adding another director results in a 2 dB increase. Additional directors have less and less effect.

A typical Yagi array used for receiving and transmitting energy is shown with a support frame in figure 4-33. This antenna is used by the military services. It operates at frequencies of from 12 to 50 megahertz and consists of two separate arrays (one high-frequency and one low-frequency antenna array) mounted on one frame. The various elements are indicated in the figure. The high-frequency (hf) array consists of one reflector, one driven element, and two directors; the low-frequency (lf) array has the same arrangement with one less director. The lengths of the elements in the high-frequency array are shorter than those in the low-frequency array. The physical lengths of the elements in the individual arrays are equal, but the electrical lengths can be varied by means of the tuning stubs at the center of the elements. The array can be rotated in any desired direction by a remotely controlled, electrically driven, antenna rotator.

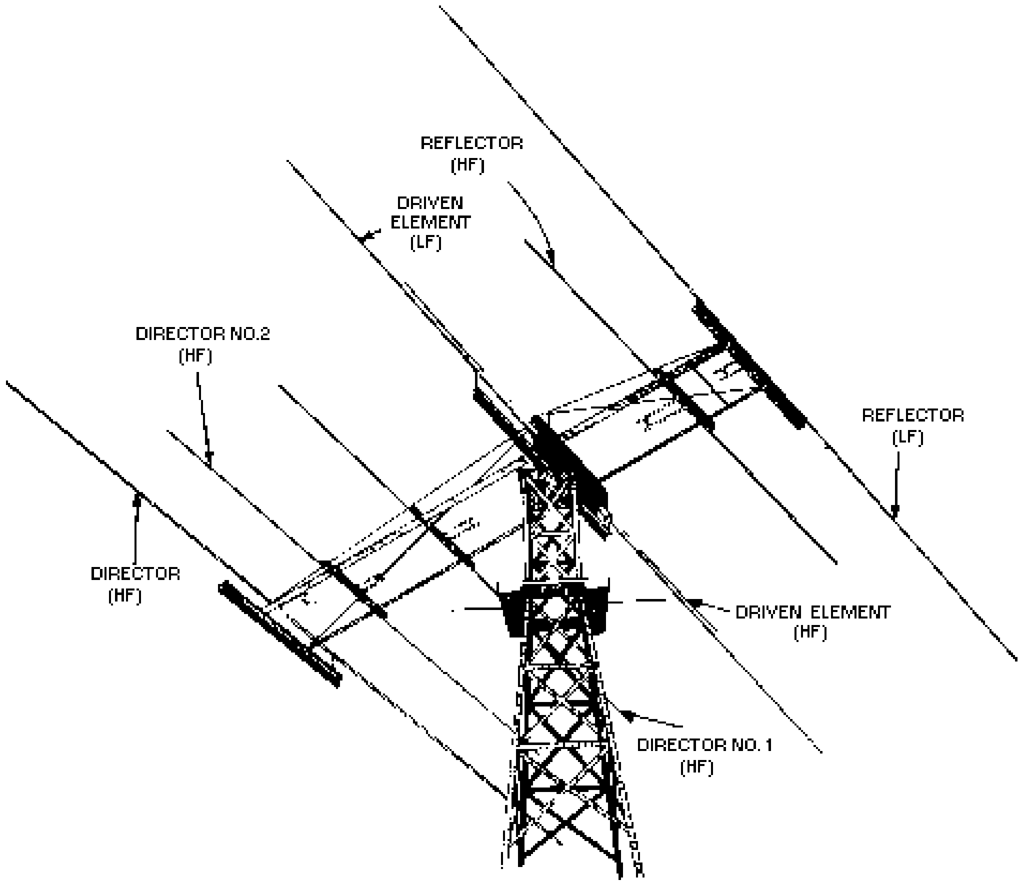


Figure 4-33.—A typical parasitic array used for transmitting and receiving.

Q42. What is the advantage of adding parasitic elements to a Yagi array?

Q43. The Yagi antenna is an example of what type of array?

SPECIAL ANTENNAS

In this section we will cover some special communications and radar antennas. Some of these antennas we touch on briefly since they are covered thoroughly in other courses.

Previously discussed antennas operate with standing waves of current and voltage along the wires. This section deals principally with antenna systems in which the current is practically uniform in all parts of the antenna. In its basic form, such an antenna consists of a single wire grounded at the far end through a resistor. The resistor has a value equal to the characteristic impedance of the antenna. This termination, just as in the case of an ordinary transmission line, eliminates standing waves. The current, therefore, decreases uniformly along the wire as the terminated end is approached. This decrease is caused by the loss of energy through radiation. The energy remaining at the end of the antenna is dissipated in the terminating resistor. For such an antenna to be a good radiator, its length must be fairly long. Also, the wire must not be too close to the ground. The return path through the ground will cause cancellation of the radiation. If the wire is sufficiently long, it will be practically nonresonant over a wide range of operating frequencies.

LONG-WIRE ANTENNA

A LONG-WIRE ANTENNA is an antenna that is a wavelength or longer at the operating frequency. In general, the gain achieved with long-wire antennas is not as great as the gain obtained from the multielement arrays studied in the previous section. But the long-wire antenna has advantages of its own. The construction of long-wire antennas is simple, both electrically and mechanically, with no particularly critical dimensions or adjustments. The long-wire antenna will work well and give satisfactory gain and directivity over a frequency range up to twice the value for which it was cut. In addition, it will accept power and radiate it efficiently on any frequency for which its overall length is not less than approximately 1/2 wavelength. Another factor is that long-wire antennas have directional patterns that are sharp in both the horizontal and vertical planes. Also, they tend to concentrate the radiation at the low vertical angles. Another type of long-wire antenna is the BEVERAGE ANTENNA, also called a WAVE ANTENNA. It is a horizontal, long-wire antenna designed especially for the reception and transmission of low-frequency, vertically polarized ground waves. It consists of a single wire, two or more wavelengths long, supported 3 to 6 meters above the ground, and terminated in its characteristic impedance, as shown in figure 4-34.

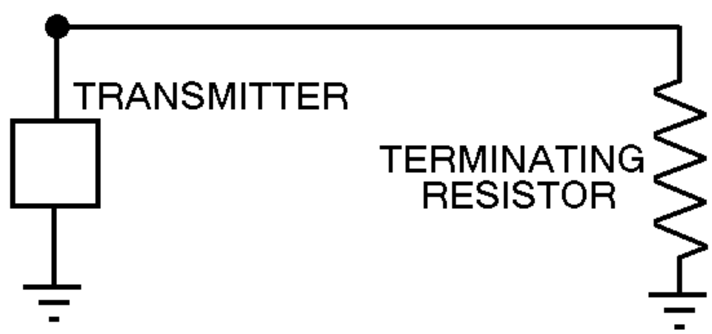


Figure 4-34.—Beverage antenna.

Q44. To radiate power efficiently, a long-wire antenna must have what minimum overall length?

Q45. What is another name for the Beverage antenna?

V ANTENNA

A V ANTENNA is a bi-directional antenna used widely in military and commercial communications. It consists of two conductors arranged to form a V. Each conductor is fed with currents of opposite polarity.

The V is formed at such an angle that the main lobes reinforce along the line bisecting the V and make a very effective directional antenna (see figure 4-35). Connecting the two-wire feed line to the apex of the V and exciting the two sides of the V 180 degrees out of phase cause the lobes to add along the line of the bisector and to cancel in other directions, as shown in figure 4-36. The lobes are designated 1, 2, 3, and 4 on leg AA', and 5, 6, 7, and 8 on leg BB'. When the proper angle between AA' and BB' is chosen, lobes 1 and 4 have the same direction and combine with lobes 7 and 6, respectively. This combination of two major lobes from each leg results in the formation of two stronger lobes, which lie along an imaginary line bisecting the enclosed angle. Lobes 2, 3, 5, and 8 tend to cancel each other, as do the smaller lobes, which are approximately at right angles to the wire legs of the V. The resultant waveform pattern is shown at the right of the V antenna in figure 4-36.

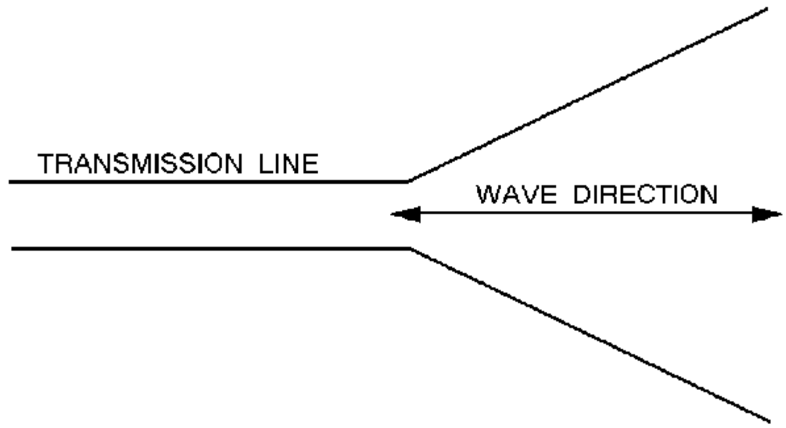


Figure 4-35.—Basic V antenna.

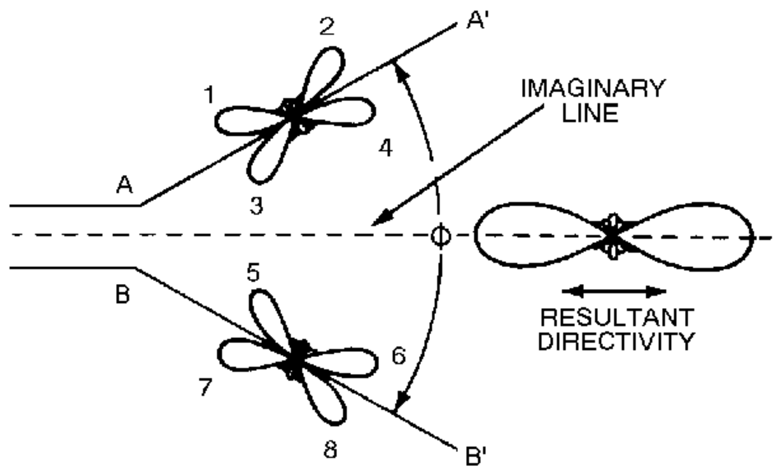


Figure 4-36.—Formation of directional radiation pattern from a resonant V antenna.

Q46. What is the polarity of the currents that feed the V antenna?

RHOMBIC ANTENNA

The highest development of the long-wire antenna is the RHOMBIC ANTENNA (see figure 4-37). It consists of four conductors joined to form a rhombus, or diamond shape. The antenna is placed end to end and terminated by a noninductive resistor to produce a uni-directional pattern. A rhombic antenna can be made of two obtuse-angle V antennas that are placed side by side, erected in a horizontal plane, and terminated so the antenna is nonresonant and unidirectional.

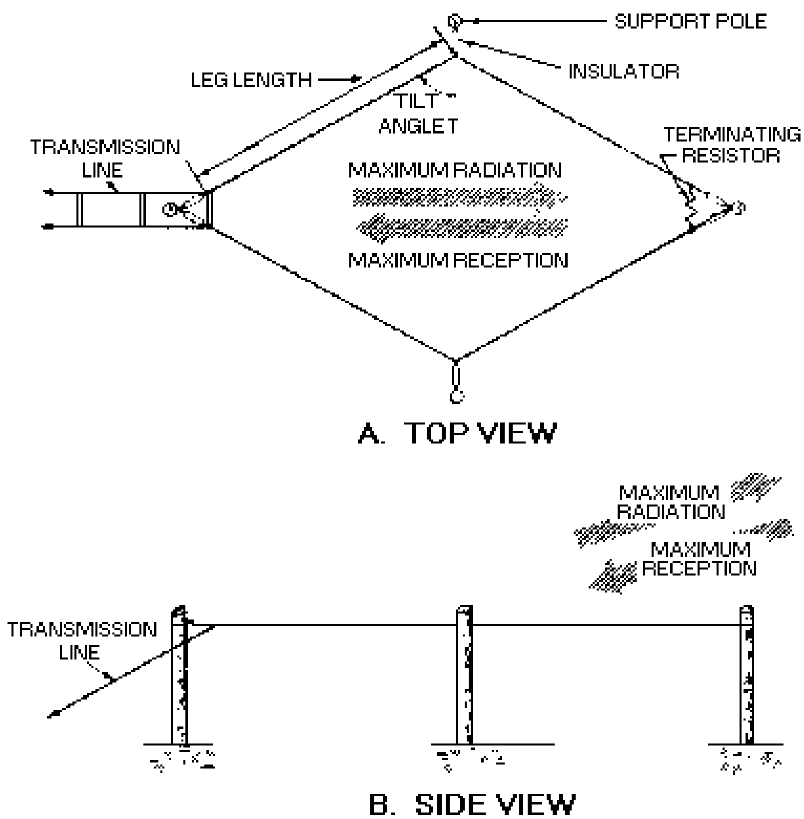


Figure 4-37.—Basic rhombic antenna.

The rhombic antenna is WIDELY used for long-distance, high-frequency transmission and reception. It is one of the most popular fixed-station antennas because it is very useful in point-to-point communications.

Advantages

The rhombic antenna is useful over a wide frequency range. Although some changes in gain, directivity, and characteristic impedance do occur with a change in operating frequency, these changes are small enough to be neglected.

The rhombic antenna is much easier to construct and maintain than other antennas of comparable gain and directivity. Only four supporting poles of common heights from 15 to 20 meters are needed for the antenna.

The rhombic antenna also has the advantage of being noncritical as far as operation and adjustment are concerned. This is because of the broad frequency characteristics of the antenna.

Still another advantage is that the voltages present on the antenna are much lower than those produced by the same input power on a resonant antenna. This is particularly important when high transmitter powers are used or when high-altitude operation is required.

Disadvantages

The rhombic antenna is not without its disadvantages. The principal one is that a fairly large antenna site is required for its erection. Each leg is made at least 1 or 2 wavelengths long at the lowest operating frequency. When increased gain and directivity are required, legs of from 8 to 12 wavelengths are used. These requirements mean that high-frequency rhombic antennas have wires of several hundred feet in length. Therefore, they are used only when a large plot of land is available.

Another disadvantage is that the horizontal and vertical patterns depend on each other. If a rhombic antenna is made to have a narrow horizontal beam, the beam is also lower in the vertical direction. Therefore, obtaining high vertical-angle radiation is impossible except with a very broad horizontal pattern and low gain. Rhombic antennas are used, however, for long-distance sky wave coverage at the high frequencies. Under these conditions low vertical angles of radiation (less than 20 degrees) are desirable. With the rhombic antenna, a considerable amount of the input power is dissipated uselessly in the terminating resistor. However, this resistor is necessary to make the antenna unidirectional. The great gain of the antenna more than makes up for this loss.

Radiation Patterns

Figure 4-38 shows the individual radiation patterns produced by the four legs of the rhombic antenna and the resultant radiation pattern. The principle of operation is the same as for the V and the half-rhombic antennas.

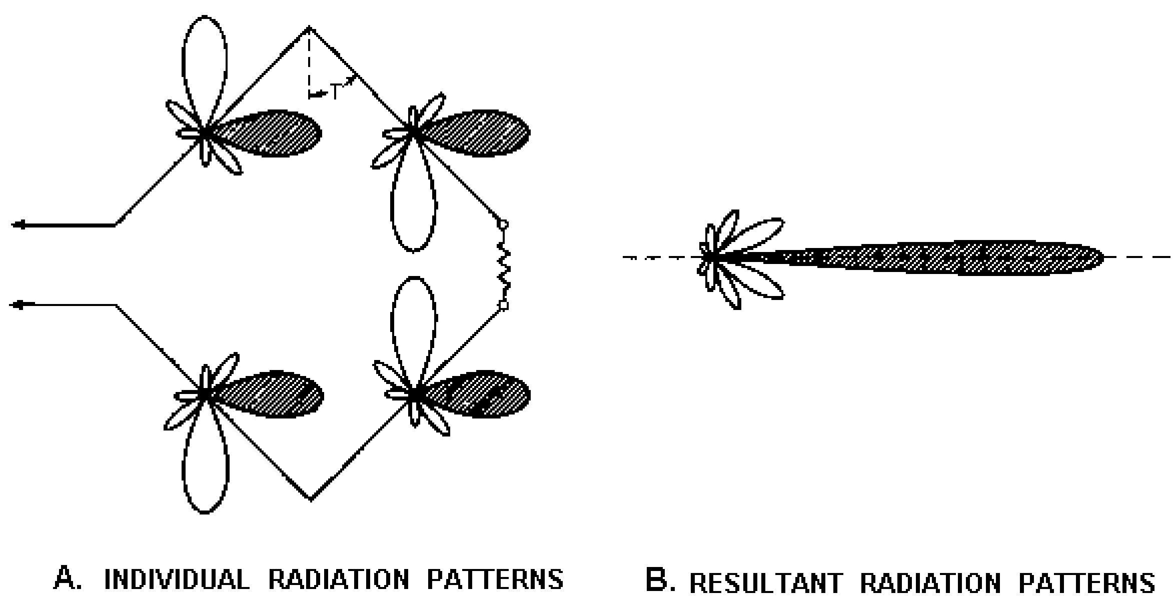


Figure 4-38.—Formation of a rhombic antenna beam.

Terminating Resistor

The terminating resistor plays an important part in the operation of the rhombic antenna. Upon it depend the unidirectivity of the antenna and the lack of resonance effects. An antenna should be properly terminated so it will have a constant impedance at its input. Terminating the antenna properly will also allow it to be operated over a wide frequency range without the necessity for changing the coupling adjustments at the transmitter. Discrimination against signals coming from the rear is of great importance for reception. The reduction of back radiation is perhaps of lesser importance for transmission. When an antenna is terminated with resistance, the energy that would be radiated backward is absorbed in the resistor.

Q47. What is the main disadvantage of the rhombic antenna?

TURNSTILE ANTENNA

The TURNSTILE ANTENNA is one of the many types that has been developed primarily for omnidirectional vhf communications. The basic turnstile consists of two horizontal half-wave antennas mounted at right angles to each other in the same horizontal plane. When these two antennas are excited with equal currents 90 degrees out of phase, the typical figure-eight patterns of the two antennas merge to produce the nearly circular pattern shown in figure 4-39, view A. Pairs of such antennas are frequently stacked, as shown in figure 4-40. Each pair is called a BAY. In figure 4-40 two bays are used and are spaced 1/2 wavelength apart, and the corresponding elements are excited in phase. These conditions cause a part of the vertical radiation from each bay to cancel that of the other bay. This results in a decrease in energy radiated at high vertical angles and increases the energy radiated in the horizontal plane. Stacking a number of bays can alter the vertical radiation pattern, causing a substantial gain in a horizontal direction without altering the overall horizontal directivity pattern. Figure 4-39, view B, compares the circular vertical radiation pattern of a single-bay turnstile with the sharp pattern of a four-bay turnstile array. A three-dimensional radiation pattern of a four-bay turnstile antenna is shown in figure 4-39, view C.

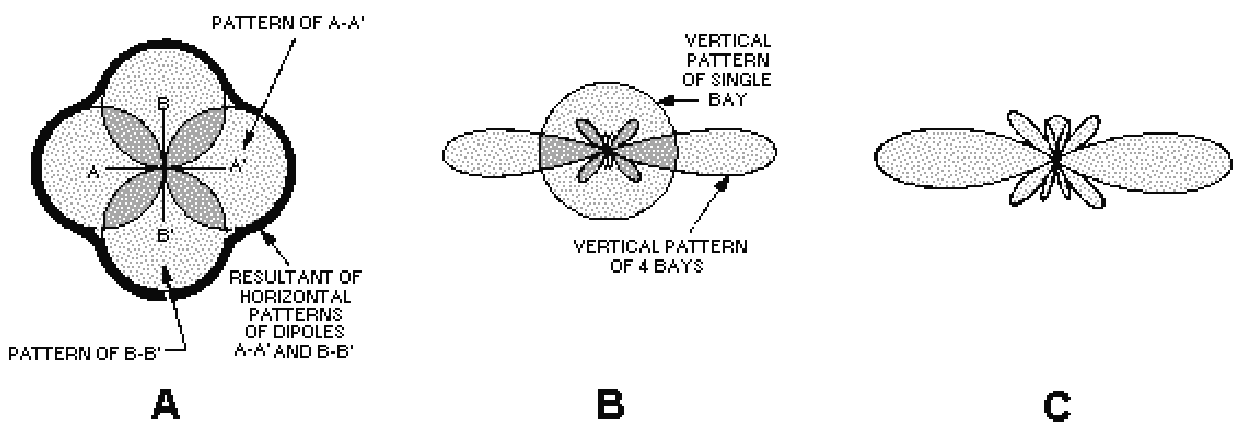


Figure 4-39.—Turnstile antenna radiation pattern.

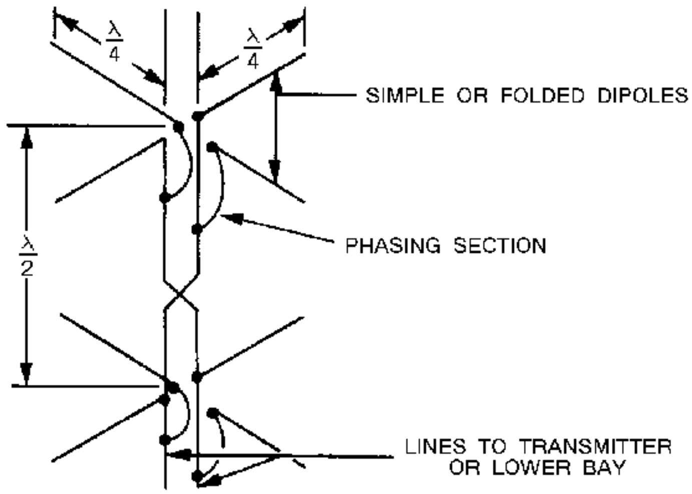


Figure 4-40.—Stacked turnstile antennas.

GROUND-PLANE ANTENNA

A vertical quarter-wave antenna several wavelengths above ground produces a high angle of radiation that is very undesirable at vhf and uhf frequencies. The most common means of producing a low angle of radiation from such an antenna is to work the radiator against a simulated ground called a GROUND PLANE. A simulated ground may be made from a large metal sheet or several wires or rods radiating from the base of the radiator. An antenna so constructed is known as a GROUND-PLANE ANTENNA. Two ground-plane antennas are shown in figure 4-41, views A and B.

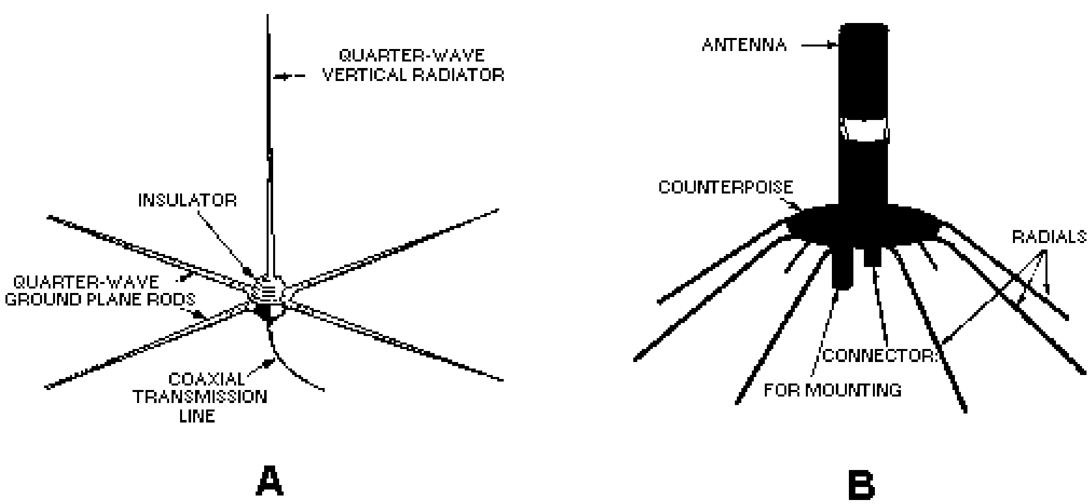


Figure 4-41.—Ground-plane antennas.

CORNER REFLECTOR

When a unidirectional radiation pattern is desired, it can be obtained by the use of a corner reflector with a half-wave dipole. A CORNER-REFLECTOR ANTENNA is a half-wave radiator with a reflector. The reflector consists of two flat metal surfaces meeting at an angle immediately behind the radiator. In other words, the radiator is set in the plane of a line bisecting the corner angle formed by the reflector sheets. The construction of a corner reflector is shown in figure 4-42. Corner-reflector antennas are mounted with the radiator and the reflector in the horizontal position when horizontal polarization is desired. In such cases the radiation pattern is very narrow in the vertical plane, with maximum signal being radiated in line with the bisector of the corner angle. The directivity in the horizontal plane is approximately the same as for any half-wave radiator having a single-rod type reflector behind it. If the antenna is mounted with the radiator and the corner reflector in the vertical position, as shown in view A, maximum radiation is produced in a very narrow horizontal beam. Radiation in a vertical plane will be the same as for a similar radiator with a single-rod type reflector behind it.

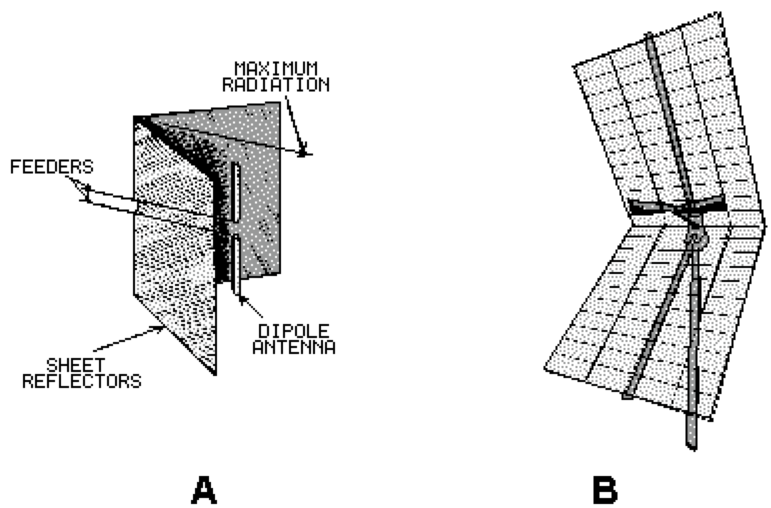


Figure 4-42.—Corner-reflector antennas.

Q48. What is the primary reason for the development of the turnstile antenna?

RF SAFETY PRECAUTIONS

Although electromagnetic radiation from transmission lines and antennas is usually of insufficient strength to electrocute personnel, it can lead to other accidents and compound injuries. Voltages may be induced in ungrounded metal objects, such as wire guys, wire cable (hawser), hand rails, or ladders. If you come in contact with these objects, you could receive a shock or rf burn. This shock can cause you to jump or fall into nearby mechanical equipment or, when working aloft, to fall from an elevated work area. Take care to ensure that all transmission lines or antennas are deenergized before working near or on them.

Either check or have someone check all guys, cables, rails, and ladders around your work area for rf shock dangers. Use working aloft "chits" and safety harnesses for your own safety. Signing a "working aloft chit" signifies that all equipment is in a nonradiating status. The person who signs the chit should ensure that no rf danger exists in areas where you or other personnel will be working.

Nearby ships or parked aircraft are another source of rf energy that you must consider when you check a work area for safety. Combustible materials can be ignited and cause severe fires from arcs or heat generated by rf energy. Also, rf radiation can detonate ordnance devices by inducing currents in the internal wiring of the devices or in the external test equipment or leads connected to them.

ALWAYS obey rf radiation warning signs and keep a safe distance from radiating antennas. The six types of warning signs for rf radiation hazards are shown in figure 4-43.

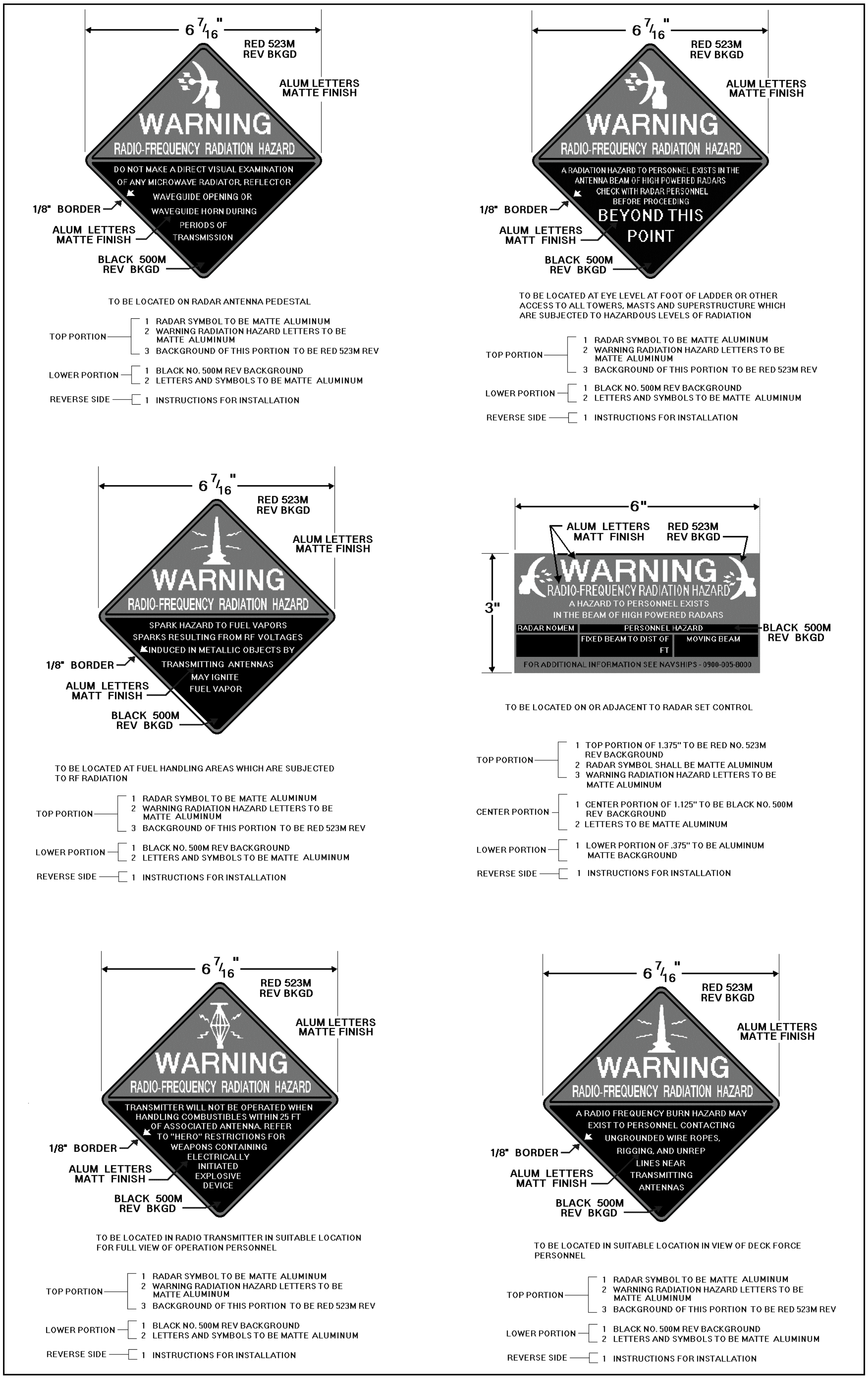


Figure 4-43.—Examples of rf radiation warning signs.

RF BURNS

Close or direct contact with rf transmission lines or antennas may result in rf burns. These are usually deep, penetrating, third-degree burns. To heal properly, these burns must heal from the inside to the skin's surface. To prevent infection, you must give proper attention to <u>all</u> rf burns, including the small "pinhole" burns. Petrolatum gauze can be used to cover these burns temporarily, before the injured person reports to medical facilities for further treatment.

DIELECTRIC HEATING

DIELECTRIC HEATING is the heating of an insulating material by placing it in a high-frequency electric field. The heat results from internal losses during the rapid reversal of polarization of molecules in the dielectric material.

In the case of a human in an rf field, the body acts as a dielectric. If the power in the rf field exceeds 10 milliwatts per centimeter, a person in that field will have a noticeable rise in body temperature. The eyes are highly susceptible to dielectric heating. For this reason, you should not look directly into devices radiating rf energy. The vital organs of the body also are susceptible to dielectric heating. For your own safety, you must NOT stand directly in the path of rf radiating devices.

PRECAUTIONS WHEN WORKING ALOFT

When radio or radar antennas are energized by transmitters, you must not go aloft unless advance tests show that little or no danger exists. A casualty can occur from even a small spark drawn from a charged piece of metal or rigging. Although the spark itself may be harmless, the "surprise" may cause you to let go of the antenna involuntarily and you may fall. There is also a shock hazard if nearby antennas are energized.

Rotating antennas also might cause you to fall when you are working aloft. Motor safety switches controlling the motion of rotating antennas must be tagged and locked open before you go aloft near such antennas.

When working near a stack, you should draw and wear the recommended oxygen breathing apparatus. Among other toxic substances, stack gas contains carbon monoxide. Carbon monoxide is too unstable to build up to a high concentration in the open, but prolonged exposure to even small quantities is dangerous.

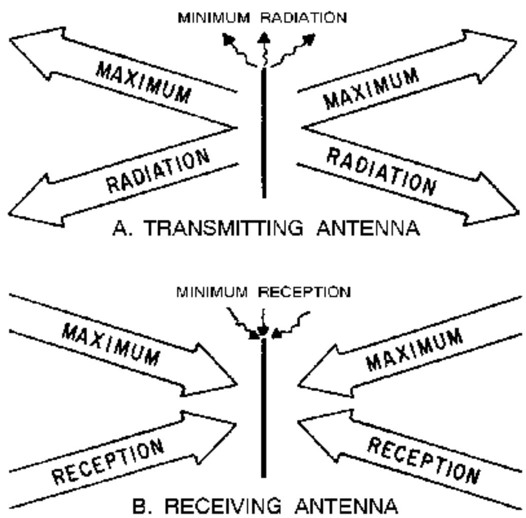
SUMMARY

This chapter has presented information on the various types of antennas. The information that follows summarizes the important points of this chapter.

An **ANTENNA** is a conductor, or system of conductors, that radiates or receives energy in the form of electromagnetic waves.

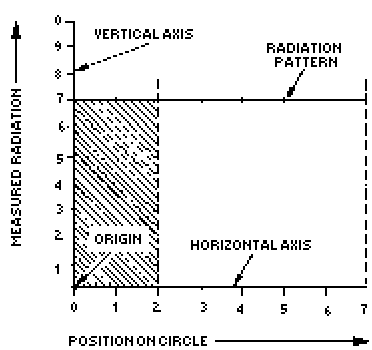
HERTZ (half-wave) and **MARCONI** (quarter-wave) are the two basic classifications of antennas.

RECIPROCITY of antennas means that the various properties of the antenna apply equally to transmitting and receiving.

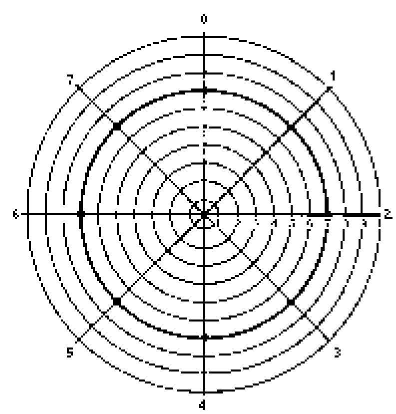


RADIATION RESISTANCE is the amount of resistance which, if inserted in place of the antenna, would consume the same amount of power that is actually radiated by the antenna.

RADIATION PATTERNS can be plotted on a rectangular- or polar-coordinate graph. These patterns are a measurement of the energy leaving an antenna.

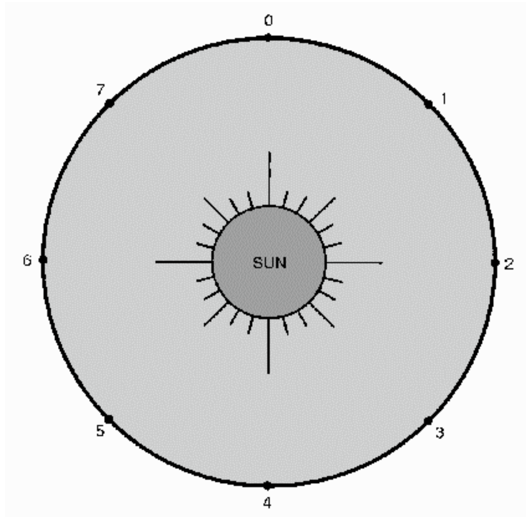


A. RECTANGULAR-COORDINATE GRAPH

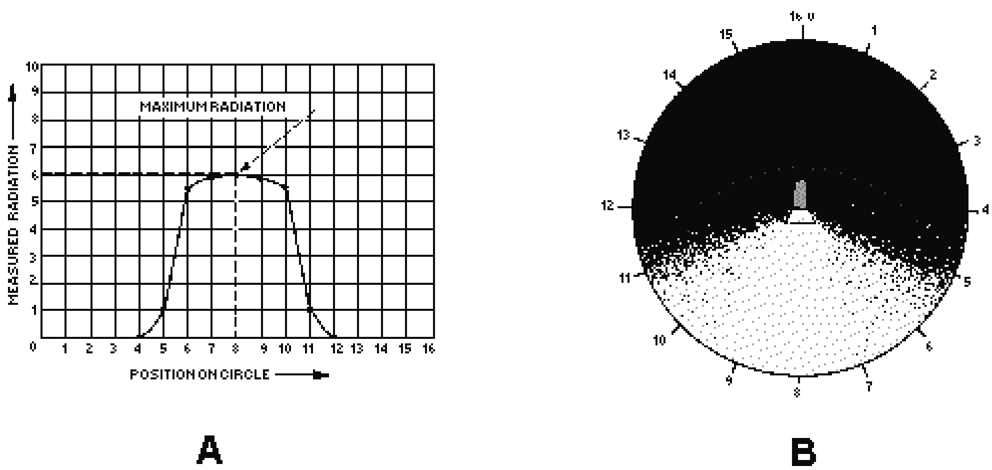


B. POLAR-COORDINATE GRAPH

An **ISOTROPIC RADIATOR** radiates energy equally in all directions.

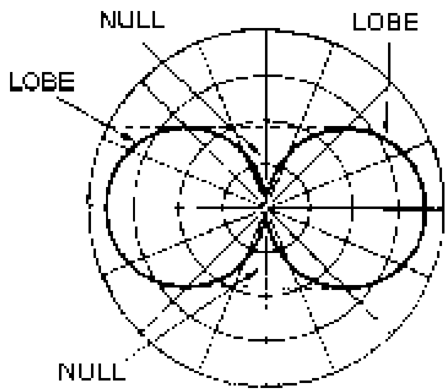


An **ANISOTROPIC RADIATOR** radiates energy directionally.

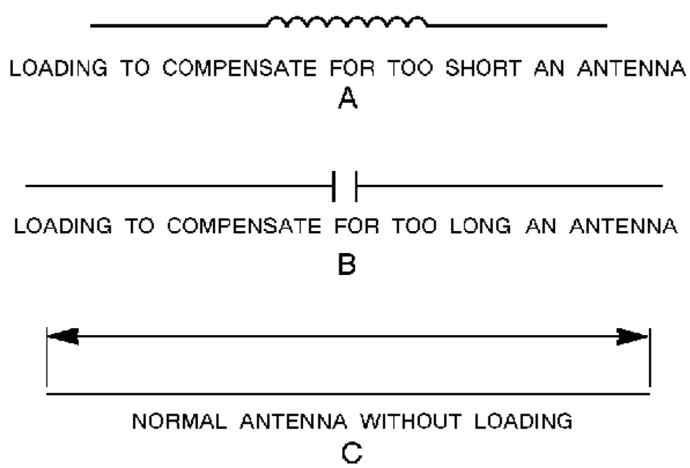


A **LOBE** is the area of a radiation pattern that is covered by radiation.

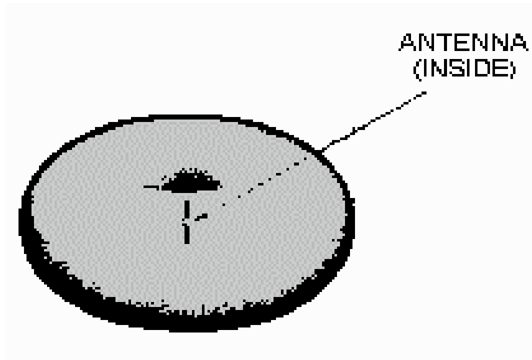
A **NULL** is the area of a radiation pattern that has minimum radiation.



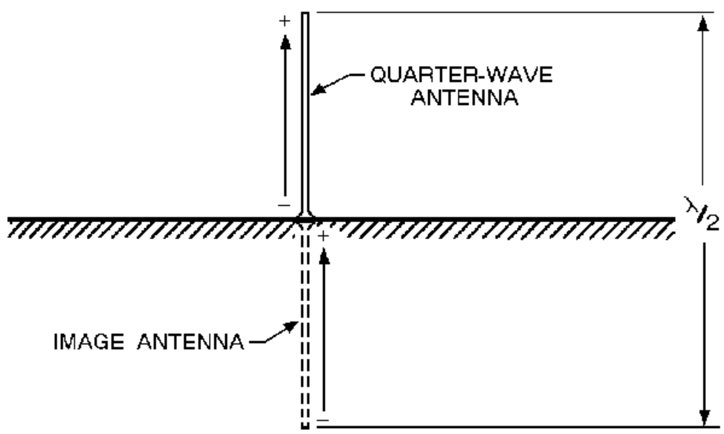
ANTENNA LOADING is the method used to change the electrical length of an antenna. This keeps the antenna in resonance with the applied frequency. It is accomplished by inserting a variable inductor or capacitor in series with the antenna.



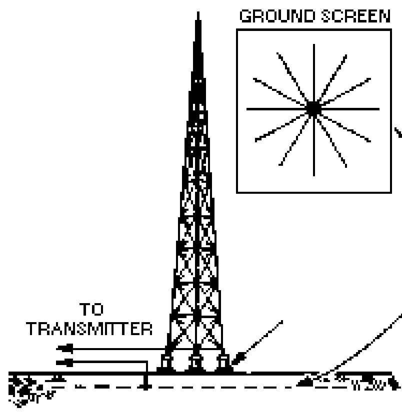
A **HALF-WAVE ANTENNA (Hertz)** consists of two lengths of rod or tubing, each a quarter-wave long at a certain frequency, which radiates a doughnut pattern.

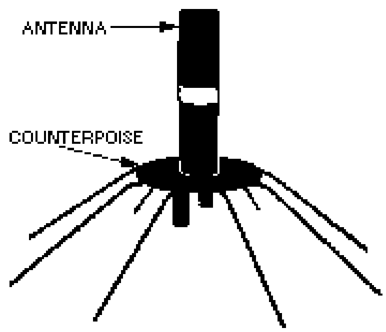


A **QUARTER-WAVE ANTENNA (Marconi)** is a half-wave antenna cut in half with one end grounded. The ground furnishes the missing half of the antenna.

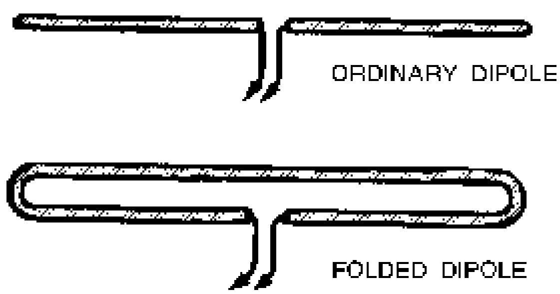


The **GROUND SCREEN** and the **COUNTERPOISE** are used to reduce losses caused by the ground in the immediate vicinity of the antenna. The ground screen is buried below the surface of the earth. The counterpoise is installed above the ground.





The **FOLDED DIPOLE** consists of a dipole radiator, which is connected in parallel at its ends to a half-wave radiator.



AN **ARRAY** is a combination of half-wave elements operating together as a single antenna. It provides more gain and greater directivity than single element antennas.

A **DRIVEN ARRAY** derives its power directly from the source.

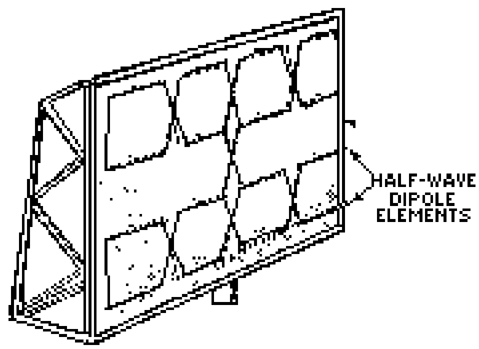
A **PARASITIC ARRAY** derives its power by coupling the energy from other elements of the antenna.

The **BIDIRECTIONAL ARRAY** radiates energy equally in two opposing directions.

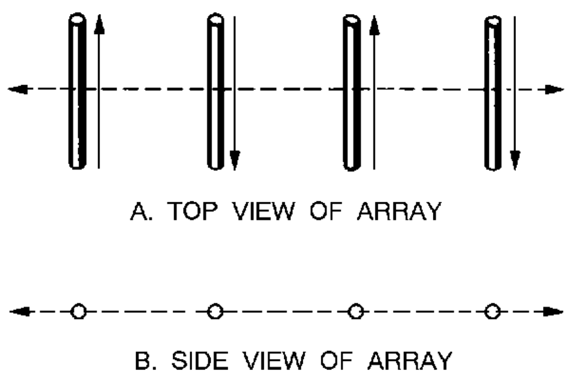
The **UNIDIRECTIONAL ARRAY** radiates energy efficiently in a single direction.

The **COLLINEAR ARRAY** has elements in a straight line. Maximum radiation occurs at right angles to this line.

The **BROADSIDE ARRAY** has elements parallel and in the same plane. Maximum radiation develops in the plane at right angles to the plane of the elements.



The **END-FIRE ARRAY** has elements parallel to each other and in the same plane. Maximum radiation occurs along the axis of the array.

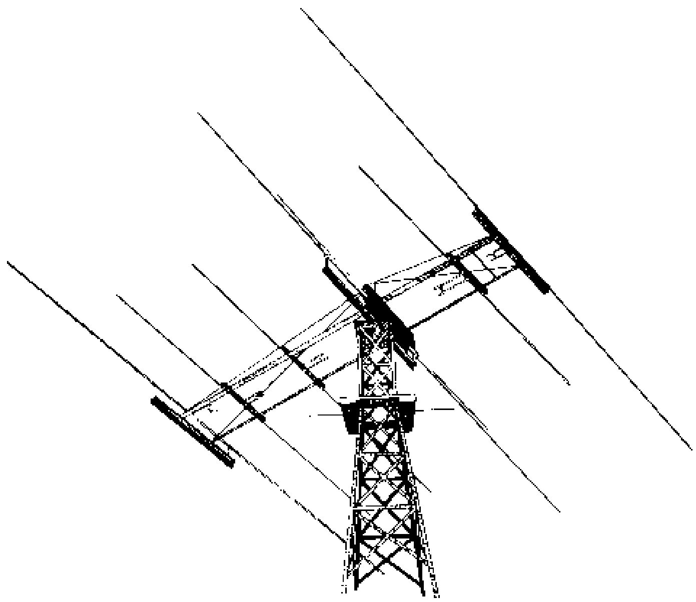


MATCHING STUBS are used between elements to maintain current in the proper phase.

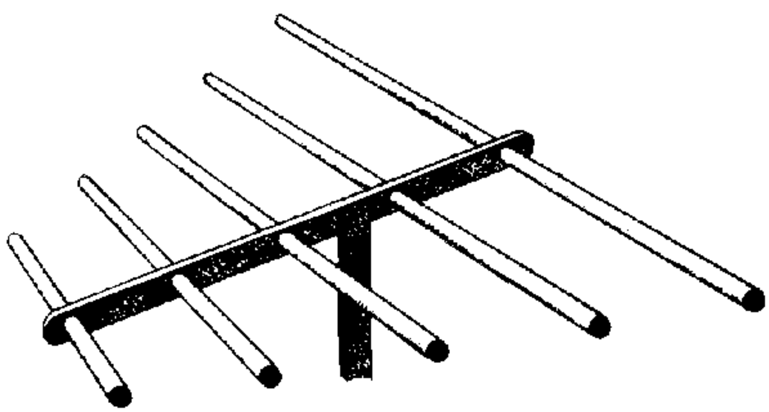
The **GAIN OF A COLLINEAR ANTENNA** is greatest when the elements are spaced from 0.4 to 0.5 wavelength apart or when the number of elements is increased.

The **OPTIMUM GAIN OF A BROADSIDE ARRAY** is obtained when the elements are spaced 0.65 wavelength apart.

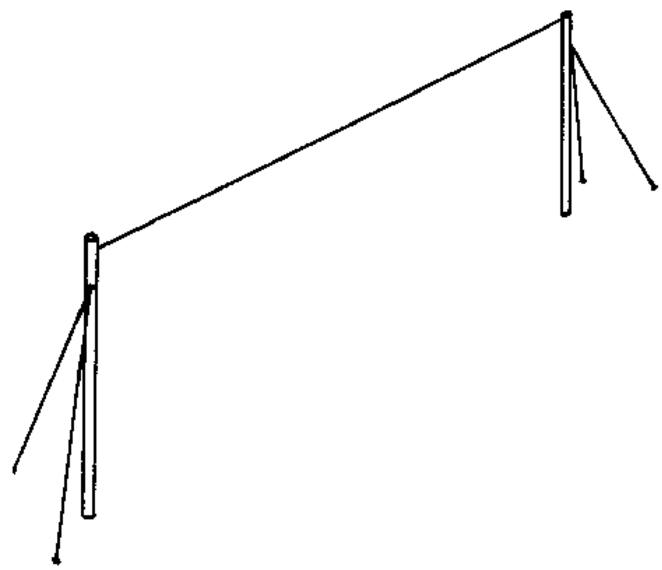
A **PARASITIC ARRAY** consists of one or more parasitic elements with a driven element. The amount of power gain and directivity depends on the lengths of the parasitic elements and the spacing between them.



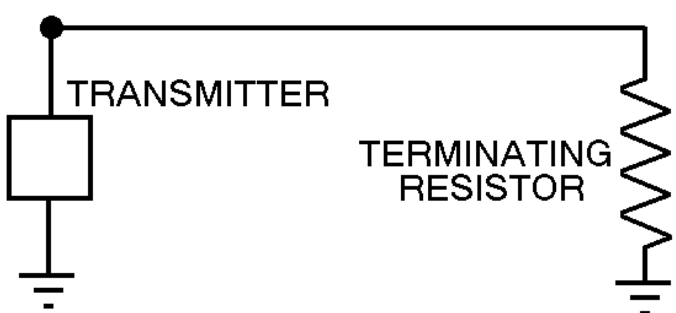
MULTIELEMENT ARRAYS, such as the YAGI, have a narrow frequency response as well as a narrow beamwidth.



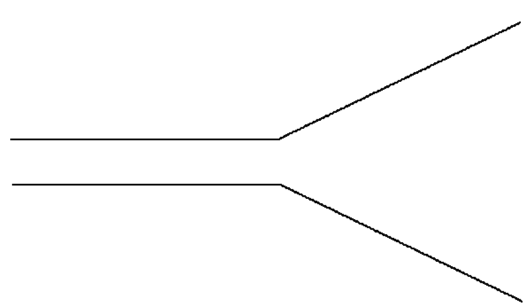
A **LONG-WIRE ANTENNA** is an antenna that is a wavelength or more long at the operating frequency. These antennas have directive patterns that are sharp in both the horizontal and vertical planes.



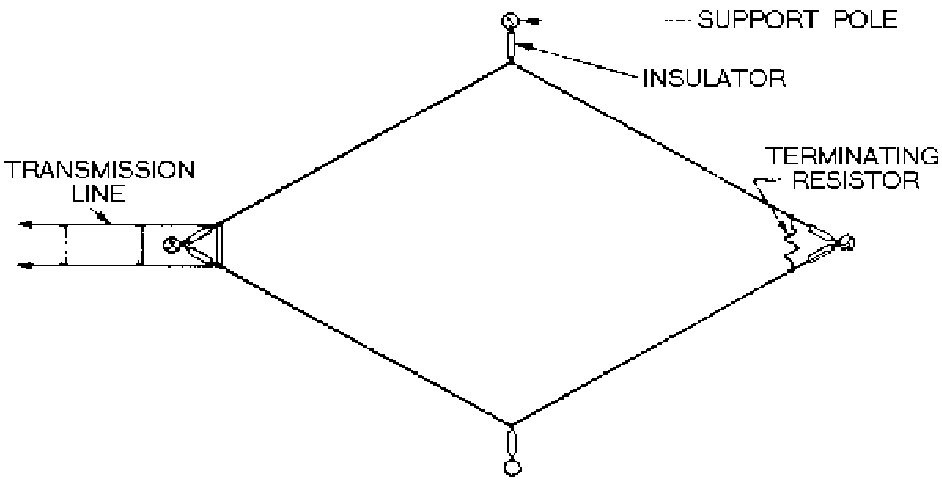
BEVERAGE ANTENNAS consist of a single wire that is two or more wavelengths long.



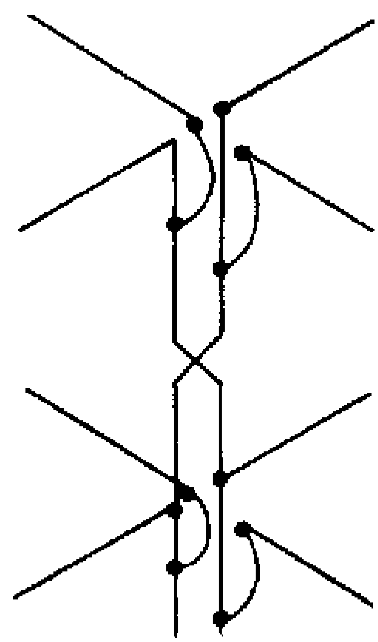
A **V ANTENNA** is a bi-directional antenna consisting of two horizontal, long wires arranged to form a V.



The **RHOMBIC ANTENNA** uses four conductors joined to form a rhombus shape. This antenna has a wide frequency range, is easy to construct and maintain, and is noncritical as far as operation and adjustment are concerned.



The **TURNSTILE ANTENNA** consists of two horizontal, half-wire antennas mounted at right angles to each other.



ANSWERS TO QUESTIONS Q1. THROUGH Q48.

A1. Half-wave (Hertz) and quarter-wave (Marconi).

A2. Coupling device, feeder, and antenna.

A3. Frequency of operation of the transmitter, amount of power to be radiated, and general direction of the receiving set.

A4. *One-half the wavelength.*

A5. *Current and voltage loops.*

A6. *Current and voltage nodes.*

A7. *Reciprocity of antennas.*

A8. *Electric (E) field.*

A9. *Circular polarization.*

A10. *Vertical polarization.*

A11. *Less interference is experienced by man-made noise sources.*

A12. *Vertical polarization.*

A13. *73 ohms.*

A14. *Anisotropic radiator.*

A15. *Isotropic radiator.*

A16. *Anisotropic radiator.*

A17. *Dipole, doublet and Hertz.*

A18. *Nondirectional.*

A19. *Vertical plane.*

A20. *The pattern would flatten.*

A21. *To connect one end through a capacitor to the final output stage of the transmitter.*

A22. *A circular radiation pattern in the horizontal plane, or same as a half wave.*

A23. *It is composed of a series of conductors arranged in a radial pattern and buried 1 to 2 feet below the ground.*

A24. *Nine times the feed-point impedance.*

A25. *Folded dipole.*

A26. *To produce desired phase relationship between connected elements.*

A27. *Major lobes have the greatest amount of radiation.*

A28. *Four.*

A29. *As more elements are added, an unbalanced condition in the system occurs which impairs efficiency.*

A30. *By increasing the lengths of the elements of the array.*

A31. Directivity increases.

A32. Lower radiation resistance.

A33. Parallel and in the same plane.

A34. They sharpen.

A35. Extremely low radiation resistance, confined to one frequency, and affected by atmospheric conditions.

A36. Along the major axis

A37. Symmetrically.

A38. Length of the parasitic element (tuning) and spacing between the parasitic and driven elements.

A39. Increased gain and directivity.

A40. Rotary array.

A41. Their adjustment is critical and they do not operate over a wide frequency range.

A42. Increased gain.

A43. Multielement parasitic array.

A44. One-half wavelength.

A45. Wave antenna.

A46. Opposite.

A47. It requires a large antenna site.

A48. For omnidirectional vhf communications.

APPENDIX I

GLOSSARY

ABSORPTION—(1) Absorbing light waves. Does not allow any reflection or refraction. (2) Atmospheric absorption of rf energy with no reflection or refraction (adversely affects long distance communications).

ACOUSTICS—The science of sound.

AMPLITUDE—The portion of a cycle measured from a reference line to a maximum value above (or to a maximum value below) the line.

ANGLE OF INCIDENCE—The angle between the incident wave and the normal.

ANGLE OF REFLECTION—The angle between the reflected wave and the normal.

ANGLE OF REFRACTION—The angle between the normal and the path of a wave through the second medium.

ANGSTROM UNIT—The unit used to define the wavelength of light waves.

ANISOTROPIC—The property of a radiator to emit strong radiation in one direction.

ANTENNA—A conductor or set of conductors used either to radiate rf energy into space or to collect rf energy from space.

ARRAY OF ARRAYS—Same as COMBINATION ARRAY.

BAY—Part of an antenna array.

BEVERAGE ANTENNA—A horizontal, longwire antenna designed for reception and transmission of low-frequency, vertically polarized ground waves.

BIDIRECTIONAL ARRAY—An array that radiates in opposite directions along the line of maximum radiation.

BROADSIDE ARRAY—An array in which the direction of maximum radiation is perpendicular to the plane containing the elements.

CENTER-FEED METHOD—Connecting the center of an antenna to a transmission line, which is then connected to the final (output) stage of the transmitter.

CHARACTERISTIC IMPEDANCE—The ratio of voltage to current at any given point on a transmission line. Represented by a value of impedance.

COAXIAL LINE—A type of transmission line that contains two concentric conductors.

COLLINEAR ARRAY—An array with all the elements in a straight line. Maximum radiation is perpendicular to the axis of the elements.

COMBINATION ARRAY—An array system that uses the characteristics of more than one array.

COMPLEMENTARY (SECONDARY) COLORS OF LIGHT—The colors of light produced when two of the primaries are mixed in overlapping beams of light. The complementary colors of light are magenta, yellow, and cyan.

COMPLEX WAVE—A wave produced by combining two or more pure tones at the same time.

COMPRESSION WAVES—Longitudinal waves that have been compressed (made more dense) as they move away from the source.

CONDUCTANCE—The opposite of resistance in transmission lines. The minute amount of resistance that is present in the insulator of a transmission line.

CONNECTED ARRAY—Another term for DRIVEN ARRAY.

COPPER LOSSES—The I^2R loss in a conductor caused by the current flow through the resistance of the conductor.

CORNER-REFLECTOR ANTENNA—A half-wave antenna with a reflector consisting of two flat metal surfaces meeting at an angle behind the radiator.

COUNTERPOISE—A network of wire that is connected to a quarter-wave antenna at one end and provides the equivalent of an additional 1/4 wavelength.

COUPLING DEVICE—A coupling coil that connects the transmitter to the feeder.

CREST (TOP)—The peak of the positive alternation (maximum value above the line) of a wave.

CRITICAL ANGLE—The maximum angle at which radio waves can be transmitted and still be refracted back to earth.

CRITICAL FREQUENCY—The maximum frequency at which a radio wave can be transmitted vertically and still be refracted back to earth.

CURRENT-FEED METHOD—Same as CENTER-FEED METHOD.

CURRENT STANDING-WAVE RATIO (ISWR)—The ratio of maximum to minimum current along a transmission line.

CYCLE—One complete alternation of a sine wave that has a maximum value above and a maximum value below the reference line.

DAMPING—Reduction of energy by absorption.

DENSITY—(1) The compactness of a substance. (2) Mass per unit volume.

DETECTOR—The device that responds to a wave or disturbance.

DIELECTRIC HEATING—The heating of an insulating material by placing it in a high frequency electric field.

DIELECTRIC LOSSES—The losses resulting from the heating effect on the dielectric material between conductors.

DIFFRACTION—The bending of the paths of waves when the waves meet some form of obstruction.

DIFFUSION—The scattering of reflected light waves (beams) from an object, such as white paper.

DIPOLE—A common type of half-wave antenna made from a straight piece of wire cut in half. Each half operates at a quarter wavelength of the output.

DIRECTIONAL—Radiation that varies with direction.

DIRECTOR—The parasitic element of an array that reinforces energy coming from the driver toward itself.

DIRECTIVITY—The property of an array that causes more radiation to take place in certain directions than in others.

DISPERSION—The refraction of light waves that causes the different frequencies to bend at slightly different angles.

DISTRIBUTED CONSTANTS—The constants of inductance, capacitance, and resistance in a transmission line. The constants are spread along the entire length of the line and cannot be distinguished separately.

DOPPLER EFFECT—The apparent change in frequency or pitch when a sound source moves either toward or away from a listener.

DOUBLET—Another name for the dipole antenna.

DRIVEN ARRAY—An array in which all of the elements are driven.

DRIVEN ELEMENT—An element of an antenna (transmitting or receiving) that is connected directly to the transmission line.

ECHO—The reflection of the original sound wave as it bounces off a distant surface.

ELASTICITY—The ability of a substance to return to its original state.

ELECTROMAGNETIC FIELD—The combination of an electric (E) field and a magnetic (H) field.

ELECTROMAGNETIC INTERFERENCE—Man-made or natural interference that degrades the quality of reception of radio waves.

ELECTROMAGNETIC RADIATION—The radiation of radio waves into space.

ELECTRIC (E) FIELD—The field produced as a result of a voltage charge on a conductor or antenna.

ELEMENT—A part of an antenna that can be either an active radiator or a parasitic radiator.

END-FEED METHOD—Connecting one end of an antenna through a capacitor to the final output stage of a transmitter.

END-FIRE ARRAY—An array in which the direction of radiation is parallel to the axis of the array.

FADING—Variations in signal strength by atmospheric conditions.

FEEDER—A transmission line that carries energy to the antenna.

FLAT LINE—A transmission line that has no standing waves. This line requires no special tuning device to transfer maximum power.

FLEXIBLE COAXIAL LINE—A coaxial line made with a flexible inner conductor insulated from the outer conductor by a solid, continuous insulating material.

FOLDED DIPOLE—An ordinary half-wave antenna (dipole) that has one or more additional conductors connected across the ends parallel to each other.

FOUR-ELEMENT ARRAY—An array with three parasitic elements and one driven element.

FREE-SPACE LOSS—The loss of energy of a radio wave because of the spreading of the wavefront as it travels from the transmitter.

FREQUENCY—The number of cycles that occur in one second. Usually expressed in hertz.

FREQUENCY DIVERSITY—Transmitting (and receiving) of radio waves on two different frequencies simultaneously.

FRONT-TO-BACK RATIO—The ratio of the energy radiated in the principal direction to the energy radiated in the opposite direction.

FUNDAMENTAL FREQUENCY—The basic frequency or first harmonic frequency.

GAIN—The ratio between the amount of energy propagated from an antenna that is directional to the energy from the same antenna that would be propagated if the antenna were not directional.

GENERATOR END—See INPUT END.

GROUND PLANE—The portion of a groundplane antenna that acts as ground.

GROUND-PLANE ANTENNA—A type of antenna that uses a ground plane as a simulated ground to produce low-angle radiation.

GROUND REFLECTION LOSS—The loss of rf energy each time a radio wave is reflected from the Earth's surface.

GROUND SCREEN—A series of conductors buried below the surface of the earth and arranged in a radial pattern. Used to reduce losses in the ground.

GROUND WAVES—Radio waves that travel near the surface of the Earth.

HALF-WAVE DIPOLE ANTENNA—An antenna consisting of two rods (1/4 wavelength each) in a straight line, that radiates electromagnetic energy.

HARMONIC—A frequency that is a whole number multiple of a smaller base frequency.

HERTZ ANTENNA—A half-wave antenna installed some distance above ground and positioned either vertically or horizontally.

HORIZONTAL AXIS—On a graph, the straight line axis plotted from left to right.

HORIZONTAL PATTERN—The part of a radiation pattern that is radiated in all directions along the horizontal plane.

HORIZONTALLY POLARIZED—Waves that are radiated with their E field component parallel to the Earth's surface.

INCIDENT WAVE—(1) The wave that strikes the surface of a medium. (2) The wave that travels from the sending end to the receiving end of a transmission line.

INDUCTION FIELD—The electromagnetic field produced about an antenna when current and voltage are present on the same antenna.

INDUCTION LOSSES—The losses that occur when the electromagnetic field around a conductor cuts through a nearby metallic object and induces a current into that object.

INFRASONIC (SUBSONIC)—Sounds below 15 hertz.

INPUT END—The end of a two-wire transmission line that is connected to a source.

INPUT IMPEDANCE—The impedance presented to the transmitter by the transmission line and its load.

INTENSITY (OF SOUND)—The measurement of the amplitude of sound energy. Sometimes mistakenly called loudness.

INTERCEPT—The point where two lines drawn on a graph cross each other.

INTERFERENCE—Any disturbance that produces an undesirable response or degrades a wave.

IONOSPHERE—The most important region of the atmosphere extending from 31 miles to 250 miles above the earth. Contains four cloud-like layers that affect radio waves.

IONOSPHERIC STORMS—Disturbances in the earth's magnetic field that make communications practical only at lower frequencies.

IONIZATION—The process of upsetting electrical neutrality.

ISOTROPIC RADIATION—The radiation of energy equally in all directions.

LEAKAGE CURRENT—The small amount of current that flows between the conductors of a transmission line through the dielectric.

LIGHT RAYS—Straight lines that represent light waves emitting from a source.

LOAD END—See OUTPUT END.

LOADING—See LUMPED-IMPEDANCE TUNING.

LOBE—An area of a radiation pattern plotted on a polar-coordinate graph that represents maximum radiation.

LONG-WIRE ANTENNA—An antenna that is a wavelength or more long at its operating frequency.

LONGITUDINAL WAVES—Waves in which the disturbance (back and forth motion) takes place in the direction of propagation. Sometimes called compression waves.

LOOP—The curves of a standing wave or antenna that represent amplitude of current or voltage.

LOWEST USABLE FREQUENCY—The minimum operating frequency that can be used for communications between two points.

LUMPED CONSTANTS—The properties of inductance, capacitance, and resistance in a transmission line.

LUMPED-IMPEDANCE TUNING—The insertion of an inductor or capacitor in series with an antenna to lengthen or shorten the antenna electrically.

MAGNETIC (H) FIELD—The field produced when current flows through a conductor or antenna.

MAJOR LOBE—The lobe in which the greatest amount of radiation occurs.

MARCONI ANTENNA—A quarter-wave antenna oriented perpendicular to the earth and operated with one end grounded.

MAXIMUM USABLE FREQUENCY— Maximum frequency that can be used for communications between two locations for a given time of day and a given angle of incidence.

MEDIUM—The substance through which a wave travels from one point to the next. Air, water, wood, etc., are examples of a medium.

MINOR LOBE—The lobe in which the radiation intensity is less than a major lobe.

MULTIELEMENT ARRAY—An array consisting of one or more arrays and classified as to directivity.

MULTIELEMENT PARASITIC ARRAY— An array that contains two or more parasitic elements and a driven element.

MULTIPATH—The multiple paths a radio wave may follow between transmitter and receiver.

NATURAL HORIZON—The line-of-sight horizon.

NEGATIVE ALTERNATION—The portion of a sine wave below the reference line.

NODE—The fixed minimum points of voltage or current on a standing wave or antenna.

NOISE (OF SOUND)—An unwanted disturbance caused by spurious waves that originate from man-made or natural sources.

NONDIRECTIONAL—See OMNIDIRECTIONAL.

NONLUMINOUS BODIES—Objects that either reflect or diffuse light that falls upon them.

NONRESONANT LINE—A transmission line that has no standing waves of current or voltage.

NORMAL—The imaginary line perpendicular to the point at which the incident wave strikes the reflecting surface. Also called the perpendicular.

NULL—On a polar-coordinate graph, the area that represents minimum or 0 radiation.

OMNIDIRECTIONAL—Transmitting in all directions.

OPAQUE—A type of substance that does not transmit any light rays.

OPEN-ENDED LINE—A transmission line that has an infinitely large terminating impedance.

OPTIMUM WORKING FREQUENCY—The most practical operating frequency that can be used with the least amount of problems; roughly 85 percent of the maximum usable frequency.

ORIGIN—The point on a graph where the vertical and horizontal axes cross each other.

OUTPUT END—The end of a transmission line that is opposite the source.

OUTPUT IMPEDANCE—The impedance presented to the load by the transmission line and its source.

PARALLEL RESONANT CIRCUIT—A circuit that acts as a high impedance at resonance.

PARALLEL-WIRE—A type of transmission line consisting of two parallel wires.

PARASITIC ARRAY—An array that has one or more parasitic elements.

PARASITIC ELEMENT—The passive element of an antenna array that is connected to neither the transmission line nor the driven element.

PERIOD—The amount of time required for completion of one full cycle.

PITCH—A term used to describe the frequency of a sound heard by the human ear.

PLANE OF POLARIZATION—The plane (vertical or horizontal) with respect to the earth in which the E field propagates.

POINT OF ZERO DISPLACEMENT—See REFERENCE LINE.

POLAR-COORDINATE GRAPH—A graph whose axes consist of a series of circles with a common center and a rotating radius extending from the center of the concentric circles.

POSITIVE ALTERNATION—The portion of a sine wave above the reference line.

POWER LOSS—The heat loss in a conductor as current flows through it.

POWER STANDING-WAVE RATIO (PSWR)—The ratio of the square of the maximum and minimum voltages of a transmission line.

PRIMARY COLORS (OF LIGHT)—The three primary colors of light (red, green, and blue), from which all other colors may be derived.

PRISM—A triangular-shaped glass that refracts and disperses light waves into component wavelengths.

PROPAGATION—Waves traveling through a medium.

QUALITY (OF SOUND)—The factor that distinguishes tones of pitch and loudness.

QUARTER-WAVE ANTENNA—Same as the Marconi antenna.

RADIATION FIELD—The electromagnetic field that detaches itself from an antenna and travels through space.

RADIATION LOSSES—The losses that occur when magnetic lines of force about a conductor are projected into space as radiation and are not returned to the conductor as the cycle alternates.

RADIATION PATTERN—A plot of the radiated energy from an antenna.

RADIATION RESISTANCE—The resistance, which if inserted in place of an antenna, would consume the same amount of power as that radiated by the antenna.

RADIO FREQUENCIES—Electromagnetic frequencies that fall between 3 kilohertz and 300 gigahertz and are used for radio communications.

RADIO HORIZON—The boundary beyond the natural horizon in which radio waves cannot be propagated over the earth's surface.

RADIO WAVE—(1) A form of radiant energy that can neither be seen nor felt. (2) An electromagnetic wave generated by a transmitter.

RAREFIED WAVE—A longitudinal wave that has been expanded or rarefied (made less dense) as it moves away from the source.

RECEIVER—The object that responds to a wave or disturbance. Same as detector.

RECEIVING ANTENNA—The device used to pick up an rf signal from space.

RECEIVING END—See OUTPUT END.

RECIPROCITY—The property of interchangeability of the same antenna for transmitting and receiving.

RECTANGULAR-COORDINATE GRAPH—A graph in which straight-line axes (horizontal and vertical) are perpendicular.

REFERENCE LINE—The position a particle of matter would occupy if it were not disturbed by wave motion.

REFLECTED WAVE—(1) The wave that reflects back from a medium. (2) Waves traveling from the load back to the generator on a transmission line. (3) The wave moving back to the sending end of a transmission line after reflection has occurred.

REFLECTION WAVES—Waves that are neither transmitted nor absorbed, but are reflected from the surface of the medium they encounter.

REFLECTOR—The parasitic element of an array that causes maximum energy radiation in a direction toward the driven element.

REFRACTION—The changing of direction as a wave leaves one medium and enters another medium of a different density.

RERADIATION—The reception and retransmission of radio waves caused by turbulence in the troposphere.

RESONANCE—The condition produced when the frequency of vibrations are the same as the natural frequency (of a cavity). The vibrations reinforce each other.

RESONANT LINE—A transmission line that has standing waves of current and voltage.

REST POSITION—See REFERENCE LINE.

REVERBERATION—The multiple reflections of sound waves.

RHOMBIC ANTENNA—A diamond-shaped antenna used widely for long-distance, high-frequency transmission and reception.

RIGID COAXIAL LINE—A coxial line consisting of a central, insulated wire (inner conductor) mounted inside a tubular outer conductor.

SCATTER ANGLE—The angle at which the receiving antenna must be aimed to capture the scattered energy of tropospheric scatter.

SELF-INDUCTION—The phenomenon caused by the expanding and collapsing fields of an electron which encircles other electrons and retards the movement of the encircled electrons.

SELF-LUMINOUS BODIES—Objects that produce their own light.

SENDING END—See INPUT END.

SERIES RESONANT CIRCUIT—A circuit that acts as a low impedance at resonance.

SHIELDED PAIR—A line consisting of parallel conductors separated from each other and surrounded by a solid dielectric.

SHORT-CIRCUITED LINE—A transmission line that has a terminating impedance equal to 0.

SINK—See OUTPUT END.

SKIN EFFECT—The flow of ac current near the surface of a conductor at rf frequencies.

SKIP DISTANCE—The distance from a transmitter to the point where the sky wave is first returned to earth.

SKIP ZONE—A zone of silence between the point where the ground wave becomes too weak for reception and the point where the sky wave is first returned to earth.

SKY WAVES—Radio waves reflected back to earth from the ionosphere.

SONIC—Pertaining to sounds capable of being heard by the human ear.

SOURCE—(1) The object that produces waves or disturbance. (2) The name given to the end of a two-wire transmission line that is connected to a source.

SPACE DIVERSITY—Reception of radio waves by two or more antennas spaced some distance apart.

SPACE WAVE—A radio wave that travels directly from the transmitter to the receiver and remains in the troposphere.

SPECTRUM—(1) The entire range of electromagnetic waves. (2) VISIBLE. The range of electromagnetic waves that stimulate the sense of sight. (3) ELECTROMAGNETIC. The entire range of electromagnetic waves arranged in order of their frequencies.

SPORADIC E LAYER—Irregular cloud-like patches of unusually high ionization. Often forms at heights near the normal E layer.

SPREADER—Insulator used with transmission lines and antennas to keep the parallel wires separated.

STANDING WAVE—The distribution of voltage and current formed by the incident and reflected waves which have minimum and maximum points on a resultant wave that appears to stand still.

STANDING-WAVE RATIO (SWR)—The ratio of the maximum (voltage, current) to the minimum (voltage, current) of a transmission line. Measures the perfection of the termination of the line.

STRATOSPHERE—Located between the troposphere and the ionosphere. Has little effect on radio waves.

STUB—Short section of a transmission line used to match the impedance of a transmission line to an antenna. Can also be used to produce desired phase relationships between connected elements of an antenna.

SUDDEN IONOSPHERIC DISTURBANCE—An irregular ionospheric disturbance that can totally blank out hf radio communications.

SUPERSONIC—Speed greater than the speed of sound.

SURFACE WAVE—A radio wave that travels along the contours of the earth, thereby being highly attenuated.

TEMPERATURE INVERSION—The condition in which warm air is formed above a layer of cool air that is near the earth's surface.

THREE-ELEMENT ARRAY—An array with two parasitic elements (reflector and director) and a driven element.

TONES—Musical sounds.

TRANSLUCENT—A type of substance, such as frosted glass, through which some light rays can pass but through which objects cannot be seen clearly.

TRANSMISSION LINE—A device designed to guide electrical energy from one point to another.

TRANSMITTING ANTENNA—The device used to send the transmitted signal energy into space.

TRANSPARENT—A type of substance, such as glass, that transmits almost all of the light waves that fall upon it.

TRANSMISSION MEDIUMS—The various types of lines and waveguides used as transmission lines.

TRANSMITTER END—See INPUT END.

TRANSVERSE WAVE MOTION—The up and down motion of a wave as the wave moves outward.

TROPOSPHERE—The portion of the atmosphere closest to the earth's surface, where all weather phenomena take place.

TROPOSPHERIC SCATTER—The propagation of radio waves in the troposphere by means of scatter.

TROUGH (BOTTOM)—The peak of the negative alternation (maximum value below the line).

TUNED LINE—Another name for the resonant line. This line uses tuning devices to eliminate the reactance and to transfer maximum power from the source to the line.

TURNSTILE ANTENNA—A type of antenna used in vhf communications that is omnidirectional and consists of two horizontal half-wave antennas mounted at right angles to each other in the same horizontal plane.

TWISTED PAIR—A line consisting of two insulated wires twisted together to form a flexible line without the use of spacers.

TWO-WIRE OPEN LINE—A parallel line consisting of two wires that are generally spaced from 2 to 6 inches apart by insulating spacers.

TWO-WIRE RIBBON (TWIN LEAD)—A parallel line similar to a two-wire open line except that uniform spacing is assured by embedding the two wires in a low-loss dielectric.

ULTRASONIC—Sounds above 20,000 hertz.

UNIDIRECTIONAL ARRAY—An array that radiates in only one general direction.

UNTUNED LINE—Another name for the flat or nonresonant line.

V ANTENNA—A bi-directional antenna, shaped like a V, which is widely used for communications.

VELOCITY—The rate at which a disturbance travels through a medium.

VERTICAL AXIS—On a graph, the straight line axis oriented from bottom to top.

VERTICAL PATTERN—The part of a radiation pattern that is radiated in the vertical plane.

VERTICALLY POLARIZED—Waves radiated with the E field component perpendicular to the earth's surface.

VOLTAGE-FEED METHOD—Same as END FEED METHOD.

VOLTAGE STANDING-WAVE RATIO (VSWR)—The ratio of maximum to minimum voltage of a transmission line.

WAVE ANTENNA—Same as BEVERAGE ANTENNA.

WAVE MOTION—A recurring disturbance advancing through space with or without the use of a physical medium.

WAVE TRAIN—A continuous series of waves with the same amplitude and wavelength.

WAVEFRONT—A small section of an expanding sphere of electromagnetic radiation, perpendicular to the direction of travel of the energy.

WAVEGUIDE—A hollow metal tube used as a transmission line to guide energy from one point to another.

WAVELENGTH—(1) The distance in space occupied by 1 cycle of a radio wave at any given instant. (2) The distance a disturbance travels during one period of vibration.

YAGI ANTENNA—A multielement parasitic array. Elements lie in the same plane as those of the end-fire array.

MODULE 10 INDEX

ASSIGNMENT 1

Textbook assignment: Chapter 1, “Wave Propagation,” pages 1-1 through 1-48.

1-1. What is the major advantage of the telegraph over earlier methods of communication?

1. Range
2. Speed
3. Security
4. Reliability

1-2. The spreading out of radio waves is referred to as propagation and is used in which of the following Navy equipment?

1. Detection
2. Communication
3. Radar and navigation
4. Each of the above

1-3. Radio-frequency waves CANNOT be seen for which of the following reasons?

1. Because radio-frequency energy is low powered
2. Because radio-frequency waves are below the sensitivity range of the human eye
3. Because the human eye detects only magnetic energy
4. Because radio-frequency waves are above the sensitivity range of the human eye

1-4. Radio waves travel at what speed?

1. Speed of sound
2. Speed of light
3. Speed of the Earth's rotation
4. Speed of the Earth's orbit around the sun

1-5. Which of the following types of energy CANNOT be seen, heard, or felt?

1. Radio waves
2. Sound waves
3. Heat waves
4. Light waves

1-6. A stone dropped into water creates a series of expanding circles on the surface of the water. This is an example of which of the following types of wave motion?

1. Transverse
2. Concentric
3. Longitudinal
4. Compression

1-7. A sound wave that moves back and forth in the direction of propagation is an example of which of the following types of wave motion?

1. Composite
2. Concentric
3. Transverse
4. Longitudinal

1-8. Which of the following terms is used for the vehicle through which a wave travels from point to point?

1. Medium
2. Source
3. Detector
4. Receiver

1-9. Which of the following is NOT an element necessary to propagate sound?

1. Medium
2. Source
3. Detector
4. Reference

1-10. If a wave has a velocity of 4,800 feet per second and a wave-length of 5 feet, what is the frequency of the wave?

1. 9.6 Hz
2. 96 Hz
3. 960 Hz
4. 9,600 Hz

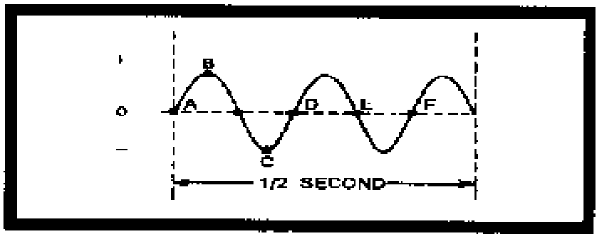


Figure 1-A.—Waveform.

IN ANSWERING QUESTIONS 1-11 THROUGH 1-15, REFER TO FIGURE 1-A.

1-11. The waveform in the figure is what type of wave?

1. Sine
2. Square
3. Sawtooth
4. Trapezoidal

1-12. The distance between which of the following points represents the completion of a full cycle of alternating current?

1. A to C
2. B to D
3. C to E
4. D to F

1-13. The distance between which of the following points represents a full wavelength?

1. A to D
2. A to E
3. D to E
4. E to F

1-14. What is the frequency of the wave?

1. 0.5 Hz
2. 2.5 Hz
3. 5.0 Hz
4. 7.5 Hz

1-15. What is the period of the wave?

1. 100 milliseconds
2. 200 milliseconds
3. 250 milliseconds
4. 500 milliseconds

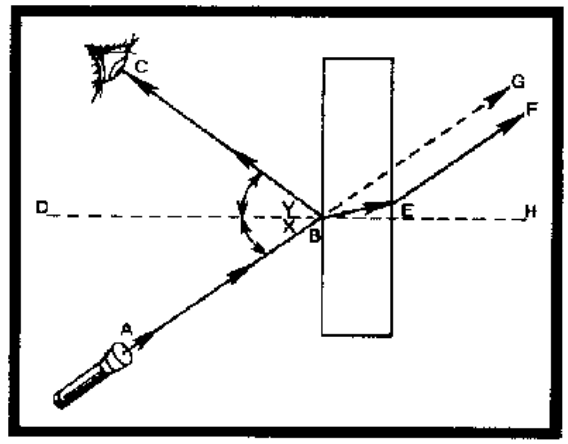


Figure 1-B.—Wave angles.

IN ANSWERING QUESTIONS 1-16 THROUGH 1-19, REFER TO FIGURE 1-B.

1-16. What line in the figure indicates the incident wave?

1. A to B
2. B to E
3. C to B
4. D to H

1-17. Angle "x" is which of the following angles?

1. Normal
2. Incidence
3. Reflection
4. Refraction

1-18. Line E to F represents which of the following waves?

1. Normal
2. Incident
3. Refracted
4. Reflected

1-19. Line D to H represents which of the following references?

1. Normal
2. Perpendicular
3. Both 1 and 2 above
4. Reflected line

1-20. Which of the following statements about a wave is the law of reflection?

1. The angle of incidence is equal to the refracted wave
2. The angle of incidence is not equal to the refracted wave
3. The angle of incidence is equal to the angle of reflection
4. The angle of incidence is not equal to the angle of reflection

1-21. If a wave passes first through a dense medium and then through a less dense medium, which of the following angle-of-refraction conditions exists?

1. The angle of refraction is greater than the angle of incidence
2. The angle of refraction is less than the angle of incidence
3. The angle of refraction is equal to the angle of incidence
4. The wave will pass through in a straight line

1-22. The reception of an AM-band radio signal over mountains can be explained by which of the following principles of wave propagation?

1. Reflection
2. Refraction
3. Diffraction
4. Doppler effect

1-23. What wave propagation principle accounts for the apparent increase in frequency as a train whistle approaches and the apparent decrease in frequency as it moves away?

1. Refraction
2. Reflection
3. Diffraction
4. Doppler effect

1-24. Longitudinal wave disturbances that travel through a medium are known as what type of waves?

1. Air
2. Sound
3. Radio
4. Light

1-25. What are the three audible frequency ranges?

1. Subsonic, sonic, and supersonic
2. Infrasonic, sonic, and ultrasonic
3. Infrasonic, subsonic, and ultrasonic
4. Infrasonic, subsonic, and supersonic

1-26. If a bell is placed in a jar and the air in the jar is replaced with a gas of a higher density, what is the effect, if any, on the speed of the sound when the bell is rung?

1. The sound stops
2. The sound travels faster
3. The sound travels slower
4. The sound is not affected

1-27. Varying which of the following wave characteristics will cause the length of sound waves to vary?

1. Phase
2. Quality
3. Amplitude
4. Frequency

1-28. What are the three basic characteristics of sound?

1. Amplitude, intensity, and quality
2. Amplitude, pitch, and tone
3. Pitch, intensity, and quality
4. Pitch, frequency, and quality

1-29. If several musical instruments are playing the same note, you should be able to distinguish one instrument from another because of which of the following characteristics of sound?

1. Quality
2. Overtones
3. Frequency
4. Intensity

1-30. Through which of the following mediums will sound travel fastest, at the indicated temperature?

1. Air at 68º F
2. Lead at 20º C
3. Steel at 32º F
4. Steel at 20º C

1-31. In sound terminology, which of the following terms is the same as a wave reflection?

1. Echo
2. Image
3. Acoustics
4. Refraction

1-32. Multiple reflections of sound waves are referred to as

1. noise
2. acoustics
3. interference
4. reverberation

1-33. Two out-of-phase waves of the same frequency that are moving through the same medium are said to present which of the following types of interference?

1. Additive
2. Constructive
3. Both 1 and 2 above
4. Subtractive

1-34. A cavity that vibrates at its own natural frequency and produces a sound that is louder than at other frequencies is demonstrating which of the following sound characteristics?

1. Noise
2. Quality
3. Resonance
4. Reverberation

1-35. Energy in the form of light can be produced through which of the following means?

1. Chemical
2. Electrical
3. Mechanical
4. Each of the above

1-36. The scientist, J. C. Maxwell, developed the theory that small packets of electromagnetic energy called photons produce

1. sound
2. noise
3. echoes
4. light

1-37. A large volume of light radiating in a given direction is referred to as a

1. ray
2. beam
3. shaft
4. pencil

1-38. Which of the following units of measurement is/are used to measure very short wavelengths of light?

1. Angstrom (Å)
2. Millimicron
3. Both 1 and 2 above
4. Millimeter

1-39. What are the primary colors of light?

1. Red, blue, and yellow
2. Red, blue, and green
3. Red, violet, and indigo
4. Blue, green, and violet

1-40. What are the secondary colors of light?

1. Orange, yellow, and blue-green
2. Magenta, yellow, and cyan
3. Purple, yellow, and black
4. Red, white, and blue

1-41. What causes sunlight to separate into different wavelengths and display a rainbow of colors when passed through a prism?

1. Refraction
2. Reflection
3. Dispersion
4. Diffraction

1-42. The sun, gas flames, and electric light filaments are visible because they are

1. opaque
2. transparent
3. nonluminous
4. self-luminous

1-43. Substances that transmit almost all of the light waves falling upon them possess which of the following properties?

1. Opaqueness
2. Transparence
3. Translucence
4. Self-lumination

1-44. Some substances are able to transmit light waves but objects cannot be seen through them. Which of the following properties does this statement describe?

1. Opaqueness
2. Transparence
3. Translucence
4. Self-lumination

1-45. The speed of light depends on the medium through which light travels. For which of the following reasons does light travel through empty space faster than through an object such as glass?

1. Space is less dense than glass
2. Space is more dense than glass
3. Glass reflects the light back to the source
4. Glass refracts the light, causing the light to travel in all directions

1-46. If a light wave strikes a sheet of glass at a perpendicular angle, what is the effect, if any, on the light wave?

1. The wave is completely absorbed
2. The wave is reflected back toward the source
3. The wave is refracted as it passes through the glass
4. The wave is unchanged and continues in a straight line

1-47. The amount of absorption of the light that strikes an object is determined by the object's

1. color
2. purity
3. density
4. complexity

1-48. In a comparison of waves of light and sound as they travel from an air into water, how is the speed of (a) light waves and (b) sound waves affected?

1. (a) Increased (b) increased
2. (a) Increased (b) decreased
3. (a) Decreased (b) decreased
4. (a) Decreased (b) increased

1-49. Which of the following waves are NOT a form of electromagnetic energy?

1. Heat waves
2. Sound waves
3. Light waves
4. Radio waves

1-50. The electromagnetic spectrum represents the entire range of electromagnetic waves arranged in the order of their

1. color
2. frequency
3. visibility
4. application

1-51. Which of the following portions of the frequency spectrum contains the highest frequency?

1. X-ray
2. Radar
3. Light
4. Cosmic

1-52. Which of the following electronic devices is used to radiate and/or collect electromagnetic waves?

1. Antenna
2. Receiver
3. Transmitter
4. Transmission line

1-53. The electric field and magnetic field combine to form which of the following types of waves?

1. Spherical
2. Elliptical
3. Electromagnetic
4. Each of the above

1-54. The magnetic field radiated from an antenna is produced by what electrical property?

1. Voltage
2. Current
3. Reactance
4. Resistance

1-55. The electric field radiated from an antenna is produced by what electrical property?

1. Voltage
2. Current
3. Reactance
4. Resistance

1-56. Applying rf energy to the elements of an antenna results in what phase relationship between voltage and current?

1. Voltage lags current by 90 degrees
2. Voltage leads current by 90 degrees
3. Voltage and current are 180 degrees out of phase
4. Voltage and current are in phase

1-57. What field exists close to the conductor of an antenna and carries the current?

1. Electric
2. Magnetic
3. Induction
4. Radiation

1-58. What field travels through space after being detached from the current-carrying rod of an antenna?

1. Electric
2. Magnetic
3. Induction
4. Radiation

1-59. Electric and magnetic fields on an antenna reach their maximum intensity at which of the following times?

1. When they are a full cycle apart
2. When they are three-quarter cycle apart
3. When they are a half-cycle apart
4. When they are a quarter-cycle apart

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Radio Wave Propagation," pages 2-1 through 2-47.

2-1. The induction field is made up of which of the following fields?

1. E field only
2. H field only
3. Both E and H fields

2-2. After the radiation field leaves an antenna, what is the relationship between the E and H fields with respect to (a) phase and (b) physical displacement in space?

1. (a) In phase (b) 90 degrees
2. (a) Out of phase (b) 90 degrees
3. (a) In phase (b) 180 degrees
4. (a) Out of phase (b) 180 degrees

2-3. What is the first harmonic of a radio wave that has a fundamental frequency of 2,000 kHz?

1. 6,000 kHz
2. 2,000 kHz
3. 3,000 kHz
4. 4,000 kHz

2-4. In a radio wave with a fundamental frequency of 1.5 kHz, which of the following frequencies is NOT a harmonic?

1. 6,000 kHz
2. 5,000 kHz
3. 3,000 kHz
4. 4,500 kHz

2-5. A radio wave with a frequency of 32 kHz is part of which of the following frequency bands?

1. The lf band
2. The mf band
3. The hf band
4. The vhf band

2-6. A frequency of 3.5 GHz falls into what rf band?

1. High
2. Very high
3. Super high
4. Extremely high

2-7. A radio wavelength expressed as 250 meters may also be expressed as how many feet?

1. 410
2. 820
3. 1,230
4. 1,640

2-8. An increase in the frequency of a radio wave will have what effect, if any, on the velocity of the radio wave?

1. Increase
2. Decrease
3. None

2-9. An increase in frequency of a radio wave will have what effect, if any, on the wavelength of the radio wave?

1. Increase
2. Decrease
3. None

2-10. What is the frequency, in kiloHertz, of a radio wave that is 40 meters long?

1. 75
2. 750
3. 7,500
4. 75,000

2-11. What is the approximate wavelength, in feet, of a radio wave with a frequency of 5,000 kHz?

1. 197 feet
2. 1,970 feet
3. 19,700 feet
4. 197,000 feet

2-12. The polarity of a radio wave is determined by the orientation of (a) what moving field with respect to (b) what reference?

1. (a) Electric (b) earth
2. (a) Electric (b) antenna
3. (a) Magnetic (b) antenna
4. (a) Magneti (b) earth

2-13. Energy radiated from an antenna is considered horizontally polarized under which of the following conditions?

1. If the wavefront is in the horizontal plane
2. If the magnetic field is in the horizontal plane
3. If the electric field is in the horizontal plane
4. If the induction field is in the horizontal plane

2-14. The ability of a reflecting surface to reflect a specific radio wave depends on which of the following factors?

1. Striking angle
2. Wavelength of the wave
3. Size of the reflecting area
4. All of the above

THIS SPACE LEFT BLANK INTENTIONALLY.

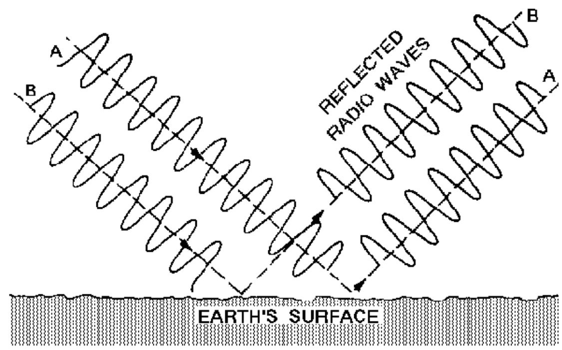


Figure 2-A.—Reflected radio waves.

IN ANSWERING QUESTION 2-15, REFER TO FIGURE 2-A.

2-15. If the two reflected radio waves shown in the figure are received at the same instant at the receiving site, what will be the effect, if any, on signal quality?

1. A stronger signal will be produced
2. A weak or fading signal will be produced
3. The signal will be completely canceled out
4. None

2-16. The bending of a radio wave because of a change in its velocity through a medium is known as

1. refraction
2. reflection
3. deflection
4. diffraction

2-17. Radio communications can be diffracted to exceptionally long distances through the use of (a) what frequency band at (b) what relative power level?

1. (a) Very low frequency
   (b) Low power
2. (a) Very high frequency
   (b) Low power
3. (a) Very low frequency
   (b) High power
4. (a) Very high frequency
   (b) High power

2-18. Electrically charged particles that affect the propagation of radio waves are found in what atmospheric layer?

1. Troposphere
2. Ionosphere
3. Chronosphere
4. Stratosphere

2-19. Most weather phenomena take place in which of the following region of the atmosphere?

1. Troposphere
2. Ionosphere
3. Chronosphere
4. Stratosphere

2-20. Radio wave propagation has the least effect because of its constancy on which of the following atmospheric layers?

1. Troposphere
2. Ionosphere
3. Chronosphere
4. Stratosphere

2-21. Long range, surface-wave communications are best achieved when the signal is transmitted over seawater with (a) what polarization at (b) what relative frequency?

1. (a) Vertical (b) Low
2. (a) Vertical (b) High
3. (a) Horizontal (b) High
4. (a) Horizontal (b) Low

2-22. The Navy's long-range vlf broadcasts are possible because of the advantages of which of the following types of propagation?

1. Diffraction
2. Ionospheric refraction
3. Repeated reflection and refraction
4. Both 2 and 3 above

2-23. A space wave (a) is primarily a result of refraction in what atmospheric layer and (b) extends approximately what distance beyond the horizon?

1. (a) Ionosphere
   (b) One-tenth farther
2. (a) Ionosphere
   (b) One-third farther
3. (a) Troposphere
   (b) One-third farther
4. (a) Troposphere
   (b) One-tenth farther

2-24. The signal of a space wave is sometimes significantly reduced at the receiving site because of which of the following interactions?

1. Space-wave refraction
2. Space-wave reflections
3. Ground-wave diffraction
4. Ground-wave reflections

2-25. For long-range communications in the hf band, which of the following types of waves is most satisfactory?

1. Sky wave
2. Space wave
3. Surface wave
4. Reflected ground wave

2-26. Ionization in the atmosphere is produced chiefly by which of the following types of radiation?

1. Alpha radiation
2. Cosmic radiation
3. Infrared radiation
4. Ultraviolet radiation

2-27. Ultraviolet waves of higher frequencies produce ionized layers at what relative altitude(s)?

1. Lower
2. Higher
3. Both 1 and 2 above

2-28. The density of ionized layers is normally greatest during which of the following periods?

1. At night
2. Before sunrise
3. Between early morning and late afternoon
4. Between afternoon and sunset

2-29. Compared to the other ionospheric layers at higher altitudes, the ionization density of the D layer is

1. about the same
2. relatively low
3. relatively high

2-30. What two layers in the ionosphere recombine and largely disappear at night?

1. D and F
2. D and E
3. E and F2
4. F1 and F2

2-31. For hf-radio communications covering long distances, what is the most important layer of the ionosphere?

1. C
2. D
3. E
4. F

2-32. Refraction of a sky wave in the ionosphere is influenced by which of the following factors?

1. Ionospheric density
2. Frequency of the wave
3. Angle of incidence of the wave
4. All of the above

2-33. A 10-MHz wave entering the ionosphere at an angle greater than its critical angle will pass through the ionosphere and be lost in space unless which of the following actions is taken?

1. The ground wave is canceled
2. The frequency of the wave is increased
3. The frequency of the wave is decreased
4. The ground wave is reinforced

2-34. The distance between the transmitter and the nearest point at which refracted waves return to earth is referred to as the

1. skip distance
2. return distance
3. reception distance
4. ground-wave distance

2-35. When ground-wave coverage is LESS than the distance between the transmitter and the nearest point at which the refracted waves return to earth, which of the following reception possibilities should you expect?

1. No sky-wave
2. Weak ground wave
3. A zone of silence
4. Strong ground wave

2-36. The greatest amount of absorption takes place in the ionosphere under which of the following conditions?

1. When sky wave intensity is the greatest
2. When collision of particles is least
3. When the density of the ionized layer is the greatest
4. When precipitation is greatest

2-37. Which of the following layers provide the greatest amount of absorption to the ionospheric wave?

1. D and E
2. D and F1
3. E and F1
4. F1 and F2

2-38. If the signal strength of an incoming signal is reduced for a prolonged period, what type of fading is most likely involved?

1. Selective
2. Multipath
3. Absorption
4. Polarization

2-39. Radio waves that arrive at a receiving site along different paths can cause signal fading if these waves have different

1. velocities
2. amplitudes
3. phase relationships
4. modulation percentages

2-40. The technique of reducing multipath fading by using several receiving antennas at different locations is known as what type of diversity?

1. Space
2. Receiver
3. Frequency
4. Modulation

2-41. The amount of rf energy lost because of ground reflections depends on which of the following factors?

1. Angle of incidence
2. Ground irregularities
3. Frequency of the wave
4. Each of the above

2-42. Receiving sites located near industrial areas can expect to have exceptionally large losses in signal quality as a result of which of the following propagation situations?

1. Absorption
2. Multihop refraction
3. Natural interference
4. Man-made interference

2-43. Which of the following ionospheric variation causes densities to vary with the axial rotation of the sun?

1. Daily variation
2. Seasonal variation
3. 27-day sunspot cycle
4. 11-year sunspot cycle

2-44. Which of the following ionospheric variation causes densities to vary with the position of the earth in its orbit around the sun?

1. Daily variation
2. Seasonal variation
3. 27-day sunspot cycle
4. 11-year sunspot cycle

2-45. Which of the following ionospheric variation causes densities to vary with the time of the day?

1. Daily variation
2. Seasonal variation
3. 27-day sunspot cycle
4. 11-year sunspot cycle

2-46. What relative range of operating frequencies is required during periods of maximum sunspot activity?

1. Lower
2. Medium
3. Higher

2-47. What factor significantly affects the frequency of occurrence of the sporadic-E layer?

1. Seasons
2. Latitude
3. Weather conditions
4. Ionospheric storms

2-48. What effect can the sporadic-E layer have on the propagation of sky waves?

1. Causes multipath interference
2. Permits long distance communications at unusually high frequencies
3. Permits short-distance communications in the normal skip zone
4. Each of the above

2-49. A sudden and intense burst of ultraviolet light is especially disruptive to communications in which of the following frequency bands?

1. Hf
2. Mf
3. Lf
4. Vlf

2-50. The density of what ionosphere layer increases because of a violent eruption on the surface of the sun?

1. D
2. E
3. F1
4. F2

2-51. Which irregular variation in ionospheric conditions can cause a waiting period of several days before communications return to normal?

1. Sporadic E
2. Ionospheric storms
3. Sudden ionospheric disturbance
4. Each of the above

2-52. For a radio wave entering the atmosphere of the earth at a given angle, the highest frequency at which refraction will occur is known by which of the following terms?

1. Usable frequency
2. Refraction frequency
3. Maximum usable frequency
4. Optimum working frequency

2-53. The most consistent communications can be expected at which of the following frequencies?

1. Critical frequency
2. Maximum usable frequency
3. Maximum working frequency
4. Optimum working frequency

2-54. If the optimum working frequency for a communications link is 4,250 kHz, what is the approximate maximum usable frequency?

1. 4,500 kHz
2. 5,000 kHz
3. 5,500 kHz
4. 6,000 kHz

2-55. In determining the success of radio transmission, which of the following factors is the LEAST predictable?

1. Antenna capabilities
2. Weather conditions along the path of communication
3. Density of ionized layers
4. Presence of ionized layers

2-56. At frequencies above 100 MHz, the greatest attenuation of rf energy from raindrops is caused by which of the following factors?

1. Ducting
2. Heat loss
3. Scattering
4. Absorption

2-57. Under certain conditions, such as ducting, line-of-sight radio waves often propagate for distances far beyond their normal ranges because of which of the following factors?

1. Low cloud masses
2. Ionospheric storms
3. Temperature inversions
4. Frequency fluctuations

2-58. When ducting is present in the atmosphere, multihop refraction of line-of-sight transmission can occur because of which of the following factors?

1. Operating frequency of the transmitter
2. Height of the transmitting antenna
3. Angle of incidence of the radio wave
4. Each of the above

2-59. A propagation technique used to extend uhf transmission range beyond the horizon uses which of the following propagation characteristics?

1. Ground reflection
2. Ionospheric scatter
3. Tropospheric scatter
4. Atmospheric refraction

2-60. Communications by tropospheric scatter can be affected by which of the following conditions?

1. Sunspot activity
2. Atmospheric conditions
3. Ionospheric disturbances
4. All of the above

2-61. What effect, if any, does the radiation angle of a transmitting antenna have on the reception of communications by tropospheric scatter?

1. The lower the angle, the weaker the signal
2. The lower the angle, the stronger the signal
3. The lower the angle, the more susceptible the signal is to distortion
4. None

2-62. Which of the following descriptions of tropospheric scatter signal reception is NOT true?

1. Receiver signal strength decreases as the turbulence height is increased
2. The level of reception depends on the number of turbulences causing scatter
3. The energy received is the portion of the wave reradiated by the turbulence
4. Increased communications distance enables more turbulence to act on the signal, thereby raising the received signal level

2-63. The tropospheric scatter signal is often characterized by very rapid fading caused by which of the following factors?

1. Extreme path lengths
2. Multipath propagation
3. Turbulence in the atmosphere
4. Angle of the transmitted beam

2-64. For which of the following communications situations would turbulence in the troposphere scatter transmission?

1. 10 MHz, range 200 miles
2. 30 MHz, range 800 miles
3. 50 MHz, range 600 miles
4. 100 MHz, range 400 miles

ASSIGNMENT 3

Textbook assignment: Chapter 3, "Principles of Transmission Lines," pages 3-1 through 3-58.

3-1. A transmission line is designed to perform which of the following functions?

1. Disperse energy in all directions
2. Detune a transmitter to match the load
3. Guide electrical energy from point to point
4. Replace the antenna in a communications system

3-2. All transmission lines must have two ends, the input end and the output end. What other name is given to the input end?

1. Sending end
2. Generator end
3. Transmitter end
4. Each of the above

3-3. A measurement of the voltage to current ratio (E_{in}/I_{in}) at the input end of a transmission line is called the

1. input-gain rate
2. input impedance
3. output impedance
4. voltage-gain ratio

3-4. Which of the following lines is NOT a transmission medium?

1. Load line
2. Coaxial line
3. Two-wire line
4. Twisted-pair line

3-5. Electrical power lines are most often made of which of the following types of transmission lines?

1. Twin-lead line
2. Shielded-pair line
3. Two-wire open line
4. Two-wire ribbon line

3-6. Uniform capacitance throughout the length of the line is an advantage of which of the following transmission lines?

1. Coaxial line
2. Twisted pair
3. Shielded pair
4. Two-wire open line

3-7. What is the primary advantage of a rigid coaxial line?

1. Low radiation losses
2. Inexpensive construction
3. Low high-frequency losses
4. Each of the above

3-8. Which of the following wave-guides is seldom used because of its large energy loss characteristics?

1. Metallic
2. Dielectric
3. Elliptical
4. Cylindrical

3-9. To some degree, transmission lines always exhibit which of the following types of losses?

1. I^2R
2. Inductor
3. Dielectric
4. Each of the above

3-10. Skin effect is classified as which of the following types of loss?

1. Copper
2. Voltage
3. Induction
4. Dielectric

3-11. What transmission-line loss is caused by magnetic lines of force not returning to the conductor?

1. Copper
2. Radiation
3. Induction
4. Dielectric

3-12. What is the electrical wave-length of 1 cycle if the frequency is 60 hertz?

1. 125,000 meters
2. 1,250,000 meters
3. 5,000,000 meters
4. 20,000,000 meters

3-13. A transmission line 10 meters in length is considered to be electrically long at which of the following frequencies?

1. 60 kilohertz
2. 600 kilohertz
3. 6 megahertz
4. 60 megahertz

3-14. The conductance value of a transmission line represents which of the following values?

1. Expected value of current flow through the insulation
2. Expected value of voltage supplied by the transmitter
3. Value of the lump and distributed constants of the line divided by impedance
4. Value of the lump and distributed constants of the line divided by impedance

3-15. Electrical constants in a transmission line are distributed in which of the following ways?

1. Into a single device
2. Along the length of the line
3. According to the thickness of the line
4. According to the cross-sectional area of the line

3-16. Leakage current in a two-wire transmission line is the current that flows through what component?

1. The resistor
2. The inductor
3. The insulator
4. The conductor

3-17. Conductance is the reciprocal of what electrical property?

1. Inductance
2. Resistance
3. Capacitance
4. Reciprocity

3-18. A transmission line that has current flowing through it has which, if any, of the following fields about it?

1. Electric field only
2. Magnetic field only
3. Both electric and magnetic fields
4. None of the above

3-19. Maximum transfer of energy from the source to the transmission line takes place when what impedance relationship exists between the source and the transmission line?

1. When the load impedance equals source impedance
2. When the load impedance is twice the source impedance
3. When the load impedance is half the source impedance
4. When the load impedance is one-fourth the source impedance

3-20. The characteristic impedance (Z_0) of a transmission line is calculated by using which of the following ratios?

1. R_s to R_{load} of the line
2. I_{max} to I_{min} at every point along the line
3. E to I at every point along the line
4. E_{in} to E_o of the line

3-21. For a given voltage, what determines the amount of current that will flow in a transmission line?

1. Conductance
2. Spacing of the wires
3. Diameter of the wires
4. Characteristic impedance

3-22. When the impedance of a transmission line is measured, which of the following values frequently is NOT considered?

1. Inductance
2. Resistance
3. Conductance
4. Capacitance

3-23. The characteristic impedance of a long transmission line may be determined by using which of the following methods?

1. Trial and error
2. Calculating the impedance of the entire line
3. Calculating the impedances at each end of the line
4. Adding the impedances of successive short sections

3-24. When should lumped values for transmission-line constants be used to calculate characteristic impedance?

1. When the line is short compared to one wavelength
2. When the line is long compared to one wavelength
3. When the line is infinitely long

3-25. In actual practice, the characteristic impedance of a transmission line is usually within which of the following resistance ranges?

1. 0 to 0.9 ohm
2. 1 to 49 ohms
3. 50 to 600 ohms
4. 601 to 1,000 ohms

3-26. The input impedance of a transmission line is affected by which of the following properties?

1. Radiation loss
2. Series inductance
3. Parallel capacitance
4. Each of the above

3-27. When a dc voltage is applied to a transmission line and the load absorbs all the energy, what is the resulting relationship between current and voltage?

1. They are in phase with each other
2. They are equal to Z_0 of the line
3. They are out of phase with each other
4. They are evenly distributed along the line

3-28. The initial waves that travel from the source to the load of a transmission line are referred to as what type of waves?

1. Incident
2. Refracted
3. Reflected
4. Diffracted

3-29. Waves that travel from the output end to the input end of a transmission line are referred to as what type of waves?

1. Incident
2. Refracted
3. Reflected
4. Diffracted

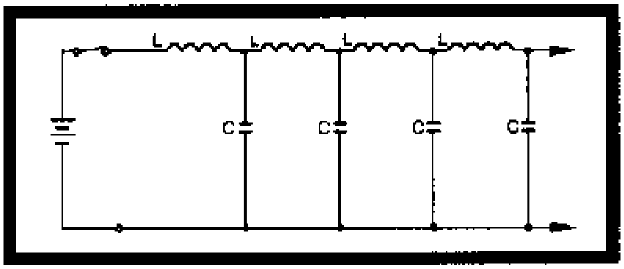


Figure 3-A.—Equivalent infinite transmission line.

IN ANSWERING QUESTION 3-30, REFER TO FIGURE 3-A.

3-30. When a dc voltage is applied to the equivalent infinite line in the figure, which of the following conditions occurs along the length of the line?

1. Standing waves of voltage form
2. Standing waves of current form
3. Current flows indefinitely
4. Voltage appears for a short time

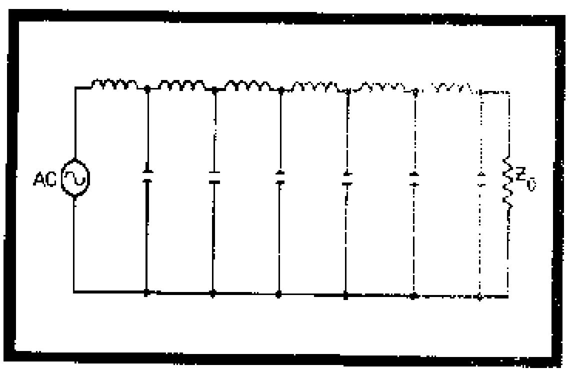


Figure 3-B.—Equivalent transmission line.

IN ANSWERING QUESTION 3-31, REFER TO FIGURE 3-B.

3-31. Compared to a dc input, what relative amount of time is required for an ac input voltage to travel the length of the line shown in the circuit?

1. Less
2. Same
3. More

3-32. The instantaneous voltage on an infinite transmission line can be plotted against time by using which of the following instruments?

1. A wavemeter
2. A multimeter
3. An oscilloscope
4. A spectrum analyzer

3-33. On an infinite transmission line with an ac voltage applied, which of the following is an accurate description of the effective voltage distribution along the line?

1. Voltage is 0 at all points
2. Voltage is constant at all points
3. Voltage varies at a sine-wave rate
4. Voltage varies at double the sine-wave rate

3-34. The velocity of propagation on a transmission line is controlled by which of the following line characteristics?

1. Conductance
2. Inductance only
3. Capacitance only
4. Capacitance and inductance

3-35. The total charge on a transmission line is equal to the current multiplied by which of the following factors?

1. Time
2. Power
3. Voltage
4. Resistance

3-36. With only capacitance and inductance of the line given, the time (T) required for a voltage change to travel down a transmission line can be found by what formula? The characteristic impedance for an infinite transmission line can be figured using which of the following ratios?

1. $T = \sqrt{\frac{L}{C}}$
2. $T = \sqrt{LC}$
3. $T = L + C$
4. $T = L - C$

3-37. The characteristic impedance for an infinite transmission line can be figured using which of the following ratios?

1. Input current to velocity
2. Input voltage to input current
3. Input voltage to line resistance
4. Input current to line resistance

3-38. The characteristic impedance of a transmission line can be figured by using which of the following formulas?

1. $Z_0 = \frac{1}{LC}$
2. $Z_0 = \sqrt{LC}$
3. $Z_0 = \frac{C}{L}$
4. $Z_0 = \sqrt{\frac{L}{C}}$

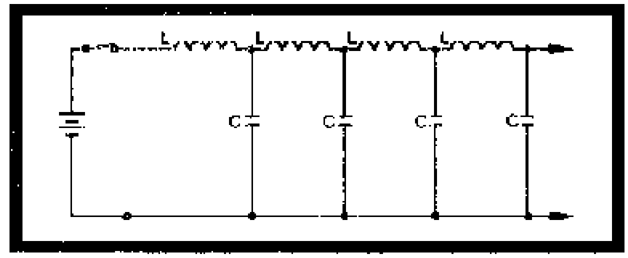


Figure 3-C.—Equivalent transmission line.

IN ANSWERING QUESTIONS 3-39 AND 3-40, REFER TO FIGURE 3-C. ASSUME THAT THE LINE IS 1,200 FEET LONG. A 150-FOOT SECTION IS MEASURED TO DETERMINE L AND C. THE 150-FOOT SECTION HAS AN INDUCTANCE OF 0.36 MILLIHENRIES AND A CAPACITANCE OF 1,000 PICOFARADS.

3-39. What is the characteristic impedance of the line?

1. 400 ohms
2. 600 ohms
3. 800 ohms
4. 900 ohms

3-40. What is the velocity of the wave on the 150-foot section?

1. 210,000,000 fps
2. 225,000,000 fps
3. 250,000,000 fps
4. 275,000,000 fps

3-41. If a transmission line is open-ended, which of the following conditions describes its terminating impedance?

1. Finite
2. Infinitely large
3. Equal to load impedance
4. Equal to source impedance

3-42. When a transmission line is not terminated in its characteristic impedance (Z_0), what happens to the incident energy that is NOT transferred to the load?

1. It is returned along the transmission line
2. It is radiated into space
3. It is absorbed by the line
4. It is converted to heat energy

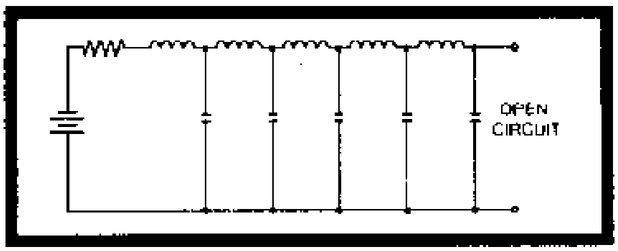


Figure 3-D.—Open-ended transmission line.

IN ANSWERING QUESTIONS 3-43 AND 3-44, REFER TO FIGURE 3-D.

3-43. When the dc voltage reaches the open end of the transmission line in the figure and is reflected, it has which, if any, of the following changes?

1. Increased amplitude
2. Decreased amplitude
3. The opposite polarity
4. None of the above

3-44. When the dc current reaches the open end of the transmission line and is reflected, it has which, if any, of the following changes?

1. Increased amplitude
2. Decreased amplitude
3. The opposite polarity
4. None of the above

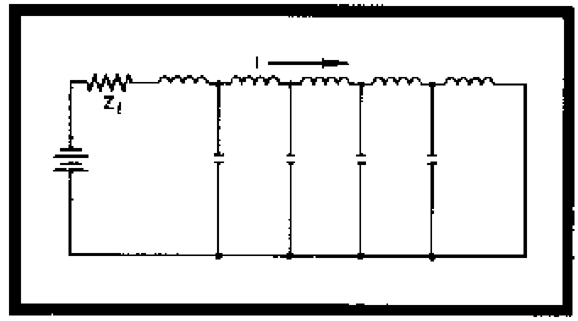


Figure 3-E.—Short-circuited transmission line.

IN ANSWERING QUESTIONS 3-45 AND 3-46, REFER TO FIGURE 3-E.

3-45. When the dc voltage reaches the shorted end of the transmission line, it is reflected. It has which, if any, of the following changes?

1. Increased amplitude
2. Decreased amplitude
3. The opposite polarity
4. None

3-46. When the dc current reaches the shorted end of the transmission line, it is reflected. It has which, if any, of the following changes?

1. Decrease amplitude
2. Increase amplitude
3. Increased polarity
4. None of the above

3-47. In an open-ended transmission line with an ac signal applied, what is the phase relationship between the incident and reflected voltage waves?

1. In phase
2. 45 degrees out of phase
3. 90 degrees out of phase
4. 180 degrees out of phase

3-48. The resultant of the incident and reflected voltage waves is called the standing wave. Its value is figured by using which of the following procedures?

1. Adding the effective values of the two waveforms
2. Algebraically adding the instantaneous values of the two waveforms
3. Algebraically subtracting the instantaneous values of the two waveforms
4. Taking the square root of the product of the incident and reflected voltages

3-49. On an open-ended transmission line that is carrying an ac signal, what is the total number of moving voltage waves?

1. One
2. Two
3. Three
4. Four

3-50. At the end of an open-ended transmission line, which, if any, of the following voltage waves is at its maximum value?

1. Incident
2. Reflected
3. Resultant
4. None

3-51. On a transmission line that is carrying an ac signal, what is the relative value of the resultant voltage wave 1/4 wavelength from the open end?

1. Maximum positive
2. Maximum negative
3. Zero

3-52. In an open-ended transmission line, the resultant ac current waveform is always zero at what point(s)?

1. At the open end only
2. 1/2 wavelength from the open-end only
3. At the open end and 1/2 wavelength from the open-end

3-53. The resultant waveform obtained by adding the incident wave to the reflected wave is referred to as a/an

1. standing wave
2. negative wave
3. algebraic wave
4. concentrated wave

3-54. On an open-ended transmission line, what is the phase relationship between the standing waves of voltage and current?

1. In phase
2. 45 degrees out of phase
3. 90 degrees out of phase
4. 180 degrees out of phase

3-55. Which of the following conditions exist at the end of a shorted transmission line?

1. Maximum voltage and minimum current
2. Maximum voltage and maximum current
3. Minimum voltage and maximum current
4. Minimum voltage and minimum current

3-56. Transmission line is considered to be nonresonant (flat) when it is terminated in which of the following ways?

1. In an impedance equal to Z_0
2. In an impedance that is infinite
3. In an inductive reactance greater than Z_0
4. In a capacitive reactance greater than Z_0

3-57. Of the following terms, which one is used for the nonresonant transmission line?

1. A tuned line
2. A shorted line
3. An untuned line
4. A terminated line

3-58. A transmission line that is resonant is sometimes referred to as which of the following types of lines?

1. Tuned
2. Matched
3. Untuned
4. Unmatched

3-59. A short-circuited section of transmission line that is an odd number of quarter-wavelengths long shows the same characteristics as which of the following devices?

1. A series-resonant circuit
2. A parallel-resonant circuit
3. An inductive reactance equal to Z_0
4. A capacitive reactance equal to Z_0

3-60. Which of the following circuits appears as a very high resistance at resonance?

1. Nonresonant
2. Series-resonant
3. Parallel-resonant
4. Each of the above

3-61. When a series-resonant circuit is resonant at a frequency above the generator frequency, it acts as what type of circuit?

1. Open
2. Resistive
3. Inductive
4. Capacitive

3-62. Which of the following sections of transmission line can be used as a parallel-resonant circuit?

1. A shorted 1/4-wavelength section
2. An open 1/4-wavelength section
3. A shorted 1/2-wavelength section
4. An open 3/4-wavelength section

3-63. A generator connected to an open-ended line greater than 1/4 wave-length but less than 1/2 wave-length senses which of the following circuit component characteristics?

1. Zero reactance
2. Low resistance
3. Inductive reactance
4. Capacitive reactance

3-64. Which of the following conditions of current (I) and impedance (Z) exist at even quarter-wave points on a shorted transmission line?

1. Low I, low Z
2. Low I, high Z
3. High I, high Z
4. High I, low Z

3-65. What is the maximum distance, in wavelengths (λ), between adjacent zero-current points on an open-circuited line?

1. 1 λ
2. 1/2 λ
3. 1/4 λ
4. 1/8 λ

3-66. When a line is terminated in a capacitance, the capacitor performs which, if any, of the following circuit actions?

1. It absorbs all the energy
2. It reflects all the energy
3. It reacts as if it were a short
4. None

3-67. When a transmission line is terminated in an inductive reactance, which, if any, of the following phase shifts takes place with respect to the current and voltage?

1. Only voltage is phase-shifted
2. Only current is phase-shifted
3. Both voltage and current are phase-shifted
4. None

3-68. When a transmission line is terminated in a resistance greater than Z_0, which of the following conditions exist?

1. The end of the line appears as an open circuit
2. Standing waves appear on the line
3. Voltage is maximum and current is minimum at the end of the line
4. Each of the above

3-69. On a transmission line, reflections begin at which of the following locations?

1. At the load
2. At the source
3. At the middle
4. At the half-wavelength point

3-70. The ratio of maximum voltage to minimum voltage on a transmission line is referred to as the

1. rswr
2. pswr
3. vswr
4. iswr

3-71. Which of the following ratios samples the magnetic field along a line?

1. Vswr
2. Pswr
3. Iswr
4. Rswr

ASSIGNMENT 4

Textbook assignment: Chapter 4, "Antennas," pages 4-1 through 4-60.

4-1. Radio energy is transmitted through which of the following mediums?

1. Rock
2. Soil
3. Water
4. Space

4-2. Energy is transmitted from a transmitter into space using which of the following devices?

1. A receiver
2. A delay line
3. A receiving antenna
4. A transmitting antenna

4-3. Transmitted rf energy takes what form as it is sent into space?

1. A magnetic field only
2. An electric field only
3. An electromagnetic field
4. A static dielectric field

4-4. The dimensions of a transmitting antenna are determined by which of the following factors?

1. Transmitted power
2. Transmitted frequency
3. Distance to the receiver
4. Antenna height above the ground

4-5. A device used to radiate or receive electromagnetic wave energy is referred to as a/an

1. feeder
2. antenna
3. transmitter
4. coupling device

4-6. An antenna that can be mounted to radiate rf energy either vertically or horizontally is classified as which of the following types?

1. Hertz
2. Marconi
3. Quarter-wave
4. Both 2 and 3 above

4-7. A complete antenna system consists of which of the following components?

1. A feeder, a coupling device, and a transmitter
2. A feeder line, a coupling device, and an antenna
3. An antenna, a transmission line, and a receiver
4. An impedance-matching device, a feeder, and a transmission line

4-8. What component in an antenna system transfers energy from the transmitter to the antenna?

1. A feeder
2. A delay line
3. A choke joint
4. A rotating joint

4-9. The type, size, and shape of an antenna are determined by which of the following factors?

1. Power output of the transmitter
2. Transmitter frequency
3. Direction to the receiver
4. Each of the above

4-10. Moving electric and magnetic fields in space have what (a) phase and (b) angular relationships?

1. (a) In phase
   (b) Perpendicular
2. (a) In phase
   (b) Displaced 45º
3. (a) Out of phase
   (b) Displaced 45º
4. (a) Out of phase
   (b) Perpendicular

4-11. What is the length of each half of the wire for a dipole antenna?

1. Wavelength
2. 3/4 wavelength
3. 1/2 wavelength
4. 1/4 wavelength

4-12. On a dipole antenna, the sinusoidal variation in charge magnitude lags the sinusoidal variation in current by what amount?

1. 1 cycle
2. 1/2 cycle
3. 1/4 cycle
4. 1/8 cycle

4-13. On a standing wave, the points of high current and voltage are identified by which of the following terms?

1. Peaks
2. Nodes
3. Poles
4. Loops

4-14. The presence of standing waves indicates which of the following conditions of an antenna?

1. Resonance
2. Saturation
3. Nonresonance
4. Minimum efficiency

4-15. The antenna property that allows the same antenna to both transmit and receive energy is

1. gain
2. resonance
3. reciprocity
4. directivity

4-16. There is a ratio between the amount of energy propagated in certain directions by a directional antenna compared to the energy that would be propagated in these directions if the antenna were not directional. This ratio is known as which of the following antenna characteristics?

1. Gain
2. Directivity
3. Reciprocity
4. Polarization

4-17. The polarization plane of the radiation field is determined by which of the following fields?

1. Electric field only
2. Magnetic field only
3. Electromagnetic field

4-18. For best reception of a signal from a horizontally polarized antenna, the receiving antenna should be mounted so that it has what relationship with the transmitting antenna?

1. 0 degrees
2. 45 degrees
3. 90 degrees
4. 135 degrees

4-19. An electric field that rotates as it travels through space exhibits what type of polarization?

1. Vertical
2. Spherical
3. Elliptical
4. Horizontal

4-20. For ground-wave transmissions, what type of polarization is required?

1. Vertical
2. Spherical
3. Elliptical
4. Horizontal

4-21. For high-frequency operation, which of the following antenna polarization patterns is preferred?

1. Vertically polarized
2. Spherically polarized
3. Elliptically polarized
4. Horizontally polarized

4-22. Omnidirectional transmission is obtained from which of the following antennas?

1. Elliptically polarized
2. Horizontal half-wave
3. Vertical half-wave
4. Each of the above

4-23. With an antenna height of 40 feet and a transmitter frequency of 90 megahertz, which of the following antenna radiation patterns is best for transmitting over bodies of water?

1. Vertically polarized
2. Spherically polarized
3. Elliptically polarized
4. Horizontally polarized

4-24. To select a desired signal and discriminate against interfering signals from strong vhf and uhf broadcast transmissions, which of the following actions should you take?

1. Increase receiver gain
2. Make the transmitting antenna bi-directional
3. Use a vertically polarized receiving antenna
4. Use narrowly directional arrays as receiving antennas

4-25. A vertically mounted transmission line is LEAST affected by which of the following antenna radiation patterns?

1. Vertically polarized
2. Spherically polarized
3. Horizontally polarized
4. Elliptically polarized

4-26. An antenna with which of the following radiation resistance values will exhibit reduced efficiency?

1. 39 ohms
2. 82 ohms
3. 107 ohms
4. 150 ohms

4-27. An isotropic radiator radiates energy in which of the following patterns?

1. Vertical
2. Bi-directional
3. Unidirectional
4. Omnidirectional

4-28. An ordinary flashlight is an example of what type of radiator?

1. Isotropic
2. Polarized
3. Anisotropic
4. Stroboscopic

THIS SPACE LEFT BLANK INTENTIONALLY.

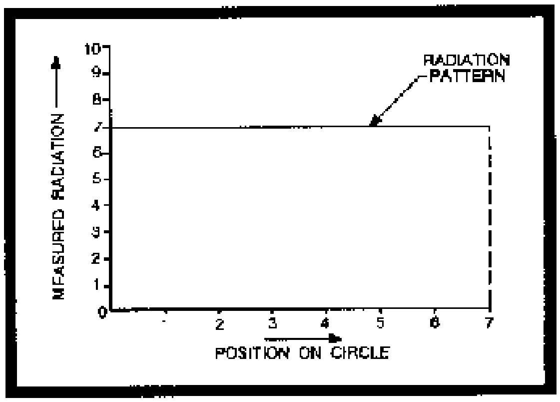


Figure 4-A.—Rectangular-coordinate graph.

IN ANSWERING QUESTION 4-29, REFER TO FIGURE 4-A.

4-29. How many points on the graph can represent the value of 7 radiation units at position 2 of the circle?

1. One
2. Two
3. Three
4. Four

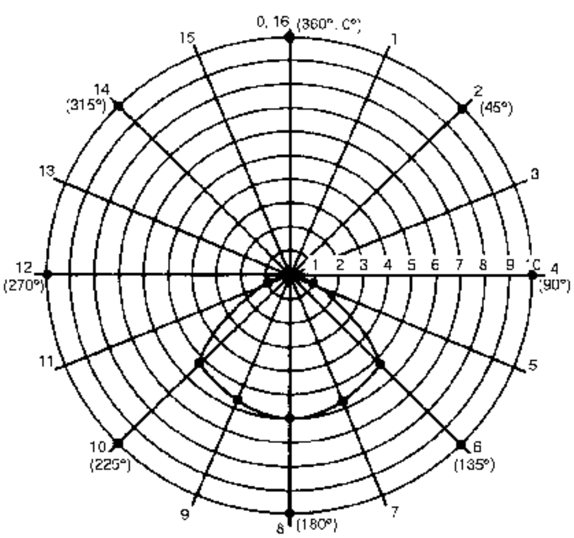


Figure 4-B.—Polar-coordinate graph.

IN ANSWERING QUESTIONS 4-30 AND 4-31, REFER TO FIGURE 4-B.

4-30. Compared with the rectangular-coordinate graph, the polar-coordinate graph has the advantage of showing which of the following antenna characteristics?

1. Polarization
2. Radiation pattern
3. Phase relationship
4. Gain versus directivity

4-31. The area enclosed by the radiation pattern is the

1. lobe
2. null
3. axis
4. coordinate

4-32. Inserting an inductor or capacitor in series with an antenna is one method of electrically changing the electrical length of an antenna. What is this method called?

1. Loading
2. Inserting
3. Unloading
4. Decoupling

4-33. Many complex antennas are constructed from what basic antenna?

1. The Marconi antenna
2. The full-wave antenna
3. The half-wave antenna
4. The quarter-wave antenna

4-34. On an energized half-wave antenna, which of the following electrical conditions exist?

1. Voltage is maximum at the ends
2. Voltage is minimum at the ends
3. Current is maximum at the ends
4. Impedance is minimum at the center

4-35. Which of the following radiation patterns is/are exhibited by a simple vertical doublet antenna?

1. Nondirectional in the horizontal plane
2. Directional in the vertical plane
3. Both 1 and 2 above
4. Spherical in all planes

4-36. A method of feeding energy to a half-wave antenna is to connect one end through a capacitor to the output stage. What is this method of feeding called?

1. End feed
2. Voltage feed
3. Both 1 and 2 above
4. Current feed

4-37. An antenna supplied by the center-feed method is fed at what point?

1. Low voltage and low current
2. Low voltage and high current
3. High voltage and low current
4. High voltage and high current

4-38. The basic Marconi antenna has which of the following characteristics?

1. One-quarter wavelength and ungrounded
2. One-half wavelength and grounded at one end
3. One-half wavelength and insulated from ground
4. One-quarter wavelength and grounded at one end

4-39. The Marconi antenna behaves as a dipole for which of the following reasons?

1. It is fed at one end
2. An image antenna is formed by reflections from the ground
3. A quarter-wavelength of conductor is buried in the ground and forms the rest of the dipole
4. The applied signal is rectified so that only half the signal will appear on the quarter-wave antenna

4-40. A series of conductors arranged in a radial pattern and buried in the ground beneath the antenna is referred to as a

1. ground spur
2. counterpoise
3. ground screen
4. ground reflector

4-41. A folded dipole can be used instead of a simple, center-fed dipole for which of the following purposes?

1. Matching voltage
2. Matching impedance
3. Increasing directivity
4. Decreasing directivity

4-42. An antenna arrangement that has elements aligned in a straight line is referred to as what type array?

1. Isotropic
2. Collinear
3. Line-of-sight
4. Unidirectional

4-43. To have current in two adjoining collinear half-wave elements in proper phase, they must be connected by which of the following stubs?

1. A shorted half-wave stub
2. An open quarter-wave stub
3. A shorted eighth-wave stub
4. A shorted quarter-wave stub

4-44. To select a desired signal and discriminate against interfering signals, the receiving antenna should have which of the following characteristics?

1. Be omnidirectional
2. Be highly directional
3. Be vertically polarized
4. Be horizontally polarized

4-45. Adding more elements to a collinear antenna array produces which of the following effects?

1. Increased gain
2. Decreased gain
3. Decreased directivity
4. Mismatched impedances

4-46. What is the maximum number of elements ordinarily used in a collinear array?

1. One
2. Two
3. Three
4. Four

4-47. Constructing a collinear array with elements longer than 1/2 wavelength has which of the following effects on antenna characteristics?

1. Increased gain
2. Decreased gain
3. Increased frequency range
4. Decreased frequency range

4-48. In a two-element collinear array, maximum gain is obtained when the center-to-center spacing between the ends of the elements is approximately what electrical distance?

1. Wavelength
2. 0.15 wavelength
3. 0.5 wavelength
4. 0.75 wavelength

4-49. Compared with collinear arrays, broadside arrays have which of the following advantages?

1. Sharper tuning
2. Broader bandwidth
3. Broader frequency response
4. Less coupling between dipole

4-50. Optimum gain is obtained from a broadside array when the spacing of its elements is which of the following distances?

1. One-half wavelength
2. One-quarter wavelength
3. Greater than one-half wavelength
4. Slightly less than one-quarter

4-51. An end-fire array physically resembles the collinear array except that it is more compact. What disadvantage does the endfire array possess?

1. It has lower gain
2. It has low radiation resistance
3. It has loose coupling
4. Each of the above

4-52. What is the range of electrical spacing between the elements of an end-fire array?

1. 3/4 to 1 wavelength
2. 1/2 to 3/4 wavelength
3. 1/4 to 1/2 wavelength
4. 1/8 to 1/4 wavelength

4-53. The end-fire array produces what type of lobes, if any, along the axis of the array?

1. Minor lobes
2. Major lobes
3. None

4-54. Assuming that the elements are correctly spaced, the directivity of an end-fire array may be increased by which of the following actions?

1. Increasing the frequency
2. Decreasing the frequency
3. Decreasing the number of elements
4. Increasing the number of elements

4-55. A unidirectional pattern can be obtained from an end-fire array by using what phase relationship between the energy fed to adjacent elements?

1. 0º
2. 45º
3. 90º
4. 180º

4-56. Energy is fed to a parasitic element using what method?

1. Direct coupling
2. Inductive coupling
3. Capacitive coupling
4. Transmission-line coupling

4-57. The directivity pattern resulting from the action of parasitic elements depends on which of the following element characteristics?

1. Length of the element
2. Diameter of the element
3. Spacing between parasitic and driven elements
4. Each of the above

4-58. The advantages of unidirectivity and increased gain can best be obtained by using which of the following elements in a parasitic array?

1. Driven elements only
2. Reflector and director elements only
3. Reflector, director, and driven elements
4. Driven and director elements only

4-59. The ratio of energy radiated by an array in the principal direction of radiation to the energy radiated in the opposite direction describes which of the following relationships?

1. Side-to-side ratio
2. Front-to-back ratio
3. Driven-to-parasitic ratio
4. Reflector-to-director ratio

4-60. The Yagi antenna is an example of what type of antenna array?

1. Driven
2. End-fire
3. Multielement parasitic
4. Single-element parasitic

4-61. The addition of parasitic elements to the Yagi antenna has which of the following effects on antenna characteristics?

1. Increased gain
2. Narrower beam width
3. Narrower frequency response
4. Each of the above

4-62. An antenna which is designed especially for vertically-polarized ground waves at low frequencies is the

1. Yagi antenna
2. Marconi antenna
3. Beverage antenna
4. V antenna

4-63. What is the phase relationship of the signals that feed the V antenna?

1. 0º
2. 45º
3. 90º
4. 180º

4-64. A rhombic antenna is essentially a combination of which of the following antennas?

1. Two stacked long-wire radiators
2. Two V antennas placed side by side
3. Two collinear arrays in parallel
4. Four parallel half-wave radiators

4-65. A rhombic antenna has which of the following advantages?

1. Simple construction
2. Wide frequency range
3. Noncritical adjustment
4. Each of the above

4-66. The principal disadvantage of the rhombic antenna is its

1. poor directivity
2. large antenna site
3. low antenna voltage
4. high-frequency inefficiency

4-67. The unidirectional radiation pattern of the rhombic antenna is caused by which of the following antenna characteristics?

1. Size
2. Shape
3. Termination resistance
4. Frequency of the input energy

4-68. Horizontal half-wave antennas mounted at right angles to each other in the same horizontal plane make up which of the following antennas?

1. Rhombic
2. Flat-top
3. Turnstile
4. Ground-plane

4-69. The most common means of obtaining a low-radiation angle from a vertical quarter-wave antenna is by what procedure?

1. Decreasing power
2. Increasing frequency
3. Adding a ground plane
4. Rotating the antenna to a horizontal plane

4-70. A corner reflector antenna is used for which of the following purposes?

1. To decrease frequency range
2. To increase frequency range
3. To produce a unidirectional pattern
4. To produce an omnidirectional pattern

4-71. If a corner-reflector antenna is horizontally polarized, its radiation pattern will take on what shape?

1. A narrow beam in the horizontal plane
2. A narrow beam in the vertical plane
3. A beam similar to a half-wave dipole in the horizontal plane
4. A beam similar to a half-wave dipole with a reflector in the vertical plane

4-72. When radio or radar antennas are energized by transmitters, you must not go aloft until which of the following requirements are met?

1. A safety harness has been issued to you
2. All transmitters are secured and tagged
3. A working aloft "chit" has been filled out and signed by proper authority
4. Each of the above

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 11—Microwave Principles

NAVEDTRA 14183

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 11 of a series.

History of the course:

Sep 1998: Original edition released.
May 2003: Administrative update released. Technical content was reviewed and minor changes *made.*

NAVSUP Logistics Tracking Number
0504-LP-026-8360

TABLE OF CONTENTS

CHAPTER 1

WAVEGUIDE THEORY AND APPLICATION

LEARNING OBJECTIVES

Upon completion of this chapter the student will be able to:

1. Describe the development of the various types of waveguides in terms of their advantages and disadvantages.
2. Describe the physical dimensions of the various types of waveguides and explain the effects of those dimensions on power and frequency.
3. Explain the propagation of energy in waveguides in terms of electromagnetic field theory.
4. Identify the modes of operation in waveguides.
5. Explain the basic input/output methods used in waveguides.
6. Describe the basic principles of waveguide plumbing.
7. Explain the reasons for and the methods of terminating waveguides.
8. Explain the basic theory of operation and applications of directional couplers.
9. Describe the basic theory of operation, construction, and applications of cavity resonators.
10. Describe the basic theory of operation of waveguide junctions.
11. Explain the operation of ferrite devices in terms of their applications.

INTRODUCTION TO WAVEGUIDE THEORY AND APPLICATION

That portion of the electromagnetic spectrum which falls between 1000 megahertz and 100,000 megahertz is referred to as the MICROWAVE region. Before discussing the principles and applications of microwave frequencies, the meaning of the term microwave as it is used in this module must be established. On the surface, the definition of a microwave would appear to be simple because, in electronics, the prefix "micro" normally means a millionth part of a unit. Micro also means small, which is a relative term, and it is used in that sense in this module. Microwave is a term loosely applied to identify electromagnetic waves above 1000 megahertz in frequency because of the short physical wavelengths of these frequencies. Short wavelength energy offers distinct advantages in many applications. For instance, excellent directivity can be obtained using relatively small antennas and low-power transmitters. These features are ideal for use in both military and civilian radar and communication applications. Small antennas and other small components are made possible by microwave frequency applications. This is an important consideration in shipboard equipment planning where space and weight are major problems. Microwave frequency usage is especially important in the design of shipboard radar because it makes possible the detection of smaller targets.

Microwave frequencies present special problems in transmission, generation, and circuit design that are not encountered at lower frequencies. Conventional circuit theory is based on voltages and currents while microwave theory is based on electromagnetic fields. The concept of electromagnetic field interaction is not entirely new, since electromagnetic fields form the basis of all antenna theory. However, many students of electronics find electromagnetic field theory very difficult to visualize and understand. This module will present the principles of microwave theory in the simplest terms possible but many of the concepts are still somewhat difficult to thoroughly understand. Therefore, you must realize that this module will require very careful study for you to properly understand microwave theory. Antenna fundamentals were covered in *NEETS*, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*.

This module will show you the solutions to problems encountered at microwave frequencies, beginning with the transmission of microwave energy and continuing through to waveguides in chapter 1. Later chapters will cover the theory of operation of microwave components, circuits, and antennas. The application of these concepts will be discussed more thoroughly in later *NEETS* modules on radar and communications.

Q-1. What is the region of the frequency spectrum from 1000 MHz to 100,000 MHz called?

Q-2. Microwave theory is based upon what concept

WAVEGUIDE THEORY

The two-wire transmission line used in conventional circuits is inefficient for transferring electromagnetic energy at microwave frequencies. At these frequencies, energy escapes by radiation because the fields are not confined in all directions, as illustrated in figure 1-1. Coaxial lines are more efficient than two-wire lines for transferring electromagnetic energy because the fields are completely confined by the conductors, as illustrated in figure 1-2.

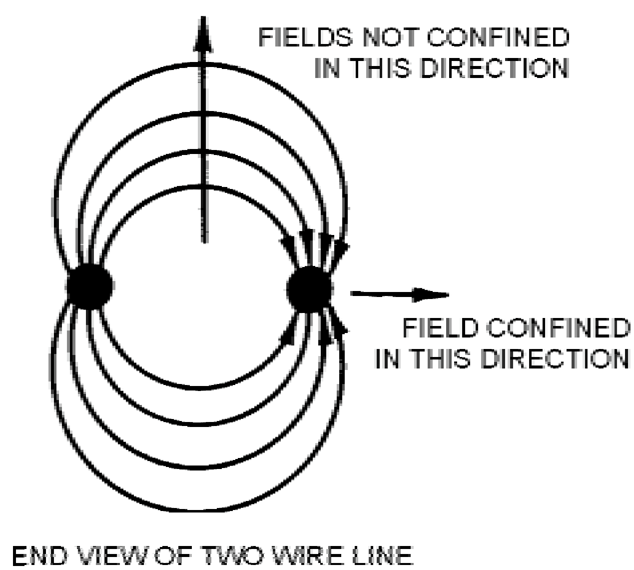


Figure 1-1.—Fields confined in two directions only.

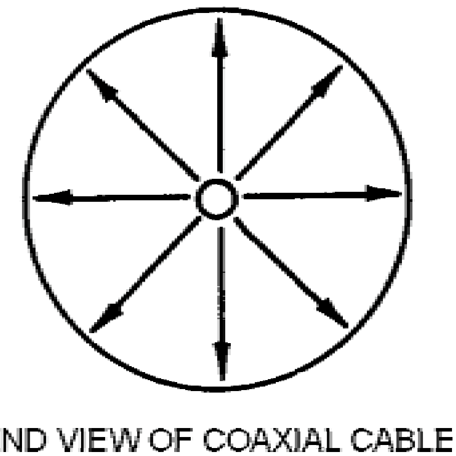


Figure 1-2.—Fields confined in all directions.

Waveguides are the most efficient way to transfer electromagnetic energy. WAVEGUIDES are essentially coaxial lines without center conductors. They are constructed from conductive material and may be rectangular, circular, or elliptical in shape, as shown in figure 1-3.

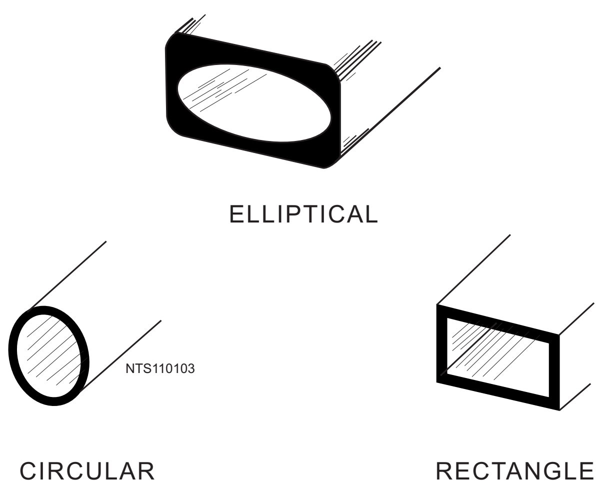


Figure 1-3.—Waveguide shapes.

Waveguide Advantages

Waveguides have several advantages over two-wire and coaxial transmission lines. For example, the large surface area of waveguides greatly reduces COPPER (I^2R) LOSSES. Two-wire transmission lines have large copper losses because they have a relatively small surface area. The surface area of the outer conductor of a coaxial cable is large, but the surface area of the inner conductor is relatively small. At microwave frequencies, the current-carrying area of the inner conductor is restricted to a very small layer at the surface of the conductor by an action called SKIN EFFECT.

Skin effect was discussed in *NEETS*, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*, Chapter 3. Skin effect tends to increase the effective resistance of the conductor. Although energy transfer in coaxial cable is caused by electromagnetic field motion, the magnitude of the field is limited by the size of the current-carrying area of the inner conductor. The small size of the center conductor is even further reduced by skin effect and energy transmission by coaxial cable becomes less efficient than by waveguides. DIELECTRIC LOSSES are also lower in waveguides than in two-wire and coaxial transmission lines. Dielectric losses in two-wire and coaxial lines are caused by the heating of the insulation between the conductors. The insulation behaves as the dielectric of a capacitor formed by the two wires of the transmission line. A voltage potential across the two wires causes heating of the dielectric and results in a power loss. In practical applications, the actual breakdown of the insulation between the conductors of a transmission line is more frequently a problem than is the dielectric loss.

This breakdown is usually caused by stationary voltage spikes or "nodes" which are caused by standing waves. Standing waves are stationary and occur when part of the energy traveling down the line is reflected by an impedance mismatch with the load. The voltage potential of the standing waves at the points of greatest magnitude can become large enough to break down the insulation between transmission line conductors.

The dielectric in waveguides is air, which has a much lower dielectric loss than conventional insulating materials. However, waveguides are also subject to dielectric breakdown caused by standing waves. Standing waves in waveguides cause arcing which decreases the efficiency of energy transfer and can severely damage the waveguide. Also since the electromagnetic fields are completely contained within the waveguide, radiation losses are kept very low.

Power-handling capability is another advantage of waveguides. Waveguides can handle more power than coaxial lines of the same size because power-handling capability is directly related to the distance between conductors. Figure 1-4 illustrates the greater distance between conductors in a waveguide.

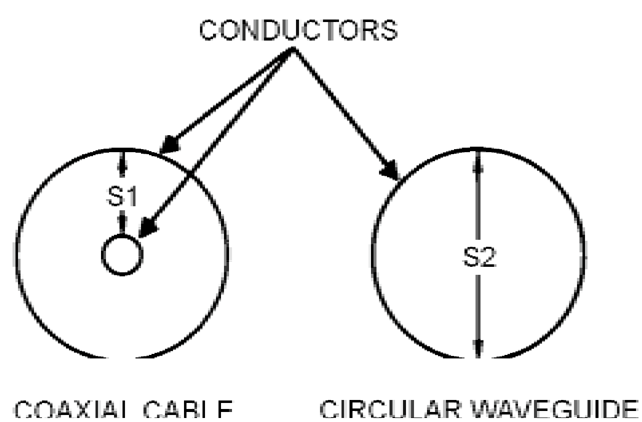


Figure 1-4.—Comparison of spacing in coaxial cable and a circular waveguide.

In view of the advantages of waveguides, you would think that waveguides should be the only type of transmission lines used. However, waveguides have certain disadvantages that make them practical for use only at microwave frequencies.

Waveguide Disadvantages

Physical size is the primary lower-frequency limitation of waveguides. The width of a waveguide must be approximately a half wavelength at the frequency of the wave to be transported. For example, a waveguide for use at 1 megahertz would be about 500 feet wide. This makes the use of waveguides at frequencies below 1000 megahertz increasingly impractical. The lower frequency range of any system using waveguides is limited by the physical dimensions of the waveguides.

Waveguides are difficult to install because of their rigid, hollow-pipe shape. Special couplings at the joints are required to assure proper operation. Also, the inside surfaces of waveguides are often plated with silver or gold to reduce skin effect losses. These requirements increase the costs and decrease the practicality of waveguide systems at any other than microwave frequencies.

Q-3. Why are coaxial lines more efficient at microwave frequencies than two-wire transmission lines?

Q-4. What kind of material must be used in the construction of waveguides?

Q-5. The large surface area of a waveguide greatly reduces what type of loss that is common in two-wire and coaxial lines?

Q-6. What causes the current-carrying area at the center conductor of a coaxial line to be restricted to a small layer at the surface?

Q-7. What is used as a dielectric in waveguides?

Q-8. What is the primary lower-frequency limitation of waveguides?

Developing the Waveguide from Parallel Lines

You may better understand the transition from ordinary transmission line concepts to waveguide theories by considering the development of a waveguide from a two-wire transmission line. Figure 1-5 shows a section of two-wire transmission line supported on two insulators. At the junction with the line, the insulators must present a very high impedance to ground for proper operation of the line. A low impedance insulator would obviously short-circuit the line to ground, and this is what happens at very high frequencies. Ordinary insulators display the characteristics of the dielectric of a capacitor formed by the wire and ground. As the frequency increases, the overall impedance decreases. A better high-frequency insulator is a quarter-wave section of transmission line shorted at one end. Such an insulator is shown in figure 1-6. The impedance of a shorted quarter-wave section is very high at the open-end junction with the two-wire transmission line. This type of insulator is known as a METALLIC INSULATOR and may be placed anywhere along a two-wire line. Note that quarter-wave sections are insulators at only one frequency. This severely limits the bandwidth, efficiency, and application of this type of two-wire line.

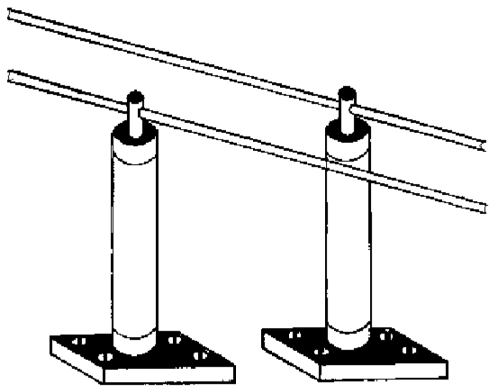


Figure 1-5.—Two-wire transmission line using ordinary insulators.

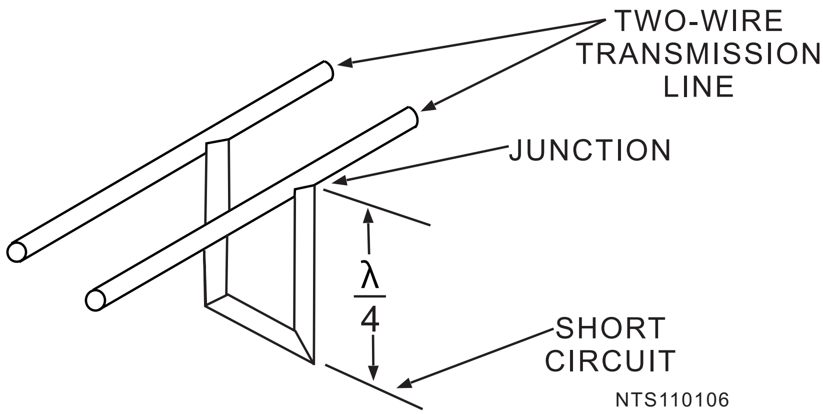


Figure 1-6.—Quarter-wave section of transmission line shorted at one end.

Figure 1-7 shows several metallic insulators on each side of a two-wire transmission line. As more insulators are added, each section makes contact with the next, and a rectangular waveguide is formed. The lines become part of the walls of the waveguide, as illustrated in figure 1-8. The energy is then conducted within the hollow waveguide instead of along the two-wire transmission line.

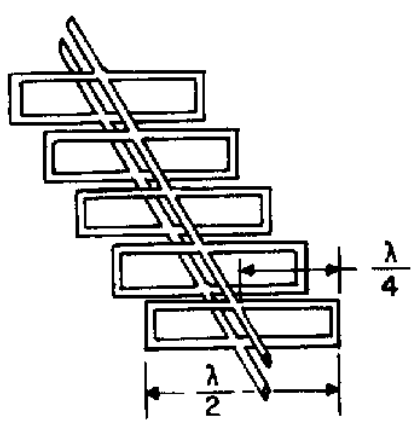


Figure 1-7.—Metallic insulators on each side of a two-wire line.

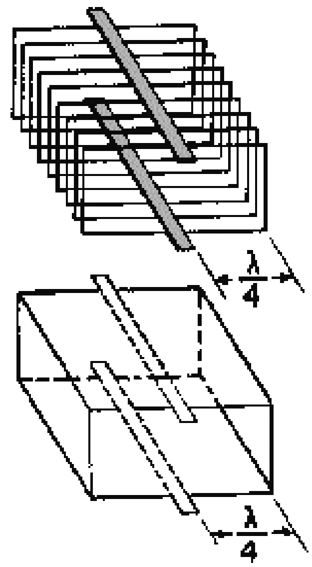


Figure 1-8.—Forming a waveguide by adding quarter-wave sections.

The comparison of the way electromagnetic fields work on a transmission line and in a waveguide is not exact. During the change from a two-wire line to a waveguide, the electromagnetic field configurations also undergo many changes. These will be discussed later in this chapter. As a result of these changes, the waveguide does not actually operate like a two-wire line that is completely shunted by quarter-wave sections. If it did, the use of a waveguide would be limited to a single-frequency wavelength that was four times the length of the quarter-wave sections. In fact, waves of this length cannot pass efficiently through waveguides. Only a small range of frequencies of somewhat shorter wavelength (higher frequency) can pass efficiently.

As shown in figure 1-9, the widest dimension of a waveguide is called the "a" dimension and determines the range of operating frequencies. The narrowest dimension determines the power-handling capability of the waveguide and is called the "b" dimension.

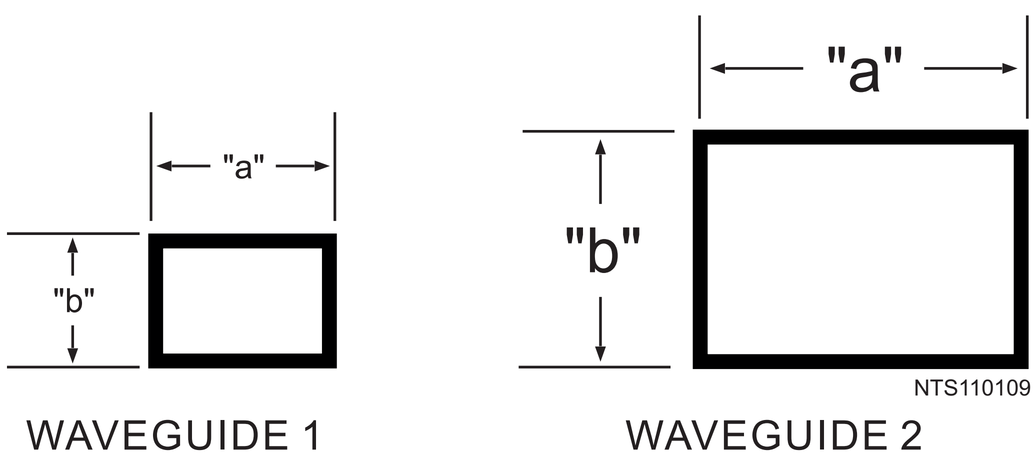


Figure 1-9.—Labeling waveguide dimensions.

NOTE

This method of labeling waveguides is not standard in all texts. Different methods may be used in other texts on microwave principles, but this method is in accordance with Navy Military Standards (MIL-STDS).

The ability of a waveguide of a given dimension to transport more than one frequency may be better understood by analyzing the actions illustrated in figure 1-10A, B, and C. A waveguide may be considered as having upper and lower quarter-wave sections and a central section which is a solid conductor called a BUS BAR. In figure 1-10A, distance mn is equal to distance pq, and both are equal to one quarter-wavelength (λ/4).

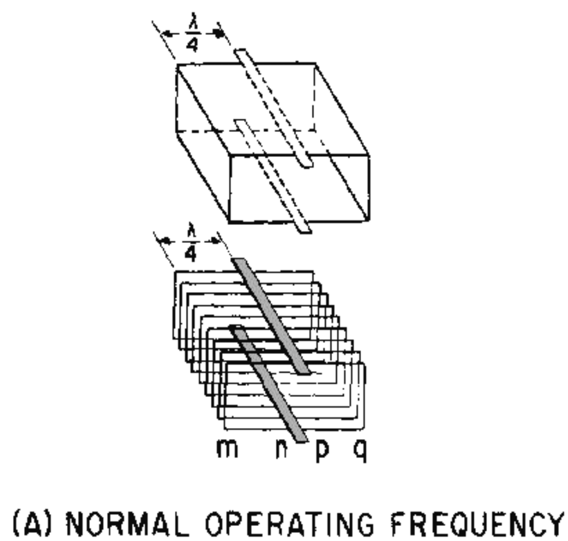
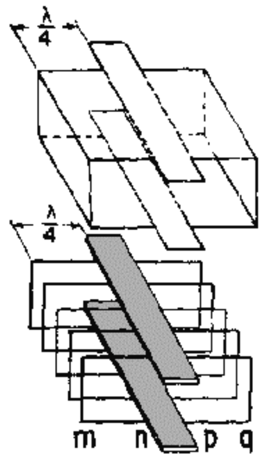


Figure 1-10A.—Frequency effects on a waveguide. NORMAL OPERATING FREQUENCY.

NOTE

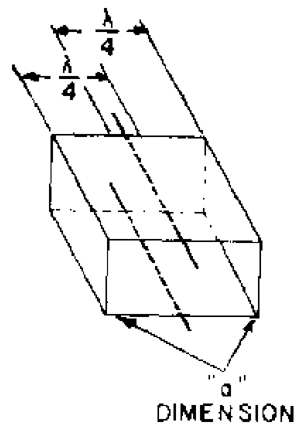
Throughout *NEETS*, 1/4λ and λ/4 are both used to represent one quarter-wavelength and are used interchangeably. Also, λ/2 and 3/2λ will be used to represent one half-wavelength and 1 1/2 wavelengths, respectively.

Distance np is the width of the bus bar. If the overall dimensions of the waveguide are held constant, the required length of the quarter-wave sections DECREASES as the frequency increases. As illustrated in figure 1-10B, this causes the width of the bus bar to INCREASE. In theory the waveguide could function at an infinite number of frequencies higher than the designed frequency; as the length of each quarter-wave section approaches zero, the bus bar continues to widen to fill the available space. However, in practice, an upper frequency limit is caused by modes of operation, which will be discussed later.



(B) INCREASING FREQUENCY

Figure 1-10B.—Frequency effects on a waveguide. INCREASING FREQUENCY.



(C) DECREASING FREQUENCY

Figure 1-10C.—Frequency effects on a waveguide. DECREASING FREQUENCY.

If the frequency of a signal is decreased so much that two quarter-wavelengths are longer than the wide dimension of a waveguide, energy will no longer pass through the waveguide. This is the lower frequency limit, or CUT-OFF FREQUENCY, of a given waveguide. In practical applications, the wide dimension of a waveguide is usually 0.7 wavelength at the operating frequency. This allows the waveguide to handle a small range of frequencies both above and below the operating frequency. The "b" dimension is governed by the breakdown potential of the dielectric, which is usually air. Dimensions ranging from 0.2 to 0.5 wavelength are common for the "b" sides of a waveguide.

Q-9. At very high frequencies, what characteristics are displayed by ordinary insulators?

Q-10. What type of insulator works well at very high frequencies?

Q-11. The frequency range of a waveguide is determined by what dimensison?

Q-12. What happens to the bus bar dimensions of the waveguide when the frequency is increased?

Q-13. When the frequency is decreased so that two quarter-wavelengths are longer than the "a" (wide) dimension of the waveguide, what will happen?

Energy Propagation in Waveguides

Since energy is transferred through waveguides by electromagnetic fields, you need a basic understanding of field theory. Both magnetic (H FIELD) and electric field (E FIELD) are present in waveguides, and the interaction of these fields causes energy to travel through the waveguide. This action is best understood by first looking at the properties of the two individual fields.

E FIELD.—An electric field exists when a difference of potential causes a stress in the dielectric between two points. The simplest electric field is one that forms between the plates of a capacitor when one plate is made positive compared to the other, as shown in figure 1-11A. The stress created in the dielectric is an electric field.

Electric fields are represented by arrows that point from the positive toward the negative potential. The number of arrows shows the relative strength of the field. In figure 1-11A, for example, evenly spaced arrows indicate the field is evenly distributed. For ease of explanation, the electric field is abbreviated E field, and the lines of stress are called E lines.

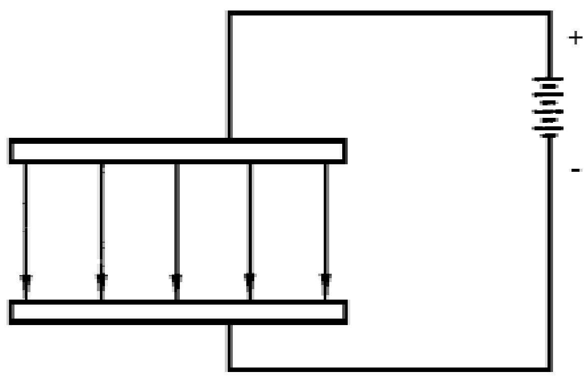


Figure 1-11A.—Simple electric fields. CAPACITOR.

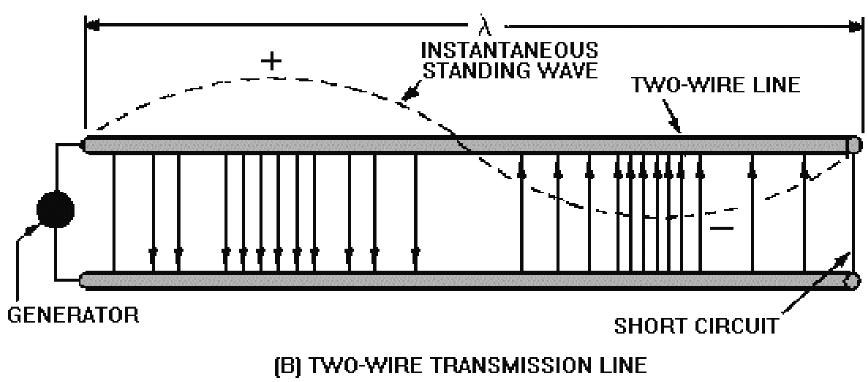


Figure 1-11B—Simple electric fields. TWO-WIRE TRANSMISSION LINE.

The two-wire transmission line, illustrated in figure 1-11B, has an instantaneous standing wave of voltage applied to it by the generator. The line is short-circuited at one-wavelength, at the positive and negative voltage peaks, but the arrows, representing each field, point in opposite directions. The voltage across the line varies sinusoidally. Therefore, the density of the E-lines varies sinusoidally.

The development of the E field in a waveguide can be illustrated by a two-wire transmission line separated by several, double quarter-wave sections, called half-wave frames, as illustrated in figure 1-12. As shown, the voltage across the two-wire line varies in a sine-wave pattern and the density of the E field also varies in a sine-wave pattern. The half-wave frames located at high-voltage points (1) and (3) have a strong E field. The frames at the zero-voltage points (2) have no E fields present. Frame (4) has a weak E field and is located at a point between maximum and minimum voltage. This illustration is a buildup to the three-dimensional aspect of the full E field in a waveguide.

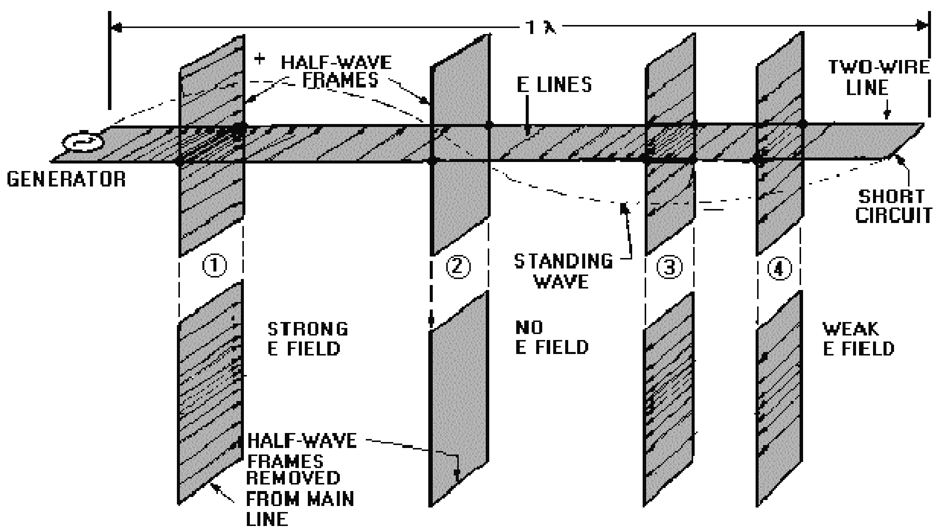


Figure 1-12.—E fields on a two-wire line with half-wave frames.

Figure 1-13, view (A), shows the E-field pattern created by a voltage sine wave applied to a one-wavelength section of waveguide shorted at one end. The electric fields are represented by the arrows shown in views (B) and (C). In the top view of view (A), the tip of each arrow is represented by a dot and the tail of each arrow is represented by an X. The E field varies in density at the same sine-wave rate as the applied voltage. This illustration represents the instant that the applied voltage wave is at its peak. At other times, the voltage and the E field in the waveguide vary continuously from zero to the peak value. Voltage and E-field polarity reverse with every reversal of the input. Note that the end view shown in view (B) shows the E field is maximum at the center and minimum near the walls of the waveguide. View (C) shows the arrangement of electromagnetic fields within a three-dimensional waveguide.

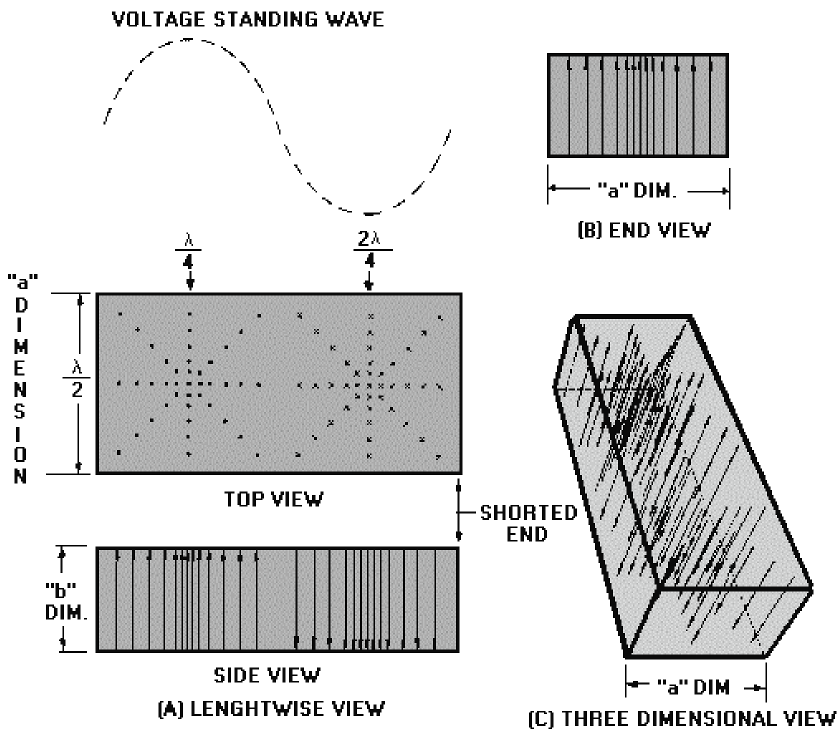


Figure 1-13.—E field of a voltage standing wave across a 1-wavelength section of a waveguide.

H FIELD.—The magnetic field in a waveguide is made up of magnetic lines of force that are caused by current flow through the conductive material of the waveguide. Magnetic lines of force, called H lines, are continuous closed loops, as shown in figure 1-14. All of the H lines associated with current are collectively called a magnetic field or H field. The strength of the H field, indicated by the number of H lines in a given area, varies directly with the amount of current.

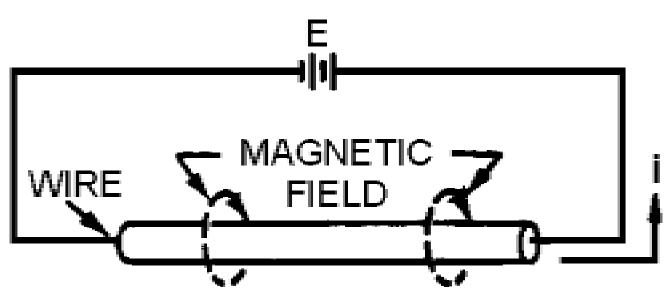


Figure 1-14.—Magnetic field on a single wire.

Although H lines encircle a single, straight wire, they behave differently when the wire is formed into a coil, as shown in figure 1-15. In a coil the individual H lines tend to form around each turn of wire. Since the H lines take opposite directions between adjacent turns, the field between the turns is cancelled. Inside and outside the coil, where the direction of each H field is the same, the fields join and form continuous H lines around the entire coil.

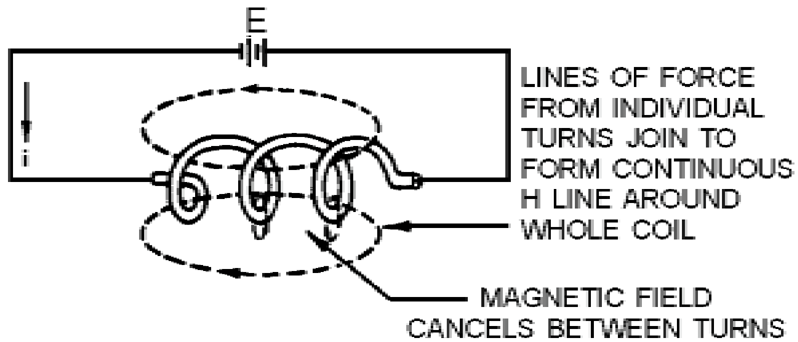


Figure 1-15.—Magnetic field on a coil.

A similar action takes place in a waveguide. In figure 1-16A, a two-wire line with quarter-wave sections is shown. Currents flow in the main line and in the quarter-wave sections. The current direction produces the individual H lines around each conductor as shown. When a large number of sections exist, the fields cancel between the sections, but the directions are the same both inside and outside the waveguide. At half-wave intervals on the main line, current will flow in opposite directions. This produces H-line loops having opposite directions. In figure 1-16A, current at the left end is opposite to the current at the right end. The individual loops on the main line are opposite in direction. All around the framework they join so that the long loop shown in figure 1-16B is formed. Outside the waveguide the individual loops cannot join to form a continuous loop. Thus, no magnetic field exists outside a waveguide.

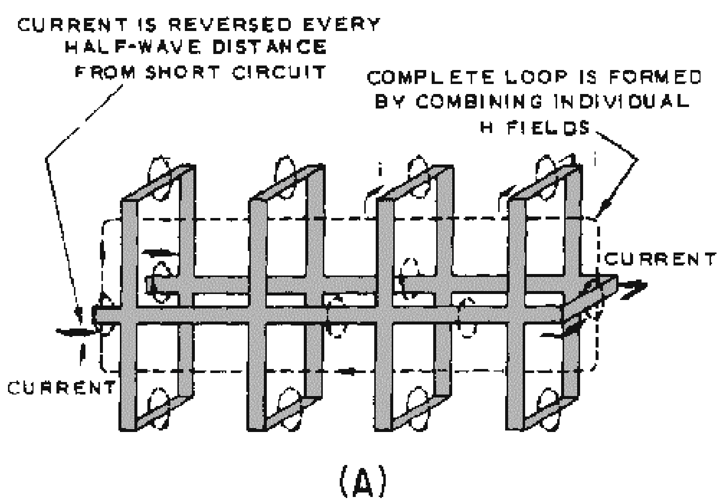


Figure 1-16A.—Magnetic fields on a two-wire line with half-wave frames.

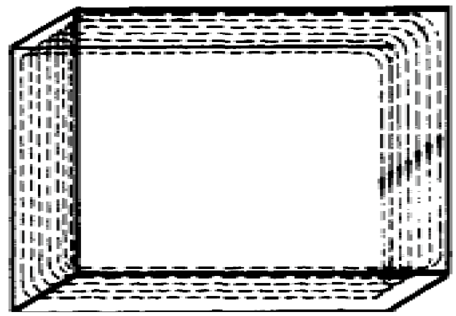


Figure 1-16B.—Magnetic fields on a two-wire line with half-wave frames.

If the two-wire line and the half-wave frames are developed into a waveguide that is closed at both ends (as shown in figure 1-16B), the distribution of H lines will be as shown in figure 1-17. If the waveguide is extended to 1 1/2λ, these H lines form complete loops at half-wave intervals with each group reversed in direction. Again, no H lines can form outside the waveguide as long as it is completely enclosed.

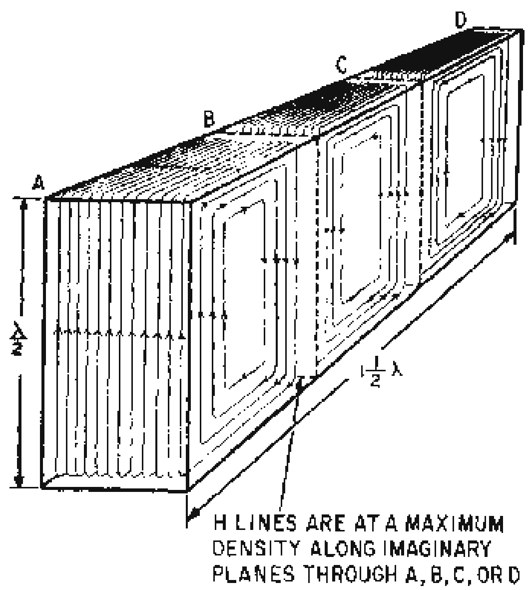


Figure 1-17.—Magnetic field pattern in a waveguide.

Figure 1-18 shows a cross-sectional view of the magnetic field pattern illustrated in figure 1-17. Note in view (A) that the field is strongest at the edges of the waveguide where the current is highest. The minimum field strength occurs at the zero-current points. View (B) shows the field pattern as it appears λ/4 from the end view of the waveguide. As with the previously discussed E fields, the H fields shown in figures 1-17 and 1-18 represent a condition that exists at only one instant in time. During the peak of the next half cycle of the input current, all field directions are reversed and the field will continue to change with changes in the input.

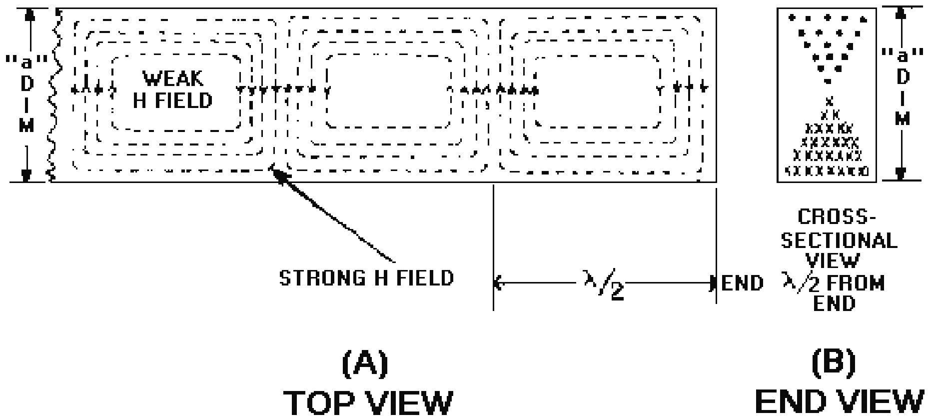


Figure 1-18.—Magnetic field in a waveguide three half-wavelengths long.

BOUNDARY CONDITIONS IN A WAVEGUIDE.—The travel of energy down a waveguide is similar, but not identical, to the travel of electromagnetic waves in free space. The difference is that the energy in a waveguide is confined to the physical limits of the guide. Two conditions, known as BOUNDARY CONDITIONS, must be satisfied for energy to travel through a waveguide.

The first boundary condition (illustrated in figure 1-19A) can be stated as follows:

For an electric field to exist at the surface of a conductor it must be perpendicular to the conductor.

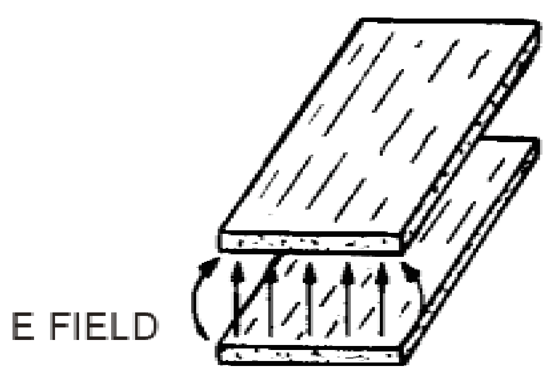


Figure 1-19A.—E field boundary condition. MEETS BOUNDARY CONDITIONS.

The opposite of this boundary condition, shown in figure 1-19B, is also true. An electric field CANNOT exist parallel to a perfect conductor.

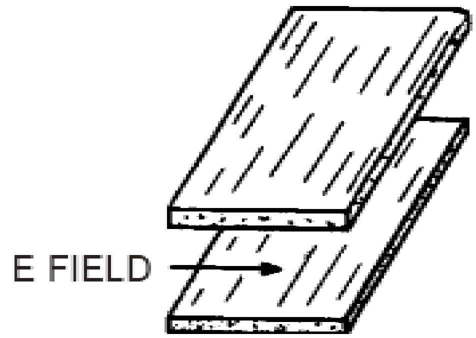


Figure 1-19B.—E field boundary condition. DOES NOT MEET BOUNDARY CONDITIONS.

The second boundary condition, which is illustrated in figure 1-20, can be stated as follows:

For a varying magnetic field to exist, it must form closed loops in parallel with the conductors and be perpendicular to the electric field.

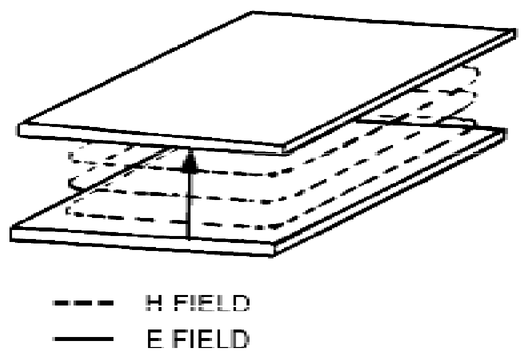


Figure 1-20.—H field boundary condition.

Since an E field causes a current flow that in turn produces an H field, both fields always exist at the same time in a waveguide. If a system satisfies one of these boundary conditions, it must also satisfy the other since neither field can exist alone. You should briefly review the principles of electromagnetic propagation in free space (*NEETS*, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas).* This review will help you understand how a waveguide satisfies the two boundary conditions necessary for energy propagation in a waveguide.

WAVEFRONTS WITHIN A WAVEGUIDE.—Electromagnetic energy transmitted into space consists of electric and magnetic fields that are at right angles (90 degrees) to each other and at right angles to the direction of propagation. A simple analogy to establish this relationship is by use of the right-hand rule for electromagnetic energy, based on the POYNTING VECTOR. It indicates that a screw (right-hand thread) with its axis perpendicular to the electric and magnetic fields will advance in the direction of propagation if the E field is rotated to the right (toward the H field). This rule is illustrated in figure 1-21.

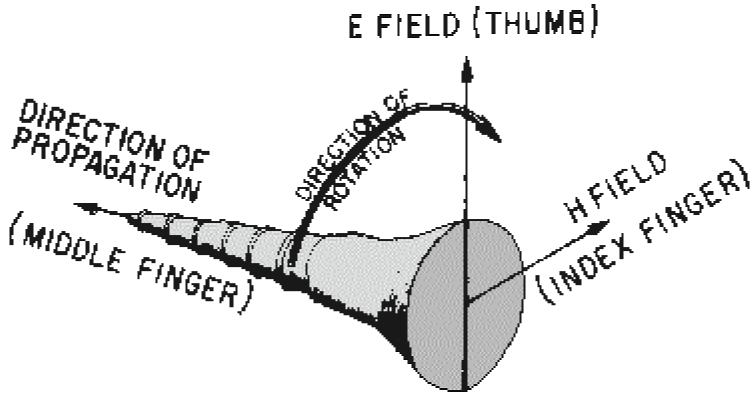


Figure 1-21.—The Poynting vector.

The combined electric and magnetic fields form a wavefront that can be represented by alternate negative and positive peaks at half-wavelength intervals, as illustrated in figure 1-22. Angle θ is the direction of travel of the wave with respect to some reference axis.

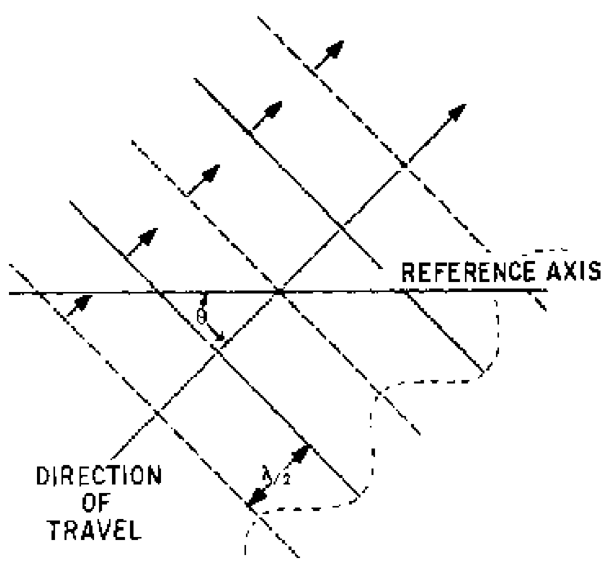


Figure 1-22.—Wavefronts in space.

If a second wavefront, differing only in the direction of travel, is present at the same time, a resultant of the two is formed. The resultant is illustrated in figure 1-23, and a close inspection reveals important characteristics of combined wavefronts. Both wavefronts add at all points on the reference axis and cancel at half-wavelength intervals from the reference axis. Therefore, alternate additions and cancellations of the two wavefronts occur at progressive half-wavelength increments from the reference axis. In figure 1-23, the lines labeled A, C, F, and H are addition points, and those labeled B, D, E, and G are cancellation points.

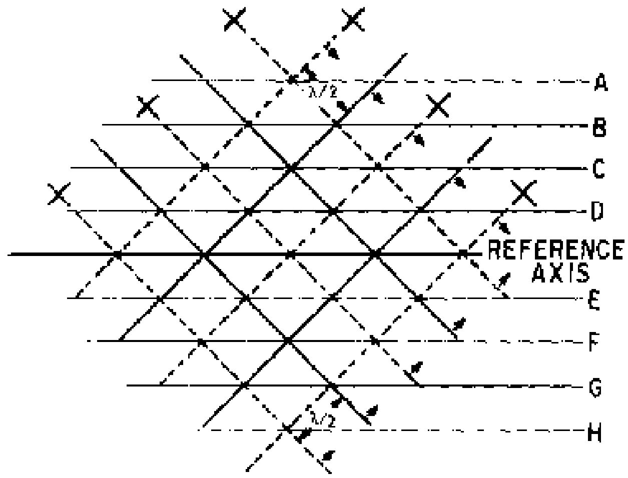


Figure 1-23.—Combined wavefronts.

If two conductive plates are placed along cancellation lines D and E or cancellation lines B and G, the first boundary condition for waveguides will be satisfied; that is, the E fields will be zero at the surface of the conductive plates. The second boundary condition is, therefore, automatically satisfied. Since these plates serve the same purpose as the "b" dimension walls of a waveguide, the "a" dimension walls can be added without affecting the magnetic or electric fields.

When a quarter-wavelength probe is inserted into a waveguide and supplied with microwave energy, it will act as a quarter-wave vertical antenna. Positive and negative wavefronts will be radiated, as shown in figure 1-24. Any portion of the wavefront traveling in the direction of arrow C will rapidly decrease to zero because it does not fulfill either of the required boundary conditions. The parts of the wavefronts that travel in the directions of arrows A and B will reflect from the walls and form reverse-phase wavefronts. These two wavefronts, and those that follow, are illustrated in figure 1-25. Notice that the wavefronts crisscross down the center of the waveguide and produce the same resultant field pattern that was shown in figure 1-23.

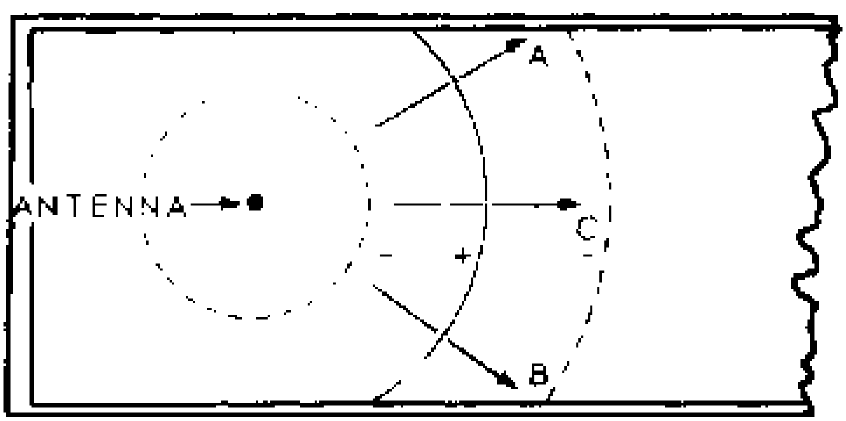
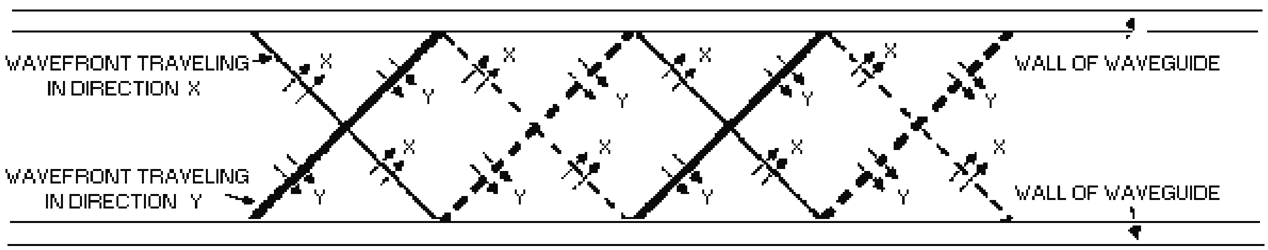
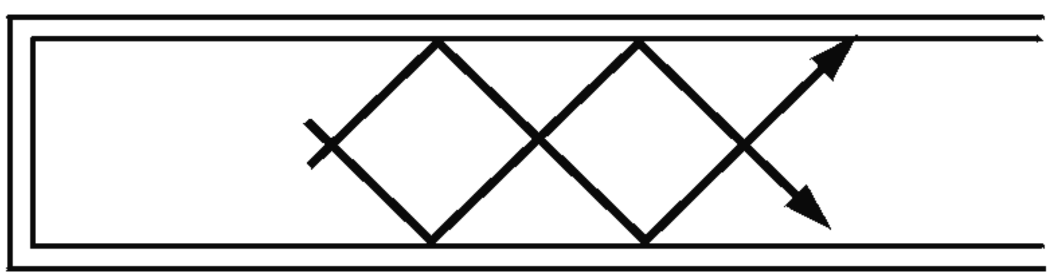


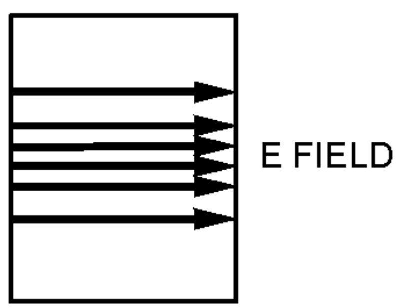
Figure 1-24.—Radiation from probe placed in a waveguide.



(A)



(B)



(C)

Figure 1-25.—Wavefronts in a waveguide.

The reflection of a single wavefront off the "b" wall of a waveguide is shown in figure 1-26. The wavefront is shown in view (A) as small particles. In views (B) and (C) particle 1 strikes the wall and is bounced back from the wall without losing velocity. If the wall is perfectly flat, the angle at which it strikes the wall, known as the angle of incidence (θ), is the same as the angle of reflection (ø) and are measured perpendicular to the waveguide surface. An instant after particle 1 strikes the wall, particle 2 strikes the wall, as shown in view (C), and reflects in the same manner. Because all the particles are traveling at the same velocity, particles 1 and 2 do not change their relative position with respect to each other. Therefore, the reflected wave has the same shape as the original. The remaining particles as shown in views (D), (E) and (F) reflect in the same manner. This process results in a reflected wavefront identical in shape, but opposite in polarity, to the incident wave.

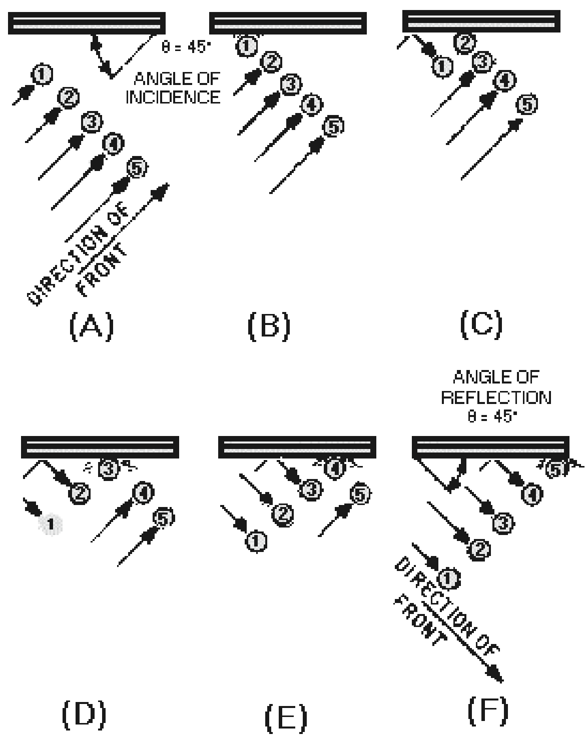


Figure 1-26.—Reflection of a single wavefront.

Figures 1-27A and 1-27B, each illustrate the direction of propagation of two different electromagnetic wavefronts of different frequencies being radiated into a waveguide by a probe. Note that only the direction of propagation is indicated by the lines and arrowheads. The wavefronts are at right angles to the direction of propagation. The angle of incidence (θ) and the angle of reflection (ø) of the wavefronts vary in size with the frequency of the input energy, but the angles of reflection are equal to each other in a waveguide. The CUTOFF FREQUENCY in a waveguide is a frequency that would cause angles of incidence and reflection to be zero degrees. At any frequency below the cutoff frequency, the wavefronts will be reflected back and forth across the guide (setting up standing waves) and no energy will be conducted down the waveguide.

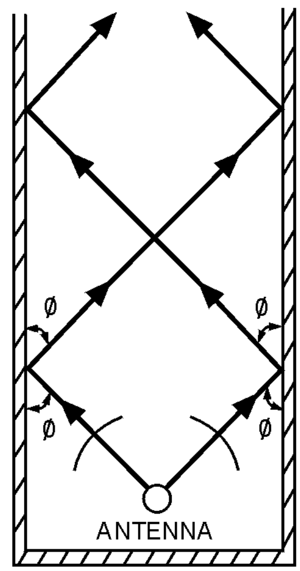


Figure 1-27A.—Different frequencies in a waveguide.

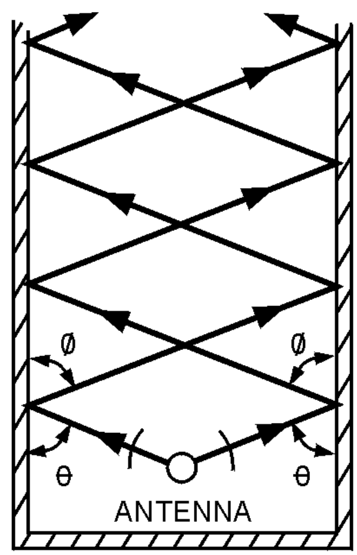


Figure 1-27B.—Different frequencies in a waveguide.

The velocity of propagation of a wave along a waveguide is less than its velocity through free space (speed of light). This lower velocity is caused by the zigzag path taken by the wavefront. The forward-progress velocity of the wavefront in a waveguide is called GROUP VELOCITY and is somewhat slower than the speed of light.

The group velocity of energy in a waveguide is determined by the reflection angle of the wavefronts off the "b" walls. The reflection angle is determined by the frequency of the input energy. This basic principle is illustrated in figures 1-28A, 1-28B, and 1-28C. As frequency is decreased, the reflection angle decreases causing the group velocity to decrease. The opposite is also true; increasing frequency increases the group velocity.

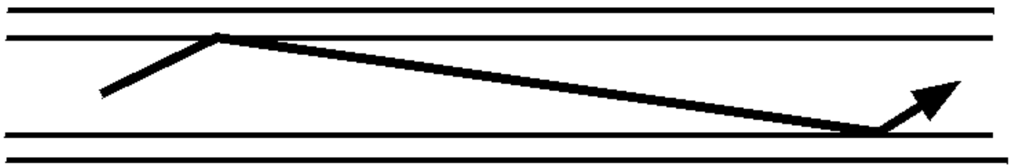


Figure 1-28A.—Reflection angle at various frequencies. LOW FREQUENCY.

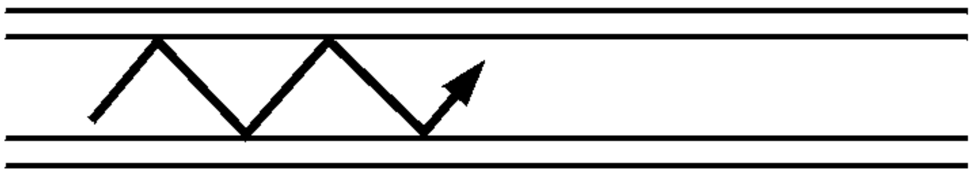


Figure 1-28B.—Reflection angle at various frequencies. MEDIUM FREQUENCY.

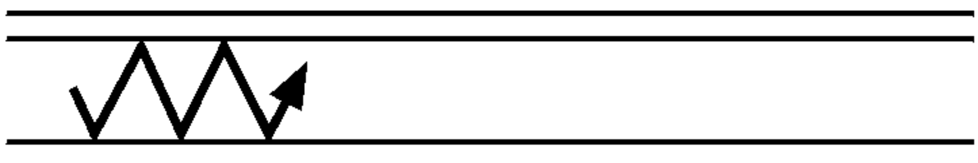


Figure 1-28C.—Reflection angle at various frequencies. HIGH FREQUENCY.

Q-14. What interaction causes energy to travel down a waveguide?

Q-15. What is indicated by the number of arrows (closeness of spacing) used to represent an electric field?

Q-16. What primary condition must magnetic lines of force meet in order to exist?

Q-17. What happens to the H lines between the conductors of a coil when the conductors are close together?

Q-18. For an electric field to exist at the surface of a conductor, the field must have what angular relationship to the conductor?

Q-19. When a wavefront is radiated into a waveguide, what happens to the portions of the wavefront that do not satisfy the boundary conditions?

Q-20. Assuming the wall of a waveguide is perfectly flat, what is the angular relationship between the angle of incidence and the angle of reflection?

Q-21. What is the frequency called that produces angles of incidence and reflection that are perpendicular to the waveguide walls?

Q-22. Compared to the velocity of propagation of waves in air, what is the velocity of propagation of waves in waveguides?

Q-23. What term is used to identify the forward progress velocity of wavefronts in a waveguide?

Waveguide Modes of Operation

The waveguide analyzed in the previous paragraphs yields an electric field configuration known as the half-sine electric distribution. This configuration, called a MODE OF OPERATION, is shown in figure 1-29. Recall that the strength of the field is indicated by the spacing of the lines; that is, the closer the lines, the stronger the field. The regions of maximum voltage in this field move continuously down the waveguide in a sine-wave pattern. To meet boundary conditions, the field must always be zero at the "b" walls.

The half-sine field is only one of many field configurations, or modes, that can exist in a rectangular waveguide. A full-sine field can also exist in a rectangular waveguide because, as shown in figure 1-30, the field is zero at the "b" walls.

Similarly, a 1 1/2 sine-wave field can exist in a rectangular waveguide because this field also meets the boundary conditions. As shown in figure 1-31, the field is perpendicular to any conducting surface it touches and is zero along the "b" walls.

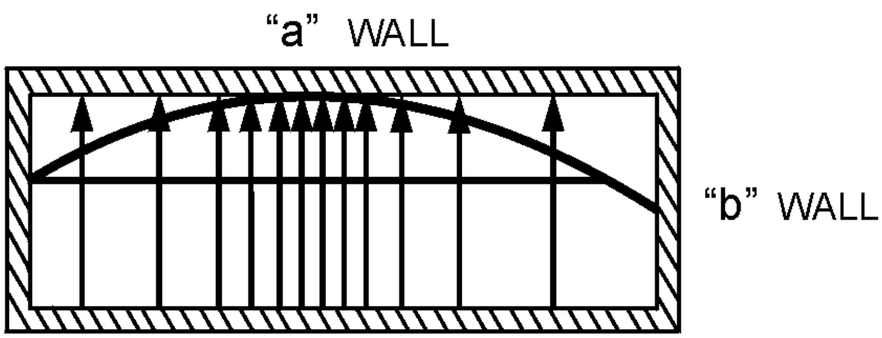


Figure 1-29.—Half-sine E field distribution.

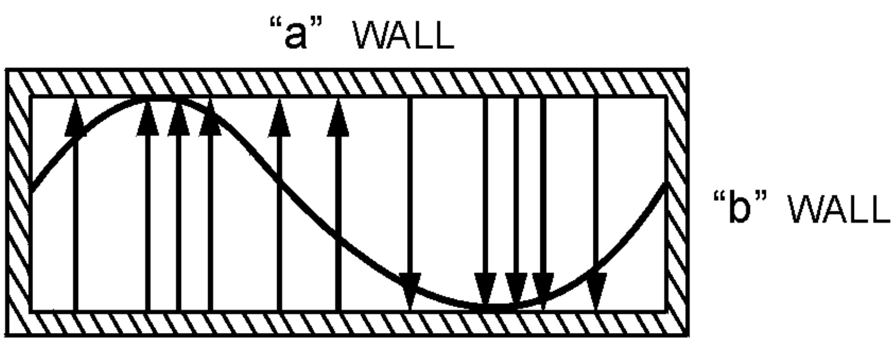


Figure 1-30.—Full-sine E field distribution.

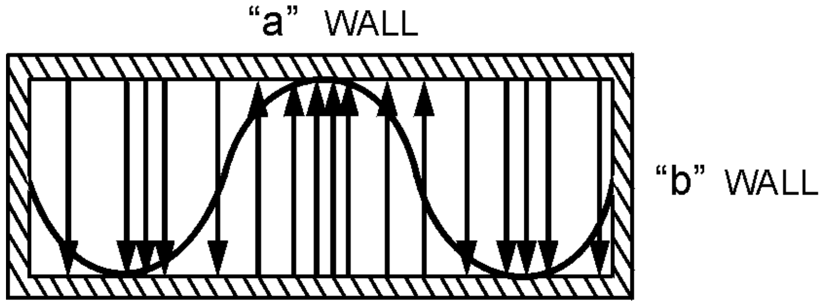


Figure 1-31.—One and one-half sine E field distribution.

The magnetic field in a rectangular waveguide is in the form of closed loops parallel to the surface of the conductors. The strength of the magnetic field is proportional to the electric field. Figure 1-32 illustrates the magnetic field pattern associated with a half-sine electric field distribution. The magnitude of the magnetic field varies in a sine-wave pattern down the center of the waveguide in "time phase" with the electric field. TIME PHASE means that the peak H lines and peak E lines occur at the same instant in time, although not necessarily at the same point along the length of the waveguide.

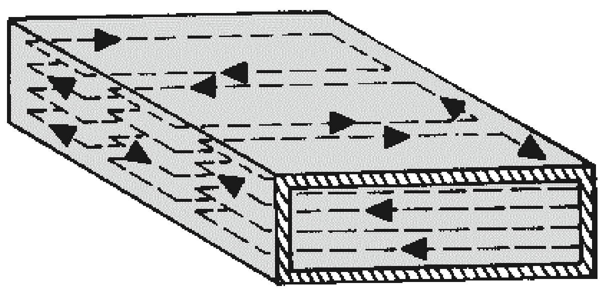


Figure 1-32.—Magnetic field caused by a half-sine E field.

An electric field in a sine-wave pattern also exists down the center of a waveguide. In figure 1-33, view (A), consider the two wavefronts, C and D. Assume that they are positive at point 1 and negative at point 2. When the wavefronts cross at points 1 and 2, each field is at its maximum strength. At these points, the fields combine, further increasing their strength. This action is continuous because each wave is always followed by a replacement wave. Figure 1-33, view (B), illustrates the resultant sine configuration of the electric field at the center of the waveguide. This configuration is only one of the many field patterns that can exist in a waveguide. Each configuration forms a separate mode of operation. The easiest mode to produce is called the DOMINANT MODE. Other modes with different field configurations may occur accidentally or may be caused deliberately.

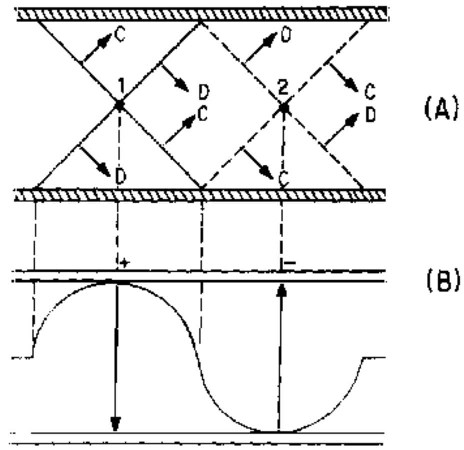


Figure 1-33.—Crisscrossing wavefronts and the resultant E field.

The dominant mode is the most efficient mode. Waveguides are normally designed so that only the dominant mode will be used. To operate in the dominant mode, a waveguide must have an "a" (wide) dimension of at least one half-wavelength of the frequency to be propagated. The "a" dimension of the waveguide must be kept near the minimum allowable value to ensure that only the dominant mode will exist. In practice, this dimension is usually 0.7 wavelength.

Of the possible modes of operation available for a given waveguide, the dominant mode has the lowest cutoff frequency. The high-frequency limit of a rectangular waveguide is a frequency at which its "a" dimension becomes large enough to allow operation in a mode higher than that for which the waveguide has been designed.

Waveguides may be designed to operate in a mode other than the dominant mode. An example of a full-sine configuration mode is shown in figures 1-34A and 1-34B. The "a" dimension of the waveguide in this figure is one wavelength long. You may assume that the two-wire line is $1/4\lambda$ from one of the "b" walls, as shown in figure 1-34A. The remaining distance to the other "b" wall is $3/4\lambda$. The three-quarter wavelength section has the same high impedance as the quarter-wave section; therefore, the two-wire line is properly insulated. The field configuration shows a complete sine-wave pattern across the "a" dimension, as illustrated in figure 1-34B.

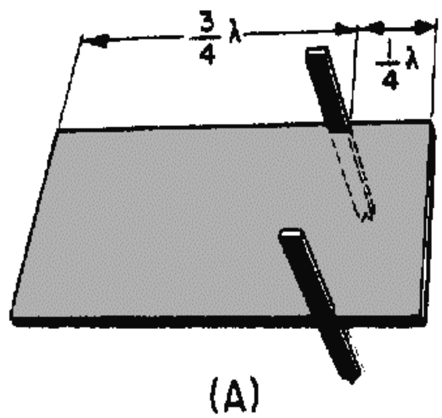


Figure 1-34A.—Waveguide operation in other than dominant mode.

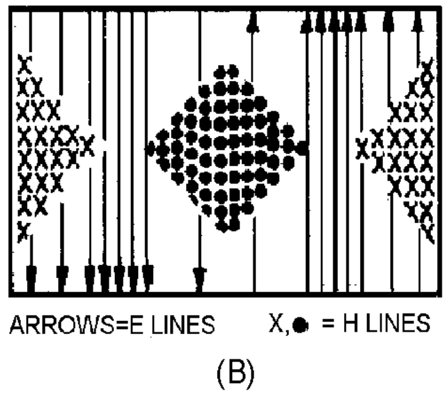


Figure 1-34B.—Waveguide operation in other than dominant mode.

Circular waveguides are used in specific areas of radar and communications systems, such as rotating joints used at the mechanical point where the antennas rotate. Figure 1-35 illustrates the dominant mode of a circular waveguide. The cutoff wavelength of a circular guide is 1.71 times the diameter of the waveguide. Since the "a" dimension of a rectangular waveguide is approximately one half-wavelength at the cutoff frequency, the diameter of an equivalent circular waveguide must be 2 ÷ 1.71, or approximately 1.17 times the "a" dimension of a rectangular waveguide.

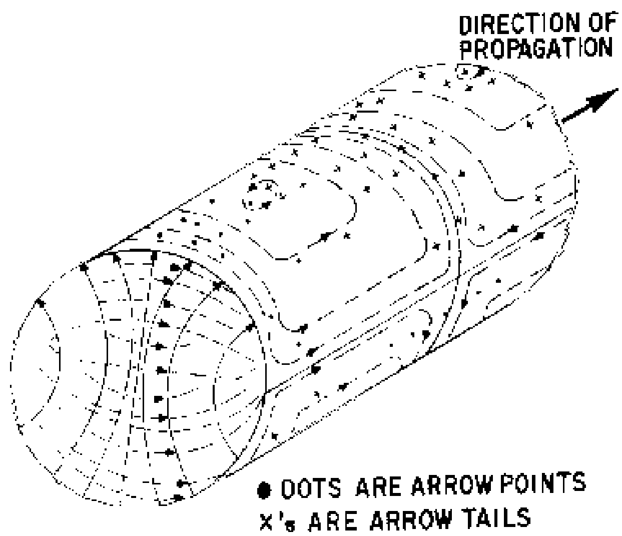


Figure 1-35.—Dominant mode in a circular waveguide.

MODE NUMBERING SYSTEMS.—So far, only the most basic types of E and H field arrangements have been shown. More complicated arrangements are often necessary to make possible coupling, isolation, or other types of operation. The field arrangements of the various modes of operation are divided into two categories: TRANSVERSE ELECTRIC (TE) and TRANSVERSE MAGNETIC (TM).

In the transverse electric (TE) mode, the entire electric field is in the transverse plane, which is perpendicular to the length of the waveguide (direction of energy travel). Part of the magnetic field is parallel to the length axis.

In the transverse magnetic (TM) mode, the entire magnetic field is in the transverse plane and has no portion parallel to the length axis.

Since there are several TE and TM modes, subscripts are used to complete the description of the field pattern. In rectangular waveguides, the first subscript indicates the number of half-wave patterns in the "a" dimension, and the second subscript indicates the number of half-wave patterns in the "b" dimension.

The dominant mode for rectangular waveguides is shown in figure 1-36. It is designated as the TE mode because the E fields are perpendicular to the "a" walls. The first subscript is 1 since there is only one half-wave pattern across the "a" dimension. There are no E-field patterns across the "b" dimension, so the second subscript is 0. The complete mode description of the dominant mode in rectangular waveguides is $TE_{1,0}$. Subsequent descriptions of waveguide operation in this text will assume the dominant ($TE_{1,0}$) mode unless otherwise noted.

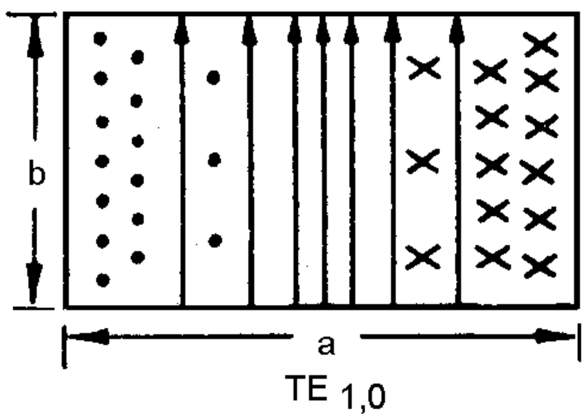


Figure 1-36.—Dominant mode in a rectangular waveguide.

A similar system is used to identify the modes of circular waveguides. The general classification of TE and TM is true for both circular and rectangular waveguides. In circular waveguides the subscripts have a different meaning. The first subscript indicates the number of full-wave patterns around the circumference of the waveguide. The second subscript indicates the number of half-wave patterns across the diameter.

In the circular waveguide in figure 1-37, the E field is perpendicular to the length of the waveguide with no E lines parallel to the direction of propagation. Thus, it must be classified as operating in the TE mode. If you follow the E line pattern in a counterclockwise direction starting at the top, the E lines go from zero, through maximum positive (tail of arrows), back to zero, through maximum negative (head of arrows), and then back to zero again. This is one full wave, so the first subscript is 1. Along the diameter, the E lines go from zero through maximum and back to zero, making a half-wave variation. The second subscript, therefore, is also 1. $TE_{1,1}$ is the complete mode description of the dominant mode in circular waveguides. Several modes are possible in both circular and rectangular waveguides. Figure 1-38 illustrates several different modes that can be used to verify the mode numbering system.

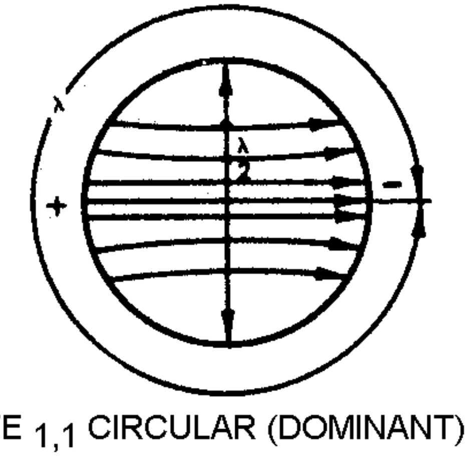


Figure 1-37.—Counting wavelengths in a circular waveguide.

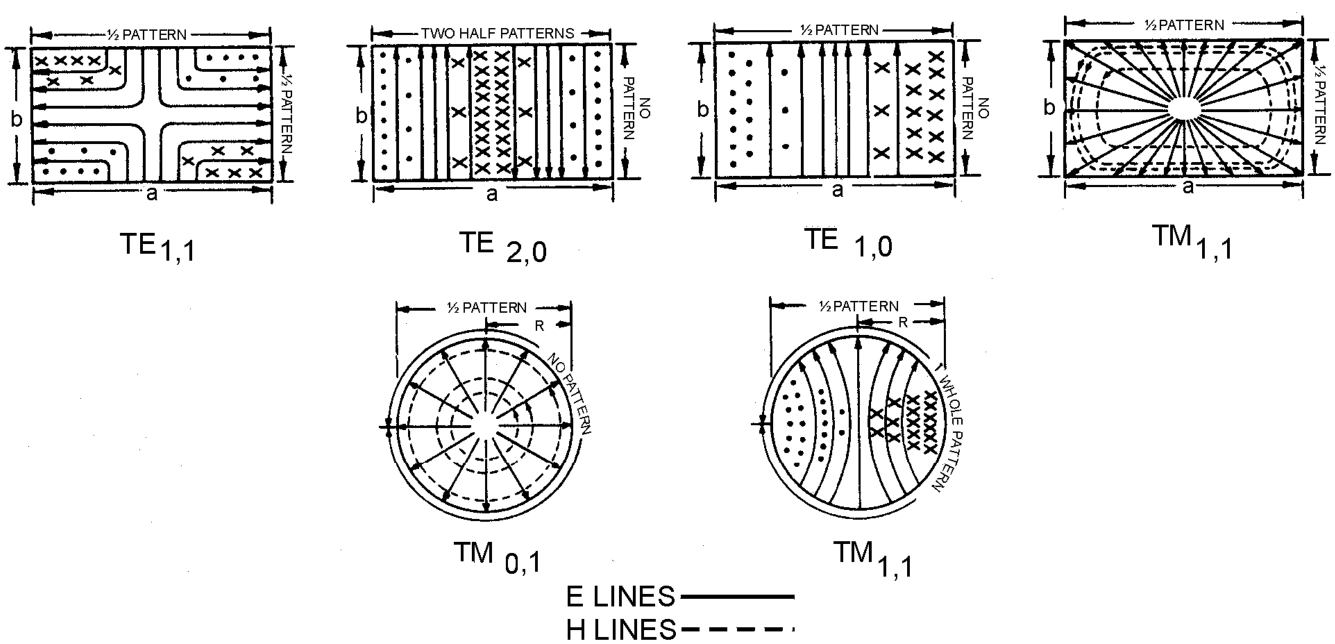


Figure 1-38.—Various modes of operation for rectangular and circular waveguides.

Waveguide Input/Output Methods

A waveguide, as explained earlier in this chapter, operates differently from an ordinary transmission line. Therefore, special devices must be used to put energy into a waveguide at one end and remove it from the other end.

The three devices used to inject or remove energy from waveguides are PROBES, LOOPS, and SLOTS. Slots may also be called APERTURES or WINDOWS.

As previously discussed, when a small probe is inserted into a waveguide and supplied with microwave energy, it acts as a quarter-wave antenna. Current flows in the probe and sets up an E field such as the one shown in figure 1-39A. The E lines detach themselves from the probe. When the probe is located at the point of highest efficiency, the E lines set up an E field of considerable intensity.

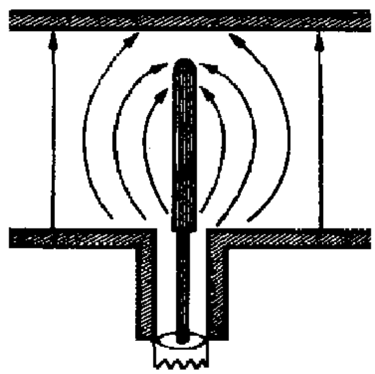


Figure 1-39A.—Probe coupling in a rectangular waveguide.

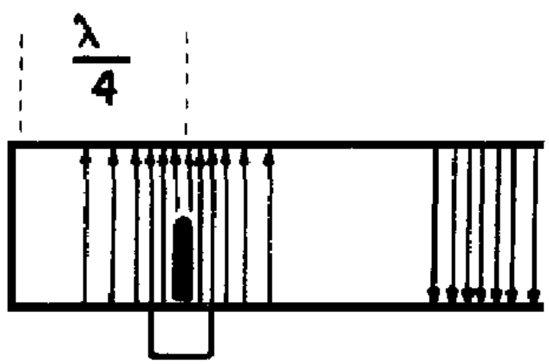


Figure 1-39B.—Probe coupling in a rectangular waveguide.

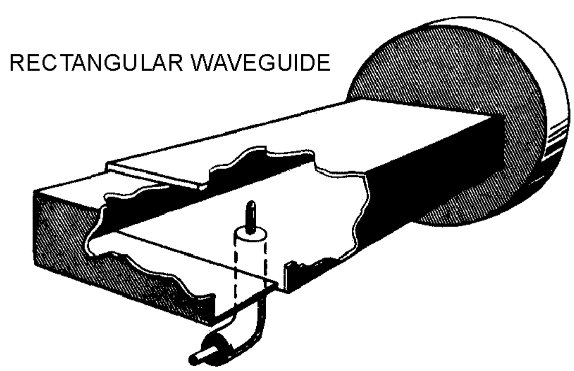


Figure 1-39C.—Probe coupling in a rectangular waveguide.

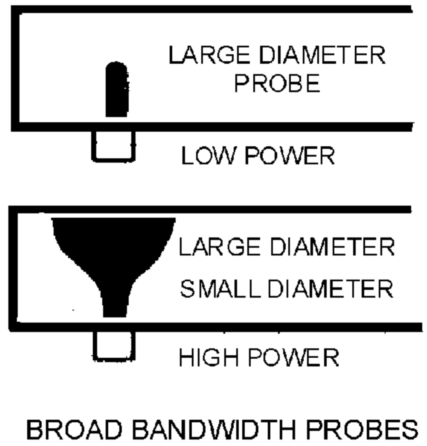


Figure 1-39D.—Probe coupling in a rectangular waveguide.

The most efficient place to locate the probe is in the center of the "a" wall, parallel to the "b" wall, and one quarter-wavelength from the shorted end of the waveguide, as shown in figure 1-39B, and figure 1-39C. This is the point at which the E field is maximum in the dominant mode. Therefore, energy transfer (coupling) is maximum at this point. Note that the quarter-wavelength spacing is at the frequency required to propagate the dominant mode.

In many applications a lesser degree of energy transfer, called loose coupling, is desirable. The amount of energy transfer can be reduced by decreasing the length of the probe, by moving it out of the center of the E field, or by shielding it. Where the degree of coupling must be varied frequently, the probe is made retractable so the length can be easily changed.

The size and shape of the probe determines its frequency, bandwidth, and power-handling capability. As the diameter of a probe increases, the bandwidth increases. A probe similar in shape to a door knob is capable of handling much higher power and a larger bandwidth than a conventional probe. The greater power-handling capability is directly related to the increased surface area. Two examples of broad-bandwidth probes are illustrated in figure 1-39D. Removal of energy from a waveguide is simply a reversal of the injection process using the same type of probe.

Another way of injecting energy into a waveguide is by setting up an H field in the waveguide. This can be accomplished by inserting a small loop which carries a high current into the waveguide, as shown in figure 1-40A. A magnetic field builds up around the loop and expands to fit the waveguide, as shown in figure 1-40B. If the frequency of the current in the loop is within the bandwidth of the waveguide, energy will be transferred to the waveguide.

For the most efficient coupling to the waveguide, the loop is inserted at one of several points where the magnetic field will be of greatest strength. Four of those points are shown in figure 1-40C.

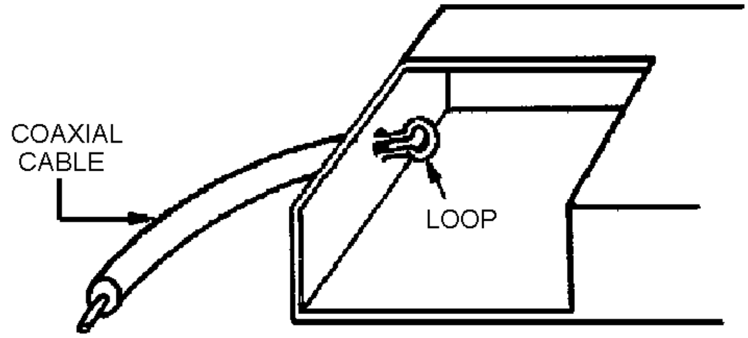


Figure 1-40A.—Loop coupling in a rectangular waveguide.

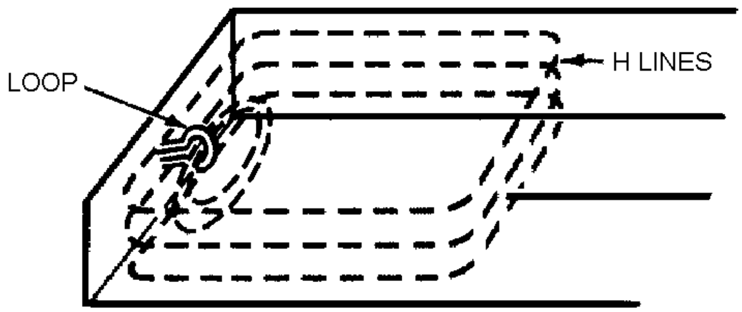


Figure 1-40B.—Loop coupling in a rectangular waveguide.

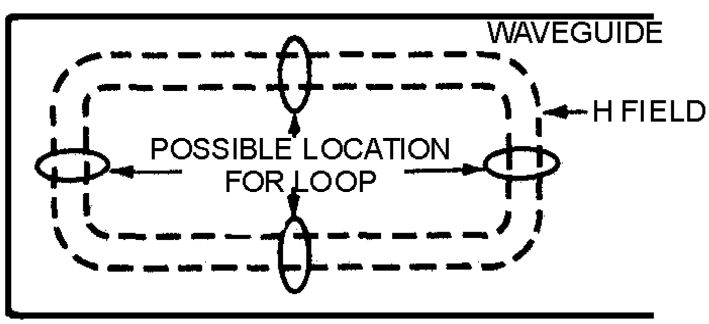


Figure 1-40C.—Loop coupling in a rectangular waveguide.

When less efficient coupling is desired, you can rotate or move the loop until it encircles a smaller number of H lines. When the diameter of the loop is increased, its power-handling capability also increases. The bandwidth can be increased by increasing the size of the wire used to make the loop.

When a loop is introduced into a waveguide in which an H field is present, a current is induced in the loop. When this condition exists, energy is removed from the waveguide.

Slots or apertures are sometimes used when very loose (inefficient) coupling is desired, as shown in figure 1-41. In this method energy enters through a small slot in the waveguide and the E field expands into the waveguide. The E lines expand first across the slot and then across the interior of the waveguide.

Minimum reflections occur when energy is injected or removed if the size of the slot is properly proportioned to the frequency of the energy.

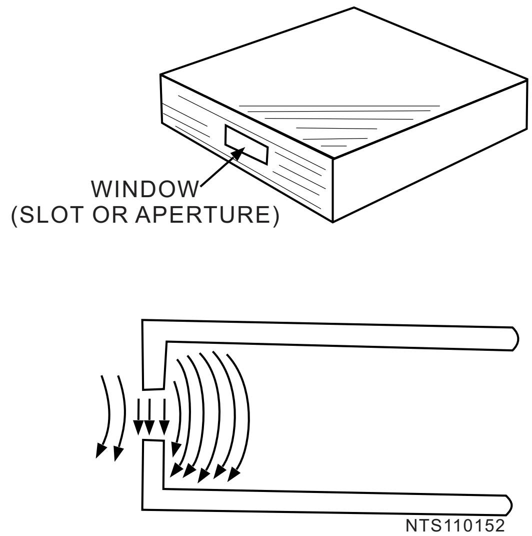


Figure 1-41.—Slot coupling in a waveguide.

After learning how energy is coupled into and out of a waveguide with slots, you might think that leaving the end open is the most simple way of injecting or removing energy in a waveguide. This is not the case, however, because when energy leaves a waveguide, fields form around the end of the waveguide. These fields cause an impedance mismatch which, in turn, causes the development of standing waves and a drastic loss in efficiency. Various methods of impedance matching and terminating waveguides will be covered in the next section.

Q-24. What term is used to identify each of the many field configurations that can exist in waveguides?

Q-25. What field configuration is easiest to produce in a given waveguide?

Q-26. How is the cutoff wavelength of a circular waveguide figured?

Q-27. The field arrangements in waveguides are divided into what two categories to describe the various modes of operation?

Q-28. The electric field is perpendicular to the "a" dimension of a waveguide in what mode?

Q-29. The number of half-wave patterns in the "b" dimension of rectangular waveguides is indicated by which of the two descriptive subscripts?

Q-30. Which subscript, in circular waveguide classification, indicates the number of full-wave patterns around the circumference?

Q-31. What determines the frequency, bandwidth, and power-handling capability of a waveguide probe?

Q-32. Loose or inefficient coupling of energy into or out of a waveguide can be accomplished by the use of what method?

Waveguide Impedance Matching

Waveguide transmission systems are not always perfectly impedance matched to their load devices. The standing waves that result from a mismatch cause a power loss, a reduction in power-handling capability, and an increase in frequency sensitivity. Impedance-changing devices are therefore placed in the waveguide to match the waveguide to the load. These devices are placed near the source of the standing waves.

Figure 1-42 illustrates three devices, called irises, that are used to introduce inductance or capacitance into a waveguide. An iris is nothing more than a metal plate that contains an opening through which the waves may pass. The iris is located in the transverse plane.

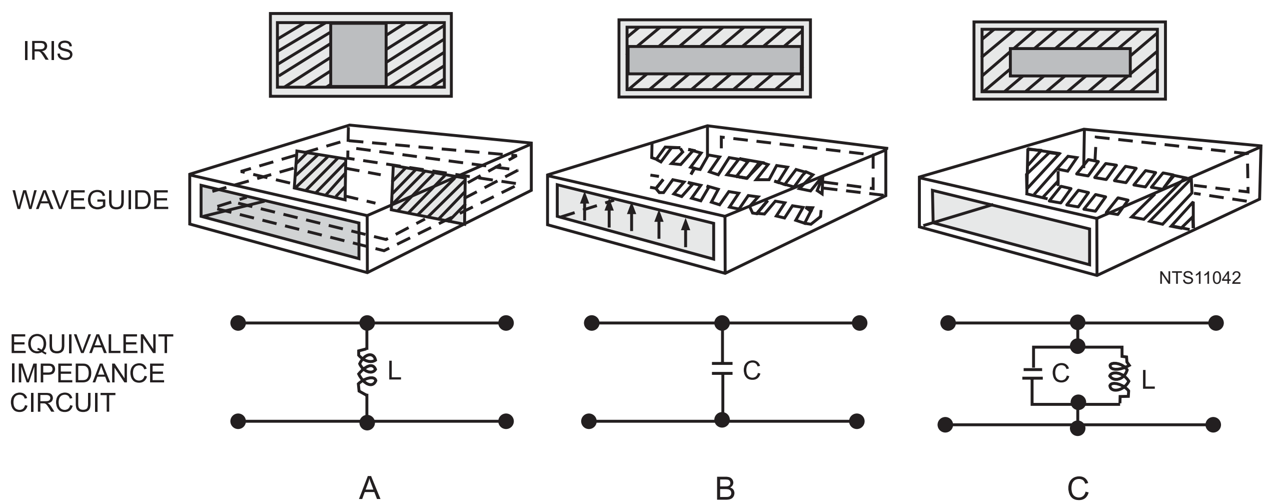


Figure 1-42.—Waveguide irises.

An inductive iris and its equivalent circuit are illustrated in figure 1-42, view (A). The iris places a shunt inductive reactance across the waveguide that is directly proportional to the size of the opening. Notice that the edges of the inductive iris are perpendicular to the magnetic plane. The shunt capacitive reactance, illustrated in view (B), basically acts the same way. Again, the reactance is directly proportional to the size of the opening, but the edges of the iris are perpendicular to the electric plane. The iris, illustrated in view (C), has portions across both the magnetic and electric planes and forms an equivalent parallel-LC circuit across the waveguide. At the resonant frequency, the iris acts as a high shunt resistance. Above or below resonance, the iris acts as a capacitive or inductive reactance.

POSTS and SCREWS made from conductive material can be used for impedance-changing devices in waveguides. Figure 1-43A and 1-43B, illustrate two basic methods of using posts and screws. A post or screw which only partially penetrates into the waveguide acts as a shunt capacitive reactance. When the post or screw extends completely through the waveguide, making contact with the top and bottom walls, it acts as an inductive reactance. Note that when screws are used the amount of reactance can be varied.

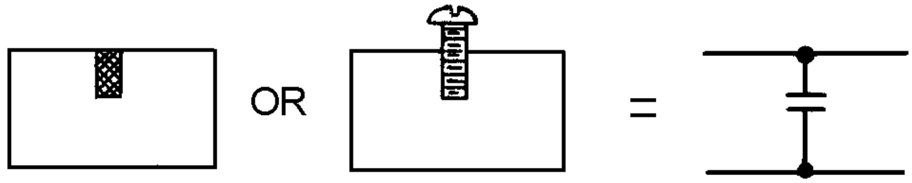


Figure 1-43A.—Conducting posts and screws. PENETRATING.

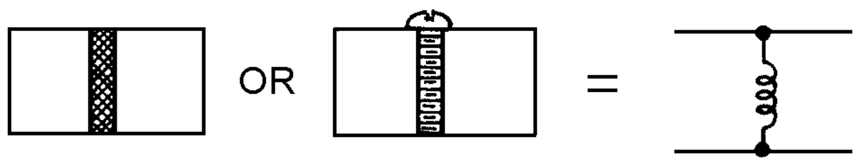


Figure 1-43B.—Conducting posts and screws. EXTENDING THROUGH.

Q-33. What is the result of an impedance mismatch in a waveguide?

Q-34. What is used to construct irises?

Q-35. An iris placed along the "b" dimension wall produces what kind of reactance?

Q-36. How will an iris that has portions along both the "a" and "b" dimension walls act at the resonant frequency?

Waveguide Terminations

Electromagnetic energy is often passed through a waveguide to transfer the energy from a source into space. As previously mentioned, the impedance of a waveguide does not match the impedance of space, and without proper impedance matching, standing waves cause a large decrease in the efficiency of the waveguide.

Any abrupt change in impedance causes standing waves, but when the change in impedance at the end of a waveguide is gradual, almost no standing waves are formed. Gradual changes in impedance can be obtained by terminating the waveguide with a funnel-shaped HORN, such as the three types illustrated in figures 1-44A, 1-44B, and 1-44C. The type of horn used depends upon the frequency and the desired radiation pattern.

Figure 1-44A.—Waveguide horns. E PLANE SECTORAL HORN.

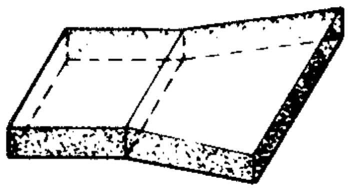


Figure 1-44B.—Waveguide horns. H PLANE SECTORAL HORN.

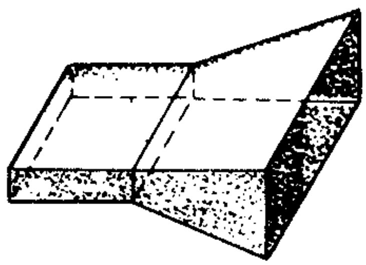


Figure 1-44C.—Waveguide horns. PYRAMID HORN.

As you may have noticed, horns are really simple antennas. They have several advantages over other impedance-matching devices, such as their large bandwidth and simple construction. The use of horns as antennas will be discussed further in chapter 3.

A waveguide may also be terminated in a resistive load that is matched to the characteristic impedance of the waveguide. The resistive load is most often called a DUMMY LOAD, because its only purpose is to absorb all the energy in a waveguide without causing standing waves.

There is no place on a waveguide to connect a fixed termination resistor; therefore, several special arrangements are used to terminate waveguides. One method is to fill the end of the waveguide with a graphite and sand mixture, as illustrated in figure 1-45A. When the fields enter the mixture, they induce a current flow in the mixture which dissipates the energy as heat. Another method figure 1-45B is to use a high-resistance rod placed at the center of the E field. The E field causes current to flow in the rod, and the high resistance of the rod dissipates the energy as a power loss, again in the form of heat.

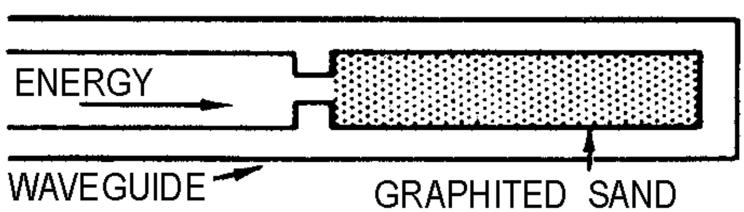


Figure 1-45A.—Terminating waveguides.

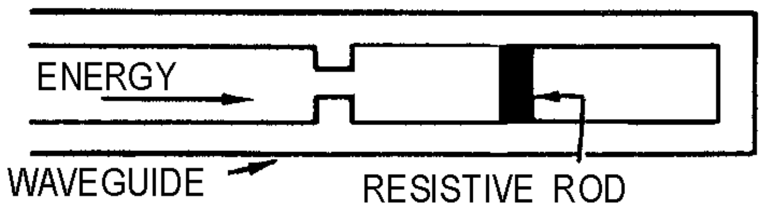


Figure 1-45B.—Terminating waveguides.

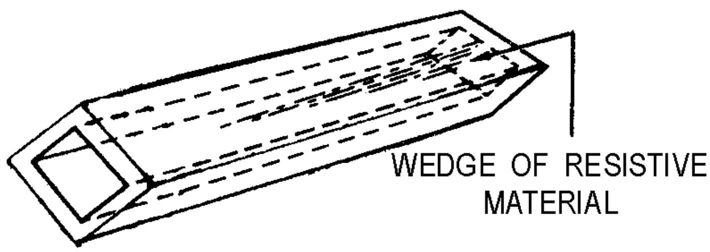


Figure 1-45C.—Terminating waveguides.

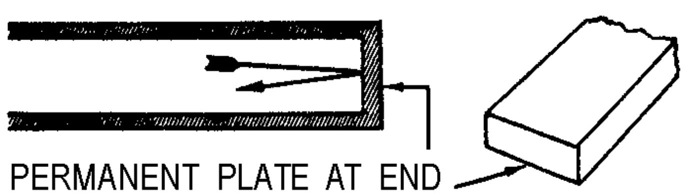


Figure 1-45D.—Terminating waveguides.

Still another method for terminating a waveguide is the use of a wedge of highly resistive material, as shown in of figure 1-45C. The plane of the wedge is placed perpendicular to the magnetic lines of force. When the H lines cut through the wedge, current flows in the wedge and causes a power loss. As with the other methods, this loss is in the form of heat. Since very little energy reaches the end of the waveguide, reflections are minimum.

All of the terminations discussed so far are designed to radiate or absorb the energy without reflections. In many instances, however, all of the energy must be reflected from the end of the waveguide. The best way to accomplish this is to permanently weld a metal plate at the end of the waveguide, as shown in figure 1-45D.

Q-37. What device is used to produce a gradual change in impedance at the end of a waveguide?

Q-38. When a waveguide is terminated in a resistive load, the load must be matched to what property of the waveguide?

Q-39. What is the primary purpose of a dummy load?

Q-40. The energy dissipated by a resistive load is most often in what form?

Waveguide Plumbing

Since waveguides are really only hollow metal pipes, the installation and the physical handling of waveguides have many similarities to ordinary plumbing. In light of this fact, the bending, twisting, joining, and installation of waveguides is commonly called waveguide plumbing. Naturally, waveguides are different in design from pipes that are designed to carry liquids or other substances. The design of a waveguide is determined by the frequency and power level of the electromagnetic energy it will carry. The following paragraphs explain the physical factors involved in the design of waveguides.

WAVEGUIDE BENDS.—The size, shape, and dielectric material of a waveguide must be constant throughout its length for energy to move from one end to the other without reflections. Any abrupt change in its size or shape can cause reflections and a loss in overall efficiency. When such a change is necessary, the bends, twists, and joints of the waveguides must meet certain conditions to prevent reflections.

Waveguides may be bent in several ways that do not cause reflections. One way is the gradual bend shown in figure 1-46. This gradual bend is known as an E bend because it distorts the E fields. The E bend must have a radius greater than two wavelengths to prevent reflections.

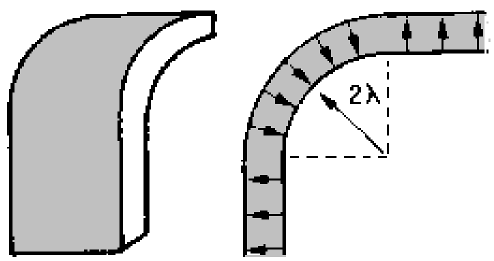


Figure 1-46.—Gradual E bend.

Another common bend is the gradual H bend (figure 1-47). It is called an H bend because the H fields are distorted when a waveguide is bent in this manner. Again, the radius of the bend must be greater than two wavelengths to prevent reflections. Neither the E bend in the "a" dimension nor the H bend in the "b" dimension changes the normal mode of operation.

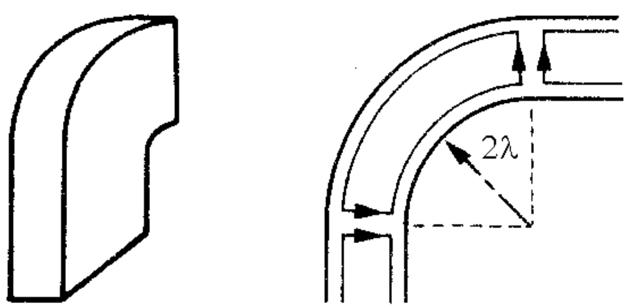


Figure 1-47.—Gradual H bend.

A sharp bend in either dimension may be used if it meets certain requirements. Notice the two 45-degree bends in figure 1-48; the bends are 1/4λ apart. The reflections that occur at the 45-degree bends cancel each other, leaving the fields as though no reflections have occurred.

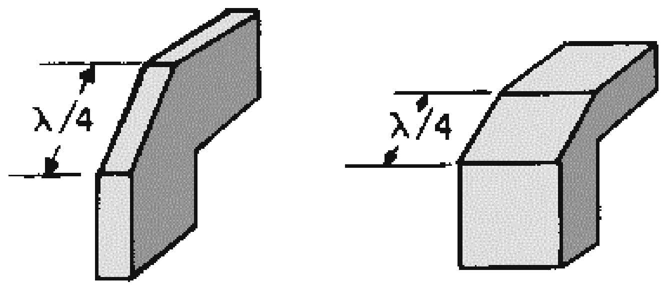


Figure 1-48.—Sharp bends.

Sometimes the electromagnetic fields must be rotated so that they are in the proper phase to match the phase of the load. This may be accomplished by twisting the waveguide as shown in figure 1-49. The twist must be gradual and greater than 2λ.

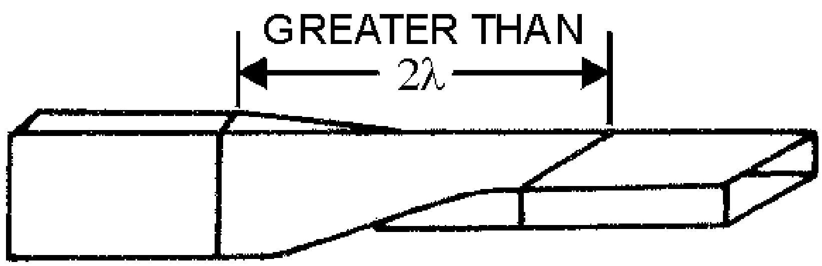


Figure 1-49.—Waveguide twist.

The flexible waveguide (figure 1-50) allows special bends which some equipment applications might require. It consists of a specially wound ribbon of conductive material, most commonly brass, with the inner surface plated with chromium. Power losses are greater in the flexible waveguide because the inner surfaces are not perfectly smooth. Therefore, it is only used in short sections where no other reasonable solution is available.

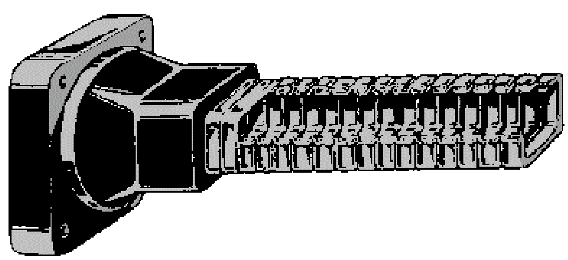


Figure 1-50.—Flexible waveguide.

WAVEGUIDE JOINTS.—Since an entire waveguide system cannot possibly be molded into one piece, the waveguide must be constructed in sections and the sections connected with joints. The three basic types of waveguide joints are the PERMANENT, the SEMIPERMANENT, and the ROTATING JOINTS. Since the permanent joint is a factory-welded joint that requires no maintenance, only the semipermanent and rotating joints will be discussed.

Sections of waveguide must be taken apart for maintenance and repair. A semipermanent joint, called a CHOKE JOINT, is most commonly used for this purpose. The choke joint provides good electromagnetic continuity between sections of waveguide with very little power loss.

A cross-sectional view of a choke joint is shown in figures 1-51A and 1-51B. The pressure gasket shown between the two metal surfaces forms an airtight seal. Notice in figure 1-51B that the slot is exactly 1/4λ from the "a" wall of the waveguide. The slot is also 1/4λ deep, as shown in figure 1-51A, and because it is shorted at point (1), a high impedance results at point (2). Point (3) is 1/4λ from point (2). The high impedance at point (2) results in a low impedance, or short, at point (3). This effect creates a good electrical connection between the two sections that permits energy to pass with very little reflection or loss.

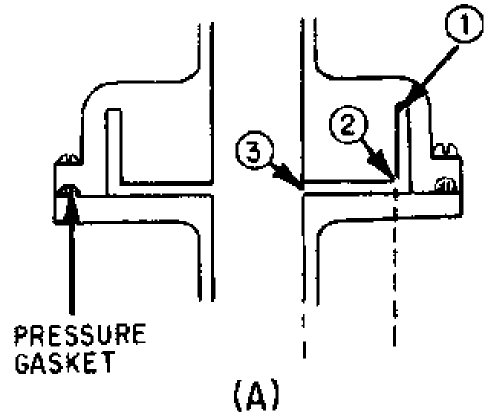


Figure 1-51A.—Choke joint.

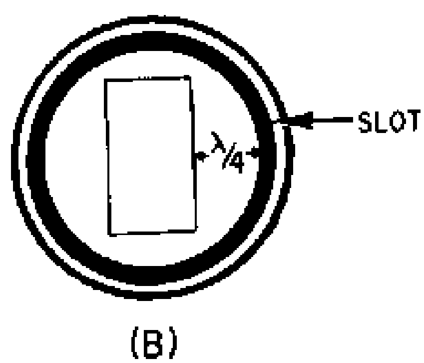


Figure 1-51B.—Choke joint.

Whenever a stationary rectangular waveguide is to be connected to a rotating antenna, a rotating joint must be used. A circular waveguide is normally used in a rotating joint. Rotating a rectangular waveguide would cause field pattern distortion. The rotating section of the joint, illustrated in figure 1-52, uses a choke joint to complete the electrical connection with the stationary section. The circular waveguide is designed so that it will operate in the $TM_{0,1}$ mode. The rectangular sections are attached as shown in the illustration to prevent the circular waveguide from operating in the wrong mode.

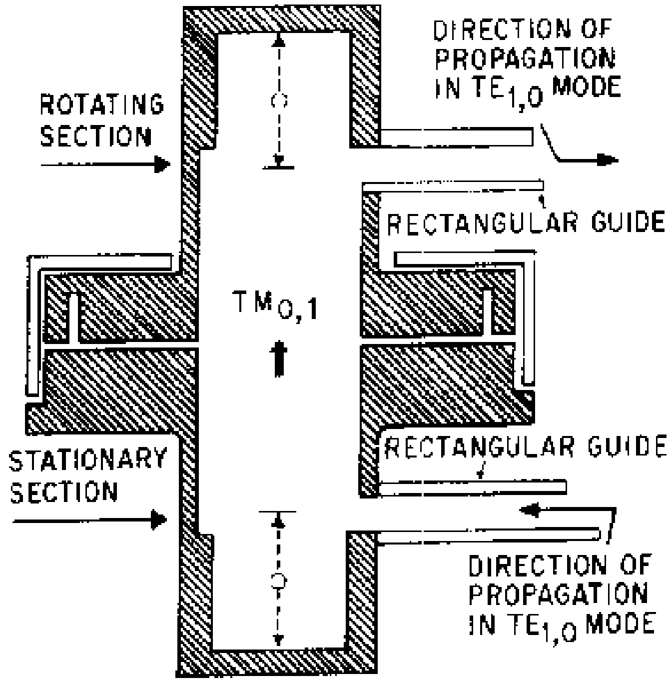


Figure 1-52.—Rotating joint.

Distance "O" is 3/4λ so that a high impedance will be presented to any unwanted modes. This is the most common design used for rotating joints, but other types may be used in specific applications.

WAVEGUIDE MAINTENANCE.—The installation of a waveguide system presents problems that are not normally encountered when dealing with other types of transmission lines. These problems often fall within the technician's area of responsibility. A brief discussion of waveguide handling, installation, and maintenance will help prepare you for this maintenance responsibility. Detailed information concerning waveguide maintenance in a particular system may be found in the technical manuals for the system.

Since a waveguide naturally has a low loss ratio, most losses in a waveguide system are caused by other factors. Improperly connected joints or damaged inner surfaces can decrease the efficiency of a system to the point that it will not work at all. Therefore, you must take great care when working with waveguides to prevent physical damage. Since waveguides are made from a soft, conductive material, such as copper or aluminum, they are very easy to dent or deform. Even the slightest damage to the inner surface of a waveguide will cause standing waves and, often, internal arcing. Internal arcing causes further damage to the waveguide in an action that is often self-sustaining until the waveguide is damaged beyond use. Part of your job as a technician will be to inspect the waveguide system for physical damage. The previously mentioned dents are only one type of physical damage that can decrease the efficiency of the system. Another problem occurs because waveguides are made from a conductive material such as copper while the structures of most ships are made from steel. When two dissimilar metals, such as copper and steel, are in direct contact, an electrical action called ELECTROLYSIS takes place that causes very rapid corrosion of the metals. Waveguides can be completely destroyed by electrolytic corrosion in a relatively short period of time if they are not isolated from direct contact with other metals. Any inspection of a waveguide system should include a detailed inspection of all support points to ensure that electrolytic corrosion is not taking place. Any waveguide that is exposed to the weather should be painted and all joints sealed. Proper painting prevents natural corrosion, and sealing the joints prevents moisture from entering the waveguide.

Moisture can be one of the worst enemies of a waveguide system. As previously discussed, the dielectric in waveguides is air, which is an excellent dielectric as long as it is free of moisture. Wet air, however, is a very poor dielectric and can cause serious internal arcing in a waveguide system. For this reason care is taken to ensure that waveguide systems are pressurized with air that is dry. Checking the pressure and moisture content of the waveguide air may be one of your daily system maintenance duties.

More detailed waveguide installation and maintenance information can be found in the technical manuals that apply to your particular system. Another good source is the *Electronics Installation and Maintenance Handbooks* (EIMB) published by Naval Sea Systems Command. *Installation Standards Handbook* EIMB, NAVSEA 0967-LP-000-0110, is the volume that deals with waveguide installation and maintenance.

Q-41. What is the result of an abrupt change in the size, shape, or dielectric of a waveguide?

Q-42. A waveguide bend must have what minimum radius?

Q-43. What is the most common type of waveguide joint?

Q-44. What is the most likely cause of losses in waveguide systems?

WAVEGUIDE DEVICES

The discussion of waveguides, up to this point, has been concerned only with the transfer of energy from one point to another. Many waveguide devices have been developed, however, that modify the energy in some fashion during transit. Some devices do nothing more than change the direction of the energy. Others have been designed to change the basic characteristics or power level of the electromagnetic energy.

This section will explain the basic operating principles of some of the more common waveguide devices, such as DIRECTIONAL COUPLERS, CAVITY RESONATORS, and HYBRID JUNCTIONS.

Directional Couplers

The directional coupler is a device that provides a method of sampling energy from within a waveguide for measurement or use in another circuit. Most couplers sample energy traveling in one direction only. However, directional couplers can be constructed that sample energy in both directions. These are called BIDIRECTIONAL couplers and are widely used in radar and communications systems.

Directional couplers may be constructed in many ways. The coupler illustrated in figure 1-53 is constructed from an enclosed waveguide section of the same dimensions as the waveguide in which the energy is to be sampled. The "b" wall of this enclosed section is mounted to the "b" wall of the waveguide from which the sample will be taken. There are two holes in the "b" wall between the sections of the coupler. These two holes are $1/4\lambda$ apart. The upper section of the directional coupler has a wedge of energy-absorbing material at one end and a pickup probe connected to an output jack at the other end. The absorbent material absorbs the energy not directed at the probe and a portion of the overall energy that enters the section.

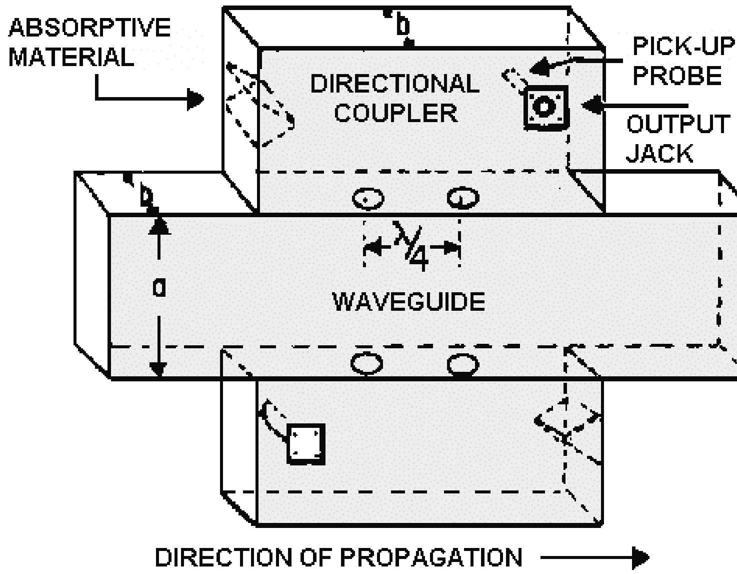


Figure 1-53.—Directional coupler.

Figure 1-54 illustrates two portions of the incident wavefront in a waveguide. The waves travel down the waveguide in the direction indicated and enter the coupler section through both holes. Since both portions of the wave travel the same distance, they are in phase when they arrive at the pickup probe. Because the waves are in phase, they add together and provide a sample of the energy traveling down the waveguide. The sample taken is only a small portion of the energy that is traveling down the waveguide. The magnitude of the sample, however, is proportional to the magnitude of the energy in the waveguide. The absorbent material is designed to ensure that the ratio between the sample energy and the energy in the waveguide is constant. Otherwise the sample would contain no useful information.

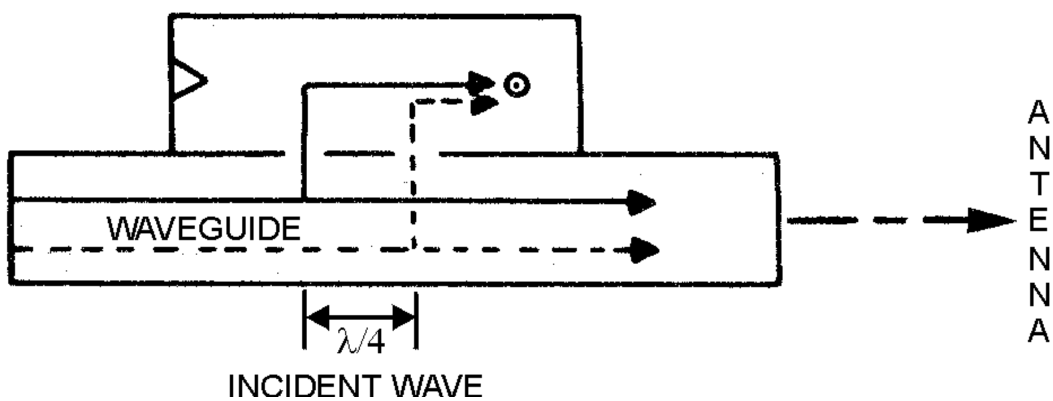


Figure 1-54.—Incident wave in a directional coupler designed to sample incident waves.

The ratio is usually stamped on the coupler in the form of an attenuation factor.

The effect of a directional coupler on any reflected energy is illustrated in figure 1-55. Note that these two waves do not travel the same distance to the pickup probe. The wave represented by the dotted line travels 1/2λ further and arrives at the probe 180 degrees out of phase with the wave represented by the solid line. Because the waves are 180 degrees out of phase at the probe, they cancel each other and no energy is induced in the pickup probe. When the reflected energy arrives at the absorbent material, it adds and is absorbed by the material.

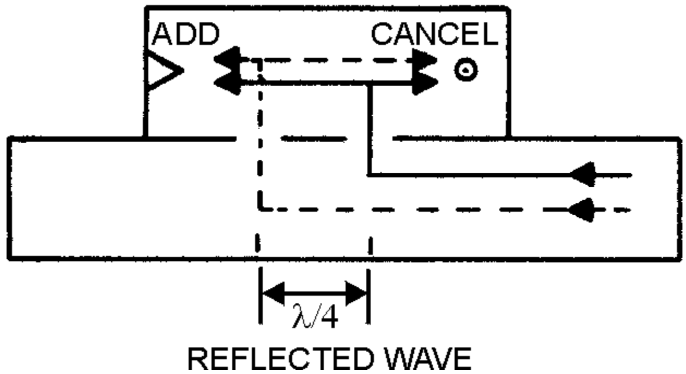


Figure 1-55.—Reflected wave in a directional coupler.

A directional coupler designed to sample reflected energy is shown in figure 1-56. The absorbent material and the probe are in opposite positions from the directional coupler designed to sample the incident energy. This positioning causes the two portions of the reflected energy to arrive at the probe in phase, providing a sample of the reflected energy. The sampled transmitted energy, however, is absorbed by the absorbent material.

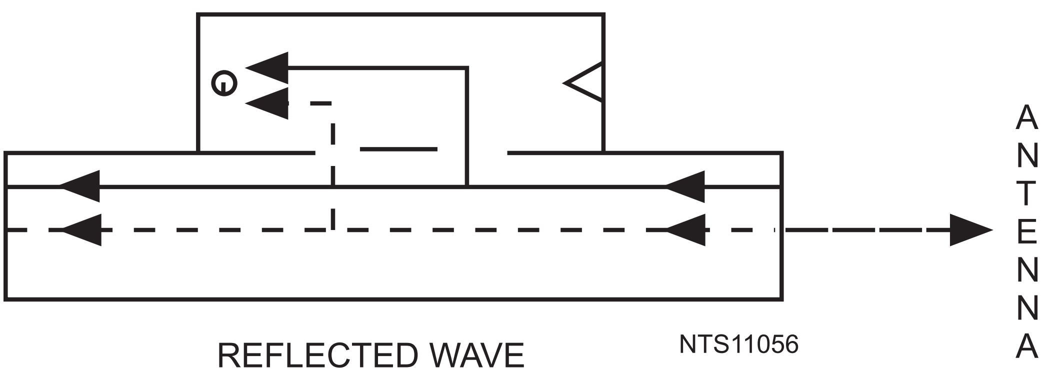


Figure 1-56.—Directional coupler designed to sample reflected energy.

A simple bidirectional coupler for sampling both transmitted and reflected energy can be constructed by mounting two directional couplers on opposite sides of a waveguide, as shown in figure 1-57.

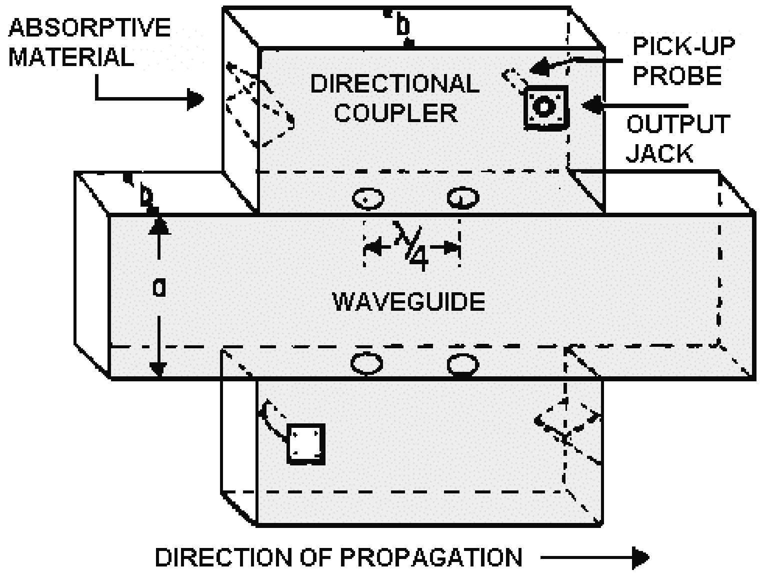


Figure 1-57.—Bidirectional coupler.

Q-45. What is the primary purpose of a directional coupler?

Q-46. How far apart are the two holes in a simple directional coupler?

Q-47. What is the purpose of the absorbent material in a directional coupler?

Q-48. In a directional coupler that is designed to sample the incident energy, what happens to the two portions of the wavefront when they arrive at the pickup probe?

Q-49. What happens to reflected energy that enters a directional coupler that is designed to sample incident energy?

Cavity Resonators

In ordinary electronic equipment a resonant circuit consists of a coil and a capacitor that are connected either in series or in parallel. The resonant frequency of the circuit is increased by reducing the capacitance, the inductance, or both. A point is eventually reached where the inductance and the capacitance can be reduced no further. This is the highest frequency at which a conventional circuit can oscillate.

The upper limit for a conventional resonant circuit is between 2000 and 3000 megahertz. At these frequencies, the inductance may consist of a coil of one-half turn, and the capacitance may simply be the stray capacitance of the coil. Tuning a one-half turn coil is very difficult and tuning stray capacitance is even more difficult. In addition, such a circuit will handle only very small amounts of current.

NEETS, Module 10, *Introduction to Wave Propagation* explained that a 1/4λ section of transmission line can act as a resonant circuit. The same is true of a 1/4λ section of waveguide. Since a waveguide is hollow, it can also be considered as a RESONANT CAVITY.

By definition, a resonant cavity is any space completely enclosed by conducting walls that can contain oscillating electromagnetic fields and possess resonant properties. The cavity has many advantages and uses at microwave frequencies. Resonant cavities have a very high Q and can be built to handle relatively large amounts of power. Cavities with a Q value in excess of 30,000 are not uncommon. The high Q gives these devices a narrow bandpass and allows very accurate tuning. Simple, rugged construction is an additional advantage.

Although cavity resonators, built for different frequency ranges and applications, have a variety of shapes, the basic principles of operation are the same for all.

One example of a cavity resonator is the rectangular box shown in figure 1-58A. It may be thought of as a section of rectangular waveguide closed at both ends by conducting plates. The frequency at which the resonant mode occurs is 1/2λ of the distance between the end plates. The magnetic and electric field patterns in the rectangular cavity are shown in figure 1-58B.

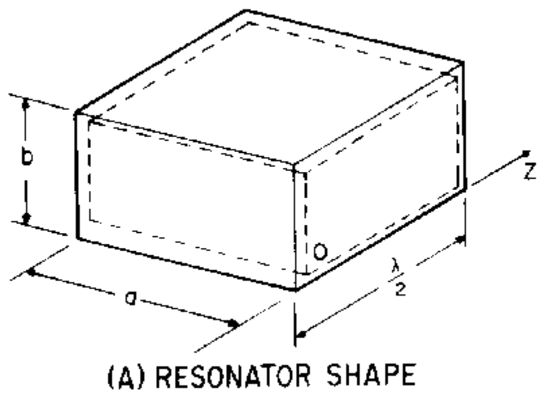


Figure 1-58A.—Rectangular waveguide cavity resonator. RESONATOR SHAPE.

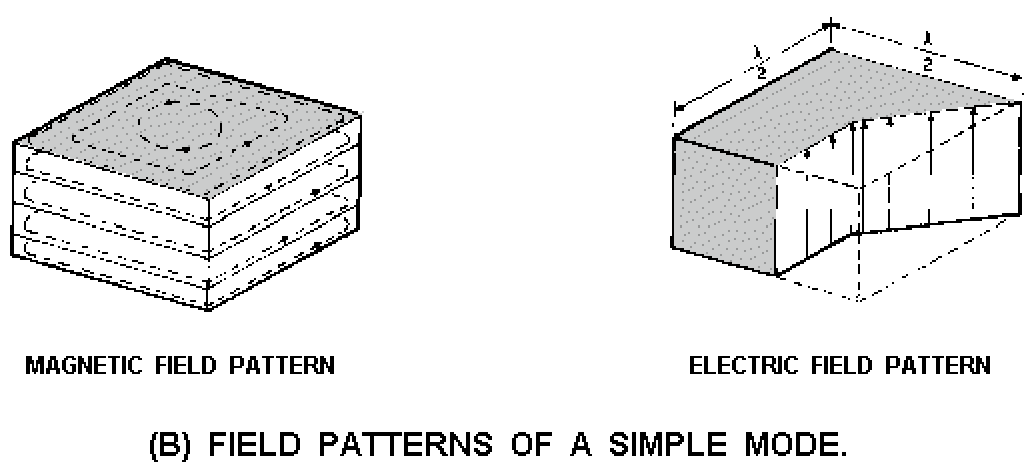


Figure 1-58B.—Rectangular waveguide cavity resonator. FIELD PATTERNS OF A SIMPLE MODE.

The rectangular cavity is only one of many cavity devices that are useful as high-frequency resonators. Figures 1-59A, 1-59B, 1-59C, and 1-59D show the development of a cylindrical resonant cavity from an infinite number of quarter-wave sections of transmission line. In figure 1-59A the 1/4λ section is shown to be equivalent to a resonant circuit with a very small amount of inductance and capacitance. Three 1/4λ sections are joined in parallel in figure 1-59B. Note that although the current-carrying ability of several 1/4λ sections is greater than that of any one section, the resonant frequency is unchanged. This occurs because the addition of inductance in parallel lowers the total inductance, but the addition of capacitance in parallel increases the total capacitance by the same proportion. Thus, the resonant frequency remains the same as it was for one section. The increase in the number of current paths also decreases the total resistance and increases the Q of the resonant circuit. Figure 1-59C shows an intermediate step in the development of the cavity. Figure 1-59D shows a completed cylindrical resonant cavity with a diameter of 1/2λ at the resonant frequency.

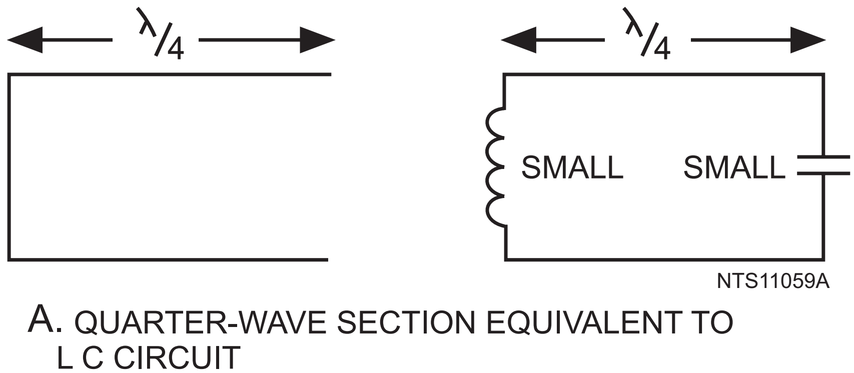


Figure 1-59A.—Development of a cylindrical resonant cavity. QUARTER-WAVE SECTION EQUIVALENT TO LC CIRCUIT.

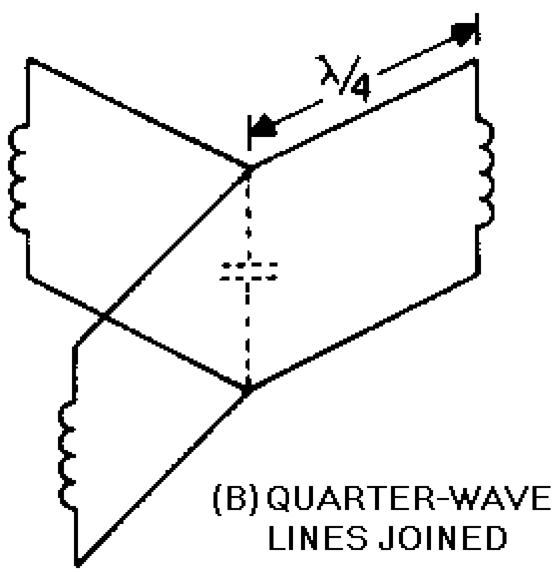


Figure 1-59B.—Development of a cylindrical resonant cavity. QUARTER-WAVE LINES JOINED.

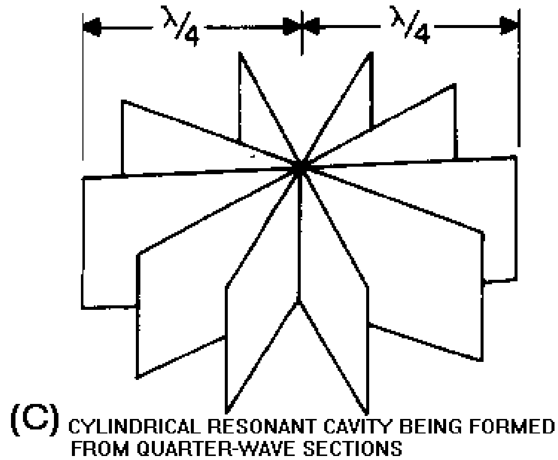


Figure 1-59C.—Development of a cylindrical resonant cavity. CYLINDRICAL RESONANT CAVITY BEING FORMED FROM QUARTER-WAVE SECTIONS.

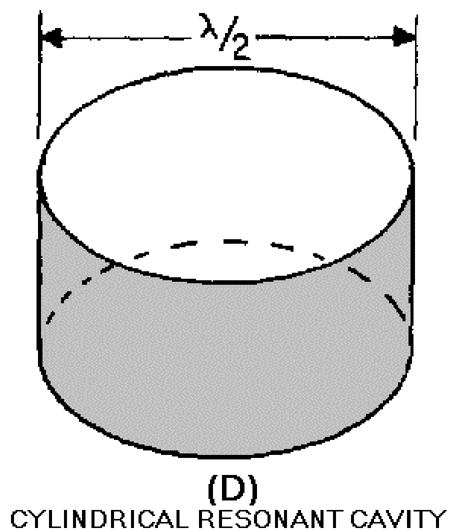


Figure 1-59D.—Development of a cylindrical resonant cavity. CYLINDRICAL RESONANT CAVITY.

There are two variables that determine the primary frequency of any resonant cavity. The first variable is PHYSICAL SIZE. In general, the smaller the cavity, the higher its resonant frequency. The second controlling factor is the SHAPE of the cavity. Figure 1-60 illustrates several cavity shapes that are commonly used. Remember from the previously stated definition of a resonant cavity that any completely enclosed conductive surface, regardless of its shape, can act as a cavity resonator.

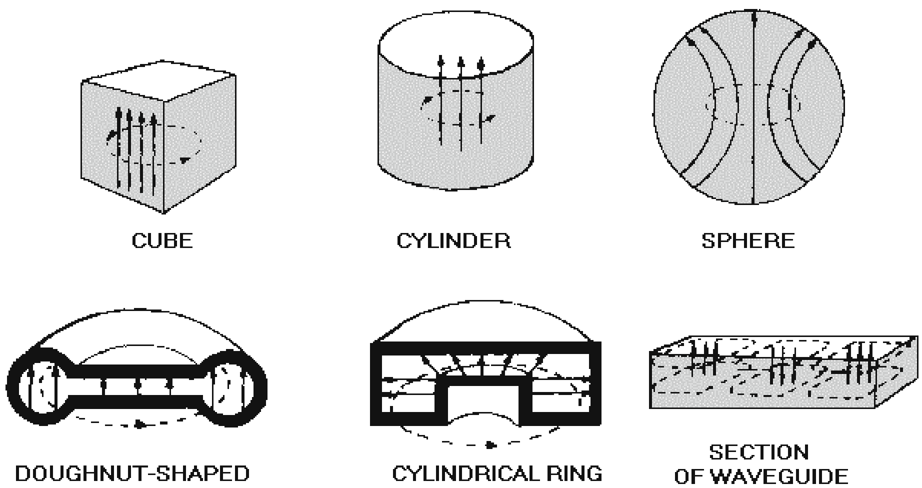


Figure 1-60.—Several types of cavities.

Cavity resonators are energized in basically the same manner as waveguides and have a similar field distribution. If the cavity shown in figure 1-61 were energized in the TE mode, the electromagnetic wave would reflect back and forth along the Z axis and form standing waves. These standing waves would form a field configuration within the cavity that would have to satisfy the same boundary conditions as those in a waveguide. Modes of operation in the cavity are described in terms of the fields that exist in the X, Y, and Z dimensions. Three subscripts are used; the first subscript indicates the number of $1/2\lambda$ along the X axis; the second subscript indicates the number of $1/2\lambda$ along the Y axis; and the third subscript indicates the number of $1/2\lambda$ along the Z axis.

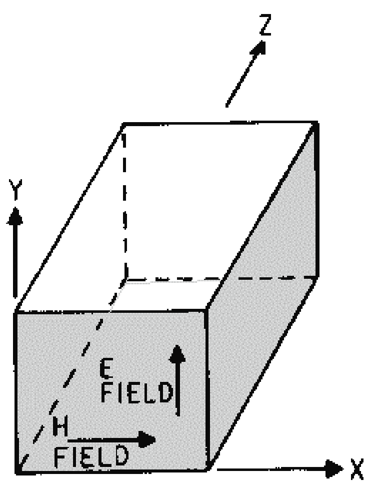


Figure 1-61.—Rectangular cavity resonator.

Energy can be inserted or removed from a cavity by the same methods that are used to couple energy into and out of waveguides. The operating principles of probes, loops, and slots are the same whether used in a cavity or a waveguide. Therefore, any of the three methods can be used with cavities to inject or remove energy.

The resonant frequency of a cavity can be varied by changing any of three parameters: cavity volume, cavity capacitance, or cavity inductance. Changing the frequencies of a cavity is known as TUNING. The mechanical methods of tuning a cavity may vary with the application, but all methods use the same electrical principles.

A mechanical method of tuning a cavity by changing the volume (VOLUME TUNING) is illustrated in figure 1-62. Varying the distance d will result in a new resonant frequency because the inductance and the capacitance of the cavity are changed by different amounts. If the volume is decreased, the resonant frequency will be higher. The resonant frequency will be lower if the volume of the cavity is made larger.

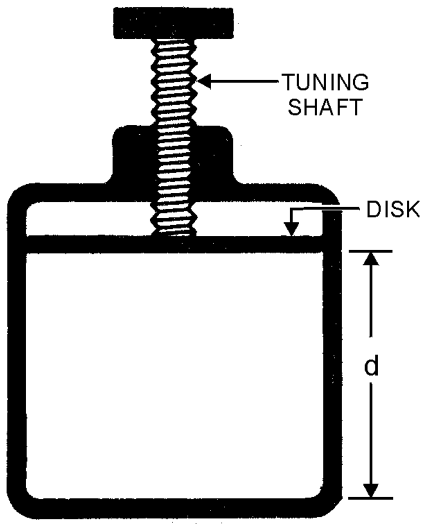


Figure 1-62.—Cavity tuning by volume.

CAPACITIVE TUNING of a cavity is shown in figure 1-63A. An adjustable slug or screw is placed in the area of maximum E lines. The distance d represents the distance between two capacitor plates. As the slug is moved in, the distance between the two plates becomes smaller and the capacitance increases. The increase in capacitance causes a decrease in the resonant frequency. As the slug is moved out, the resonant frequency of the cavity increases.

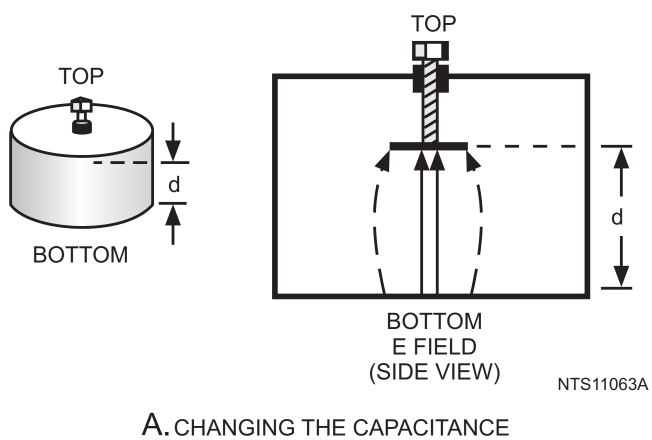


Figure 1-63A.—Methods of changing the resonant frequency of a cavity. CHANGING THE CAPACITANCE.

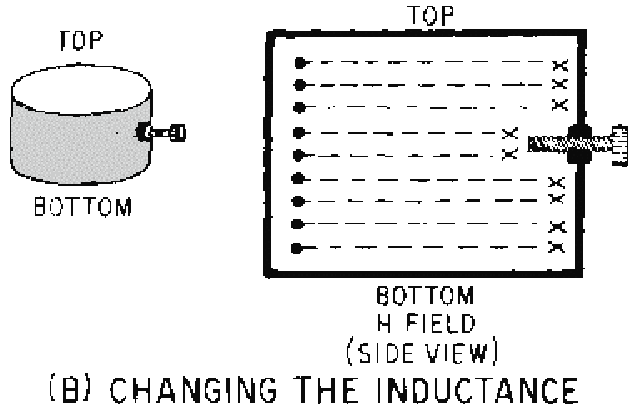


Figure 1-63B.—Methods of changing the resonant frequency of a cavity. CHANGING THE INDUCTANCE.

INDUCTIVE TUNING is accomplished by placing a nonmagnetic slug in the area of maximum H lines, as shown in figure 1-63B. The changing H lines induce a current in the slug that sets up an opposing H field. The opposing field reduces the total H field in the cavity, and therefore reduces the total inductance. Reducing the inductance, by moving the slug in, raises the resonant frequency. Increasing the inductance, by moving the slug out, lowers the resonant frequency.

Resonant cavities are widely used in the microwave range, and many of the applications will be studied in chapter 2. For example, most microwave tubes and transmitting devices use cavities in some form to generate microwave energy. Cavities are also used to determine the frequency of the energy traveling in a waveguide, since conventional measurement devices do not work well at microwave frequencies.

Q-50. What two variables determine the primary frequency of a resonant cavity?

Q-51. Energy can be inserted or removed from a cavity by what three methods?

Q-52. Inductive tuning of a resonant cavity is accomplished by placing a nonmagnetic slug in what area?

Waveguide Junctions

You may have assumed that when energy traveling down a waveguide reaches a junction, it simply divides and follows the junction. This is not strictly true. Different types of junctions affect the energy in different ways. Since waveguide junctions are used extensively in most systems, you need to understand the basic operating principles of those most commonly used.

The T JUNCTION is the most simple of the commonly used waveguide junctions. T junctions are divided into two basic types, the E-TYPE and the H-TYPE. HYBRID JUNCTIONS are more complicated developments of the basic T junctions. The MAGIC-T and the HYBRID RING are the two most commonly used hybrid junctions.

E-TYPE T JUNCTION.—An E-type T junction is illustrated in figure 1-64, view (A). It is called an E-type T junction because the junction arm extends from the main waveguide in the same direction as the E field in the waveguide.

Figure 1-64, view (B), illustrates cross-sectional views of the E-type T junction with inputs fed into the various arms. For simplicity, the magnetic lines that are always present with an electric field have been omitted. In view (K), the input is fed into arm b and the outputs are taken from the a and c arms. When the E field arrives between points 1 and 2, point 1 becomes positive and point 2 becomes negative. The positive charge at point 1 then induces a negative charge on the wall at point 3. The negative charge at point 2 induces a positive charge at point 4. These charges cause the fields to form 180 degrees out of phase in the main waveguide; therefore, the outputs will be 180 degrees out of phase with each other. In view (L), two in-phase inputs of equal amplitude are fed into the a and c arms. The signals at points 1 and 2 have the same phase and amplitude. No difference of potential exists across the entrance to the b arm, and no energy will be coupled out. However, when the two signals fed into the a and c arms are 180 degrees out of phase, as shown in view (M), points 1 and 2 have a difference of potential. This difference of potential induces an E field from point 1 to point 2 in the b arm, and energy is coupled out of this arm. Views (N) and (P) illustrate two methods of obtaining two outputs with only one input.

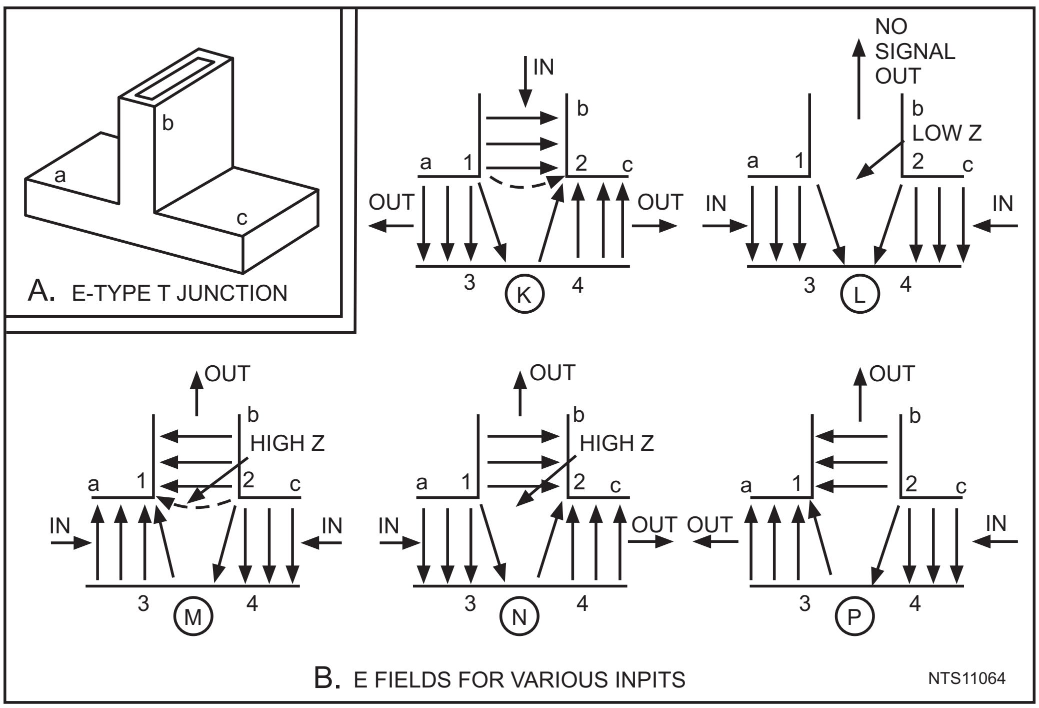


Figure 1-64.—E fields in an E-type T junction.

H-TYPE T JUNCTION.—An H-type T junction is illustrated in figure 1-65A. It is called an H-type T junction because the long axis of the "b" arm is parallel to the plane of the magnetic lines of force in the waveguide. Again, for simplicity, only the E lines are shown in this figure. Each X indicates an E line moving away from the observer. Each dot indicates an E line is moving toward the observer.

In view (1) of figure 1-65B, the signal is fed into arm b and in-phase outputs are obtained from the a and c arms. In view (2), in-phase signals are fed into arms a and c and the output signal is obtained from the b arm because the fields add at the junction and induce E lines into the b arm. If 180-degree-out-of-phase signals are fed into arms a and c, as shown in view (3), no output is obtained from the b arm because the opposing fields cancel at the junction. If a signal is fed into the a arm, as shown in view (4), outputs will be obtained from the b and c arms. The reverse is also true. If a signal is fed into the c arm, outputs will be obtained from the a and b arms.

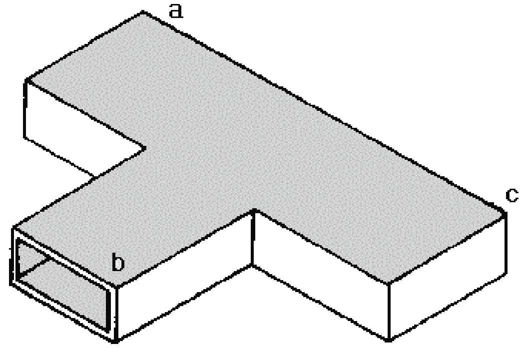


Figure 1-65A.—E fields in an H-type junction. H-TYPE T JUNCTION.

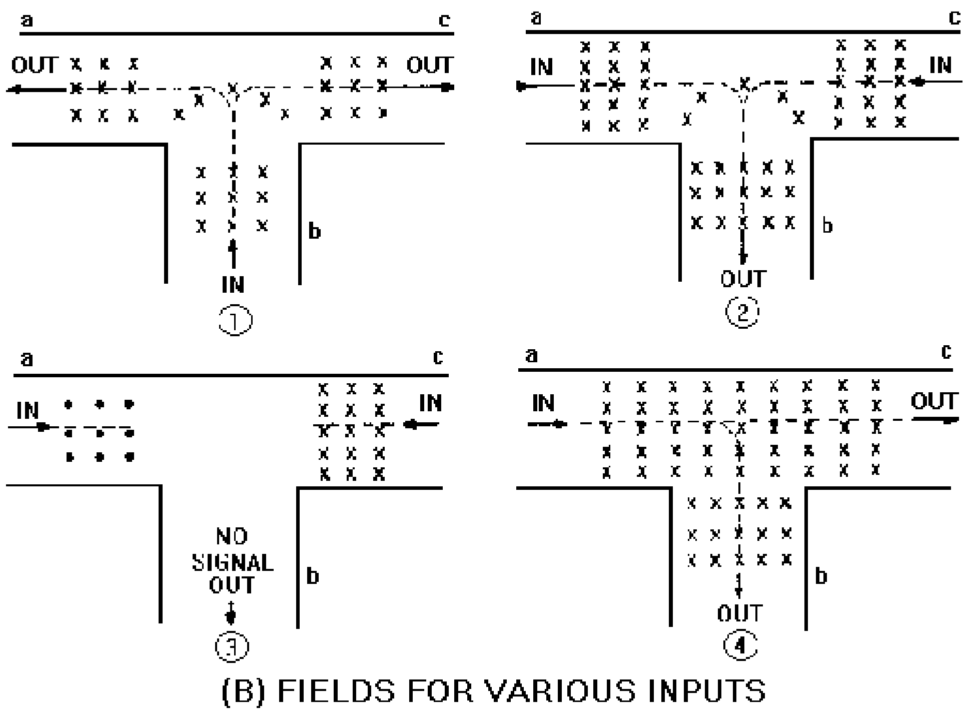


Figure 1-65B.—E fields in an H-type junction. FIELDS FOR VARIOUS INPUTS.

MAGIC-T HYBRID JUNCTION.—A simplified version of the magic-T hybrid junction is shown in figure 1-66. The magic-T is a combination of the H-type and E-type T junctions. The most common application of this type of junction is as the mixer section for microwave radar receivers. Its operation as a mixer will be discussed in later *NEETS* modules. At present, only the fields within the magic-T junction will be discussed.

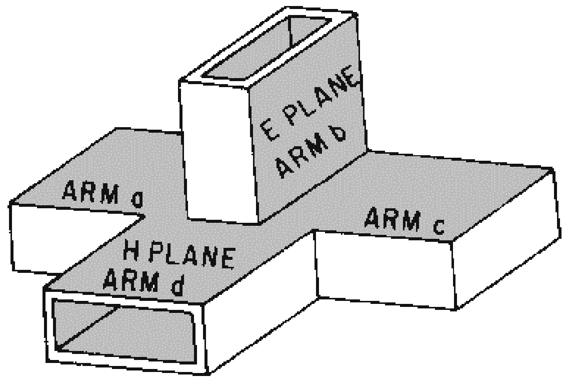


Figure 1-66.—Magic-T hybrid junction.

If a signal is fed into the b arm of the magic- T, it will divide into two out-of-phase components. As shown in figure 1-67A, these two components will move into the a and c arms. The signal entering the b arm will not enter the d arm because of the zero potential existing at the entrance of the d arm. The potential must be zero at this point to satisfy the boundary conditions of the b arm. This absence of potential is illustrated in figures 1-67B and 1-67C where the magnitude of the E field in the b arm is indicated by the length of the arrows. Since the E lines are at maximum in the center of the b arm and minimum at the edge where the d arm entrance is located, no potential difference exists across the mouth of the d arm.

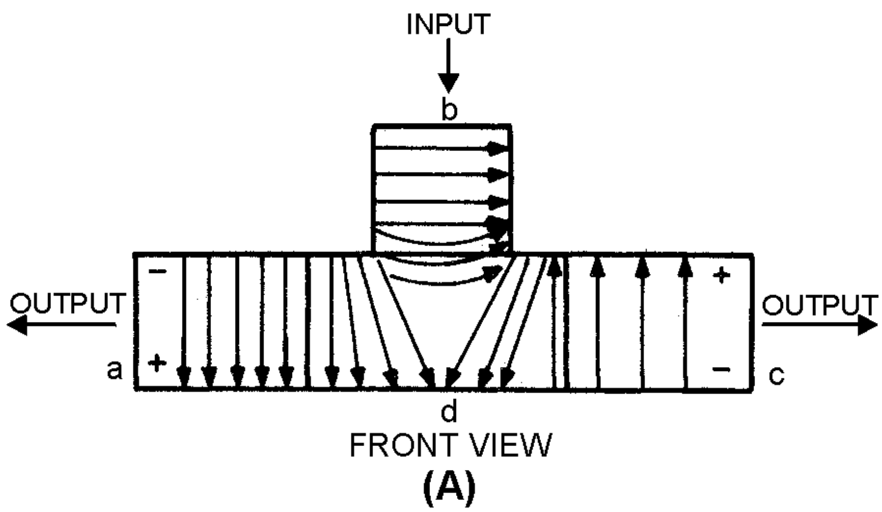


Figure 1-67A.—Magic-T with input to arm b.

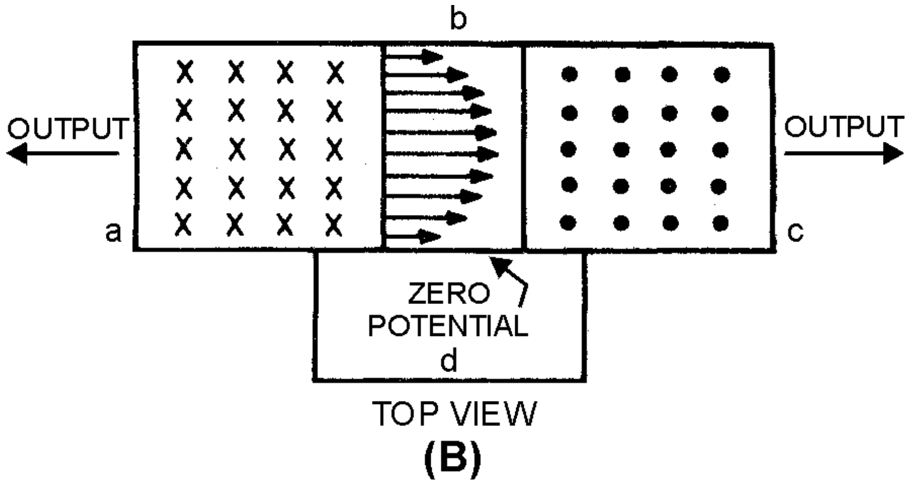


Figure 1-67B.—Magic-T with input to arm b.

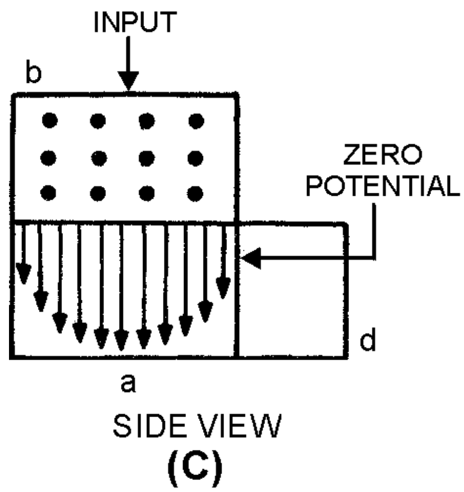


Figure 1-67C.—Magic-T with input to arm b.

In summary, when an input is applied to arm b of the magic-T hybrid junction, the output signals from arms a and c are 180 degrees out of phase with each other, and no output occurs at the d arm.

The action that occurs when a signal is fed into the d arm of the magic-T is illustrated in figure 1-68. As with the H-type T junction, the signal entering the d arm divides and moves down the a and c arms as outputs which are in phase with each other and with the input. The shape of the E fields in motion is shown by the numbered curved slices. As the E field moves down the d arm, points 2 and 3 are at an equal potential. The energy divides equally into arms a and c, and the E fields in both arms become identical in shape. Since the potentials on both sides of the b arm are equal, no potential difference exists at the entrance to the b arm, resulting in no output.

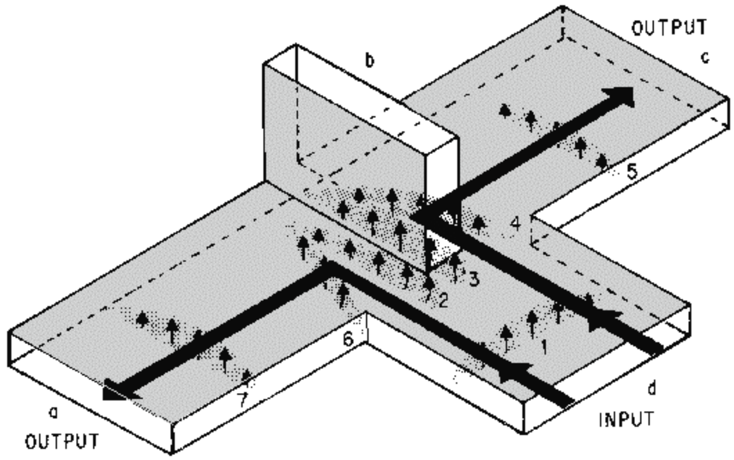


Figure 1-68.—Magic-T with input to arm d.

When an input signal is fed into the a arm as shown in figure 1-69, a portion of the energy is coupled into the b arm as it would be in an E-type T junction. An equal portion of the signal is coupled through the d arm because of the action of the H-type junction. The c arm has two fields across it that are out of phase with each other. Therefore the fields cancel, resulting in no output at the c arm. The reverse of this action takes place if a signal is fed into the c arm, resulting in outputs at the b and d arms and no output at the a arm.

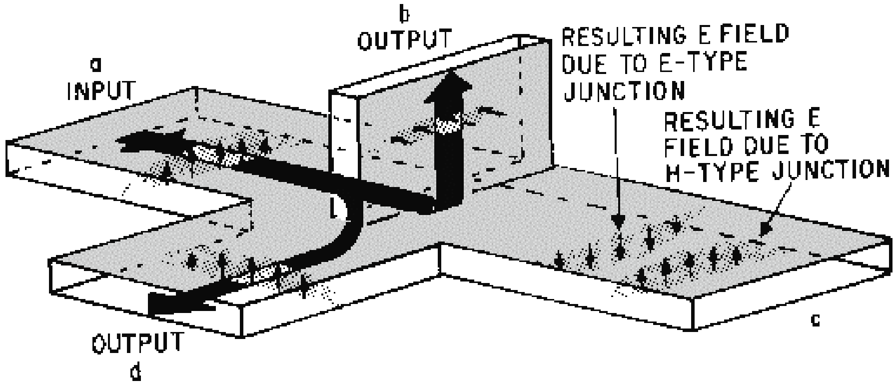


Figure 1-69.—Magic-T with input to arm a.

Unfortunately, when a signal is applied to any arm of a magic-T, the flow of energy in the output arms is affected by reflections. Reflections are caused by impedance mismatching at the junctions. These reflections are the cause of the two major disadvantages of' the magic-T. First, the reflections represent a power loss since all the energy fed into the junction does not reach the load which the arms feed. Second, the reflections produce standing waves that can result in internal arching. Thus the maximum power a magic-T can handle is greatly reduced.

Reflections can be reduced by using some means of' impedance matching that does not destroy the shape of' the junctions. One method is shown in figure 1-70. A post is used to match the H plane, and an iris is used to match the E plane. Even though this method reduces reflections, it lowers the power-handling capability even further.

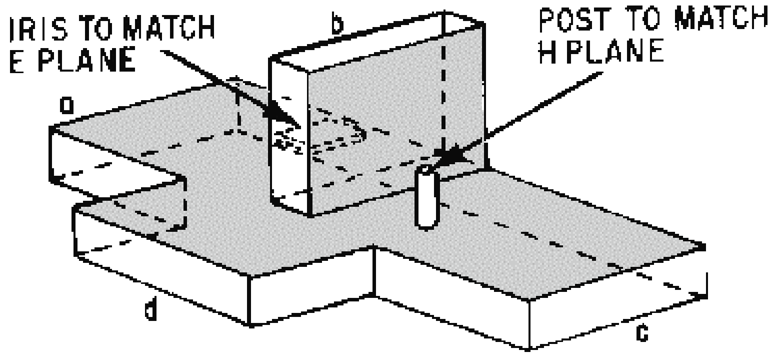


Figure 1-70.—Magic-T impedance matching.

HYBRID RING.—A type of hybrid junction that overcomes the power limitation of the magic-T is the hybrid ring, also called a RAT RACE. The hybrid ring, illustrated in figure 1-71A, is actually a modification of the magic-T. It is constructed of rectangular waveguides molded into a circular pattern. The arms are joined to the circular waveguide to form E-type T junctions. Figure 1-71B shows, in wavelengths, the dimensions required for a hybrid ring to operate properly.

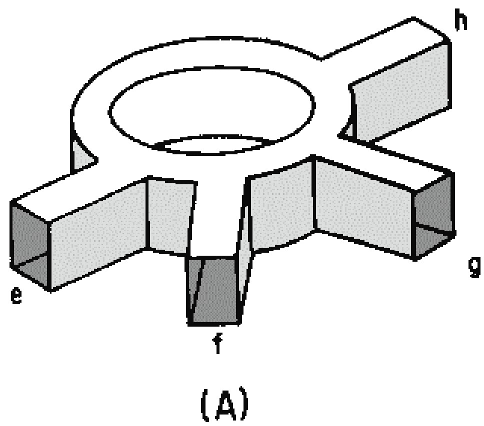


Figure 1-71A.—Hybrid ring with wavelength measurements.

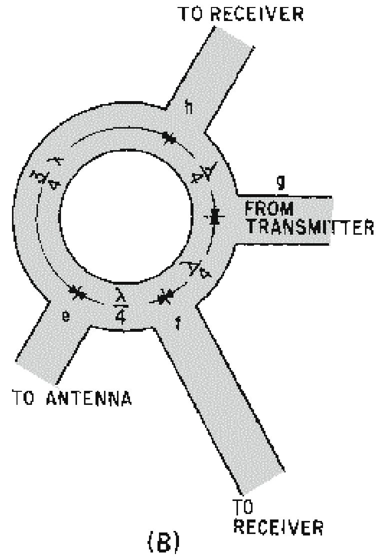


Figure 1-71B.—Hybrid ring with wavelength measurements.

The hybrid ring is used primarily in high-powered radar and communications systems to perform two functions. During the transmit period, the hybrid ring couples microwave energy from the transmitter to the antenna and allows no energy to reach the receiver. During the receive cycle, the hybrid ring couples energy from the antenna to the receiver and allows no energy to reach the transmitter. Any device that performs both of these functions is called a DUPLEXER. A duplexer permits a system to use the same antenna for both transmitting and receiving. Since the only common application of the hybrid ring is as a duplexer, the details of hybrid ring operation will be explained in later *NEETS* modules concerning duplexers.

Q-53. What are the two basic types of T junctions?

Q-54. Why is the H-type T junction so named?

Q-55. The magic-T is composed of what two basic types of T junctions?

Q-56. What are the primary disadvantages of the magic-T?

Q-57. What type of junctions are formed where the arms of a hybrid ring meet the main ring?

Q-58. Hybrid rings are used primarily for what purpose?

Ferrite Devices

A FERRITE is a device that is composed of material that causes it to have useful magnetic properties and, at the same time, high resistance to current flow. The primary material used in the construction of ferrites is normally a compound of iron oxide with impurities of other oxides added. The compound of iron oxide retains the properties of the ferromagnetic atoms, and the impurities of the other oxides increase the resistance to current flow. This combination of properties is not found in conventional magnetic materials. Iron, for example, has good magnetic properties but a relatively low resistance to current flow. The low resistance causes eddy currents and significant power losses at high frequencies (You may want to review *NEETS*, Module 2, *Introduction to Alternating Current and Transformers*, Chapter 5). Ferrites, on the other hand, have sufficient resistance to be classified as semiconductors.

The compounds used in the composition of ferrites can be compared to the more familiar compounds used in transistors. As in the construction of transistors, a wide range of magnetic and electrical properties can be produced by the proper choice of atoms in the right proportions. An example of a ferrite device is shown in figure 1-72.

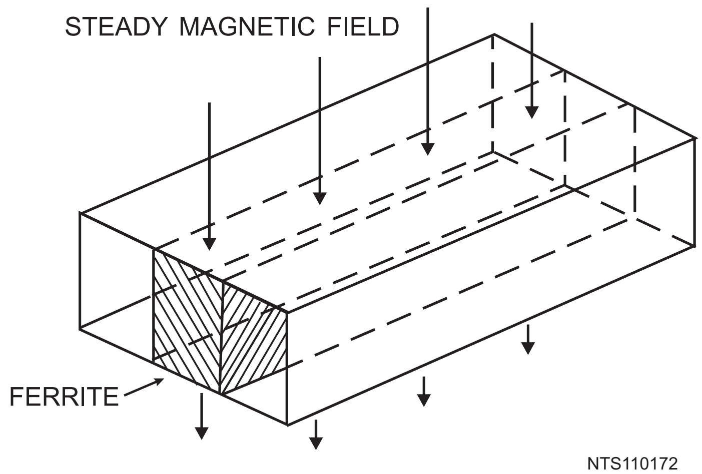


Figure 1-72.—Ferrite attenuator.

Ferrites have long been used at conventional frequencies in computers, television, and magnetic recording systems. The use of ferrites at microwave frequencies is a relatively new development and has had considerable influence on the design of microwave systems. In the past, the microwave equipment was made to conform to the frequency of the system and the design possibilities were limited. The unique properties of ferrites provide a variable reactance by which microwave energy can be manipulated to conform to the microwave system. At present, ferrites are used as LOAD ISOLATORS, PHASE SHIFTERS, VARIABLE ATTENUATORS, MODULATORS, and SWITCHES in microwave systems. The operation of ferrites as isolators, attenuators, and phase shifters will be explained in the following paragraphs. The operation of ferrites in other applications will be explained in later *NEETS* modules. Ferromagnetism is a continuation of the conventional domain theory of magnetism that was explained in *NEETS*, Module 1, *Matter, Energy, and Direct Current*. A review of the section on magnetism might be helpful to you at this point.

The magnetic property of any material is a result of electron movement within the atoms of the material. Electrons have two basic types of motion. The most familiar is the ORBITAL movement of the electron about the nucleus of the atom. Less familiar, but even more important, is the movement of the electron about its own axis, called ELECTRON SPIN.

You will recall that magnetic fields are generated by current flow. Since current is the movement of electrons, the movement of the electrons within an atom create magnetic fields. The magnetic fields caused by the movement of the electrons about the nucleus have little effect on the magnetic properties of a material. The magnetic fields caused by electron spin combine to give a material magnetic properties. The different types of electron movement are illustrated in figure 1-73. In most materials the spin axes of the electrons are so randomly arranged that the magnetic fields largely cancel out and the material displays no significant magnetic properties. The electron spin axes within some materials, such as iron and nickel, can be caused to align by applying an external magnetic field. The alignment of the electrons within a material causes the magnetic fields to add, and the material then has magnetic properties.

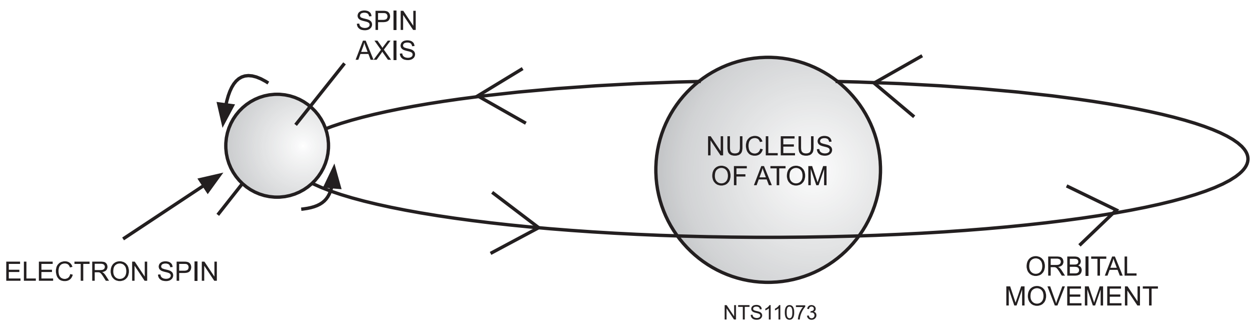


Figure 1-73.—Two types of electron movement.

In the absence of an external force, the axis of any spinning object tends to remain pointed in one direction. Spinning electrons behave the same way. Therefore, once the electrons are aligned, they tend to remain aligned even when the external field is removed. Electron alignment in a ferrite is caused by the orbital motion of the electrons about the nucleus and the force that holds the atom together. When a static magnetic field is applied, the electrons try to align their spin axes with the new force. The attempt of the electrons to balance between the interaction of the new force and the binding force causes the electrons to wobble on their axes, as shown in figure 1-74. The wobble of the electrons has a natural resonant WOBBLE FREQUENCY that varies with the strength of the applied field. Ferrite action is based on this behavior of the electrons under the influence of an external field and the resulting wobble frequency.

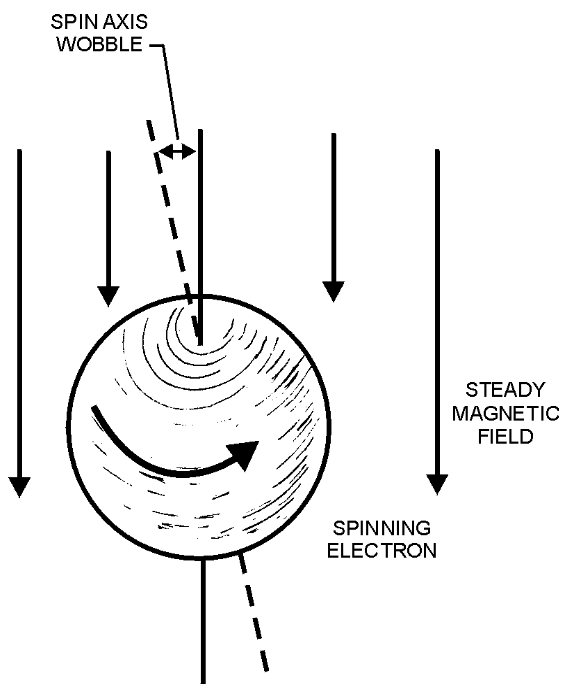


Figure 1-74.—Electron wobble in a magnetic field.

FERRITE ATTENUATORS.—A ferrite attenuator can be constructed that will attenuate a particular microwave frequency and allow all others to pass unaffected. This can be done by placing a ferrite in the center of a waveguide, as shown in figure 1-72. The ferrite must be positioned so that the magnetic fields caused by its electrons are perpendicular to the energy in the waveguide. A steady external field causes the electrons to wobble at the same frequency as the energy that is to be attenuated.

Since the wobble frequency is the same as the energy frequency, the energy in the waveguide always adds to the wobble of the electrons. The spin axis of the electron changes direction during the wobble motion and energy is used. The force causing the increase in wobble is the energy in the waveguide. Thus, the energy in the waveguide is attenuated by the ferrite and is given off as heat. Energy in the waveguide that is a different frequency from the wobble frequency of the ferrite is largely unaffected because it does not increase the amount of electron wobble. The resonant frequency of electron wobble can be varied over a limited range by changing the strength of the applied magnetic field.

FERRITE ISOLATORS.—An isolator is a ferrite device that can be constructed so that it allows microwave energy to pass in one direction but blocks energy in the other direction in a waveguide. This isolator is constructed by placing a piece of ferrite off-center in a waveguide, as shown in figure 1-75. A magnetic field is applied by the magnet and adjusted to make the electron wobble frequency of the ferrite equal to the frequency of the energy traveling down the waveguide. Energy traveling down the waveguide from left to right will set up a rotating magnetic field that rotates through the ferrite material in the same direction as the natural wobble of the electrons. The aiding magnetic field increases the wobble of the ferrite electrons so much that almost all of the energy in the waveguide is absorbed and dissipated as heat. The magnetic fields caused by energy traveling from right to left rotate in the opposite direction through the ferrite and have very little effect on the amount of electron wobble. In this case the fields attempt to push the electrons in the direction opposite the natural wobble and no large movements occur. Since no overall energy exchange takes place, energy traveling from right to left is affected very little.

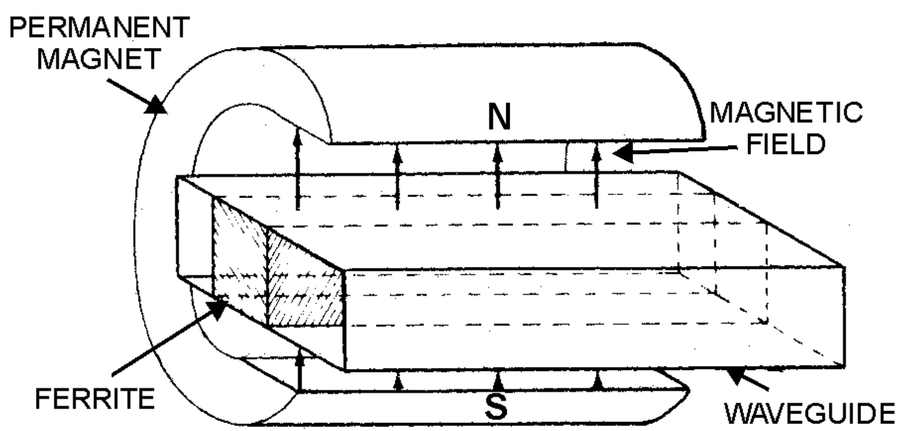


Figure 1-75.—One-way isolator.

FERRITE PHASE SHIFTER.—When microwave energy is passed through a piece of ferrite in a magnetic field, another effect occurs. If the frequency of the microwave energy is much greater than the electron wobble frequency, the plane of polarization of the wavefront is rotated. This is known as the FARADAY ROTATION EFFECT and is illustrated in figure 1-76. A ferrite rod is placed along the axis of the waveguide, and a magnetic field is set up along the axis by a coil. As a wavefront enters the section containing the ferrite, it sets up a limited motion in the electrons. The magnetic fields of the wavefront and the wobbling electrons interact, and the polarization of the wavefront is rotated. The amount of rotation depends upon the length of the ferrite rod. The direction of rotation depends upon the direction of the external magnetic field and can be reversed by reversing the field. The direction of rotation will remain constant, no matter what direction the energy in the waveguide travels, as long as the external field is not changed.

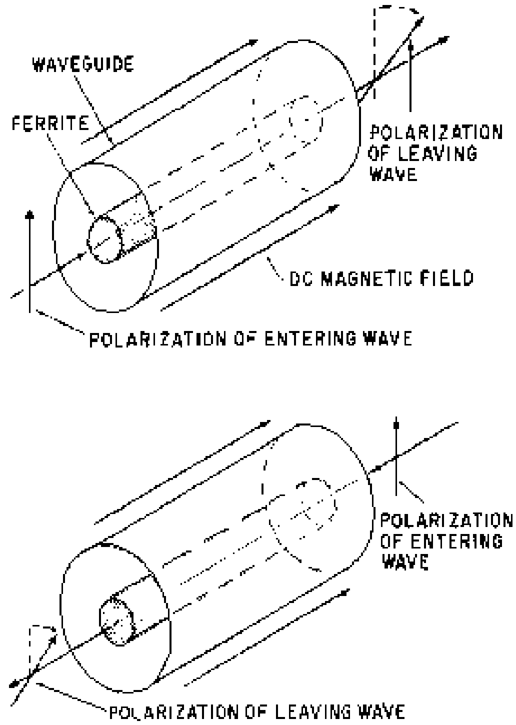


Figure 1-76.—Faraday rotation.

Q-59. Ferrite devices are useful in microwave applications because they possess what properties?

Q-60. Which of the two types of electron motion (orbital movement and electron spin) is more important in the explanation of magnetism?

Q-61. The interaction between an external field and the binding force of an atom causes electrons to do what?

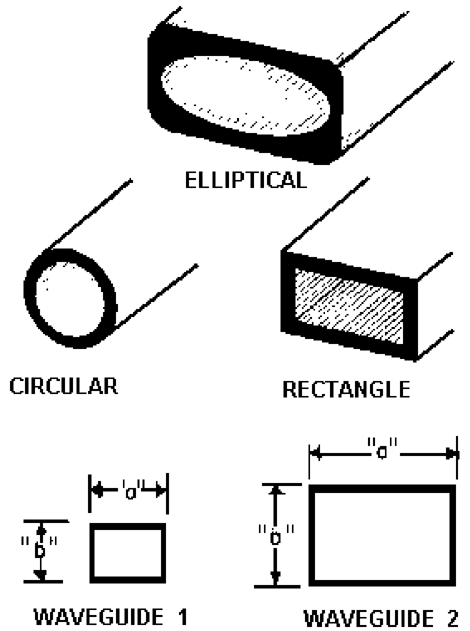
Q-62. The resonant frequency of electron wobble can be changed by variation of what force?

Q-63. Rotating the plane of polarization of a wavefront by passing it through a ferrite device is called what?

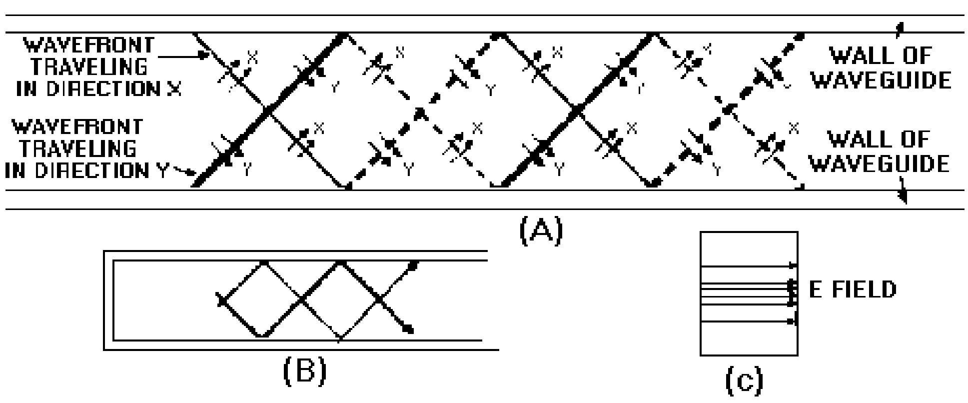
SUMMARY

This chapter has presented information on waveguide theory and application. The information that follows summarizes the important points of this chapter.

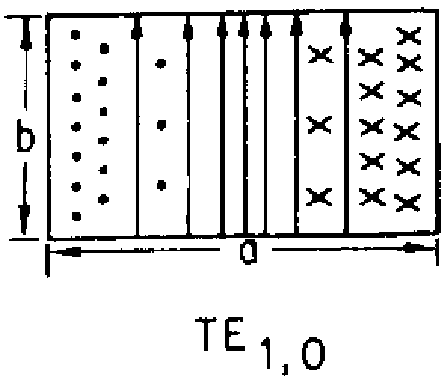
WAVEGUIDES are the primary methods of transporting microwave energy. Waveguides have fewer losses and greater power-handling capability than transmission lines. The physical size of the waveguides becomes too great for use at frequencies less than 1000 megahertz. Waveguides are made in three basic shapes, as shown in the first illustration. The wide, or "a," dimension determines the frequency range of the waveguide, and the narrow, or "b," dimension determines power-handling capability as shown in the second illustration. Waveguides handle a small range of frequencies both above and below the primary operating frequency. Energy is transported through waveguides by the interaction of electric and magnetic fields, abbreviated E FIELD and H FIELD, respectively. The density of the E field varies at the same rate as the applied voltage. If energy is to travel through a waveguide, two BOUNDARY CONDITIONS must be met: (1) An electric field, to exist at the surface of a conductor, must be perpendicular to the conductor, and (2) a varying magnetic field must exist in closed loops parallel to the conductors and perpendicular to the electric field.



WAVEFRONTS travel down a waveguide by reflecting from the side walls in a zigzag pattern, as shown in the figure. The striking angle, or angle of incidence (θ), is the same as the angle of reflection (θ), causing the reflected wavefront to have the same shape as the incident wavefront. The velocity of wavefronts traveling down a waveguide is called the GROUP VELOCITY because of the zigzag path of these wavefronts. The group velocity is slower than the velocity of wavefronts through space.

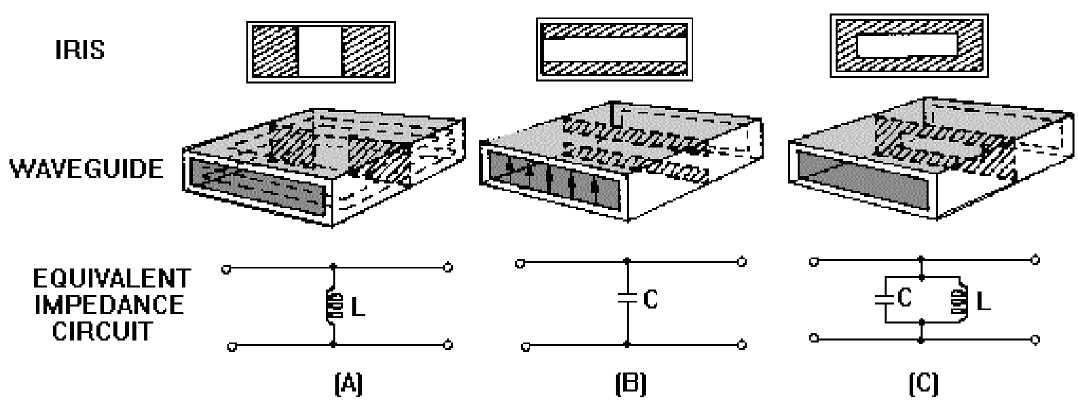


The **MODES** in waveguides are divided into two categories: (1) the TRANSVERSE ELECTRIC (TE) mode and (2) the TRANSVERSE MAGNETIC (TM) mode. Subscripts are used to complete the description of the various TE and TM modes. The dominant mode for rectangular waveguides is shown in the figure.



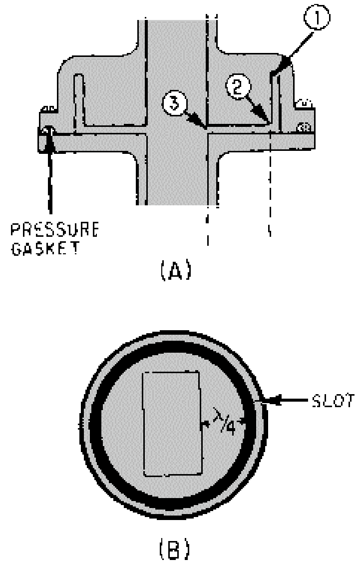
WAVEGUIDE INPUT/OUTPUT METHODS are divided into three basic categories: PROBES, LOOPS, and SLOTS. Size, shape, and placement in the waveguide are critical factors in the efficiency of all three input/output methods.

WAVEGUIDE/IMPEDANCE MATCHING is often necessary to reduce reflections caused by a MISMATCH between the waveguide and the load. Matching devices called IRISES, shown in the illustration, are used to introduce either capacitance or inductance (or a combination of both) into a waveguide. Conductive POSTS and SCREWS can also be used for impedance matching in waveguides.

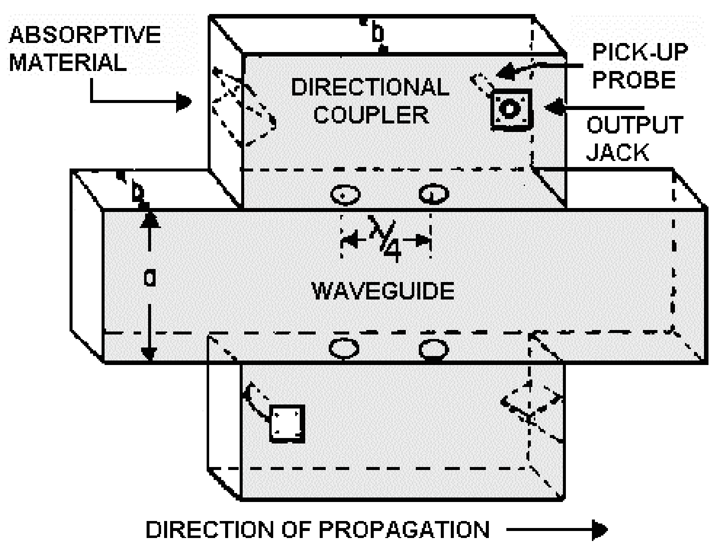


WAVEGUIDE TERMINATIONS prevent standing waves at the end of a waveguide system. They are usually specially constructed HORNS or absorptive loads called DUMMY LOADS.

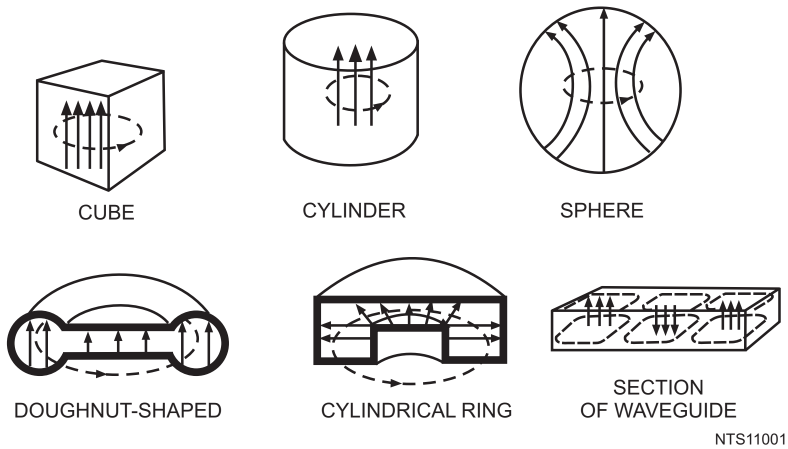
WAVEGUIDE PLUMBING refers to the bends, twists, and joints necessary to install waveguides. E bends, H bends, and twists must have a radius greater than two wavelengths. The CHOKE JOINT, shown in the figure, is most often used to connect two pieces of waveguide. The ROTATING JOINT is used when a waveguide must be connected to a rotating load such as an antenna.



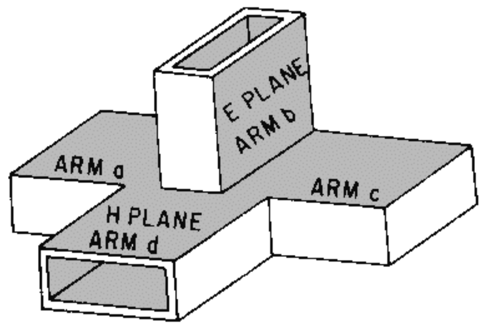
DIRECTIONAL COUPLERS are devices that permit the sampling of the energy in a waveguide. Directional couplers may be constructed to sample energy in one direction only or in both directions. The energy removed by the directional coupler is a small sample that is proportional to the magnitude of the energy in the waveguide. An example of a directional coupler is shown in the illustration.



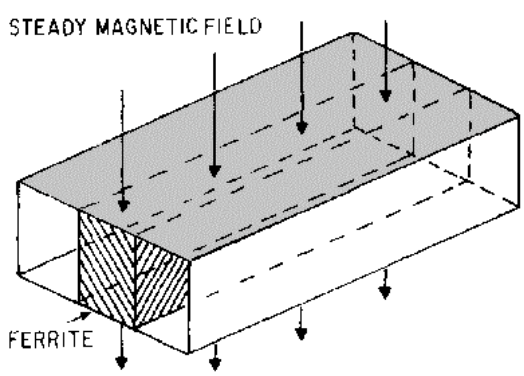
A **RESONANT CAVITY** is any space completely enclosed by conductive walls that can contain oscillating electromagnetic fields and can possess resonant properties. Several cavity shapes are shown in the illustration.



WAVEGUIDE JUNCTIONS are of several basic types. The T-JUNCTION may be either of the E-TYPE or the H-TYPE. The effect on the input energy depends upon which arm is used as the input. The MAGIC-T HYBRID JUNCTION, shown at the right, is a combination of the E- and H-type T junctions.



FERRITE DEVICES combine magnetic properties with a high resistance to current flow. Ferrites are constructed from compounds of ferrous metal oxides to achieve the desired characteristics. The fact that the spin axes of electrons will wobble at a natural resonant frequency when subjected to an external magnetic field is the basic principle of operation of ferrite devices. The position of a typical ferrite device within a waveguide is shown in the figure.



ANSWERS TO QUESTIONS Q1. THROUGH Q63.

A-1. Microwave region.

A-2. Electromagnetic field theory.

A-3. The electromagnetic fields are completely confined.

A-4. Conductive material.

A-5. Copper loss.

A-6. Skin effect.

A-7. Air.

A-8. Physical size.

A-9. The characteristics of the dielectric of a capacitor.

A-10. A shorted quarter-wave section called a metallic insulator.

A-11. The "a" dimension.

A-12. The bus bar becomes wider.

A-13. Energy will no longer pass through the waveguide.

A-14. The interaction of the electric and magnetic fields.

A-15. The relative strength of the field.

A-16. Magnetic lines of force must form a continuous closed loop.

A-17. The H lines cancel.

A-18. The field must be perpendicular to the conductors.

A-19. Decrease to zero.

A-20. The angles are equal.

A-21. Cutoff frequency.

A-22. Slower.

A-23. Group velocity.

A-24. Mode of operation.

A-25. Dominant mode.

A-26. 1.71 times the diameter.

A-27. Transverse electric (TE) and transverse magnetic (TM).

A-28. TE.

A-29. Second.

A-30. First.

A-31. Size and shape.

A-32. Slots and apertures.

A-33. Standing waves that cause power losses, a reduction in power-handling capability, and an increase in frequency and sensitivity.

A-34. Metal plates.

A-35. Inductive.

A-36. As a shunt resistance.

A-37. Horn.

A-38. Characteristic impedance.

A-39. Absorb all energy without producing standing waves.

A-40. Heat.

A-41. Reflections.

A-42. Greater than 2 wavelengths.

A-43. Choke joint.

A-44. Improperly connected joints or damaged inner surface.

A-45. Sampling energy within a waveguide.

A-46. 1/4 wavelength.

A-47. Absorb the energy not directed at the pick-up probe and a portion of the overall energy.

A-48. The wavefront portions add.

A-49. The reflected energy adds at the absorbent material and is absorbed.

A-50. Size and shape of the cavity.

A-51. Probes, loops, and slots.

A-52. The area of maximum H lines.

A-53. E-type and H-type.

A-54. The junction arm extends in a direction parallel to the H lines in the main waveguide.

A-55. E-type and H-type.

A-56. Low power-handling capability and power losses.

A-57. Basic E-type junctions.

A-58. High-power duplexes.

A-59. Magnetic properties and high resistance.

A-60. Electron spin.

A-61. Wobble at a natural resonant frequency.

A-62. The applied magnetic field.

A-63. Faraday rotation.

CHAPTER 2

MICROWAVE COMPONENTS AND CIRCUITS

LEARNING OBJECTIVES

Upon completion of this chapter the student will be able to:

1. Explain the basic principles of microwave tubes and describe the limitations of conventional tubes.
2. Describe the basic principles of velocity modulation.
3. Outline the development of microwave tubes.
4. Describe the basic theory of operation of klystrons including multicavity and reflex klystrons.
5. Explain the basic theory of operation of traveling-wave tubes and backward-wave oscillators.
6. Describe the construction, basic theories of operation, and typical applications of magnetrons and amplitrons.
7. Describe the basic theory of operation of tunnel diodes when used in oscillator-, amplifier-, and frequency-converter circuits.
8. Explain the operation of varactors when used in parametric amplifiers and frequency converters.
9. State the basic principles of operation of bulk-effect diodes and the gunn oscillator.
10. Explain the basic operation of passive microwave diodes in terms of theory and application.
11. Explain the basic operation of microwave transistors in terms of theory and application.

MICROWAVE COMPONENTS

The waveguides discussed in chapter 1 serve to transport microwave energy from one place to another. Energy is transported after it has been generated or amplified in a previous stage of the circuit. In this chapter you will be introduced to the special components used in those circuits.

Microwave energy is used in both radar and communications applications. The fact that the frequencies are very high and the wavelengths very short presents special problems in circuit design. Components that were previously satisfactory for signal generation and amplification use are no longer useful in the microwave region. The theory of operation for these components is discussed in this chapter. Because the theory of operation is sometimes difficult to understand, you need to pay particular attention to detail as you study this chapter. It is written in the simplest manner possible while retaining the necessary technical complexity.

MICROWAVE TUBE PRINCIPLES

The efficiency of conventional tubes is largely independent of frequency up to a certain limit. When frequency increases beyond that limit, several factors combine to rapidly decrease tube efficiency. Tubes that are efficient in the microwave range usually operate on the theory of VELOCITY MODULATION, a concept that avoids the problems encountered in conventional tubes. Velocity modulation is more easily understood if the factors that limit the frequency range of a conventional tube are thoroughly understood. Therefore, the frequency limitations of conventional tubes will be discussed before the concepts and applications of velocity modulation are explained. You may want to review *NEETS*, Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* Chapters 1 and 2, for a refresher on vacuum tubes before proceeding.

Frequency Limitations of Conventional Tubes

Three characteristics of ordinary vacuum tubes become increasingly important as frequency rises. These characteristics are interelectrode capacitance, lead inductance, and electron transit time.

The INTERELECTRODE CAPACITANCES in a vacuum tube, at low or medium radio frequencies, produce capacitive reactances that are so large that no serious effects upon tube operation are noticeable. However, as the frequency increases, the reactances become small enough to materially affect the performance of a circuit. For example, in figure 2-1A, a 1-picofarad capacitor has a reactance of 159,000 ohms at 1 megahertz. If this capacitor was the interelectrode capacitance between the grid and plate of a tube, and the rf voltage between these electrodes was 500 volts, then 3.15 milliamperes of current would flow through the interelectrode capacitance. Current flow in this small amount would not seriously affect circuit performance. On the other hand, at a frequency of 100 megahertz the reactance would decrease to approximately 1,590 ohms and, with the same voltage applied, current would increase to 315 milliamperes (figure 2-1B). Current in this amount would definitely affect circuit performance.

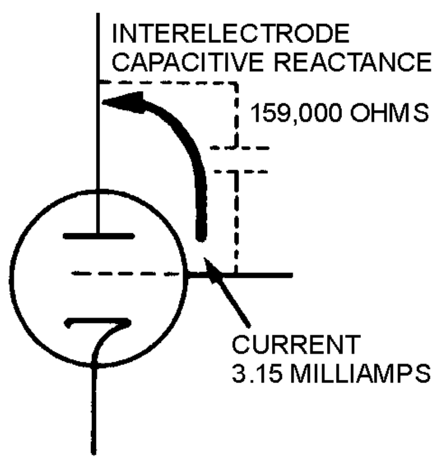


Figure 2-1A.—Interelectrode capacitance in a vacuum tube. 1 MEGAHERTZ.

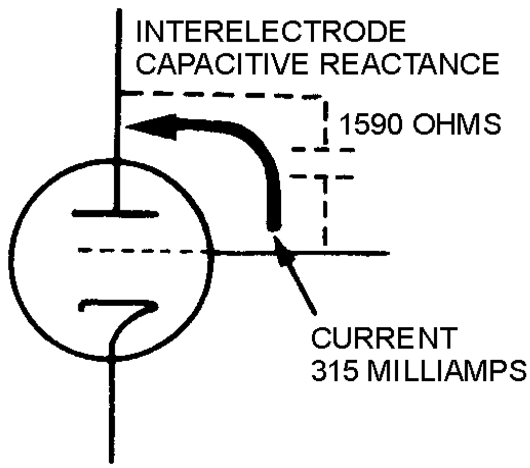


Figure 2-1B.—Interelectrode capacitance in a vacuum tube. 100 MEGAHERTZ.

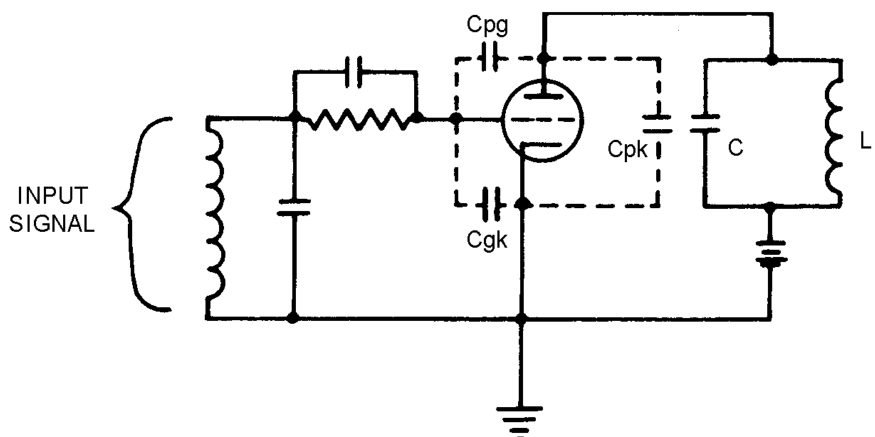


Figure 2-1C.—Interelectrode capacitance in a vacuum tube. INTERELECTRODE CAPACITANCE IN A TUNED-PLATE TUNED-GRID OSCILLATOR.

A good point to remember is that the higher the frequency, or the larger the interelectrode capacitance, the higher will be the current through this capacitance. The circuit in figure 2-1C, shows the interelectrode capacitance between the grid and the cathode (Cgk) in parallel with the signal source. As the frequency of the input signal increases, the effective grid-to-cathode impedance of the tube decreases because of a decrease in the reactance of the interelectrode capacitance. If the signal frequency is 100 megahertz or greater, the reactance of the grid-to-cathode capacitance is so small that much of the signal is short-circuited within the tube. Since the interelectrode capacitances are effectively in parallel with the tuned circuits, as shown in figures 2-1A, B, and C, they will also affect the frequency at which the tuned circuits resonate.

Another frequency-limiting factor is the LEAD INDUCTANCE of the tube elements. Since the lead inductances within a tube are effectively in parallel with the interelectrode capacitance, the net effect is to raise the frequency limit. However, the inductance of the cathode lead is common to both the grid and plate circuits. This provides a path for degenerative feedback which reduces overall circuit efficiency.

A third limitation caused by tube construction is TRANSIT TIME. Transit time is the time required for electrons to travel from the cathode to the plate. While some small amount of transit time is required for electrons to travel from the cathode to the plate, the time is insignificant at low frequencies. In fact, the transit time is so insignificant at low frequencies that it is generally not considered to be a hindering factor. However, at high frequencies, transit time becomes an appreciable portion of a signal cycle and begins to hinder efficiency. For example, a transit time of 1 nanosecond, which is not unusual, is only 0.001 cycle at a frequency of 1 megahertz. The same transit time becomes equal to the time required for an entire cycle at 1,000 megahertz. Transit time depends on electrode spacing and existing voltage potentials. Transit times in excess of 0.1 cycle cause a significant decrease in tube efficiency. This decrease in efficiency is caused, in part, by a phase shift between plate current and grid voltage.

If the tube is to operate efficiently, the plate current must be in phase with the grid-signal voltage and 180 degrees out of phase with the plate voltage. When transit time approaches 1/4 cycle, this phase relationship between the elements does not hold true. A positive swing of a high-frequency grid signal causes electrons to leave the cathode and flow to the plate. Initially this current is in phase with the grid voltage. However, since transit time is an appreciable part of a cycle, the current arriving at the plate now lags the grid-signal voltage. As a result, the power output of the tube decreases and the plate power dissipation increases. Another loss of power occurs because of ELECTROSTATIC INDUCTION.

The electrons forming the plate current also electrostatically induce potentials in the grid as they move past it. This electrostatic induction in the grid causes currents of positive charges to move back and forth in the grid structure. This back and forth action is similar to the action of hole current in semiconductor devices. When transit-time effect is not a factor (as in low frequencies), the current induced in one side of the grid by the approaching electrons is equal to the current induced on the other side by the receding electrons. The net effect is zero since the currents are in opposite directions and cancel each other. However, when transit time is an appreciable part of a cycle, the number of electrons approaching the grid is not always equal to the number going away. As a result, the induced currents do not cancel. This uncancelled current produces a power loss in the grid that is considered resistive in nature. In other words, the tube acts as if a resistor were connected between the grid and the cathode. The resistance of this imaginary resistor decreases rapidly as the frequency increases. The resistance may become so low that the grid is essentially short-circuited to the cathode, preventing proper operation of the tube.

Several methods are available to reduce the limitations of conventional tubes, but none work well when frequency increases beyond 1,000 megahertz. Interelectrode capacitance can be reduced by moving the electrodes further apart or by reducing the size of the tube and its electrodes. Moving the electrodes apart increases the problems associated with transit time, and reducing the size of the tube lowers the power-handling capability. You can see that efforts to reduce certain limitations in conventional tubes are compromises that are often in direct opposition to each other. The net effect is an upper limit of approximately 1,000 megahertz, beyond which conventional tubes are not practical.

Q-1. What happens to the impedance of interelectrode capacitance as frequency increases?

Q-2. What undesirable effect is caused by the inductance of the cathode lead?

Q-3. How does transit time affect the relationship of the grid voltage and the plate current at high frequencies?

Q-4. Moving tube electrodes apart to decrease interelectrode capacitance causes an increase in the effect of what property?

Velocity Modulation

The microwave tube was developed when the use of the frequency spectrum went beyond 1,000 megahertz and into the microwave range. The microwave tube uses transit time in the conversion of dc power to radio-frequency (rf) power. The interchange of power is accomplished by using the principle of electron VELOCITY MODULATION and low-loss resonant cavities in the microwave tube.

A clear understanding of microwave tubes must start with an understanding of how electrons and electric fields interact. An electron has mass and thus exhibits kinetic energy when in motion. The amount of kinetic energy in an electron is directly proportional to its velocity; that is, the higher the velocity, the higher the energy level. The basic concept of the electron energy level being directly related to electron velocity is the key principle of energy transfer and amplification in microwave tubes.

An electron can be accelerated or decelerated by an electrostatic field. Figure 2-2 shows an electron moving in an electrostatic field. The direction of travel (shown by the heavy arrow) is against the electrostatic lines of force which are from positive to negative. The negatively charged electron will be attracted to the positively charged body and will increase in velocity. As its velocity increases, the energy level of the electron will also increase. Where does the electron acquire its additional energy? The only logical source is from the electrostatic field. Thus, the conclusion is clear. An electron traveling in a direction opposite to electrostatic lines of force will absorb energy and increase in velocity (accelerate).

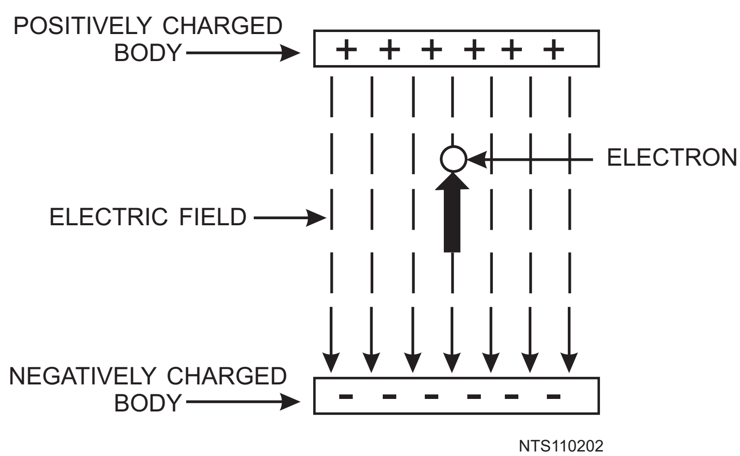


Figure 2-2.—Moving electron gaining velocity and energy.

As figure 2-3 illustrates, the opposite condition is also true. An electron traveling in the same direction as the electrostatic lines of force will decelerate by giving up energy to the field. The negatively charged body will repel the electron and cause it to decrease in velocity. When the velocity is reduced, the energy level is also reduced. The energy lost by the electron is gained by the electrostatic field.

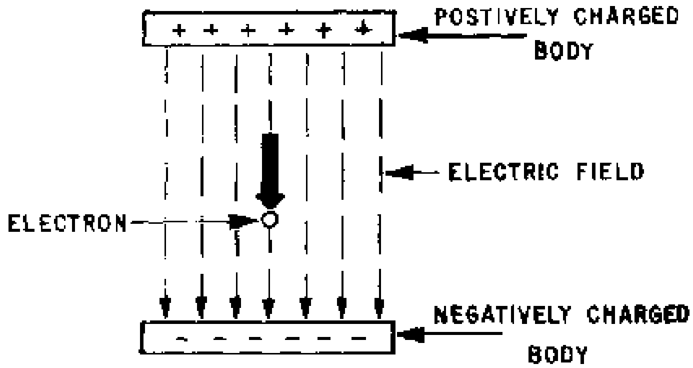


Figure 2-3.—Moving electron losing energy and velocity.

The operation of a velocity-modulated tube depends on a change in the velocity of the electrons passing through its electrostatic field. A change in electron velocity causes the tube to produce BUNCHES of electrons. These bunches are separated by spaces in which there are relatively few electrons. Velocity modulation is then defined as that variation in the velocity of a beam of electrons caused by the alternate speeding up and slowing down of the electrons in the beam. This variation is usually caused by a voltage signal applied between the grids through which the beam must pass.

The first requirement in obtaining velocity modulation is to produce a stream of electrons which are all traveling at the same speed. The electron stream is produced by an electron gun. A simplified version of an electron gun is shown in figure 2-4A. Electrons emitted from the cathode are attracted toward the positive accelerator grid and all but a few of the electrons pass through the grid and form a beam. The electron beam then passes through a pair of closely spaced grids, called BUNCHER GRIDS. Each grid is connected to one side of a tuned circuit. The parallel-resonant tuned circuit (figure 2-4A) in the illustration represents the doughnut-shaped resonant cavity surrounding the electron stream (figure 2-4B). The buncher grids are the dashed lines at the center of the cavity and are at the same dc potential as the accelerator grid. The alternating voltage which exists across the resonant circuit causes the velocity of the electrons leaving the buncher grids to differ from the velocity of the electrons arriving at the buncher grids. The amount of difference depends on the strength and direction of the electrostatic field within the resonant cavity as the electrons pass through the grids.

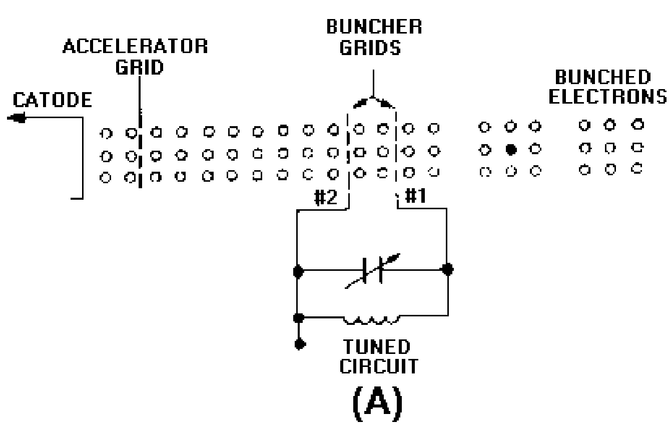


Figure 2-4A.—Electron gun with buncher grids.

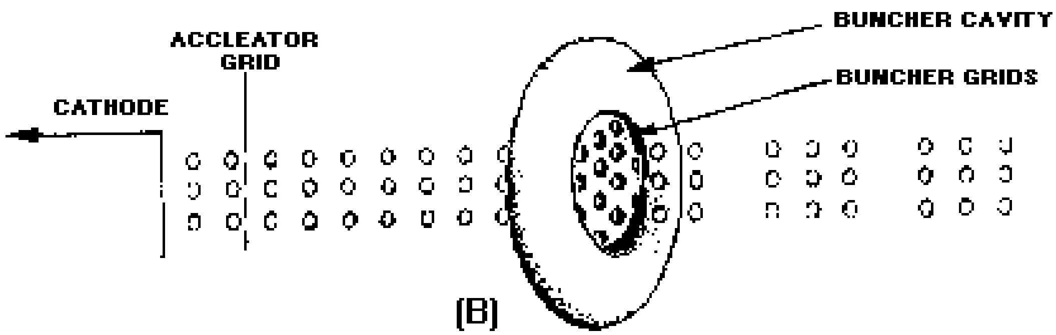


Figure 2-4B.—Electron gun with buncher grids.

The manner in which the buncher produces bunches of electrons is better understood by considering the motions of individual electrons, as illustrated in figure 2-5A.

When the voltage across the grids is negative, as shown in figure 2-5B, electron 1 crossing the gap at that time is slowed. Figure 2-5C shows the potential across the gap at 0 volts; electron 2 is not affected. Electron 3 enters the gap (figure 2-5D) when the voltage across the gap is positive and its velocity is increased. The combined effect is shown in figure 2-5E. All of the electrons in the group have been bunched closer together.

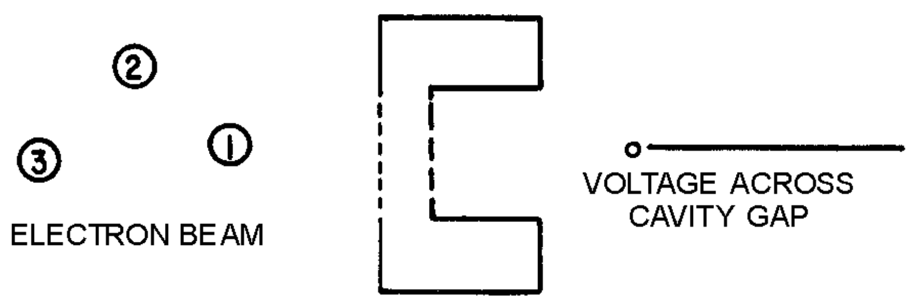


Figure 2-5A.—Buncher cavity action. BUNCHER CAVITY.

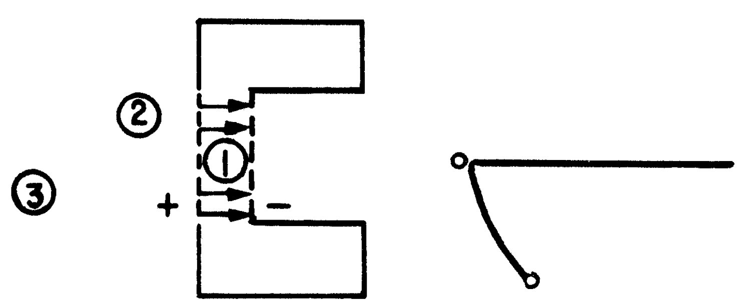


Figure 2-5B.—Buncher cavity action. ELECTRON #1 DECELERATED.

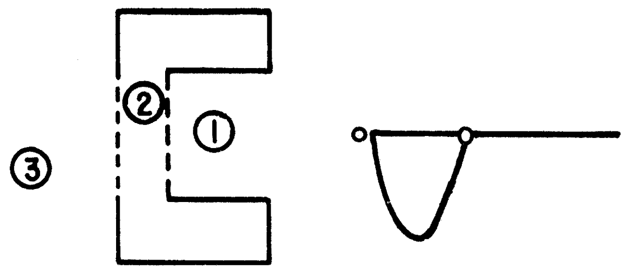


Figure 2-5C.—Buncher cavity action. ELECTRON #2 VELOCITY UNCHANGED.

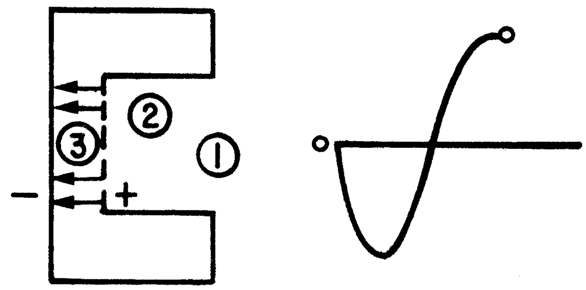


Figure 2-5D.—Buncher cavity action. ELECTRON #3 ACCELERATED.

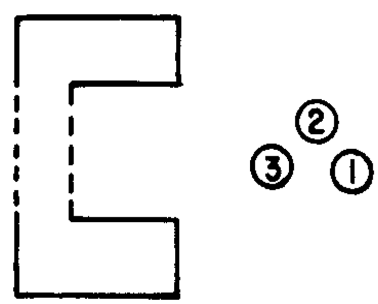


Figure 2-5E.—Buncher cavity action. ELECTRONS BEGINNING TO BUNCH, DUE TO VELOCITY DIFFERENCES.

The velocity modulation of the beam is merely a means to an end. No useful power has been produced at this point. The energy gained by the accelerated electrons is balanced by the energy lost by the decelerated electrons. However, a new and useful beam distribution will be formed if the velocity-modulated electrons are allowed to drift into an area that has no electrostatic field.

As the electrons drift into the field-free area beyond the buncher cavity, bunches continue to form because of the new velocity relationships between the electrons. Unless the beam is acted upon by some other force, these bunches will tend to form and disperse until the original beam distribution is eventually reformed. The net effect of velocity modulation is to form a current-density modulated beam that varies at the same rate as the grid-signal frequency. The next step is to take useful power from the beam.

The current-modulated (bunched) electron beam in figure 2-6A and B is shown in various stages of formation and dispersion. A second cavity, called a CATCHER CAVITY, must be placed at a point of maximum bunching to take useful energy from the beam (shown in figure 2-6B). The physical position of the catcher cavity is determined by the frequency of the buncher-grid signal because this signal determines the transit time of the electron bunches. Note also that both cavities are resonant at the buncher-grid frequency. The electron bunches will induce an rf voltage in the grid gap of the second cavity causing it to oscillate. Proper placement of the second cavity will cause the induced grid-gap voltage to decelerate the electron bunches as they arrive at the gap. Since the largest concentration of electrons is in the bunches, slowing the bunches causes a transfer of energy to the output cavity. The balance of energy has been disturbed because the placement of the catcher cavity is such that bunches are slowed down when they arrive at the cavity. The areas between bunches arrive at the cavity at just the right time. At this time the voltage is of the correct polarity to increase the velocity of the electrons and the beam absorbs energy. The areas between the bunches have very few electrons, so the energy removed from the beam is much greater than the energy required to speed up the electrons between the bunches. Therefore, if the second cavity is properly positioned, useful energy can be removed from a velocity-modulated electron beam.

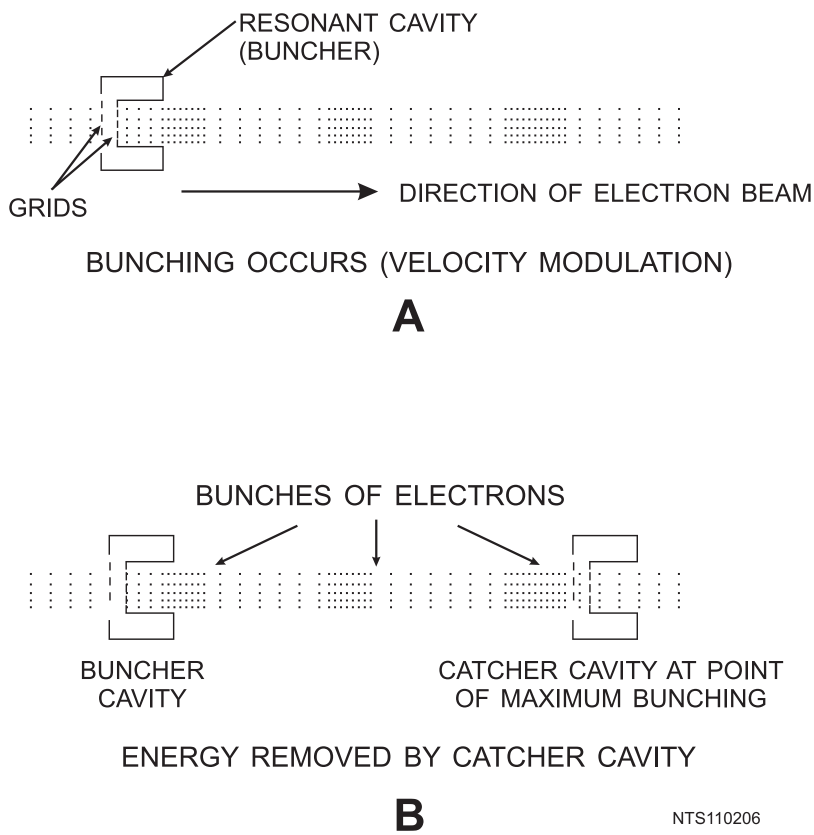


Figure 2-6A-B.—Removing energy from a velocity-modulated beam.

Q-5. The kinetic energy of an electron is directly proportional to what property?

Q-6. What will be the effect upon an electron traveling in the opposite direction to the lines of force in an electrostatic field?

Q-7. How is a beam of electrons velocity-modulated?

Q-8. What portion of an electron gun causes the electrons to accelerate or decelerate?

Q-9. What is the effect upon an electron that enters the buncher gap when the potential across the grids is at 0 volts?

Q-10. What determines the placement of the catcher cavity?

MICROWAVE TUBES

Microwave tubes perform the same functions of generation and amplification in the microwave portion of the frequency spectrum that vacuum tubes perform at lower frequencies. This section will explain the basic operation of the most widely used microwave tubes, including klystrons, traveling-wave tubes, backward-wave oscillators, magnetrons, and crossed-field amplifiers. The variations of these tubes for use in specific applications are so numerous that all of them cannot be discussed in this module. However, general principles of operation are similar in all of the variations so the explanations will be restricted to the general principles of operation.

The Basic Two-Cavity Klystron

Klystrons are velocity-modulated tubes that are used in radar and communications equipment as oscillators and amplifiers. Klystrons make use of the transit-time effect by varying the velocity of an electron beam in much the same manner as the previously discussed velocity-modulation process. Strong electrostatic fields are necessary in the klystron for efficient operation. This is necessary because the interaction of the signal and the electron beam takes place in a very short distance.

The construction and essential components of a TWO-CAVITY KLYSTRON are shown in figure 2-7A. Figure 2-7B is a schematic representation of the same tube. When the tube is energized, the cathode emits electrons which are focused into a beam by a low positive voltage on the control grid. The beam is then accelerated by a very high positive dc potential that is applied in equal amplitude to both the accelerator grid and the buncher grids. The buncher grids are connected to a cavity resonator that superimposes an ac potential on the dc voltage. Ac potentials are produced by oscillations within the cavity that begin spontaneously when the tube is energized. The initial oscillations are caused by random fields and circuit imbalances that are present when the circuit is energized. The oscillations within the cavity produce an oscillating electrostatic field between the buncher grids that is at the same frequency as the natural frequency of the cavity. The direction of the field changes with the frequency of the cavity. These changes alternately accelerate and decelerate the electrons of the beam passing through the grids. The area beyond the buncher grids is called the DRIFT SPACE. The electrons form bunches in this area when the accelerated electrons overtake the decelerated electrons.

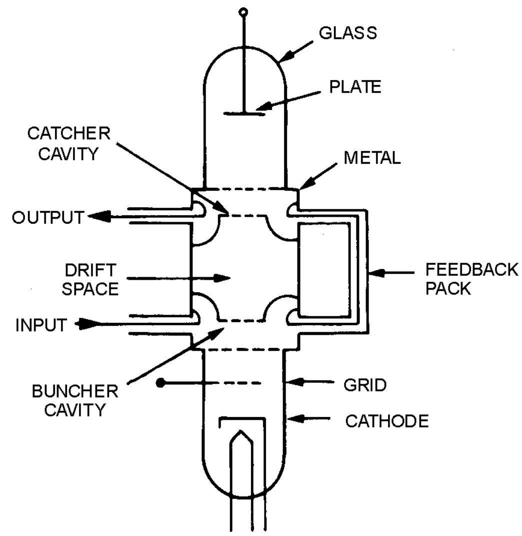


Figure 2-7A.—Functional and schematic diagram of a two-cavity klystron.

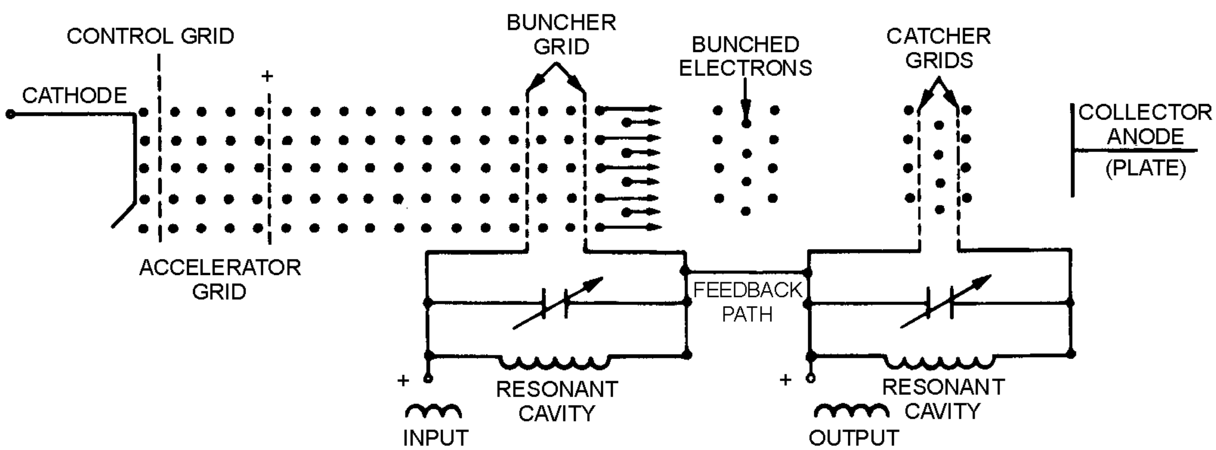


Figure 2-7B.—Functional and schematic diagram of a two-cavity klystron.

The function of the CATCHER GRIDS is to absorb energy from the electron beam. The catcher grids are placed along the beam at a point where the bunches are fully formed. The location is determined by the transit time of the bunches at the natural resonant frequency of the cavities (the resonant frequency of the catcher cavity is the same as the buncher cavity). The location is chosen because maximum energy transfer to the output (catcher) cavity occurs when the electrostatic field is of the correct polarity to slow down the electron bunches.

The two-cavity klystron in figure 2-7A and B may be used either as an oscillator or an amplifier. The configuration shown in the figure is correct for oscillator operation. The feedback path provides energy of the proper delay and phase relationship to sustain oscillations. A signal applied at the buncher grids will be amplified if the feedback path is removed.

Q-11. What is the basic principle of operation of a klystron?

Q-12. The electrons in the beam of a klystron are speeded up by a high dc potential applied to what elements?

Q-13. The two-cavity klystron uses what cavity as an output cavity?

Q-14. A two-cavity klystron without a feedback path will operate as what type of circuit?

The Multicavity Power Klystron

Klystron amplification, power output, and efficiency can be greatly improved by the addition of intermediate cavities between the input and output cavities of the basic klystron. Additional cavities serve to velocity-modulate the electron beam and produce an increase in the energy available at the output. Since all intermediate cavities in a multicavity klystron operate in the same manner, a representative THREE-CAVITY KLYSTRON will be discussed.

A three-cavity klystron is illustrated in figure 2-8. The entire DRIFT-TUBE ASSEMBLY, the three CAVITIES, and the COLLECTOR PLATE of the three-cavity klystron are operated at ground potential for reasons of safety. The electron beam is formed and accelerated toward the drift tube by a large negative pulse applied to the cathode. MAGNETIC FOCUS COILS are placed around the drift tube to keep the electrons in a tight beam and away from the side walls of the tube. The focus of the beam is also aided by the concave shape of the cathode in high-powered klystrons.

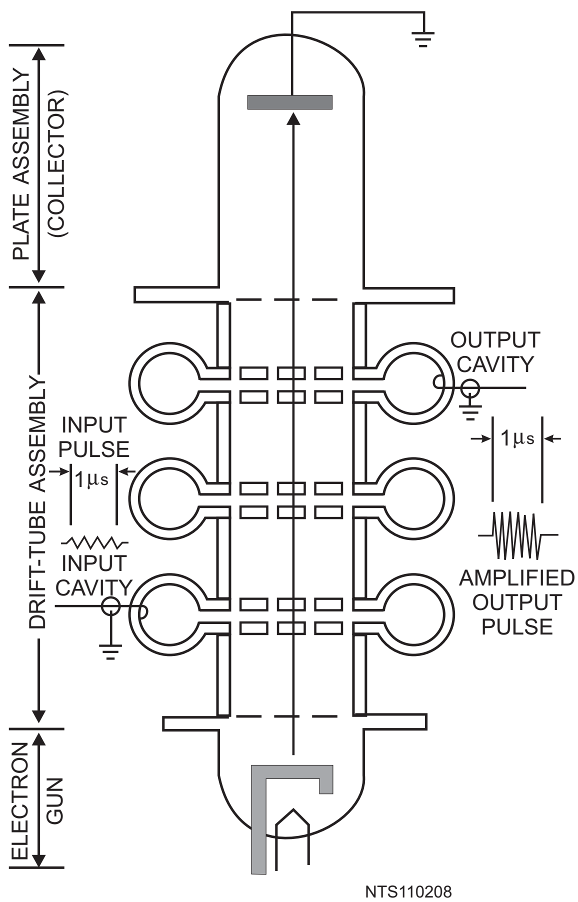


Figure 2-8.—Three-cavity klystron.

The output of any klystron (regardless of the number of cavities used) is developed by velocity modulation of the electron beam. The electrons that are accelerated by the cathode pulse are acted upon by rf fields developed across the input and middle cavities. Some electrons are accelerated, some are decelerated, and some are unaffected. Electron reaction depends on the amplitude and polarity of the fields across the cavities when the electrons pass the cavity gaps. During the time the electrons are traveling through the drift space between the cavities, the accelerated electrons overtake the decelerated electrons to form bunches. As a result, bunches of electrons arrive at the output cavity at the proper instant during each cycle of the rf field and deliver energy to the output cavity.

Only a small degree of bunching takes place within the electron beam during the interval of travel from the input cavity to the middle cavity. The amount of bunching is sufficient, however, to cause oscillations within the middle cavity and to maintain a large oscillating voltage across the input gap. Most of the velocity modulation produced in the three-cavity klystron is caused by the voltage across the input gap of the middle cavity. The high voltage across the gap causes the bunching process to proceed rapidly in the drift space between the middle cavity and the output cavity. The electron bunches cross the gap of the output cavity when the gap voltage is at maximum negative. Maximum energy transfer from the electron beam to the output cavity occurs under these conditions. The energy given up by the electrons is the kinetic energy that was originally absorbed from the cathode pulse.

Klystron amplifiers have been built with as many as five intermediate cavities in addition to the input and output cavities. The effect of the intermediate cavities is to improve the electron bunching process which improves amplifier gain. The overall efficiency of the tube is also improved to a lesser extent. Adding more cavities is roughly the same as adding more stages to a conventional amplifier. The overall amplifier gain is increased and the overall bandwidth is reduced if all the stages are tuned to the same frequency. The same effect occurs with multicavity klystron tuning. A klystron amplifier tube will deliver high gain and a narrow bandwidth if all the cavities are tuned to the same frequency. This method of tuning is called SYNCHRONOUS TUNING. If the cavities are tuned to slightly different frequencies, the gain of the amplifier will be reduced but the bandwidth will be appreciably increased. This method of tuning is called STAGGERED TUNING.

Q-15. What can be added to the basic two-cavity klystron to increase the amount of velocity modulation and the power output?

Q-16. How is the electron beam of a three-cavity klystron accelerated toward the drift tube?

Q-17. Which cavity of a three-cavity klystron causes most of the velocity modulation?

Q-18. In a multicavity klystron, tuning all the cavities to the same frequency has what effect on the bandwidth of the tube?

Q-19. The cavities of a multicavity klystron are tuned to slightly different frequencies in what method of tuning?

The Reflex Klystron

Another tube based on velocity modulation, and used to generate microwave energy, is the REFLEX KLYSTRON (figure 2-9). The reflex klystron contains a REFLECTOR PLATE, referred to as the REPELLER, instead of the output cavity used in other types of klystrons. The electron beam is modulated as it was in the other types of klystrons by passing it through an oscillating resonant cavity, but here the similarity ends. The feedback required to maintain oscillations within the cavity is obtained by reversing the beam and sending it back through the cavity. The electrons in the beam are velocity-modulated before the beam passes through the cavity the second time and will give up the energy required to maintain oscillations. The electron beam is turned around by a negatively charged electrode that repels the beam. This negative element is the repeller mentioned earlier. This type of klystron oscillator is called a reflex klystron because of the reflex action of the electron beam.

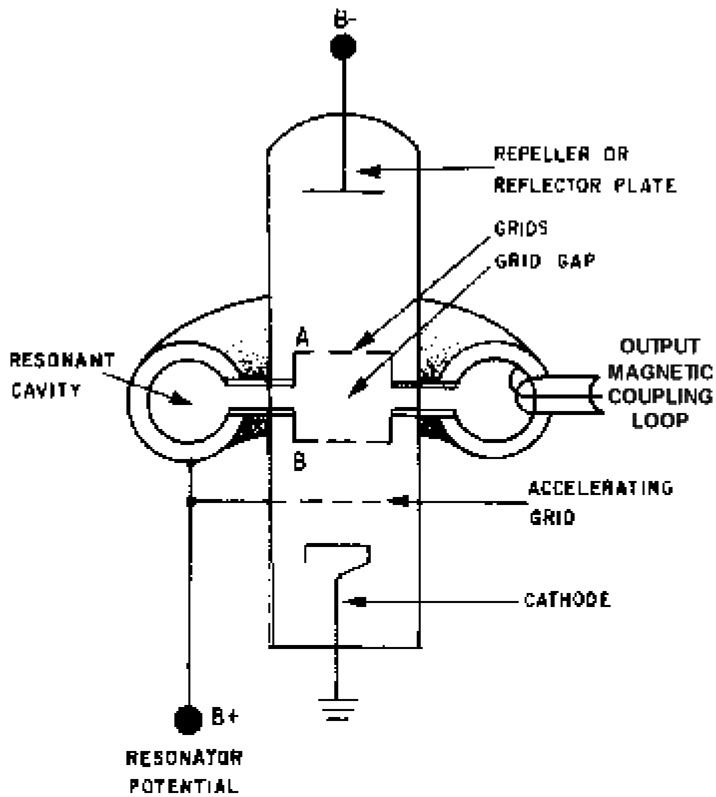


Figure 2-9.—Functional diagram of a reflex klystron.

Three power sources are required for reflex klystron operation: (1) filament power, (2) positive resonator voltage (often referred to as beam voltage) used to accelerate the electrons through the grid gap of the resonant cavity, and (3) negative repeller voltage used to turn the electron beam around. The electrons are focused into a beam by the electrostatic fields set up by the resonator potential (B+) in the body of the tube. Note in figure 2-9 that the resonator potential is common to the resonator cavity, the accelerating grid, and the entire body of the tube.

The resonator potential also causes the resonant cavity to begin oscillating at its natural frequency when the tube is energized. These oscillations cause an electrostatic field across the grid gap of the cavity that changes direction at the frequency of the cavity. The changing electrostatic field affects the electrons in the beam as they pass through the grid gap. Some are accelerated and some are decelerated, depending upon the polarity of the electrostatic field as they pass through the gap. Figure 2-10, view (A), illustrates the three possible ways an electron can be affected as it passes through the gap (velocity increasing, decreasing, or remaining constant). Since the resonant cavity is oscillating, the grid potential is an alternating voltage that causes the electrostatic field between the grids to follow a sine-wave curve as shown in figure 2-10, view (B). As a result, the velocity of the electrons passing through the gap is affected uniformly as a function of that sine wave. The amount of velocity change is dependent on the strength and polarity of the grid voltage.

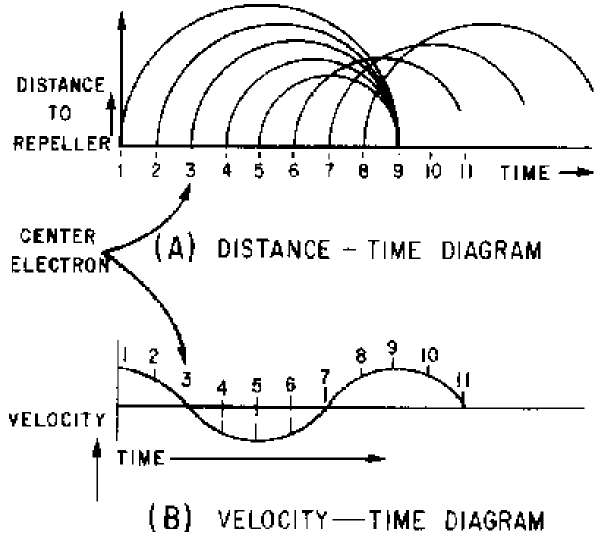


Figure 2-10.—Electron bunching diagram.

The variation in grid voltage causes the electrons to enter the space between the grid and the repeller at various velocities. For example, in figure 2-10, views (A) and (B), the electrons at times 1 and 2 are speeded up as they pass through the grid. At time 3, the field is passing through zero and the electron is unaffected. At times 4 and 5, the grid field is reversed; the electrons give up energy because their velocity is reduced as they pass through the grids.

The distance the electrons travel in the space separating the grid and the repeller depends upon their velocity. Those moving at slower velocities, such as the electron at time 4, move only a short distance from the grid before being affected by the repeller voltage. When this happens, the electron is forced by the repeller voltage to stop, reverse direction, and return toward the grid. The electrons moving at higher velocities travel further beyond the grid before reversing direction because they have greater momentum. If the repeller voltage is set at the correct value, the electrons will form a bunch around the constant-speed electrons. The electrons will then return to the grid gap at the instant the electrostatic field is at the correct polarity to cause maximum deceleration of the bunch. This action is also illustrated in figure 2-10, view (A). When the grid field provides maximum deceleration, the returning electrons release maximum energy to the grid field which is in phase with cavity current. Thus, the returning electrons supply the regenerative feedback required to maintain cavity oscillations.

The constant-speed electrons must remain in the reflecting field space for a minimum time of 3/4 cycle of the grid field for maximum energy transfer. The period of time the electrons remain in the repeller field is determined by the amount of negative repeller voltage. The reflex klystron will continue to oscillate if the electrons remain in the repeller field longer than 3/4 cycle (as long as the electrons return to the grid gap when the field is of the proper polarity to decelerate the electrons). Figure 2-11 shows the effect of the repeller field on the electron bunch for 3/4 cycle and for 1 3/4 cycles. Although not shown in the figure, the constant-velocity electrons may remain in the repeller field for any number of cycles over the minimum 3/4 cycle. If the electrons remain in the field for longer than 3/4 cycle, the difference in electron transit time causes the tube performance characteristics to change. The differences in operating characteristics are identified by MODES OF OPERATION.

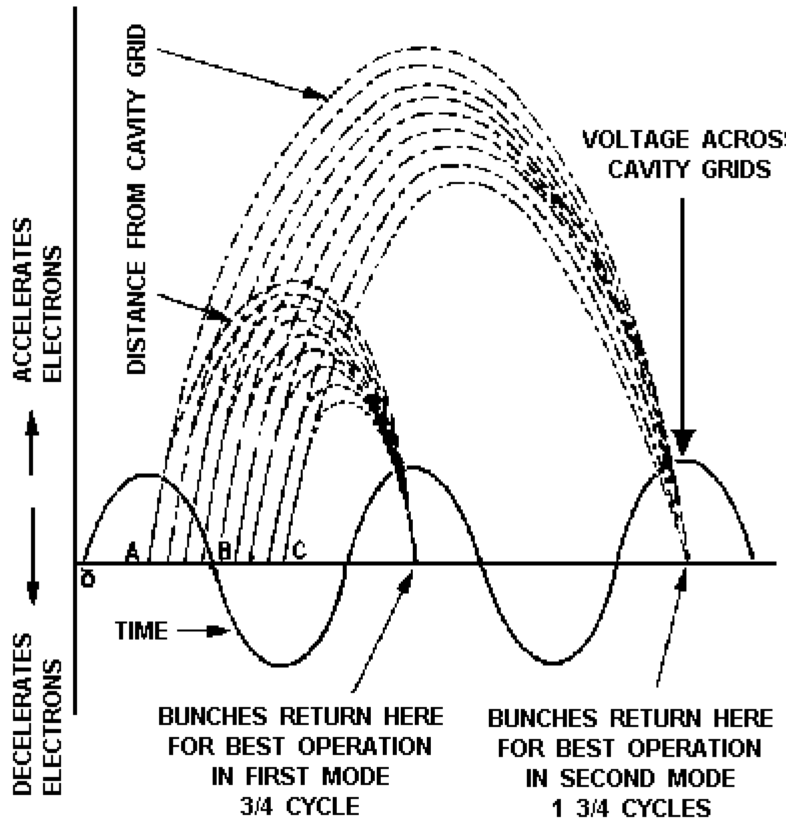


Figure 2-11.—Bunching action of a reflex klystron.

The reflex klystron operates in a different mode for each additional cycle that the electrons remain in the repeller field. Mode 1 is obtained when the repeller voltage produces an electron transit time of 3/4 cycle. Additional modes follow in sequence. Mode 2 has an electron transit time of 1 3/4 cycles; mode 3 has an electron transit time of 2 3/4 cycles; etc. The physical design of the tube limits the number of modes possible in practical applications. A range of four modes of operation are normally available. The actual mode used (1 3/4 cycles through 4 3/4 cycles, 2 3/4 cycles through 6 3/4 cycles, etc.) depends upon the application. The choice of mode is determined by the difference in power available from each mode and the band of frequencies over which the circuit can be tuned.

OUTPUT POWER.—The variation in output power for different modes of operation can be explained by examining the factors which limit the amplitude of oscillations. Power and amplitude limitations are caused by the DEBUNCHING process of the electrons in the repeller field space. Debunching is simply the spreading out of the electron bunches before they reach electrostatic fields across the cavity grid . The lower concentration of electrons in the returning bunches provides less power for delivery to the oscillating cavity. This reduced power from the bunches, in turn, reduces the amplitude of the cavity oscillations and causes a decrease in output power. In higher modes of operation the electron bunches are formed more slowly. They are more likely to be affected by debunching because of mutual repulsion between the negatively charged electrons. The long drift time in the higher modes allows more time for this electron interaction and, as a result, the effects of debunching are more severe. The mutual repulsion changes the relative velocity between the electrons in the bunches and causes the bunches to spread out.

Figure 2-12 illustrates the ELECTRONIC TUNING (tuning by altering the repeller voltage) range and output power of a reflex klystron. Each mode has a center frequency of 3,000 megahertz which is predetermined by the physical size of the cavity. The output power increases as the repeller voltage is made more negative. This is because the transit time of the electron bunches is decreased.

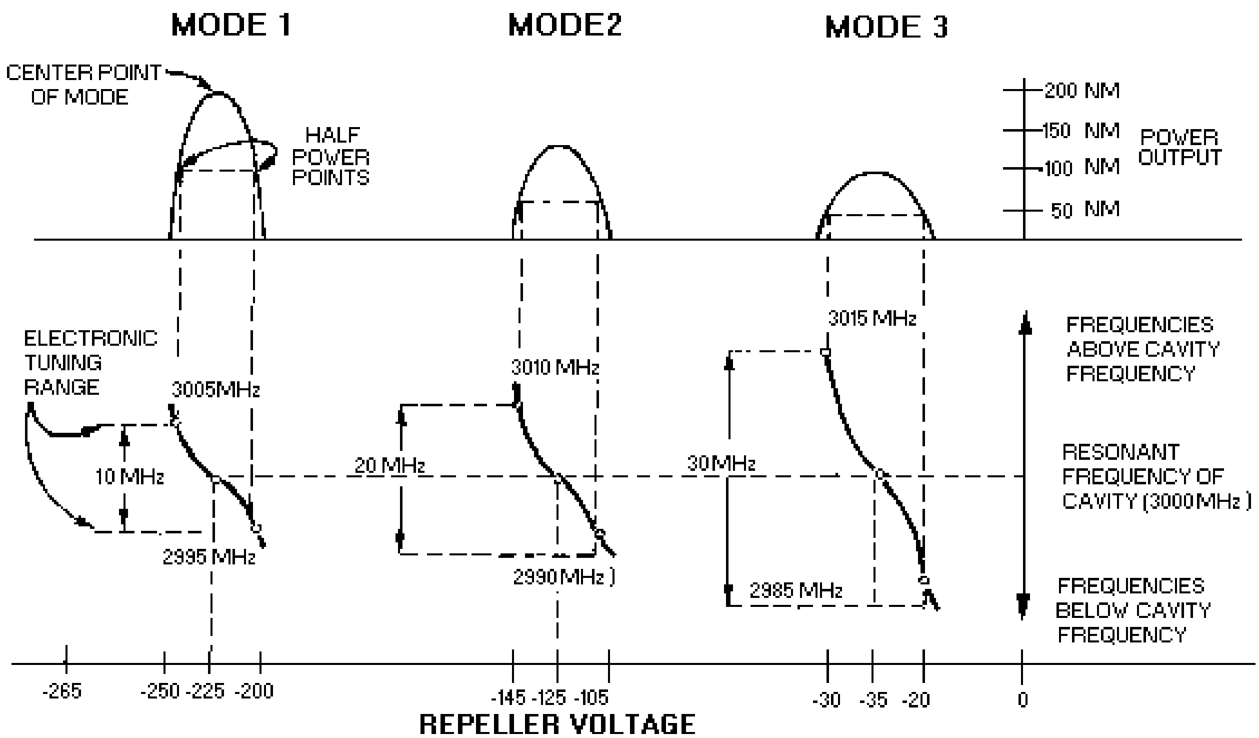


Figure 2-12.—Electronic tuning and output power of a reflex klystron.

Electronic tuning does not change the center frequency of the cavity, but does vary the frequency within the mode of operation. The amount the frequency can be varied above or below the center frequency is limited by the half-power points of the mode, as shown in figure 2-12. The center frequency can be changed by one of two methods One method, GRID-GAP TUNING, varies the cavity frequency by altering the distance between the grids to change the physical size of the cavity. This method varies the capacitance of the cavity by using a tuning screw to change the distance between the grids mechanically. The cavity can also be tuned by PADDLES or SLUGS that change the inductance of the cavity.

Q-20. What element of the reflex klystron replaces the output cavity of a normal klystron?

Q-21. When the repealer potential is constant, what property of the electron determines how long it will remain in the drift space of the reflex klystron?

Q-22. The constant-speed electrons of an electron bunch in a reflex klystron must remain in the repeller field for what minimum time?

Q-23. If the constant-speed electrons in a reflex klystron remain in the repeller field for 1 3/4 cycles, what is the mode of operation?

Q-24. Debunching of the electron bunches in the higher modes of a reflex klystron has what effect on output power?

Q-25. What limits the tuning range around the center frequency of a reflex klystron in a particular mode of operation?

The Decibel Measurement System

Because of the use of the decibel measurement system in the following paragraphs, you will be introduced to it at this point. Technicians who deal with communications and radar equipment most often speak of the gain of an amplifier or a system in terms of units called DECIBELS (dB). Throughout your Navy career you will use decibels as an indicator of equipment performance; therefore, you need to have a basic understanding of the decibel system of measurement. Because the actual calculation of decibel measurements is seldom required in practical applications, the explanation given in this module is somewhat simplified. Most modern test equipment is designed to measure and indicate decibels directly which eliminates the need for complicated mathematical calculations. Nevertheless, a basic explanation of the decibel measurement system is necessary for you to understand the significance of dB readings and equipment gain ratings which are expressed in decibels.

The basic unit of measurement in the system is not the decibel, but the bel, named in honor of the American inventor, Alexander Graham Bell. The bel is a unit that expresses the logarithmic ratio between the input and output of any given component, circuit, or system and may be expressed in terms of voltage, current, or power. Most often it is used to show the ratio between input and output power. The formula is as follows:

$$N = \log_{10} \frac{P1}{P2} \text{bel}$$

The gain of an amplifier can be expressed in bels by dividing the output (P1) by the input (P2) and taking the base 10 logarithm of the resulting quotient. Thus, if an amplifier doubles the power, the quotient will be 2. If you consult a logarithm table, you will find that the base 10 logarithm of 2 is 0.3; so the power gain of the amplifier is 0.3 bel. Experience has taught that because the bel is a rather large unit, it is difficult to apply. A more practical unit that can be applied more easily is the decibel (1/10 bel). Any figure expressed in bels can easily be converted to decibels by multiplying the figure by 10 or simply by moving the decimal one place to the right. The previously found ratio of 0.3 is therefore equal to 3 decibels.

The reason for using the decibel system when expressing signal strength may be seen in the power ratios in table 2-1. For example, to say that a reference signal has increased 50 dB is much easier than to say the output has increased 100,000 times. The amount of increase or decrease from a chosen reference level is the basis of the decibel measurement system, not the reference level itself. Whether the input power is increased from 1 watt to 100 watts or from 1,000 watts to 100,000 watts, the amount of increase is still 20 decibels.

Table 2-1.—Decibel Power Ratios

Source Level (dB)		Power Ratio
1	=	1.3
3	=	2.0
5	=	3.2
6	=	4.0
7	=	5.0
10	=	10 = 10_1
20	=	100 = 10_2
30	=	1000 = 10_3
40	=	10,000 = 10_4
50	=	100,000 = 10_5
60	=	1,000,000 = 10_6
70	=	10,000,000 = 10_7
100	=	10_{10}
110	=	10_{11}
140	=	10_{14}

Examine table 2-1 again, and take particular note of the power ratios for source levels of 3 dB and 6 dB. As the table illustrates, an increase of 3 dB represents a doubling of power. The reverse is also true. If a signal decreases by 3 dB, half the power is lost. For example, a 1,000 watt signal decreased by 3 dB will equal 500 watts while a 1,000 watt signal increased by 3 dB equals 2,000 watts.

The attenuator is a widely used piece of test equipment that can be used to demonstrate the importance of the decibel as a unit of measurement. Attenuators are used to reduce a signal to a smaller level for use or measurement. Most attenuators are rated by the number of decibels the signal is reduced. The technician's job is to know the relationship between the dB rating and the power reduction it represents. This is so important, in fact, that every student of electronics should memorize the relationships in table 2-1 through the 60 dB range. The technician will have to apply this knowledge to prevent damage to valuable equipment. A helpful hint is to note that the first digit of the source level (on the chart) is the same number as the corresponding power of 10 exponent; i.e., 40 dB = 1×10^4 or 10,000. A 20 dB attenuator, for example, will reduce an input signal by a factor of 100. In other words, a 100-milliwatt signal will be reduced to 1 milliwatt. A 30 dB attenuator will reduce the same 100-milliwatt signal by a factor of 1,000 and produce an output of 0.1 milliwatt. When an attenuator of the required size is not available, attenuators of several smaller sizes may be added directly together to reach the desired amount of attenuation. A 10 dB attenuator and a 20 dB attenuator add directly to equal 30 dB of attenuation. The same relationship exists with amplifier stages as well. If an amplifier has two stages rated at 10 dB each, the total amplifier gain will be 20 dB.

When you speak of the dB level of a signal, you are really speaking of a logarithmic comparison between the input and output signals. The input signal is normally used as the reference level. However, the application sometimes requires the use of a standard reference signal. The most widely used reference level is a 1-milliwatt signal. The standard decibel abbreviation of dB is changed to dBm to indicate the use of the 1-milliwatt standard reference. Thus, a signal level of +3 dBm is 3 dB above 1 milliwatt, and a signal level of −3 dBm is 3 dB below 1 milliwatt. Whether using dB or dBm, a plus (+) sign (or no sign at all) indicates the output signal is larger than the reference; a minus (−) sign indicates the output signal is less than the reference.

The Navy student of electronics will encounter the dBm system of measurement most often as a figure indicating the receiver sensitivity of radar or communications equipment. Typically, a radar receiver will be rated at approximately −107 dBm, which means the receiver will detect a signal 107 dB below 1 milliwatt.

The importance of understanding the decibel system of measurement can easily be seen in the case of receiver-sensitivity measurements. At first glance a loss of 3 dBm from a number as large as −107 dBm seems insignificant; however, it becomes extremely important when the number indicates receiver sensitivity in the decibel system. When the sensitivity falls to −104 dBm, the receiver will only detect a signal that is twice as large as a signal at −107 dBm.

The Traveling-Wave Tube

The TRAVELING-WAVE TUBE (twt) is a high-gain, low-noise, wide-bandwidth microwave amplifier. It is capable of gains greater than 40 dB with bandwidths exceeding an octave. (A bandwidth of 1 octave is one in which the upper frequency is twice the lower frequency.) Traveling-wave tubes have been designed for frequencies as low as 300 megahertz and as high as 50 gigahertz. The twt is primarily a voltage amplifier. The wide-bandwidth and low-noise characteristics make the twt ideal for use as an rf amplifier in microwave equipment.

The physical construction of a typical twt is shown in figure 2-13. The twt contains an electron gun which produces and then accelerates an electron beam along the axis of the tube. The surrounding magnet provides a magnetic field along the axis of the tube to focus the electrons into a tight beam. The HELIX, at the center of the tube, is a coiled wire that provides a low-impedance transmission line for the rf energy within the tube. The rf input and output are coupled onto and removed from the helix by directional couplers that have no physical connection to the helix. If the rf energy is transported on coaxial cables, the coaxial couplers are wound in a helical manner similar to that shown in figure 2-13. If the rf energy is transported in waveguides, waveguide directional couplers are used. The attenuator prevents any reflected waves from traveling back down the helix.

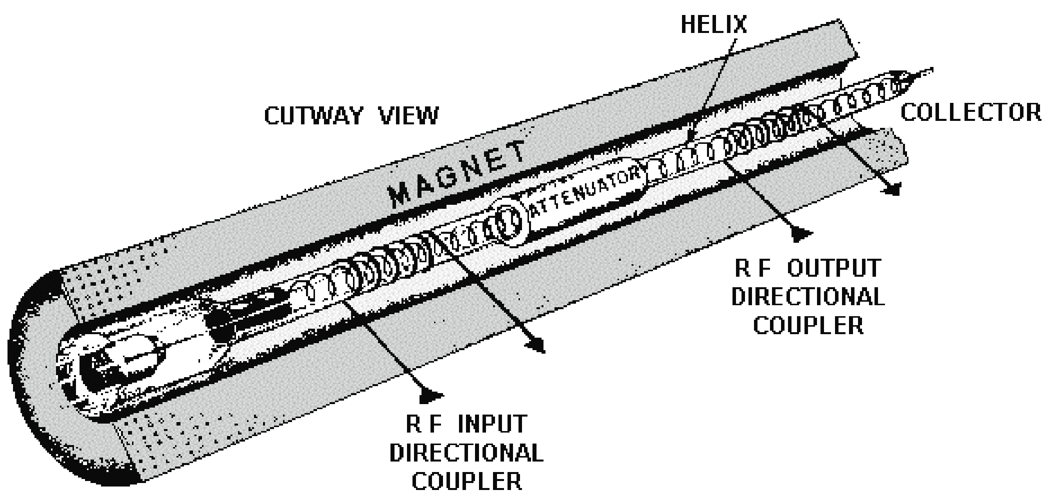


Figure 2-13.—Physical construction of a twt.

A simplified version of twt operation is shown in figure 2-14. In the figure, an electron beam is passing along a nonresonant transmission line represented by a straight wire. The input to the transmission line is an rf wave which travels on the line from input to output. The line will transport a wide range of rf frequencies if it is terminated in the characteristic impedance of the line. The electromagnetic waves traveling down the line produce electric fields that interact with the electrons of the beam.

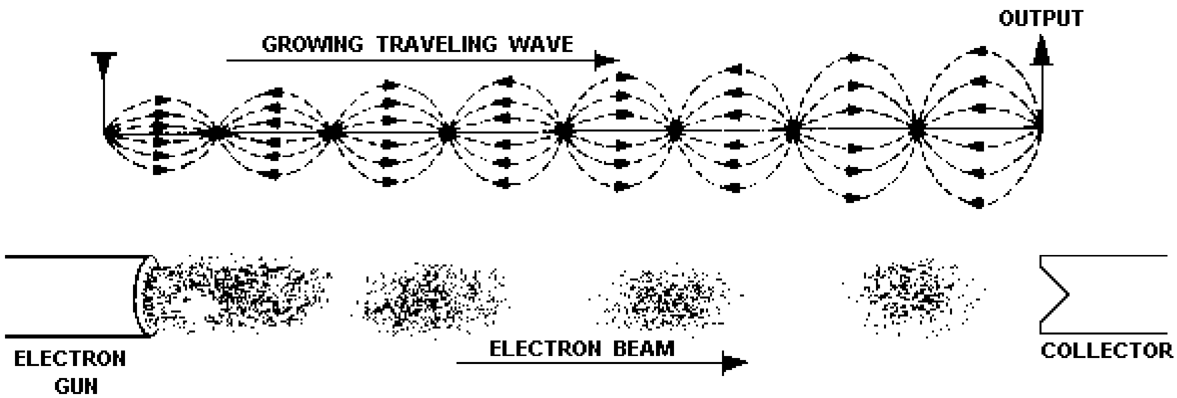


Figure 2-14.—Simplified twt.

If the electrons of the beam were accelerated to travel faster than the waves traveling on the wire, bunching would occur through the effect of velocity modulation. Velocity modulation would be caused by the interaction between the traveling-wave fields and the electron beam. Bunching would cause the electrons to give up energy to the traveling wave if the fields were of the correct polarity to slow down the bunches. The energy from the bunches would increase the amplitude of the traveling wave in a progressive action that would take place all along the length of the twt, as shown in figure 2-14.

However, because the waves travel along the wire at the speed of light, the simple twt shown in figure 2-14 will not work. At present no way is known to accelerate an electron beam to the speed of light. Since the electron beam cannot travel faster than the wave on the wire, bunching will not take place and the tube will not work. The twt is therefore designed with a delay structure to slow the traveling wave down to or below the speed of the electrons in the beam. A common twt delay structure is a wire, wound in the form of a long coil or helix, as shown in figure 2-15, view (A). The shape of the helix slows the effective velocity of the wave along the common axis of the helix and the tube to about one-tenth the speed of light. The wave still travels down the helix wire at the speed of light, but the coiled shape causes the wave to travel a much greater total distance than the electron beam. The speed at which the wave travels down the tube can be varied by changing the number of turns or the diameter of the turns in the helix wire. The helical delay structure works well because it has the added advantage of causing a large proportion of electric fields that are parallel to the electron beam. The parallel fields provide maximum interaction between the fields and the electron beam.

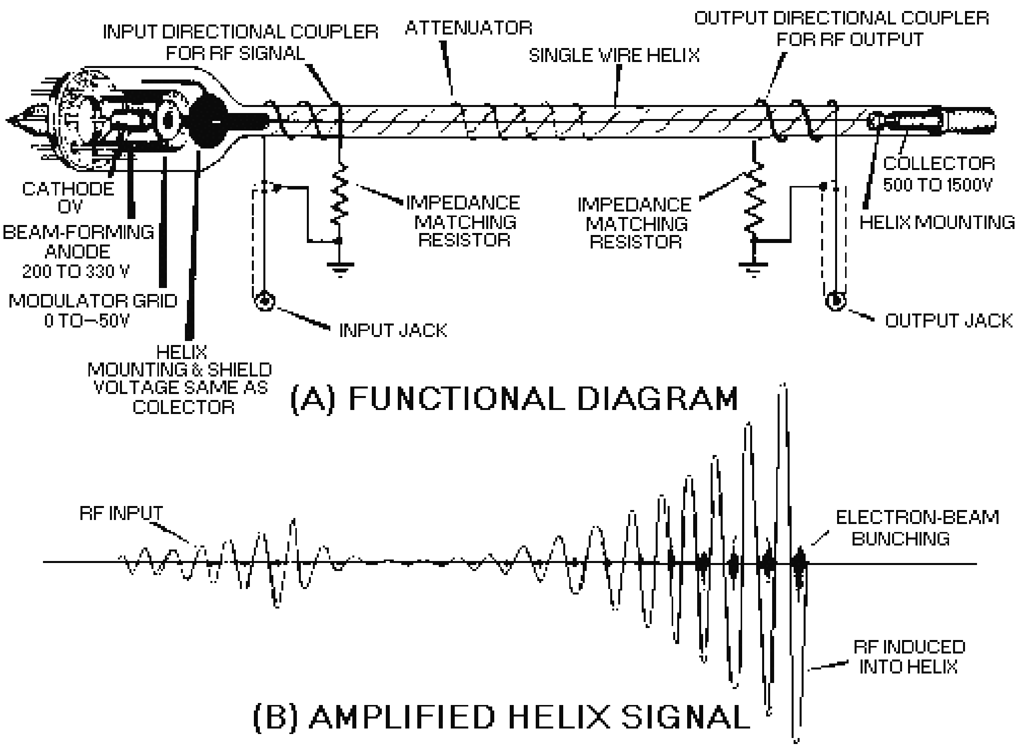


Figure 2-15.—Functional diagram of a twt.

In a typical twt, the electron beam is directed down the center of the helix while, at the same time, an rf signal is coupled onto the helix. The electrons of the beam are velocity-modulated by the electric fields produced by the rf signal.

Amplification begins as the electron bunches form and release energy to the signal on the helix. The slightly amplified signal causes a denser electron bunch which, in turn, amplifies the signal even more. The amplification process is continuous as the rf wave and the electron beam travel down the length of the tube.

Any portion of the twt output signal that reflects back to the input will cause oscillations within the tube which results in a decrease in amplification. Attenuators are placed along the length of the helix to prevent reflections from reaching the input. The attenuator causes a loss in amplitude, as can be seen in figure 2-15, view (B), but it can be placed so as to minimize losses while still isolating the input from the output.

The relatively low efficiency of the twt partially offsets the advantages of high gain and wide bandwidth. The internal attenuator reduces the gain of the tube, and the power required to energize the focusing magnet is an operational loss that cannot be recovered. The twt also produces heat which must be dissipated by either air-conditioning or liquid-cooling systems. All of these factors reduce the overall efficiency of the twt, but the advantages of high gain and wide bandwidth are usually enough to overcome the disadvantages.

The Backward-Wave Oscillator

The BACKWARD-WAVE OSCILLATOR (bwo) is a microwave-frequency, velocity-modulated tube that operates on the same principle as the twt. However, a traveling wave that moves from the electron gun end of the tube toward the collector is not used in the bwo. Instead, the bwo extracts energy from the electron. beam by using a backward wave that travels from the collector toward the electron gun (cathode). Otherwise, the electron bunching action and energy extraction from the electron beam is very similar to the actions in a twt.

The typical bwo is constructed from a folded transmission line or waveguide that winds back and forth across the path of the electron beam, as shown in figure 2-16. The folded waveguide in the illustration serves the same purpose as the helix in a twt. The fixed spacing of the folded waveguide limits the bandwidth of the bwo. Since the frequency of a given waveguide is constant, the frequency of the bwo is controlled by the transit time of the electron beam. The transit time is controlled by the collector potential. Thus, the output frequency can be changed by varying the collector voltage, which is a definite advantage. As in the twt, the electron beam in the bwo is focused by a magnet placed around the body of the tube.

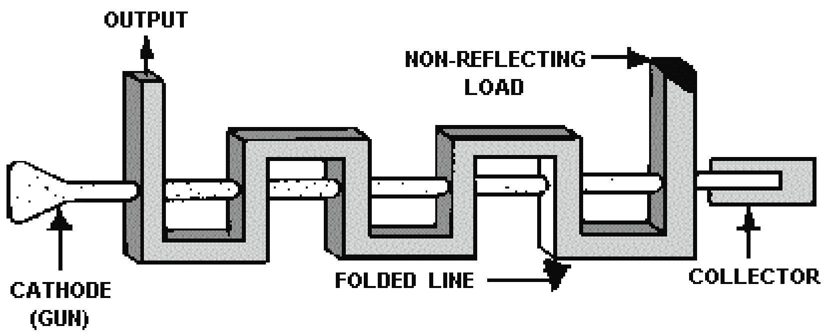


Figure 2-16.—Typical bwo.

Q-26. What is the primary use of the twt?

Q-27. The magnet surrounding the body of a twt serves what purpose?

Q-28. How are the input and output directional couplers in a twt connected to the helix?

Q-29. What relationship must exist between the electron beam and the traveling wave for bunching to occur in the electron beam of a twt?

Q-30. What structure in the twt delays the forward progress of the traveling wave?

The Magnetron

The MAGNETRON, shown in figure 2-17A, is a self-contained microwave oscillator that operates differently from the linear-beam tubes, such as the twt and the klystron. Figure 2-17B is a simplified drawing of the magnetron. CROSSED-ELECTRON and MAGNETIC fields are used in the magnetron to produce the high-power output required in radar and communications equipment.

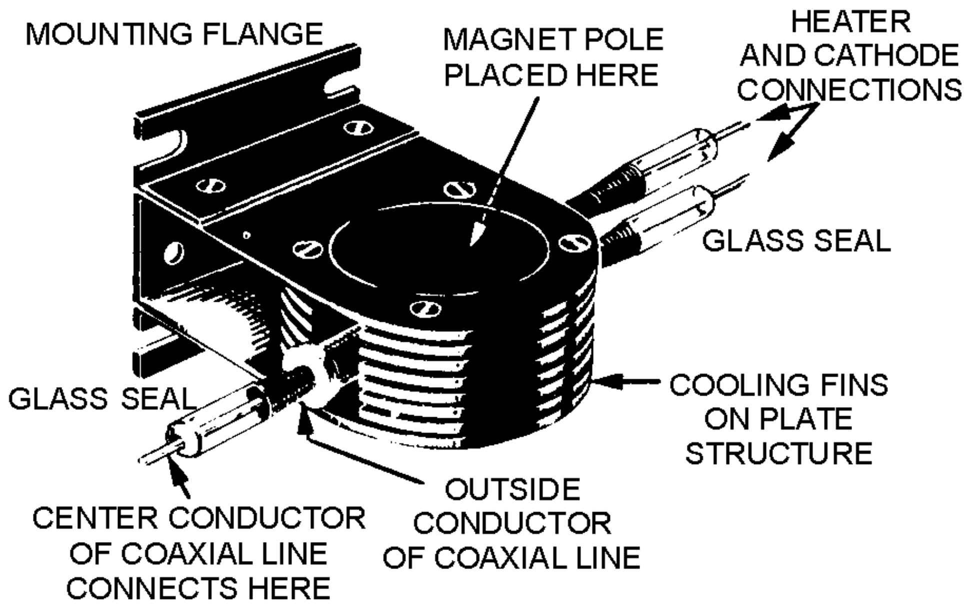


Figure 2-17A.—Magnetron.

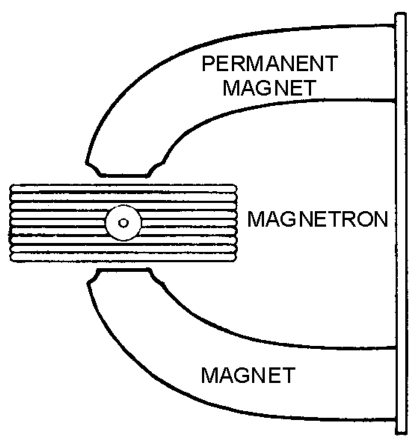


Figure 2-17B.—Magnetron.

The magnetron is classed as a diode because it has no grid. A magnetic field located in the space between the plate (anode) and the cathode serves as a grid. The plate of a magnetron does not have the same physical appearance as the plate of an ordinary electron tube. Since conventional inductive-capacitive (LC) networks become impractical at microwave frequencies, the plate is fabricated into a cylindrical copper block containing resonant cavities which serve as tuned circuits. The magnetron base differs considerably from the conventional tube base. The magnetron base is short in length and has large diameter leads that are carefully sealed into the tube and shielded.

The cathode and filament are at the center of the tube and are supported by the filament leads. The filament leads are large and rigid enough to keep the cathode and filament structure fixed in position. The output lead is usually a probe or loop extending into one of the tuned cavities and coupled into a waveguide or coaxial line. The plate structure, shown in figure 2-18, is a solid block of copper. The cylindrical holes around its circumference are resonant cavities. A narrow slot runs from each cavity into the central portion of the tube dividing the inner structure into as many segments as there are cavities. Alternate segments are strapped together to put the cavities in parallel with regard to the output. The cavities control the output frequency. The straps are circular, metal bands that are placed across the top of the block at the entrance slots to the cavities. Since the cathode must operate at high power, it must be fairly large and must also be able to withstand high operating temperatures. It must also have good emission characteristics, particularly under return bombardment by the electrons. This is because most of the output power is provided by the large number of electrons that are emitted when high-velocity electrons return to strike the cathode. The cathode is indirectly heated and is constructed of a high-emission material. The open space between the plate and the cathode is called the INTERACTION SPACE. In this space the electric and magnetic fields interact to exert force upon the electrons.

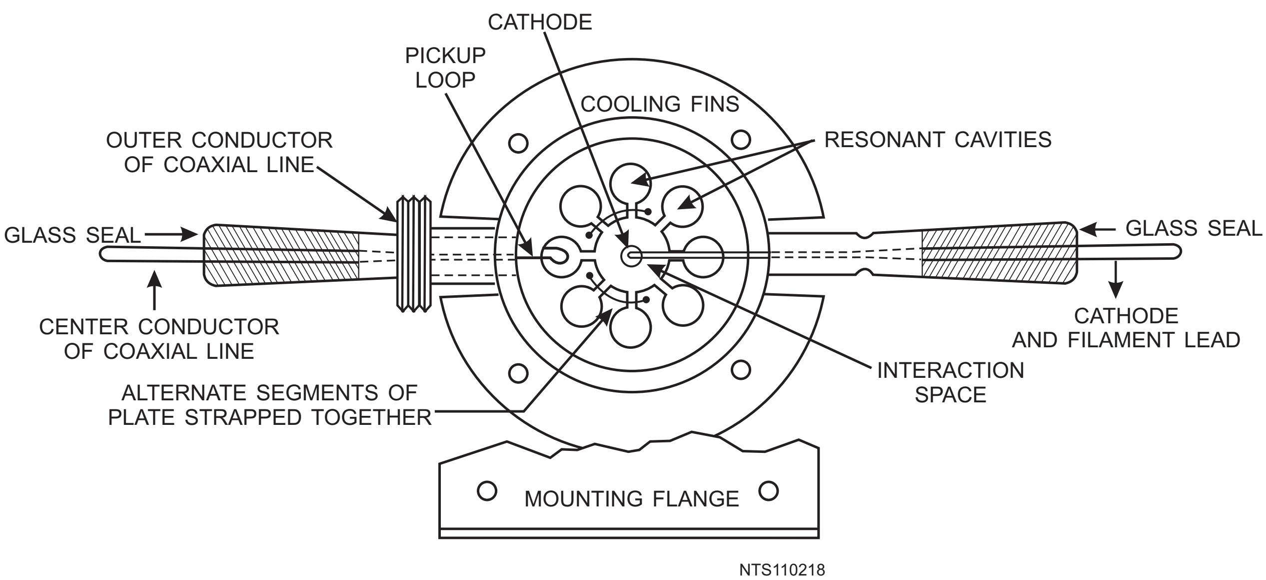


Figure 2-18.—Cutaway view of a magnetron.

The magnetic field is usually provided by a strong, permanent magnet mounted around the magnetron so that the magnetic field is parallel with the axis of the cathode. The cathode is mounted in the center of the interaction space.

BASIC MAGNETRON OPERATION.—Magnetron theory of operation is based on the motion of electrons under the influence of combined electric and magnetic fields. The following information presents the laws governing this motion.

The direction of an electric field is from the positive electrode to the negative electrode. The law governing the motion of an electron in an electric field (E field) states:

The force exerted by an electric field on an electron is proportional to the strength of the field. Electrons tend to move from a point of negative potential toward a positive potential.

This is shown in figure 2-19. In other words, electrons tend to move against the E field. When an electron is being accelerated by an E field, as shown in figure 2-19, energy is taken from the field by the electron.

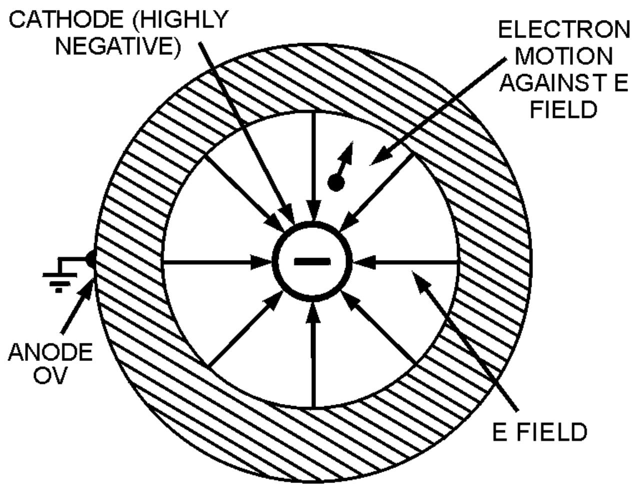


Figure 2-19.—Electron motion in an electric field.

The law of motion of an electron in a magnetic field (H field) states:

The force exerted on an electron in a magnetic field is at right angles to both the field and the path of the electron. The direction of the force is such that the electron trajectories are clockwise when viewed in the direction of the magnetic field.

This is shown in figure 2-20.

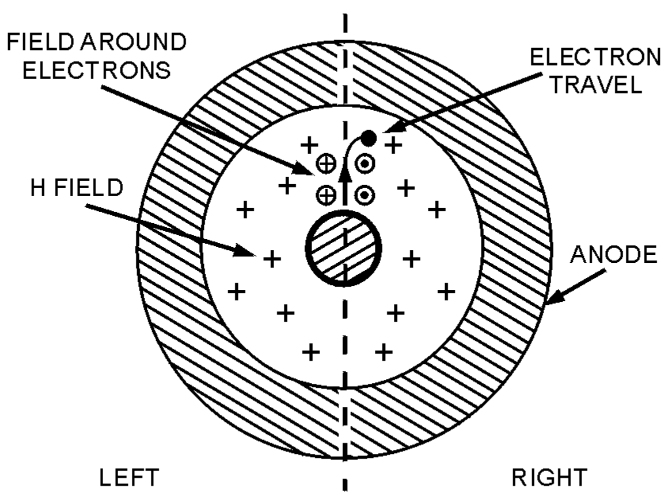


Figure 2-20.—Electron motion in a magnetic field.

In figure 2-20, assume that a south pole is below the figure and a north pole is above the figure so that the magnetic field is going into the paper. When an electron is moving through space, a magnetic field builds around the electron just as it would around a wire when electrons are flowing through a wire. In figure 2-20 the magnetic field around the moving electron adds to the permanent magnetic field on the left side of the electron's path and subtracts from the permanent magnetic field on the right side. This action weakens the field on the right side; therefore, the electron path bends to the right (clockwise). If the strength of the magnetic field is increased, the path of the electron will have a sharper bend. Likewise, if the velocity of the electron increases, the field around it increases and the path will bend more sharply.

A schematic diagram of a basic magnetron is shown in figure 2-21A. The tube consists of a cylindrical plate with a cathode placed along the center axis of the plate. The tuned circuit is made up of cavities in which oscillations take place and are physically located in the plate.

When no magnetic field exists, heating the cathode results in a uniform and direct movement of the field from the cathode to the plate, as illustrated in figure 2-21B. However, as the magnetic field surrounding the tube is increased, a single electron is affected, as shown in figure 2-22. In figure 2-22, view (A), the magnetic field has been increased to a point where the electron proceeds to the plate in a curve rather than a direct path.

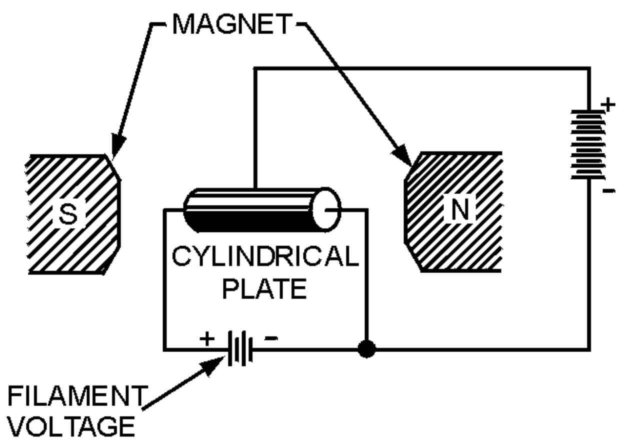


Figure 2-21A.—Basic magnetron. SIDE VIEW.

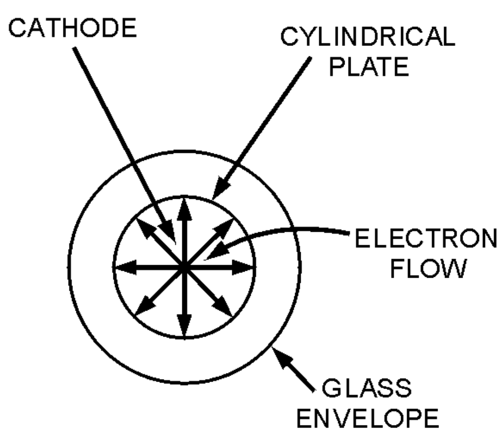


Figure 2-21B.—Basic magnetron. END VIEW OMITTING MAGNETS.

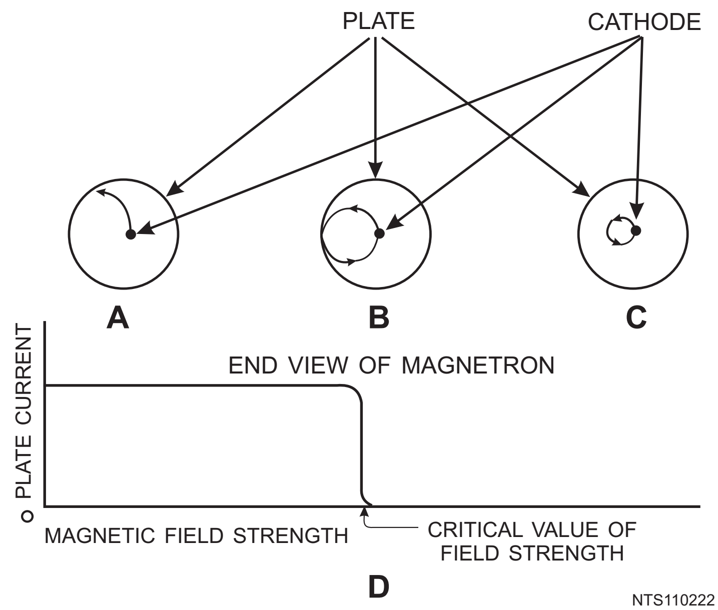


Figure 2-22.—Effect of a magnetic field on a single electron.

In view (B) of figure 2-22, the magnetic field has reached a value great enough to cause the electron to just miss the plate and return to the filament in a circular orbit. This value is the CRITICAL VALUE of field strength. In view (C), the value of the field strength has been increased to a point beyond the critical value; the electron is made to travel to the cathode in a circular path of smaller diameter.

View (D) of figure 2-22. shows how the magnetron plate current varies under the influence of the varying magnetic field. In view (A), the electron flow reaches the plate, so a large amount of plate current is flowing. However, when the critical field value is reached, as shown in view (B), the electrons are deflected away from the plate and the plate current then drops quickly to a very small value. When the field strength is made still greater, as shown in view (C), the plate current drops to zero.

When the magnetron is adjusted to the cutoff, or critical value of the plate current, and the electrons just fail to reach the plate in their circular motion, it can produce oscillations at microwave frequencies. These oscillations are caused by the currents induced electrostatically by the moving electrons. The frequency is determined by the time it takes the electrons to travel from the cathode toward the plate and back again. A transfer of microwave energy to a load is made possible by connecting an external circuit between the cathode and the plate of the magnetron. Magnetron oscillators are divided into two classes: NEGATIVE-RESISTANCE and ELECTRON-RESONANCE MAGNETRON OSCILLATORS.

A negative-resistance magnetron oscillator is operated by a static negative resistance between its electrodes. This oscillator has a frequency equal to the frequency of the tuned circuit connected to the tube.

An electron-resonance magnetron oscillator is operated by the electron transit time required for electrons to travel from cathode to plate. This oscillator is capable of generating very large peak power outputs at frequencies in the thousands of megahertz. Although its average power output over a period of time is low, it can provide very high-powered oscillations in short bursts of pulses.

Q-31. The folded waveguide in a bwo serves the same purpose as what component in a twt?

Q-32. What serves as a grid in a magnetron?

Q-33. A cylindrical copper block with resonant cavities around the circumference is used as what component of a magnetron?

Q-34. What controls the output frequency of a magnetron?

Q-35. What element in the magnetron causes the curved path of electron flow?

Q-36. What is the term used to identify the amount of field strength required to cause the electrons to just miss the plate and return to the filament in a circular orbit?

Q-37. A magnetron will produce oscillations when the electrons follow what type of path?

NEGATIVE-RESISTANCE MAGNETRON.—The split-anode, negative-resistance magnetron is a variation of the basic magnetron which operates at a higher frequency. The negative-resistance magnetron is capable of greater power output than the basic magnetron. Its general construction is similar to the basic magnetron except that it has a split plate, as shown in figure 2-23A and B. These half plates are operated at different potentials to provide an electron motion, as shown in figure 2-24. The electron leaving the cathode and progressing toward the high-potential plate is deflected by the magnetic field and follows the path shown in figure 2-24. After passing the split between the two plates, the electron enters the electrostatic field set up by the lower-potential plate.

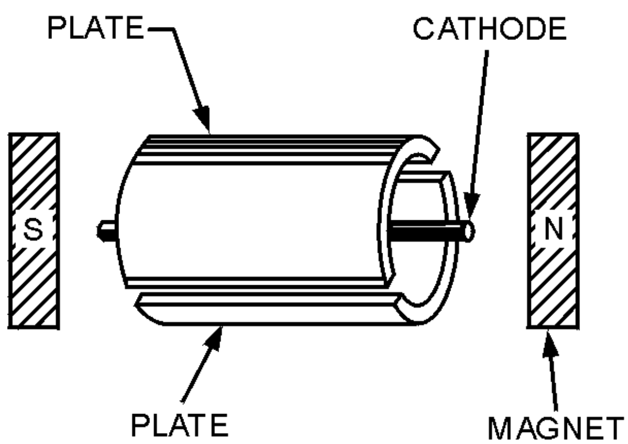


Figure 2-23A.—Split-anode magnetron.

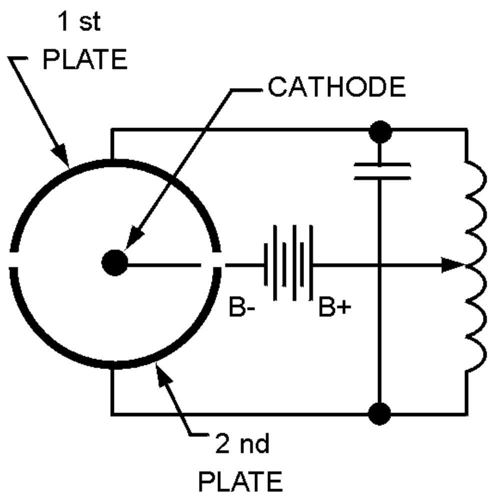


Figure 2-23B.—Split-anode magnetron.

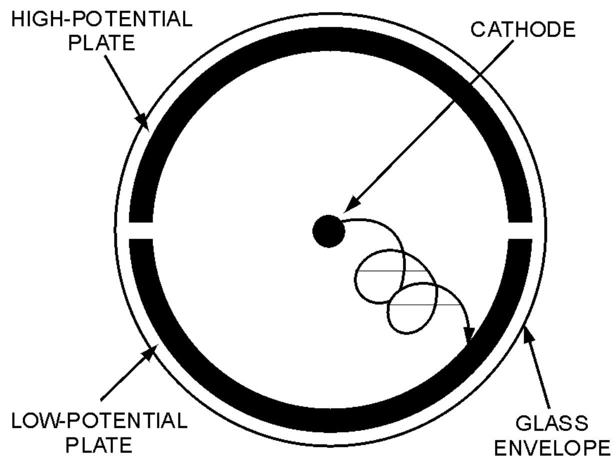


Figure 2-24.—Movement of an electron in a split-anode magnetron.

Here the magnetic field has more effect on the electron and deflects it into a tighter curve. The electron then continues to make a series of loops through the magnetic field and the electric field until it finally arrives at the low-potential plate.

Oscillations are started by applying the proper magnetic field to the tube. The field value required is slightly higher than the critical value. In the split-anode tube, the critical value is the field value required to cause all the electrons to miss the plate when its halves are operating at the same potential. The alternating voltages impressed on the plates by the oscillations generated in the tank circuit will cause electron motion, such as that shown in figure 2-24, and current will flow. Since a very concentrated magnetic field is required for the negative-resistance magnetron oscillator, the length of the tube plate is limited to a few centimeters to keep the magnet at reasonable dimensions. In addition, a small diameter tube is required to make the magnetron operate efficiently at microwave frequencies. A heavy-walled plate is used to increase the radiating properties of the tube. Artificial cooling methods, such as forced-air or water-cooled systems, are used to obtain still greater dissipation in these high-output tubes.

The output of a magnetron is reduced by the bombardment of the filament by electrons which travel in loops, shown in figure 2-22, views (B) and (C). This action causes an increase of filament temperature under conditions of a strong magnetic field and high plate voltage and sometimes results in unstable operation of the tube. The effects of filament bombardment can be reduced by operating the filament at a reduced voltage. In some cases, the plate voltage and field strength are also reduced to prevent destructive filament bombardment.

ELECTRON-RESONANCE MAGNETRON.—In the electron-resonance magnetron, the plate is constructed to resonate and function as a tank circuit. Thus, the magnetron has no external tuned circuits. Power is delivered directly from the tube through transmission lines, as shown in figure 2-25. The constants and operating conditions of the tube are such that the electron paths are somewhat different from those in figure 2-24. Instead of closed spirals or loops, the path is a curve having a series of sharp points, as illustrated in figure 2-26. Ordinarily, this type of magnetron has more than two segments in the plate. For example, figure 2-26 illustrates an eight-segment plate.

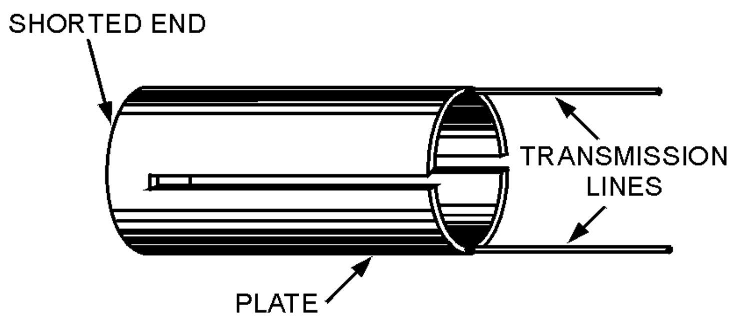


Figure 2-25.—Plate tank circuit of a magnetron.

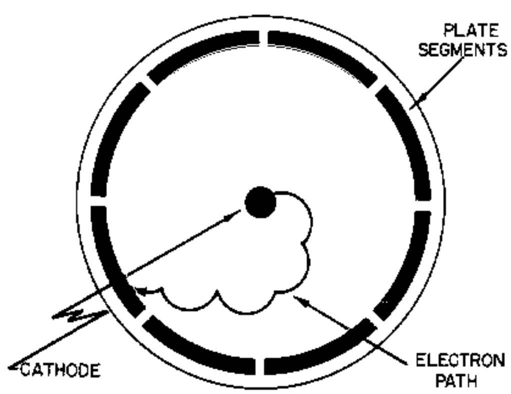


Figure 2-26.—Electron path in an electron-resonance magnetron.

The electron-resonance magnetron is the most widely used for microwave frequencies because it has reasonably high efficiency and relatively high output. The average power of the electron-resonance magnetron is limited by the amount of cathode emission, and the peak power is limited by the maximum voltage rating of the tube components. Three common types of anode blocks used in electron-resonance magnetrons are shown in figure 2-27.

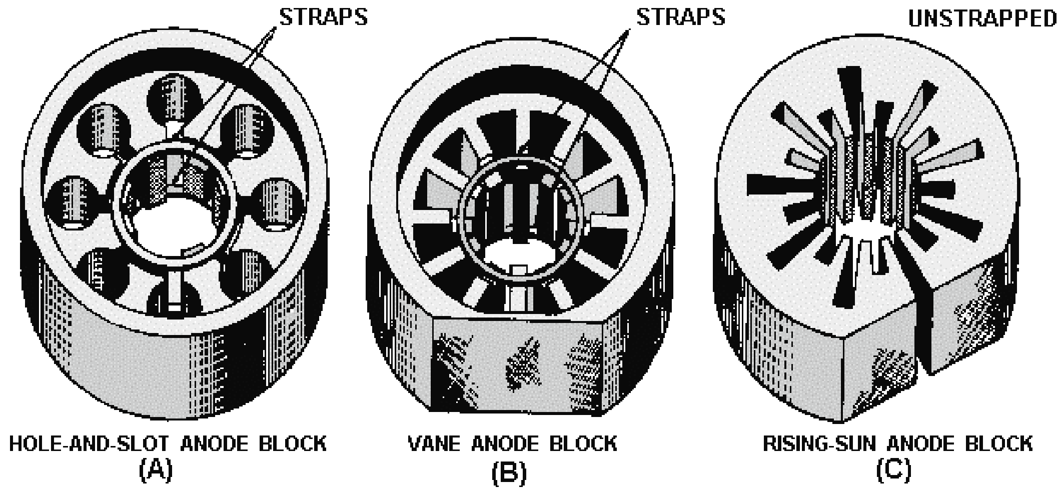


Figure 2-27.—Common types of anode blocks.

The anode block shown in figure 2-27, view (A), has cylindrical cavities and is called a HOLE-AND-SLOT ANODE. The anode block in view (B) is called the VANE ANODE which has trapezoidal cavities. The first two anode blocks operate in such a way that alternate segments must be connected, or strapped, so that each segment is opposite in polarity to the segment on either side, as shown in figure 2-28. This also requires an even number of cavities.

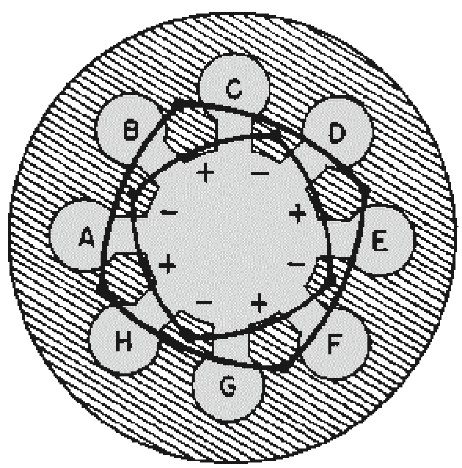


Figure 2-28.—Strapping alternate segments.

The anode block illustrated in figure 2-27, view (C), is called a RISING-SUN BLOCK. The alternate large and small trapezoidal cavities in this block result in a stable frequency between the resonant frequencies of the large and small cavities.

Figure 2-29A, shows the physical relationships of the resonant cavities contained in the hole-and-slot anode (figure 2-27, view (A)). This will be used when analyzing the operation of the electron-resonance magnetron.

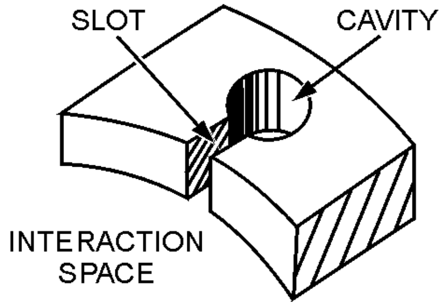


Figure 2-29A.—Equivalent circuit of a hole-and-slot cavity.

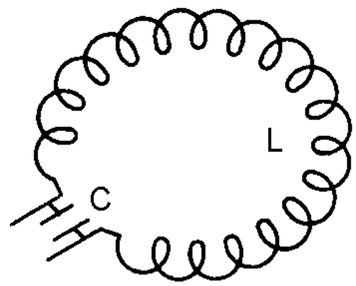


Figure 2-29B.—Equivalent circuit of a hole-and-slot cavity.

Electrical Equivalent.—Notice in figure 2-29A, that the cavity consists of a cylindrical hole in the copper anode and a slot which connects the cavity to the interaction space.

The equivalent electrical circuit of the hole and slot is shown in figure 2-29B. The parallel sides of the slot form the plates of a capacitor while the walls of the hole act as an inductor. The hole and slot thus form a high-Q, resonant LC circuit. As shown in figure 2-27, the anode of a magnetron has a number of these cavities.

An analysis of the anodes in the hole-and-slot block reveals that the LC tanks of each cavity are in series (assuming the straps have been removed), as shown in figure 2-30. However, an analysis of the anode block after alternate segments have been strapped reveals that the cavities are connected in parallel because of the strapping. Figure 2-31 shows the equivalent circuit of a strapped anode.

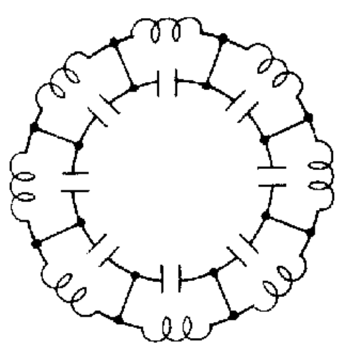


Figure 2-30.—Cavities connected in series.

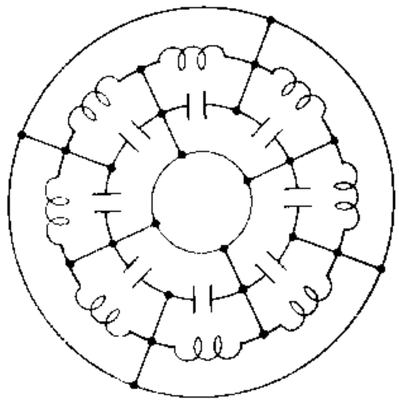


Figure 2-31.—Cavities in parallel because of strapping.

Electric Field.—The electric field in the electron-resonance oscillator is a product of ac and dc fields. The dc field extends radially from adjacent anode segments to the cathode, as shown in figure 2-32. The ac fields, extending between adjacent segments, are shown at an instant of maximum magnitude of one alternation of the rf oscillations occurring in the cavities.

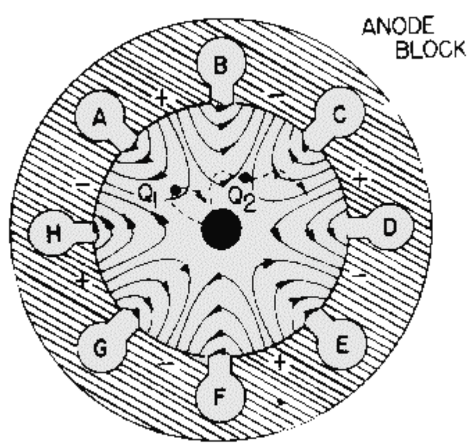


Figure 2-32.—Probable electron paths in an electron-resonance magnetron oscillator.

The strong dc field going from anode to cathode is created by a large, negative dc voltage pulse applied to the cathode. This strong dc field causes electrons to accelerate toward the plate after they have been emitted from the cathode. Recall that an electron moving against an E field is accelerated by the field and takes energy from the field. Also, an electron gives up energy to a field and slows down if it is moving in the same direction as the field (positive to negative). Oscillations are sustained in a magnetron because as electrons pass through the ac and dc fields, they gain energy from the dc field and give up energy to the ac field. The electrons that give up energy to the ac field are called WORKING ELECTRONS. However, not all of the electrons give up energy to the ac field. Some electrons take energy from the ac field, which is an undesirable action.

In figure 2-32, consider electron Q1, which is shown entering the field around the slot entrance to cavity A. The clockwise rotation of the electron path is caused by the interaction of the magnetic field around the moving electron with the permanent magnetic field. The permanent magnetic field is assumed to be going into the paper in figure 2-32 (the action of an electron moving in an H field was explained earlier). Notice that electron Q1 is moving against the ac field around cavity A. The electron takes energy from the ac field and then accelerates, turning more sharply when its velocity increases. Thus, electron Q1 turns back toward the cathode. When it strikes the cathode, it gives up the energy it received from the ac field. This bombardment also forces more electrons to leave the cathode and accelerate toward the anode. Electron Q2 is slowed down by the field around cavity B and gives up some of its energy to the ac field. Since electron Q2 loses velocity, the deflective force exerted by the H field is reduced. The electron path then deviates to the left in the direction of the anode, rather than returning to the cathode as did electron Q1.

The cathode to anode potential and the magnetic field strength determine the amount of time for electron Q2 to travel from a position in front of cavity B to a position in front of cavity C. Cavity C is equal to approximately 1/2 cycle of the rf oscillations of the cavities. When electron Q2 reaches a position in front of cavity C, the ac field of cavity C is reversed from that shown. Therefore, electron Q2 gives up energy to the ac field of cavity C and slows down even more. Electron Q2 actually gives up energy to each cavity as it passes and eventually reaches the anode when its energy is expended. Thus, electron Q2 has helped sustain oscillations because it has taken energy from the dc field and given it to the ac field. Electron Q1, which took energy from the ac field around cavity A, did little harm because it immediately returned to the cathode.

The cumulative action of many electrons returning to the cathode while others are moving toward the anode forms a pattern resembling the moving spokes of a wheel known as a SPACE-CHARGE WHEEL, as indicated in figure 2-33. Electrons in the spokes of the wheel are the working electrons.

The space-charge wheel rotates about the cathode at an angular velocity of 2 poles (anode segments) per cycle of the ac field. This phase relationship enables the concentration of electrons to continuously deliver energy to sustain the rf oscillations. Electrons emitted from the area of the cathode between the spokes are quickly returned to the cathode.

In figure 2-33 the alternate segments between cavities are assumed to be at the same potential at the same instant. An ac field is assumed to exist across each individual cavity. This mode of operation is called the PI MODE, since adjacent segments of the anode have a phase difference of 180 degrees, or one-pi radian. Several other modes of oscillation are possible, but a magnetron operating in the pi mode has greater power and output and is the most commonly used.

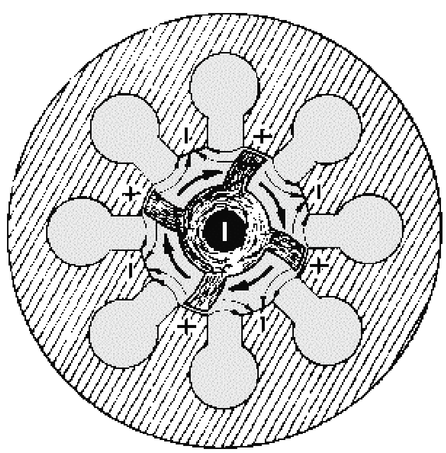


Figure 2-33.—Rotating space-charge wheel in an eight-cavity magnetron.

An even number of cavities, usually six or eight, are used and alternate segments are strapped to ensure that they have identical polarities. The frequency of the pi mode is separated from the frequency of the other modes by strapping.

For the pi mode, all parts of each strapping ring are at the same potential; but the two rings have alternately opposing potentials, as shown in figure 2-34. Stray capacitance between the rings adds capacitive loading to the resonant mode. For other modes, however, a phase difference exists between the successive segments connected to a given strapping ring which causes current to flow in the straps.

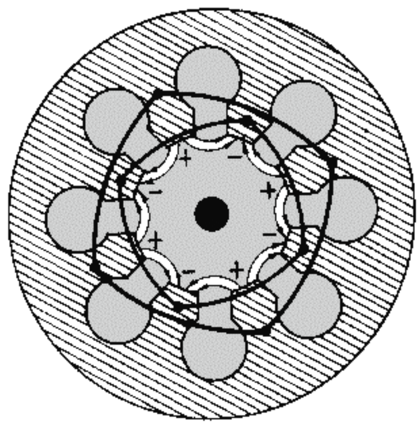


Figure 2-34.—Alternate segments connected by strapping rings.

The straps contain inductance, and an inductive shunt is placed in parallel with the equivalent circuit. This lowers the inductance and increases the frequency at modes other than the pi mode.

Q-38. What is the primary difference in construction between the basic magnetron and the negative-resistance magnetron?

Q-39. What starts the oscillations in a negative-resistance magnetron?

Q-40. Why is the negative-resistance magnetron often operated with reduced filament voltage?

Q-41. What type of electron-resonance anode block does not require strapping?

Q-42. Without strapping, the resonant cavities of a hole-and-slot anode are connected in what manner?

Q-43. What are the electrons called that give up energy to the ac field in a magnetron?

COUPLING METHODS.—Energy (rf) can be removed from a magnetron by means of a COUPLING LOOP. At frequencies lower than 10,000 megahertz, the coupling loop is made by bending the inner conductor of a coaxial cable into a loop. The loop is then soldered to the end of the outer conductor so that it projects into the cavity, as shown in figure 2-35A. Locating the loop at the end of the cavity, as shown in figure 2-35B, causes the magnetron to obtain sufficient pickup at higher frequencies.

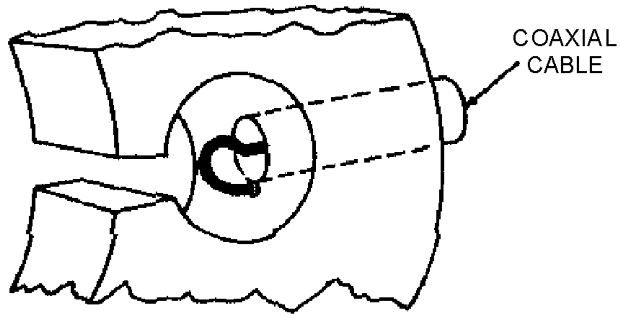


Figure 2-35A.—Magnetron coupling methods.

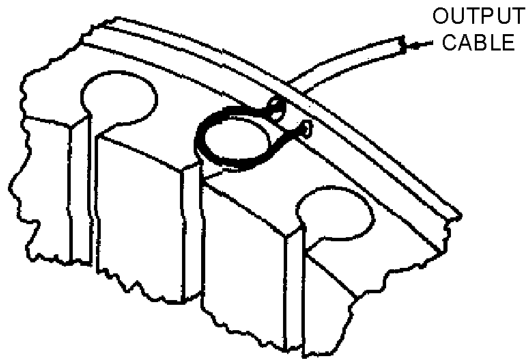


Figure 2-35B.—Magnetron coupling methods.

The SEGMENT-FED LOOP METHOD is shown in figure 2-35C. The loop intercepts the magnetic lines passing between cavities. The STRAP-FED LOOP METHOD (figure 2-35D), intercepts the energy between the strap and the segment. On the output side, the coaxial line feeds another coaxial line directly or feeds a waveguide through a choke joint. The vacuum seal at the inner conductor helps to support the line. APERTURE, OR SLOT, COUPLING is illustrated in figure 2-35E. Energy is coupled directly to a waveguide through an iris.

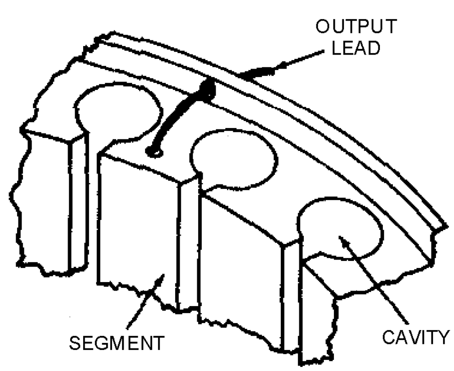


Figure 2-35C.—Magnetron coupling methods.

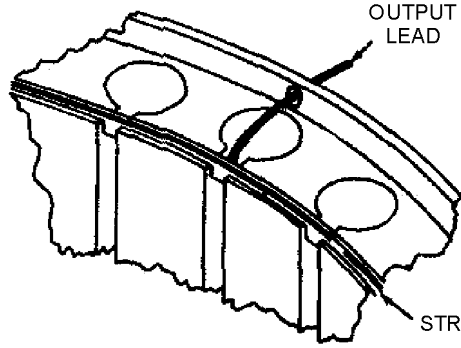


Figure 2-35D.—Magnetron coupling methods.

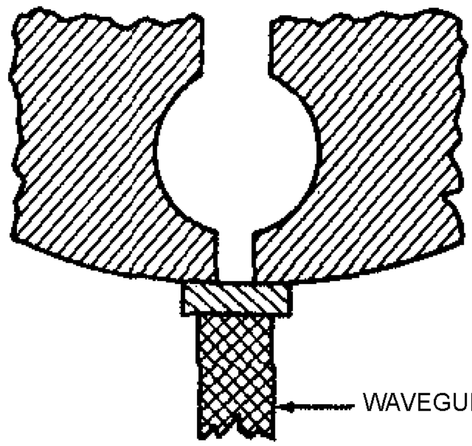


Figure 2-35E.—Magnetron coupling methods.

MAGNETRON TUNING.—A tunable magnetron permits the system to be operated at a precise frequency anywhere within a band of frequencies, as determined by magnetron characteristics.

The resonant frequency of a magnetron may be changed by varying the inductance or capacitance of the resonant cavities. In figure 2-36, an inductive tuning element is inserted into the hole portion of the hole-and-slot cavities. It changes the inductance of the resonant circuits by altering the ratio of surface area to cavity volume in a high-current region. The type of tuner illustrated in figure 2-36 is called a SPROCKET TUNER or CROWN-OF-THORNS TUNER. All of its tuning elements are attached to a frame which is positioned by a flexible bellows arrangement. The insertion of the tuning elements into each anode hole decreases the inductance of the cavity and therefore increases the resonant frequency. One of the limitations of inductive tuning is that it lowers the unloaded Q of the cavities and therefore reduces the efficiency of the tube.

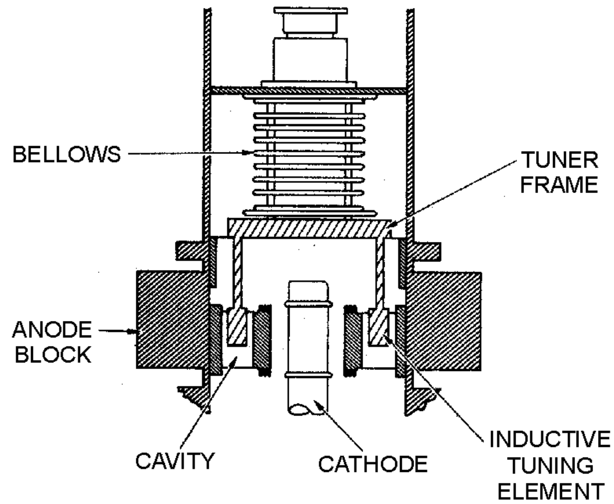


Figure 2-36.—Inductive magnetron tuning.

The insertion of an element (ring) into the cavity slot, as shown in figure 2-37, increases the slot capacitance and decreases the resonant frequency. Because the gap is narrowed in width, the breakdown voltage is lowered. Therefore, capacitively tuned magnetrons must be operated with low voltages and at low-power outputs. The type of capacitive tuner illustrated in figure 2-37 is called a COOKIE-CUTTER TUNER. It consists of a metal ring inserted between the two rings of a double-strapped magnetron, which serves to increase the strap capacitance. Because of the mechanical and voltage breakdown problems associated with the cookie-cutter tuner, it is more suitable for use at longer wavelengths. Both the capacitance and inductance tuners described are symmetrical; that is, each cavity is affected in the same manner, and the pi mode is preserved.

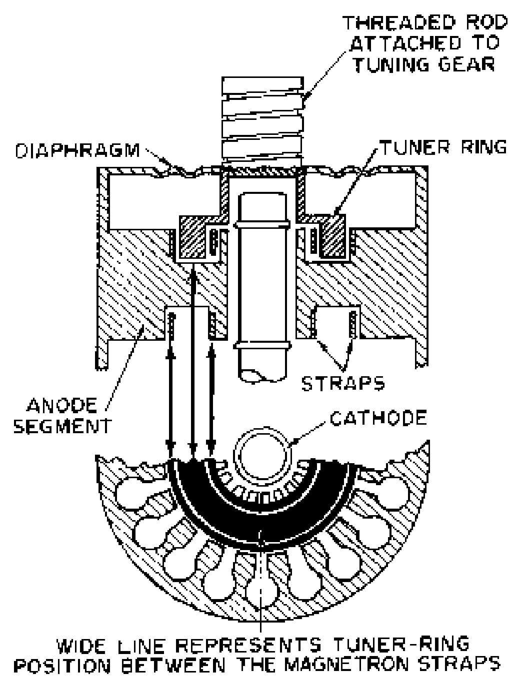


Figure 2-37.—Capacitive magnetron tuning.

A 10-percent frequency range may be obtained with either of the two tuning methods described above. Also, the two tuning methods may be used in combination to cover a larger tuning range than is possible with either one alone.

ARCING IN MAGNETRONS.—During initial operation a high-powered magnetron arcs from cathode to plate and must be properly BROKEN IN or BAKED IN. Actually, arcing in magnetrons is very common. It occurs with a new tube or following long periods of idleness.

One of the prime causes of arcing is the release of gas from tube elements during idle periods. Arcing may also be caused by the presence of sharp surfaces within the tube, mode shifting, and by drawing excessive current. While the cathode can withstand considerable arcing for short periods of time, continued arcing will shorten the life of the magnetron and may destroy it entirely. Therefore, each time excessive arcing occurs, the tube must be baked in again until the arcing ceases and the tube is stabilized.

The baking-in procedure is relatively simple. Magnetron voltage is raised from a low value until arcing occurs several times a second. The voltage is left at that value until arcing dies out. Then the voltage is raised further until arcing again occurs and is left at that value until the arcing again ceases. Whenever the arcing becomes very violent and resembles a continuous arc, the applied voltage is excessive and should be reduced to permit the magnetron to recover. When normal rated voltage is reached and the magnetron remains stable at the rated current, the baking-in is complete. A good maintenance practice is to bake-in magnetrons left idle in the equipment or those used as spares when long periods of nonoperating time have accumulated.

The preceding information is general in nature. The recommended times and procedures in the technical manuals for the equipment should be followed when baking-in a specific type magnetron.

The Crossed-Field Amplifier (Amplitron)

The CROSSED-FIELD AMPLIFIER (cfa), commonly known as an AMPLITRON and sometimes referred to as a PLATINOTRON, is a broadband microwave amplifier that can also be used as an oscillator. The cfa is similar in operation to the magnetron and is capable of providing relatively large amounts of power with high efficiency. The bandwidth of the cfa, at any given instant, is approximately plus or minus 5 percent of the rated center frequency. Any incoming signals within this bandwidth are amplified. Peak power levels of many megawatts and average power levels of tens of kilowatts average are, with efficiency ratings in excess of 70 percent, possible with crossed-field amplifiers.

Because of the desirable characteristics of wide bandwidth, high efficiency, and the ability to handle large amounts of power, the cfa is used in many applications in microwave electronic systems. This high efficiency has made the cfa useful for space-telemetry applications, and the high power and stability have made it useful in high-energy, linear atomic accelerators. When used as the intermediate or final stage in high-power radar systems, all of the advantages of the cfa are used.

Since the cfa operates in a manner so similar to the magnetron, the detailed theory is not presented in this module. Detailed information of cfa operation is available in NAVSHIPS 0967-443-2230, *Handling, Installation and Operation of Crossed-Field Amplifiers*. As mentioned earlier, crossed-field amplifiers are commonly called Amplitrons. You should note, however, that Amplitron is a trademark of the Raytheon Manufacturing Company for the Raytheon line of crossed-field amplifiers. An illustration of a crossed-field amplifier is shown in figure 2-38.

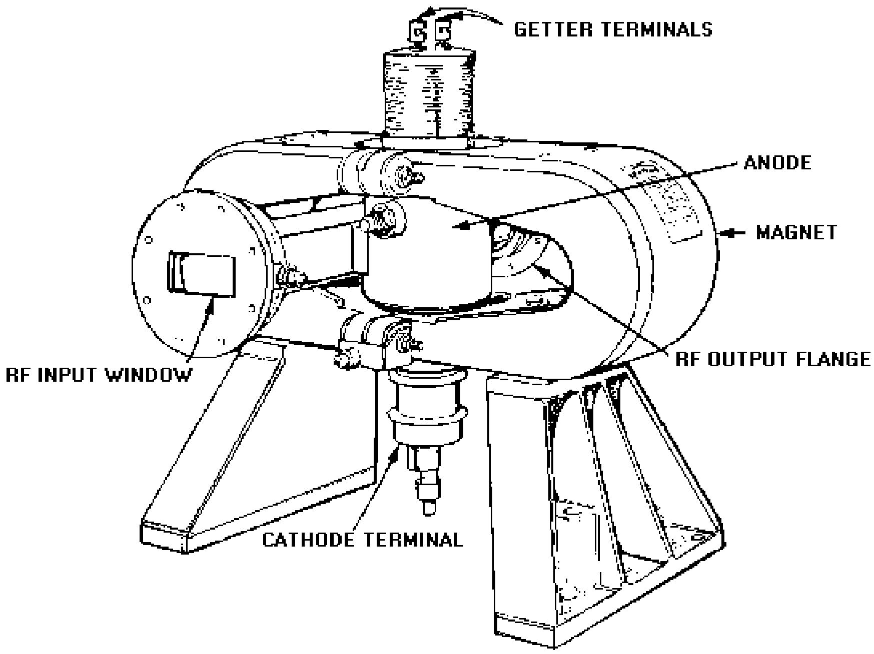


Figure 2-38.—Crossed-field amplifier (Amplitron).

Q-44. Why is the pi mode the most commonly used magnetron mode of operation?

Q-45. What two methods are used to couple energy into and out of magnetrons?

Q-46. Magnetron tuning by altering the surface-to-volume ratio of the hole portion of a hole-and-slot cavity is what type of tuning?

Q-47. Capacitive tuning by inserting a ring into the cavity slot of a magnetron is accomplished by what type of tuning mechanism?

SOLID-STATE MICROWAVE DEVICES

As with vacuum tubes, the special electronics effects encountered at microwave frequencies severely limit the usefulness of transistors in most circuit applications. The need for small-sized microwave devices has caused extensive research in this area. This research has produced solid-state devices with higher and higher frequency ranges. The new solid-state microwave devices are predominantly active, two-terminal diodes, such as tunnel diodes, varactors, transferred-electron devices, and avalanche transit-time diodes. This section will describe the basic theory of operation and some of the applications of these relatively new solid-state devices.

Tunnel Diode Devices

The TUNNEL DIODE is a pn junction with a very high concentration of impurities in both the p and n regions. The high concentration of impurities causes it to exhibit the properties of a negative-resistance element over part of its range of operation, as shown in the characteristic curve in figure 2-39. In other words, the resistance to current flow through the tunnel diode increases as the applied voltage increases over a portion of its region of operation. Outside the negative-resistance region, the tunnel diode functions essentially the same as a normal diode. However, the very high impurity density causes a junction depletion region so narrow that both holes and electrons can transfer across the pn junction by a quantum mechanical action called TUNNELING. Tunneling causes the negative-resistance action and is so fast that no transit-time effects occur even at microwave frequencies. The lack of a transit-time effect permits the use of tunnel diodes in a wide variety of microwave circuits, such as amplifiers, oscillators, and switching devices. The detailed theory of tunnel-diode operation and the negative-resistance property exhibited by the tunnel diode was discussed in *NEETS*, Module 7, *Introduction to Solid-State Devices and Power Supplies*, Chapter 3.

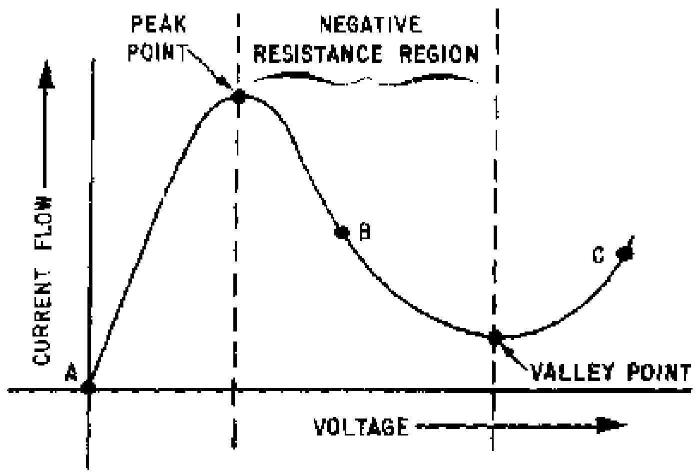


Figure 2-39.—Tunnel-diode characteristic curve.

TUNNEL-DIODE OSCILLATORS.—A tunnel diode, biased at the center point of the negative-resistance range (point B in figure 2-39) and coupled to a tuned circuit or cavity, produces a very stable oscillator. The oscillation frequency is the same as the tuned circuit or cavity frequency.

Microwave tunnel-diode oscillators are useful in applications that require microwatts or, at most, a few milliwatts of power, such as local oscillators for microwave superheterodyne receivers. Tunnel-diode oscillators can be mechanically or electronically tuned over frequency ranges of about one octave and have a top-end frequency limit of approximately 10 gigahertz.

Tunnel-diode oscillators that are designed to operate at microwave frequencies generally use some form of transmission line as a tuned circuit. Suitable tuned circuits can be built from coaxial lines, transmission lines, and waveguides.

An example of a highly stable tunnel-diode oscillator is shown in figure 2-40. A tunnel-diode is loosely coupled to a high-Q tunable cavity. Loose coupling is achieved by using a short, antenna feed probe placed off-center in the cavity. Loose coupling is used to increase the stability of the oscillations and the output power over a wider bandwidth.

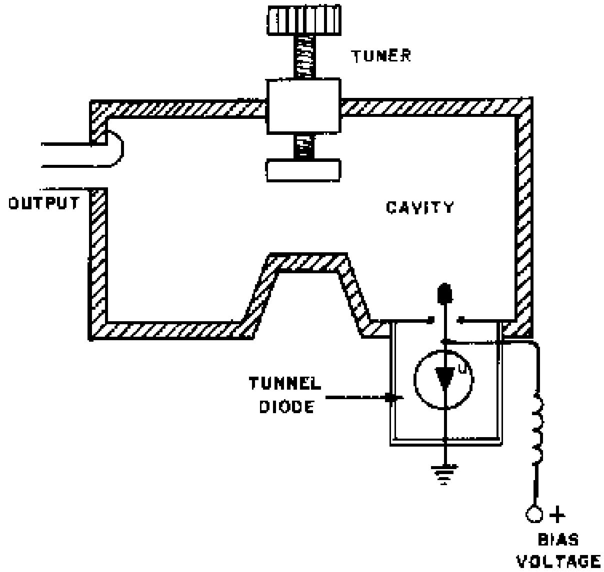


Figure 2-40.—Tunnel-diode oscillator.

The output power produced is in the range of a few hundred microwatts, sufficient for many microwave applications. The frequency at which the oscillator operates is determined by the physical positioning of the tuner screw in the cavity. Changing the output frequency by this method is called MECHANICAL TUNING. In addition to mechanical tuning, tunnel-diode oscillators may be tuned electronically. One method is called BIAS TUNING and involves nothing more than changing the bias voltage to change the bias point on the characteristic curve of the tunnel-diode. Another method is called VARACTOR TUNING and requires the addition of a varactor to the basic circuit. Varactors were discussed in *NEETS*, Module 7, *Introduction to Solid-State Devices, and Power Supplies*, Chapter 3. Tuning is achieved by changing the voltage applied across the varactor which alters the capacitance of the tuned circuit.

TUNNEL-DIODE AMPLIFIERS.—Low-noise, tunnel-diode amplifiers represent an important microwave application of tunnel diodes. Tunnel-diode amplifiers with frequencies up to 85 gigahertz have been built in waveguides, coaxial lines, and transmission lines. The low-noise generation, gain ratios of up to 30 dB, high reliability, and light weight make these amplifiers ideal for use as the first stage of amplification in communications and radar receivers.

Most microwave tunnel-diode amplifiers are REFLECTION-TYPE, CIRCULATOR-COUPLED AMPLIFIERS. As in oscillators, the tunnel diode is biased to the center point of its negative-resistance region, but a CIRCULATOR replaces the tuned cavity.

A circulator is a waveguide device that allows energy to travel in one direction only, as shown in figure 2-41. The tunnel diode in figure 2-41 is connected across a tuned-input circuit. This arrangement normally produces feedback that causes oscillations if the feedback is allowed to reflect back to the tuned-input circuit. The feedback is prevented because the circulator carries all excess energy to the absorptive load (R_L). In this configuration the tunnel diode cannot oscillate, but will amplify.

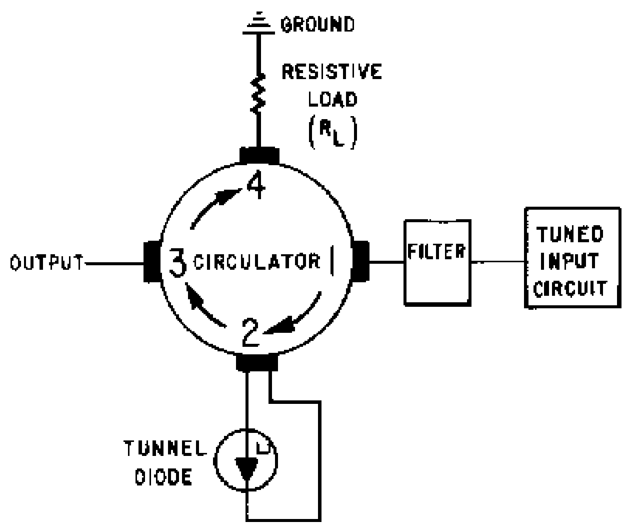


Figure 2-41.—Tunnel-diode amplifier.

The desired frequency input signal is fed to port 1 of the circulator through a bandpass filter. The filter serves a dual purpose as a bandwidth selector and an impedance-matching device that improves the gain of the amplifiers. The input energy enters port 2 of the circulator and is amplified by the tunnel diode. The amplified energy is fed from port 2 to port 3 and on to the mixer. If any energy is reflected from port 3, it is passed to port 4, where it is absorbed by the matched load resistance.

TUNNEL-DIODE FREQUENCY CONVERTERS AND MIXERS.—Tunnel diodes make excellent mixers and frequency converters because their current-voltage characteristics are highly nonlinear. While other types of frequency converters usually have a conversion power loss, tunnel-diode converters can actually have a conversion power gain. A single tunnel diode can also be designed to act as both the nonlinear element in a converter and as the negative-resistance element in a local oscillator at the same time.

Practical tunnel-diode frequency converters usually have either a unity conversion gain or a small conversion loss. Conversion gains as high as 20 dB are possible if the tunnel diode is biased near or into the negative-resistance region. Although high gain is useful in some applications, it presents problems in stability. For example, the greatly increased sensitivity to variations in input impedance can cause high-gain converters to be unstable unless they are protected by isolation circuitry.

As with tunnel-diode amplifiers, low-noise generation is one of the more attractive characteristics of tunnel-diode frequency converters. Low-noise generation is a primary concern in the design of today's extremely sensitive communications and radar receivers. This is one reason tunnel-diode circuits are finding increasingly wide application in these fields.

Q-48. Name the procedure used to reduce excessive arcing in a magnetron?

Q-49. What causes the negative-resistance property of tunnel diodes?

Q-50. What determines the frequency of a tunnel-diode oscillator?

Q-51. Why is the tunnel diode loosely coupled to the cavity in a tunnel-diode oscillator?

Q-52. What is the purpose of the circulator in a tunnel-diode amplifier?

Varactor Devices

The VARACTOR is another of the active two-terminal diodes that operates in the microwave range. Since the basic theory of varactor operation was presented in *NEETS*, Module 7, *Introduction to Solid-State Devices and Power Supplies*, Chapter 3, only a brief review of the basic principles is presented here.

The varactor is a semiconductor diode with the properties of a voltage-dependent capacitor. Specifically, it is a variable-capacitance, pn-junction diode that makes good use of the voltage dependency of the depletion-area capacitance of the diode.

In figure 2-42A, two materials are brought together to form a pn-junction diode. The different energy levels in the two materials cause a diffusion of the holes and electrons through both materials which tends to balance their energy levels. When this diffusion process stops, the diode is left with a small area on either side of the junction, called the depletion area, which contains no free electrons or holes. The movement of electrons through the materials creates an electric field across the depletion area that is described as a barrier potential and has the electrical characteristics of a charged capacitor.

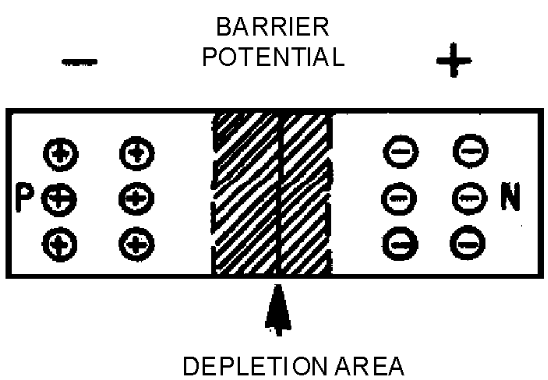


Figure 2-42A.—Pn-junction diode as a variable capacitor.

External bias, applied in either the forward or reverse direction, as shown in figure 2-42B and C, affects the magnitude, barrier potential, and width of the depletion area. Enough forward or reverse bias will overcome the barrier potential and cause current to flow through the diode. The width of the depletion region can be controlled by keeping the bias voltage at levels that do not allow current flow. Since the depletion area acts as a capacitor, the diode will perform as a variable capacitor that changes with the applied bias voltage. The capacitance of a typical varactor can vary from 2 to 50 picofarads for a bias variation of just 2 volts.

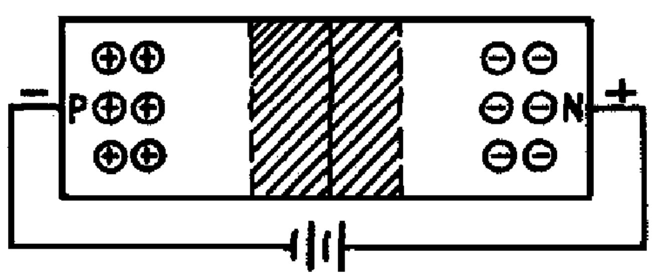


Figure 2-42B.—Pn-junction diode as a variable capacitor.

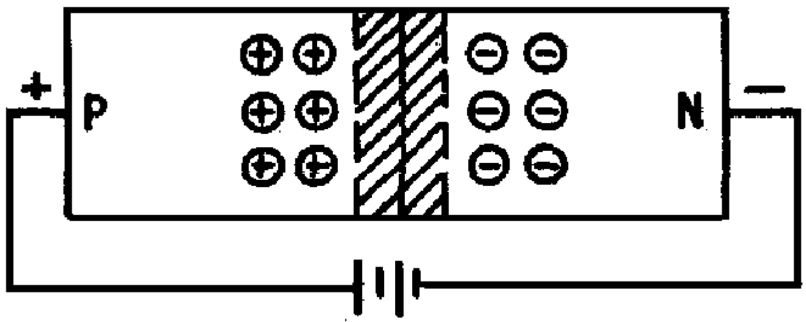


Figure 2-42C.—Pn-junction diode as a variable capacitor.

The variable capacitance property of the varactor allows it to be used in circuit applications, such as amplifiers, that produce much lower internal noise levels than circuits that depend upon resistance properties. Since noise is of primary concern in receivers, circuits using varactors are an important development in the field of low-noise amplification. The most significant use of varactors to date has been as the basic component in parametric amplifiers.

PARAMETRIC AMPLIFIERS.—The parametric amplifier is named for the time-varying parameter, or value of capacitance, associated with the operation. Since the underlying principle of operation is based on reactance, the parametric amplifier is sometimes called a REACTANCE AMPLIFIER.

The conventional amplifier is essentially a variable resistance that uses energy from a dc source to increase ac energy. The parametric amplifier uses a nonlinear variable reactance to supply energy from an ac source to a load. Since reactance does not add thermal noise to a circuit, parametric amplifiers produce much less noise than most conventional amplifiers.

Because the most important feature of the parametric amplifier is the low-noise characteristic, the nature of ELECTRONIC NOISE and the effect of this type of noise on receiver operation must first be discussed. Electronic noise is the primary limitation on receiver sensitivity and is the name given to very small randomly fluctuating voltages that are always present in electronic circuits. The sensitivity limit of the receiver is reached when the incoming signal falls below the level of the noise generated by the receiver circuits. At this point the incoming signal is hidden by the noise, and further amplification has no effect because the noise is amplified at the same rate as the signal. The effects of noise can be reduced by careful circuit design and control of operating conditions, but it cannot be entirely eliminated. Therefore, circuits such as the parametric amplifier are important developments in the fields of communication and radar.

The basic theory of parametric amplification centers around a capacitance that varies with time. Consider the simple series circuit shown in figure 2-43. When the switch is closed, the capacitor charges to value (Q). If the switch is opened, the isolated capacitor has a voltage across the plates determined by the charge Q divided by the capacitance C.

$$V = \frac{Q}{C}$$

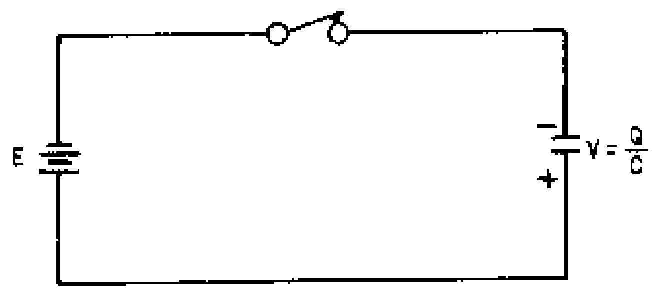


Figure 2-43.—Voltage amplification from a varying capacitor.

An increase in the charge Q or a decrease in the capacitance C causes an increase in the voltage across the plates. Thus, a voltage increase, or amplification, can be obtained by mechanically or electronically varying the amount of capacitance in the circuit. In practice a voltage-variable capacitance, such as a varactor, is used. The energy required to vary the capacitance is obtained from an electrical source called a PUMP.

Figure 2-44, view (A), shows a circuit application using a voltage-variable capacitor and a pump circuit. The pump circuit decreases the capacitance each time the input signal (E) across the capacitor reaches maximum. The decreased capacitance causes a voltage buildup as shown by the dotted line in view (B). Therefore, each time the pump decreases capacitance (view (C)), energy transfers from the pump circuit to the input signal. The step-by-step buildup of the input-signal energy level is shown in view (D).

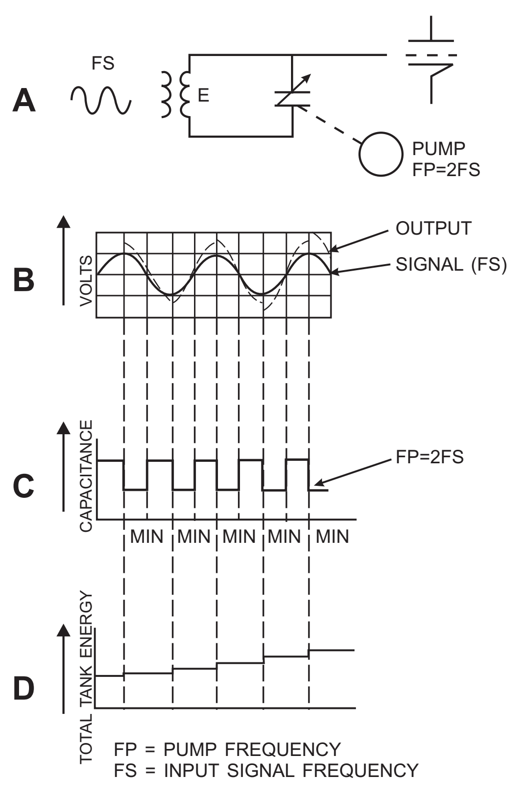


Figure 2-44.—Energy transfer from pump signal to input signal.

Proper phasing between the pump and the input signal is crucial in this circuit. The electrical pump action is simply a sine-wave voltage applied to a varactor located in a resonant cavity. For proper operation, the capacitance must be decreased when the input voltage is maximum and increased when the input voltage is minimum. In other words, the pump signal frequency must be exactly double the frequency of the input signal. This relationship can be seen when you compare views (B) and (C). A parametric amplifier of the type shown in figure 2-44 is quite phase-sensitive. The input signal and the capacitor variation are often in the wrong phase for long periods of time.

A parametric amplifier that is not phase-sensitive, referred to as a NONDEGENERATIVE PARAMETRIC AMPLIFIER, uses a pump circuit with a frequency higher than twice the input signal. The higher-frequency pump signal mixes with the input signal and produces additional frequencies that represent both the sum and difference of the input signal and pump frequencies.

Figure 2-45A, is a diagram of a typical nondegenerative parametric amplifier with the equivalent circuit shown in figure 2-45B. The pump signal (fp) is applied to the varactor. The cavity on the left is resonant at the input frequency (fs), and the cavity on the right is resonant at the difference frequency (fp-fs). The difference frequency is called the IDLER- or LOWER-SIDEBAND frequency. The varactor is located at the high-voltage points of the two cavities and is reverse biased by a small battery. The pump signal varies the bias above and below the fixed-bias level.

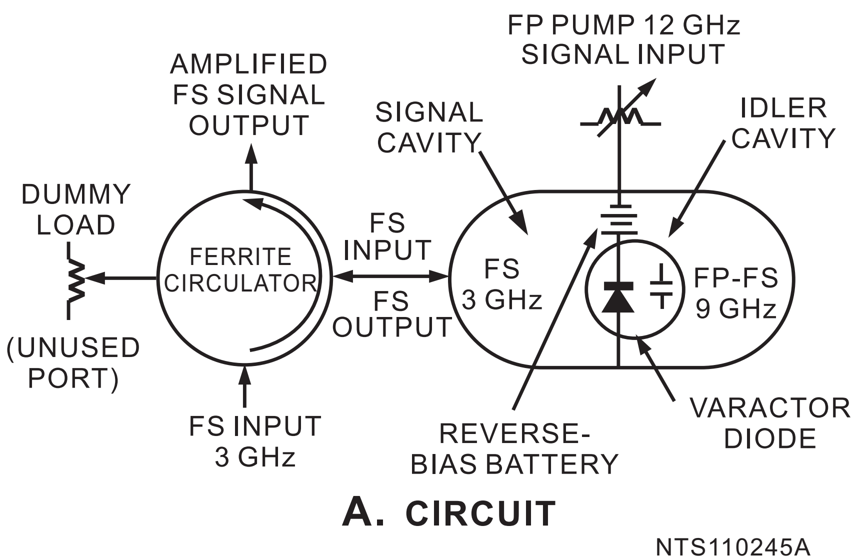


Figure 2-45A.—Nondegenerative parametric amplifier. CIRCUIT.

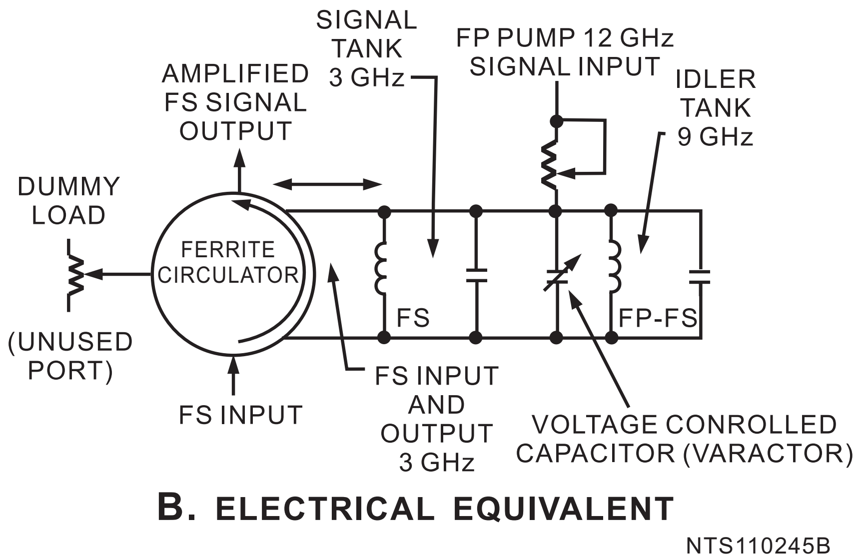


Figure 2-45B.—Nondegenerative parametric amplifier. ELECTRICAL EQUIVALENT.

The pump signal causes the capacitor in figure 2-45A to vary at a 12-gigahertz rate. The 3-gigahertz input signal enters via a four-port ferrite circulator, is developed in the signal cavity, and applied across the varactor. The nonlinear action of the varactor produces a 9-gigahertz difference frequency (fp-fs) with an energy-level higher than the original input signal.

The difference (idler) frequency is reapplied to the varactor to increase the gain and to produce an output signal of the correct frequency. The 9-gigahertz idler frequency recombines with the 12-gigahertz pump signal and produces a 3-gigahertz difference signal that has a much larger amplitude than the original 3-gigahertz input signal. The amplified signal is sent to the ferrite circulator for transfer to the next stage.

As with tunnel-diode amplifiers, the circulator improves stability by preventing reflection of the signal back into the amplifier. Reflections would be amplified and cause uncontrollable oscillations. The ferrite circulator also serves as an isolator to prevent source and load impedance changes from affecting gain.

Typically, the gain of a parametric amplifier is about 20 dB. The gain can be controlled with a variable attenuator that changes the amount of pump power applied to the varactor.

Parametric amplifiers are relatively simple in construction. The only component is a varactor diode placed in an arrangement of cavities and waveguides. The most elaborate feature of the amplifier is the mechanical tuning mechanism. Figure 2-46 illustrates an actual parametric amplifier.

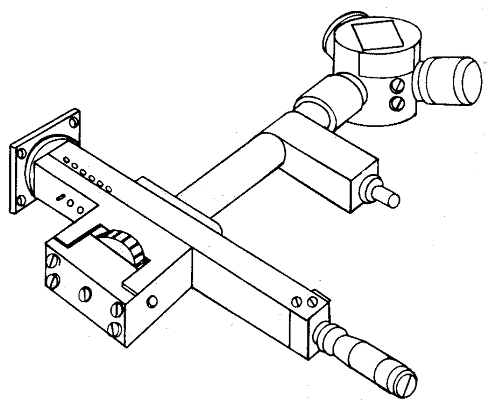


Figure 2-46.—Parametric amplifier.

PARAMETRIC FREQUENCY CONVERTERS.—Parametric frequency converters, using varactors, are of three basic types. The UPPER-SIDEBAND PARAMETRIC UP-CONVERTER produces an output frequency that is the SUM of the input frequency and the pump frequency. The LOWER-SIDEBAND PARAMETRIC DOWN-CONVERTER produces an output frequency that is the DIFFERENCE between the pump frequency and the input frequency. The DOUBLE-SIDEBAND PARAMETRIC UP-CONVERTER produces an output in which both the SUM and the DIFFERENCE of the pump and input frequencies are available.

Parametric frequency converters are very similar to parametric amplifiers in both construction and operation. Figure 2-47 is a functional diagram of a parametric down-converter.

The parametric frequency converter operates in the same manner as the parametric amplifier except that the sideband frequencies are not reapplied to the varactor. Therefore, the output is one or both of the sideband frequencies and is not the same as the input frequency. The output frequency is determined by the cavity used as an output. For example, the idler cavity in figure 2-47 could be replaced by a cavity that is resonant at the upper-sideband frequency (22 gigahertz) to produce an upper-sideband parametric up-converter. Since input and output signals are at different frequencies, the parametric frequency converter does not require a ferrite circulator. However, a ferrite isolator is used to isolate the converter from changes in source impedance.

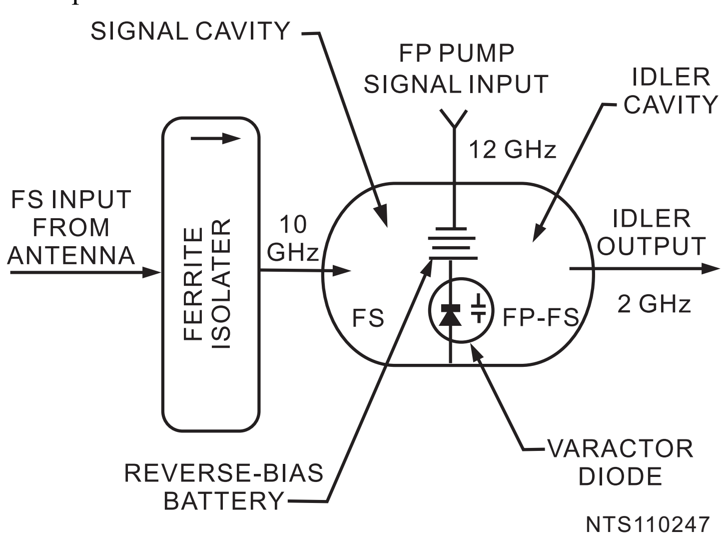


Figure 2-47.—Lower-sideband parametric down-converter.

Q-53. What limits the usefulness of high-gain, tunnel-diode frequency converters?

Q-54. The varactor is a pn junction that acts as what type of electronic device?

Q-55. The underlying principle of operation of the parametric amplifier is based on what property?

Q-56. What is the most important feature of the parametric amplifier?

Q-57. How is amplification achieved in the circuit shown in figure 2-43?

Q-58. What is the purpose of the pump in a parametric amplifier?

Q-59. The pump signal frequency must be of what value when compared to the input signal of a simple parametric amplifier?

Q-60. What is the primary difference between the pump signal of a simple parametric amplifier and the pump signal of a nondegenerative parametric amplifier?

Q-61. In a nondegenerative parametric amplifier the difference between the input frequency and the pump frequency is called what?

Bulk-Effect Semiconductors

BULK-EFFECT SEMICONDUCTORS are unlike normal pn-junction diodes in both construction and operation. Some types have no junctions and the processes necessary for operation occur in a solid block of semiconductor material. Other types have more than one junction but still use bulk-effect action. Bulk-effect devices are among the latest of developments in the field of microwave semiconductors and new applications are being developed rapidly. They seem destined to revolutionize the field of high-power, solid-state microwave generation because they can produce much larger microwave power outputs than any currently available pn-junction semiconductors.

Bulk-effect semiconductors are of two basic types: the transferred-electron devices and the avalanche transit-time devices.

TRANSFERRED-ELECTRON SEMICONDUCTORS.—The discovery that microwaves could be generated by applying a steady voltage across a chip of n-type gallium-arsenide (GaAs) crystal was made in 1963 by J.B. Gunn. The device is operated by raising electrons in the crystal to conduction-band energy levels that are higher than the level they normally occupy. The overall effect is called the transferred-electron effect.

In a gallium-arsenide semiconductor, empty electron conduction bands exist that are at a higher energy level than the conduction bands occupied by most of the electrons. Any electrons that do occupy the higher conduction band essentially have no mobility. If an electric field of sufficient intensity is applied to the semiconductor electrons, they will move from the low-energy conduction band to the high-energy conduction band and become essentially immobile. The immobile electrons no longer contribute to the current flow and the applied voltage progressively increases the rate at which the electrons move from the low band to the high band. As the curve in figure 2-48 shows, the maximum current rate is reached and begins to decrease even though the applied voltage continues to increase. The point at which the current on the curve begins to decrease is called the THRESHOLD. This point is the beginning of the negative-resistance region. Negative resistance is caused by electrons moving to the higher conduction band and becoming immobile.

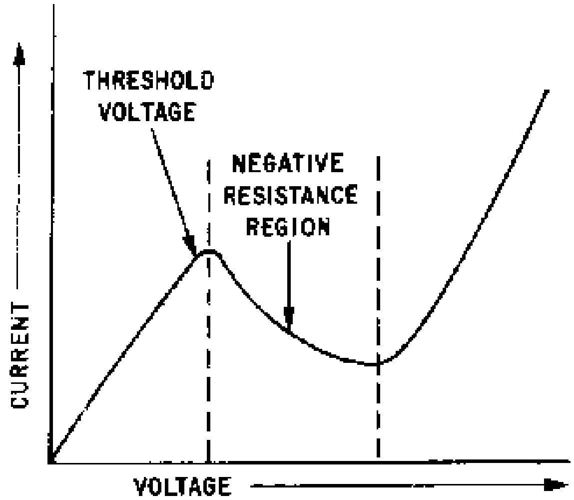


Figure 2-48.—Characteristic curve for a bulk-effect semiconductor.

If an increase in voltage is applied to a gallium-arsenide semiconductor, which is biased to operate in the negative-resistance region, it divides into regions of varying electric fields. A tiny region, known as a DOMAIN, forms that has an electric field of much greater intensity than the fields in the rest of the semiconductor. The applied voltage causes the domain to travel across the semiconductor chip from the cathode to the anode. The high field intensity of the domain is caused by the interaction of the slow electrons in the high-energy band and the faster electrons in the low-energy band. The electrons in the low-energy band travel faster than the moving domain and continually catch up during the transit from cathode to anode. When the fast electrons catch up to the domain, the high field intensity forces them into the higher band where they lose most of their mobility. This also causes them to fall behind the moving domain. Random scattering causes the electrons to lose some energy and drop back into the lower, faster, energy band and race again after the moving domain. The movement from the low-energy band to the high-energy band causes the electrons to bunch up at the back of the domain and to provide the electron-transfer energy that creates the high field intensity in the domain.

The domains form at or near the cathode and move across the semiconductor to the anode, as shown in figure 2-49. As the domain disappears at the anode, a new domain forms near the cathode and repeats the process.

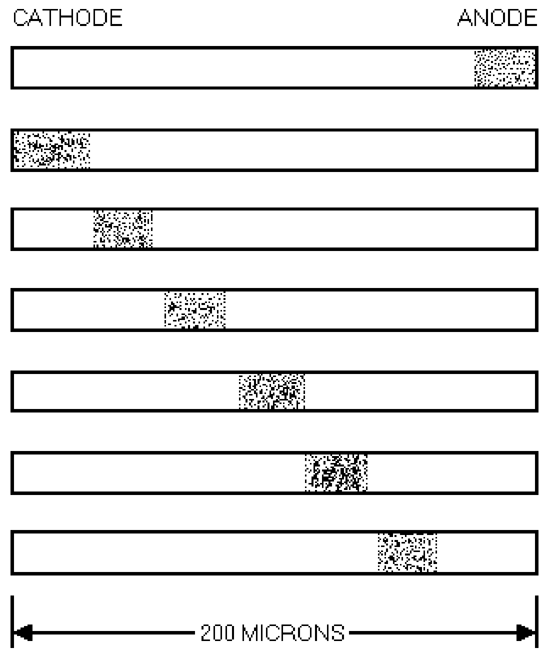


Figure 2-49.—Gallium-arsenide semiconductor domain movement.

The GUNN OSCILLATOR is a source of microwave energy that uses the bulk-effect, gallium-arsenide semiconductor. The basic frequency of a gunn oscillator is inversely proportional to the transit time of a domain across the semiconductor. The transit time is proportional to the length of semiconductor material, and to some extent, the voltage applied. Each domain causes a pulse of current at the output; thus, the output is a frequency determined by the physical length of the semiconductor chip.

The gunn oscillator can deliver continuous power up to about 65 milliwatts and pulsed outputs of up to about 200 watts peak. The power output of a solid chip is limited by the difficulty of removing heat from the small chip. Much higher power outputs have been achieved using wafers of gallium-arsenide as a single source.

AVALANCHE TRANSIT-TIME DIODES.—Avalanche transit-time diodes, also called IMPATT (Impact Avalanche and Transit-Time) diodes, are multilayer diodes of several different types used to generate microwave power. The earliest of the avalanche transit-time diodes consists of four layers in a pnin arrangement.The intrinsic (i) layer has neither p nor n properties.

The pn junction for the pnin diode, shown in figure 2-50, is strongly reverse biased to cause an avalanche in its depletion layer when the positive half cycle of a microwave signal is applied. The avalanche effect causes the electrons in the n region, which is very thin, to cross over to the intrinsic layer. The intrinsic layer is constructed so that the drift transit time causes the current to lag the signal voltage by more than 90 degrees at the desired frequency. Such a lag represents a negative resistance at the desired frequency. The pnin avalanche transit-time diode, when inserted in a microwave cavity with the proper dc bias, amplifies microwave signals introduced to the cavity.

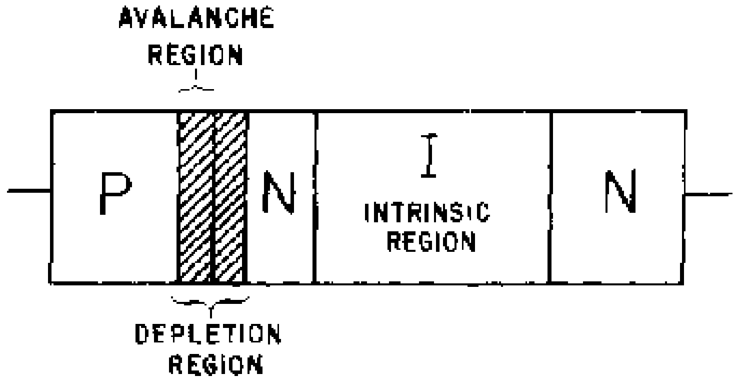


Figure 2-50.—Avalanche transit time for a pnin diode.

More recent research has shown that pin-junction diodes and simple pn-junction diodes can show negative resistance and amplification at microwave frequencies when they are reverse biased into an avalanche condition. The negative resistance in a simple pn-junction or pin diode is the result of a more complicated internal mechanism than in the pnin diode. The avalanche region and the drift region of the pnin diode are physically separate. Diodes of the pn and pin type must use the same physical region for both avalanche and drift-time control. In all types of avalanche transit-time diodes, the negative-resistance property causes dc bias energy to be absorbed by electrons in the avalanche process and given up to the applied microwave field.

Q-62. What is the output frequency of an upper-sideband parametric-frequency converter?

Q-63. What is the primary advantage of bulk-effect devices over normal pn-junction semiconductors?

Q-64. What happens to the electrons of a gallium-arsenide semiconductor when they move from the normal low-energy conduction band to the high-energy conduction band?

Q-65. The point on the current curve of a gallium-arsenide semiconductor at which it begins to exhibit negative resistance is called what?

Q-66. The domain in a gallium-arsenide semiconductor has what type of electrical field when compared to the other regions across the body of a semiconductor?

Q-67. What characteristic of a gunn oscillator is inversely proportional to the transit time of the domain across the semiconductor?

Q-68. What is the junction arrangement of the original avalanche transit-time diode?

Q-69. What causes dc bias energy to be absorbed by avalanche electrons and given up to the microwave field applied to an avalanche transit-time diode?

The Point-Contact Diode

POINT-CONTACT DIODES, commonly called CRYSTALS, are the oldest microwave semiconductor devices. They were developed during World War II for use in microwave receivers and are still in widespread use as receiver mixers and detectors.

Unlike the pn-junction diode, the point-contact diode depends on the pressure of contact between a point and a semiconductor crystal for its operation. Figure 2-51A and B, illustrate a point-contact diode. One section of the diode consists of a small rectangular crystal of n-type silicon. A fine beryllium-copper, bronze-phosphor, or tungsten wire called the CATWHISKER presses against the crystal and forms the other part of the diode. During the manufacture of the point contact diode, a relatively large current is passed from the catwhisker to the silicon crystal. The result of this large current is the formation of a small region of p-type material around the crystal in the vicinity of the point contact. Thus, a pn-junction is formed which behaves in the same way as a normal pn-junction.

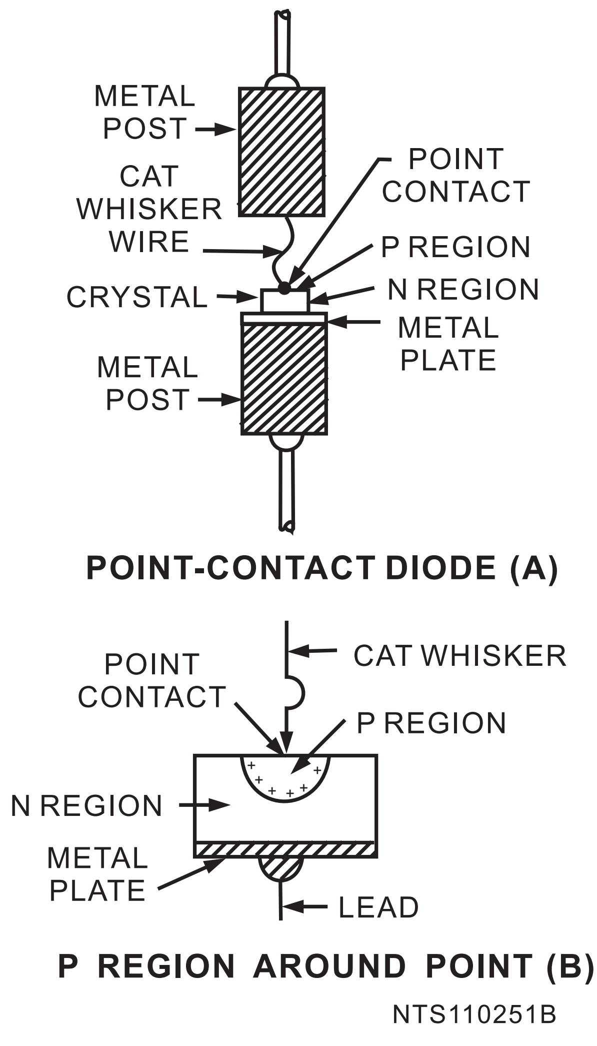


Figure 2-51A-B.—Point-contact diode. P REGION AROUND POINT.

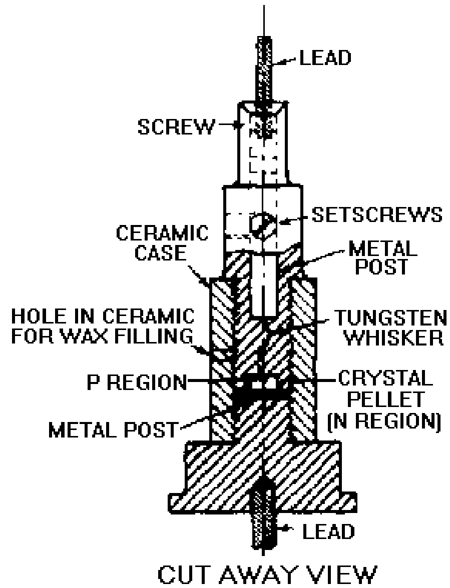


Figure 2-51C.—Point-contact diode. CUT AWAY VIEW.

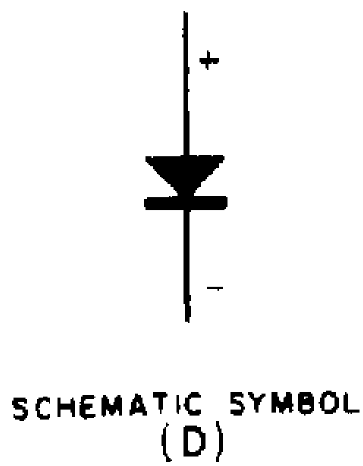


Figure 2-51D.—Point-contact diode. SCHEMATIC SYMBOL.

The pointed wire is used instead of a flat metal plate to produce a high-intensity electric field at the point contact without using a large external source voltage. It is not possible to apply large voltages across the average semiconductor because of the excessive heating.

The end of the catwhisker is one of the terminals of the diode. It has a low-resistance contact to the external circuit. A flat metal plate on which the crystal is mounted forms the lower contact of the diode with the external circuit. Both contacts with the external circuit are low-resistance contacts.

The characteristics of the point-contact diode under forward and reverse bias are somewhat different from those of the junction diode.

With forward bias, the resistance of the point-contact diode is higher than that of the junction diode. With reverse bias, the current flow through a point-contact diode is not as independent of the voltage applied to the crystal as it is in the junction diode. The point-contact diode has an advantage over the junction diode because the capacitance between the catwhisker and the crystal is less than the capacitance between the two sides of the junction diode. As such, the capacitive reactance existing across the point-contact diode is higher and the capacitive current that will flow in the circuit at high frequencies is smaller. A cutaway view of the entire point-contact diode is shown in figure 2-51C. The schematic symbol of a point-contact diode is shown in figure 2-51D.

Schottky Barrier Diode

The SCHOTTKY BARRIER DIODE is actually a variation of the point-contact diode in which the metal semiconductor junction is a surface rather than a point contact. The large contact area, or barrier, between the metal and the semiconductor in the Schottky barrier diode provides some advantages over the point-contact diode. Lower forward resistance and lower noise generation are the most important advantages of the Schottky barrier diode. The applications of the Schottky barrier diode are the same as those of the point-contact diode. The low noise level generated by Schottky diodes makes them especially suitable as microwave receiver detectors and mixers.

The Schottky barrier diode is sometimes called the HOT-ELECTRON or HOT-CARRIER DIODE because the electrons flowing from the semiconductor to the metal have a higher energy level than the electrons in the metal. The effect is the same as it would be if the metal were heated to a higher temperature than normal. Figure 2-52 is an illustration of the construction of a Schottky barrier diode.

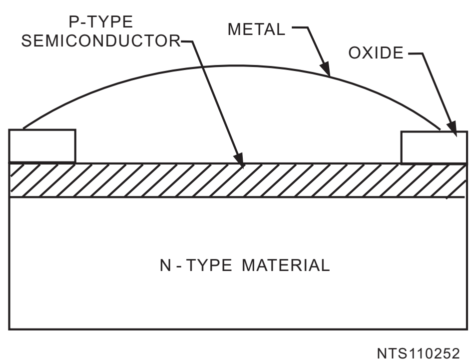


Figure 2-52.—Schottky-barrier diode.

PIN Diodes

The pin diode consists of two narrow, but highly doped, semiconductor regions separated by a thicker, lightly-doped material called the intrinsic region. As suggested in the name, pin, one of the heavily doped regions is p-type material and the other is n-type. The same semiconductor material, usually silicon, is used for all three areas. Silicon is used most often for its power-handling capability and because it provides a highly resistive intrinsic (i) region. The pin diode acts as an ordinary diode at frequencies up to about 100 megahertz, but above this frequency the operational characteristics change.

The large intrinsic region increases the transit time of electrons crossing the region. Above 100 megahertz, electrons begin to accumulate in the intrinsic region. The carrier storage in the intrinsic region causes the diode to stop acting as a rectifier and begin acting as a variable resistance. The equivalent circuit of a pin diode at microwave frequencies is shown in figure 2-53A. A resistance versus voltage characteristic curve is shown in figure 2-53B.

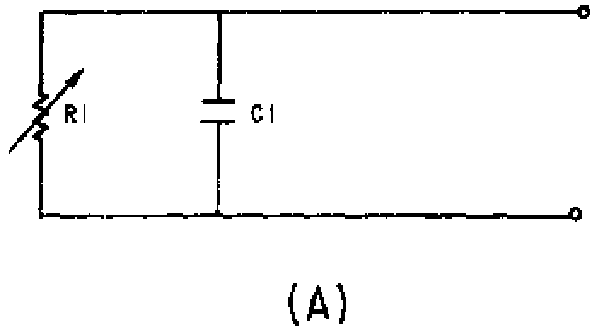


Figure 2-53A.—Diode equivalent circuit (pin).

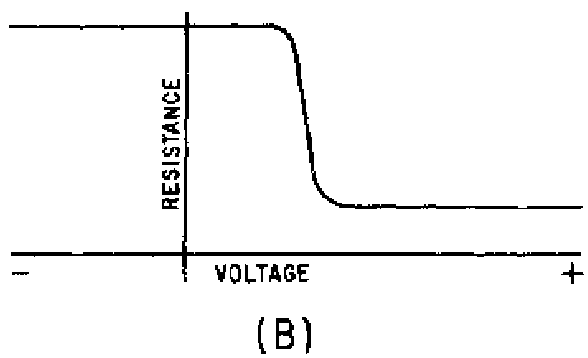


Figure 2-53B.—Diode equivalent circuit (pin).

When the bias on a pin diode is varied, the microwave resistance changes from a typical value of 6 kilohms under negative bias to about 5 ohms when the bias is positive. Thus when the diode is mounted across a transmission line or waveguide, the loading effect is insignificant while the diode is reverse biased, and the diode presents no interference to power flow. When the diode is forward biased, the resistance drops to approximately 5 ohms and most power is reflected. In other words, the diode acts as a switch when mounted in parallel with a transmission line or waveguide. Several diodes in parallel can switch power in excess of 150 kilowatts peak. The upper power limit is determined by the ability of the diode to dissipate power. The upper frequency limit is determined by the shunt capacitance of the pn junction, shown as C1 in figure 2-53A. Pin diodes with upper limit frequencies in excess of 30 gigahertz are available.

Q-70. During the manufacture of a point-contact diode, what is the purpose of passing a relatively large current from the catwhisker to the silicon crystal?

Q-71. What is the capacitive reactance across a point-contact diode as compared to a normal junction diode?

Q-72. What are the most important advantages of the Schottky barrier diode?

Q-73. At frequencies above 100 megahertz, the intrinsic (i) region causes a pin diode to act as what?

Q-74. The pin diode is primarily used for what purpose?

Microwave Transistors

Transistors, like vacuum tubes, have had a very limited application in the microwave range. Many of the same problems encountered with vacuum tubes, such as transit-time effects, also limit the upper frequency range of transistors. However, research in the area of microwave transistors, and especially MICROWAVE INTEGRATED CIRCUITS (ICs), is proceeding rapidly.

GALLIUM-ARSENIDE FET AMPLIFIERS have been developed which provide low-noise amplification up to about 30 dB in the 7- to 18-gigahertz range. The power output of many of these amplifiers is relatively low, approximately 20 to 200 milliwatts, but that is satisfactory for many microwave applications. Research has extended both the frequency range and the power output of gallium-arsenide FET amplifiers to frequencies as high as 26.5 gigahertz and power levels in excess of 1 watt in multistage amplifiers.

SILICON BIPOLAR-TRANSISTOR AMPLIFIERS in integrated circuit form have been developed that provide up to 40 watts peak power in the 1- to 1.5-gigahertz range. Other types of microwave transistor amplifiers combined into multistage modules are capable of providing power outputs approaching 100 watts.

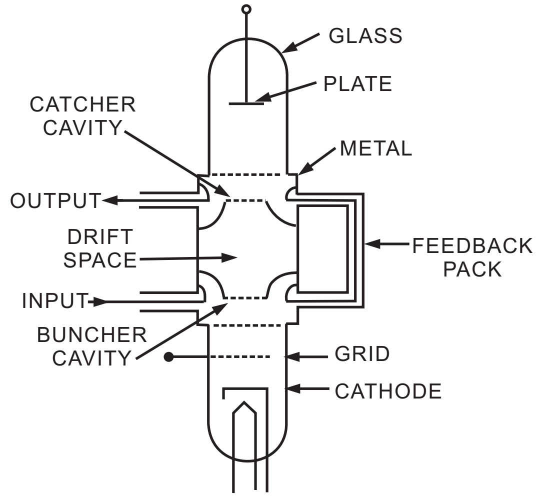
Microwave transistor amplifiers, because of their stability, light weight, and long life, are rapidly replacing microwave tubes in the first stages of high-powered radar and communications transmitters. In the future new systems will be almost completely solid state.

SUMMARY

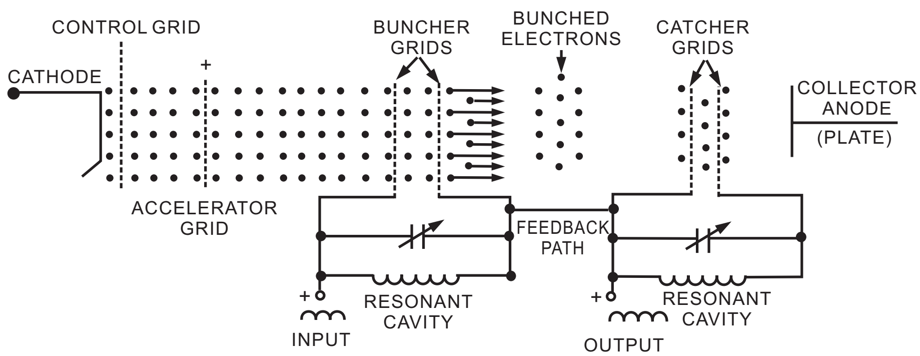
The information that follows summarizes the important points presented in this chapter.

The use of microwave frequencies forced the development of special tubes to offset the limitations caused by interelectrode capacitance, lead inductance, and electron transit-time effects in conventional tubes. Microwave tubes, such as the klystron and twt, take advantage of transit-time effects through the use of VELOCITY MODULATION to amplify and generate microwave energy.

The **KLYSTRON** is a velocity-modulated tube which may be used as an amplifier or oscillator. The klystron, when used as an amplifier, requires at least two resonant cavities, the buncher and the catcher. A diagram of a basic klystron is shown at the right.



A

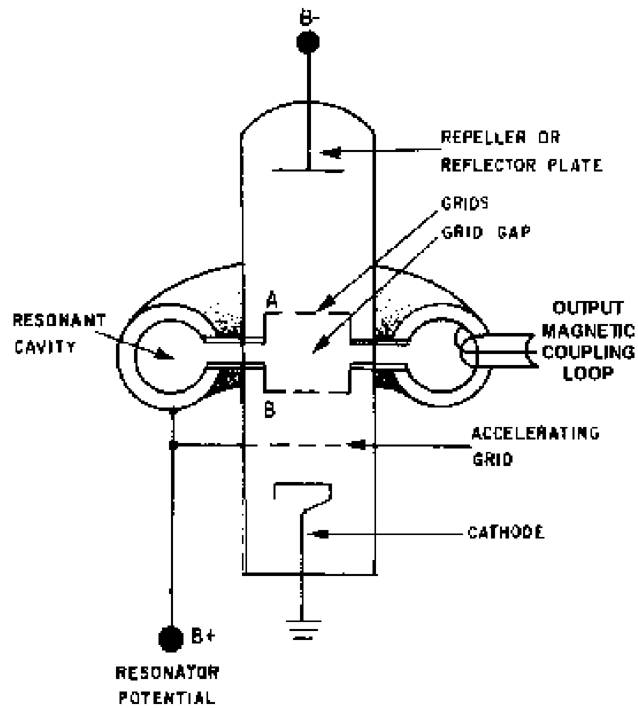


NOTE: RESONANT CAVITIES SHOWN AS CONVENTIONAL PARALLEL RESONANT CIRCUIT

B

NTS1102I-2

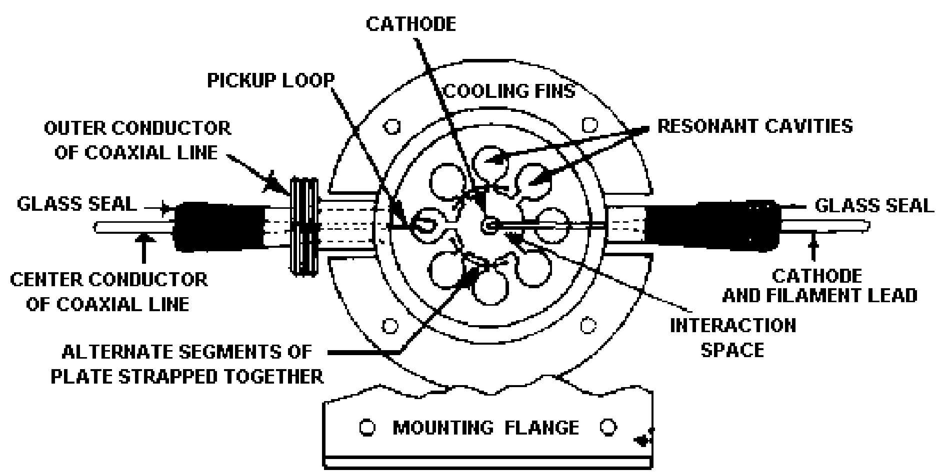
The **REFLEX KLYSTRON**, shown at the right, is used only as an oscillator and uses only one cavity to bunch and collect the electrons. The frequency is determined by the size and shape of the cavity. The reflex klystron has several possible modes of operation which are determined by electron transit time. Electron transit time is controlled by the REPELLER voltage.



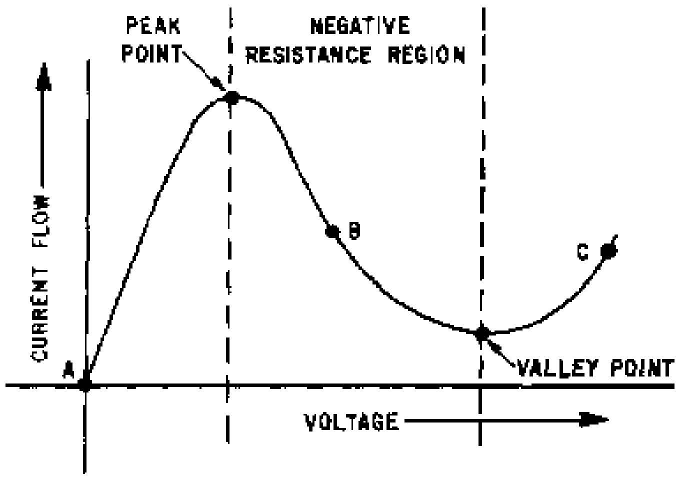
The **TWT** is a wide-bandwidth, velocity-modulated tube used primarily as an amplifier. The electron beam is bunched by a signal applied to the HELIX. The bunching causes an energy transfer from the electron beam to the traveling wave on the helix.

The **MAGNETRON** is a **DIODE OSCILLATOR** capable of delivering microwave energy at very high power levels. Three fields exist within a magnetron that influence operation: (1) the DC ELECTRIC FIELD between the anode and cathode; (2) the AC ELECTRIC FIELD produced by the oscillating resonant cavities and on the same plane as the dc field; and (3) the MAGNETIC FIELD produced by the permanent magnet which is perpendicular to the dc electric field.

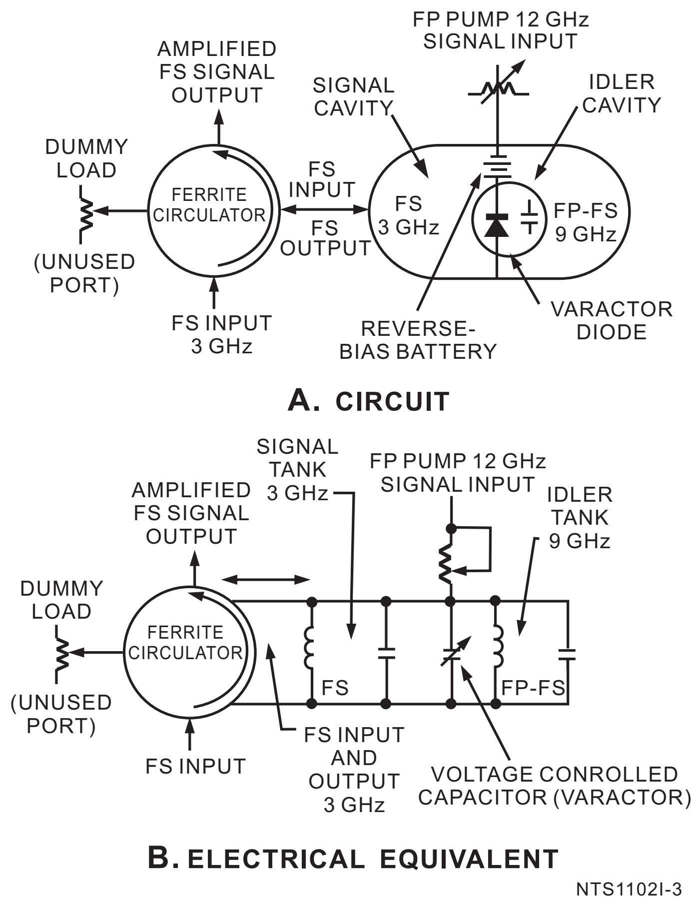
Magnetrons are of two basic types, the NEGATIVE-RESISTANCE MAGNETRON and the ELECTRON-RESONANCE MAGNETRON. A diagram of a magnetron is shown at the right.



SOLID-STATE MICROWAVE DEVICES are becoming increasingly widespread in microwave equipment with new developments almost daily. Most of the currently available solid-state devices are two-terminal diodes with the capability to generate or amplify microwave energy. Many of the solid-state devices, such as the TUNNEL DIODE and the BULK-EFFECT DIODE, apply the property of NEGATIVE RESISTANCE to amplify microwave signals or generate microwave energy. A characteristic curve illustrating the negative-resistance property of the tunnel diode is shown at the right.



The **VARACTOR** is a two-terminal diode that acts as a variable capacitance and is the active element of PARAMETRIC AMPLIFIERS. The parametric amplifier is a low-noise microwave amplifier that uses variable reactance to amplify microwave signals. The illustration shows an example of a NONDEGENERATIVE PARAMETRIC AMPLIFIER.



ANSWERS TO QUESTIONS Q1. THROUGH Q74.

A-1. *Impedance decreases.*

A-2. *Degenerative feedback.*

A-3. *Transit time causes the grid voltage and plate current to be out of phase.*

A-4. *Transit time.*

A-5. *Velocity.*

A-6. *The electron will be accelerated.*

A-7. *By alternately speeding up or slowing down the electrons.*

A-8. *The buncher grids.*

A-9. *There is no effect.*

A-10. *The frequency period of the buncher grid signal.*

A-11. *Velocity modulation.*

A-12. The accelerator grid and the buncher grids.

A-13. The catcher cavity.

A-14. Amplifier.

A-15. Intermediate cavities between the input and output cavities.

A-16. A large negative pulse is applied to the cathode.

A-17. The middle cavity.

A-18. The bandwidth decreases.

A-19. Stagger tuning.

A-20. The reflector or repeller.

A-21. Velocity.

A-22. Three-quarter cycle.

A-23. Mode 2.

A-24. Power is reduced.

A-25. The half-power points of the mode.

A-26. Voltage amplification.

A-27. Used to focus the electrons into a tight beam.

A-28. The directional couplers are not physically connected to the helix.

A-29. The traveling wave must have a forward velocity equal to or less than the speed of the electrons in the beam.

A-30. The helix.

A-31. Helix.

A-32. A magnetic field.

A-33. Anode or plate.

A-34. The resonant cavities.

A-35. The permanent magnet.

A-36. The critical value of field strength.

A-37. Circular.

A-38. The negative-resistance magnetron has a split plate.

A-39. The application of the proper magnetic field.

A-40. To reduce the effects of filament bombardment.

A-41. Rising-sun block.

A-42. Series.

A-43. Working electrons.

A-44. Greater power output.

A-45. Loops and slots.

A-46. Inductive.

A-47. A cookie-cutter tuner.

A-48. Baking in.

A-49. The tunneling action.

A-50. The tuned circuit or cavity frequency.

A-51. To increase the stability.

A-52. Prevent feedback to the tuned input circuit.

A-53. Stability problems.

A-54. Variable capacitor.

A-55. Reactance.

A-56. The low-noise characteristic.

A-57. By varying the amount of capacitance in the circuit.

A-58. Supplies the electrical energy required to vary the capacitance.

A-59. Exactly double the input frequency.

A-60. The pump signal of a nondegenerative parametric amplifier is higher than twice the input signal.

A-61. Idler- or lower-sideband frequency.

A-62. The sum of the input frequency and the pump frequency.

A-63. Larger microwave power outputs.

A-64. The electrons become immobile.

A-65. Threshold.

A-66. A field of much greater intensity.

A-67. The frequency.

A-68. Pnin.

A-69. The negative-resistance property.

A-70. To form a small region of p-type material.

A-71. Lower.

A-72. Lower forward resistance and low noise.

A-73. Variable resistance.

A-74. A switching device.

CHAPTER 3

MICROWAVE ANTENNAS

LEARNING OBJECTIVES

Upon completion of this section the student will be able to:

1. Explain the basic characteristics of coupling, directivity, reciprocity, and efficiency in microwave antennas.
2. Describe the construction and basic theory of operation of reflector antennas and horn radiators.
3. Explain construction and operation of microwave lens antennas.
4. Describe the construction and theory of operation of driven and parasitic antenna arrays.
5. Explain the basic operation and applications of frequency-sensitive antennas.

INTRODUCTION

In this chapter you will study the general characteristics of microwave antennas that are widely used in radar and communications applications. The basic principles of operation of microwave antennas are similar to those of antennas used at lower frequencies. You might want to review the principles presented in *NEETS*, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*, at this time. Pay particular attention to basic antenna principles in chapter 4 for a review of microwave antennas.

Antennas are devices used to radiate electromagnetic energy into space. The characteristics of transmitting and receiving antennas are similar, so a good transmitting antenna is often a good receiving antenna. A single antenna performs both functions in many modern applications.

ANTENNA CHARACTERISTICS

Since the operating principles of low-frequency and microwave antennas are essentially the same, the electrical characteristics are also very similar. You will need a fundamental knowledge of radar and communications antenna electrical theory in your shipboard antenna maintenance work. Antenna theory is primarily a design consideration of antenna size and shape requirements that depend on the frequency used. A brief description of antenna electrical characteristics is sufficient for the needs of most students of electronics.

Antenna Efficiency

The effectiveness of an antenna depends upon its ability to couple or radiate energy into the air. An efficient antenna is one which wastes very little energy during the radiation process. The efficiency of an antenna is usually referred to as the POWER GAIN or POWER RATIO as compared to a standard reference antenna. The power gain of an antenna is a ratio of the radiated power to that of the reference antenna, which is usually a basic dipole. Both antennas must be fed rf energy in the same manner and must be in the same position when the energy is radiated. The power gain of a single dipole without a reflector is unity (one). An array of several dipoles in the same position as the single dipole, and fed with the same line, has a power gain of more than one.

The effectiveness of an entire transmitting/ receiving system depends largely on impedance matching between the elements of the system. Impedance matching is particularly critical at the antenna connection. If a good impedance match is maintained between the system and the antenna throughout the operating frequency band, power transfer to and from the antenna is always maximum. The transmission line or waveguide used to transport energy to and from the antenna should have a characteristic impedance equal to that of the antenna. A proper impedance match allows all available power to be absorbed and radiated by the antenna without reflections back down the line.

If you have a transmission line or waveguide with an impedance mismatch at the termination, standing waves are set up by the reflections. Standing waves cause losses in the form of unwanted radiations, heat losses in transmission lines, and arcing in waveguides.

The STANDING-WAVE RATIO, abbreviated swr, is a way to measure the degree of mismatch between the transmission line and its load. The swr can be expressed as a ratio of the maximum and minimum values of the current or voltage in the standing waves that are set up on the lines as follows:

$$vswr = \frac{E_{max}}{E_{min}}$$

or

$$iswr = \frac{I_{max}}{I_{min}}$$

A transmission line or waveguide approaches a perfectly matched condition when the swr approaches a value of 1. A ratio that is a little higher than 1 is usually acceptable in practical applications.

Measurement of swr is the only practical method of detecting an impedance mismatch between a transmitting/receiving system and its antenna. As such, the system swr is an important indication of the overall efficiency of the system during operation.

The line impedance can usually be matched to the antenna at only one frequency. However, the swr will NOT become too high if the antenna is used over a small range of frequencies and the line is matched to the center frequency.

Antenna Directivity

You can divide antennas into two general classes based on directivity, omnidirectional and directional. OMNIDIRECTIONAL antennas radiate and receive energy from all directions at once (SPHERICAL WAVEFRONT). They are seldom used in modern radar systems as the primary antenna, but are commonly used in radio equipment and iff (identification friend or foe) receivers. DIRECTIONAL antennas radiate energy in LOBES (or BEAMS) that extend outward from the antenna in either one or two directions. The radiation pattern contains small minor lobes, but these lobes are weak and normally have little effect on the main radiation pattern. Directional antennas also receive energy efficiently from only one or two directions, depending upon whether it is unidirectional or bidirectional.

Directional antennas have two characteristics that are important to you in radar and communications systems. One is DIRECTIVITY and the other is POWER GAIN. The directivity of an antenna refers to the NARROWNESS of the radiated beam. If the beam is NARROW in either the horizontal or vertical plane, the antenna has a high degree of directivity in that plane. An antenna may be designed for high directivity in one plane only or in both planes, depending on the application. The power gain of an antenna increases as the degree of directivity increases because the power is concentrated into a narrow beam and less power is required to cover the same distance.

Since microwave antennas are predominantly unidirectional, the examples you will study in this chapter are all of the unidirectional type.

Reciprocity

You read in this chapter that an antenna is able to both transmit and receive electromagnetic energy. This is known as RECIPROCITY. Antenna reciprocity is possible because antenna characteristics are essentially the same regardless of whether an antenna is transmitting or receiving electromagnetic energy. Reciprocity allows most radar and communications systems to operate with only one antenna. An automatic switch, called a DUPLEXER, connects either the transmitter or the receiver to the antenna at the proper time. Duplexer operation will be explained in later *NEETS* modules dealing with radar and communications systems. Because of the reciprocity of antennas, this chapter will discuss antennas from the viewpoint of the transmitting cycle. However, you should understand that the same principles apply on the receiving cycle.

Radar Fundamentals

Radio, television, radar, and the human eye have much in common because they all process the same type of electromagnetic energy. The major difference between the light processed by the human eye and the radio-frequency energy processed by radio and radar is frequency. For example, radio transmitters send out signals in all directions. These signals can be detected by receivers tuned to the same frequency. Radar works somewhat differently because it uses reflected energy (echo) instead of directly transmitted energy. The echo, as it relates to sound, is a familiar concept to most of us. An experienced person can estimate the distance and general direction of an object causing a sound echo. Radar uses microwave electromagnetic energy in much the same way.

Radar transmits microwave energy that reflects off an object and returns to the radar. The returned portion of the energy is called an ECHO, as it is in sound terminology. It is used to determine the direction and distance of the object causing the reflection. Determination of direction and distance to an object is the primary function of most radar systems.

Telescopes and radars, in terms of locating objects in space, have many common problems. Both have a limited field of view and both require a geographic reference system to describe the position of an object (target). The position of an object viewed with a telescope is usually described by relating it to a familiar object with a known position. Radar uses a standard system of reference coordinates to describe the position of an object in relation to the position of the radar. Normally ANGULAR measurements are made from true north in an imaginary flat plane called the HORIZONTAL PLANE. All angles in the UP direction are measured in a second imaginary plane perpendicular to the horizontal plane called the VERTICAL PLANE. The center of the coordinate system is the radar location. As shown in figure 3-1, the target position with respect to the radar is defined as 60 degrees true, 10 degrees up, and 10 miles distant. The line directly from the radar to the target is called the LINE OF SIGHT. The distance from point 1 to point 2, measured along the line of sight, is called TARGET RANGE. The angle between the horizontal plane and the line of sight is known as the ELEVATION ANGLE. The angle measured in a clockwise direction in the horizontal plane between true north and the line of sight is known as BEARING (sometimes referred to as AZIMUTH). These three coordinates of range, bearing, and elevation determine the location of the target with respect to the radar.

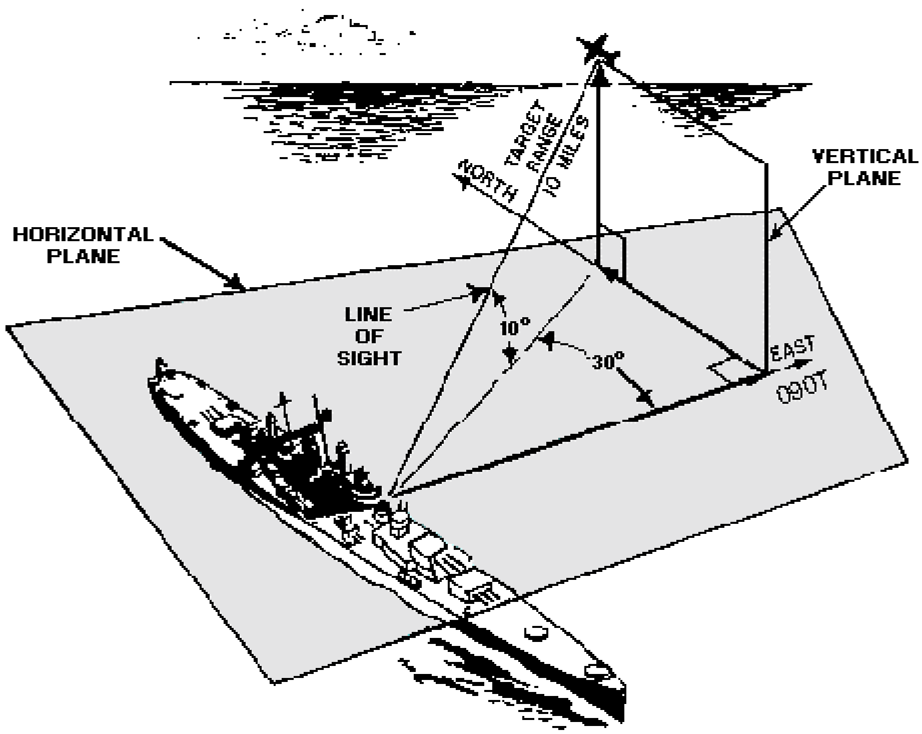


Figure 3-1.—Radar target position.

Bearing and elevation angles are determined by measuring the angular position of the radar antenna (the transmitted beam) when it is pointing directly at the target. Range is more difficult to determine because it cannot be directly measured. The radar system is designed to measure range as a function of time. Since the speed of electromagnetic energy is the same as the speed of light, range is determined by measuring the time required for a pulse of energy to reach the target and return to the radar. Because the speed of the pulse is known, the two-way distance can be determined by multiplying the time by the speed of travel. The total must be divided by two to obtain the one-way range because the time value used initially is the time required for the pulse to travel to the target and return.

The discussion of microwave antennas in this chapter requires only the most basic understanding of radar concepts! Radar fundamentals will be discussed in more detail in a later *NEETS* module.

Q-1. Microwave antennas and low-frequency antennas are similar in what ways?

Q-2. What term is used to express the efficiency of an antenna?

Q-3. What term is used to express the measurement of the degree of mismatch between a line and its load?

Q-4. What type of antenna radiates in and receives energy from all directions at once?

Q-5. What is the term that is used to describe narrowness in the radiated beam of an antenna?

Q-6. What characteristic allows the same antenna to both transmit and receive?

REFLECTOR ANTENNAS

A spherical wavefront (one in which the energy spreads out in all directions) spreads out as it travels away from the antenna and produces a pattern that is not very directional. A wavefront that exists in only one plane does not spread because all of the wavefront moves forward in the same direction. For an antenna to be highly directive, it must change the normally spherical wavefront into a plane wavefront. Many highly directive microwave antennas produce a plane wavefront by using a reflector to focus the radiated energy. The PARABOLIC REFLECTOR is most often used for high directivity.

Microwaves travel in straight lines as do light rays. They can also be focused and reflected just as light rays can, as illustrated by the antenna shown in figure 3-2. A microwave source is placed at focal point F. The field leaves this antenna as a spherical wavefront. As each part of the wavefront reaches the reflecting surface, it is phase-shifted 180 degrees. Each part is then sent outward at an angle that results in all parts of the field traveling in parallel paths. Because of the special shape of a parabolic surface, all paths from F to the reflector and back to line XY are the same length. Therefore, when the parts of the field are reflected from the parabolic surface, they travel to line XY in the same amount of time.

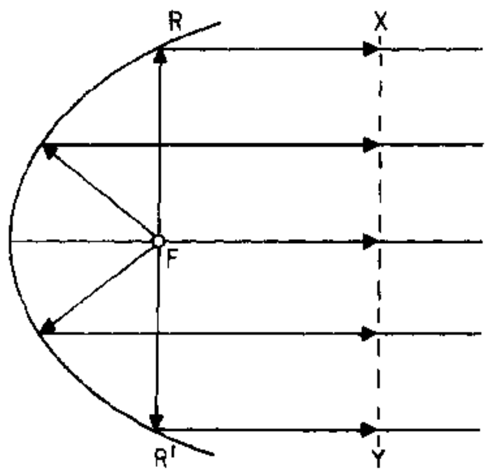


Figure 3-2.—Parabolic reflector radiation.

If a dipole is used as the source of transmission, energy will be radiated from the antenna into space as well as toward the reflector. Energy which is not directed toward the paraboloid has a wide-beam characteristic which will destroy the narrow pattern of the parabolic reflector. However, a HEMISPHERICAL SHIELD (not shown) may be used to direct most of the radiation toward the parabolic surface and thus prevent the destruction of the narrow pattern. Direct radiation into space is eliminated, the beam is made sharper, and more power is concentrated in the beam. Without the shield, some of the radiated field would leave the radiator directly. Since this part of the field that would leave the radiator would not be reflected, it would not become a part of the main beam and could serve no useful purpose.

In figure 3-3 the radiation pattern of a paraboloid reflector contains a major lobe and several minor lobes. The major lobe is directed along the axis of revolution. Very narrow beams are possible with this type of reflector. Figure 3-4 illustrates the basic paraboloid reflector.

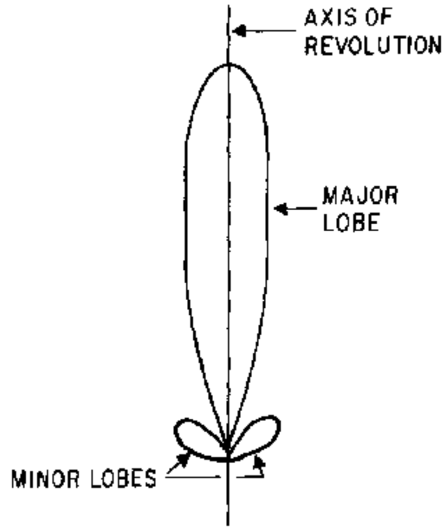


Figure 3-3.—Parabolic radiation pattern.

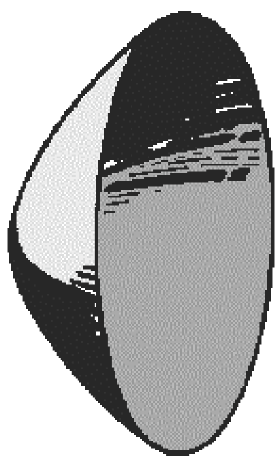


Figure 3-4.—Basic paraboloid reflector.

You may see several variations of the basic paraboloid reflector used to produce different beam shapes required by special applications. The basic characteristics of the most commonly used paraboloids are presented in the following paragraphs.

Truncated Paraboloid

Figure 3-5A, shows a TRUNCATED PARABOLOID. Since the reflector is parabolic in the horizontal plane, the energy is focused into a narrow beam. With the reflector TRUNCATED (cut) so that it is shortened vertically, the beam spreads out vertically instead of being focused. This fan-shaped beam is used in radar detection applications for the accurate determination of bearing. Since the beam is spread vertically, it will detect aircraft at different altitudes without changing the tilt of the antenna. The truncated paraboloid also works well for surface search radar applications to compensate for the pitch and roll of the ship.

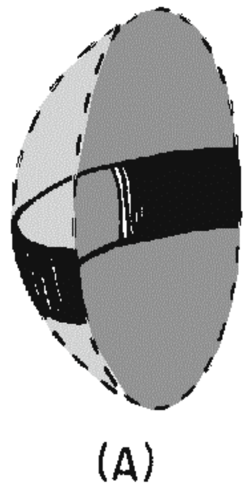


Figure 3-5A.—Truncated paraboloid.

The truncated paraboloid may be used in target height-finding systems if the reflector is rotated 90 degrees, as shown in figure 3-5B. Since the reflector is now parabolic in the vertical plane, the energy is focused vertically into a narrow beam. If the reflector is truncated, or cut, so that it is shortened horizontally, the beam will spread out horizontally instead of being focused. Such a fan-shaped beam is used to accurately determine elevation.

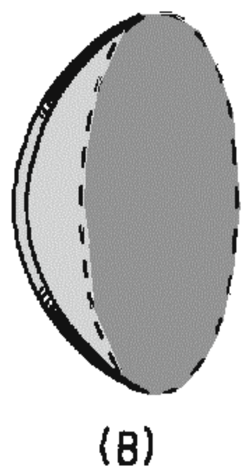


Figure 3-5B.—Truncated paraboloid.

Orange-Peel Paraboloid

A section of a complete circular paraboloid, often called an ORANGE-PEEL REFLECTOR because of its orange-peel shape, is shown in figure 3-6. Since the reflector is narrow in the horizontal plane and wide in the vertical plane, it produces a beam that is wide in the horizontal plane and narrow in the vertical plane. In shape, the beam resembles a huge beaver tail. The microwave energy is sent into the parabolic reflector by a horn radiator (not shown) which is fed by a waveguide. The horn radiation pattern covers nearly the entire shape of the reflector, so almost all of the microwave energy strikes the reflector and very little escapes at the sides. Antenna systems which use orange-peel paraboloids are often used in height-finding equipment.

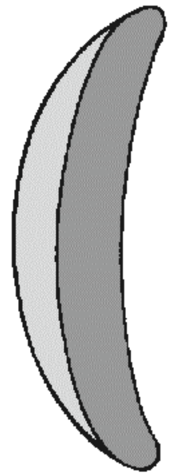


Figure 3-6.—Orange-peel paraboloid.

Cylindrical Paraboloid

When a beam of radiated energy that is noticeably wider in one cross-sectional dimension than in another is desired, a cylindrical paraboloidal section which approximates a rectangle can be used. Figure 3-7 illustrates such an antenna. A PARABOLIC CYLINDER has a parabolic cross section in just one dimension which causes the reflector to be directive in one plane only. The cylindrical paraboloid reflector is fed either by a linear array of dipoles, a slit in the side of a waveguide, or by a thin waveguide radiator. It also has a series of focal points forming a straight line rather than a single focal point. Placing the radiator, or radiators, along this focal line produces a directed beam of energy. As the width of the parabolic section is changed, different beam shapes are obtained. You may see this type of antenna system used in search radar systems and in ground control approach (gca) radar systems.

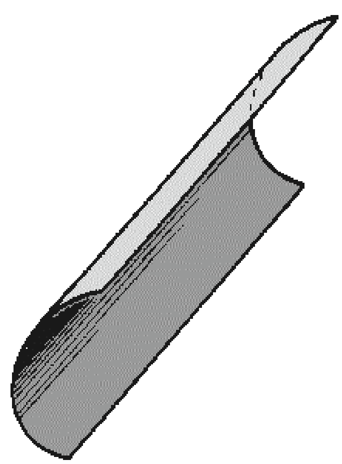


Figure 3-7.—Cylindrical paraboloid.

Corner Reflector

The CORNER-REFLECTOR ANTENNA consists of two flat conducting sheets that meet at an angle to form a corner, as shown in figure 3-8. The corner reflector is normally driven by a HALF-WAVE RADIATOR located on a line which bisects the angle formed by the sheet reflectors.

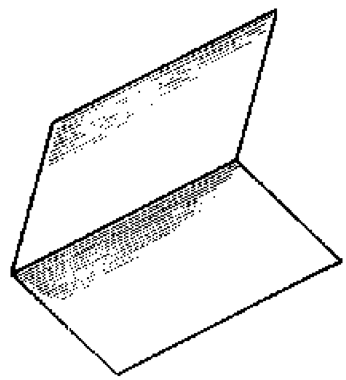


Figure 3-8.—Corner reflector.

Q-7. What type of reflector is most often used in directive antennas?

Q-8. Microwaves can be focused and reflected in the same way as what other type of waves?

Q-9. How many major lobes are radiated by a parabolic reflector?

Q-10. A horizontally truncated paraboloid antenna is used for what purpose?

Q-11. The beam from a horizontally positioned cylindrical paraboloid is narrow in what plane?

HORN RADIATORS

Like parabolic reflectors, you can use HORN RADIATORS to obtain directive radiation at microwave frequencies. Because they do not use resonant elements, horns have the advantage of being useful over a wide frequency band.

The operation of a horn as an rf radiating device is similar to that of an automobile horn radiating sound waves. However, the throat of an automobile horn usually is sized much smaller than the sound wavelengths for which it is used. The throat of the rf radiating horn is sized to be comparable to the wavelength being used.

Horn radiators are used with waveguides because they serve both as an impedance-matching device and as a directional radiator. Horn radiators may be fed by coaxial and other types of lines.

Horn radiators are constructed in a variety of shapes, as illustrated in figure 3-9. The shape of the horn determines the shape of the field pattern. The ratio of the horn length to the size of its mouth determines the beam angle and directivity. In general, the larger the mouth of the horn, the more directive is the field pattern.

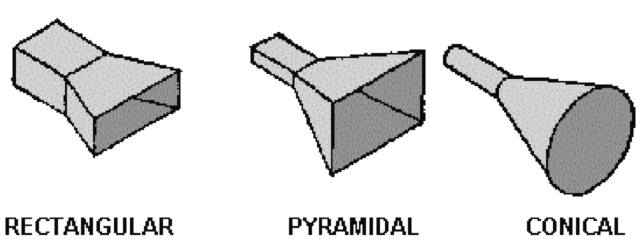


Figure 3-9.—Horn radiators.

LENS ANTENNAS

With a LENS ANTENNA you can convert spherically radiated microwave energy into a plane wave (in a given direction) by using a point source (open end of the waveguide) with a COLLIMATING LENS. A collimating lens forces all radial segments of the spherical wavefront into parallel paths. The point source can be regarded as a gun which shoots the microwave energy toward the lens. The point source is often a horn radiator or a simple dipole antenna.

Waveguide Type

The WAVEGUIDE-TYPE LENS is sometimes referred to as a conducting-type. It consists of several parallel concave metallic strips which are placed parallel to the electric field of the radiated energy fed to the lens, as shown in figure 3-10A and 3-10B. These strips act as waveguides in parallel for the incident (radiated) wave. The strips are placed slightly more than a half wavelength apart.

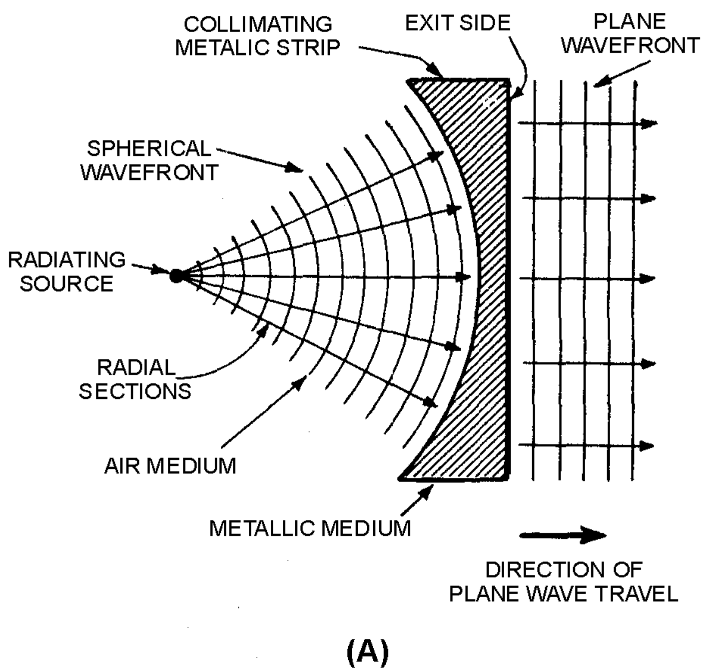
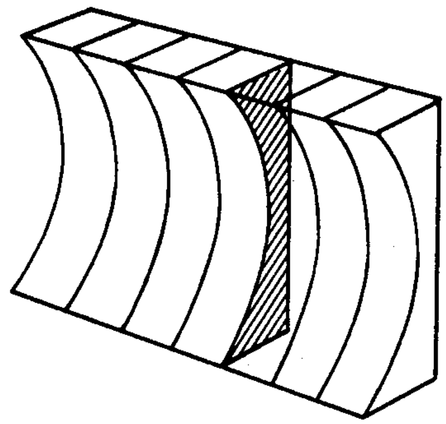


Figure 3-10A.—Waveguide lens.



THREE DIMENSIONAL VIEW OF PARALLEL-PLATE LENS

(B)

Figure 3-10B.—Waveguide lens.

The radiated energy consists of an infinite number of RADIAL SECTIONS (RAYS). Each of the radial sections contains mutually perpendicular E and H lines and both are perpendicular to the direction of travel. Because each of the radial sections travels in a different direction, the point source, in itself, has poor directivity. The purpose of the lens is to convert the input spherical microwave segment (which consists of all of the radial sections) into parallel (collimated) lines in a given direction at the exit side of the lens. The focusing action of the lens is accomplished by the refracting qualities of the metallic strips. The collimating effect of the lens is possible because the velocity of electromagnetic energy propagation through metals is greater than its velocity through air. Because of the concave construction of the lens, wavefronts arriving near the ends of the lens travel farther in the same amount of time than do those at the center. Thus, the wavefront emerging from the exit side of the lens appears as a plane wave. It consists of an infinite number of parallel sections (with both the E field and H field components) mutually perpendicular to the direction of travel.

Delay lens

Another type of lens that you may see is the DIELECTRIC or METALLIC DELAY LENS shown in figure 3-11. The delay lens, as its name implies, slows down the phase propagation (velocity) as the wave passes through the lens. The delay lens is convex and is constructed of dielectric material. The delay in the phase of the wave passing through the lens is determined by the DIELECTRIC CONSTANT (REFRACTIVE INDEX) of the material. In most cases, artificial dielectrics, consisting of conducting rods or spheres that are small compared to the wavelength, are used. (Artificial dielectrics are of three-dimensional construction and act as a dielectric to electromagnetic waves.) In this case the inner portion of the transmitted wave is decelerated for a longer interval of time than the outer portions. The delay causes the radiated wave to be collimated.

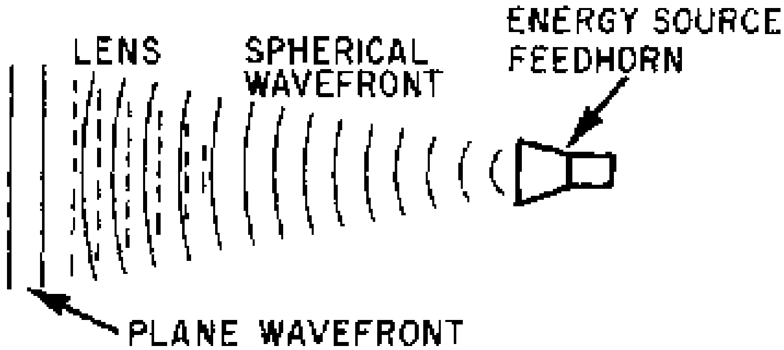


Figure 3-11.—Delay-type lens.

Loaded Microwave Lens

The LOADED MICROWAVE LENS, shown in figure 3-12, is a multi-cellular array of thousands of cells. Each cell contains a slow-wave (delayed), serrated-metal, plastic-supported waveguide element which acts as a phase-controlling device. A loaded lens can focus microwave energy in much the same way as the waveguide type. The reason is that the speed of propagation is higher in the region between parallel plates than in free space. The parallel plates support the cells.

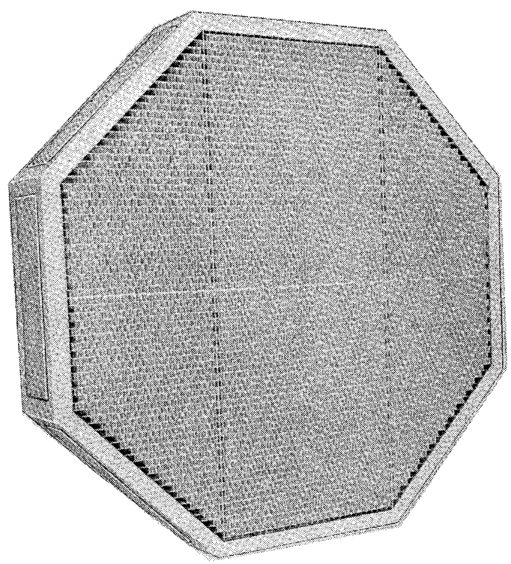


Figure 3-12.—Loaded lens.

The lens shown in figure 3-12 has an egg-crate appearance because it is really two lenses occupying the same volume. Vertical plates make up a lens that focuses a vertically polarized beam, and horizontal plates handle beams which are horizontally polarized. In other words, this type of construction can be used in multiple-beam applications where the polarization of the beams is different.

Q-12. What is the purpose of a collimating lens?

Q-13. How does a waveguide-type lens focus spherical wavefront microwave energy?

Q-14. What type of lens decelerates a portion of a spherical wavefront?

ANTENNA ARRAYS

Sharply directive antennas can be constructed from two or more simple half-wave dipole elements. They must be positioned so that the fields from the elements add in some directions and cancel in others. Such a set of antenna elements is called an ANTENNA ARRAY. When a reflector is placed behind the dipole array, radiation occurs in one direction with a pattern similar to the one shown in figure 3-13.

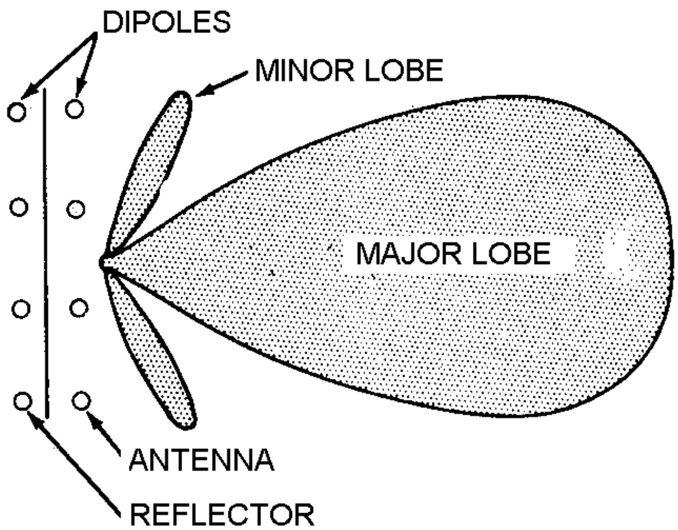


Figure 3-13.—Field pattern of an antenna array.

You will encounter two basic types of antenna arrays, PARASITIC and DRIVEN. Both types of antenna arrays were explained in *NEETS*, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*. Only a brief review is presented in this chapter.

The parabolic reflector antennas previously discussed and the antenna shown in figure 3-13 are examples of parasitic arrays. Notice that the reflector in figure 3-13 is not directly connected to the energy source. Driven arrays, in which all the radiating elements are connected to the energy source, have smaller losses than parasitic arrays while retaining some of the narrow-beam characteristics. Parasitic arrays, such as the parabolic reflector, are used primarily as antennas in fire control radars and other installations, such as microwave communication systems, that require very accurate (narrow) beams. Driven arrays are used primarily as search-radar antennas because extremely narrow beams are less critical than low losses.

If you position a number of driven half-wave antenna elements with respect to each other so that energy from the individual elements will add in certain directions and cancel in other directions, then the antenna system is directional.

Signals from a number of different sources may contribute to or subtract from the overall effect. By properly phasing the energy fed to the antenna elements, and by properly locating the elements, you can control the direction of the energy. You can cause the energy to add in the desired direction and to be out of phase (cancel) in the undesired direction.

Driven arrays are usually made up of a number of half-wave dipoles positioned and phased so that the desired directional pattern will be achieved. Figure 3-14, view (A), shows a simple antenna array consisting of two horizontally mounted elements, each a half wavelength long and fed in phase. The resulting radiation pattern is in a direction at right angles to the plane containing the antenna conductor.

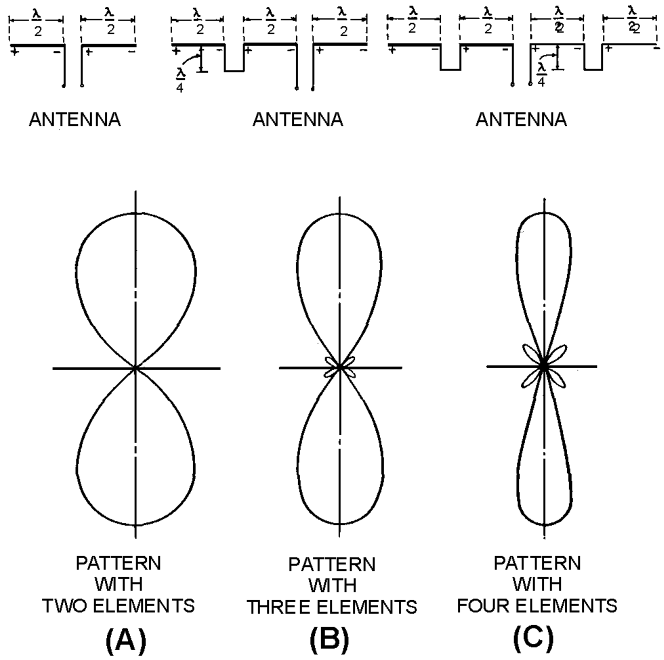


Figure 3-14.—Horizontal array field patterns.

Three- and four-element arrays are shown in figure 3-14, views (B) and (C), respectively. The field pattern of each array is shown beneath it. Note that the beam becomes sharper as the number of elements is increased. If a still-narrower beam is desired, you may add additional elements. The field patterns of the antennas in the figure are bidirectional. Unidirectional patterns may be obtained with a parasitic reflector mounted behind the driven antenna elements.

The BEDSPRING ARRAY (figure 3-15), so called because of its resemblance to a bedspring, is an example of a unidirectional antenna. It consists of a stacked dipole array with an untuned reflector. The more dipoles that are used or stacked in one dimension (horizontal, for example), the more narrow the beam of radiated energy becomes in that plane. Consequently, the size of the antenna is not the same for all installations. Antennas such as the bedspring array are commonly used in TWO-DIMENSIONAL SEARCH RADARS that obtain the range and bearing information of a target.

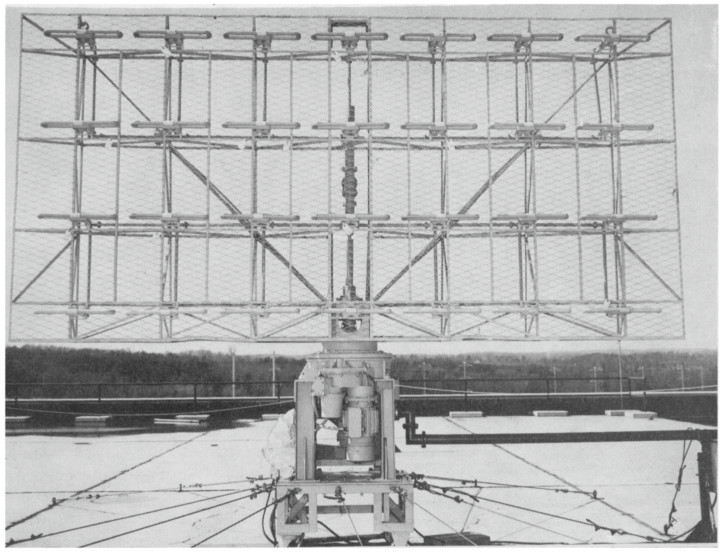


Figure 3-15.—Bedspring array.

FREQUENCY-SENSITIVE ANTENNA

The radar antenna in figure 3-16 uses a feed section to drive horizontally stacked array sections which radiate the applied rf pulses. The same array sections receive the target returns. Each array contains slots cut to radiate and receive a particular frequency. Bearing data is obtained by mechanically rotating the antenna 360 degrees. Elevation data is obtained by electronic scanning of the beam in elevation. The radar antenna is frequency sensitive and radiates pulses at an elevation angle determined by the applied frequency. When the frequency is increased, the beam elevation angle decreases. Conversely, when the applied frequency is decreased, the beam elevation angle increases. The beam elevation angle is therefore selected by the application of a frequency corresponding to the desired angle of elevation. The physical length of the antenna feed section, called the SERPENTINE SECTION (figure 3-17), in relation to the wavelength of the applied energy determines the direction of the radiated beam. You may understand this more clearly if you consider how the beam is shifted. The shift occurs with a change in frequency because the positive and negative peaks of the energy arrive at adjacent slotted arrays at different times. The change in the field pattern is such that the angle of departure (angle at which the radiated beam leaves the antenna) of the beam is changed. Note that a change in phase of the applied rf energy would cause the same effect.

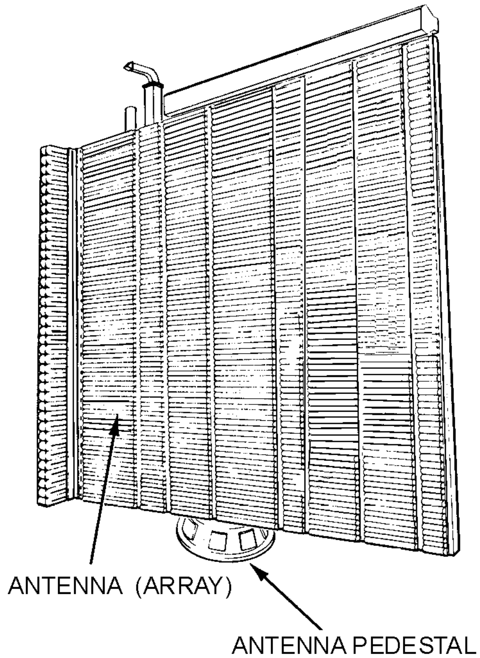


Figure 3-16.—Frequency-sensitive antenna.

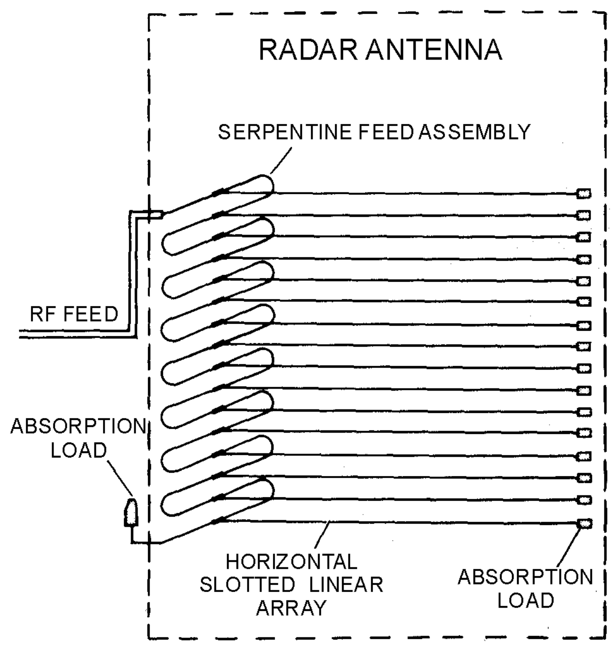


Figure 3-17.—Serpentine feed.

A SLOT ANTENNA exhibits many of the characteristics of a conventional dipole antenna. When arranged in arrays, a high degree of directivity can be obtained. Also, the beam can be caused to scan a volume of space by changing either the frequency or phase of the energy driving the antenna elements.

Basic Slot Antenna and Its Complementary Dipole

The slot antenna consists of a radiator formed by cutting a narrow slot in a large metal surface. Such an antenna is shown in figure 3-18. The slot length is a half wavelength at the desired frequency and the width is a small fraction of a wavelength. The antenna is frequently compared to a conventional half-wave dipole consisting of two flat metal strips. The physical dimensions of the metal strips are such that they would just fit into the slot cut out of the large metal sheet.

This type of antenna is called the COMPLEMENTARY DIPOLE.

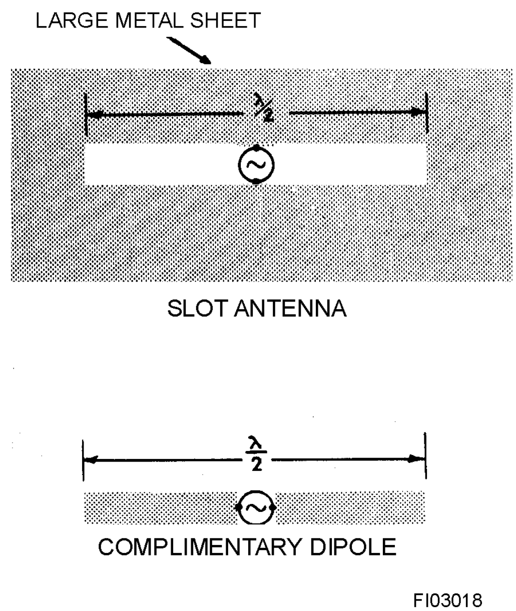


Figure 3-18.—Slot antenna and complementary dipole.

The slot antenna is compared to its complementary dipole to illustrate that the radiation patterns produced by a slot antenna cut into an infinitely large metal sheet and that of the complementary dipole antenna are the same.

Several important differences exist between the slot antenna and its complementary antenna. First, the electric and magnetic fields are interchanged. In the case of the dipole antenna shown in figure 3-18, the electric lines are horizontal while the magnetic lines form loops in the vertical plane. With the slot antenna, the magnetic lines are horizontal and the electric lines are vertical. The electric lines are built up across the narrow dimensions of the slot. As a result, the polarization of the radiation produced by a horizontal slot is vertical. If a vertical slot is used, the polarization is horizontal.

A second difference between the slot antenna and its complementary dipole is that the direction of the lines of electric and magnetic force abruptly reverse from one side of the metal sheet to the other. In the case of the dipole, the electric lines have the same general direction while the magnetic lines form continuous closed loops.

When energy is applied to the slot antenna, currents flow in the metal sheet. These currents are not confined to the edges of the slot but rather spread out over the sheet. Radiation then takes place from both sides of the sheet. In the case of the complementary dipole, however, the currents are more confined; so a much greater magnitude of current is required to produce a given power output using the dipole antenna.

The current distribution of the dipole resembles the voltage distribution of the slot. The edges on the slot have a high voltage concentration and relatively low current distribution; the complementary dipole has a high current concentration and relatively low voltage.

Slot antennas are adaptable for the vhf and uhf ranges. One of their practical advantages is that the feed section which energizes the slot may be placed below the large metal surface in which the slot is cut. Thus, nothing needs to extend from the surface. In addition, the slot itself may be covered by a section of insulating material to provide a seal so that the antenna can be pressurized with dry air. Dry air pressurization reduces moisture in the waveguide and prevents arcing.

Many of the new radar systems reaching the fleet over the next few years will use frequency- or phase-sensitive antennas. Some of the new radars will use antennas that electronically scan the azimuth as well as elevation, eliminating the moving antenna.

Q-15. What is a set of antenna elements called?

Q-16. What type of antenna has all elements connected to the same energy source?

Q-17. What determines the beam elevation angle of an antenna that is electronically scanned in elevation?

Q-18. What is the polarization of the energy radiated by a vertical slot?

SUMMARY

This chapter has presented information on the characteristics of microwave antennas. The information that follows summarizes the important points of this chapter.

The **ANTENNA CHARACTERISTICS** of microwave and low-frequency antennas are essentially the same. The efficiency of an antenna is expressed as a POWER GAIN or POWER RATIO as compared to a standard reference antenna.

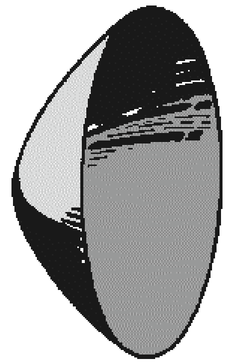
The **STANDING WAVE RATIO (swr)** is a measurement of the impedance mismatch between a transmission line and its load and is an indicator of overall system efficiency.

DIRECTIVITY refers to the direction in which an antenna radiates and the narrowness of the radiated beam in DIRECTIONAL ANTENNAS.

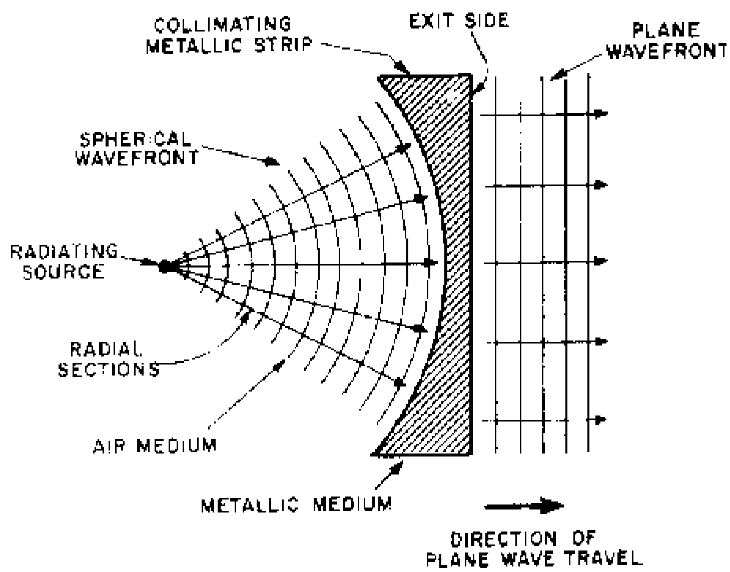
OMNIDIRECTIONAL ANTENNAS radiate and receive in all directions at once.

RECIPROCITY is the ability of an antenna to both transmit and receive electromagnetic energy.

REFLECTOR ANTENNAS are antennas that use a reflector to focus electromagnetic energy into a beam that is directional in either the vertical plane, the horizontal plane, or both planes at once. The basic PARABOLIC REFLECTOR shown in the illustration, or one of its variations, is most often used.



LENS ANTENNAS use a **COLLIMATING LENS** to force the spherical components of a wavefront into parallel (focused) paths by delaying or accelerating portions of the wavefronts, as shown in the illustration.



An **ANTENNA ARRAY** is a set of antenna elements and may be one of two basic types, the DRIVEN ARRAY or the PARASITIC ARRAY.

FREQUENCY-SENSITIVE ANTENNAS use frequency-sensitive slots as radiation sources to achieve directivity. The angle at which the radiated beam leaves the antenna is determined by the frequency of the radiated energy. Currently the most common frequency-sensitive antennas use this feature to achieve elevation coverage while azimuth coverage is achieved by rotating the antennas. New systems will use stationary frequency-sensitive antennas to achieve both azimuth and elevation coverage.

ANSWERS TO QUESTIONS Q1. THROUGH Q18.

A-1. Operating principles and electrical characteristics.

A-2. Power gain or power ratio.

A-3. Standing-wave ratio (swr).

A-4. Omnidirectional.

A-5. Antenna directivity.

A-6. Reciprocity.

A-7. Parabolic.

A-8. Light waves.

A-9. One.

A-10. Determine elevation.

A-11. Vertical.

A-12. Forces the radial segments of a wavefront into parallel paths.

A-13. Some wavefronts are accelerated so that all wavefronts exit the lens at the same time.

A-14. Delay lens.

A-15. Antenna Array.

A-16. Driven Array.

A-17. Frequency or phase of radiated energy.

A-18. Horizontal.

APPENDIX I

GLOSSARY

APERTURE—See slot.

BOUNDARY CONDITIONS—The two conditions that the E-field and H-field within a waveguide must meet before energy will travel down the waveguide. The E-field must be perpendicular to the walls and the H-field must be in closed loops, parallel to the walls, and perpendicular to the E-field.

BEARING—An angular measurement that indicates the direction of an object in degrees from true north. Also called azimuth.

BUNCHER CAVITY—The input resonant cavity in a conventional klystron oscillator.

BUNCHER GRID—In a velocity-modulated tube, the grid which concentrates the electrons in the electron beam into bunches.

CATCHER GRID—In a velocity-modulated tube, a grid on which the spaced electron groups induce a signal. The output of the tube is taken from the catcher grid.

CAVITY RESONATOR—A space totally enclosed by a metallic conductor and supplied with energy in such a way that it becomes a source of electromagnetic oscillations. The size and shape of the enclosure determine the resonant frequency.

CHOKE JOINT—A joint between two sections of waveguide that provides a good electrical connection without power losses or reflections.

COOKIE-CUTTER TUNER—Mechanical magnetron tuning device that changes the frequency by changing the capacitance of the anode cavities.

COPPER LOSS—Power loss in copper conductors caused by the internal resistance of the conductors to current flow. Also called I^2R loss.

CROWN-OF-THORNS TUNER—See Sprocket Tuner.

CUTOFF FREQUENCY—The frequency at which the attenuation of a waveguide increases sharply and below which a traveling wave in a given mode cannot be maintained. A frequency with a half wavelength that is greater than the wide dimension of a waveguide.

DIELECTRIC CONSTANT—The ratio of a given dielectric to the dielectric value of a vacuum.

DIELECTRIC LOSSES—The electric energy that is converted to heat in a dielectric subjected to a varying electric field.

DIRECTIONAL COUPLER—A device that samples the energy traveling in a waveguide for use in another circuit.

DIRECTIVITY—The narrowness of the radiated beam from an antenna.

DOMINANT MODE—The easiest mode to produce in a waveguide, and also, the most efficient mode in terms of energy transfer.

DRIFT SPACE—In an electron tube, a region free of external fields in which relative electron position depends on velocity.

DUMMY LOAD—A device used at the end of a transmission line or waveguide to convert transmitted energy into heat so no energy is radiated outward or reflected back.

E-FIELD—Electric field that exists when a difference in electrical potential causes a stress in the dielectric between two points.

E-TYPE T-JUNCTION—A waveguide junction in which the junction arm extends from the main waveguide in the same direction as the E-field in the waveguide.

ELECTRIC FIELD—See E-field.

ELECTRONIC TUNING—In a reflex klystron, changing the frequency and output power of the tube by altering the repeller voltage.

ELECTROLYSIS—Chemical changes produced by passing an electrical current from one substance (electrode) to another (electrolyte).

ELECTRON ORBITAL MOVEMENT—The movement of an electron around the nucleus of an atom.

ELECTRON SPIN—The movement of an electron around its axis.

ELEVATION ANGLE—The angle between the line of sight to an object and the horizontal plane.

FARADAY ROTATION—The rotation of the plane of polarization of electromagnetic energy when it passes through a substance influenced by a magnetic field that has a component in the direction of propagation.

FERRITE—A powdered and compressed ferric oxide material that has both magnetic properties and resistance to current flow.

FERRITE SWITCH—A ferrite device that blocks the flow of energy through a waveguide by rotating the electric field 90 degrees. The rotated energy is then reflected or absorbed.

GRID-GAP TUNING—A method of changing the center frequency of a resonant cavity by physically changing the distance between the cavity grids.

GROUP VELOCITY—The forward progress velocity of a wave front in a waveguide.

H-FIELD—Any space or region in which a magnetic force is exerted. The magnetic field may be produced by a current-carrying coil or conductor, by a permanent magnet, or by the earth itself.

H-TYPE T-JUNCTION—A waveguide junction in which the junction arm is parallel to the magnetic lines of force in the main waveguide.

HELIX—A spirally wound transmission line used in a traveling-wave tube to delay the forward progress of the input traveling wave.

HORIZONTAL PLANE—An imaginary plane tangent to and touching the Earth's surface as established by a stable element, such as a gyroscope.

HORN—A funnel-shaped section of waveguide used as a termination device and as a radiating antenna.

HOT CARRIER—A current carrier, which may be either a hole or an electron, that has relatively high energy with respect to the current carriers normally found in majority-carrier devices.

HOT-CARRIER DIODE—A semiconductor diode in which hot carriers are emitted from a semiconductor layer into the metal base. Also called a hot-electron diode. An example is the Schottky-Barrier diode.

HYBRID JUNCTION—A waveguide junction that combines two or more basic T-junctions.

HYBRID RING—A hybrid-waveguide junction that combines a series of E-type T-junctions in a ring configuration.

IDLER FREQUENCY—In a parametric amplifier, the difference between the input signal and the pump signal frequency. Also called the lower-sideband frequency.

INTERACTION SPACE—The region in an electron tube where the electrons interact with an alternating electromagnetic field.

INTERELECTRODE CAPACITANCE—The capacitance between the electrodes of an electron tube.

I^2R LOSS—See Copper Loss.

IRIS—A metal plate with an opening through which electromagnetic waves may pass. Used as an impedance matching device in waveguides.

LEAD INDUCTANCE—The inductance of the lead wires connecting the internal components of an electron tube.

LOAD ISOLATOR—A passive attenuator in which the loss in one direction is much greater than that in the opposite direction. An example is a ferrite isolator for waveguides that allows energy to travel in only one direction.

LOOP—A curved conductor that connects the ends of a coaxial cable or other transmission line and projects into a waveguide or resonant cavity for the purpose of injecting or extracting energy.

LOOSE COUPLING—Inefficient coupling of energy from one circuit to another that is desirable in some applications. Also called weak coupling.

MAGIC-T JUNCTION—A combination of the H-type and E-type T-junctions.

MAGNETIC FIELD—See H-field.

METALLIC INSULATOR—A shorted quarter-wave section of transmission line.

MICROWAVE REGION—The portion of the electromagnetic spectrum from 1,000 megahertz to 100,000 megahertz.

MODULATOR—A device that produces modulation; i.e., varies the amplitude, frequency, or phase of an ac signal.

NEGATIVE-RESISTANCE ELEMENT—A component having an operating region in which an increase in the applied voltage increases the resistance and produces a proportional decrease in current. Examples include tunnel diodes and silicon unijunction transistors.

NONDEGENERATIVE-PARAMETRIC AMPLIFIER—A parametric amplifier that uses a pump signal frequency that is higher than twice the frequency of the input signal.

PHASE SHIFTER—A device used to change the phase relationship between two ac signals.

POWER GAIN—The ratio of the radiated power of an antenna compared to the output power of a standard antenna. A measure of antenna efficiency usually expressed in decibels. Also referred to as POWER RATIO.

POWER RATIO—See Power Gain.

PROBE—A metal rod that projects into, but is insulated from, a waveguide or resonant cavity and used to inject or extract energy.

PUMP—Electrical source of the energy required to vary the capacitance of a parametric amplifier.

RANGE—Distance, as measured from a point of reference, such as a radar, to a target or other object.

REACTANCE AMPLIFIER—A low-noise amplifier that uses a nonlinear variable reactance as the active element instead of a variable resistance. Also called a parametric amplifier.

RECIPROCITY—The ability of an antenna to both transmit and receive electromagnetic energy.

REFLEX KLYSTRON—A klystron with a reflector (repeller) electrode in place of a second resonant cavity to redirect the velocity-modulated electrons back through the cavity which produced the modulation.

REFRACTIVE INDEX—The ratio of the phase velocity of a wave in free space to the phase velocity of the wave in a given substance (dielectric).

REPELLER—Sometimes called a reflector. An electrode in a reflex klystron with the primary purpose of reversing the direction of the electron beam.

ROTATING JOINT—A joint that permits one section of a transmission line or waveguide to rotate continuously with respect to another while passing energy through the joint. Also called a rotary coupler.

SKIN EFFECT—The tendency for alternating current to concentrate in the surface layer of a conductor. The effect increases with frequency and serves to increase the effective resistance of the conductor.

SLOT—Narrow opening in a waveguide wall used to couple energy in or out of the waveguide. Also called an aperture or a window.

SPROCKET TUNER—Mechanical tuning device for magnetron tubes that changes the frequency of the cavities by changing the inductance. Also called a crown-of-thorns tuner.

STAGGER TUNING—A method of klystron tuning in which the resonant cavities are tuned to slightly different frequencies to increase the bandwidth of the amplifier.

STANDING WAVE RATIO—The ratio of the maximum to the minimum amplitudes of corresponding components of a field, voltage, or current along a transmission line or waveguide in the direction of propagation measured at a given frequency.

SYNCHRONOUS TUNING—In a klystron amplifier, a method of tuning which tunes all the resonant cavities to the same frequency. High gain is achieved, but the bandwidth is narrow.

TRANSIT TIME—The time an electron takes to cross the distance between the cathode and anode.

TRANSVERSE ELECTRIC MODE—The entire electric field in a waveguide is perpendicular to the wide dimension and the magnetic field is parallel to the length. Also called the TE mode.

TRANSVERSE MAGNETIC MODE—The entire magnetic field in a waveguide is perpendicular to the wide dimension ("a" wall) and some portion of the electric field is parallel to the length. Also called the TM mode.

TUNNELING—The piercing of a potential barrier in a semiconductor by a particle (current carrier) that does not have sufficient energy to go over the barrier.

TUNNEL DIODE—A heavily doped junction diode that has negative resistance in the forward direction over a portion of its operating range. See NEGATIVE-RESISTANCE ELEMENT.

VARACTOR—A pn-junction semiconductor designed for microwave frequencies in which the capacitance varies with the applied bias voltage.

VARIABLE ATTENUATOR—An attenuator for reducing the strength of an ac signal either continuously or in steps, without causing signal distortion.

VELOCITY MODULATION—Modification of the velocity of an electron beam by the alternate acceleration and deceleration of electrons.

VERTICAL PLANE—An imaginary plane that is perpendicular to the horizontal plane.

WAVEGUIDE—A rectangular, circular, or elliptical metal pipe designed to transport electromagnetic waves through its interior.

WAVEGUIDE MODE OF OPERATION—Particular field configuration in a waveguide that satisfies the boundary conditions. Usually divided into two broad types: the transverse electric (TE) and the transverse magnetic (TM).

WAVEGUIDE POSTS—A rod of conductive material used as impedance-changing devices in waveguides.

WAVEGUIDE SCREW—A screw that projects into a waveguide for the purpose of changing the impedance.

WINDOW—See Slot.

WOBBLE FREQUENCY—The frequency at which an electron wobbles on its axis under the influence of an external magnetic field of a given strength.

MODULE 11 INDEX

T

W

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Waveguide Theory and Applications," pages 1-1 through 1-68.

1-1. The portion of the electromagnetic spectrum which falls between 1,000 and 100,000 megahertz is referred to as which of the following regions?

1. X-ray
2. Infrared
3. Microwave
4. Ultra-violet

1-2. Microwave theory is based on the action of which of the following fields?

1. Electric field only
2. Magnetic field only
3. Electromagnetic field

1-3. Coaxial lines are more efficient than two-wire lines at microwave frequencies for which of the following reasons?

1. Because electromagnetic fields are completely confined in coaxial lines
2. Because electromagnetic fields are not completely confined in coaxial lines
3. Because coaxial lines have less resistance to current flow than two-wire transmission lines
4. Each of the above

1-4. The most efficient transfer of electromagnetic energy can be provided by which of the following mediums?

1. Waveguides
2. Twin-lead flat lines
3. Single-conductor lines
4. Coaxial transmission lines

1-5. Copper (I^2R) losses are reduced by what physical property of waveguides?

1. Small surface area
2. Large surface area
3. Shape of the waveguide
4. Waveguide material used

1-6. In a coaxial line, the current-carrying area of the inner conductor is restricted to a small surface layer because of which of the following properties?

1. Skin effect
2. Copper loss
3. Conductor density
4. Temperature effect

1-7. Which of the following dielectrics is used in waveguides?

1. Air
2. Mica
3. Insulating oil
4. Insulating foam

1-8. Which of the following characteristics of a waveguide causes the lower-frequency limitation?

1. I^2R loss
2. Physical size
3. Wall thickness
4. Dielectric loss

1-9. At very high frequencies, ordinary insulators in a two-wire transmission line display the characteristics of what electrical component?

1. An inductor
2. A resistor
3. A capacitor
4. A transformer

1-10. At very high frequencies, which of the following devices works best as an insulator?

1. Open half-wave section
2. Open quarter-wave section
3. Shorted half-wave section
4. Shorted quarter-wave section

1-11. The range of operating frequencies is determined by which of the following wave-guide dimensions?

1. The widest
2. The longest
3. The shortest
4. The narrowest

1-12. If frequency is decreased, what change, if any, will be required in the dimensions of the wave-guide bus bar?

1. Decrease in dimensions
2. Increase in dimensions
3. None

1-13. The cutoff frequency for a wave-guide is controlled by the physical dimensions of the wave-guide and is defined as the frequency at which two quarter-wavelengths are

1. shorter than the "a" dimension
2. shorter than the "b" dimension
3. longer than the "b" dimension
4. longer than the "a" dimension

1-14. In practical applications, which of the following dimensions describes the wide dimension of the wave-guide at the operating frequency?

1. 0.1 wavelength
2. 0.2 wavelength
3. 0.5 wavelength
4. 0.7 wavelength

1-15. Which of the following fields is/are present in wave guides?

1. E field only
2. H field only
3. E and H fields
4. Stationary field

1-16. A difference in potential across a dielectric causes which of the following fields to develop?

1. Electric field only
2. Magnetic field only
3. Electromagnetic field

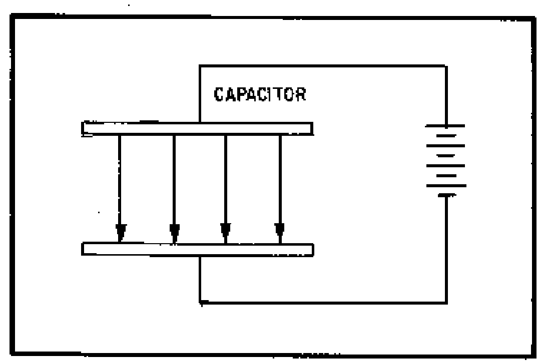


Figure 1A.—Electric field.

IN ANSWERING QUESTION 1-17, REFER TO FIGURE 1A.

1-17. What information is indicated by the number of arrows between the plates of the capacitor?

1. The amount of capacitance
2. The amount of current flow
3. The strength of the electric field
4. The strength of the magnetic field

1-18. H lines have which of the following distinctive characteristics?

1. They are continuous straight lines
2. They are generated by voltage
3. They form closed loops
4. They form only in the wave-guide

1-19. What minimum number of boundary conditions must be satisfied for energy to travel down a waveguide?

1. One
2. Two
3. Three
4. Four

1-20. For an electric field to exist at the surface of a conductor, the field must have what angular relationship to the conductor?

1. 0 degrees
2. 30 degrees
3. 45 degrees
4. 90 degrees

1-21. What, if anything, happens to the amplitude of the wavefronts within a waveguide that DO NOT meet boundary conditions?

1. They Increase rapidly to maximum
2. They decrease slowly to the half-power point
3. They decrease rapidly to zero
4. Nothing

1-22. If the wall of a wave-guide is perfectly flat, the angle of reflection is equal to which of the following angles?

1. Angle of cutoff
2. Angle of incidence
3. Angle of refraction
4. Angle of penetration

1-23. THIS QUESTION HAS BEEN DELETED.

1-24. How does the group velocity of an electromagnetic field in a waveguide compare to the velocity of a wavefront through free space?

1. Group velocity is faster
2. Group velocity is slower
3. Their velocities are the same

1-25. The group velocity of a wavefront in a waveguide may be increased by which of the following actions?

1. Decreasing the frequency of the input energy
2. Increasing the frequency of the input energy
3. Increasing the power of the input energy
4. Decreasing the power of the input energy

1-26. The various field configurations that can exist in a waveguide are referred to as

1. wavefronts
2. modes of operation
3. fields of operation
4. fields of distribution

1-27. The most efficient transfer of energy occurs in a waveguide in the what mode?

1. Sine
2. Dominant
3. Transverse
4. Time-phase

1-28. How is the cutoff wavelength for a circular waveguide figured?

1. 1.17 times the radius of the waveguide
2. 1.17 times the diameter of the waveguide
3. 1.71 times the diameter of the waveguide
4. 1.71 times the radius of the waveguide

1-29. The field configuration in waveguides is divided into what two categories?

1. Half-sine and dominant
2. Transverse electric and transverse magnetic
3. Transverse electric and dominant
4. Transverse magnetic and half-sine

1-30. With a mode description of $TE_{1,0}$, what maximum number of half-wave patterns exist across the "a" dimension of a waveguide?

1. One
2. Two
3. Three
4. Four

1-31. With the mode description, $TE_{1,1}$, what maximum number of half-wave patterns exist across the diameter of a circular waveguide?

1. One
2. Two
3. Three
4. Four

1-32. To inject or remove energy from a waveguide, which of the following devices could you use?

1. Slot
2. Loop
3. Probe
4. Each of the above

1-33. Loose coupling is a method used to reduce the amount of energy being transferred from a waveguide. How is loose coupling achieved when using a probe?

1. By doubling the size of the probe
2. By increasing the length of the probe
3. By decreasing the length of the probe
4. By placing the probe directly in the center of energy field

1-34. Loop coupling is most efficient when the loop is placed at what point in which of the following fields?

1. At the point of maximum electric field
2. At the point of minimum electric field
3. At the point of minimum magnetic field
4. At the point of maximum magnetic field

1-35. Increasing the size of the loop wire increases which of the following loop capabilities?

1. Efficiency
2. Bandwidth coverage
3. Power-handling capability
4. Each of the above

1-36. A waveguide which is not perfectly impedance matched to its load is not efficient. Which of the following conditions in a waveguide causes this inefficiency?

1. Sine waves
2. Dominant waves
3. Standing waves
4. Transverse waves

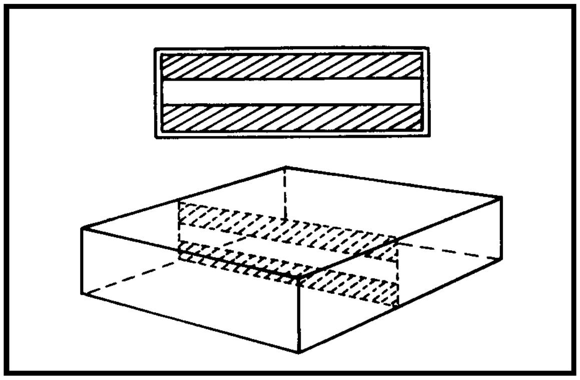


Figure 1B.—Waveguide iris.

IN ANSWERING QUESTION 1-37, REFER TO FIGURE 1B.

1-37. The iris shown in the figure has what type of equivalent circuit?

1. Parallel-LC
2. Shunt-resistive
3. Shunt-inductive
4. Shunt-capacitive

1-38. A waveguide iris that covers part of both the electric and magnetic planes acts as what type of equivalent circuit at the resonant frequency?

1. As a shunt inductive reactance
2. As a shunt resistance
3. As a shunt capacitive reactance
4. Each of the above

1-39. A horn can be used as a waveguide termination device because it provides which of the following electrical functions?

1. A reflective load
2. An absorptive load
3. An abrupt change in impedance
4. A gradual change in impedance

1-40. For a waveguide to be terminated with a resistive load, that load must be matched to which of the following properties of the waveguide?

1. The bandwidth
2. The frequency
3. The inductance
4. The characteristic impedance

1-41. A resistive device with the sole purpose of absorbing all the energy in a waveguide without causing reflections is a/an

1. iris
2. horn
3. antenna
4. dummy load

1-42. A resistive load most often dissipates energy in which of the following forms?

1. Heat
2. Light
3. Magnetic
4. Electrical

1-43. Reflections will be caused by an abrupt change in which of the following waveguide physical characteristics?

1. Size
2. Shape
3. Dielectric material
4. Each of the above

1-44. A waveguide bend which is in the E or H plane must be greater than two wavelengths to prevent

1. cracking
2. reflections
3. energy gaps
4. electrolysis

1-45. A flexible waveguide is used in short sections because of the power-loss disadvantages. What is the cause of this power loss?

1. Walls are not smooth
2. E and H fields are not perpendicular
3. Cannot be terminated in its characteristics impedance
4. Wall size cannot be kept consistent

1-46. The choke joint is used for what purpose in a waveguide?

1. To reduce standing waves
2. To restrict the volume of electron flow
3. To prevent the field from rotating
4. To provide a temporary joint in a waveguide during maintenance or repair

1-47. A circular waveguide is normally used in a rotating joint because rotating a rectangular waveguide would cause which of the following unwanted conditions?

1. Oscillation
2. Large power loss
3. Decrease in bandwidth
4. Field-pattern distortion

1-48. In your waveguide inspections, you should be alert for which of the following problems?

1. Corrosion
2. Damaged surface
3. Improperly sealed joints
4. Each of the above

1-49. What type of corrosion occurs when dissimilar metals are in contact?

1. Contact corrosion
2. Metallic corrosion
3. Electrical corrosion
4. Electrolytic corrosion

1-50. Internal arcing in a waveguide is usually a symptom of which of the following conditions?

1. Change in mode
2. Electrolysis at a joint
3. Moisture in the waveguide
4. Gradual change in frequency

1-51. What is the primary purpose of a directional coupler?

1. To sample the energy in a waveguide
2. To change the phase of the energy in the waveguide
3. To change the direction of energy travel in the waveguide
4. To allow energy in the waveguide to travel in one direction only

1-52. What is the electrical distance between the two holes in a simple directional coupler?

1. 1/8 wavelength
2. 1/4 wavelength
3. 1/2 wavelength
4. 3/4 wavelength

1-53. When the two portions of a reflected wave reach the pickup probe of an incident-wave directional coupler, what is their phase relationship?

1. 45º out of phase
2. 90º out of phase
3. 120º out of phase
4. 180º out of phase

1-54. The highest frequency at which a conventional circuit can oscillate is reached when which of the following values can be reduced no further?

1. Total resistance
2. Total inductance only
3. Total capacitance only
4. The total capacitance and inductance

1-55. For a device to be considered a resonant cavity, it must fulfill which of the following requirements?

1. Be enclosed by conducting walls
2. Possess resonant properties
3. Contain oscillating electromagnetic fields
4. All of the above

1-56. What property gives a resonant cavity a narrow bandpass and allows very accurate tuning?

1. Low Q
2. High Q
3. Inductive reactance
4. Capacitive reactance

1-57. What factor(s) determines the primary frequency of a resonant cavity?

1. Size only
2. Shape only
3. Size and shape
4. Q of the cavity

1-58. Tuning is the process of changing what property of a resonant cavity?

1. The Q
2. The power output
3. The cutoff frequency
4. The resonant frequency

1-59. An adjustable slug or screw placed in the area of maximum E lines in a resonant cavity provides what type of tuning?

1. Volume
2. Inductive
3. Resistive
4. Capacitive

1-60. What are the two basic types of waveguide T junctions?

1. H-type and T-type
2. H-type and E-type
3. H-type and magic T
4. E-type and magic T

1-61. A waveguide junction in which the arm area extends from the main waveguide in the same direction as the electric field is an example of what type junction?

1. E-type magic T
2. H-type magic T
3. H-type T junction
4. E-type T junction

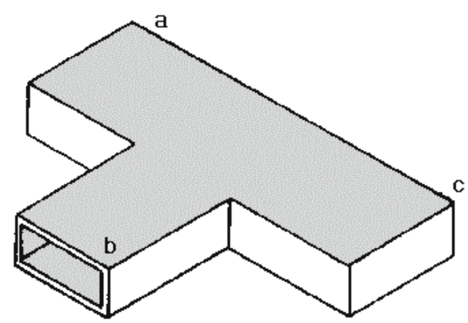


Figure 1C.—H-type T junction.

IN ANSWERING QUESTION 1-62, REFER TO FIGURE 1C.

1-62. When an input is fed into the "b" arm in the figure, which of the following output signal arrangements is/are available?

1. Out-of-phase signals from arms "a" and "c"
2. In-phase signals from arms "a" and "c"
3. An output from the "a" arm only
4. An output from the "c" arm only

1-63. E-type and H-type junctions are combined in which of the following devices?

1. Magic T
2. Rat race
3. Feed horn
4. Hybrid ring

1-64. Low power-handling capabilities and internal power losses are the primary disadvantages of which of the following devices?

1. Magic T
2. Rat race
3. Duplexer
4. Hybrid ring

1-65. The hybrid ring is usually used as what type of device in radar systems?

1. Mixer
2. Detector
3. Duplexer
4. Impedance matcher

1-66. Ferrite devices are useful in electronic and microwave applications because they possess magnetic properties and offer which of the following other properties?

1. Negative resistance to current flow
2. Low resistance to current flow
3. High resistance to current flow
4. High conductance for current flow

1-67. Electrons exhibit which of the following types of motion?

1. Spin
2. Orbital
3. Both 1 and 2 above
4. Linear

1-68. Electrons in a ferrite can be caused to wobble on their axes by which of the following actions?

1. Decreasing the internal resistance
2. Increasing the internal resistance
3. Applying a magnetic field
4. Applying an electric field

1-69. The energy in a ferrite attenuator that is attenuated is dissipated as which of the following energy forms?

1. Heat
2. Light
3. Magnetic
4. Electrical

1-70. The amount of rotation in a Faraday-rotation type ferrite phase shifter is dependent upon which of the following ferrite properties?

1. Length of the material
2. Diameter of the material
3. Strength of the material
4. Internal resistance of the material

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Microwave Principles," pages 2-1 through 2-66. Chapter 3, "Microwave Antennas," pages 3-1 through 3-20.

2-1. As the frequency in a conventional vacuum tube is increased, what is the effect on the capacitive reactance of the tube?

1. It increases
2. It decreases
3. It remains the same

2-2. Undesirable degenerative feedback in conventional vacuum tubes can be caused by which of the following tube factors?

1. Grid bias
2. Transit time
3. Plate voltage
4. Lead inductance

2-3. A decrease in efficiency in a conventional vacuum tube caused by a phase shift between plate current and grid voltage can be the result of excessive

1. transit time
2. lead inductance
3. capacitive reactance
4. interelectrode capacitance

2-4. Moving tube electrodes further apart to decrease interelectrode capacitance causes which of the following tube characteristics to increase?

1. Transit time
2. Lead inductance
3. Inductive reactance
4. Capacitive reactance

2-5. Which of the following properties of an electron is directly proportional to its velocity?

1. Mass
2. Kinetic energy
3. Potential energy
4. All of the above

2-6. An electron that enters an electrostatic field and travels in the same direction as the lines of force will react in what way?

1. It will accelerate
2. It will decelerate
3. It will spin faster
4. It will spin slower

2-7. The alternate speeding up and slowing down of electrons in a beam that produces electron bunches is known as which of the following modulation terms?

1. Carrier modulation
2. Velocity modulation
3. Amplitude modulation
4. Frequency modulation

2-8. How is an electron affected, if at all, if it enters the buncher-grid gap when the potential across the grids is positive?

1. It is deflected
2. It is accelerated
3. It is decelerated

2-9. The point in a stream of electron bunches at which the catcher cavity will most efficiently remove power is determined by which of the following factors?

1. Electron spin velocity
2. The size of the cavity
3. The size of the bunches
4. Frequency of the buncher-grid signal

2-10. Which of the following electronic interactions is used in klystron operation?

1. Velocity-modulation
2. Voltage and current
3. Variable-capacitance
4. Crossed electromagnetic-field

2-11. In a klystron, an ac potential is superimposed on the dc voltage that is applied to the buncher grids by what component?

1. The cathode
2. The accelerator grid
3. The buncher-grid cavity resonator
4. The catcher-grid cavity resonator

2-12. A two-cavity klystron that has a feedback path from the catcher cavity to the buncher cavity will operate as what type of circuit?

1. Modulator
2. Amplifier
3. Oscillator
4. Discriminator

2-13. The input signal to a two-cavity klystron amplifier is applied to which of the following components?

1. The cathode
2. The catcher grids
3. The buncher grids
4. The accelerator grid

2-14. In a klystron, the placement of additional cavities between the buncher cavity and catcher cavity increases the power output by causing which of the following electronic actions?

1. Increased velocity modulation
2. Decreased velocity modulation
3. Increased electron-beam speed
4. Decreased electron-beam speed

2-15. What is the purpose of applying a large negative pulse to the cathode of a three-cavity klystron?

1. To focus the electron beam
2. To accelerate the electron beam
3. To decelerate the electron beam
4. To modulate the electron beam

2-16. In a three-cavity klystron, what cavity, if any, contributes most to the velocity modulation of the electron beam?

1. The middle cavity
2. The catcher cavity
3. The buncher cavity
4. None

2-17. The bandwidth of a multicavity klystron can be increased by using which of the following tuning methods?

1. Varactor tuning
2. Staggered tuning
3. Electronic tuning
4. Synchronous tuning

2-18. The repeller of a reflex klystron replaces what component in other types of klystrons?

1. The input cavity
2. The output cavity
3. The buncher cavity
4. The intermediate cavity

2-19. In a reflex klystron, what type(s) of electrical charge, if any, does the repeller have?

1. Positive only
2. Negative only
3. Alternately positive and negative
4. None

2-20. In a reflex klystron, the length of time a constant speed electron remains in the space separating the grid and repeller is determined by which of the following factor(s)?

1. Repeller voltage
2. Electron velocity
3. Both 1 and 2 above
4. The distance between grid and repeller

2-21. A reflex klystron in which the constant-speed electrons remain in the repeller field for 3/4 cycle is operating in what mode?

1. Mode 1
2. Mode 2
3. Mode 3
4. Mode 4

2-22. In a reflex klystron, the choice of operating mode is determined by which of the following circuit factors?

1. The voltage required
2. The power available
3. The frequency range available
4. Both 2 and 3 above

2-23. In the higher modes, power and amplitude limitations in a reflex klystron are caused by which of the following actions?

1. Electron debunching
2. Frequency fluctuation
3. Decreasing transit time
4. Increasing electron density

2-24. The term "bel" is a unit of measurement that expresses which of the following relationships?

1. Ratio of voltage and resistance
2. Logarithmic ratio between input and output
3. Geometric progression from input to output
4. Ratio of voltage to current

2-25. The term "dBm" is based on what standard reference level?

1. 1 watt
2. 1 volt
3. 1 milliwatt
4. 1 millivolt

2-26. Which of the following twt characteristics makes it ideal for use as an rf amplifier?

1. High-noise and narrow-bandwidth
2. Low-noise and wide-bandwidth
3. High-noise and wide-bandwidth
4. Low-noise and narrow-bandwidth

2-27. In a twt, what is the primary purpose of the helix?

1. To increase the forward velocity of the input
2. To decrease the forward velocity of the input
3. To decrease the reflected velocity of the output
4. To increase the reflected velocity of the output

2-28. Velocity modulation of the electron beam in a twt is achieved by what action?

1. By the action of resonant cavities
2. By interaction of the electron beam with the permanent magnet field
3. By interaction of the electron beam with the electric field in the helix
4. All of the above

2-29. In a twt, what is the purpose of the attenuators along the helix?

1. To focus the beam
2. To limit the input
3. To limit the output
4. To prevent reflections

2-30. A microwave tube that extracts energy from a wave that travels from the collector toward the cathode is referred to as the

1. klystron
2. traveling-wave tube
3. crossed-field amplifier
4. backward-wave oscillator

2-31. In a magnetron, the magnetic field between the plate and cathode serves what purpose?

1. Acts as a grid
2. Provides a plate load
3. Acts as a space-charge suppressor
4. Provides filament power

2-32. What property in a magnetron is controlled by the cavities in the plate?

1. Input power
2. Output power
3. Input voltage
4. Output frequency

2-33. In a magnetron, what causes the path of an electron to curve when it is moving from the cathode to the plate?

1. The cathode pulse
2. The electric field
3. The resonant cavities
4. The permanent magnetic field

2-34. The critical value of magnetic field strength in a magnetron causes which of the following electronic actions?

1. Plate cavities stop oscillating
2. Output power decreases to zero
3. Electrons strike the plate and return to the cathode
4. Electrons miss the plate and return to the cathode

2-35. Magnetron oscillators are divided into what total number of classes?

1. One
2. Two
3. Three
4. Four

2-36. THIS QUESTION HAS BEEN DELETED.

2-37. In a negative-resistance magnetron, which, if any, of the following values of magnetic field strengths is required to start oscillations?

1. Critical value
2. Slightly lower than critical value
3. Slightly higher than critical value
4. None of the above

2-38. In magnetrons, the effect of filament bombardment can be reduced by which of the following actions?

1. Increasing plate voltage
2. Reducing filament voltage
3. Reducing signal frequency
4. Increasing signal frequency

2-39. In electron-resonance magnetrons, which of the following anode blocks is/are used?

1. Vane anode
2. Rising-sun anode
3. Hole-and-slot anode
4. Each of the above

2-40. In the electron-resonance magnetron, the total electric field is produced by which of the following field combinations?

1. The dc field only
2. The ac and dc fields
3. The ac and magnetic fields
4. The dc and magnetic fields

2-41. Energy from working electrons is received by which of the following magnetron fields?

1. The ac field only
2. The dc field only
3. Both the ac and dc fields
4. The magnetic field

2-42. In a magnetron, the total action of many electrons returning to the cathode while others are moving toward the anode forms which of the following patterns?

1. Vertical wavefront
2. Space-charge wheel
3. Spherical wavefront
4. Horizontal space charge

2-43. The greatest power output is produced in what magnetron mode of operation?

1. Mode 1
2. Mode 2
3. The pi mode
4. The radian mode

2-44. Magnetic lines of force passing between cavities are intercepted in which, if any, of the following magnetron coupling methods?

1. Slot coupling method
2. Strap-fed loop method
3. Segment-fed loop method
4. None of the above

2-45. Inductive tuning in a magnetron is accomplished by which of the following actions?

1. Changing the cavity resistance
2. Altering the cavity surface-to-volume ratio
3. Decreasing the size of the slot gap
4. Decreasing the space charge

2-46. A magnetron should be "baked in" under which of the following conditions?

1. Before each use
2. After periods of idleness
3. After initial installation
4. Both 2 and 3 above

2-47. The crossed-field amplifier produces which of the following desirable circuit characteristics?

1. Wide bandwidth
2. High efficiency
3. Large power-handling capability
4. All of the above

2-48. The tunneling action of the tunnel diode produces which of the following useful properties?

1. Long transit time
2. Positive resistance
3. Negative resistance
4. Variable inductance

2-49. The tuned circuit in a tunnel-diode oscillator determines which of the following circuit values?

1. Frequency
2. Resistance
3. Power range
4. Capacitance

2-50. Tuning a tunnel-diode oscillator by changing the capacitance of the tuned circuit can be accomplished by which of the following tuning methods?

1. Bias tuning
2. Slot tuning
3. Current tuning
4. Varactor tuning

2-51. In the reflection-type, circulator-coupled tunnel-diode amplifier, what component prevents feedback to the tuned input circuit?

1. The circulator
2. The output loop
3. The tuned cavity
4. The tunnel diode

2-52. Stability is a problem in which, if any, of the following tunnel-diode frequency converters?

1. High-gain converter
2. Unity-gain converter
3. Conversion-loss converter
4. None of the above

2-53. The varactor is a type of pn junction that acts as which of the following types of electronic devices?

1. Fixed resistor
2. Fixed capacitor
3. Variable resistor
4. Variable capacitor

2-54. What is the most important feature of the parametric amplifier?

1. Low noise
2. High gain
3. Power output
4. Frequency range

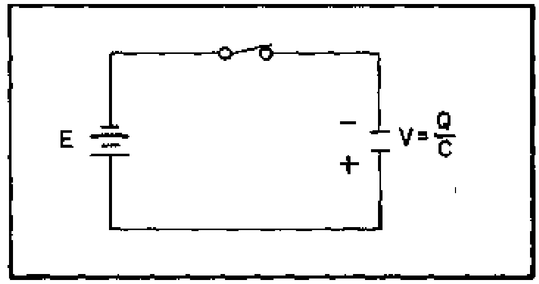


Figure 2A.—Voltage amplification circuit.

IN ANSWERING QUESTION 2-55, REFER TO FIGURE 2A.

2-55. How is amplification accomplished in the circuit?

1. By opening the switch
2. By varying the frequency
3. By varying the resistance
4. By varying the capacitance

2-56. In a nondegenerative parametric amplifier, how does the pump frequency compare to the input signal frequency?

1. The pump frequency is half the input signal frequency
2. The pump frequency is double the input signal frequency
3. The pump frequency is more than double the input signal
4. The pump frequency is equal to the input signal frequency

2-57. In a nondegenerative parametric amplifier with a pump frequency of 12 gigahertz and an idler frequency of 9 gigahertz, what is the input signal frequency?

1. 3 gigahertz
2. 18 gigahertz
3. 21 gigahertz
4. 108 gigahertz

2-58. Bulk-effect semiconductors have what primary advantage over normal pn-junction semiconductors?

1. Smaller size
2. Lower frequency
3. Simpler construction
4. Greater power output

2-59. Gallium-arsenide semiconductors begin to exhibit which of the following electrical characteristics at the threshold point?

1. Variable inductance
2. Positive resistance
3. Negative resistance
4. Variable capacitance

2-60. In an avalanche transit-time diode, what causes the dc bias energy previously absorbed by avalanche electrons to be given up to the microwave field?

1. The electron velocity
2. The negative-resistance property
3. The electron transit time
4. The amount of dc bias energy available

2-61. In a point-contact diode, passing a large current from the catwhisker to the silicon crystal produces what region?

1. A domain region
2. A small p region
3. A small n region
4. An avalanche region

2-62. A Schottky barrier diode has which of the following advantages over a point-contact diode?

1. Lower frequency range
2. Higher frequency range
3. Lower noise generation
4. Higher noise generation

2-63. The PIN diode begins acting as a variable resistance at what minimum frequency?

1. 100 megahertz
2. 200 megahertz
3. 300 megahertz
4. 400 megahertz

2-64. Power ratio is a term used to express what property of an antenna?

1. Efficiency
2. Reciprocity
3. Sensitivity
4. Power output

2-65. A standing-wave ratio (swr) describes which of the following quantities?

1. Transmission-to-reception efficiency value
2. The amount of output power
3. The amount of mismatch between a transmission line and its load
4. The amount of characteristic impedance

2-66. Directivity refers to which of the following properties of a radiated beam?

1. Power gain
2. Standing-wave ratio
3. Narrowness of the beam
4. Polarization of the beam

2-67. Surface angular measurements for antenna directivity in radar and communications systems are made in relationship to which of the following references?

1. Vertical plane only
2. Horizontal plane only
3. Horizontal plane and true north
4. Vertical plane and true north

2-68. Radar range is determined as a function of which of the following measures?

1. Pulse travel time
2. Elevation angle of the antenna
3. Distance to the horizon
4. Angular velocity of the energy

2-69. The parabolic reflector is often used because it produces a radiation pattern with which of the following antenna characteristics?

1. Omnidirectional
2. Highly directive
3. Many equal lobes
4. Spherical wavefronts

2-70. In a lens antenna, what is the purpose of the collimating lens?

1. It produces spherical wavefronts
2. It produces an omnidirectional pattern
3. It forces parallel segments of the wavefront into spherical paths
4. It forces radial segments of the wavefront into parallel paths

2-71. A lens antenna which accelerates some portion of the wavefronts so that all wavefronts exit the lens at the same time is referred to as the

1. metallic delay lens
2. waveguide-type lens
3. loaded microwave lens
4. dielectric delay lens

2-72. In a delay-type lens, the amount of delay in the phase of the wave passing through the lens is determined by which of the following characteristics?

1. The dielectric constant of the lens
2. The characteristic impedance of the lens
3. The physical size of the lens
4. All of the above

2-73. What is/are the basic type(s) of antenna array(s) in common use?

1. Driven only
2. Parasitic only
3. Parasitic and driven

2-74. In a frequency-sensitive antenna, the physical length of the serpentine section and its relationship to the wavelength of the applied energy determines which of the following characteristics?

1. Output power
2. Beam narrowness
3. Antenna efficiency
4. Direction of the beam

2-75. In a horizontal-slot antenna, the polarization of the energy is radiated in what direction?

1. Vertical
2. Spherical
3. Horizontal
4. Omnidirectional

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 12—Modulation

NAVEDTRA 14184

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 12 of a series.

History of the course:

Sep 1998: Original edition released.
May 2003: Administrative update released. Technical content was reviewed and minor changes *made.*

NAVSUP Logistics Tracking Number
0504-LP-026-8370

TABLE OF CONTENTS

CHAPTER 1

AMPLITUDE MODULATION

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the OCC/ECC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completion of this chapter, you will be able to:

1. Discuss the generation of a sine wave by describing its three characteristics: amplitude, phase, and frequency.
2. Describe the process of heterodyning.
3. Discuss the development of continuous-wave (cw) modulation.
4. Describe the two primary methods of cw communications keying.
5. Discuss the radio frequency (rf) spectrum usage by cw transmissions.
6. Discuss the advantages and disadvantages of cw transmissions.
7. Explain the operation of typical cw transmitter circuitry.
8. Discuss the method of changing sound waves into electrical impulses.
9. Describe the rf usage of an AM signal.
10. Calculate the percent of modulation for an AM signal.
11. Discuss the difference between high- and low-level modulation.
12. Describe the circuit description, operation, advantages, and disadvantages of the following common AM tube/transistor modulating circuits: plate/collector, control grid/base, and cathode/emitter.
13. Discuss the advantages and disadvantages of AM communications.

INTRODUCTION TO MODULATION PRINCIPLES

People have always had the desire to communicate their ideas to others. Communications have not only been desired from a social point of view, but have been an essential element in the building of civilization. Through communications, people have been able to share ideas of mutual benefit to all mankind. Early attempts to maintain communications between distant points were limited by several factors. For example, the relatively short distance sound would carry and the difficulty of hand-carrying messages over great distances hampered effective communications.

As the potential for the uses of electricity were explored, scientists in the United States and England worked to develop the telegraph. The first practical system was established in London, England, in 1838. Just 20 years later, the final link to connect the major countries with electrical communications was completed when a transatlantic submarine cable was connected. Commercial telegraphy was practically worldwide by 1890. The telegraph key, wire lines, and Morse code made possible almost instantaneous communications between points at great distances. Submarine cables solved the problems of transoceanic communications, but communications with ships at sea and mobile forces were still poor.

In 1897 Marconi demonstrated the first practical wireless transmitter. He sent and received messages over a distance of 8 miles. By 1898 he had demonstrated the usefulness of wireless telegraph communications at sea. In 1899 he established a wireless telegraphic link across the English Channel. His company also established general usage of the wireless telegraph between coastal light ships (floating lighthouses) and land. The first successful transatlantic transmissions were achieved in 1902. From that time to the present, radio communication has grown at an extraordinary rate. Early systems transmitted a few words per minute with doubtful reliability. Today, communications systems reliably transmit information across millions of miles.

The desire to communicate directly by voice, at a higher rate of speed than possible through basic telegraphy, led to further research. That research led to the development of MODULATION. Modulation is the ability to impress intelligence upon a TRANSMISSION MEDIUM, such as radio waves. A transmission medium can be described as light, smoke, sound, wire lines, or radio-frequency waves. In this module, you will study the basic principles of modulation and DEMODULATION (removing intelligence from the medium).

In your studies, you will learn about modulation as it applies to radio-frequency communications. To modulate is to impress the characteristics (intelligence) of one waveform onto a second waveform by varying the amplitude, frequency, phase, or other characteristics of the second waveform. First, however, you will review the characteristics and generation of a sine wave. This review will help you to better understand the principles of modulation. Then, an important principle called HETERODYNING (mixing two frequencies across a nonlinear impedance) will be studied and applied to modulation. Nonlinear impedance will be discussed in the heterodyning section. You will also study several methods of modulating a radio-frequency carrier. You will come to a better understanding of the demodulation principle by studying the various circuits used to demodulate a modulated carrier.

Q-1. What is modulation?

Q-2. What is a transmission medium?

Q-3. What is heterodyning?

Q-4. What is demodulation?

SINE WAVE CHARACTERISTICS

The basic alternating waveform for all complex waveforms is the sine wave. Therefore, an understanding of sine wave characteristics and how they can be acted upon is essential for you to understand modulation. You may want to review sine waves in chapter 1 of *NEETS*, Module 2, *Introduction to Alternating Current and Transformers* at this point.

GENERATION OF SINE WAVES

Since numbers represent individual items in a group, arrows can be used to represent quantities that have magnitude and direction. This may be done by using an arrow and a number, as illustrated in figure 1-1, view (A). The number represents the magnitude of force and the arrow represents the direction of the force.

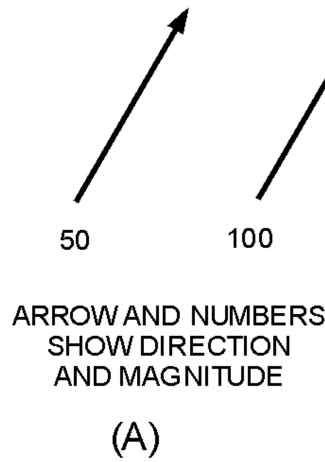


Figure 1-1A.—Vectors representing magnitude and direction.

View (B) illustrates a simpler method of representation. In this method, the length of the arrow is proportional to the magnitude of force, and the direction of force is indicated by the direction of the arrow. Thus, if an arrow 1-inch long represents 50 pounds of force, then an arrow 2-inches long would represent 100 pounds of force. This method of showing both magnitude and direction is called a VECTOR. To more clearly show the relationships between the amplitude, phase, and frequency of a sine wave, we will use vectors.

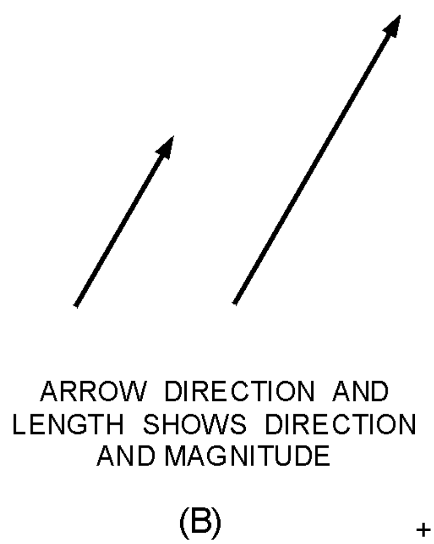


Figure 1-1B.—Vectors representing magnitude and direction.

Vector Applied to Sine-Wave Generation

As covered, in *NEETS*, Module 2, *Introduction to Alternating Current and Transformers*, an alternating current is generated by rotating a coil in the magnetic field between two magnets. As long as the magnetic field is uniform, the output from the coil will be a sine wave, as shown in figure 1-2. This wave shape is called a sine wave because the voltage of the coil depends on its angular position in the magnetic field.

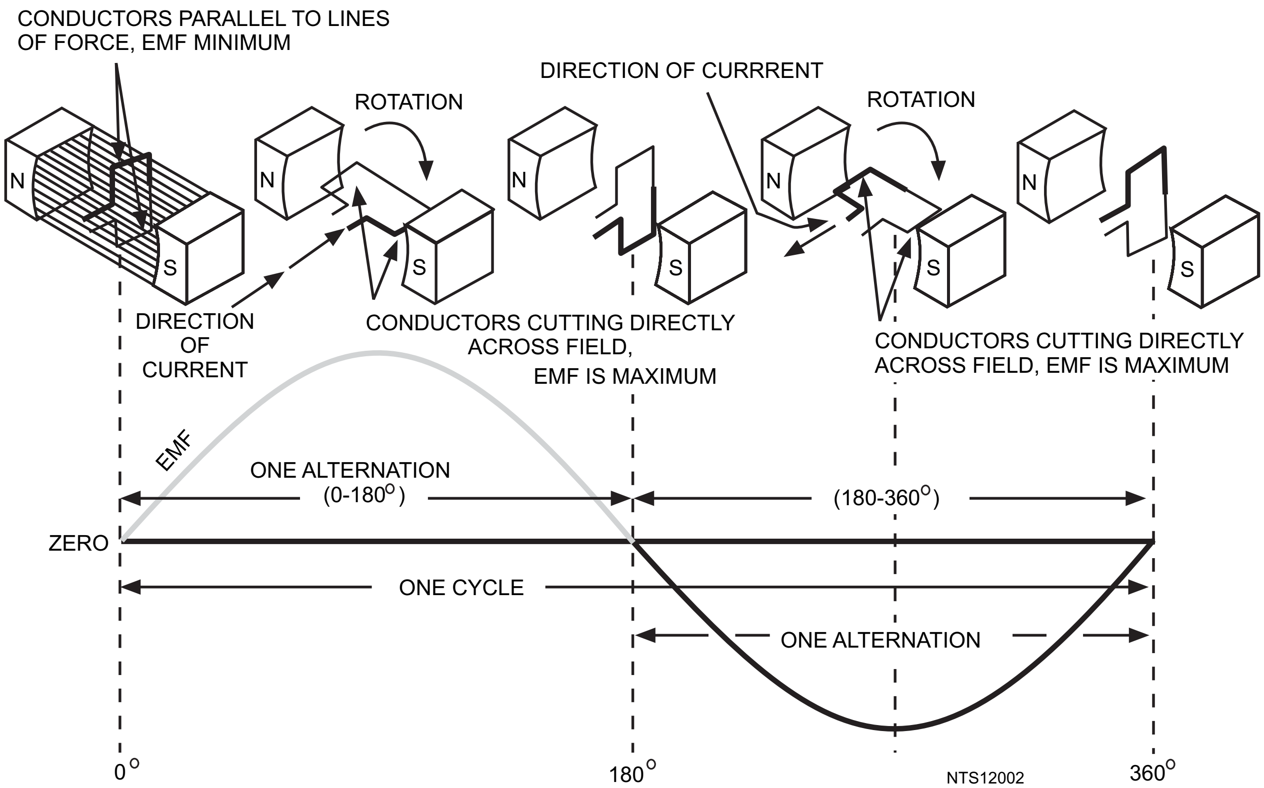


Figure 1-2.—Sine-wave generator.

This relationship can be expressed mathematically by the formula:

$$e = E_{max} \text{ sine} \theta$$

Where: e = instantaneous value of the voltage developed when the coil is at some angle (θ)

E_{max} = maximum value of the voltage

θ = angular position of the coil

You should recall that the trigonometric ratio (inset in figure 1-3) for the sine in a right triangle (a triangle in which one angle is 90 degrees) is:

$$\text{sine } \theta = \frac{\text{opposite side}}{\text{hypotenuse}}$$

Where:

θ = acute angle

opposite side = side of the triangle taht is opposite the angle θ

hypotenuse = the longest side of the triangle

When an alternating waveform is generated, the coil is represented by a vector which has a length that is equal to the maximum output voltage (E_{max}). The output voltage at any given angle can be found by applying the above trigonometric function. Because the output voltage is in direct relationship with the sine of the angle !, it is commonly called a sine wave.

You can see this relationship more clearly in figure 1-3 where the coil positions in relation to time are represented by the numbers 0 through 12. The corresponding angular displacements, shown as !, are shown along the horizontal time axis. The induced voltages (V_1 through V_{12}) are plotted along this axis. Connecting the induced voltage points, shown in the figure, forms a sine-wave pattern. This relationship can be proven by taking any coil position and applying the trigonometric function to an equivalent right triangle. When the vector is placed horizontally (position 0), the angle ! is 0 degrees. Since e = E_{max} sine !, and the sine of 0 degrees is 0, the output voltage is 0 volts, as shown below:

Where:

E_{max} = 100V

θ = 0 degrees

sine θ = 0

Solution:

$e = E_{max} \text{ sine } \theta$

$e = (100V)(0)$

$e = 0 V$

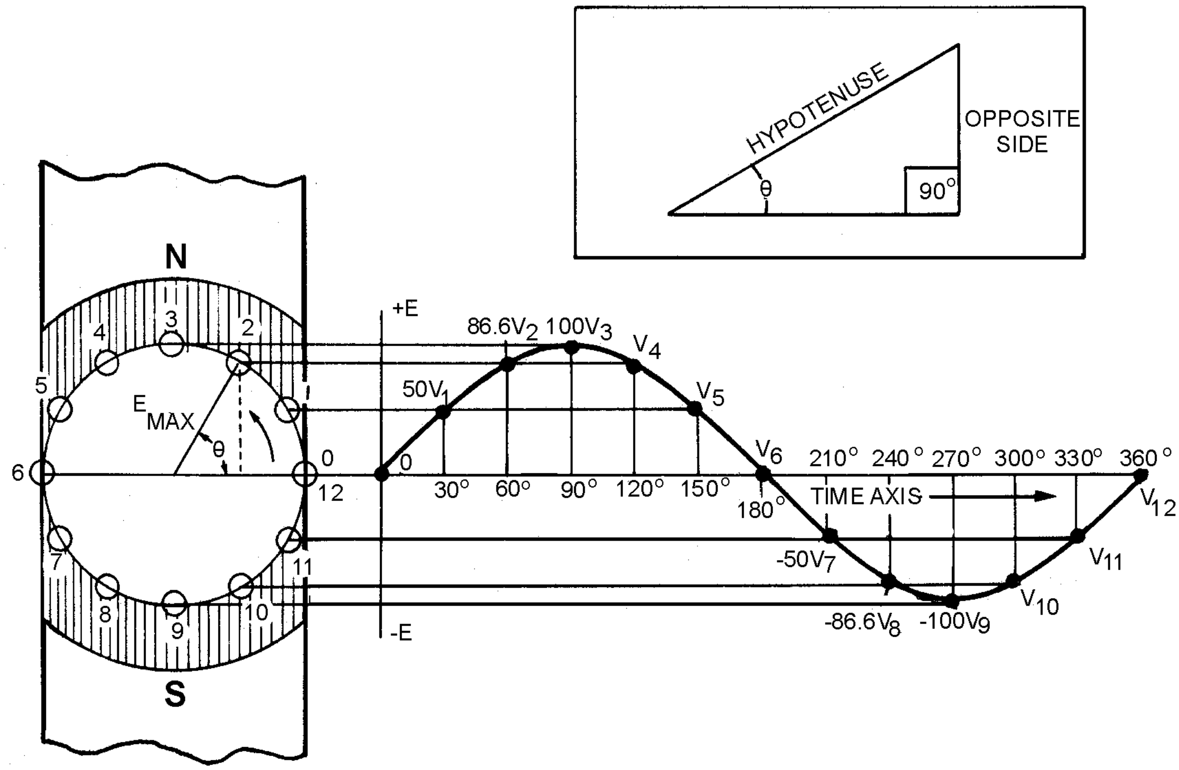


Figure 1-3.—Generation of sine-wave voltage.

At position 2, the sine of 60 degrees is 0.866 and an output of 86.6 volts is developed.

Where:

$E_{max} = 100V$

$\text{sine } \theta = 0.866$

Solution:

$e = (100V)(0.886)$

$e = 86.6V$

This relationship is plotted through 360 degrees of rotation. A continuous line is drawn through the successive points and is known as 1 CYCLE of a sine wave. If the time axis were extended for a second revolution of the vector plotted, you would see 2 cycles of the sine wave. The 0-degree point of the second cycle would be the same point as the 360-degree point of the first cycle.

Q-5. What waveform is the basis of all complex waveforms?

Q-6. What is the purpose of using vectors?

Q-7. What is the trigonometric ratio for the sine of an angle?

Q-8. What is the mathematical formula for computing the output voltage from a moving coil in a magnetic field?

AMPLITUDE

A sine wave is used to represent values of electrical current or voltage. The greater its height, the greater the value it represents. As you have studied, a sine wave alternately rises above and then falls below the reference line. That part above the line represents a positive value and is referred to as a POSITIVE ALTERNATION. That part of the cycle below the line has a negative value and is referred to as a NEGATIVE ALTERNATION. The maximum value, above or below the reference line, is called the PEAK AMPLITUDE. The value at any given point along the reference line is called the INSTANTANEOUS AMPLITUDE.

PHASE

PHASE or PHASE ANGLE indicates how much of a cycle has been completed at any given instant. This merely describes the angle that exists between the starting point of the vector and its position at that instant. The number of degrees of vector rotation and the number of degrees of the resultant sine wave that have been completed will be the same. For example, at time position 2 of figure 1-3, the vector has rotated to 60 degrees and 60 degrees of the resultant sine wave has been completed. Therefore, both are said to have an instantaneous phase angle of 60 degrees.

FREQUENCY

The rate at which the vector rotates determines the FREQUENCY of the sine wave that is generated; that is, the faster the vector rotates, the more cycles completed in a given time period. The basic time period used is 1 second. If a vector completes one revolution per 1 second, the resultant sine wave has a frequency 1 cycle per second (1 hertz). If the rate of rotation is increased to 1,000 revolutions per second, the frequency of the sine wave generated will be 1,000 cycles per second (1 kilohertz).

PERIOD

Another term that is important in the discussion of a sine wave is its duration, or PERIOD. The period of a cycle is the elapsed time from the beginning of a cycle to its completion. If the vector shown in figure 1-3 were to make 1 revolution per second, each cycle of the resultant sine wave would have a period of 1 second. If it were rotating at a speed of 1,000 revolutions per second, each revolution would require 1/1,000 of a second and the period of the resultant sine wave would be 1/1,000 of a second. This illustrates that the period is related to the frequency. As the number of cycles completed in 1 second increases, the period of each cycle will decrease proportionally. This relationship is shown in the following formulas:

Where:

t = period in seconds
f = frequency in hertz

$$f = \frac{1}{t}$$

or

$$t = \frac{1}{f}$$

WAVELENGTH

The WAVELENGTH of a sine wave is determined by its physical length. During the period a wave is being generated, its leading edge is moving away from the source at 300,000,000 meters per second. The physical length of the sine wave is determined by the amount of time it takes to complete one full cycle. This wavelength is an important factor in determining the size of equipments used to generate and transmit radio frequencies.

To help you understand the magnitude of the distance a wavefront (the initial part of a wave) travels during 1 cycle, we will compute the wavelengths (l;) of several frequencies. Consider a vector that rotates at 1 revolution per second. The resultant sine wave is transmitted into space by an antenna. As the vector moves from its 0-degree starting position, the wavefront begins to travel away from the antenna. When the vector reaches the 360-degree position, and the sine wave is completed, the sine wave is stretched out over 300,000,000 meters. The reason the sine wave is stretched over such a great distance is that the wavefront has been moving away from the antenna at 300,000,000 meters per second. This is shown in the following example:

$$\lambda = \text{rate of travel} \times \text{period}$$

$$\lambda = (300{,}000{,}000 \frac{\text{meters}}{\text{second}})(1 \text{ second})$$

$$\lambda = 300{,}000{,}000 \text{ meters}$$

If a vector were rotating at 1,000 revolutions per second, its period would be 0.001 second. By applying the formula for wavelength, you would find that the wavelength is 300,000 meters:

$$\lambda = \left(300{,}000{,}000 \frac{\text{meters}}{\text{second}}\right)(0.001 \quad \text{second})$$

$$\lambda = 300{,}000 \quad \text{meters}$$

Since we normally know the frequency of a sine wave instead of its period, the wavelength is easier to find using the frequency:

$$\lambda = \frac{\text{rate of travel}}{\text{frequency}}$$

Thus, for a sine wave with a frequency of 1,000,000 hertz (1 megahertz), the wavelength would be 300 meters, as shown below:

$$\lambda = \frac{\text{rate of travel}}{\text{frequency}}$$

$$\lambda = \frac{300{,}000{,}000 \ \frac{\text{meters}}{\text{second}}}{1{,}000{,}000 \text{ hertz}}$$

$$\lambda = 300 \text{ meters}$$

The *higher* the frequency, the *shorter* the wavelength of a sine wave. This important relationship between frequency and wavelength is illustrated in table 1-1.

Table 1-1.—Radio frequency versus wavelength

FREQUENCY	WAVELENGTH	
	METRIC	U.S.
300,000 MHz	.001 m	.04 in
EHF-		
30,000 MHz	.01 m	.39 in
SHF-		
3,000 MHz	.1 m	3.94 in
UHF--		
300 MHz	1 m	39.37 in
VHF---		
30 MHz	10 m	10.93 yd
HF----		
3 MHz	100 m	109.4 yd
MF----		
300 kHz	1 km	.62 mi
LF----		
30 kHz	10 km	6.2 mi
VLF-----		
3 kHz	100 km	62 mi

Q-9. What is the instantaneous amplitude of a sine wave?

Q-10. What term describes how much of a cycle has been completed?

Q-11. What determines the frequency of a sine wave?

Q-12. What is the period of a cycle?

Q-13. How do you calculate the wavelength of a sine wave?

HETERODYNING

Information waveforms are produced by many different sources and are generally quite low in frequency. A good example is the human voice. The frequencies involved in normal speech vary from one individual to another and cover a wide range. This range can be anywhere from a low of about 90 hertz for a deep bass to as high as 10 kilohertz for a high soprano.

The most important speech frequencies almost entirely fall below 3 kilohertz. Higher frequencies merely help to achieve more perfect sound production. The range of frequencies used to transmit voice intelligence over radio circuits depends on the degree of FIDELITY (the ability to faithfully reproduce the input in the output) that is desired. The minimum frequency range that can be used for the transmission of speech is 500 to 2,000 hertz. The average range used on radiotelephone circuits is 250 to 2,750 hertz.

Frequencies contained within the human voice can be transmitted over telephone lines without difficulty, but transmitting them via radio circuits is not practical. This is because of their extremely long wavelengths and the fact that antennas would have to be constructed with long physical dimensions to transmit or radiate these wavelengths. Generally, antennas have radiating elements that are 1/4, 1/2, 1, or more full wavelengths of the frequency to be radiated. The wavelengths of voice frequencies employed on radiotelephone circuits range from 1,200,000 meters at 250 hertz to 109,090 meters at 2,750 hertz. Even a quarter-wave antenna would require a large area, be expensive to construct, and consume enormous amounts of power.

As studied in *NEETS*, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*, radio frequencies do not have the limitations just described for voice frequencies. Radio waves, given a suitable antenna, can often radiate millions of miles into space. Several methods of modulation can be used to impress voices frequencies onto radio waves for transmission through space.

In the modulation process, waves from the information source are impressed onto a radio-frequency sine wave called a CARRIER. This carrier is sufficiently high in frequency to have a wavelength short enough to be radiated from an antenna of practical dimensions. For example, a carrier frequency of 10 megahertz has a wavelength of 30 meters, as shown below:

$$\lambda = \frac{\text{rate of travel}}{\text{frequency}}$$

$$\lambda = \frac{300{,}000{,}000 \dfrac{\text{meters}}{\text{second}}}{10{,}000{,}000 \text{ hertz}}$$

$$\lambda = 30 \text{ meters } (98.4 \text{ feet})$$

Construction of an antenna related to that wavelength does not cause any problems.

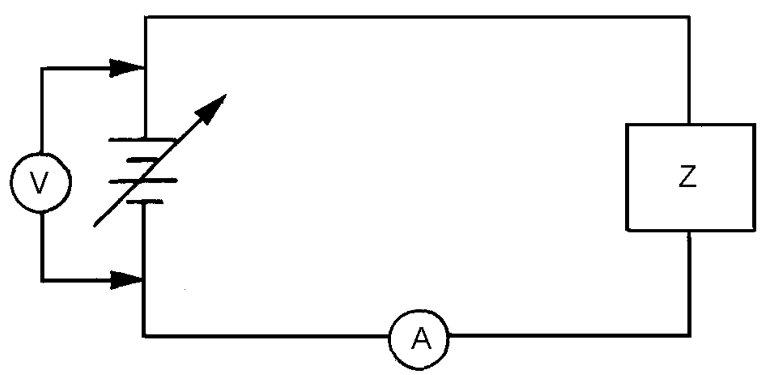
An information wave is normally referred to as a MODULATING WAVE. When a modulating wave is impressed on a carrier, the voltages of the modulating wave and the carrier are combined in such a manner as to produce a COMPLEX WAVE (a wave composed of two or more parts). This complex wave is referred to as the MODULATED WAVE and is the waveform that is transmitted through space. When the modulated wave is received and demodulated, the original component waves (carrier and modulating waves) are reproduced with their respective frequencies, phases, and amplitudes unchanged.

Modulation of a carrier can be achieved by any of several methods. Generally, the methods are named for the sine-wave characteristic that is altered by the modulation process. In this module, you will study AMPLITUDE MODULATION, which includes CONTINUOUS-WAVE MODULATION. You will also learn about two forms of ANGLE MODULATION (FREQUENCY MODULATION and PHASE MODULATION). A special type of modulation, known as PULSE MODULATION, will also be discussed. Before we present the methods involved in developing modulation, you need to study a process that is essential to the modulation of a carrier, known as heterodyning.

To help you understand the operation of heterodyning circuits, we will begin with a discussion of LINEAR and NONLINEAR devices. In linear devices, the output rises and falls directly with the input. In nonlinear devices, the output does *not* rise and fall directly with the input.

LINEAR IMPEDANCE

Whether the impedance of a device is linear or nonlinear can be determined by comparing the change in *current through* the device to the change in *voltage applied* to the device. The simple circuit shown in view (A) of figure 1-4 is used to explain this process.



(A)

Figure 1-4A.—Circuit with one linear impedance.

First, the current through the device must be measured as the voltage is varied. Then the current and voltage values can be plotted on a graph, such as the one shown in view (B), to determine the impedance of the device. For example, assume the voltage is varied from 0 to 200 volts in 50-volt steps, as shown in view (B). At the first 50-volt point, the ammeter reads 0.5 ampere. These ordinates are plotted as point **a** in view (B). With 100 volts applied, the ammeter reads 1 ampere; this value is plotted as point **b**. As these steps are continued, the values are plotted as points **c** and **d**. These points are connected with a straight line to show the linear relationship between current and voltage. For every change in voltage applied to the device, a proportional change occurs in the current through the device. When the change in current is proportional to the change in applied voltage, the impedance of the device is linear and a straight line is developed in the graph.

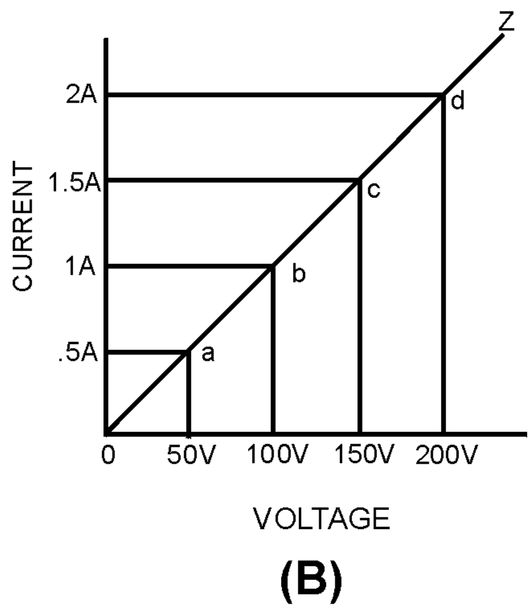


Figure 1-4B.—Circuit with one linear impedance.

The principle of linear impedance can be extended by connecting two impedance devices in series, as shown in figure 1-5, view (A). The characteristics of both individual impedances are determined as explained in the preceding section. For example, assume voltmeter V1 shows 50 volts and the ammeter shows 0.5 ampere. Point **a** in view (B) represents this ordinate. In the same manner, increasing the voltage in increments of 50 volts gives points **b**, **c**, and **d**. Lines Z1 and Z2 show the characteristics of the two impedances. The total voltage of the series combination can be determined by adding the voltages across Z1 and Z2. For example, at 0.5 ampere, point **a** (50 volts) plus point **e** (75 volts) produces point **i** (125 volts). Also, at 1 ampere, point **b** plus point **f** produces point **j**. Line Z1 + Z2 represents the combined voltage-current characteristics of the two devices.

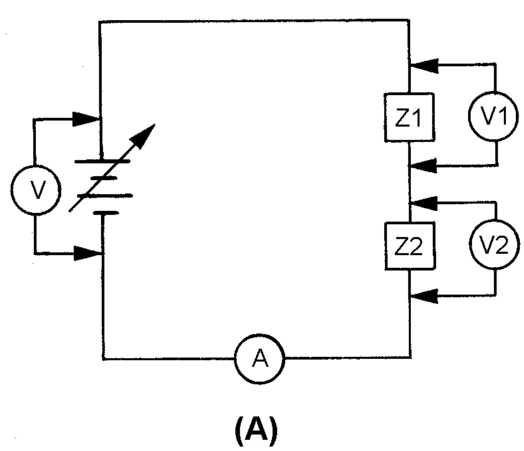


Figure 1-5A.—Circuit with two linear impedances.

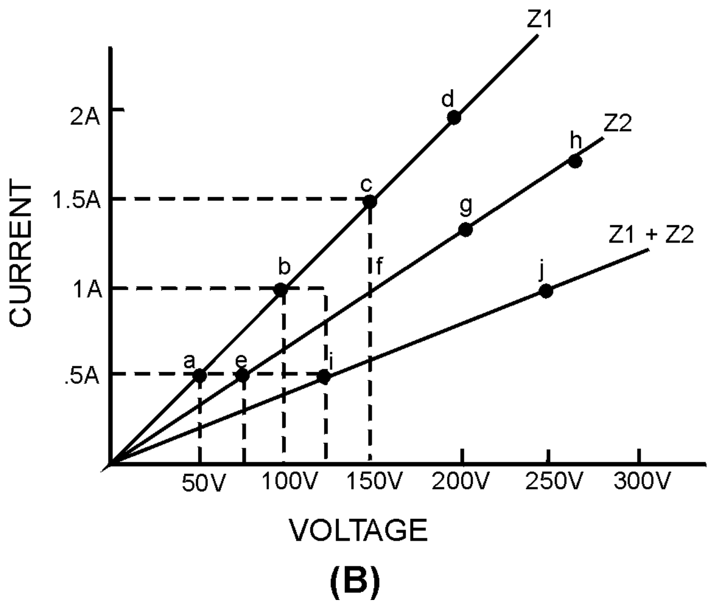


Figure 1-5B.—Circuit with two linear impedances.

View (A) of figure 1-6 shows two impedances in parallel. View (B) plots the impedances both individually (Z1 and Z2) and combined (Z1 x Z2)/(Z1 + Z2). Note that Z1 and Z2 are not equal. At 100 volts, Z1 has 1 ampere of current plotted at point **b** and Z2 has 0.5 ampere plotted at point **f**. The coordinates of the equivalent impedance of the parallel combination are found by adding the current through Z1 to the current through Z2. For example, at 100 volts, point **b** is added to point **f** to determine point **j** (1.5 amperes).

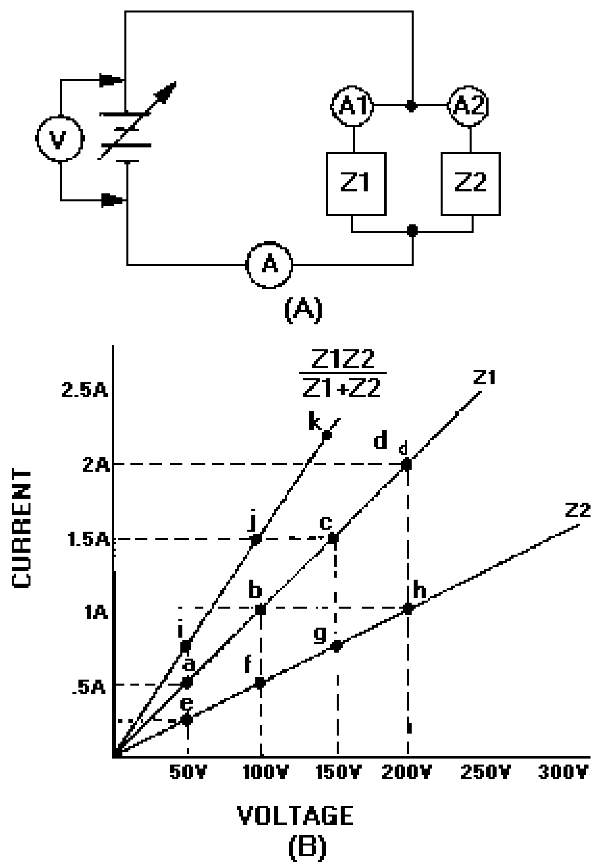


Figure 1-6.—Circuit with parallel linear impedances.

Positive or negative voltage values can be used to plot the voltage-current graph. Figure 1-7 shows an example of this situation. First, the voltage versus current is plotted with the battery polarity as shown in view (A). Then the battery polarity is reversed and the remaining voltage versus current points are plotted. As a result, the line shown in view (C) is obtained.

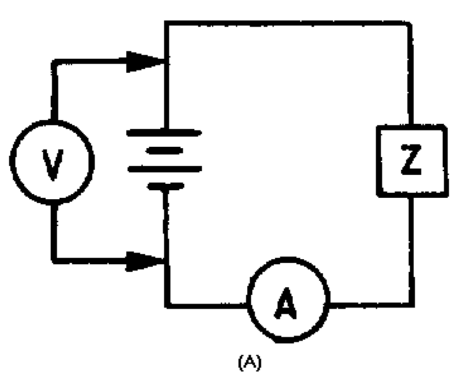


Figure 1-7A.—Linear impedance circuit.

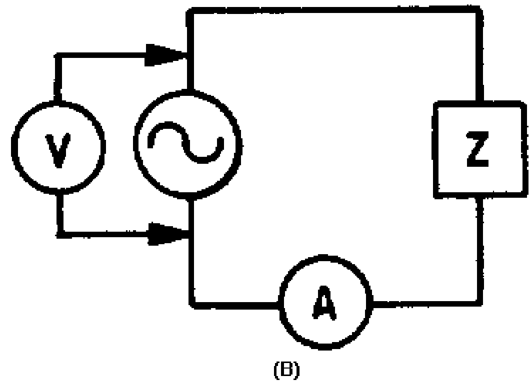


Figure 1-7B.—Linear impedance circuit.

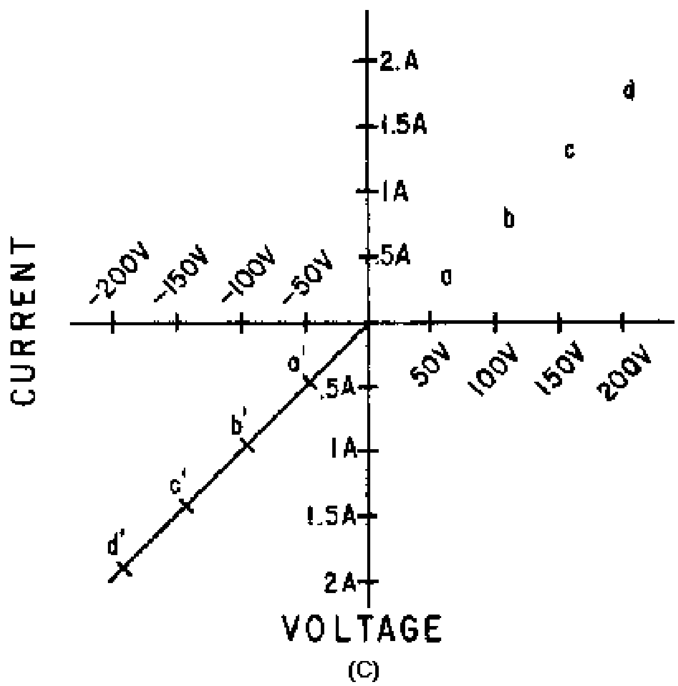


Figure 1-7C.—Linear impedance circuit

The battery in view (A) could be replaced with an ac generator, as shown in view (B), to plot the characteristic chart. The same linear voltage-current chart would result. Current flow in either direction is directly proportional to the change in voltage.

In conclusion, when dc or sine-wave voltages are applied to a linear impedance, the current through the impedance will vary directly with a change in the voltage. The device could be a resistor, an air-core inductor, a capacitor, or any other linear device. In other words, if a sine-wave generator output is applied to a combination of linear impedances, the resultant current will be a sine wave which is directly proportional to the change in voltage of the generator. The linear impedances do not alter the waveform of the sine wave. The amplitude of the voltage developed across each linear component may vary, or the phase of the wave may shift, but the shape of the wave will remain the same.

NONLINEAR IMPEDANCE

You have studied that a linear impedance is one in which the resulting current is directly proportional to a change in the applied voltage. A nonlinear impedance is one in which the resulting current is not directly proportional to the change in the applied voltage. View (A) of figure 1-8 illustrates a circuit which contains a nonlinear impedance (Z), and view (B) shows its voltage-current curve.

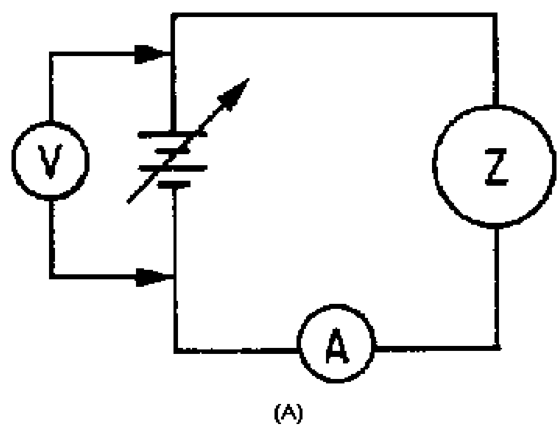


Figure 1-8A.—Nonlinear impedance circuit.

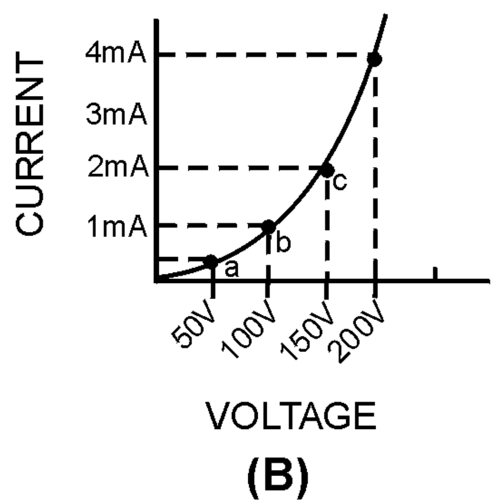


Figure 1-8B.—Nonlinear impedance circuit.

As the applied voltage is varied, ammeter readings which correspond with the various voltages can be recorded. For example, assume that 50 volts yields 0.4 milliampere (point **a**), 100 volts produces 1 milliampere (point **b**), and 150 volts causes 2.2 milliamperes (point **c**). Current through the nonlinear impedance does not vary proportionally with the voltage; the chart is not a straight line. Therefore, Z is a nonlinear impedance; that is, the current through the impedance does not faithfully follow the change in voltage. Various combinations of voltage and current for this particular nonlinear impedance may be obtained by use of this voltage-current curve.

COMBINED LINEAR AND NONLINEAR IMPEDANCES

The series combination of a linear and a nonlinear impedance is illustrated in view (A) of figure 1-9. The voltage-current charts of Z1 and Z2 are shown in view (B). A chart of the combined impedance can be plotted by adding the amount of voltage required to produce a particular current through linear impedance Z1 to the amount of voltage required to produce the same amount of current through nonlinear impedance Z2. The total will be the amount of voltage required to produce that particular current through the series combination. For example, point **a** (25 volts) is added to point **c** (50 volts) which yields point **e** (75 volts); and point **b** (50 volts) is added to point **d** (100 volts) which yields point **f** (150 volts). Intermediate points may be determined in the same manner and the resultant characteristic curve (Z1 + Z2) is obtained for the series combination.

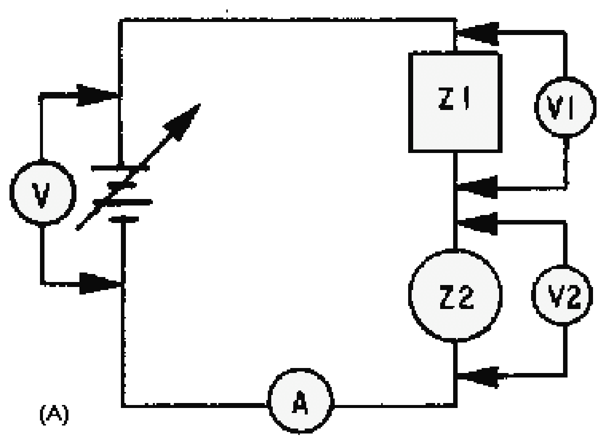


Figure 1-9A.—Combined linear and nonlinear impedances.

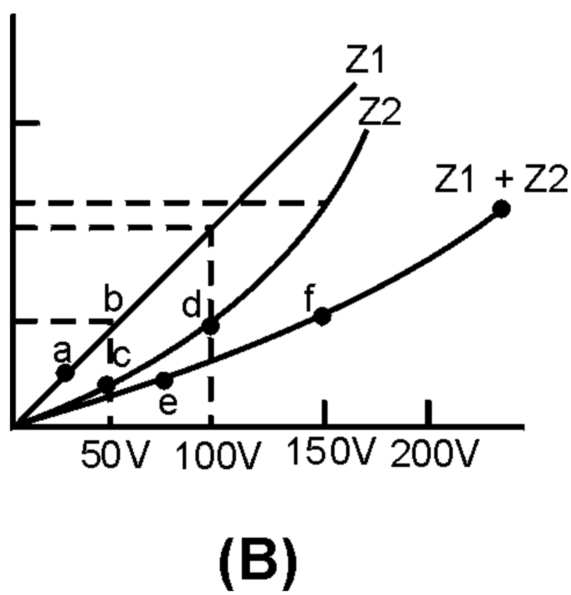


Figure 1-9B.—Combined linear and nonlinear impedances.

You should see from this graphic analysis that when a linear impedance is combined with a nonlinear impedance, the resulting characteristic curve is nonlinear. Some examples of nonlinear impedances are crystal diodes, transistors, iron-core transformers, and electron tubes.

AC APPLIED TO LINEAR AND NONLINEAR IMPEDANCES

Figure 1-10 illustrates an ac sine-wave generator applied to a circuit containing several linear impedances. A sine-wave voltage applied to linear impedances will cause a sine wave of current through them. The wave shape across each linear impedance will be identical to the applied waveform.

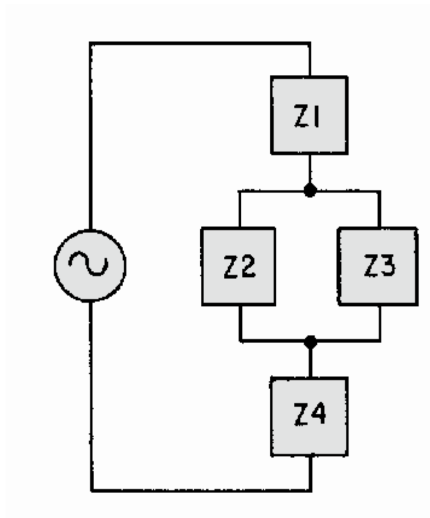
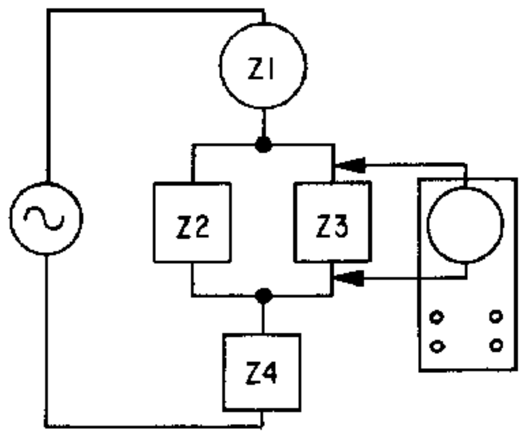


Figure 1-10.—Sine wave generator applied to several impedances.

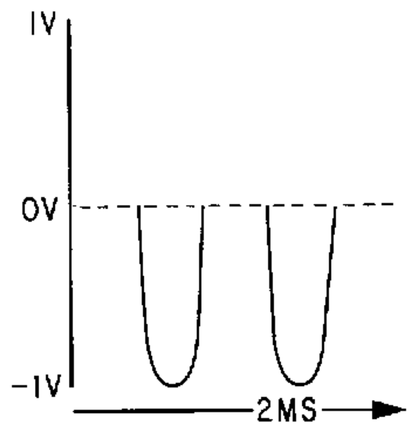
The amplitude, on the other hand, may differ from the amplitude of the applied voltage. Furthermore, the phase of the voltage developed by any of the impedances may not be identical to the phase of the voltage across any of the other impedances or the phase of the applied voltage. If an impedance is a reactive component (coil or capacitor), voltage or current may lead or lag, but the wave shape will remain the same. In a linear circuit, the output of the generator is not distorted. The frequency remains the same throughout the entire circuit and no new frequencies are generated.

View (A) of figure 1-11 illustrates a circuit that contains a combination of linear and nonlinear impedances with a sine wave of voltage applied. Impedances Z2, Z3, and Z4 are linear; and Z1 is nonlinear. The result of a linear and nonlinear combination of impedances is a nonlinear waveform. The curve Z, shown in view (B), is the nonlinear curve for the circuit of view (A). Because of the nonlinear impedance, current can flow in the circuit only during the positive alternation of the sine-wave generator. If an oscilloscope is connected, as shown in view (A), the waveform across Z3 will not be a sine wave. Figure 1-12, view (A), illustrates the sine wave from the generator and view (B) shows the waveform across the linear impedance Z3. Notice that the nonlinear impedance Z1 has eliminated the negative half cycles.



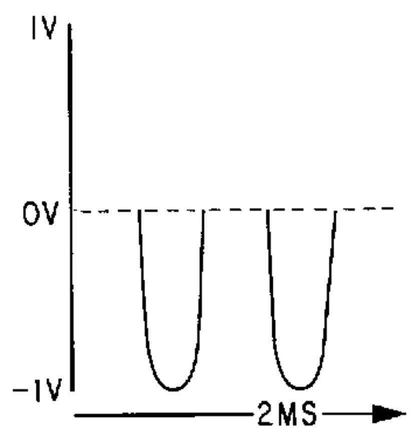
(A)

Figure 1-11A.—Circuit with nonlinear impedances.



(A)

Figure 1-11B.—Circuit with nonlinear impedances.



(A)

Figure 1-12A.—Waveform in a circuit with nonlinear impedances.

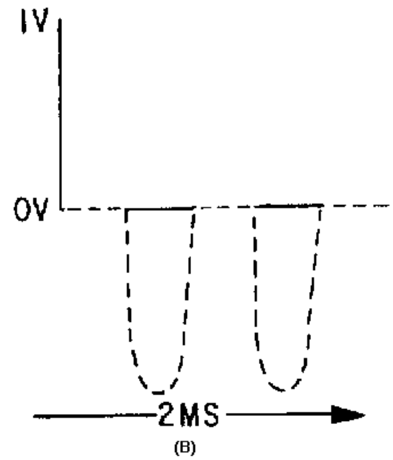


Figure 1-12B.—Waveform in a circuit with nonlinear impedances.

The waveform in view (B) is no longer identical to that of view (A) and the nonlinear impedance network has generated HARMONIC FREQUENCIES. The waveform now consists of the fundamental frequency and its harmonics. (Harmonics were discussed in *NEETS*, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*.)

TWO SINE WAVE GENERATORS IN LINEAR CIRCUITS

A circuit composed of two sine-wave generators, G1 and G2, and two linear impedances, Z1 and Z2, is shown in figure 1-13. The voltage applied to Z1 and Z2 will be the vector sum of the generator voltages. The sum of the individual instantaneous voltages across each impedance will equal the applied voltages.

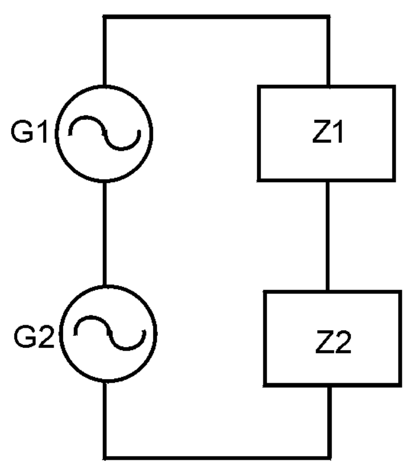


Figure 1-13.—Two sine-wave generators with linear impedances.

If the two generator outputs are of the same frequency, then the waveform across Z1 and Z2 will be a sine wave, as shown in figure 1-14, views (A) and (B). No new frequencies will be created. Relative amplitude and phase will be determined by the relative values and types of the impedances.

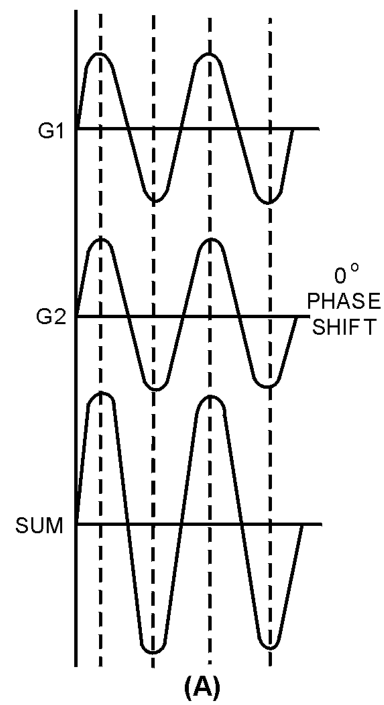


Figure 1-14A.—Waveforms across two nonlinear impedances.

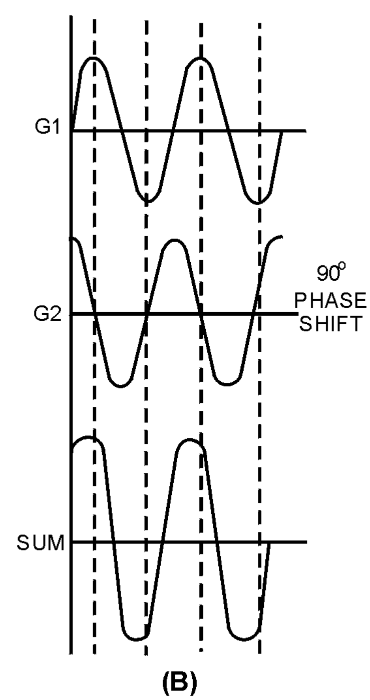


Figure 1-14B.—Waveforms across two nonlinear impedances.

If the two sine wave generators are of different frequencies, then the sum of the instantaneous values will appear as a complex wave across the impedances, as shown in figure 1-15, views (A) and (B). To determine the wave shape across each individual impedance, assume only one generator is connected at a time and compute the sine-wave voltage developed across each impedance for that generator input. Then, combine the instantaneous voltages (caused by each generator input) to obtain the complex waveform across each impedance. The nature of the impedance (resistive or reactive) will determine the shape of the complex waveform. Because the complex waveform is the sum of two individual sine waves, the composite waveform contains only the two original frequencies.

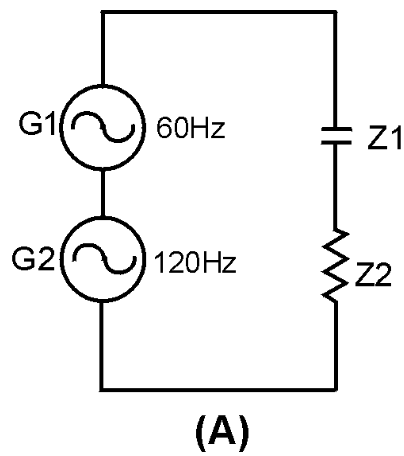


Figure 1-15A.—Sine-wave generators with different frequencies and linear impedances.

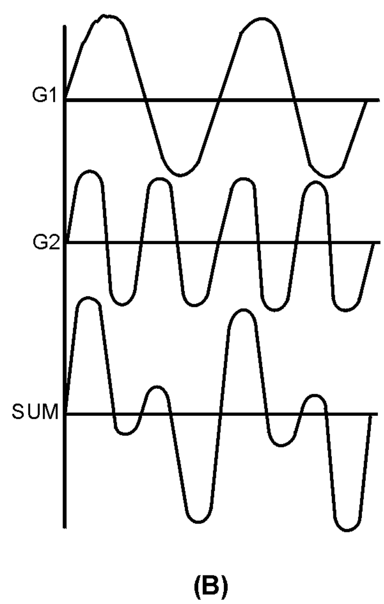


Figure 1-15B.—Sine-wave generators with different frequencies and linear impedances.

Linear impedances may alter complex waveforms, but they do not produce new frequencies. The output of one generator does not influence the output of the other generator.

TWO SINE WAVE GENERATORS AND A COMBINATION OF LINEAR AND NONLINEAR IMPEDANCES

Figure 1-16 illustrates a circuit that contains two sine-wave generators (G1 and G2), linear impedance Z1, and nonlinear impedance Z2, in series. When a single sine-wave voltage is applied to a combined linear and nonlinear impedance circuit, the voltages developed across the impedances are complex waveforms.

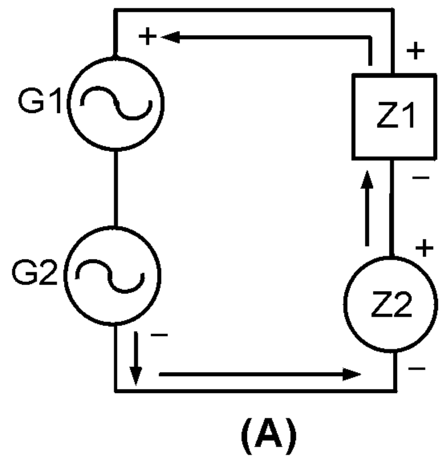


Figure 1-16.—Sine-wave generators with a combination of impedances.

When two sine wave voltages are applied to a circuit, as in figure 1-16, nonlinear impedance Z2 reshapes the two sine-wave inputs and their harmonics, resulting in a very complex waveform.

Assume that nonlinear impedance Z2 will allow current to flow only when the sum of the two sine-wave generators (G1 and G2) has the polarity indicated. The waveforms present across the linear impedance will appear as a varying waveform. This will be a complex waveform consisting of:

- a dc level
- the two fundamental sine wave frequencies
- the harmonics of the two fundamental frequencies
- the sum of the fundamental frequencies
- the difference between frequencies

The sum and difference frequencies occur because the phase angles of the two fundamentals are constantly changing. If generator G1 produces a 10-hertz voltage and generator G2 produces an 11-hertz voltage, the waveforms produced because of the nonlinear impedance will be as shown in the following list:

- a 10-hertz voltage
- an 11-hertz voltage
- harmonics of 10 hertz and 11 hertz (the higher the harmonic, the lower its strength)
- the sum of 10 hertz and 11 hertz (21 hertz)
- the difference between 10 hertz and 11 hertz (1 hertz)

Figure 1-17 illustrates the relationship between the two frequencies (10 and 11 hertz). Since the waveforms are not of the same frequency, the 10 hertz of view (B) and the 11 hertz of view (A) will be in phase at some points and out of phase at other points. You can see this by closely observing the two waveforms at different instants of time. The result of the differences in phase of the two sine waves is shown in view (C). View (D) shows the waveform that results from the nonlinearity in the circuit.

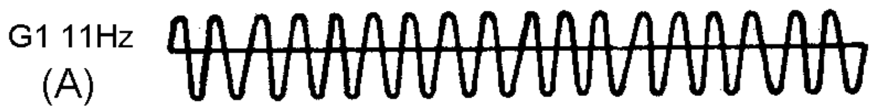


Figure 1-17A.—Frequency relationships.

Figure 1-17B.—Frequency relationships.

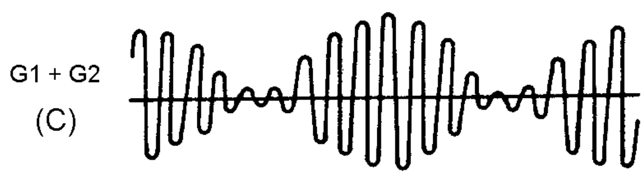


Figure 1-17C.—Frequency relationships.

Figure 1-17D.—Frequency relationships.

The most important point to remember is that when varying voltages are applied to a circuit which contains a nonlinear impedance, the resultant waveform contains frequencies which are not present at the input source.

The process of combining two or more frequencies in a nonlinear impedance results in the production of new frequencies. This process is referred to as heterodyning.

SPECTRUM ANALYSIS

The heterodyning process can be analyzed by using SPECTRUM ANALYSIS (the display of electromagnetic energy arranged according to wavelength or frequency). As shown in figure 1-18, spectrum analysis is an effective way of viewing the energy in electronic circuits. It clearly shows the relationships between the two fundamental frequencies (10 and 11 hertz) and their sum (21 hertz) and difference (1 hertz) frequencies. It also allows you to view the BANDWIDTH (the amount of the frequency spectrum that signals occupy) of the signal you are studying.

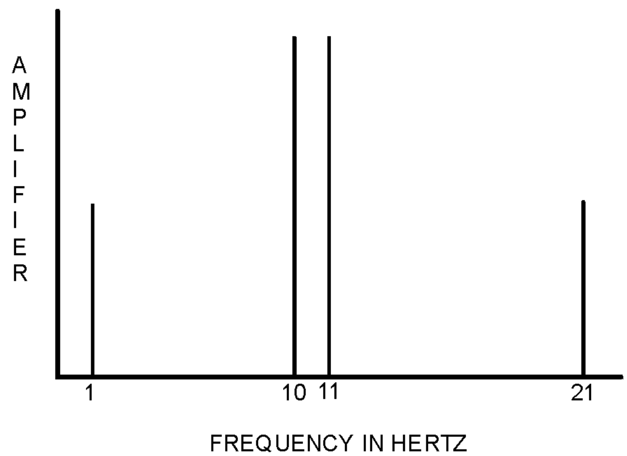


Figure 1-18.—Spectrum analysis of heterodyned signal.

TYPICAL HETERODYNING CIRCUIT

Two conditions must be met in a circuit for heterodyning to occur. First, at least two different frequencies must be applied to the circuit. Second, these signals must be applied to a nonlinear impedance. These two conditions will result in new frequencies (sum and difference) being produced. Any one of the frequencies can be selected by placing a frequency-selective device (such as a tuned tank circuit) in series with the nonlinear impedance in the circuit.

Figure 1-19 illustrates a basic heterodyning circuit. The diode D1 serves as the nonlinear impedance in the circuit. Generators G1 and G2 are signal sources of different frequencies. The primary of T1, with its associated capacitance, serves as the frequency-selective device.

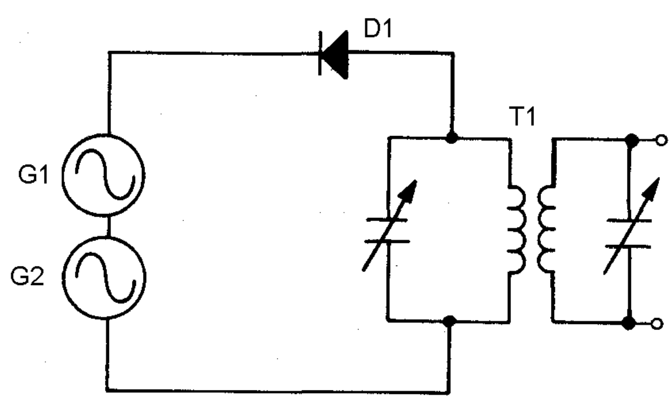


Figure 1-19.—Typical heterodyning circuit.

The principles of this circuit are similar to those of the block diagram circuit of figure 1-16. Notice in figure 1-19 that the two generators are connected in series. Therefore, the resultant waveform of their combined frequencies will determine when the cathode of D1 will be negative with respect to the anode, thereby controlling the conduction of the diode. The new frequencies that are generated by applying these signals to nonlinear impedance D1 are the sum and difference of the two original frequencies. The frequency-selective device T1 may be tuned to whichever frequency is desired for use in later circuit stages. Heterodyning action takes place, intentionally or not, whenever these conditions exist. Heterodyning (MIXING) circuits are found in most electronic transmitters and receivers. These transmitter and receiver circuits will be explained in detail later in this module.

Q-14. Define the heterodyne principle.

Q-15. What is a nonlinear impedance?

Q-16. What is spectrum analysis?

Q-17. What two conditions are necessary for heterodyning to take place?

AMPLITUDE-MODULATED SYSTEMS

Amplitude modulation refers to any method of varying the amplitude of an electromagnetic carrier frequency in accordance with the intelligence to be transmitted by the carrier. The CARRIER frequency is a radio-frequency wave suitable for modulation by the intelligence to be transmitted. One form of this method of modulation is simply to interrupt the carrier in accordance with a prearranged code.

CONTINUOUS WAVE (CW)

The "on-off" KEYING of a continuous wave (cw) carrier frequency was the principal method of modulating a carrier in the early days of electrical communications. The intervals of time when a carrier either was present or absent conveyed the desired intelligence. This is still used in modern communications. When applied to a continuously oscillating radio-frequency source, on-off keying is referred to as cw signaling. This type of communication is sometimes referred to as an interrupted continuous wave (icw).

Development

The use of a cw transmitter can be very simple. All that is required for the transmitter to work properly is a device to generate the oscillations, a method of keying the oscillations on and off, and an antenna to radiate the energy. Continuous wave was the first type of modulation used. It is still extensively used for long-range communications. When Marconi and others were attempting the transfer of intelligence between two points, without reliance on a conducting path, they employed the use of a practical coding system known as Morse code. You probably know that Morse code is a system of on-off keying developed for telegraph that is capable of passing intelligence over wire at an acceptable rate. Morse code consists only of periods of signal and no-signal.

Figure 1-20 is the International Morse code used with telegraphy and cw modulation. Each character in the code is made up of a series of elements referred to as DOTS or DASHES. These are short (dot) and long (dash) bursts of signal separated by intervals of no signal. The dot is the basic time element of the code. The dash has three times the duration of a dot interval. The waveforms for both are shown in figure 1-21. The elements within each character are separated by intervals of no signal with a time duration of one dot. The characters are separated by a no-signal interval equal in duration to one dash. Each interval during which signal is present is called the MARKING interval, and the period of no signal is called the SPACING interval. Figure 1-22 shows the relationships between the rf carrier view (A), the on-off keying waveform view (B), and the resultant carrier wave view (C).

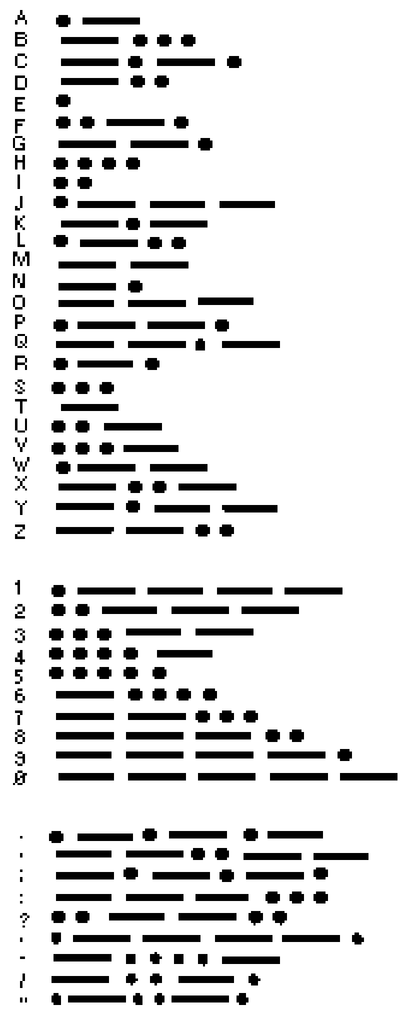


Figure 1-20.—International Morse code.

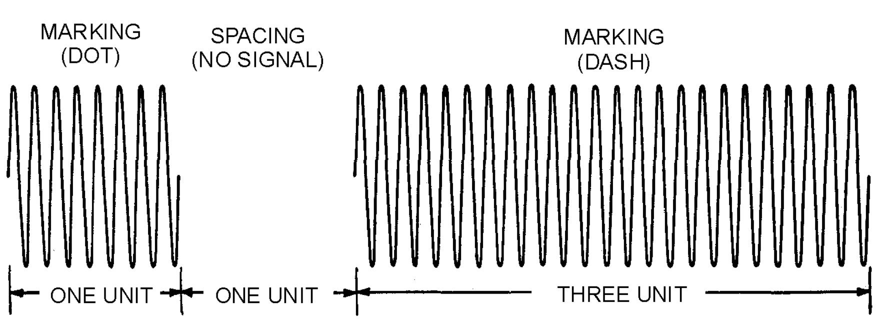
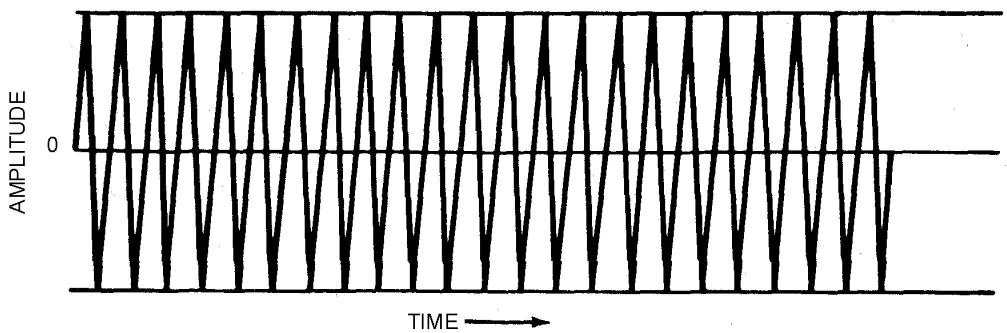


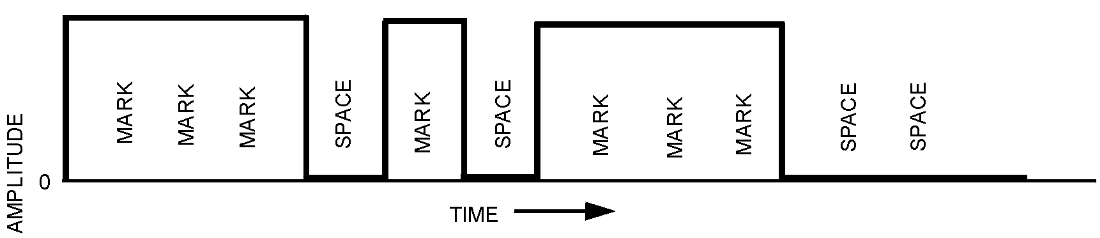
Figure 1-21.—Dot and dash in radiotelegraph code.



RF CARRIER (CW) (EACH CYCLE REPRESENTS SEVERAL THOUSAND OR SEVERAL MILLION CYCLES)

(A)

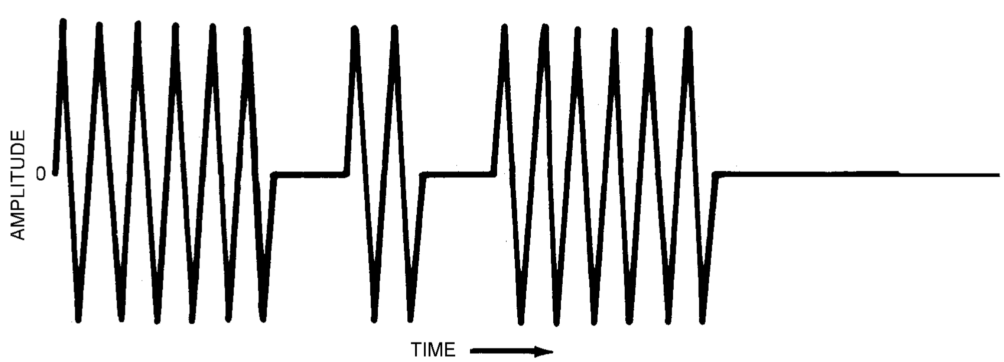
Figure 1-22A.—Essential elements of ON-OFF keying.



ON-OFF KEYED MORSE CODE CHARACTER (LETTER “K”)

(B)

Figure 1-22B.—Essential elements of ON-OFF keying.



RESULTANT CARRIER WAVE TRANSMITTED ALONG CONNECTING PATH.

(C)

Figure 1-22C.—Essential elements of ON-OFF keying.

Keying Methods

Keying a transmitter causes an rf signal to be radiated only when the key contacts are closed. When the key contacts are open, the transmitter does not radiate energy. Keying is accomplished in either the oscillator or amplifier stage of a transmitter. A number of different keying systems are used in Navy transmitters.

In most Navy transmitters, the hand telegraph key is at a low-voltage potential with respect to ground. A keying bar is usually grounded to protect the operator. Generally, a keying relay, with its contacts in the center-tap lead of the filament transformer, is used to key the equipment. Because one or more stages use the same filament transformer, these stages are also keyed. A class C final amplifier, when operated with fixed bias, is usually not keyed. This is because no output occurs when no excitation is applied in class C operation. Keying the final amplifier along with the other stages is not necessary in this case.

OSCILLATOR KEYING.—Two methods of OSCILLATOR KEYING are shown in figure 1-23. In view (A) the grid circuit is closed at all times. The key (K) opens and closes the negative side of the plate circuit. This system is called PLATE KEYING. When the key is open, no plate current can flow and the circuit does not oscillate. In view (B), the cathode circuit is open when the key is open and neither grid current nor plate current can flow. Both circuits are closed when the key is closed. This system is called CATHODE KEYING. Although the circuits of figure 1-23 may be used to key amplifiers, other keying methods are generally employed because of the high values of plate current and voltage encountered.

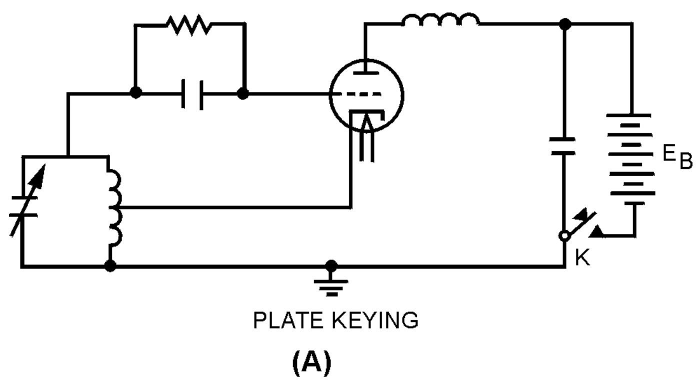


Figure 1-23A.—Oscillator keying.

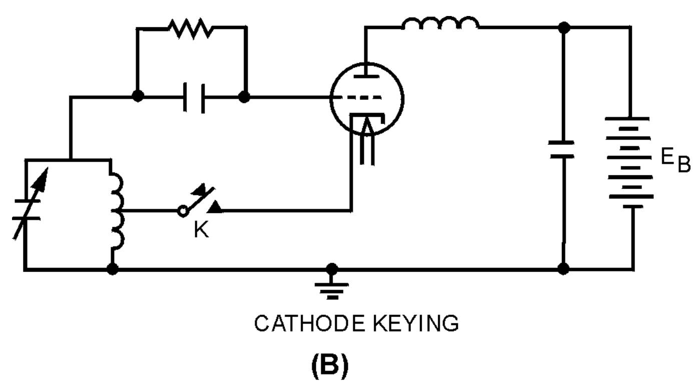


Figure 1-23B.—Oscillator keying.

BLOCKED-GRID KEYING.—Two methods of BLOCKED-GRID KEYING are shown in figure 1-24. The key in view (A) shorts cathode resistor R1 allowing normal plate current to flow. With the key open, reduced plate current flows up through resistor R1 making the end connected to grid resistor R_g negative. If R1 has a high enough value, the bias developed is sufficient to cause cutoff of plate current. Depressing the key short-circuits R1. This increases the bias above cutoff and allows the normal flow of plate current. Grid resistor R_g is the usual grid-leak resistor for normal biasing.

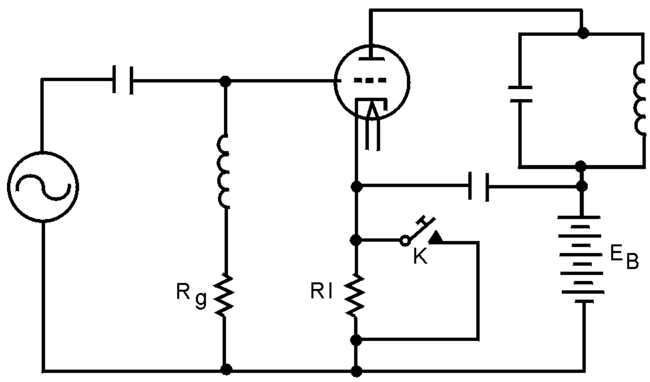


Figure 1-24A.—Blocked-grid keying.

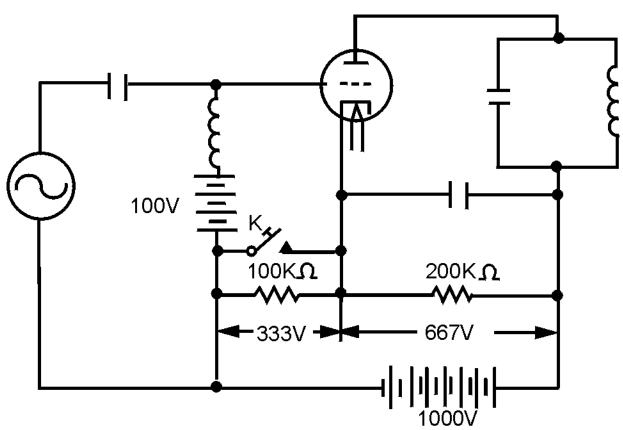


Figure 1-24B.—Blocked-grid keying.

The blocked-grid keying method in view (B) affords a complete cutoff of plate current. It is one of the best methods for keying amplifier stages in transmitters. In the voltage divider with the key open, two-thirds of 1,000 volts, or 667 volts, is developed across the 200-kilohm resistor; one-third of 1,000 volts, or 333 volts, is developed across the 100-kilohm resistor. The grid bias is the sum of 100 volts and 333 volts, or 433 volts. This sum is below cutoff and no plate current flows. The plate voltage is 667 volts. With the key closed, the 200-kilohm resistor drops 1,000 volts. The plate voltage becomes 1,000 volts at the same time the grid bias becomes 100 volts. Grid bias is reduced enough so that the triode amplifier will conduct only on the peaks of the drive signal.

When greater frequency stability is required, the oscillator should not be keyed, but should remain in continuous operation; other transmitter circuits may be keyed. This procedure keeps the oscillator tube at a normal operating temperature and offers less chance for frequency variations to occur each time the key is closed.

KEYING RELAYS.—In transmitters using a crystal-controlled oscillator, the keying is almost always in a circuit stage following the oscillator. In large transmitters (75 watts or higher), the ordinary hand key cannot accommodate the plate current without excessive arcing.

WARNING

BECAUSE OF THE HIGH PLATE POTENTIALS USED, OPERATING A HAND KEY IN THE PLATE CIRCUIT IS DANGEROUS. A SLIGHT SLIP OF THE HAND BELOW THE KEY KNOB COULD RESULT IN SEVERE SHOCK OR, IN THE CASE OF DEFECTIVE RF PLATE CHOKES, A SEVERE RF BURN.

In larger transmitters, some local low-voltage supply, such as a battery, is used with the hand key to open and close a circuit through the coils of a KEYING RELAY. The relay contacts open and close the keying circuit of the amplifier. A schematic diagram of a typical relay-operated keying system is shown in figure 1-25. The hand key closes the circuit from the low-voltage supply through the coil (L) of the keying relay. The relay armature closes the relay contacts as a result of the magnetic pull exerted on the armature. The armature moves against the tension of a spring. When the hand key is opened, the relay coil is deenergized and the spring opens the relay contacts.

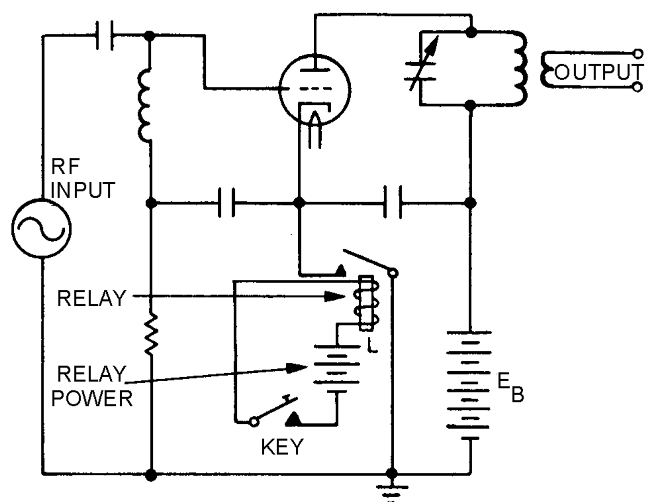


Figure 1-25.—Relay-operated keying system.

KEY CLICKS.—Ideally, cw keying a transmitter should instantly start and stop radiation of the carrier completely. However, the sudden application and removal of power causes rapid surges of current which may cause interference in nearby receivers. Even though such receivers are tuned to frequencies far removed from that of the transmitter, interference may be present in the form of "clicks" or "thumps." KEY-CLICK FILTERS are used in the keying systems of radio transmitters to prevent such interference. Two types of key-click filters are shown in figure 1-26.

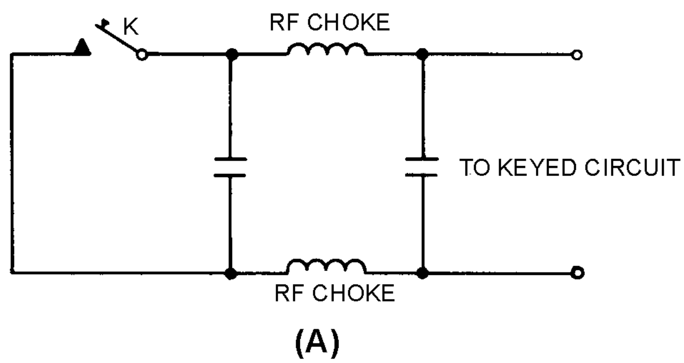


Figure 1-26A.—Key-click filters.

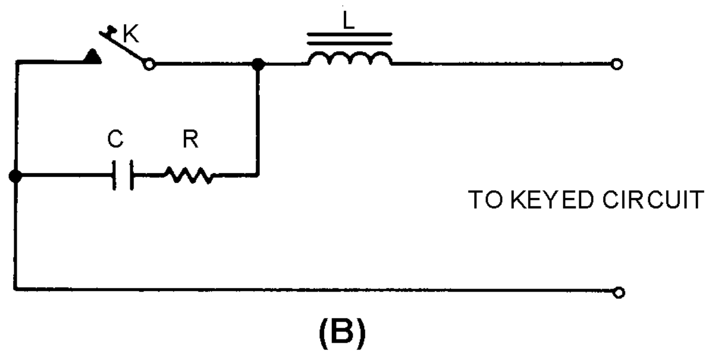


Figure 1-26B.—Key-click filters.

The capacitors and rf chokes in figure 1-26, views (A) and (B), prevent surges of current. In view (B), the choke coil causes a lag in the current when the key is closed, and the current builds up gradually instead of instantly. The capacitor charges as the key is opened and slowly releases the energy stored in the magnetic field of the inductor. The resistor controls the rate of charge of the capacitor and also prevents sparking at the key contacts by the sudden discharge of the capacitor when the key is closed.

MACHINE KEYING.—The speed with which information can be transmitted using a hand key depends on the keying ability of the operator. Early communicators turned to mechanized methods of keying the transmitters to speed transmissions. More information could be passed in a given time by replacing the hand-operated key with a keying device capable of reading information from punched tape. Using this method, several operators could prepare tapes at their normal operating speed. The tapes could then be read through the keying device at a higher rate of speed and more information could be transmitted in a given amount of time.

Spectrum Analysis

Continuous-wave transmission has the disadvantage of being a relatively slow transmission method. Still, it has several advantages. Some of the advantages of cw transmission are a high degree of clarity under severe noise conditions, long-range operation, and narrow bandwidth. A highly skilled operator can pick out and read a cw signal even though it has a high degree of background noise or interference. Since only a single-carrier frequency is being transmitted, all of the transmitter power can be concentrated in the intelligence. This concentration of power gives the transmission a greater range. The use of spectrum analysis (figure 1-27) illustrates the transmitted frequency characteristics of a cw signal. Because the cw signal is a pure sine wave, it occupies only a single frequency in the rf spectrum and the system is relatively simple.

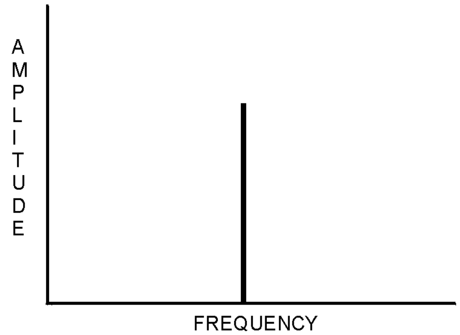


Figure 1-27.—Carrier-wave signal spectrum analysis.

Q-18. What is amplitude modulation?

Q-19. What are the three requirements for cw transmission?

Q-20. Name two methods of oscillator keying.

Q-21. State the method used to increase the speed of keying in a cw transmitter.

Q-22. Name three advantages of cw transmission.

Single-Stage Transmitters

A simple, single-tube cw transmitter can be made by coupling the output of an oscillator directly to an antenna (figure 1-28). The primary purpose of the oscillator is to develop an rf voltage which has a constant frequency and is immune to outside factors which may cause its frequency to shift. The output of this simple transmitter is controlled by placing a telegraph key at point **K** in series with the voltage supply. Since the plate supply is interrupted when the key is open, the circuit oscillates only as long as the key is closed. Although the transmitter shown uses a Colpitts oscillator, any of the oscillators previously described in *NEETS*, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits* can be used.

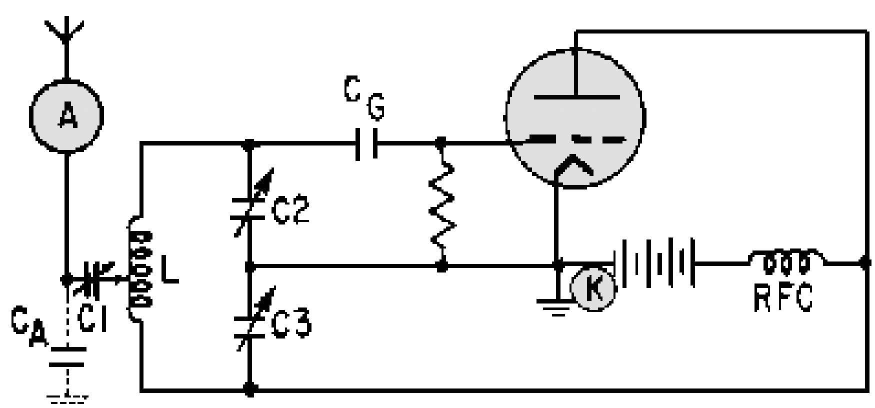


Figure 1-28.—Simple electron-tube transmitter.

Capacitors C2 and C3 can be GANGED (mechanically linked together) to simplify tuning. Capacitor C1 is used to tune (resonate) the antenna to the transmitter frequency. C_A is the effective capacitance existing between the antenna and ground. This antenna-to-ground capacitance is in parallel with the tuning capacitors, C2 and C3. Since the antenna has capacitance, any change in its length or position, such as that caused by swaying of the antenna, changes the value of C_A and causes the oscillator to change frequency. Because these frequency changes are undesirable for reliable communications, the multistage transmitter was developed to increase reliability.

Multistage Transmitters

The simple, single-tube transmitter, shown in figure 1-28, is rarely used in practical equipment. Most of the transmitters you will see use a number of tubes or stages. The number used depends on the frequency, power, and application of the equipment. For your study, the following three categories of cw transmitters are discussed: (1) master oscillator power amplifier (mopa) transmitters, (2) multistage, high-power transmitters, (3) high- and very-high frequency transmitters.

The mopa is both an oscillator and a power amplifier. Power-amplifying stages and frequency-multiplying stages must be used to increase power and raise the frequency from those achievable in a mopa. The main difference between many low- and high-power transmitters is in the number of power-amplifying stages that are used. Similarly, the main difference between many high- and very-high frequency transmitters is in the number of frequency-multiplying stages used.

MASTER OSCILLATOR POWER AMPLIFIER.—For a transmitter to be stable, its oscillator must not be LOADED DOWN. This means that its antenna (which can present a varying impedance) must not be connected directly to the oscillator circuit. The rf oscillations must be sent through another circuit before they are fed to the antenna for good frequency stability to be obtained. That additional circuit is an rf power amplifier. Its purpose is to raise the amplitude of rf oscillations to the required output power level and isolate the oscillator from the antenna. Any transmitter consisting of an oscillator and a single-amplifier stage is called a master oscillator power amplifier transmitter (mopa), as shown in figure 1-29.

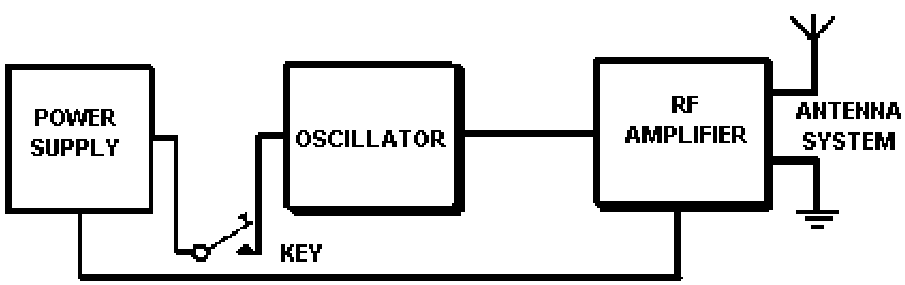


Figure 1-29.—Block diagram of a master oscillator power amplifier transmitter (mopa).

Most mopa transmitters have only one tube amplifier in the power-amplifier stage. However, the oscillator may not produce sufficient power to drive a power-amplifier tube to the power output level required for the antenna. In such cases, the power-amplifier stages are designed to use two or more amplifiers which can be driven by the oscillator. Two or more amplifiers can be connected in parallel (with similar elements of each tube connected) or in a push-pull arrangement. In a push-pull amplifier, the grids are fed equal rf voltages that are 180 degrees out of phase.

The main advantage of a mopa transmitter is that the power-amplifier stage isolates the oscillator from the antenna. This prevents changes in antenna-to-ground capacitance from affecting the oscillator frequency. A second advantage is that the rf power amplifier is operated so that a small change in the voltage applied to its grid circuit will produce a large change in the power developed in its plate circuit.

Rf power amplifiers require that a specific amount of power be fed into the grid circuit. Only in this way can the tube deliver an amplified power output. However, the stable oscillator can produce only limited amounts of power. Therefore, the mopa transmitter is limited in the amount of power it can develop. This is one of the disadvantages of the mopa transmitter. Another disadvantage is that it often is impractical for use at very- and ultra-high frequencies. The reason is that the stability of self-excited oscillators decreases rapidly as the operating frequency increases. Circuit tuning capacitances are small at high frequencies and stray capacitances adversely affect frequency stability.

MULTISTAGE HIGH-POWER TRANSMITTERS.—The power amplifier of a high-power transmitter may require far more driving power than can be supplied by an oscillator. Therefore, one or more low-power intermediate amplifiers may be inserted between the oscillator and the final power amplifier to boost power to the antenna. In some types of equipment, a VOLTAGE AMPLIFIER, called a BUFFER is used between the oscillator and the first intermediate amplifier. The ideal buffer is operated class A and is biased negatively to prevent grid current flow during the excitation cycle. Therefore, it does not require driving power from, nor does it load down, the oscillator. The purpose of the buffer is to isolate the oscillator from the following stages and to minimize changes in oscillator frequency that occur with changes in loading. A buffer is required when keying takes place in an intermediate or final amplifier operating at comparatively high power. Look at the block diagrams of several medium-frequency transmitters in figure 1-30. The input and output powers are given for each stage. You should be able to see that the power output rating of a transmitter can be increased by adding amplifier tubes capable of delivering the power required.

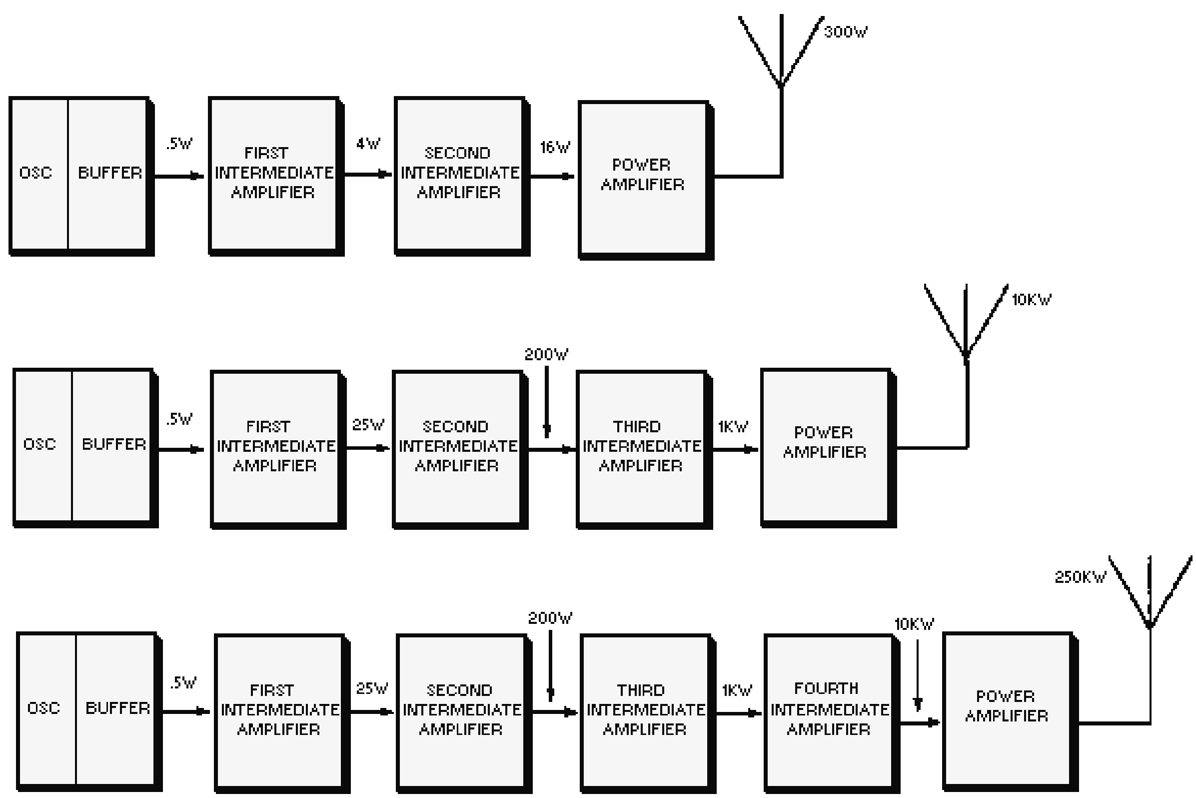


Figure 1-30.—Block diagram of several medium-frequency transmitters.

HF AND VHF TRANSMITTERS.—Oscillators are too unstable for direct frequency control in very- and ultra-high frequency transmitters. Therefore, these transmitters have oscillators operating at comparatively low frequencies, sometimes as low as 1/100 of the output frequency. The oscillator frequency is raised to the required output frequency by passing it through one or more FREQUENCY MULTIPLIERS. Frequency multipliers are special rf power amplifiers which multiply the input frequency. In practice, the MULTIPLICATION FACTOR (number of times the input frequency is multiplied) is seldom larger than five in any one stage. The block diagram of a typical VHF transmitter, designed for continuous tuning between 256 and 288 megahertz, is shown in figure 1-31.

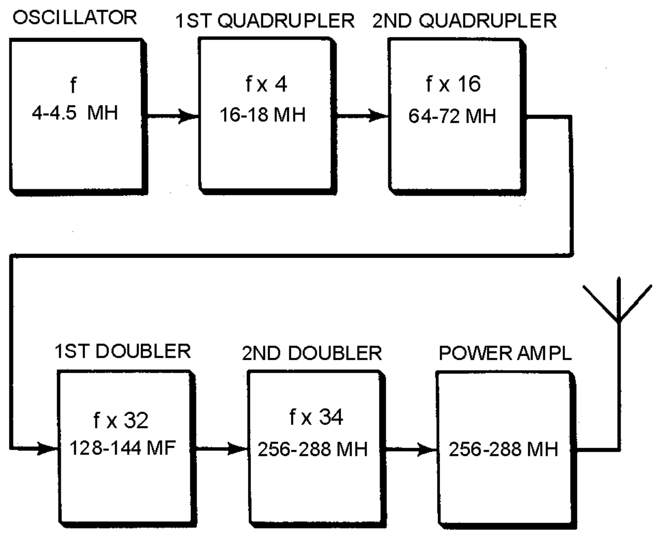


Figure 1-31.—Block diagram of a vhf transmitter.

The stages which multiply the frequency by two are DOUBLERS; those which multiply by four are QUADRUPLERS. The oscillator is tunable from 4 to 4.5 megahertz. The multiplier stages increase the frequency by multiplying successively by 4, 4, 2, and 2, for a total factor of 64. In high-power, high-frequency transmitters, one or more intermediate amplifiers may be used between the last frequency multiplier and the power amplifier.

Q-23. Name a disadvantage of a single-stage cw transmitter.

Q-24. What is the purpose of the power-amplifier stage in a master oscillator power amplifier cw transmitter?

Q-25. What is the purpose of frequency-multiplier stages in a VHF transmitter?

AMPLITUDE MODULATION

The telegraph and radiotelegraph improved man's ability to communicate by allowing speedy passage of information between two distant points. However, it failed to satisfy one of man's other communications needs; that is, the ability to hear and be heard, by voice, at a great distance. In an effort to improve on the telegraph, Alexander Graham Bell developed the principles on which modern communications are built. He developed the modulation of an electric current by complex waveforms, the demodulation of the resulting wave, and recovery of the original waveform. This section will examine the process of varying an electric current in amplitude at an audio frequency.

Microphones

If an rf carrier is to convey intelligence, some feature of the carrier must be varied in accordance with the information to be transmitted. In the case of speech intelligence, sound waves must be converted to electrical energy.

A MICROPHONE is an energy converter that changes sound energy into electrical energy. A diaphragm in the microphone moves in and out in accordance with the compression and rarefaction of the atmosphere caused by sound waves. The diaphragm is connected to a device that causes current flow in proportion to the instantaneous pressure delivered to it. Many devices can perform this function. The particular device used in a given application depends on the characteristics desired, such as sensitivity, frequency response, impedance matching, power requirements, and ruggedness.

The SENSITIVITY or EFFICIENCY of a microphone is usually expressed in terms of the electrical power level which the microphone delivers to a matched-impedance load compared to the sound level being converted. The sensitivity is rated in dB and must be as high as possible. A high microphone output requires less gain in the amplifiers used with the microphone. This keeps the effects of thermal noise, amplifier hum, and noise pickup at a minimum.

For good quality sound reproduction, the electrical signal from the microphone must correspond in frequency content to the original sound waves. The microphone response should be uniform, or flat, within its frequency range and free from the electrical or mechanical generation of new frequencies.

The impedance of a microphone is important in that it must be matched to the microphone cable and to the amplifier input as well as to the amplifier input load. Exact impedance matching is not always possible, especially in the case where the impedance of the microphone increases with an increase in frequency. A long microphone cable tends to seriously attenuate the high frequencies if the microphone impedance is high. This attenuation is caused by the increased capacitive action of the line at higher frequencies. If the microphone has a low impedance, a lower voltage is developed in the microphone, and more voltage is available at the load. Because many microphone lines used aboard ship are long, low-impedance microphones must be used to preserve a sufficiently high voltage level- over the required frequency range.

The symbol used to represent a microphone in a schematic diagram is shown in figure 1-32. The schematic symbol identifies neither the type of microphone used nor its characteristics.

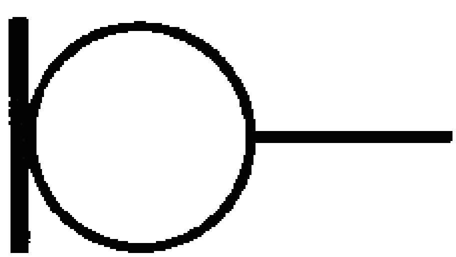


Figure 1-32.—Microphone schematic symbol.

CARBON MICROPHONE.—Operation of the SINGLE-BUTTON CARBON MICROPHONE figure 1-33, view (A) is based on varying the resistance of a pile of carbon granules located within the microphone. An insulated cup, referred to as the button, holds the loosely piled granules. It is so mounted that it is in constant contact with the thin metal diaphragm. Sound waves striking the diaphragm vary the pressure on the button which varies the pressure on the pile of carbon granules. The dc resistance of the carbon granule pile is varied by this pressure. This varying resistance is in series with a battery and the primary of a transformer. The changing resistance of the carbon pile produces a corresponding change in the current of the circuit. The varying current in the transformer primary produces an alternating voltage in the secondary. The transformer steps up the voltage and matches the low impedance of the microphone to the high impedance of the first amplifier. The voltage across the secondary may be as high as 25 volts peak. The impedance of this type of microphone varies from 50 to 200 ohms. This effect is caused by the pressure of compression and rarefaction of sound waves, discussed in chapter 1 of *NEETS*, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas.*

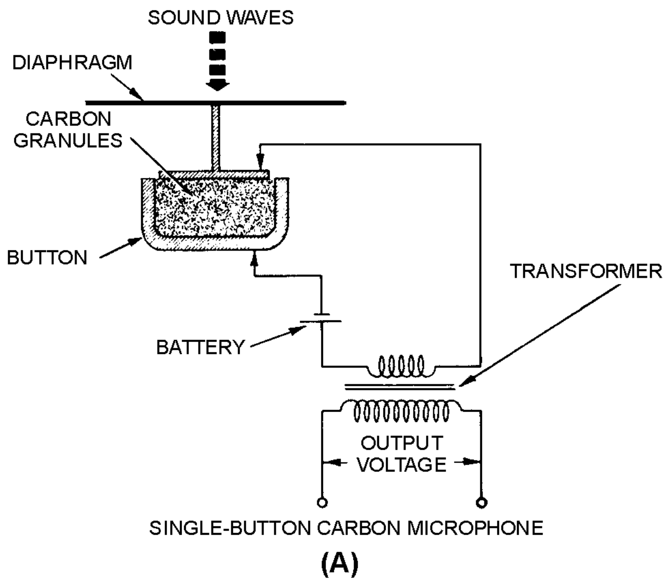


Figure 1-33A.—Carbon microphones. SINGLE-BUTTON CARBON MICROPHONE.

The DOUBLE-BUTTON CARBON MICROPHONE is shown in figure 1-33, view (B). Here, one button is positioned on each side of the diaphragm so that an increase in resistance on one side is accompanied by a simultaneous decrease in resistance on the other. Each button is in series with the battery and one-half of the transformer primary. The decreasing current in one-half of the primary and the increasing current in the other half produces an output voltage in the secondary winding. The output voltage is proportional to the sum of the primary winding signal components. This action is similar to that of push-pull amplifiers.

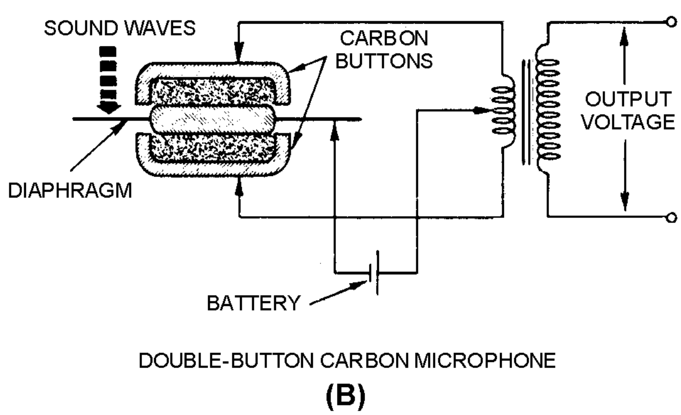


Figure 1-33B.—Carbon microphones. DOUBLE-BUTTON CARBON MICROPHONE.

One disadvantage of carbon microphones is that of a constant BACKGROUND HISS (hissing noise) which results from random changes in the resistance between individual carbon granules. Other disadvantages are reduced sensitivity and distortion that may result from the granules packing or sticking together. The carbon microphone also has a limited frequency response. Still another disadvantage is a requirement for an external voltage source.

The disadvantages, however, are offset by advantages that make its use in military applications widespread. It is lightweight, rugged, and can produce an extremely high output.

CRYSTAL MICROPHONE.—The CRYSTAL MICROPHONE uses the PIEZOELECTRIC EFFECT of Rochelle salt, quartz, or other crystalline materials. This means that when mechanical stress is placed upon the material, a voltage electromagnetic force (EMF) is generated. Since Rochelle salt has the largest voltage output for a given mechanical stress, it is the most commonly used crystal in microphones. View (A) of figure 1-34 is a crystal microphone in which the crystal is mounted so that the sound waves strike it directly. View (B) has a diaphragm that is mechanically linked to the crystal so that the sound waves are indirectly coupled to the crystal.

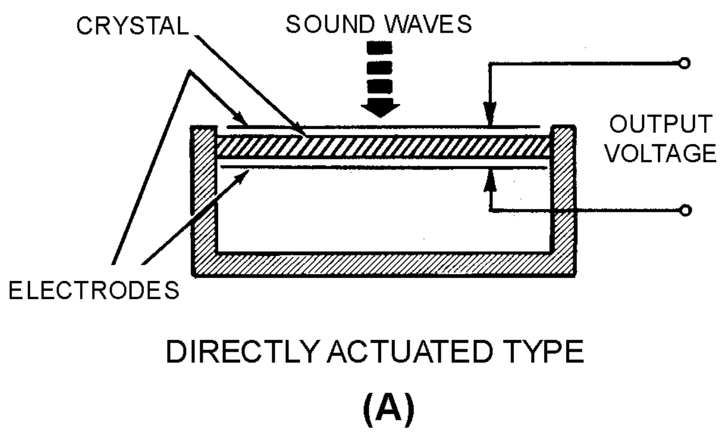


Figure 1-34A.—Crystal microphones.

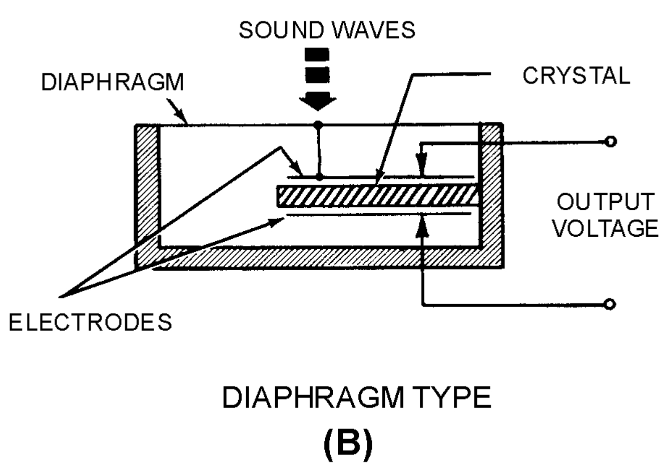


Figure 1-34B.—Crystal microphones.

A crystal microphone has a high impedance and does not require an external voltage source. It can be connected directly into the input circuit of a high-gain amplifier. However, because its output is low, several stages of high-gain amplification are required. Crystal microphones are delicate and must be handled with care. Exposure to temperatures above 52 degrees Celsius (125 degrees Fahrenheit) may permanently damage the crystal unit. Crystals are also soluble in water and other liquids and must be protected from moisture and excessive humidity.

DYNAMIC MICROPHONE.—A cross section of the DYNAMIC or MOVING-COIL MICROPHONE is shown in figure 1-35. A coil of fine wire is mounted on the back of the diaphragm and located in the magnetic field of a permanent magnet. When sound waves strike the diaphragm, the coil moves back and forth cutting the magnetic lines of force. This induces a voltage into the coil that is an electrical reproduction of the sound waves.

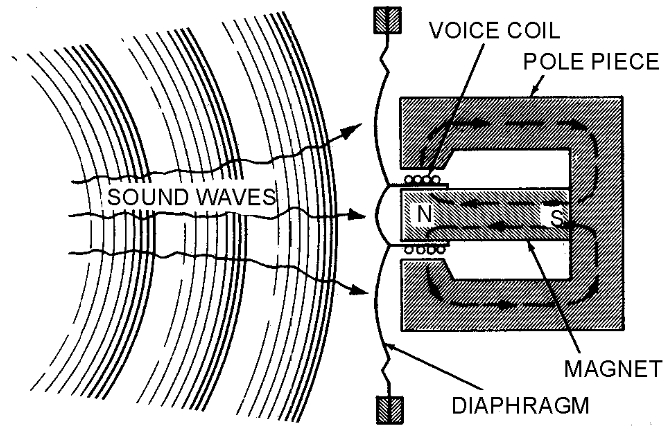


Figure 1-35.—Dynamic microphone.

The sensitivity of the dynamic microphone is almost as high as that of the carbon type. It is lightweight and requires no external voltage. The dynamic microphone is rugged and can withstand the effects of vibration, temperature, and moisture. This microphone has a uniform response over a frequency range that extends from 40 to 15,000 hertz. The impedance is very low (generally 50 ohms or less). A transformer is required to match its impedance to that of the input of an af amplifier.

MAGNETIC MICROPHONE.—The MAGNETIC or MOVING-ARMATURE MICROPHONE (figure 1-36) consists of a coil wound on an armature that is mechanically connected to the diaphragm with a driver rod. The coil is located between the pole pieces of the permanent magnet. Any vibration of the diaphragm vibrates the armature at the same rate. This varies the magnetic flux in the armature and through the coil.

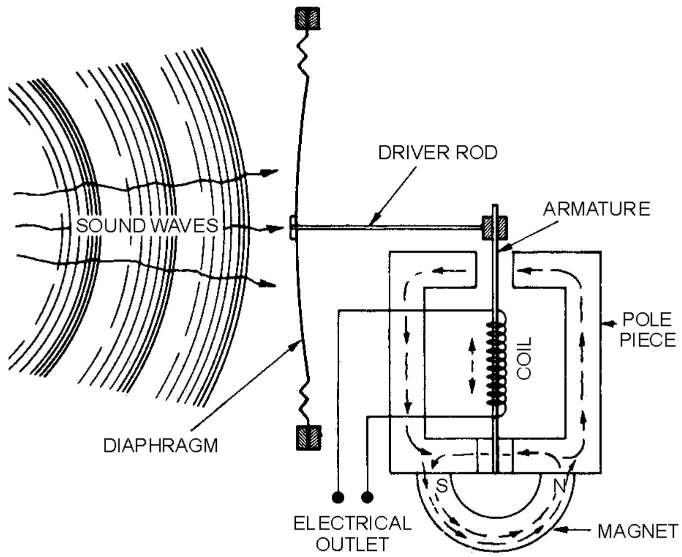


Figure 1-36.—Magnetic microphone action.

When the armature is in its resting position (midway between the two poles), the magnetic flux is established across the air gap. However, no resultant flux is established in the armature. When a compression wave strikes the diaphragm, the armature is deflected to the right. Most of the flux continues to move in the direction of the arrows. However, some flux now flows from the north pole of the magnet across the reduced gap at the upper right, down through the armature, and around to the south pole of the magnet.

When a rarefaction wave occurs at the diaphragm, the armature is deflected to the left. Some flux is now directed from the north pole of the magnet, up through the armature, through the reduced gap at the upper left, and back to the south pole.

The vibrations of the diaphragm cause an alternating flux in the armature which induces an alternating voltage in the coil. This voltage has the same waveform as that of the sound waves striking the diaphragm.

The magnetic microphone is very similar to the dynamic microphone in terms of impedance, sensitivity, and frequency response. However, it is more resistant to vibration, shock, and rough handling than other types of microphones.

Changing sound waves into electrical impulses is the first step in voice communications. It is common to all the transmission media you will study in the remainder of this chapter. We will discuss the various types of modulation that arc used to transfer this information to a transmission medium in the following sections.

Q-26. What is a microphone?

Q-27. What special electromechanical effect is the basis for carbon microphone operation?

Q-28. What is a major disadvantage of a carbon microphone?

Q-29. What property of a crystalline material is used in a crystal microphone?

Q-30. What is the difference between a dynamic microphone and a magnetic microphone?

AM TRANSMITTER PRINCIPLES

In this section we will describe the methods used to apply voice signals (intelligence) to a carrier wave by the process of amplitude modulation (AM).

An AM transmitter can be divided into two major sections according to the frequencies at which they operate, radio-frequency (rf) and audio-frequency (af) units. The rf unit is the section of the transmitter used to generate the rf carrier wave. As illustrated in figure 1-37, the carrier originates in the master oscillator stage where it is generated as a constant-amplitude, constant-frequency sine wave. The carrier is not of sufficient amplitude and must be amplified in one or more stages before it attains the high power required by the antenna. With the exception of the last stage, the amplifiers between the oscillator and the antenna are called INTERMEDIATE POWER AMPLIFIERS (ipa). The final stage, which connects to the antenna, is called the FINAL POWER AMPLIFIER (fpa).

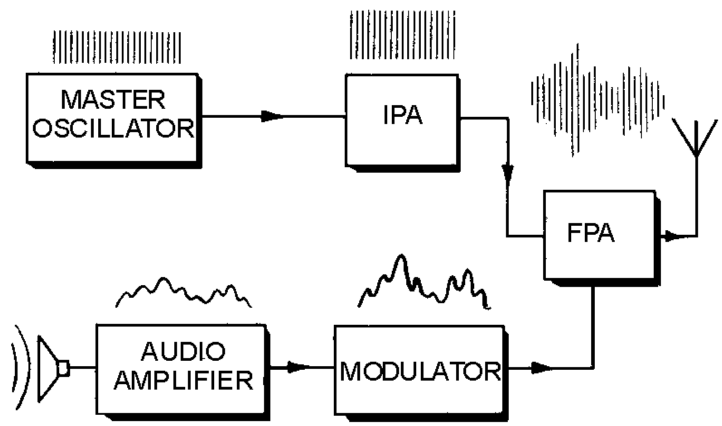


Figure 1-37.—Block diagram of an AM transmitter.

The second section of the transmitter contains the audio circuitry. This section of the transmitter takes the small signal from the microphone and increases its amplitude to the amount necessary to fully modulate the carrier. The last audio stage is the MODULATOR. It applies its signal to the carrier in the final power amplifier. In this way, intelligence is included in the radiated rf waveform.

The Modulated Wave

The frequencies present in a signal can be conveniently represented by a graph of the frequency spectrum, shown in figure 1-38. In this graph, each individual frequency is portrayed as a vertical line. The position of the line along the horizontal axis indicates the frequency of the signal. The height of the frequency line is proportional to the amplitude of the signal. The rf spectrum in figure 1-38 shows the frequencies present when heterodyning occurs between frequencies of 5 and 100 kilohertz.

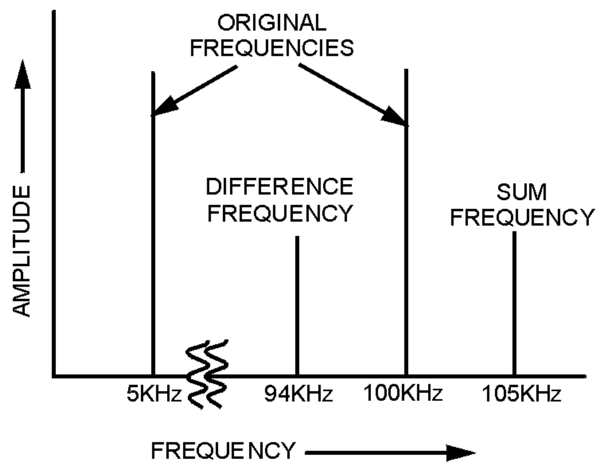


Figure 1-38.—Radio-frequency spectrum.

Radiating energy at audio frequencies (discussed earlier in this chapter) is not practical. The heterodyning principle, however, makes possible the conversion of an af signal (intelligence) to an rf signal (with af intelligence) which can be radiated or transmitted through space.

Look again at figure 1-38. The sum and difference frequencies are located very near the rf signal (100 kilohertz), while the audio signal (5 kilohertz) is spaced a considerable distance away. Because of this frequency separation, the audio frequency can be easily removed by filter circuits, leaving just three radio frequencies of 95, 100, and 105 kilohertz. These three radio frequencies are radiated through space to the receiving station. At the receiver, the process is reversed. The frequency of 95 kilohertz, for example, is heterodyned with the frequency of 100 kilohertz and the sum and difference frequencies are again produced. (A similar process occurs between the frequencies of 100 and 105 kilohertz.) Of the resultant frequencies (95, 100, 105, and 5 kilohertz), all are filtered out except the 5 kilohertz difference frequency. This frequency, which is identical to the original 5 kilohertz audio applied at the transmitter, is retained and amplified. Thus, the 5 kilohertz audio tone *appears* to have been radiated through space from the transmitter to the receiver.

In the process just described, the 100 kilohertz frequency is referred to as the CARRIER FREQUENCY, and the sum and difference frequencies are referred to as SIDE FREQUENCIES. Since the sum frequency appears above the carrier frequency, it is referred to as the UPPER SIDE FREQUENCY. The difference frequency appears below the carrier and is referred to as the LOWER SIDE FREQUENCY.

When a carrier is modulated by voice or music signals, a large number of sum and difference frequencies are produced. All of the sum frequencies above the carrier are spoken of collectively as the UPPER SIDEBAND. All the difference frequencies below the carrier, also considered as a group, are called the LOWER SIDEBAND.

If the carrier and the modulating signal are constant in amplitude, the sum and difference frequencies will also be constant in amplitude. However, when the carrier and sidebands are combined in a single impedance and viewed simultaneously with an oscilloscope, the resultant waveform appears as shown in figure 1-39. This resultant wave is called the MODULATION ENVELOPE. The modulation envelope has the same frequency as the carrier. However, it rises and falls in amplitude with the continual phase shift between the carrier and sidebands. This causes these signals to first aid and then oppose one another. These cyclic variations in the amplitude of the envelope have the same frequency as the audio-modulating voltage. The audio intelligence is actually contained in the spacing or difference between the carrier and sideband frequencies.

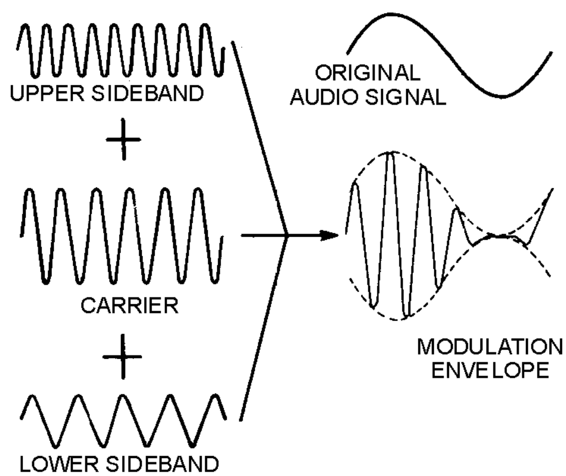


Figure 1-39.—Formation of the modulation envelope.

BANDWIDTH OF AN AM WAVE.—An ideal carrier wave contains a single frequency and occupies very little of the frequency spectrum. When the carrier is amplitude modulated, sideband frequencies are created both above and below the carrier frequency. This causes the signal to use up a greater portion of the frequency spectrum. The amount of space in the frequency spectrum required by the signal is called the BANDWIDTH of the signal.

The bandwidth of a modulated wave is a function of the frequencies contained in the modulating signal. For example, when a 100-kilohertz carrier is modulated by a 5-kilohertz audio tone, sideband frequencies are created at 95 and 105 kilohertz. This signal requires 10 kilohertz of space in the spectrum.

If the same 100-kilohertz carrier is modulated by a 10-kilohertz audio tone, sideband frequencies will appear at 90 and 110 kilohertz and the signal will have a bandwidth of 20 kilohertz. Notice that as the modulating signal becomes higher in frequency, the bandwidth required also becomes greater. As illustrated by the above examples, the bandwidth of an amplitude-modulated wave at any instant is two times the highest modulating frequency applied at that time. Thus, if a 400-kilohertz carrier is modulated with 3, 5, and 8 kilohertz simultaneously, sideband frequencies will appear at 392, 395, 397, 403, 405, and 408 kilohertz. This signal extends from 392 to 408 kilohertz and has a bandwidth of 16 kilohertz, twice the highest modulating frequency of 8 kilohertz.

Musical instruments produce complex sound waves containing a great number of frequencies. The frequencies produced by a piano, for example, range from approximately 27 to 4,200 hertz with harmonic frequencies extending beyond 10 kilohertz. Modulating frequencies of up to 15 kilohertz must be included in the signal to transmit a musical passage with a high degree of fidelity. This requires a bandwidth of at least 30 kilohertz to prevent attenuation of higher-order harmonic frequencies.

If the signal to be transmitted contains voice frequencies only, and fidelity is of minor importance, the bandwidth requirement is much smaller. A baritone voice includes frequencies of approximately 100 to 350 hertz, or 250 hertz. Intelligible voice communications can be carried out as long as the communications system retains audio frequencies up to several thousand hertz. Comparing the conditions for transmitting voice signals with those for transmitting music reveals that much less spectrum space is required for voice communications.

Radio stations in the standard broadcast band are assigned carrier frequencies by the Federal Communications Commission (FCC). When two stations are located near each other, their carriers must be spaced some minimum distance apart in the radio spectrum. Otherwise, the sideband frequencies of one station will interfere with sideband frequencies of the other station. The standard AM broadcast band starts at 535 kilohertz and ends at 1,605 kilohertz. Carrier assignments start at 540 kilohertz and continue in a succession of 10-kilohertz increments until the upper limit of the broadcast band is reached. This adds up to a total of 107 carrier assignments, or CHANNELS, over the entire broadcast band. If stations were assigned to all 107 channels (in a given geographical area), each station would be allotted a channel width of 10 kilohertz. This leaves 5 kilohertz on each side of each carrier for sidebands. Since interference between such closely spaced stations would be nearly impossible to prevent, the FCC avoids assigning adjacent channels to stations in the same area. As a consequence of this policy, one or more vacant channels normally exist between stations in the broadcast band. In the interest of better fidelity, the stations are permitted to use modulating frequencies higher than 5 kilohertz as long as no interference with other stations is produced.

Q-31. What are the two major sections of a typical AM transmitter?

Q-32. When 100 kilohertz and 5 kilohertz are heterodyned, what frequencies are present?

Q-33. What is the upper sideband of an AM transmission?

Q-34. Where is the intelligence in an AM transmission located?

Q-35. What determines the bandwidth of an AM transmission?

ANALYSIS OF AN AM WAVE.—A significant amount of information concerning the basic principles of amplitude modulation can be obtained from a study of the properties of the modulation envelope.

A carrier wave which has been modulated by voice or music signals is accompanied by two sidebands; each sideband contains individual frequencies that vary continuously. Since a wave of this nature is nearly impossible to analyze, you can assume in the following sections that the modulating signal, unless otherwise qualified, is a single-frequency, constant-amplitude sine wave.

PERCENT OF MODULATION IN AN AM WAVE.—The degree of modulation is defined in terms of the maximum permissible amount of modulation. Thus, a fully modulated wave is said to be 100-PERCENT MODULATED. The modulation envelope in figure 1-40, view (A), shows the conditions for 100-percent sine-wave modulation. For this degree of modulation, the peak audio voltage must be equal to the dc supply voltage to the final power amplifier. Under these conditions, the rf output voltage will reach 0 on the negative peak of the modulating signal; on the positive peak, it will rise to twice the amplitude of the unmodulated carrier.

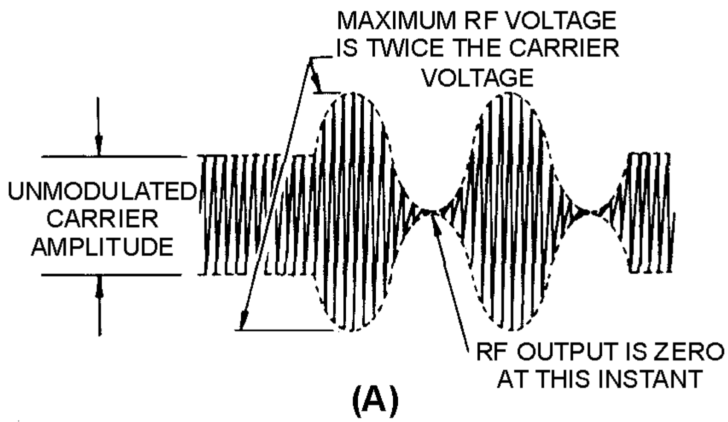


Figure 1-40A.—Conditions for 100-percent modulation.

When analyzed, the modulation envelope consists of the unmodulated rf carrier voltage plus the combined voltage of the two sidebands. The combined sideband voltages are approximately equal to the rf carrier voltage since each sideband frequency contains one-half the carrier voltage, as shown in view (B). This condition is known as 100-percent modulation and the maximum modulated rf voltage is twice the carrier voltage. The audio-modulating voltage can be increased beyond the amount required to produce 100-percent modulation. When this happens, the negative peak of the modulating signal becomes larger in amplitude than the dc plate-supply voltage to the final power amplifier. This causes the final plate voltage to be negative for a short period of time near the negative peak of the modulating signal. For the duration of the negative plate voltage, no rf energy is developed across the plate tank circuit and the rf output voltage remains at 0, as shown in figure 1-41, view (A).

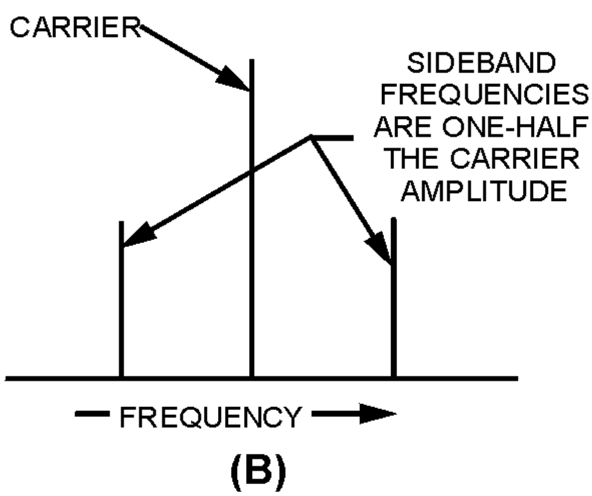


Figure 1-40B.—Conditions for 100-percent modulation.

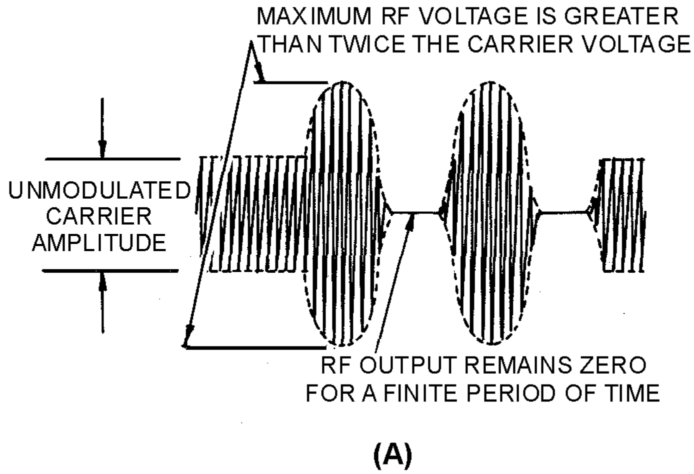


Figure 1-41A.—Overmodulation conditions.

Look carefully at the modulation envelope in view (A). It shows that the negative peak of the modulating signal has effectively been limited. If the signal were demodulated (detected in the receiver), it would have an appearance somewhat similar to a square wave. This condition, known as OVERMODULATION, causes the signal to sound severely distorted (although this will depend on the degree of overmodulation).

Overmodulation will generate unwanted (SPURIOUS) sideband frequencies. This effect can easily be detected by tuning a receiver near, but somewhat outside the desired frequency. You would likely be able to tune to one or more of these undesired sideband frequencies, but the reception would be severely distorted, possibly unintelligible. (Without overmodulation, no such unwanted sideband frequencies would exist and you would be able to tune only to the desired frequency.) These unwanted frequencies will appear for a considerable range both above and below the desired channel. This effect is sometimes called SPLATTER. These spurious frequencies, shown in view (B), cause interference with other stations operating on adjacent channels. You should clearly understand that overmodulation, and its attendant distortion and interference is to be avoided.

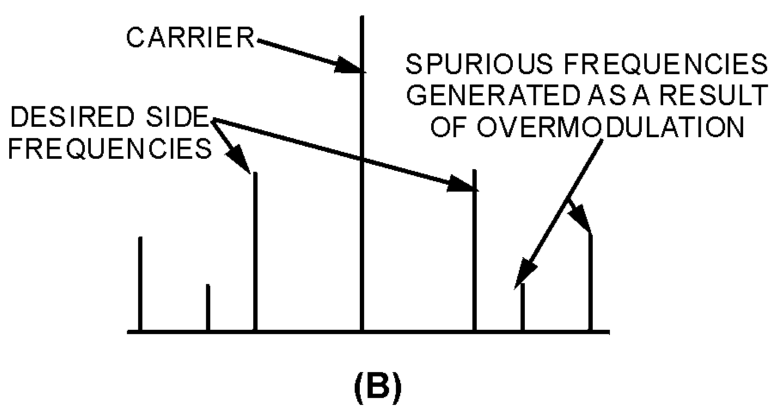


Figure 1-41B.—Overmodulation conditions.

In addition to the above problems, overmodulation also causes abnormally large voltages and currents to exist at various points within the transmitter. Therefore, sufficient overload protection by circuit breakers and fuses should be provided. When this protection is not provided, the excessive voltages can cause arcing between transformer windings and between the plates of capacitors, which will permanently destroy the dielectric material. Excessive currents can also cause overheating of tubes and other components.

Ideally, you will want to operate a transmitter at 100-percent modulation so that you can provide the maximum amount of energy in the sideband. However, because of the large and rapid fluctuations in amplitude that these signals normally contain, this ideal condition is seldom possible. When the modulator is properly adjusted, the loudest parts of the transmission will produce 100-percent modulation. The quieter parts of the signal then produce lesser degrees of modulation.

To measure degrees of modulation less than 100 percent, you can use a MODULATION FACTOR (M) to indicate the relative magnitudes of the rf carrier and the audio-modulating signal. Numerically, the modulation factor is:

$$M = \frac{E_m}{E_c}$$

Where:

M = the modulation factor

E_m = the peak, peak-to-peak, or rms value of the modulating voltage

E_c = the peak, peak-to-peak,

To illustrate this use of the equation, assume that a carrier wave with a peak amplitude of 400 volts is modulated by a 3-kilohertz sine wave with a peak amplitude of 200 volts. The modulation factor is figured as follows:

$$M = \frac{E_m}{E_c}$$

$$M = \frac{200}{400}$$

$$M = 0.5$$

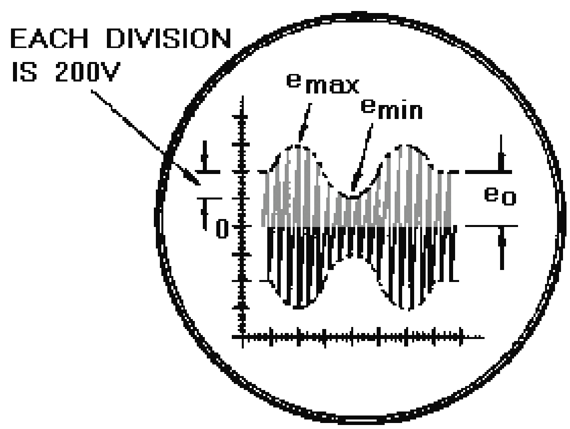
If the modulation factor were multiplied by 100, the resultant quantity would be the PERCENT OF MODULATION (%M):

$$\%M = \frac{E_m}{E_c} \times 100$$

$$\%M = \frac{200}{400} \times 100$$

$$\%M = 50 \text{ percent}$$

By using the correct equation, you can determine the percent of modulation from the modulation envelope pattern. This method is useful when the percent of modulation is to be determined using the pattern on the screen of an oscilloscope. For example, assume that your oscilloscope is connected to the output of a modulator circuit and produces the screen pattern shown in figure 1-42. According to the setting of the calibration control, each large division on the vertical scale is equal to 200 volts. By using this scale, you can see that the peak carrier amplitude (unmodulated portion) is 400 volts. The peak amplitude of the carrier is designated as e_o in figure 1-42.



$$\%M = \frac{e_{max} - e_{min}}{2e_o} \times 100$$

-OR-

$$\%M = \frac{e_{max} - e_{min}}{e_{max} + e_{min}} \times 100$$

Figure 1-42.—Computing percent of modulation from the modulation envelope.

The amplitude of the audio-modulating voltage can also be determined from amplitude variations in the envelope pattern. Notice that the peak-to-peak variations in envelope amplitude ($e_{max} - e_{min}$) is equal to 400 volts on the scale. Note then that the peak amplitude of the audio voltage is 200 volts. If these rf and audio voltage values are inserted into the equation, the pattern in figure 1-42 is found to represent 50-percent modulation.

If E_m and E_c in the equation are assumed to represent peak-to-peak values, the following formula results:

$$\%M = \frac{E_m}{E_c} \times 100$$

Since the peak-to-peak value of E_m in figure 1-42 is $e_{max} - e_{min}$, we can substitute as follows:

$$\%M = \frac{e_{max} - e_{min}}{E_c} \times 100$$

Also, since the peak-to-peak value of the carrier E_c is 2 times e_o, we can subsititute $2e_o$ for E_c as follows:

$$\%M = \frac{e_{max} - e_{min}}{2e_o} \times 100$$

Linear vertical distance represents voltage on the screen of a cathode-ray tube. Vertical distance units can be used in place of voltage in equations. Thus, if only the percent of modulation is required, the oscilloscope need not be calibrated and the actual circuit voltages are not required. In figure 1-42, e_{max} represents 600 volts (3 large divisions); e_{min} is 200 volts (1 division); and e_o is 400 volts (2 divisions). Using the equation and the dimensions of the screen pattern, you can figure the percent of modulation as follows:

$$\%M = \frac{e_{max} - e_{min}}{2e_o} \times 100$$

$$\%M = \frac{3 - 1}{2 \times 2} \times 100$$

$$\%M = \frac{2}{4} \times 100$$

$$\%M = 50 \text{ percent}$$

When e_o of the equation is difficult to measure, an alternative solution can be obtained with the equation below:

$$\%M = \frac{e_{max} - e_{min}}{e_{max} + e_{min}} \times 100$$

VECTOR ANALYSIS OF AN AM WAVE.—You studied earlier in this chapter that the modulation envelope results when the instantaneous sums of the carrier and sideband voltages are plotted with respect to time. An attempt to add these three voltages, point-by-point, would prove to be a huge task. The same end result can be obtained by using a rotating vector to represent each of the three frequencies in the composite envelope. In the following analysis, vectors will be scaled to indicate the peak voltage value of the frequencies they represent.

The analysis has been simplified further by using a frequency of 8 hertz to represent the carrier frequency. Each cycle of the carrier then requires 1/8 of a second to complete 360 degrees. The carrier will be 100-percent modulated by a sine wave having a frequency of 1 hertz, thereby producing sideband frequencies of 7 and 9 hertz.

Envelope Development from Vectors.—The modulating signal, upper sideband, carrier, and lower sideband waveforms are illustrated in views (A) through (D), respectively, in figure 1-43. Notice that the vertical lines passing through the figure divide each waveform into segments of 1/8 of a second each. These lines also coincide with the starting and ending points of each cycle of the carrier wave.

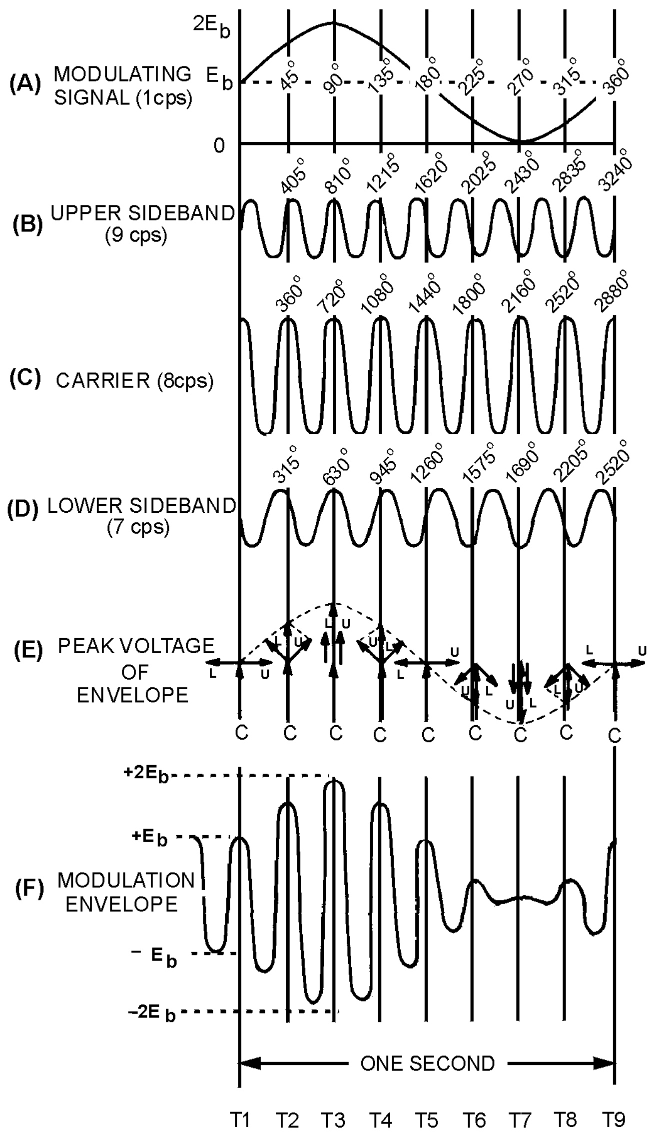


Figure 1-43.—Formation of the modulation envelope by the addition of vectors representing the carrier and sidebands.

During the first 1/8 of a second (T1 to T2), the carrier wave completes exactly 1 cycle, or 360 degrees, as shown in view (C). The upper sideband, which has a frequency of 9 hertz, will complete each cycle in less than 1/8 of a second. Therefore, during the time required for the carrier to complete 1 cycle of 360 degrees, the upper sideband [view (B)] is able to complete 1 cycle of 360 degrees plus an additional 45 degrees of the next cycle, for a total of 405 degrees.

The lower sideband [view (D)] has a frequency of 7 hertz and cannot complete an entire cycle in 1/8 of a second. During the time interval required for the carrier wave to progress through 360 degrees, the lower sideband frequency of 7 hertz can complete only 315 degrees, 45 degrees short of a full cycle.

Keeping these factors in mind, you should be able to see that the phase angles between the two sideband frequencies, and between each sideband frequency and the carrier frequency, will continually shift. At an instant in time (T3), the carrier and sidebands will be in phase [view (E)], causing the envelope amplitude [view (F)] to be twice the amplitude of the carrier. At another instant in time (T7), the sidebands are out of phase with the carrier [view (E)], causing complete cancellation of the rf voltage. The envelope amplitude will become 0 at this point. You should see that, although the carrier and sideband frequencies have constant amplitudes, the ever-changing phase differences between them causes the modulation envelope to vary continuously in amplitude.

The vector analysis of the modulation envelope will be developed with the aid of figure 1-44. In figure 1-44, view (A), a vertical vector (C) has been drawn to represent the carrier wave in figure 1-43. At T1 in figure 1-43, the upper and lower sideband frequencies are of opposite phase with respect to each other, and 90 degrees out of phase with respect to the carrier. This condition is illustrated in figure 1-44, view (A), by sideband vectors U and L drawn in opposite directions along the horizontal axis. Since the upper sideband U is equal in amplitude but opposite in phase to lower sideband L, the two sideband voltages cancel one another; the amplitude of the envelope at T1 is equal to the amplitude of the carrier. The same vector diagram is shown on a smaller scale in figure 1-43, view (E).

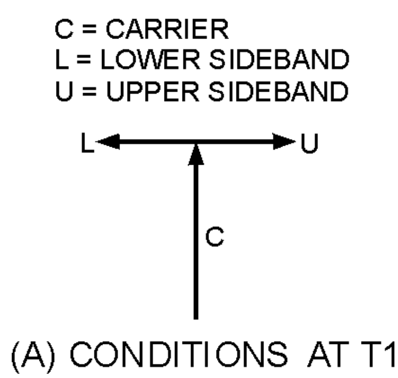


Figure 1-44A.—Vector diagrams for T1 and T2.

During the 1/8 of a second time interval between T1 and T2, all three vectors rotate in a counterclockwise direction at a velocity determined by their respective frequencies. The vector representing the carrier, for example, has made one complete rotation of 360 degrees and is back in its original position, as shown in figure 1-44, view (B). The upper sideband frequency, however, will complete 405 degrees in this same 1/8 of a second. Notice in view (B) that vector U has made one complete counterclockwise rotation of 360 degrees, plus an additional 45 degrees for a total rotation of

405 degrees. Vector L, representing the lower sideband, rotates at a velocity less than that of either the carrier or the upper sideband. In 1/8 of a second, vector L completes only 315 degrees, which is 45 degrees short of one complete rotation. At the end of 1/8 of a second, the three vectors have advanced to the positions shown in view (B).

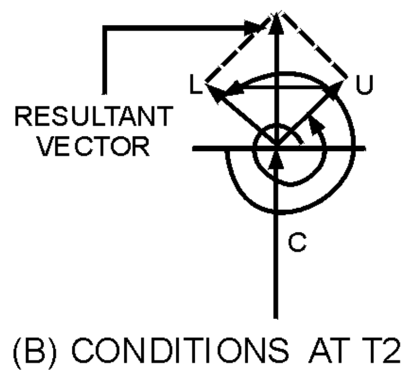


Figure 1-44B.—Vector diagrams for T1 and T2.

The resultant vector in view (B) is obtained by adding vector U to vector L. Since each sideband has one-half the amplitude of the carrier, and the two sidebands differ in phase by 90 degrees, the amplitude of the resultant vector can be computed. This computation (not shown) would show the resultant vector to have an amplitude that is approximately 70 percent that of the carrier. Thus, at T2 the amplitude of the modulation envelope is about 1.7 times the amplitude of the carrier. This condition is shown in figure 1-43, view (F).

By a similar procedure, vector diagrams can be constructed for time intervals T3 through T9. This has been done in figure 1-43, view (E). From these nine individual vector diagrams, the complete modulation envelope in figure 1-43, view (F), can be constructed.

Notice in particular the vector diagrams for T3 and T7. At T3, all three waves, and therefore all three vectors, are in phase. The modulation envelope at this instant must, therefore, be equal to twice the amplitude of the carrier since each sideband frequency has one-half the amplitude of the carrier.

At T7, the two sideband frequencies are in phase with each other but 180 degrees out of phase with the carrier. This causes the combined sideband voltage to cancel the carrier voltage, and the modulation envelope becomes 0 at that instant. Note that for the transmitter output to be 0 at T7, both the carrier and sideband frequencies must be present. If any one of these three frequencies were missing, complete cancellation would not occur and rf energy would be present in the output.

Although this vector analysis was made for frequencies of 7, 8, and 9 hertz, the same description could be applied to the frequencies actually present at the output of a transmitter.

Modulation Level of an AM Wave

As stated earlier, the modulating signal can be introduced into any active element of a tube. In addition to the various arrangements possible within a single stage, the modulating signal can also be applied to any of the rf stages in the transmitter. For example, the modulating signal could be applied to the control grid or plate of one of the intermediate power amplifiers.

A modulator circuit is usually placed into one of two categories, high- or low-level modulation. Circuits are categorized according to the level of the carrier wave at the point in the system where the modulation is applied. The FCC defines HIGH-LEVEL MODULATION in the Code of Federal Regulations as "modulation produced in the plate circuit of the last radio stage of the system." This same document defines LOW-LEVEL MODULATION as "modulation produced in an earlier stage than the final."

Q-36. What is percent of modulation?

Q-37. With a single modulating tone, what is the amplitude of the sideband frequencies at 100-percent modulation?

Q-38. What is the formula for percent of modulation?

Q-39. What is high-level modulation?

MODULATION SYSTEMS

To complete your understanding of AM modulation, we are now going to analyze the operation of a typical plate modulator. Detailed circuit descriptions will be used to give you an understanding of a basic AM plate modulator. In addition, we will cover basic circuit descriptions for cathode and grid electron-tube modulators and for base, emitter, and collector transistor modulators in this chapter.

Plate Modulator

Figure 1-45 is a basic plate-modulator circuit. Plate modulation permits the transmitter to operate with high efficiency. It is the simplest of the modulators available and is also the easiest to adjust for proper operation. The modulator is coupled to the plate circuit of the final rf amplifier through the modulation transformer. For 100-percent modulation, the modulator must supply enough power to cause the plate voltage of the final rf amplifier to vary between 0 and twice the dc operating plate voltage. The modulator tube (V2) is a power amplifier biased so that it operates class A. The final rf power amplifier (V1) is biased in the nonlinear portion of its operating range (class C). This provides for efficient operation of V1 and produces the necessary heterodyning action between the rf carrier and the af modulating frequencies.

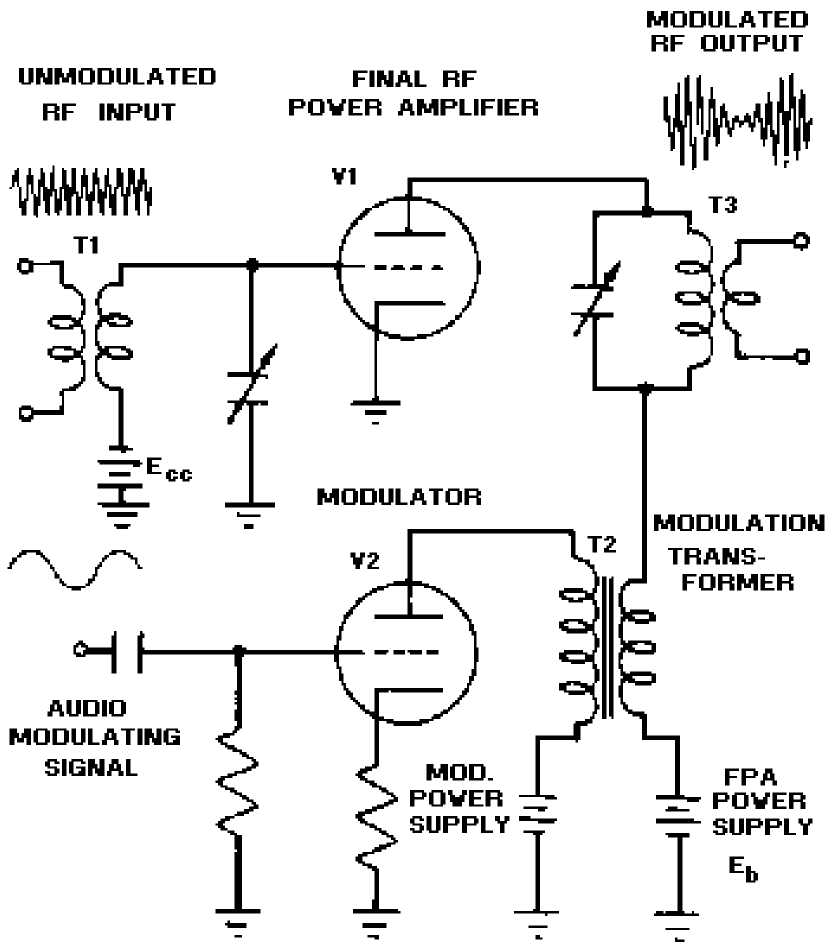


Figure 1-45.—Plate-modulation circuit.

PLATE MODULATOR CIRCUIT OPERATION.—Figure 1-46, views (A) through (E), shows the waveforms associated with the plate-modulator circuit shown in figure 1-45. Refer to these two figures throughout the following discussion.

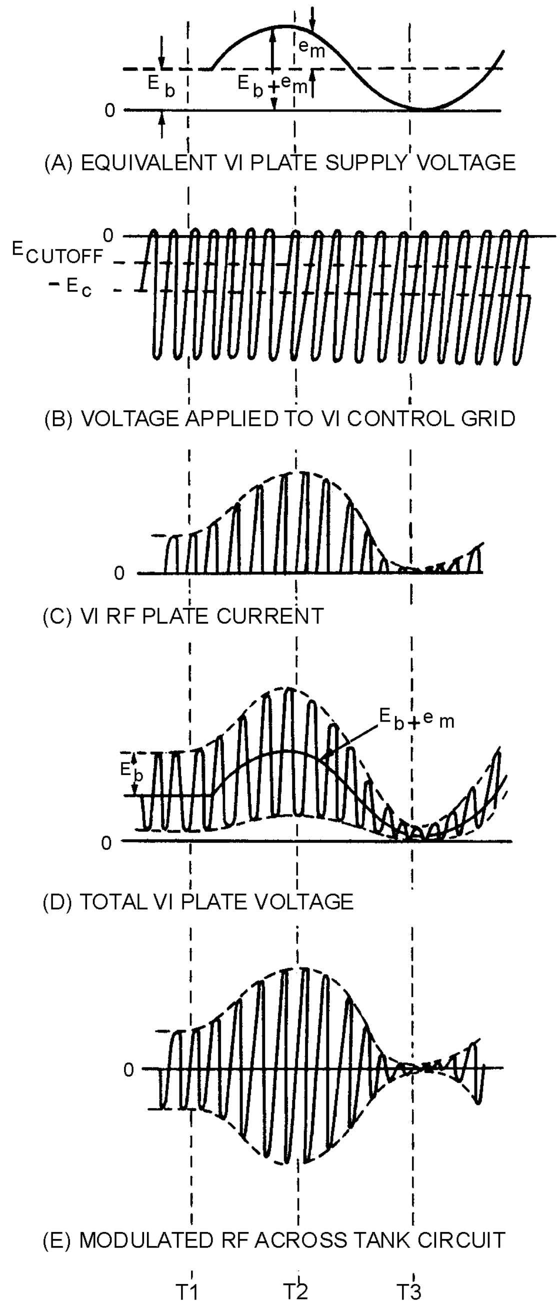


Figure 1-46.—Plate-modulator waveforms.

The rf power amplifier (V1) acts as a class C amplifier when no modulation is present in the plate circuit. V2 is the modulator which transfers the modulating voltage to the plate circuit of V1. Let's see how this circuit produces a modulated rf output.

View (A) of figure 1-46 shows the plate supply voltage for V1 as a constant dc value (E_b) at time 1 with no modulating signal applied. V1 is biased at cutoff at this time. The incoming rf carrier [view (A)] is applied to the grid of V1 by transformer T1 and causes the plate circuit current to PULSE (SURGE) each time the grid is driven positive. These rf pulses are referred to as current pulses and are shown in view (C). The plate tank output circuit (T3) is shocked into oscillation by these current pulses and the rf output waveform shown in view (E) is developed. The rf plate voltage waveform is shown in view (D).

An audio-modulating voltage applied to the grid of V2 is amplified by the modulator and coupled to the plate of V1 by modulation transformer T2. The secondary of T2 is in series with the plate-supply voltage (E) of V1. The modulating voltage will either add to or subtract from the plate voltage of V1. This is shown in view (A) at time 2 and time 3. At time 2 in view (A), the plate supply voltage for V1 increases to twice its normal value and the rf plate current pulses double, as shown in view (C). At time 3 in view (A), the supply voltage is reduced to 0 and the rf plate current decreases to 0, as shown in view (C). These changes in rf plate current cause rf tank T3 voltage to double at time 2 and to decrease to 0 at time 3, as shown in view (E). This action results in the modulation envelope shown in view (E) that represents 100-percent modulation. This is transformer-coupled out of tank circuit T3 to an antenna. Because of the oscillating action of tank circuit T3, V1 has to be rated to handle at least four times its normal plate supply voltage (E_b), as shown by the plate voltage waveform in view (D).

Heterodyning the audio frequency intelligence from the modulator (V2) with the carrier in the plate circuit of the final power amplifier (V1) requires a large amount of audio power. All of the power or voltage that contains the intelligence must come from the modulator stage. This is why plate modulation is called high-level modulation.

The heterodyning action in the plate modulator effectively changes an audio frequency to a different part of the frequency spectrum. This action allows antennas and equipment of practical sizes to be used to transmit the intelligence. Now, let's look at several other typical modulators.

Collector-Injection Modulator

The COLLECTOR-INJECTION MODULATOR is the transistor equivalent of the electron-tube AM plate modulator. This transistor modulator can be used for low-level or relatively high-level modulation. It is referred to as relatively high-level modulation because, at the present time, transistors are limited in their power-handling capability. As illustrated in figure 1-47, the circuit design for a transistor collector-injection modulator is very similar to that of a plate modulator. The collector-injection modulator is capable of 100-percent modulation with medium power-handling capabilities.

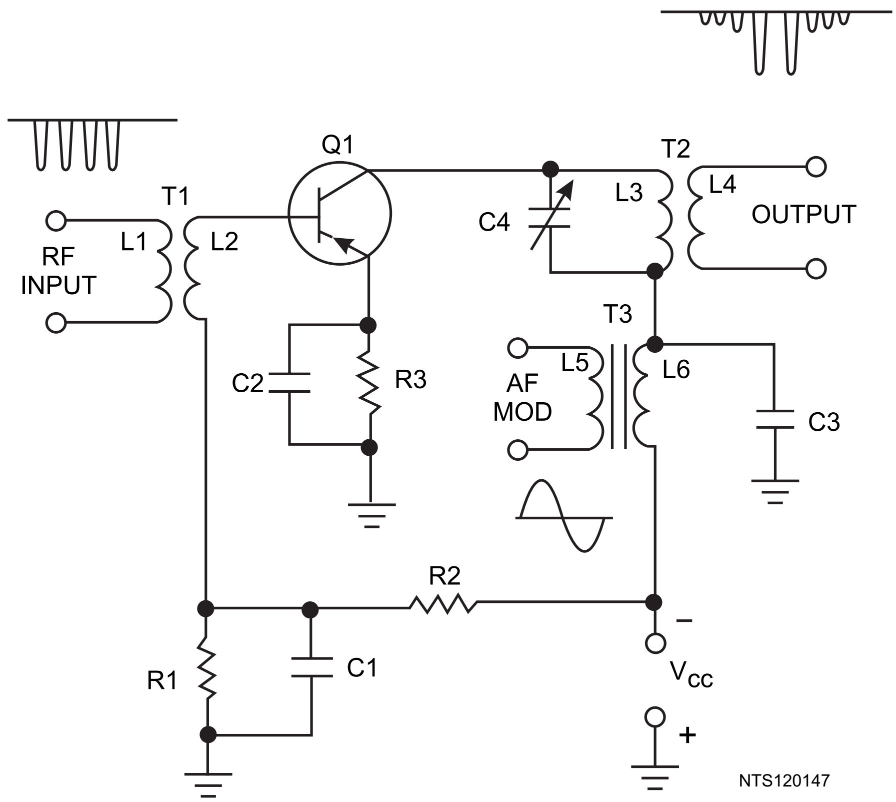


Figure 1-47.—Collector-injection modulator.

In figure 1-47, the rf carrier is applied to the base of modulator Q1. The modulating signal is applied to the collector in series with the collector supply voltage through T3. The output is then taken from the secondary of T2. With no modulating signal, Q1 acts as an rf amplifier for the carrier frequency. When the modulation signal is applied, it adds to or subtracts from the collector supply voltage. This causes the rf current pulses of the collector to vary in amplitude with the collector supply voltage. These collector current pulses cause oscillations in the tank circuit (C4 and the primary of T2). The tank circuit is tuned to the carrier frequency. During periods when the collector current is high, the tank circuit oscillates strongly. At times when the collector current is small, or entirely absent, little or no energy is supplied to the tank and oscillations become weak or die out. Thus, the modulation envelope is developed as it was in a plate modulator.

As transistor technology continues to develop, higher power applications of transistor collector-injection modulation will be employed. Plate and collector-injection modulation are the most commonly used types of modulation because the modulating signal can be applied in the final stages of rf amplification. This allows the majority of the rf amplifier stages to be operated class C for maximum efficiency. The plate and collector-injection modulators also require large amounts of af modulating power since the modulator stage must supply the power contained in the sidebands.

Q-40. For what class of operation is the final rf power amplifier of a plate-modulator circuit biased?

Q-41. The modulator is required to be what kind of a circuit stage in a plate modulator?

Q-42. How much must the fpa plate current vary to produce 100-percent modulation in a plate modulator?

Q-43. The collector-injection modulator is similar to what type of tube modulator?

Control-Grid Modulator

In cases when the use of a minimum of af modulator power is desired, a form of low-level modulation is necessary. The CONTROL-GRID MODULATOR is used widely in portable and mobile equipment to reduce size and power requirements. It is also used in extremely high-power, wideband equipment such as television transmitters where high-level or plate modulation is difficult and costly to achieve. Figure 1-48 is a basic schematic for a typical control-grid modulator.

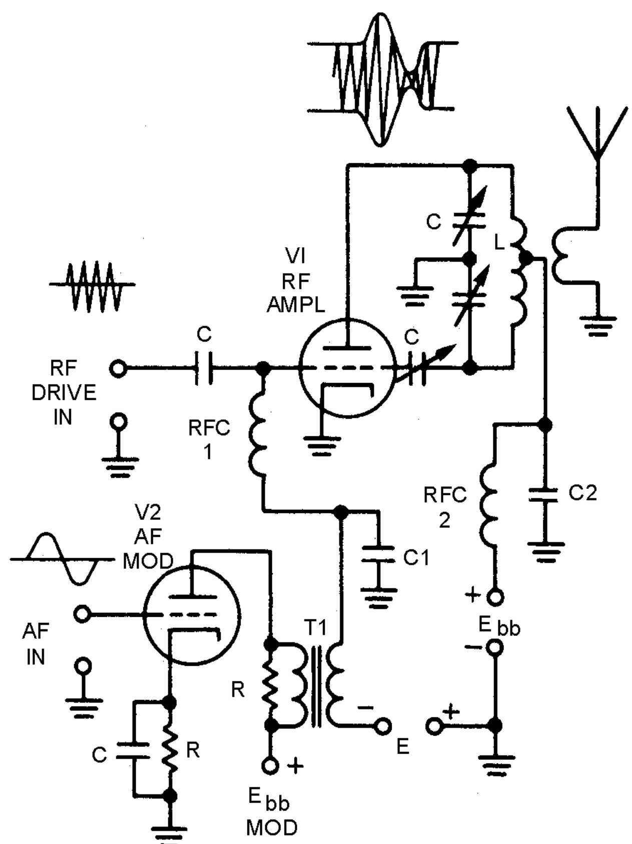


Figure 1-48.—Control-grid modulator.

The control-grid modulator uses a variation of grid bias (at the frequency of the modulating signal) to vary the instantaneous plate voltage and current. These variations cause modulation of the carrier frequency. The carrier frequency is introduced through coupling capacitor $_{Cc}$. The modulating frequency is introduced in series with the grid bias through T1. As the modulating signal increases and decreases (positive and negative), it will add to or subtract from the bias on rf amplifier V1. This change in bias causes a corresponding change in plate voltage and current. These changes in plate voltage and current add vectorially to the carrier frequency and provide a modulation envelope in the same fashion as does the plate modulator. Since changes in the plate circuit of the rf amplifier are controlled by changes in the grid bias, the gain of the tube requires only a low-level modulating signal. Even when the input signals are at these low levels, occasional modulation voltage peaks will occur that will cause V1 to saturate. This creates distortion in the output. Care must be taken to bias the rf amplifier tube for maximum power out while maintaining minimum distortion. The power to develop the modulation envelope comes from the rf amplifier. Because the rf amplifier has to be capable of supplying this additional power, it is biased for (and driven by the carrier frequency at) a much lower output level than its rating. This reduced efficiency is necessary during nonmodulated periods to provide the tube with the power to develop the sidebands.

Compared to plate modulation, grid modulation is less efficient, produces more distortion, and requires the rf power amplifier to supply all the power in the output signal. Grid modulation has the advantage of not requiring much power from the modulator.

Base-Injection Modulator

The BASE-INJECTION MODULATOR is similar to the control-grid modulator in electron-tube circuits. It is used to produce low-level modulation in equipment operating at very low power levels.

In figure 1-49, the bias on Q1 is established by the voltage divider R1 and R2. With the rf carrier input at T1, and no modulating signal, the circuit acts as a standard rf amplifier. When a modulating signal is injected through C1, it develops a voltage across R1 that adds to or subtracts from the bias on Q1. This change in bias changes the gain of Q1, causing more or less energy to be supplied to the collector tank circuit. The tank circuit develops the modulation envelope as the rf frequency and af modulating frequency are mixed in the collector circuit. Again, this action is identical to that in the plate modulator.

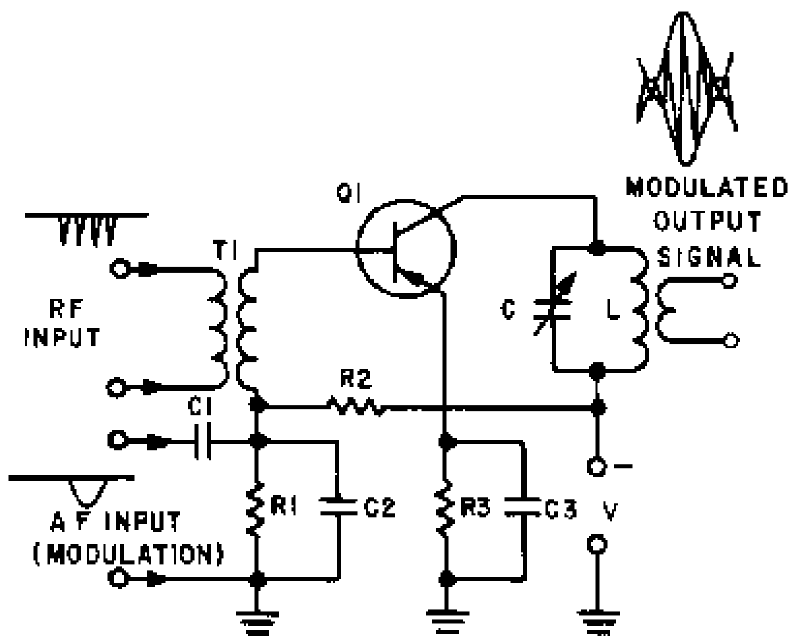


Figure 1-49.—Base-injection modulator.

Because of the extremely low-level signals required to produce modulation, the base-injection modulator is well suited for use in small, portable equipment, such as "walkie-talkies," and test equipment.

Cathode Modulator

Another low-level modulator, the CATHODE MODULATOR, is generally employed where the audio power is limited and distortion of the grid-modulated circuit cannot be allowed. The cathode modulator varies the voltage of the cathode to produce the modulation envelope. Since the cathode is in series with the grid and plate circuits, you should be able see that changing the cathode voltage will effectively change the voltage of the other tube elements. By properly controlling the voltages on the tube, you can cause the cathode modulator to operate in a form of plate modulation with high efficiency. Usually, the cathode modulator is designed to perform about midway between plate and grid modulator levels, using the advantages of each type. When operated between the two levels, the modulator provides a more linear output with moderate efficiency and a modest audio power requirement.

In figure 1-50, the rf carrier is applied to the grid of V1 and the modulating signal is applied in series with the cathode through T1. Since the modulating signal is effectively in series with the grid and plate voltage, the level of modulating voltage required will be determined by the relationships of the three voltages. The modulation takes place in the plate circuit with the plate tank developing the modulation envelope, just as it did in the plate modulator.

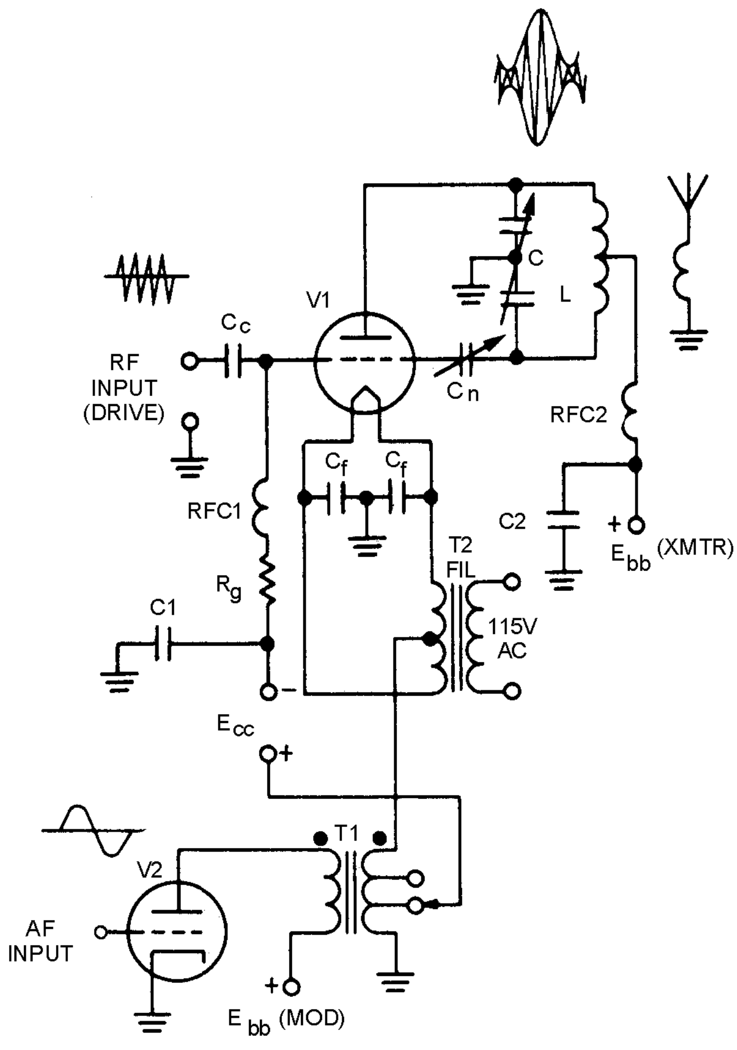


Figure 1-50.—Cathode modulator.

Emitter-Injection Modulator

This is the transistor equivalent of the cathode modulator. The EMITTER-INJECTION MODULATOR has the same characteristics as the base-injection modulator discussed earlier. It is an extremely low-level modulator that is useful in portable equipment. In emitter-injection modulation, the gain of the rf amplifier is varied by the changing voltage on the emitter. The changing voltage is caused by the injection of the modulating signal into the emitter circuitry of Q1, as shown in figure 1-51. Here the modulating voltage adds to or subtracts from transistor biasing. The change in bias causes a change in collector current and results in a heterodyning action. The modulation envelope is developed across the collector-tank circuit.

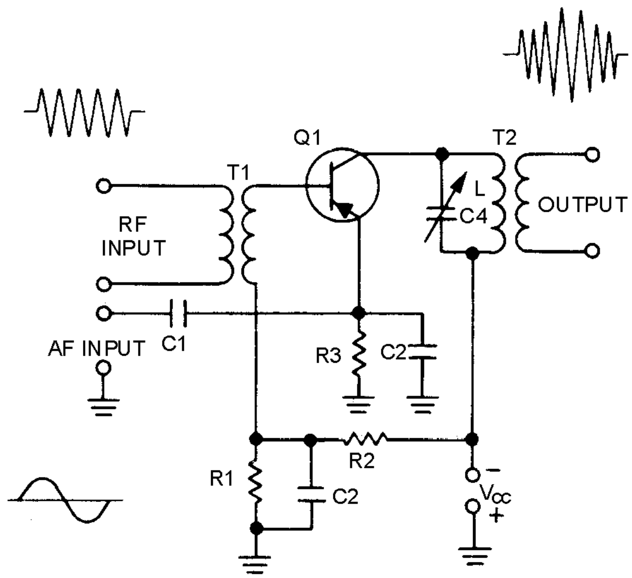


Figure 1-51.—Emitter-injection modulator.

Q-44. When is a control-grid modulator used?

Q-45. What type of modulator is the cathode modulator (low- or high-level)?

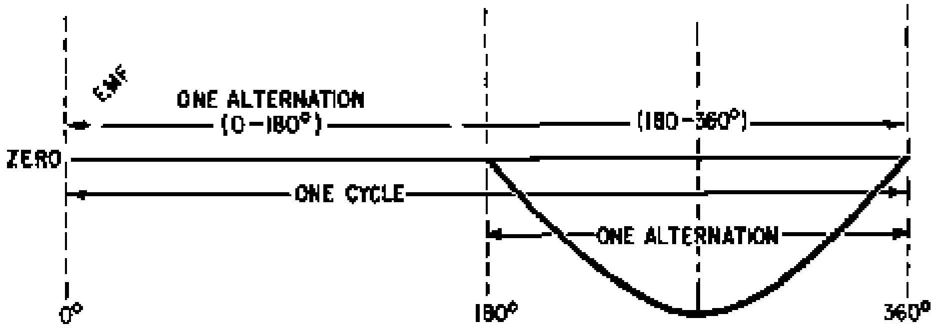
Q-46. What causes the change in collector current in an emitter-injection modulator?

You have studied six methods of amplitude modulation. These are not the only methods available, but they are the most common. All methods of AM modulation use the same theory of heterodyning across a nonlinear device. AM modulation is one of the easiest and least expensive types of modulation to achieve. The primary disadvantages of AM modulation are susceptibility to noise interference and the inefficiency of the transmitter. Power is wasted in the transmission of the carrier frequency because it contains no AM intelligence. In the next chapter, you will study other forms of modulation that have been developed to overcome these disadvantages.

SUMMARY

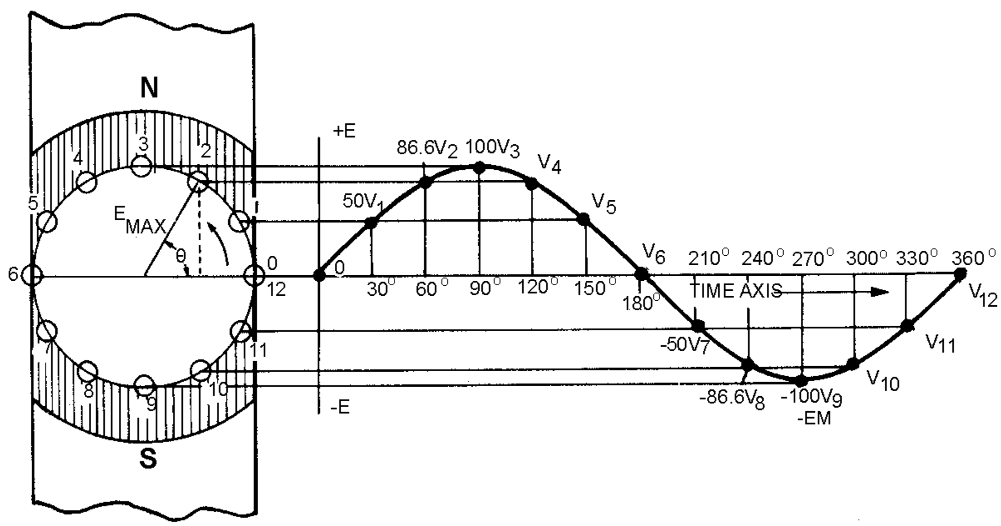
Now that you have completed this chapter, a short review of what you have learned is in order. The following summary will refresh your memory of amplitude modulation, its basic principles, and typical circuitry used to generate this modulation.

The **SINE WAVE** is the basis for all complex waveforms and is generated by moving a coil through a magnetic field.



AMPLITUDE (instantaneous voltage) of a coil is found by the formula:

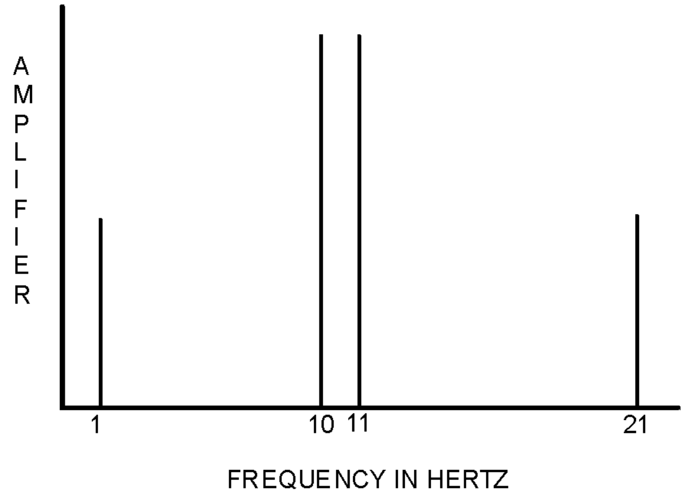
$$e = E_{max} \sin e\,\theta$$



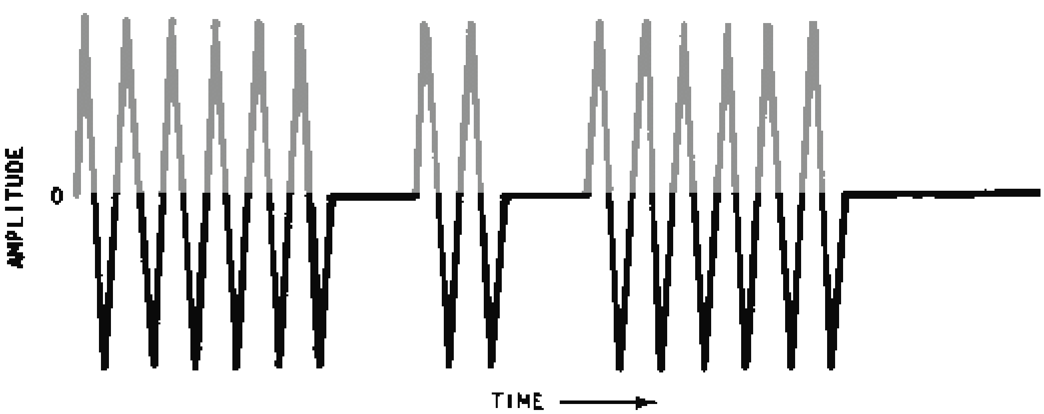
PHASE or **PHASE ANGLE** is the angle that exists between the starting position of a vector generating the sine wave and its position at a given instant.

FREQUENCY is the rate at which the vector rotates.

HETERODYNING is the process of mixing two different frequencies across a nonlinear impedance to give the ORIGINAL frequencies, a SUM frequency, and a DIFFERENCE frequency.



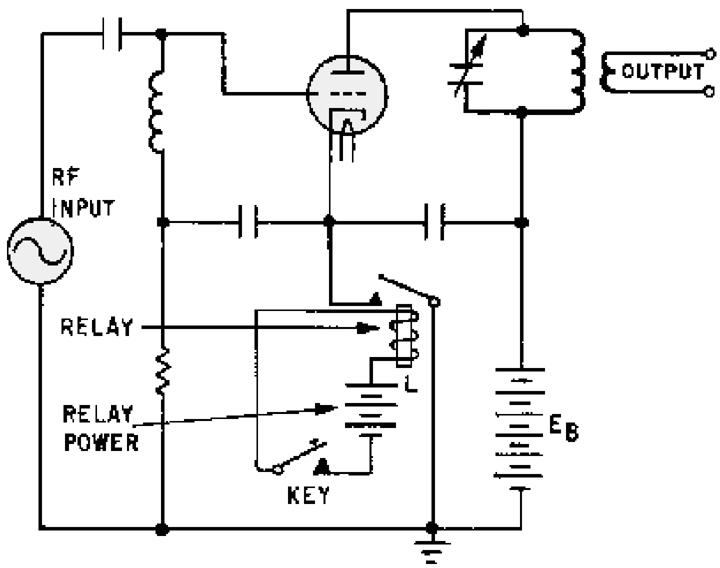
CONTINUOUS-WAVE MODULATION is the basic form of rf communications. It is essentially on-off keying of an rf carrier.



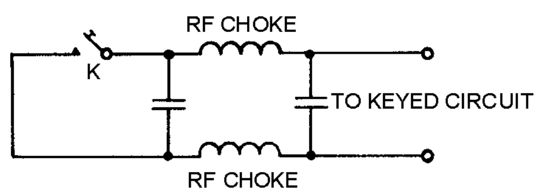
(C) RESULTANT CARRIER WAVE TRANSMITTED ALONG CONNECTING PATH.

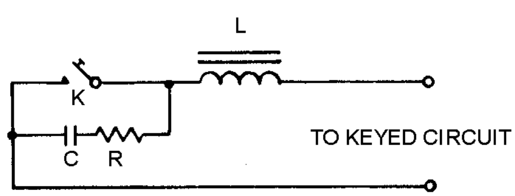
HAND-OPERATED and MACHINE KEYING are two types of cw keying. PLATE, CATHODE, and BLOCKED-GRID KEYING are circuits commonly used in hand-operated and machine keying.

KEYING RELAYS are used for safety and to handle the current requirements in high-power transmitters.



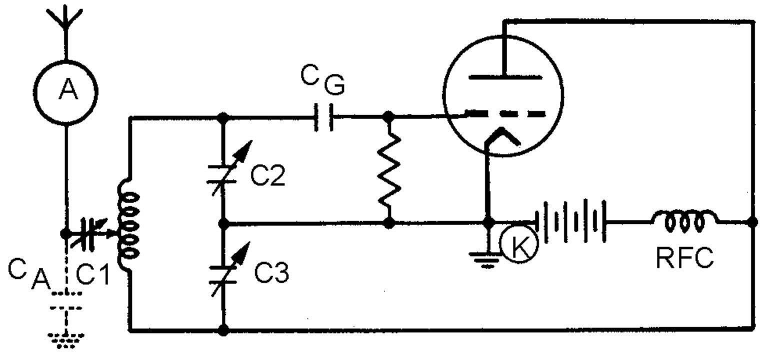
KEY-CLICK FILTERS are used to prevent interference in cw transmitters.



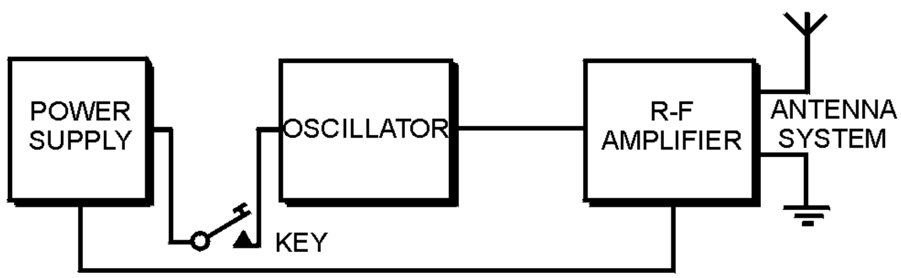


Although it is a relatively slow transmission method, CW COMMUNICATIONS is highly reliable under severe noise conditions for long-range operation.

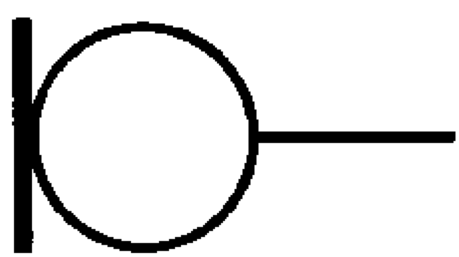
SINGLE-STAGE CW TRANSMITTERS can be made by coupling the output of an oscillator to an antenna.



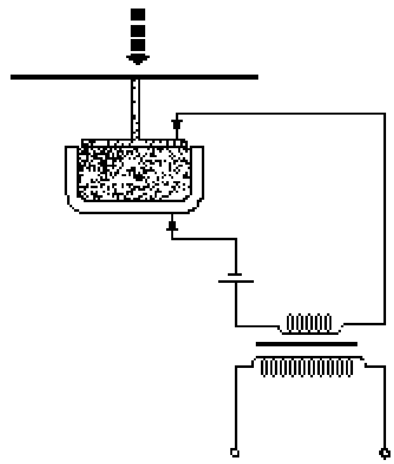
MULTISTAGE CW TRANSMITTERS are used to improve frequency stability and increase output power.



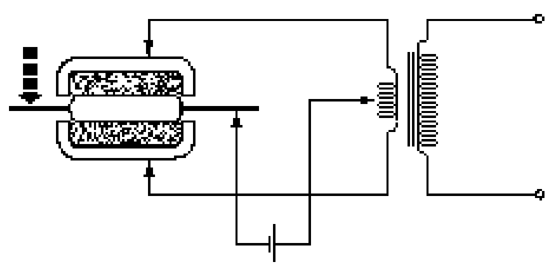
A **MICROPHONE** is an energy converter that changes sound energy into electrical energy.



A **CARBON MICROPHONE** uses carbon granules and an external battery supply to generate af voltages from sound waves.

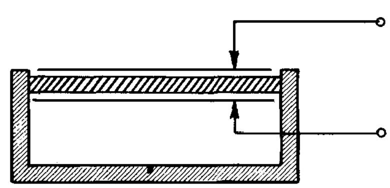


SINGLE-BUTTON CARBON MICROPHONE

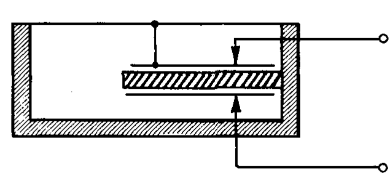


DOUBLE-BUTTON CARBON MICROPHONE

A **CRYSTAL MICROPHONE** uses the piezoelectric effect to generate an output voltage.

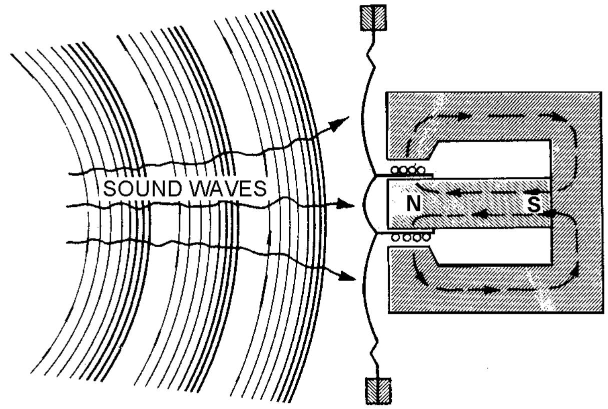


DIRECTLY ACTUATED TYPE

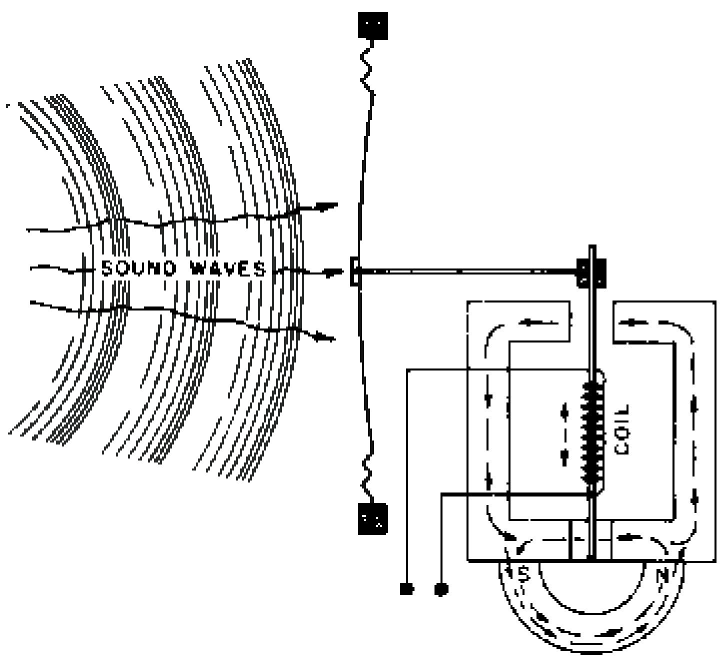


DIAPHRAGM TYPE

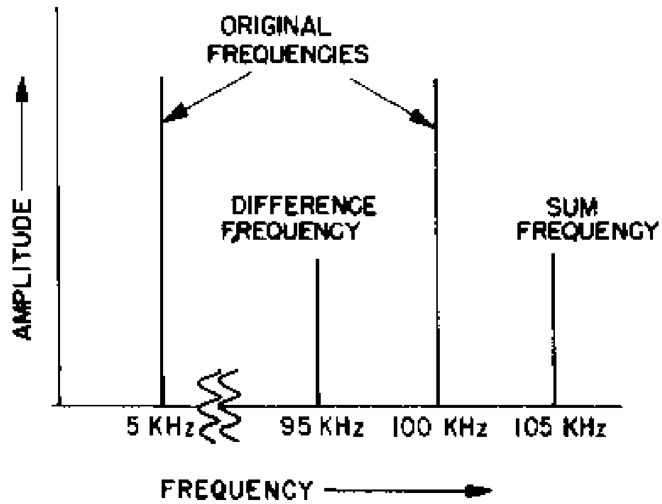
A **DYNAMIC MICROPHONE** uses a coil of fine wire mounted on the back of a diaphragm located in the magnetic field of a permanent magnet.



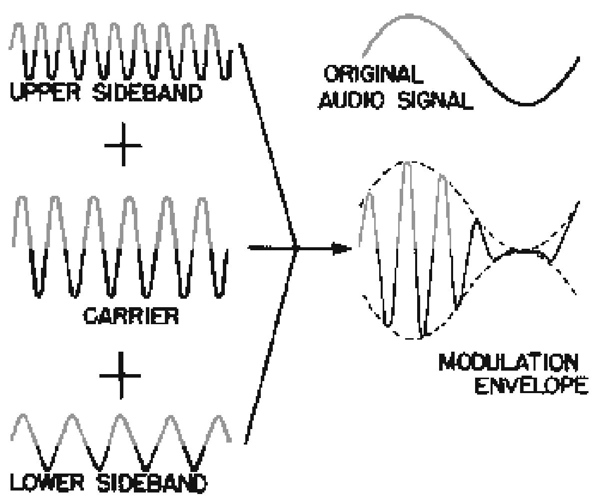
A **MAGNETIC MICROPHONE** uses a moving armature in a magnetic field to generate an output.



The **FREQUENCY SPECTRUM** of a modulated wave can be conveniently illustrated in graph form as frequency versus amplitude.

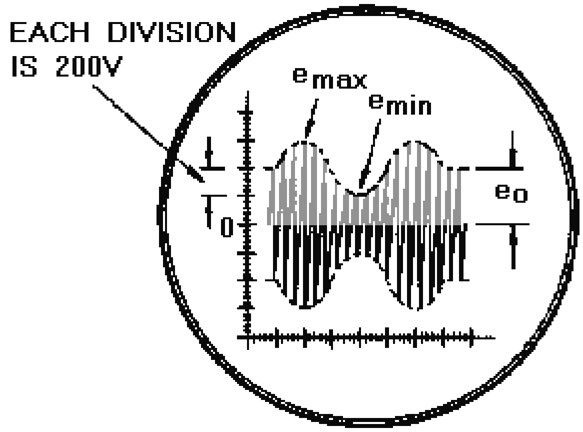


The **MODULATION ENVELOPE** is the waveform observed when the CARRIER, UPPER SIDEBAND, and LOWER SIDEBAND are combined in a single impedance and observed as time versus amplitude.



The **BANDWIDTH** of an rf signal is the amount of space in the frequency spectrum used by the signal.

PERCENT OF MODULATION is a measure of the relative magnitudes of the rf carrier and the af modulating signal.



$$\%M = \frac{e_{max} - e_{min}}{2e_o} \times 100$$

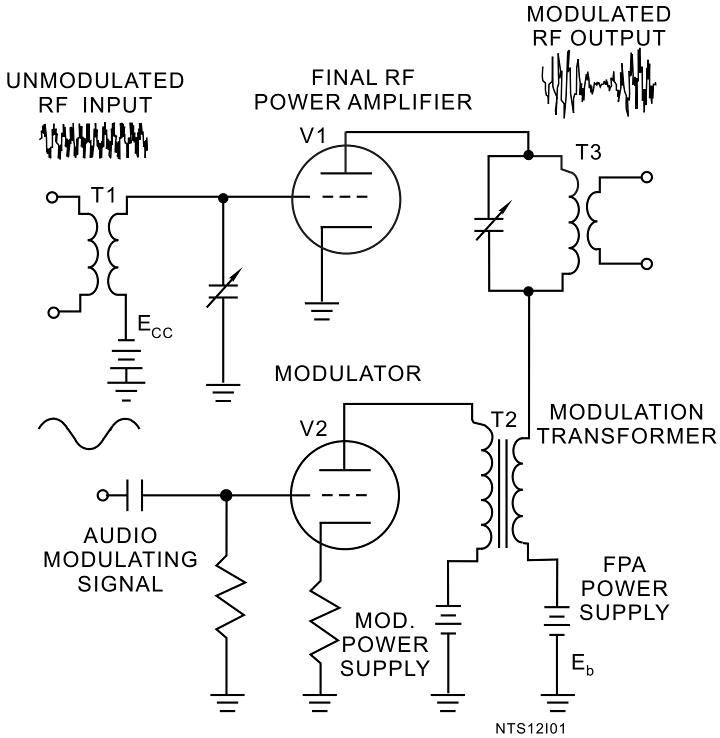
-OR-

$$\%M = \frac{e_{max} - e_{min}}{e_{max} + e_{min}} \times 100$$

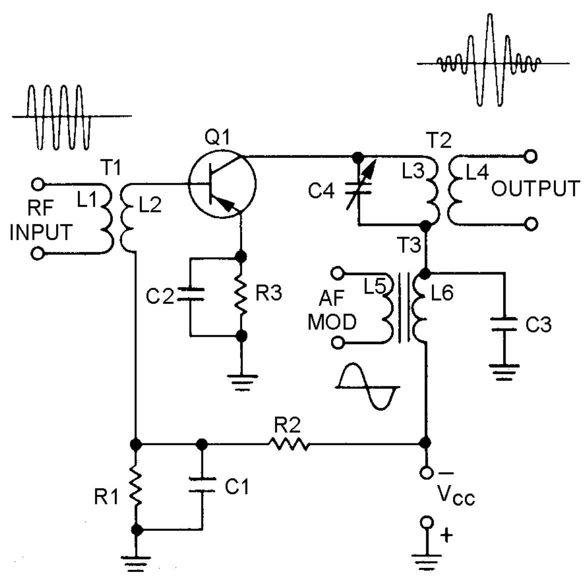
HIGH-LEVEL MODULATION is modulation produced in the plate circuit of the last radio stage of the system.

LOW-LEVEL MODULATION is modulation produced in an earlier stage than the final power amplifier.

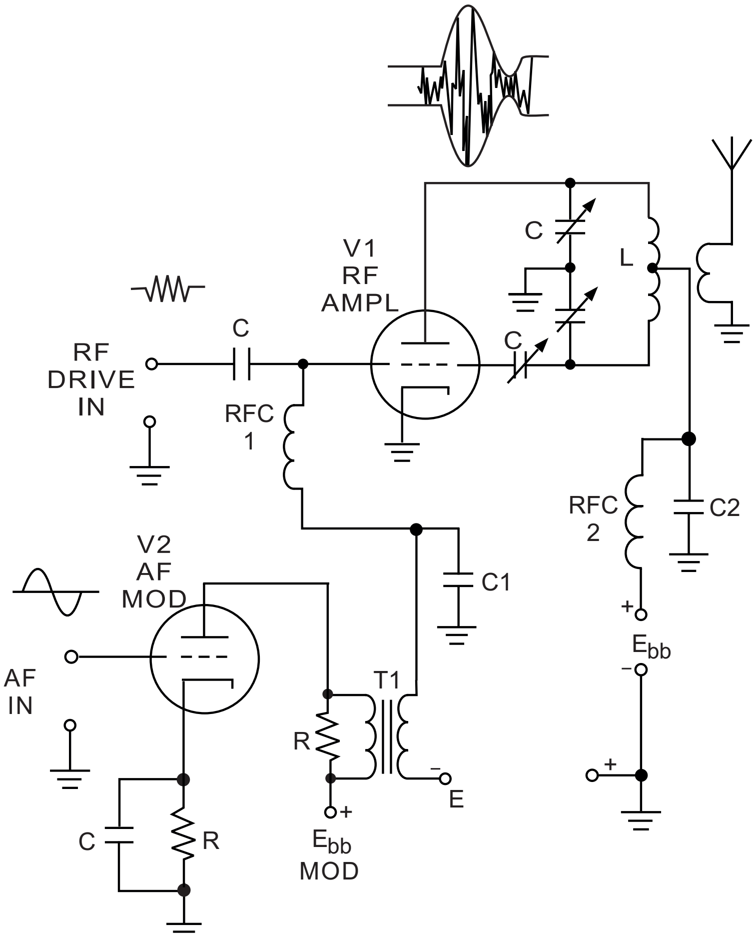
The **PLATE MODULATOR** is a high-level modulator. The modulator tube must be capable of varying the plate-supply voltage of the final power amplifier. It must vary the plate voltage so that the plate current pulses will vary between 0 and nearly twice their unmodulated value to achieve 100-percent modulation.



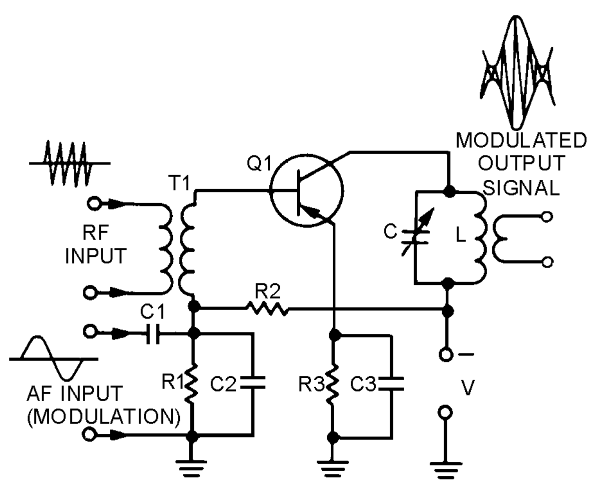
A **COLLECTOR-INJECTION MODULATOR** is a transistorized version of the plate modulator. It is classified as a high-level modulator, although present state-of-the-art transistors limit them to medium-power applications.



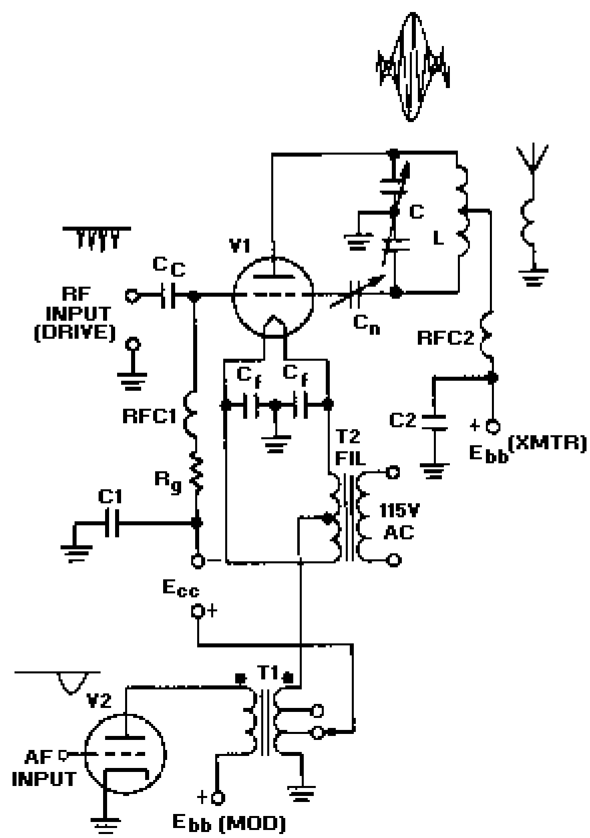
A **CONTROL-GRID MODULATOR** is a low-level modulator that is used where a minimum of af modulator power is desired. It is less efficient than a plate modulator and produces more distortion.



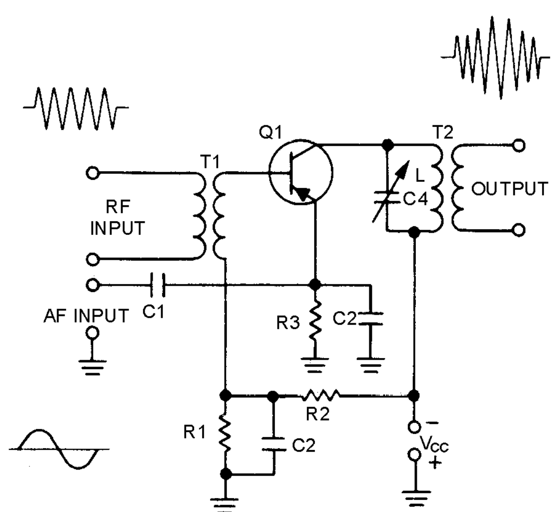
A **BASE-INJECTION MODULATOR** is used to produce low-level modulation in equipment operating at very low power levels. It is often used in small portable equipment and test equipment.



The **CATHODE MODULATOR** is a low-level modulator employed where the audio power is limited and the inherent distortion of the grid modulator cannot be tolerated.



The **EMITTER-INJECTION MODULATOR** is an extremely low-level modulator that is useful in portable equipment.



The primary disadvantages of AM modulation are susceptibility to NOISE INTERFERENCE and the INEFFICIENCY of the transmitter.

ANSWERS TO QUESTIONS Q1. THROUGH Q46.

A-1. Modulation is the impressing of intelligence on a transmission medium.

A-2. May be anything that transmits information, such as light, smoke, sound, wire lines, or radio-frequency waves.

A-3. Mixing two frequencies across a nonlinear impedance.

A-4. The process of recovering intelligence from a modulated carrier.

A-5. The sine wave.

A-6. To represent quantities that have both magnitude and direction.

A-7. Sine ! = opposite side ÷ hypotenuse.

A-8. $e = E_{max}$ *sine ! .*

A-9. The value at any given point on the sine wave.

A-10. Phase or phase angle.

A-11. The rate at which the vector which is generating the sine wave is rotating.

A-12. The elapsed time from the beginning of cycle to its completion.

A-13. Wavelength = rate of travel × period.

A-14. Process of combining two signal frequencies in a nonlinear device.

A-15. An impedance in which the resulting current is not proportional to the applied voltage.

A-16. The display of electromagnetic energy that is arranged according to wavelength or frequency.

A-17. At least two different frequencies applied to a nonlinear impedance.

A-18. Any method of modulating an electromagnetic carrier frequency by varying its amplitude in accordance with the intelligence.

A-19. A method of generating oscillations, a method of turning the oscillations on and off (keying), and an antenna to radiate the energy.

A-20. Plate keying and cathode keying.

A-21. Machine keying.

A-22. A high degree of clarity even under severe noise conditions, long-range operation, and narrow bandwidth.

A-23. Antenna-to-ground capacitance can cause the oscillator frequency to vary.

A-24. To isolate the oscillator from the antenna and increase the amplitude of the rf oscillations to the required output level.

A-25. To raise the low frequency of a stable oscillator to the vhf range.

A-26. An energy converter that changes sound energy into electrical energy.

A-27. The changing resistance of carbon granules as pressure is applied to them.

A-28. Background hiss resulting from random changes in the resistance between individual carbon granules.

A-29. The piezoelectric effect.

A-30. A dynamic microphone has a moving coil and the magnetic microphone has a moving armature.

A-31. Rf and af units.

A-32. 100 kilohertz, 5 kilohertz, 95 kilohertz, and 105 kilohertz.

A-33. All of the sum frequencies above the carrier.

A-34. The intelligence is contained in the spacing between the carrier and sideband frequencies.

A-35. The highest modulating frequency.

A-36. The depth or degree of modulation.

A-37. One-half the amplitude of the carrier.

A-38.

$$\%M = \frac{E_m}{E_c} \times 100$$

A-39. Modulation produced in the plate circuit of the last radio stage of the system.

A-40. Class C.

A-41. Power amplifier.

A-42. Between 0 and nearly two times its unmodulated value.

A-43. Plate modulator.

A-44. In cases when the use of a minimum of af modulator power is desired.

A-45. Low-level.

A-46. Gain is varied by changing the voltage on the emitter.

CHAPTER 2

ANGLE AND PULSE MODULATION

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Describe frequency-shift keying (fsk) and methods of providing this type of modulation.
2. Describe the development of frequency modulation (fm) and methods of frequency modulating a carrier.
3. Discuss the development of phase modulation (pm) and methods of phase modulating a carrier.
4. Describe phase-shift keying (psk), its generation, and application.
5. Discuss the development and characteristics of pulse modulation.
6. Describe the operation of the spark gap and thyratron modulators.
7. Discuss the characteristics of a pulse train that may be varied to provide communications capability.
8. Describe pulse-amplitude modulation (pam) and generation.
9. Describe pulse-duration modulation (pdm) and generation.
10. Describe pulse-position modulation (ppm) and generation.
11. Describe pulse-frequency modulation (pfm) and generation.
12. Describe pulse-code modulation (pcm) and generation.

INTRODUCTION

In chapter 1 you learned that modulation of a carrier frequency was necessary to allow fast communications between two points. As the volume of transmissions increased, a need for more reliable methods of communication was realized. In this chapter you will study angle modulation and pulse modulation. These two types of modulation have been developed to overcome one of the main disadvantages of amplitude modulation - susceptibility to noise interference. In addition, a special application of pulse type modulation for ranging and detection equipment will be discussed.

ANGLE MODULATION

ANGLE MODULATION is modulation in which the angle of a sine-wave carrier is varied by a modulating wave. FREQUENCY MODULATION (fm) and PHASE MODULATION (pm) are two types of angle modulation. In frequency modulation the modulating signal causes the carrier frequency to vary. These variations are controlled by both the frequency and the amplitude of the modulating wave. In phase modulation the phase of the carrier is controlled by the modulating waveform. Let's study these modulation methods for an understanding of their similarities and differences.

FREQUENCY-MODULATION SYSTEMS

In frequency modulation an audio signal is used to shift the frequency of an oscillator at an audio rate. The simplest form of this is seen in FREQUENCY-SHIFT KEYING (fsk). Frequency-shift keying is somewhat similar to continuous-wave keying (cw) in AM transmissions.

Frequency-Shift Keying

Consider figure 2-1, views (A) through (D). View (A) is a radio frequency (rf) carrier which is actually several thousand or million hertz. View (B) represents the intelligence to be transmitted as MARKS and SPACES. Recall that in cw transmission, this intelligence was applied to the rf carrier by interrupting the signal, as shown in view (C). The amplitude of the rf alternated between maximum and 0 volts. By comparing views (B) and (C), you can see the mark/space intelligence of the Morse code character on the rf. The spacing of the waveform in view (D) is an example of the same intelligence as it is applied to the frequency instead of the amplitude of the rf. This is simple frequency-shift keying of the same Morse code character.

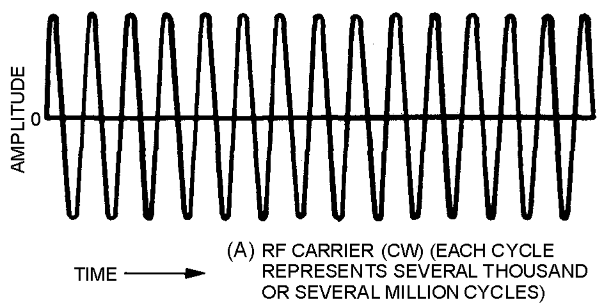


Figure 2-1A.—Comparison of ON-OFF and frequency-shift keying. RF CARRIER (CW) (EACH CYCLE REPRESENTS SEVERAL THOUSAND OR SEVERAL MILLION CYCLES).

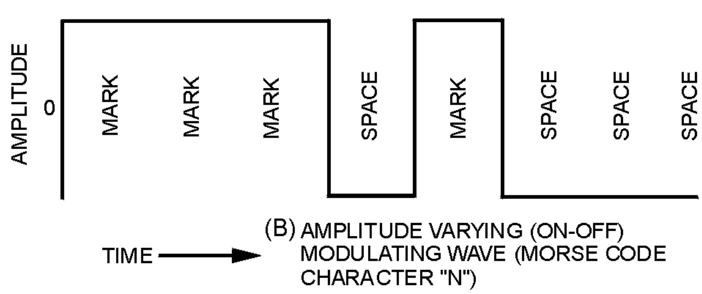


Figure 2-1B.—Comparison of ON-OFF and frequency-shift keying. AMPLITUDE VARYING (ON-OFF) MODULATING WAVE (MORSE CODE CHARACTER "N").

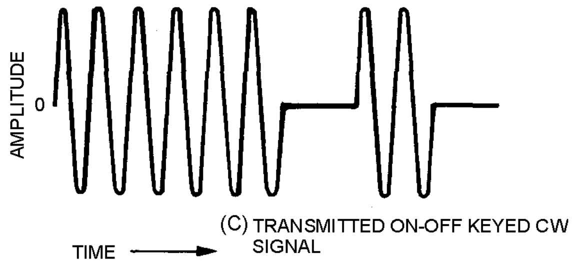


Figure 2-1C.—Comparison of ON-OFF and frequency-shift keying. TRANSMITTED ON-OFF KEYED CW SIGNAL.

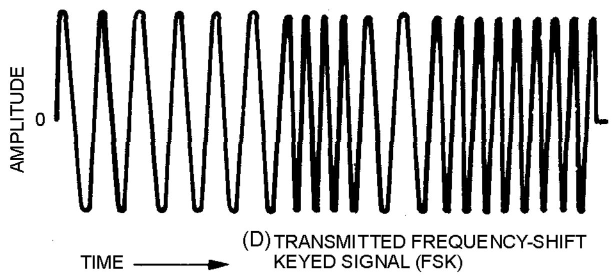


Figure 2-1D.—Comparison of ON-OFF and frequency-shift keying. TRANSMITTED FREQUENCY-SHIFT KEYED SIGNAL (FSK).

In fsk the output is abruptly changed between two differing frequencies by opening and closing the key. This is shown in view (D). For illustrative purposes, the spacing frequency in view (D) is shown as double the marking frequency. However, in practice the difference is usually less than 1,000 hertz, even when operating at several megahertz. You should also note that the limit of frequency shift is determined without reference to the amplitude of the keying signal in the fsk system. The frequency shift may be set at plus or minus 425 hertz from the allocated channel frequency. The total shift between mark and space would be 850 hertz. Either the mark or space may use the higher of the two frequencies. The upper frequency of the transmitted signal is usually the spacing interval and the lower frequency is the marking interval.

COMPARING FSK AND CW SIGNALS.—A comparison of on-off keyed cw (figure 2-2), (view (A), view (B), view (C)), and fsk (figure 2-3), (view (A), view (B), view (C)), signals will show clearly the principal features of fsk and give us a basis on which frequency modulation can be discussed. Let's use views (A), (B), and (C) of both figures to show the Morse code character "F" for an example. Figures 2-2 and 2-3 are graphic drawings of the two types of keying. Time and amplitude are known dimensions of AM; but to explain fsk properly, we have added the third dimension of frequency.

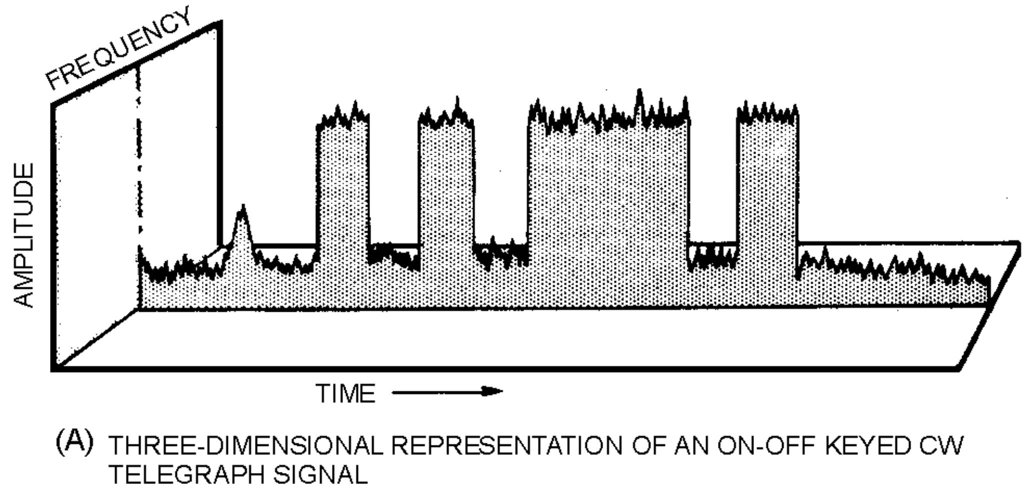


Figure 2-2A.—Comparison of AM and fm receiver response to an AM signal. THREE-DIMENSIONAL REPRESENTATION OF AN ON-OFF KEYED CW TELEGRAPH SIGNAL.

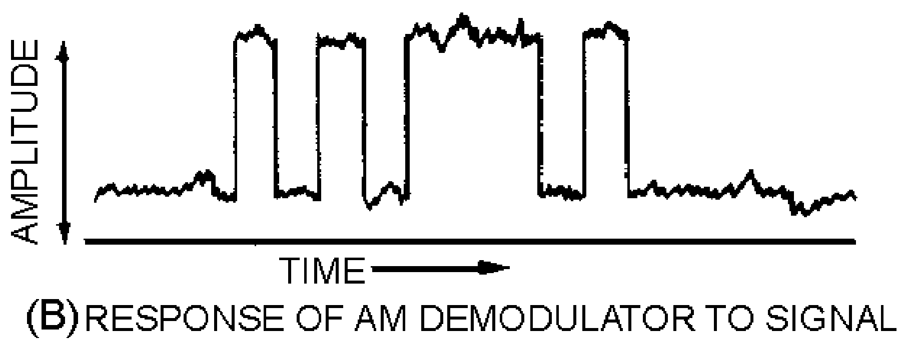


Figure 2-2B.—Comparison of AM and fm receiver response to an AM signal. RESPONSE OF AM DEMODULATOR TO SIGNAL.

Figure 2-2C.—Comparison of AM and fm receiver response to an AM signal.RESPONSE OF FM DISCRIMINATOR TO SIGNAL.

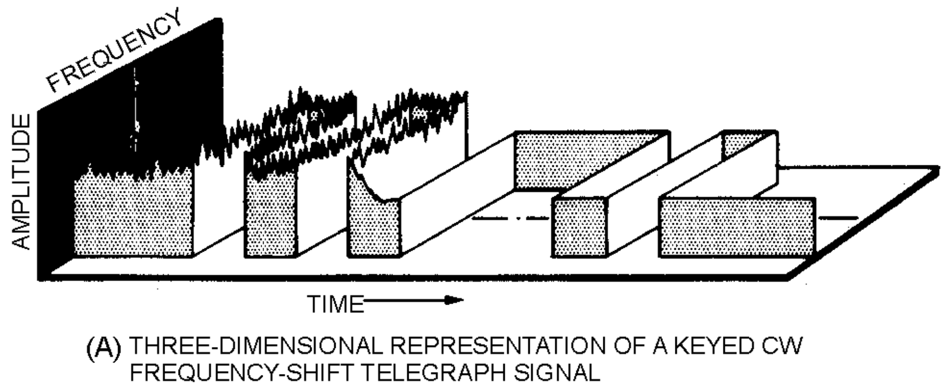


Figure 2-3A.—Comparison of AM and fm receiver response to an fm signal. THREE-DIMENSIONAL REPRESENTATION OF A FREQUENCY-SHIFT KEYED TELEGRAPH SIGNAL.

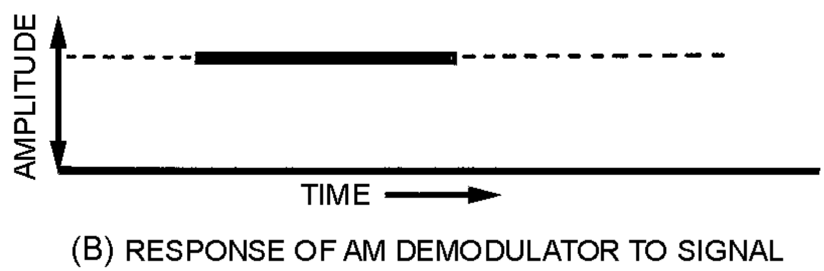


Figure 2-3B.—Comparison of AM and fm receiver response to an fm signal. RESPONSE OF AM DEMODULATOR TO SIGNAL.

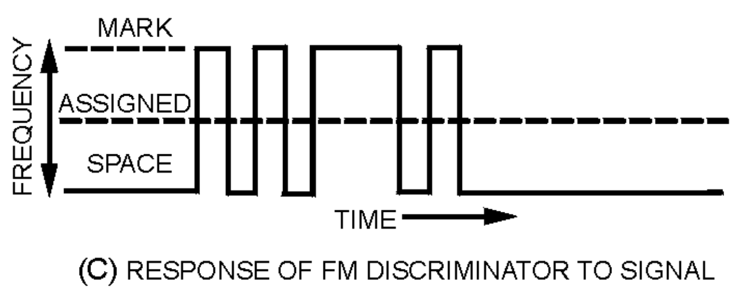


Figure 2-3C.—Comparison of AM and fm receiver response to an fm signal. RESPONSE OF FM DISCRIMINATOR TO SIGNAL.

CW SIGNALS.—Since cw signals are of essentially constant frequency, there is no variation along the frequency axis in view (A) of figure 2-2. The complete intelligence is carried as variations in the amplitude of the signal. To receive the intelligence carried by such a signal, the receiving equipment must be able to scan the signal along the time and amplitude axes, which carry the information. When scanned along the time and amplitude axes [shown in view (B)], the intelligence appears as large changes in amplitude. If the circuit were perfect, these variations would be from 0 amplitude to some maximum value (established by transmitter power, distance, and so forth) depending on whether the key were open or closed. However, interfering components of energy caused by atmospherics, interfering stations, and electrical machinery appear as additional variations along the amplitude axis. When these amplitude variations approach or exceed the variation caused by the keyed intelligence, the signal is blanked out by interference. We have all heard this happen on our AM radios during storms or when near operating machinery.

View (C) of figure 2-2 represents the same signal when scanned along the time and frequency axes as it would be in an fm receiver. Variations in signal amplitude have no effect on the frequency and no intelligence can be received. Note that the noise and interference components have also been suppressed so that they have little effect on the received signal. Thus, if the intelligence variations were impressed as changes along the frequency axis, and the receiving equipment were designed to respond to this type of signal, then the effects of noise and interference would be practically eliminated. Frequency-shift keyed circuits fulfill these conditions.

FSK SIGNALS.—In fsk the rf signal is shifted in frequency (not amplitude) between "key-open" and "key-closed" conditions. The signal amplitude remains essentially constant. View (A) of figure 2-3 represents the letter "F" keyed as a shift in frequency between mark and space. The normal frequency condition with the key open is a space. Recall that this may be either the lower or higher frequency. When the key is closed, the frequency instantly changes to the mark value and remains constant during the marking interval. Opening the key again returns the frequency to the space frequency. Midway between the mark and space frequencies is the assigned channel frequency.

Also shown in view (A) is the variation along the amplitude axis caused by the same noise and interference mentioned earlier. The right-hand portion of view (A) illustrates the elimination of this noise by the receiving equipment. View (B) clearly shows that scanning the signal along the amplitude and time axes reproduces no amplitude variations from signal interference. However, if the scanning is accomplished along the frequency and time axes, the intelligence is reproduced, as shown in view (C). By this system, the intelligence can be recovered at the receiving station in its original form; it will be nearly unaffected by conditions in the radio path other than fading. As a matter of fact, fsk resists the effects of fading better than cw.

FREQUENCY-SHIFT KEYING.—In its simplest form, frequency-shift keying of a transmitter can be accomplished by shunting a capacitor (or an inductor) and key (in series) across the oscillator circuit. By locking the normal key of the transmitter and operating only the oscillator circuit key, you can change the oscillator frequency. The shift in frequency between mark (key-closed) and space (key-open) conditions is determined by the effect of the additional capacitance (or inductance) on the oscillator frequency. The frequency multiplication factor in the transmitter amplifiers must be taken into consideration when determining the oscillator frequency shift. Thus, if the desired shift is the conventional 850 hertz at the transmission frequency, and this frequency is four times the oscillator frequency (that is, doubled in two stages), then the effect of the additional capacitance (or inductance) on the oscillator must be limited to 212.5 hertz as shown below:

$$\text{total frequency shift desired} = 850 \text{ hertz}$$

$$\text{multiplication factor per stage} = \times 2$$

$$\text{number of amplifier stages} = 2$$

$$\text{total multiplication factor} = 2 \times 2$$

$$\begin{aligned}\text{limiting value of capacitance/inductance} & \\ \text{in terms of frequency variation)} &= 4\,\overline{)\,850 \text{ hertz}} \\ &= 212.5 \text{ hertz}\end{aligned}$$

Frequency-shift keyers are, of course, more complicated than this simple illustration would seem to show, but the basic principles are the same. Still, the keyer does change the oscillator frequency by a certain number of cycles. Further, this change must be correlated with the multiplication factor of the transmitter to cause the desired shift between mark and space frequencies.

METHODS OF FREQUENCY SHIFTING.—Frequency-shift keyers operate on either of two general principles. First, the keyer may take the output of the transmitter's master oscillator and modulate it with the output of another oscillator that is frequency-shift keyed. This action will result in two frequencies that are used to excite the first amplifier stage of the transmitter. This system is illustrated in view (A) of figure 2-4. View (B) illustrates the second method of frequency-shift keyer operation. In this method the transmitter's master oscillator is itself shifted in frequency by the mark and space impulses from the keyer unit.

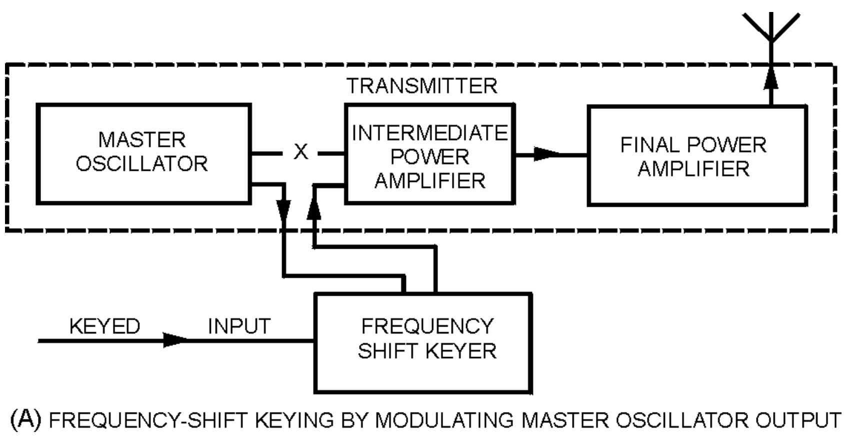


Figure 2-4A.—Two methods of frequency-shift keying (fsk). FREQUENCY-SHIFT KEYING BY MODULATING MASTER OSCILLATOR OUTPUT.

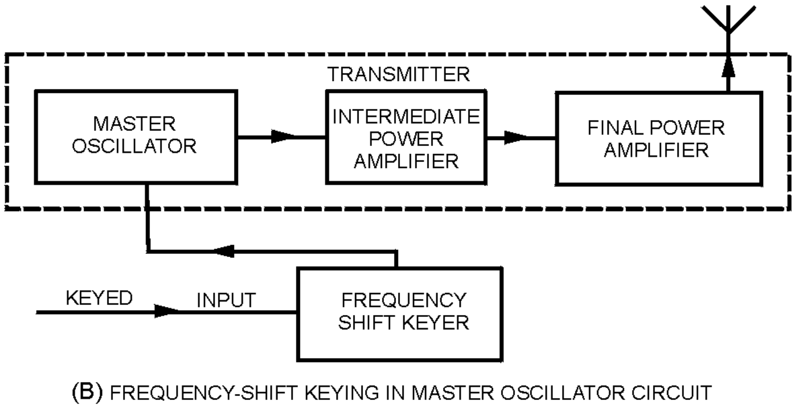


Figure 2-4B.—Two methods of frequency-shift keying (fsk). FREQUENCY-SHIFT KEYING IN MASTER OSCILLATOR CIRCUIT.

ADVANTAGES OF FSK OVER AM.—Frequency-shift keying is used in all single-channel, radiotelegraph systems that use automatic printing systems. The advantage of fsk over on-off keyed cw is that it rejects unwanted signals (noise) that are weaker than the desired signal. This is true of all fm systems. Also, since a signal is always present in the fsk receiver, automatic volume control methods may be used to minimize the effects of signal fading caused by ionospheric variations. The amount of inherent signal-to-noise ratio improvement of fsk over AM is approximately 3 to 4 dB. This improvement is because the signal energy of fsk is always present while signal energy is present for only one-half the time in AM systems. Noise is continuously present in both fsk and AM, but is eliminated in fsk reception. Under the rapid fading and high-noise conditions that commonly exist in the high frequency (hf) region, fsk shows a marked advantage over AM. Overall improvement is sometimes expressed as the RATIO OF TRANSMITTED POWERS required to give equivalent transmission results over the two systems. Such a ratio varies widely, depending on the prevailing conditions. With little fading, the ratio may be entirely the result of the improvement in signal-to-noise ratio and may be under 5 dB. However, under severe fading conditions, large amounts of power often fail to give good results for AM transmission. At the same time, fsk may be satisfactory at nominal power. The power ratio (fsk versus AM) would become infinite in such a case.

Another application of fsk is at low and very low frequencies (below 300 kilohertz). At these frequencies, keying speeds are limited by the "flywheel" effect of the extremely large capacitance and inductance of the antenna circuits. These circuits tend to oscillate at their resonant frequencies. Frequency-shifting the transmitter and changing the antenna resonance by the same keying impulses will result in much greater keying speeds. As a result, the use of these expensive channels is much more efficient.

Q-1. What are the two types of angle modulation?

Q-2. Name the modulation system in which the frequency alternates between two discrete values in response to the opening and closing of a key?

Q-3. What is the primary advantage of an fsk transmission system?

Frequency Modulation

In frequency modulation, the instantaneous frequency of the radio-frequency wave is varied in accordance with the modulating signal, as shown in view (A) of figure 2-5. As mentioned earlier, the amplitude is kept constant. This results in oscillations similar to those illustrated in view (B). The number of times per second that the instantaneous frequency is varied from the average (carrier frequency) is controlled by the *frequency* of the modulating signal. The amount by which the frequency departs from the average is controlled by the *amplitude* of the modulating signal. This variation is referred to as the FREQUENCY DEVIATION of the frequency-modulated wave. We can now establish two clear-cut rules for frequency deviation rate and amplitude in frequency modulation:

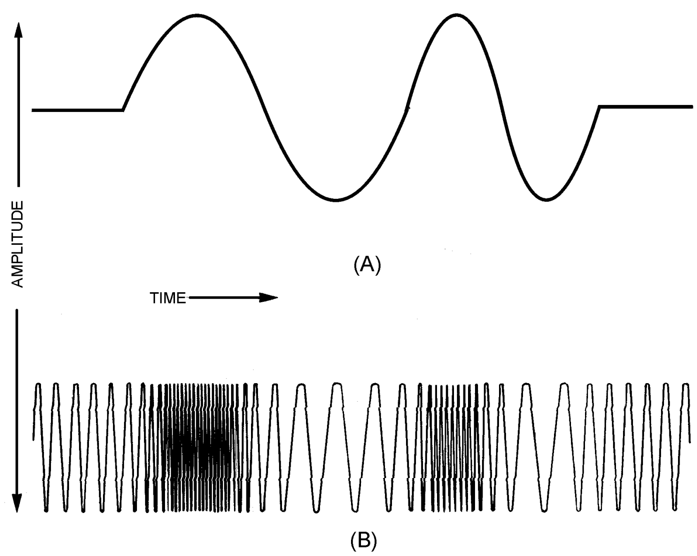


Figure 2-5.—Effect of frequency modulation on an rf carrier.

- *AMOUNT* OF FREQUENCY SHIFT IS PROPORTIONAL TO THE AMPLITUDE OF THE MODULATING SIGNAL

(This rule simply means that if a 10-volt signal causes a frequency shift of 20 kilohertz, then a 20-volt signal will cause a frequency shift of 40 kilohertz.)

- RATE OF FREQUENCY SHIFT IS PROPORTIONAL TO THE FREQUENCY OF THE MODULATING SIGNAL

(This second rule means that if the carrier is modulated with a 1-kilohertz tone, then the carrier is changing frequency 1,000 times each second.)

Figure 2-6 illustrates a simple oscillator circuit with the addition of a condenser microphone (M) in shunt with the oscillator tank circuit. Although the condenser microphone capacitance is actually very low, the capacitance of this microphone will be considered near that of the tuning capacitor (C). The frequency of oscillation in this circuit is, of course, determined by the LC product of all elements of the circuit; but, the product of the inductance (L) and the combined capacitance of C and M are the primary frequency components. When no sound waves strike M, the frequency is the rf carrier frequency. Any excitation of M will alter its capacitance and, therefore, the frequency of the oscillator circuit. Figure 2-7 illustrates what happens to the capacitance of the microphone during excitation. In view (A), the audio-frequency wave has three levels of intensity, shown as **X**, a whisper; **Y**, a normal voice; and **Z**, a loud voice. In view (B), the same conditions of intensity are repeated, but this time at a frequency twice that of view (A). Note in each case that the capacitance changes both positively and negatively; thus the frequency of oscillation alternates both above and below the resting frequency. The amount of change is determined by the change in capacitance of the microphone. The change is caused by the amplitude of the sound wave exciting the microphone. The rate at which the change in frequency occurs is determined by the rate at which the capacitance of the microphone changes. This rate of change is caused by the frequency of the sound wave. For example, suppose a 1,000-hertz tone of a certain loudness strikes the microphone. The frequency of the carrier will then shift by a certain amount, say plus and minus 40 kilohertz. The carrier will be shifted 1,000 times per second. Now assume that with its loudness unchanged, the frequency of the tone is changed to 4,000 hertz. The carrier frequency will still shift plus and minus 40 kilohertz; but now it will shift at a rate of 4,000 times per second. Likewise, assume that at the same loudness, the tone is reduced to 200 hertz. The carrier will continue to shift plus and minus 40 kilohertz, but now at a rate of 200 times per second. If the loudness of any of these modulating tones is reduced by one-half, the frequency of the carrier will be shifted plus and minus 20 kilohertz. The carrier will then shift at the same rate as before. This fulfills all requirements for frequency modulation. Both the frequency and the amplitude of the modulating signal are translated into variations in the frequency of the rf carrier.

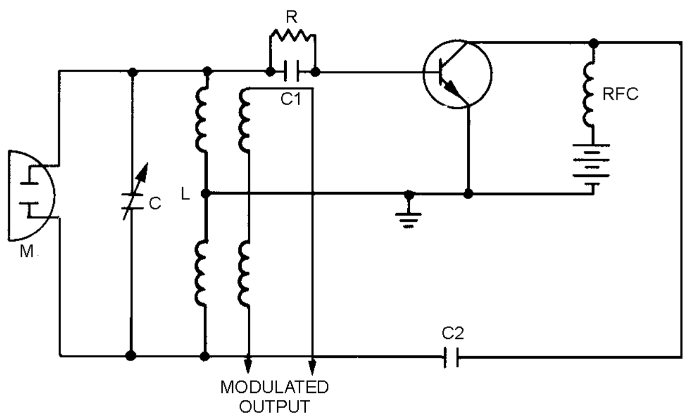


Figure 2-6.—Oscillator circuit illustrating frequency modulation.

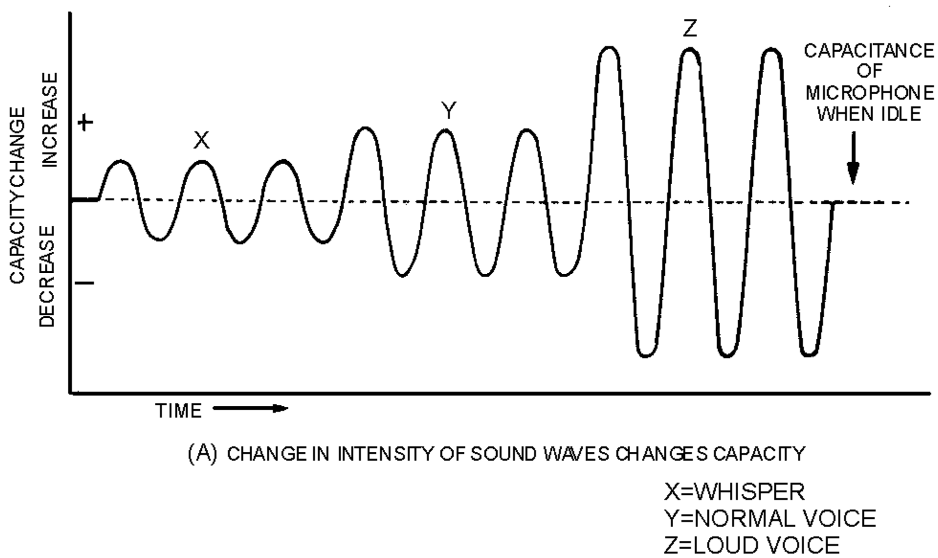


Figure 2-7A.—Capacitance change in an oscillator circuit during modulation. CHANGE IN INTENSITY OF SOUND WAVES CHANGES CAPACITY.

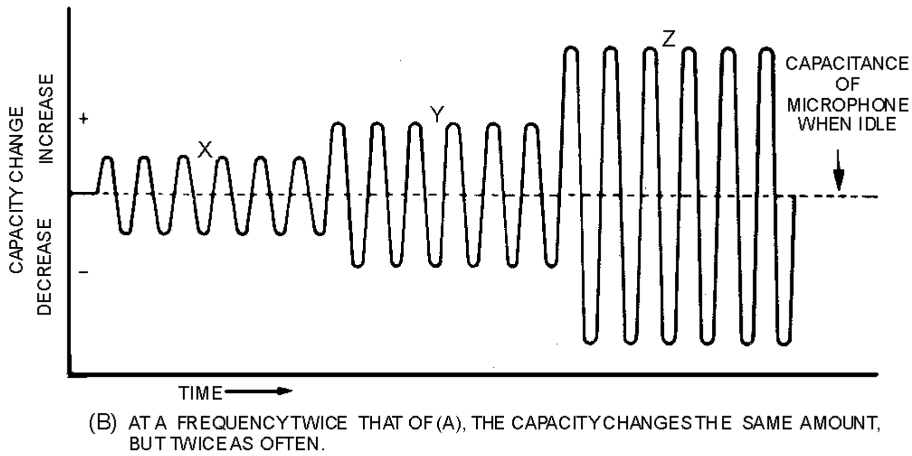


Figure 2-7B.—Capacitance change in an oscillator circuit during modulation. AT A FREQUENCY TWICE THAT OF (A), THE CAPACITY CHANGES THE SAME AMOUNT, BUT TWICE AS OFTEN.

Figure 2-8 shows how the frequency shift of an fm signal goes through the same variations as does the modulating signal. In this figure the dimension of the constant amplitude is omitted. (As these remaining waveforms are presented, be sure you take plenty of time to study and digest what the figures tell you. Look each one over carefully, noting everything you can about them. Doing this will help you understand this material.) If the maximum frequency deviation is set at 75 kilohertz above and below the carrier, the audio amplitude of the modulating wave must be so adjusted that its peaks drive the frequency only between these limits. This can then be referred to as 100-PERCENT MODULATION, although the term is only remotely applicable to fm. Projections along the vertical axis represent deviations in frequency from the resting frequency (carrier) in terms of audio amplitude. Projections along the horizontal axis represent time. The distance between **A** and **B** represents 0.001 second. This means that carrier deviations from the resting frequency to plus 75 kilohertz, then to minus 75 kilohertz, and finally back to rest would occur 1,000 times per second. This would equate to an audio frequency of 1,000 hertz. Since the carrier deviation for this period (**A** to **B**) extends to the full allowable limits of plus and minus 75 kilohertz, the wave is fully modulated. The distance from **C** to **D** is the same as that from **A** to **B**, so the time interval and frequency are the same as before. Notice, however, that the amplitude of the modulating wave has been decreased so that the carrier is driven to only plus and minus 37.5 kilohertz, one-half the allowable deviation. This would correspond to only 50-percent modulation if the system were AM instead of fm. Between **E** and **F**, the interval is reduced to 0.0005 second. This indicates an increase in frequency of the modulating signal to 2,000 hertz. The amplitude has returned to its maximum allowable value, as indicated by the deviation of the carrier to plus and minus 75 kilohertz. Interval **G** to **H** represents the same frequency at a lower modulation amplitude (66 percent). Notice the GUARD BANDS between plus and minus 75 kilohertz and plus and minus 100 kilohertz. These bands isolate the modulation extremes of this particular channel from that of adjacent channels.

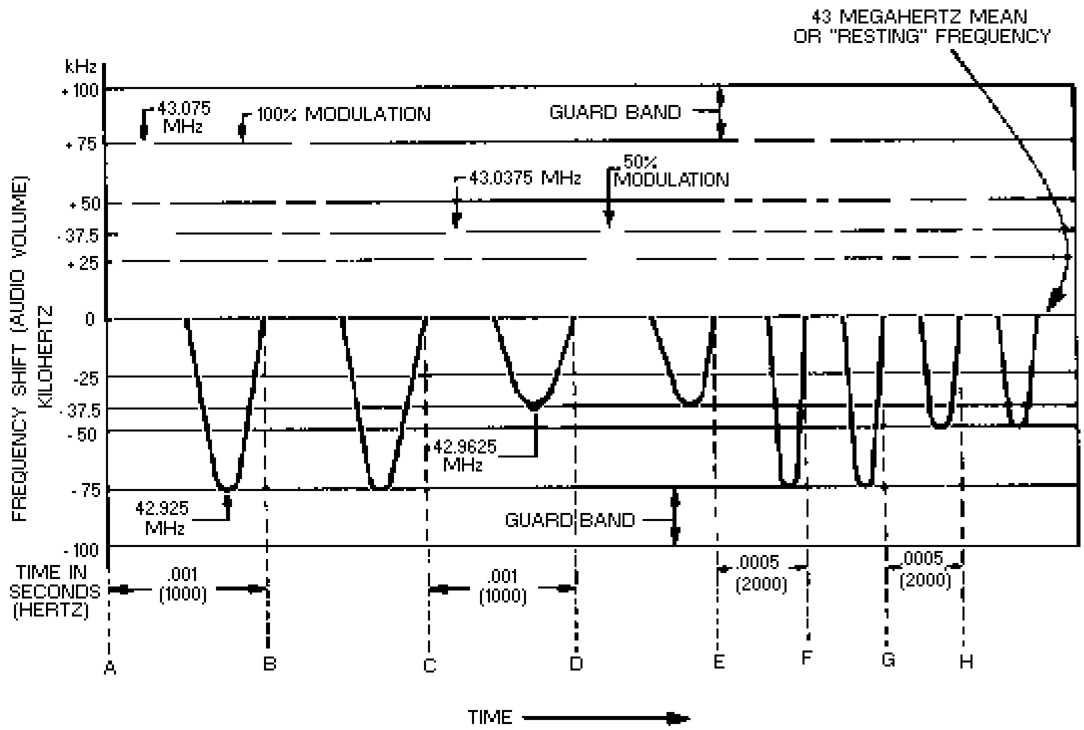


Figure 2-8.—Frequency-modulating signal.

PERCENT OF MODULATION.—Before we explain 100-percent modulation in an fm system, let's review the conditions for 100-percent modulation of an AM wave. Recall that 100-percent modulation for AM exists when the amplitude of the modulation envelope varies between 0 volts and twice its normal unmodulated value. At 100-percent modulation there is a power increase of 50 percent. Because the modulating wave is not constant in voice signals, the degree of modulation constantly varies. In this case the vacuum tubes in an AM system cannot be operated at maximum efficiency because of varying power requirements.

In frequency modulation, 100-percent modulation has a meaning different from that of AM. The modulating signal varies only the frequency of the carrier. Therefore, tubes do not have varying power requirements and can be operated at maximum efficiency and the fm signal has a constant power output. In fm a modulation of 100 percent simply means that the carrier is deviated in frequency by the full permissible amount. For example, an 88.5-megahertz fm station operates at 100-percent modulation when the modulating signal deviation frequency band is from 75 kilohertz above to 75 kilohertz below the carrier (the maximum allowable limits). This maximum deviation frequency is set arbitrarily and will vary according to the applications of a given fm transmitter. In the case given above, 50-percent modulation would mean that the carrier was deviated 37.5 kilohertz above and below the resting frequency (50 percent of the 150-kilohertz band divided by 2). Other assignments for fm service may limit the allowable deviation to 50 kilohertz, or even 10 kilohertz. Since there is no fixed value for comparison, the term "percent of modulation" has little meaning for fm. The term MODULATION INDEX is more useful in fm modulation discussions. Modulation index is frequency deviation divided by the frequency of the modulating signal.

MODULATION INDEX.—This ratio of frequency deviation to frequency of the modulating signal is useful because it also describes the ratio of amplitude to tone for the audio signal. These factors determine the number and spacing of the side frequencies of the transmitted signal. The modulation index formula is shown below:

$$\text{modulation index} = \frac{\Delta f}{f_m}$$

Where:

Δf = frequency deviation
f_m = modulating frequency

Views (A) and (B) of figure 2-9 show the frequency spectrum for various fm signals. In the four examples of view (A), the modulating frequency is constant; the deviation frequency is changed to show the effects of modulation indexes of 0.5, 1.0, 5.0, and 10.0. In view (B) the deviation frequency is held constant and the modulating frequency is varied to give the same modulation indexes.

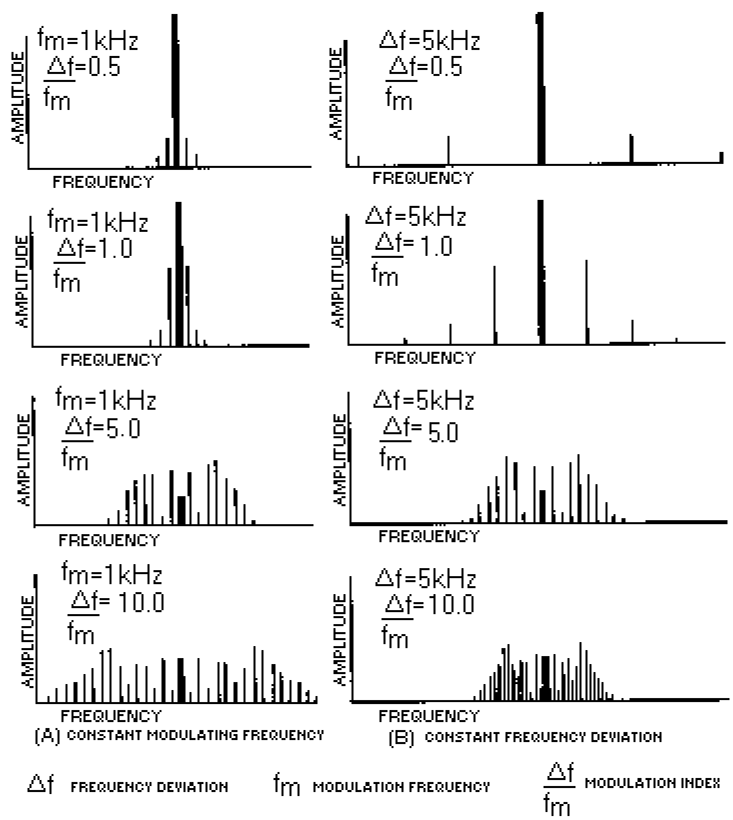


Figure 2-9.—Frequency spectra of fm waves under various conditions.

You can determine several facts about fm signals by studying the frequency spectrum. For example, table 2-1 was developed from the information in figure 2-9. Notice in the top spectrums of both views (A) and (B) that the modulation index is 0.5. Also notice as you look at the next lower spectrums that the modulation index is 1.0. Next down is 5.0, and finally, the bottom spectrums have modulation indexes of

10.0. This information was used to develop table 2-1 by listing the modulation indexes in the left column and the number of significant sidebands in the right. SIGNIFICANT SIDEBANDS (those with significantly large amplitudes) are shown in both views of figure 2-9 as vertical lines on each side of the carrier frequency. Actually, an infinite number of sidebands are produced, but only a small portion of them are of sufficient amplitude to be important. For example, for a modulation index of 0.5 [top spectrums of both views (A) and (B)], the number of significant sidebands counted is 4. For the next spectrums down, the modulation index is 1.0 and the number of sidebands is 6, and so forth. This holds true for any combination of deviating and modulating frequencies that yield identical modulating indexes.

Table 2-1.—Modulation index table

MODULATION INDEX	SIGNIFICANT SIDEBANDS
.01	2
.4	2
.5	4
1.0	6
2.0	8
3.0	12
4.0	14
5.0	16
6.0	18
7.0	22
8.0	24
9.0	26
10.0	28
11.0	32
12.0	32
13.0	36
14.0	38
15.0	38

You should be able to see by studying figure 2-9, views (A) and (B), that the modulating frequency determines the spacing of the sideband frequencies. By using a significant sidebands table (such as table 2-1), you can determine the bandwidth of a given fm signal. Figure 2-10 illustrates the use of this table. The carrier frequency shown is 500 kilohertz. The modulating frequency is 15 kilohertz and the deviation frequency is 75 kilohertz.

$$\Delta f = 75kHz$$

$$f_m = 15kHz$$

$$MI = \frac{\Delta f}{f_m}$$

$$MI = \frac{75kHz}{15kHz}$$

$$MI = 5$$

77

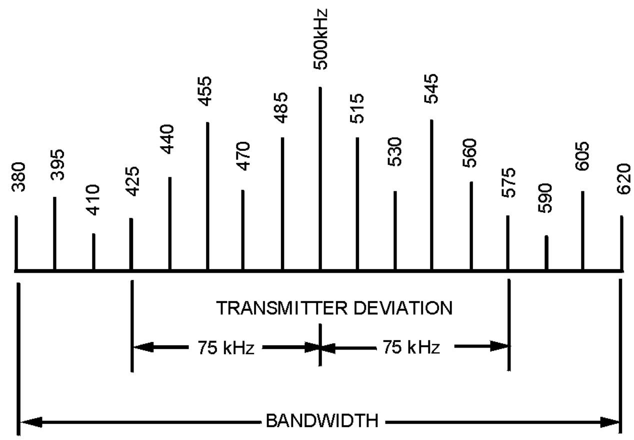


Figure 2-10.—Frequency deviation versus bandwidth.

From table 2-1 we see that there are 16 significant sidebands for a modulation index of 5. To determine total bandwidth for this case, we use:

$$bw = \text{modulating frequency} \times \text{no. of significant sidebands}$$

$$bw = 15kHz \times 16$$

$$bw = 240kHz$$

The use of this math is to illustrate that the actual bandwidth of an fm transmitter (240 kHz) is greater than that suggested by its maximum deviation bandwidth (±75 kHz or 150 kHz). This is important to know when choosing operating frequencies or designing equipment.

Q-4. What characteristic of a carrier wave is varied in frequency modulation?

Q-5. How is the degree of modulation expressed in an fm system?

Q-6. What two values may be used to determine the bandwidth of an fm wave?

METHODS OF FREQUENCY MODULATION.—The circuit shown earlier in figure 2-6 and the discussion in previous paragraphs were for illustrative purposes only. In reality, such a circuit would not be practical. However, the basic principle involved (the change in reactance of an oscillator circuit in accordance with the modulating voltage) constitutes one of the methods of developing a frequency-modulated wave.

Reactance-Tube Modulation.—In direct modulation, an oscillator is frequency modulated by a REACTANCE TUBE that is in parallel (SHUNT) with the oscillator tank circuit. (The terms "shunt" or "shunting" will be used in this module to mean the same as "parallel" or "to place in parallel with" components.) This is illustrated in figure 2-11. The oscillator is a conventional Hartley circuit with the reactance-tube circuit in parallel with the tank circuit of the oscillator tube. The reactance tube is an ordinary pentode. It is made to act either capacitively or inductively; that is, its grid is excited with a voltage which either leads or lags the oscillator voltage by 90 degrees.

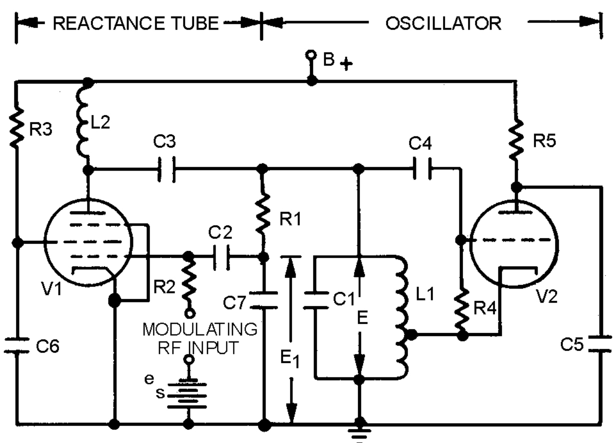


Figure 2-11.—Reactance-tube fm modulator.

When the reactance tube is connected across the tank circuit with no modulating voltage applied, it will affect the frequency of the oscillator. The voltage across the oscillator tank circuit (L1 and C1) is also in parallel with the series network of R1 and C7. This voltage causes a current flow through R1 and C7. If R1 is at least five times larger than the capacitive reactance of C7, this branch of the circuit will be essentially resistive. Voltage E_1, which is across C7, will lag current by 90 degrees. E_1 is applied to the control grid of reactance tube V1. This changes plate current (I_p), which essentially flows only through the LC tank circuit. This is because the value of R1 is high compared to the impedance of the tank circuit. Since current is inversely proportional to impedance, most of the plate current coupled through C3 flows through the tank circuit.

At resonance, the voltage and current in the tank circuit are in phase. Because E_1 lags E by 90 degrees and I is in phase with grid voltage E_1, the superimposed current through the tank circuit lags the original tank current by 90 degrees. Both the resultant current (caused by I_p) and the tank current lag tank voltage and current by some angle depending on the relative amplitudes of the two currents. Because this resultant current is a lagging current, the impedance across the tank circuit cannot be at its maximum unless something happens within the tank to bring current and voltage into phase. Therefore, this situation continues until the frequency of oscillations in the tank circuit changes sufficiently so that the voltages across the tank and the current flowing into it are again in phase. This action is the same as would be produced by adding a reactance in parallel with the L1C1 tank. Because the superimposed current lags voltage E by 90 degrees, the introduced reactance is inductive. In *NEETS*, Module 2, *Introduction to*

Alternating Current and Transformers, you learned that total inductance decreases as additional inductors are added in parallel. Because this introduced reactance effectively reduces inductance, the frequency of the oscillator increases to a new fixed value.

Now let's see what happens when a modulating signal is applied. The magnitude of the introduced reactance is determined by the magnitude of the superimposed current through the tank. The magnitude of I_p for a given E_1 is determined by the transconductance of V1. (Transconductance was covered in *NEETS*, Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies.*) Therefore, the value of reactance introduced into the tuned circuit varies directly with the transconductance of the reactance tube. When a modulating signal is applied to the grid of V1, both E_1 and I change, causing transconductance to vary with the modulating signal. This causes a variable reactance to be introduced into the tuned circuit. This variable reactance either adds to or subtracts from the fixed value of reactance that is introduced in the absence of the modulating signal. This action varies the reactance across the oscillator which, in turn, varies the instantaneous frequency of the oscillator. These variations in the oscillator frequency are proportional to the instantaneous amplitude of the modulating voltage. Reactance-tube modulators are usually operated at low power levels. The required output power is developed in power amplifier stages that follow the modulators.

The output of a reactance-tube modulated oscillator also contains some unwanted amplitude modulation. This unwanted modulation is caused by stray capacitance and the resistive component of the RC phase splitter. The resistance is much less significant than the desired X_C, but the resistance does allow some plate current to flow which is not of the proper phase relationship for good tube operation. The small amplitude modulation that this produces is easily removed by passing the oscillator output through a limiter-amplifier circuit.

Semiconductor Reactance Modulator.—The SEMICONDUCTOR-REACTANCE MODULATOR is used to frequency modulate low-power semiconductor transmitters. Figure 2-12 shows a typical frequency-modulated oscillator stage operated as a reactance modulator. Q1, along with its associated circuitry, is the oscillator. Q2 is the modulator and is connected to the circuit so that its collector-to-emitter capacitance (C_{CE}) is in parallel with a portion of the rf oscillator coil, L1. As the modulator operates, the output capacitance of Q2 is varied. Thus, the frequency of the oscillator is shifted in accordance with the modulation the same as if C1 were varied.

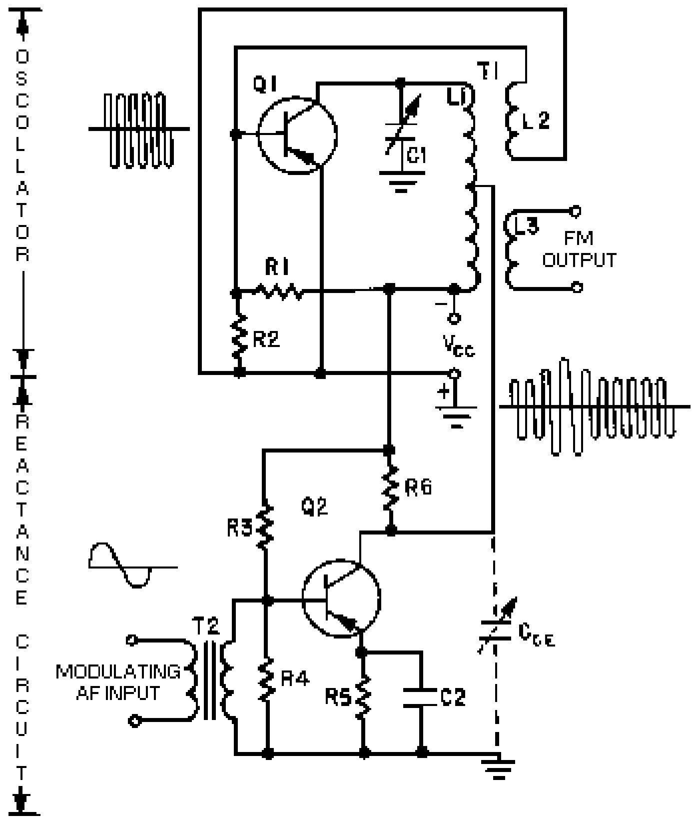


Figure 2-12.—Reactance-semiconductor fm modulator.

When the modulating signal is applied to the base of Q2, the emitter-to-base bias varies at the modulation rate. This causes the collector voltage of Q2 to vary at the same modulating rate. When the collector voltage increases, output capacitance C_{CE} decreases; when the collector voltage decreases, C_{CE} increases. An *increase* in collector voltage has the effect of spreading the plates of C_{CE} farther apart by increasing the width of the barrier. A *decrease* of collector voltage reduces the width of the pn junction and has the same effect as pushing the capacitor plates together to provide more capacitance.

When the output capacitance *decreases*, the instantaneous frequency of the oscillator tank circuit increases (acts the same as if C1 were decreased). When the output capacitance *increases*, the instantaneous frequency of the oscillator tank circuit decreases. This decrease in frequency produces a lower frequency in the output because of the shunting effect of C_{CE}. Thus, the frequency of the oscillator tank circuit increases and decreases at an audio frequency (af) modulating rate. The output of the oscillator, therefore, is a frequency modulated rf signal.

Since the audio modulation causes the collector voltage to increase and decrease, an AM component is induced into the output. This produces both an fm and AM output. The amplitude variations are then removed by placing a limiter stage after the reactance modulator and only the frequency modulation remains.

Frequency multipliers or mixers (discussed in chapter 1) are used to increase the oscillator frequency to the desired output frequency. For high-power applications, linear rf amplifiers are used to increase the steady-amplitude signal to a higher power output. With the initial modulation occurring at low levels, fm represents a savings of power when compared to conventional AM. This is because fm noise-reducing properties provide a better signal-to-noise ratio than is possible with AM.

Multivibrator Modulator.—Another type of frequency modulator is the astable multivibrator illustrated in figure 2-13. Inserting the modulating af voltage in series with the base-return of the multivibrator transistors causes the gate length, and thus the fundamental frequency of the multivibrator, to vary. The amount of variation will be in accordance with the amplitude of the modulating voltage. One requirement of this method is that the fundamental frequency of the multivibrator be high in relation to the highest modulating frequencies. A factor of at least 100 provides the best results.

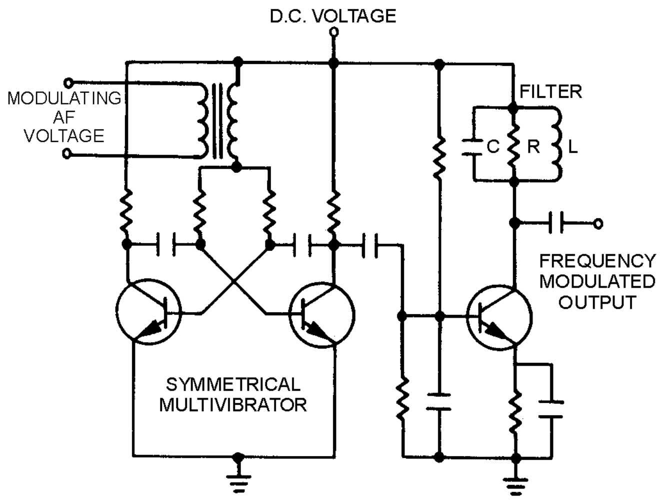


Figure 2-13.—Astable multivibrator and filter circuit for generating an fm carrier.

Recall that a multivibrator output consists of the fundamental frequency and all of its harmonics. Unwanted even harmonics are eliminated by using a SYMMETRICAL MULTIVIBRATOR circuit, as shown in figure 2-13. The desired fundamental frequency, or desired odd harmonics, can be amplified after all other odd harmonics are eliminated in the LCR filter section of figure 2-13. A single frequency-modulated carrier is then made available for further amplification and transmission.

Proper design of the multivibrator will cause the frequency deviation of the carrier to faithfully follow (referred to as a "linear" function) the modulating voltage. This is true up to frequency deviations which are considerable fractions of the fundamental frequency of the multivibrator. The principal design consideration is that the RC coupling from one multivibrator transistor base to the collector of the other has a time constant which is greater than the actual gate length by a factor of 10 or more. Under these conditions, a rise in base voltage in each transistor is essentially linear from cutoff to the bias at which the transistor is switched on. Since this rise in base voltage is a linear function of time, the gate length will change as an inverse function of the modulating voltage. This action will cause the frequency to change as a linear function of the modulating voltage.

The multivibrator frequency modulator has the advantage over the reactance-type modulator of a greater linear frequency deviation from a given carrier frequency. However, multivibrators are limited to frequencies below about 1 megahertz. Both systems are subject to drift of the carrier frequency and must, therefore, be stabilized. Stabilization may be accomplished by modulating at a relatively low frequency and translating by heterodyne action to the desired output frequency, as shown in figure 2-14. A 1-megahertz signal is heterodyned with 49 megahertz from the crystal-controlled oscillator to provide a stable 50-megahertz output from the mixer. If a suitably stable heterodyning oscillator is used, the frequency stability can be greatly improved. For instance, at the frequencies shown in figure 2-14, the stability of the unmodulated 50-megahertz carrier would be 50 times better than that which harmonic multiplication could provide.

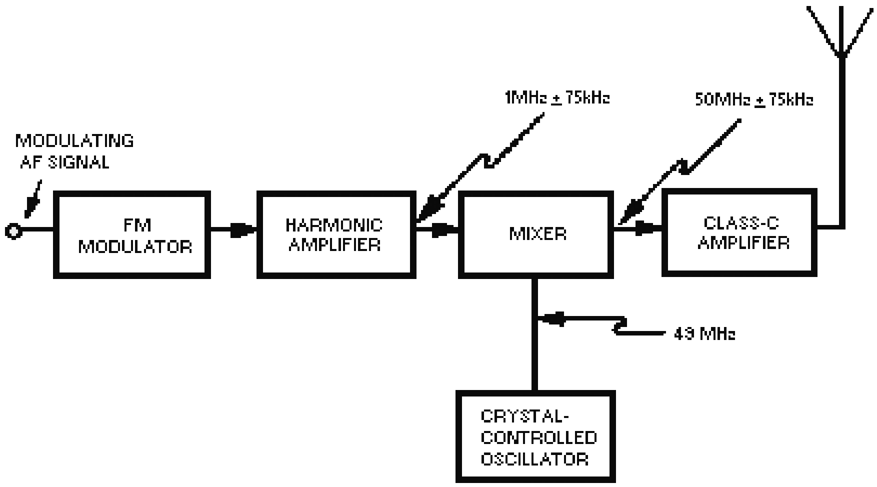


Figure 2-14.—Method for improving frequency stability of fm system.

Varactor FM Modulator.—Another fm modulator which is widely used in transistorized circuitry uses a voltage-variable capacitor (VARACTOR). The varactor is simply a diode, or pn junction, that is designed to have a certain amount of capacitance between junctions. View (A) of figure 2-15 shows the varactor schematic symbol. A diagram of a varactor in a simple oscillator circuit is shown in view (B). This is not a working circuit, but merely a simplified illustration. The capacitance of a varactor, as with regular capacitors, is determined by the area of the capacitor plates and the distance between the plates. The depletion region in the varactor is the dielectric and is located between the p and n elements, which serve as the plates. Capacitance is varied in the varactor by varying the reverse bias which controls the thickness of the depletion region. The varactor is so designed that the change in capacitance is linear with the change in the applied voltage. This is a special design characteristic of the varactor diode. The varactor must not be forward biased because it cannot tolerate much current flow. Proper circuit design prevents the application of forward bias.

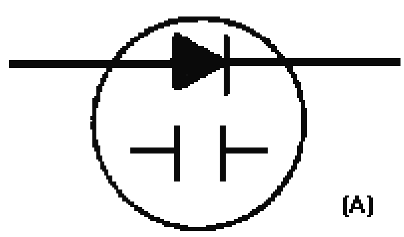


Figure 2-15A.—Varactor symbol and schematic. SCHEMATIC SYMBOL.

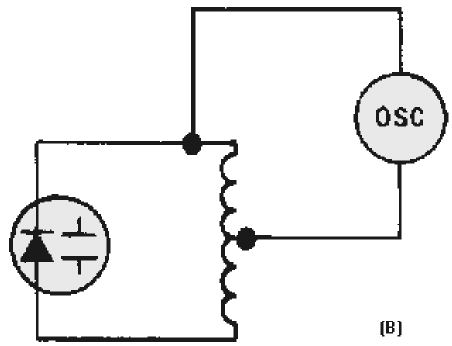


Figure 2-15B.—Varactor symbol and schematic. SIMPLIFIED CIRCUIT.

Notice the simplicity of operation of the circuit in figure 2-16. An af signal that is applied to the input results in the following actions: (1) On the positive alternation, reverse bias increases and the dielectric (depletion region) width increases. This decreases capacitance which increases the frequency of the oscillator. (2) On the negative alternation, the reverse bias decreases, which results in a decrease in oscillator frequency.

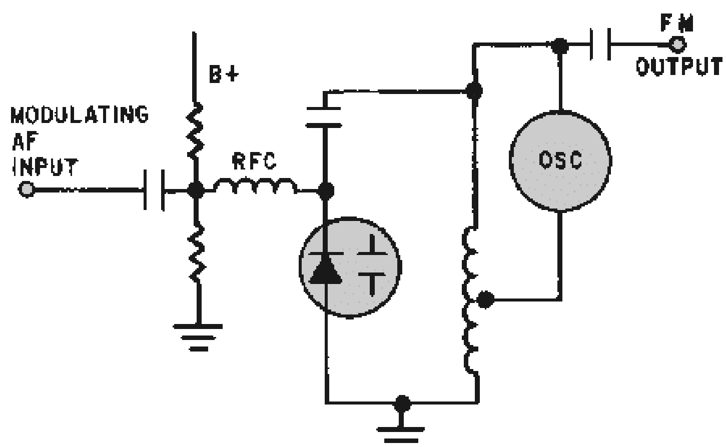


Figure 2-16.—Varactor fm modulator.

Many different fm modulators are available, but they all use the basic principles you have just studied. The main point to remember is that an oscillator must be used to establish the reference (carrier) frequency. Secondly, some method is needed to cause the oscillator to change frequency in accordance with an af signal. Anytime this can be accomplished, we have a frequency modulator.

Q-7. How does the reactance-tube modulator impress intelligence onto an rf carrier?

Q-8. What characteristic of a transistor is varied in a semiconductor-reactance modulator?

Q-9. What circuit section is required in the output of a multivibrator modulator to eliminate unwanted output frequencies?

Q-10. What characteristic of a varactor is used in an fm modulator?

PHASE MODULATION

Frequency modulation requires the oscillator frequency to deviate both above and below the carrier frequency. During the process of frequency modulation, the peaks of each successive cycle in the modulated waveform occur at times other than they would if the carrier were unmodulated. This is actually an incidental phase shift that takes place along with the frequency shift in fm. Just the opposite action takes place in phase modulation. The af signal is applied to a PHASE MODULATOR in pm. The resultant wave from the phase modulator shifts in phase, as illustrated in figure 2-17. Notice that the time period of each successive cycle varies in the modulated wave according to the audio-wave variation. Since frequency is a function of time period per cycle, we can see that such a phase shift in the carrier will cause its frequency to change. The frequency change in fm is vital, but in pm it is merely incidental. The amount of frequency change has nothing to do with the resultant modulated wave shape in pm. At this point the comparison of fm to pm may seem a little hazy, but it will clear up as we progress.

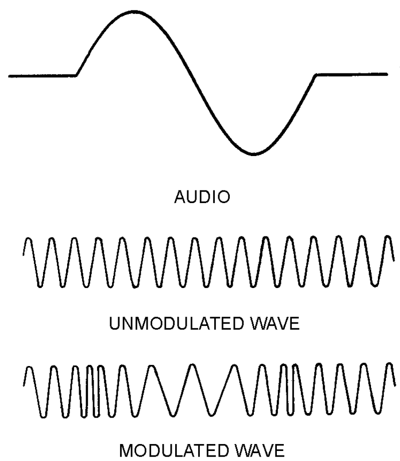


Figure 2-17.—Phase modulation.

Let's review some voltage phase relationships. Look at figure 2-18 and compare the three voltages (**A**, **B**, and **C**). Since voltage **A** begins its cycle and reaches its peak before voltage **B**, it is said to *lead* voltage **B**. Voltage **C**, on the other hand, *lags* voltage **B** by 30 degrees. In phase modulation the phase of the carrier is caused to shift at the rate of the af modulating signal. In figure 2-19, note that the unmodulated carrier has constant phase, amplitude, and frequency. The dotted wave shape represents the modulated carrier. Notice that the phase on the second peak leads the phase of the unmodulated carrier. On the third peak the shift is even greater; however, on-the fourth peak, the peaks begin to realign phase with each other. These relationships represent the effect of 1/2 cycle of an af modulating signal. On the negative alternation of the af intelligence, the phase of the carrier would lag and the peaks would occur at times later than they would in the unmodulated carrier.

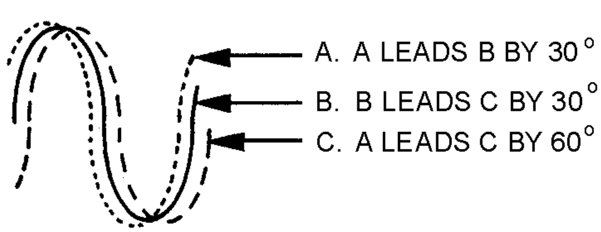


Figure 2-18.—Phase relationships.

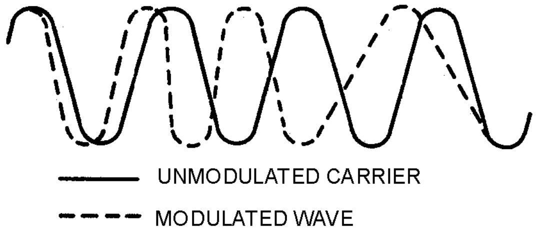


Figure 2-19.—Carrier with and without modulation.

The presentation of these two waves together does not mean that we transmit a modulated wave together with an unmodulated carrier. The two waveforms were drawn together only to show how a modulated wave looks when compared to an unmodulated wave.

Now that you have seen the phase and frequency shifts in both fm and pm, let's find out exactly how they differ. First, only the phase shift is important in pm. It is proportional to the af modulating signal. To visualize this relationship, refer to the wave shapes shown in figure 2-20. Study the composition of the fm and pm waves carefully as they are modulated with the modulating wave shape. Notice that in fm, the carrier frequency deviates when the modulating wave changes polarity. With each alternation of the modulating wave, the carrier advances or retards in frequency and remains at the new frequency for the duration of that cycle. In pm you can see that between one alternation and the next, the carrier phase must change, and the frequency shift that occurs does so *only* during the transition time; the frequency then returns to its normal rate. Note in the pm wave that the frequency shift occurs only when the modulating wave is changing polarity. The frequency during the constant amplitude portion of each alternation is the REST FREQUENCY.

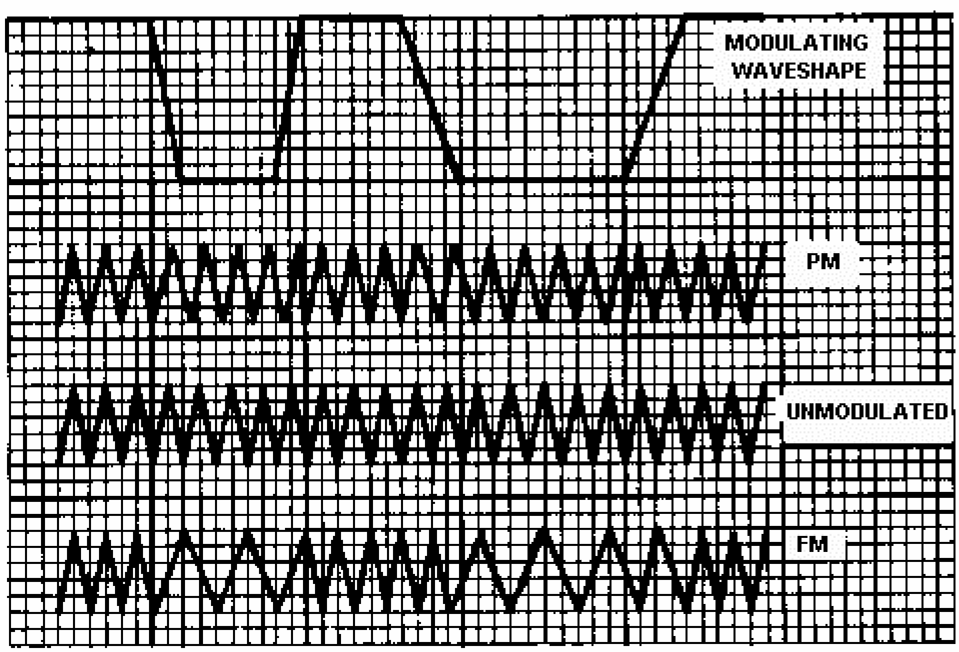


Figure 2-20.—Pm versus fm.

The relationship, in pm, of the modulating af to the change in the phase shift is easy to see once you understand AM and fm principles. Again, we can establish two clear-cut rules of phase modulation:

- AMOUNT OF PHASE SHIFT IS PROPORTIONAL TO THE AMPLITUDE OF THE MODULATING SIGNAL.

(If a 10-volt signal causes a phase shift of 20 degrees, then a 20-volt signal causes a phase shift of 40 degrees.)

- RATE OF PHASE SHIFT IS PROPORTIONAL TO THE FREQUENCY OF THE MODULATING SIGNAL.

(If the carrier were modulated with a 1-kilohertz tone, the carrier would advance and retard in phase 1,000 times each second.)

Phase modulation is also similar to frequency modulation in the number of sidebands that exist within the modulated wave and the spacing between sidebands. Phase modulation will also produce an infinite number of sideband frequencies. The spacing between these sidebands will be equal to the frequency of the modulating signal. However, one factor is very different in phase modulation; that is, the distribution of power in pm sidebands is not similar to that in fm sidebands, as will be explained in the next section.

Modulation Index

Recall from frequency modulation that modulation index is used to calculate the number of significant sidebands existing in the waveform. The higher the modulation index, the greater the number of sideband pairs. The modulation index is the ratio between the amount of oscillator deviation and the frequency of the modulating signal:

$$MI = \frac{\text{transmitter deviation}}{\text{modulating frequency}}$$

In frequency modulation, we saw that as the frequency of the modulating signal increased (assuming the deviation remained constant) the number of significant sideband pairs decreased. This is shown in views (A) and (B) of figure 2-21. Notice that although the total number of significant sidebands decreases with a higher frequency-modulating signal, the sidebands spread out relative to each other; the total bandwidth increases.

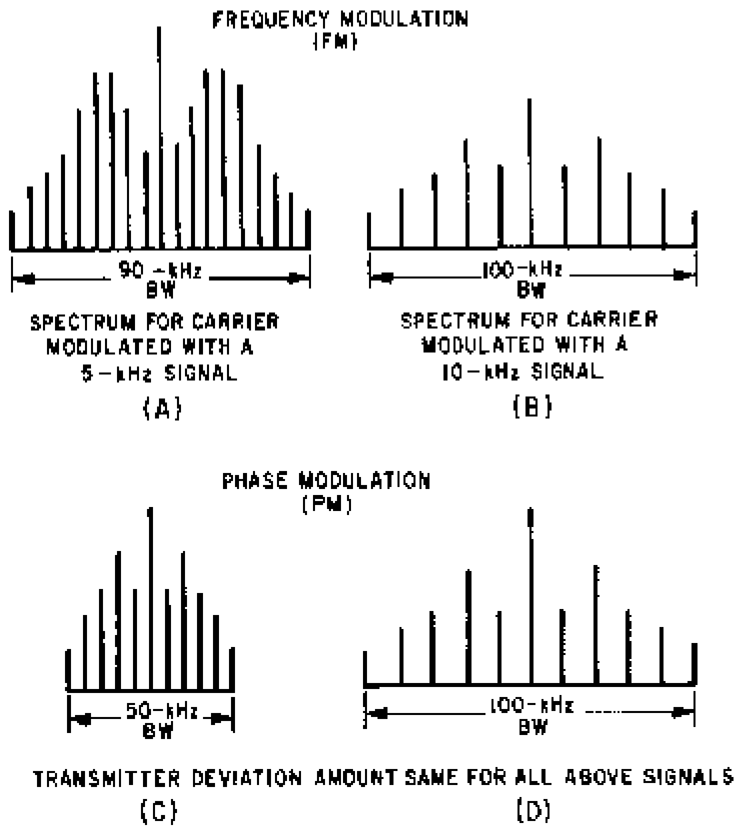


Figure 2-21.—Fm versus pm spectrum distribution.

In phase modulation the oscillator does not deviate, and the power in the sidebands is a function of the amplitude of the modulating signal. Therefore, two signals, one at 5 kilohertz and the other at 10 kilohertz, used to modulate a carrier would have the same sideband power distribution. However, the 10-kilohertz sidebands would be farther apart, as shown in views (C) and (D) of figure 2-21. When compared to fm, the bandwidth of the pm transmitted signal is greatly increased as the frequency of the modulating signal is increased.

As we pointed out earlier, phase modulation cannot occur without an incidental change in frequency, nor can frequency modulation occur without an incidental change in phase. The term fm is loosely used when referring to any type of angle modulation, and phase modulation is sometimes incorrectly referred to as "indirect fm." This is a definition that you should disregard to avoid confusion. Phase modulation is just what the words imply — phase modulation of a carrier by an af modulating signal. You will develop a better understanding of these points as you advance in your study of modulation.

Basic Modulator

In phase modulation you learned that varying the phase of a carrier at an intelligence rate caused that carrier to contain variations which could be converted back into intelligence. One circuit that can cause this phase variation is shown in figure 2-22.

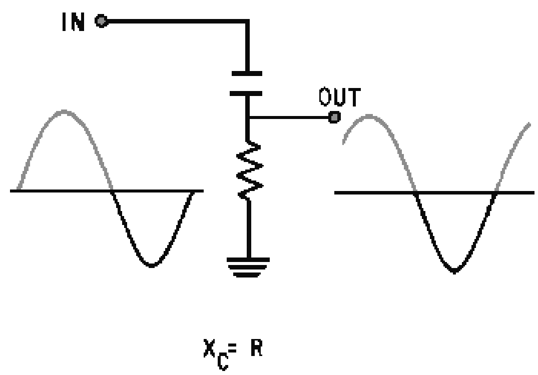


Figure 2-22.—Phase shifting a sine wave.

The capacitor in series with the resistor forms a phase-shift circuit. With a constant frequency rf carrier applied at the input, the output across the resistor would be 45 degrees out of phase with the input if $X_C = R$.

Now, let's vary the resistance and observe how the output is affected in figure 2-23. As the resistance reaches a value greater than 10 times X_C, the phase difference between input and output is nearly 0 degrees. For all practical purposes, the circuit is resistive. As the resistance is decreased to 1/10 the value of X_C, the phase difference approaches 90 degrees. The circuit is now almost completely capacitive. By replacing the resistor with a vacuum tube, as shown in view (A) of figure 2-24, we can vary the resistance (vacuum-tube impedance) by varying the voltage applied to the grid of the tube. The frequency applied to the circuit (from a crystal-controlled master oscillator) will be shifted in phase by 45 degrees with no audio input [view (B)]. With the application of an audio signal, the phase will shift as the impedance of the tube is varied.

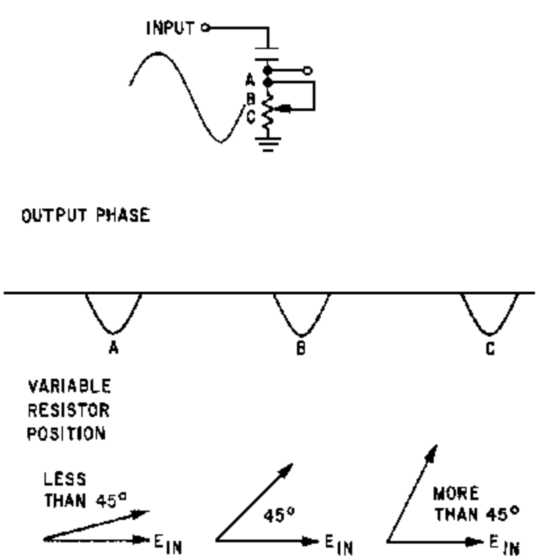


Figure 2-23.—Control over the amount of phase shift.

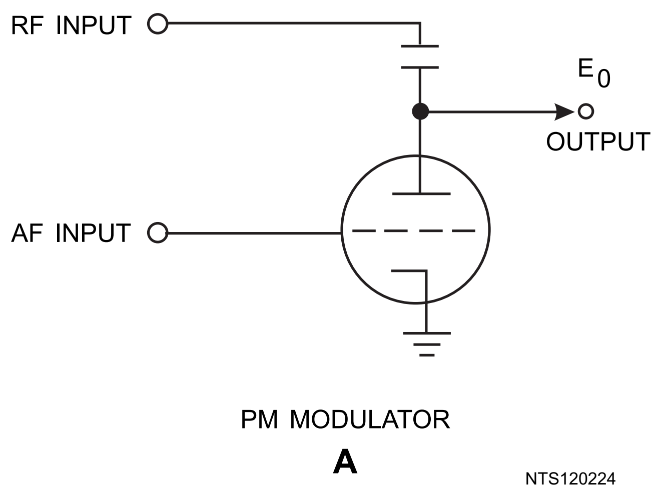


Figure 2-24A.—Phase modulator.

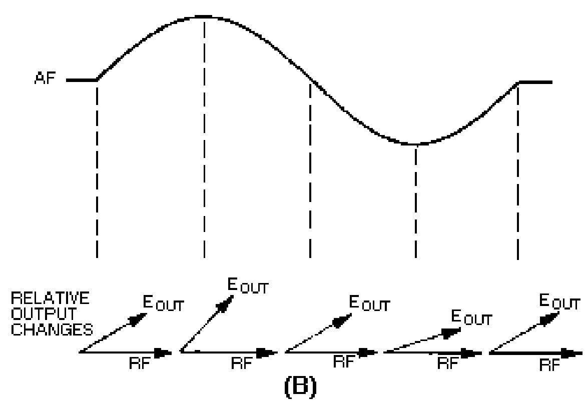


Figure 2-24B.—Phase modulator.

In practice, a circuit like this could not provide enough phase shift to produce the desired results in the output. Several of these circuits are arranged in cascade to provide the desired amount of phase shift. Also, since the output of this circuit will vary in amplitude, the signal is fed to a limiter to remove amplitude variations.

The major advantage of this type modulation circuit over frequency modulation is that this circuit uses a crystal-controlled oscillator to maintain a stable carrier frequency. In fm the oscillator cannot be crystal controlled because it is actually required to vary in frequency. That means that an fm oscillator will require a complex automatic frequency control (afc) system. An afc system ensures that the oscillator stays on the same carrier frequency and achieves a high degree of stability. The afc circuit will be covered in a later module.

Phase-Shift Keying

Phase-shift keying (psk) is similar to ON-OFF cw keying in AM systems and frequency-shift keying in fm systems. Psk is most useful when the code elements are all of equal length; that is, all marks and spaces, whether message elements or synchronizing signals, occupy identical elements of time. It is not fully suitable for use on start-stop teletypewriter circuits where the stop pulse is 1.42 times longer than the other pulses. Neither is it applicable to those pulsed systems in which the duration or position of the pulses are varied by the modulation frequency. In its simplest form, psk operates on the principle of phase reversal of the carrier. Each time a mark is received, the phase is reversed. No phase reversal takes place when a space is received. In binary systems, marks and spaces are called ONES and ZEROS, respectively, so that a ONE causes a 180-degree phase shift, and a ZERO has no effect on the incoming signal. Figure 2-25 shows the application of phase-shift keying to an unmodulated carrier [view (A)] in the af range. For transmission over other than a conductive path, the wave shown in view (D) must be used as the modulating signal for some other system of modulating an rf carrier.

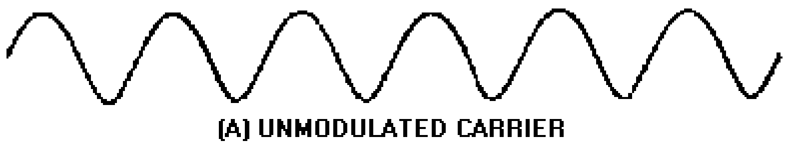


Figure 2-25A.—Phase-shift keying. UNMODULATED CARRIER.

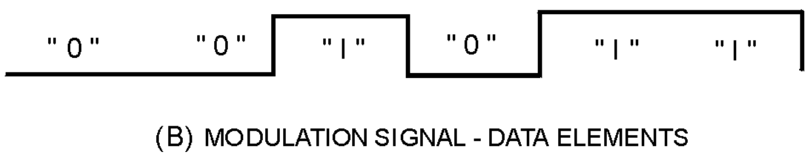


Figure 2-25B.—Phase-shift keying. MODULATION SIGNAL - DATA ELEMENTS.

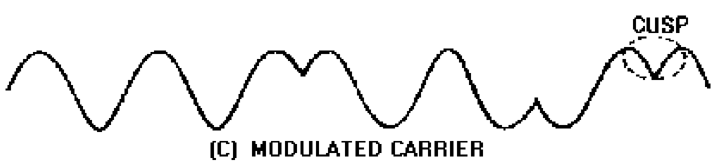


Figure 2-25C.—Phase-shift keying. MODULATED CARRIER.

Figure 2-25D.—Phase-shift keying. MODULATED CARRIER AFTER FILTERING.

The modulating signal in view (B) consists of a bit stream of ZEROS and ONES. A ZERO does not affect the carrier frequency which is usually set to equal the bit rate. For example, a data stream of 1,200 bits per second would have a carrier of 1,200 hertz. When a data bit ONE occurs, the phase of the carrier frequency is shifted 180 degrees. In view (C) we find that the third, fifth, and sixth cycles (all ONE) have been reversed in phase. This phase reversal produces CUSPS (sharp phase reversals) which are usually removed by filtering before transmission or further modulation. This filtering action limits the bandwidth of the output signal frequencies. The resulting wave is shown in view (D).

The exact waveform of figure 2-25, view (D), can be obtained by logic operations of timing and data. This is illustrated in figure 2-26, where a timing signal [view (A)] is used rather than a carrier frequency. The data (intelligence) is shown in view (B) and is combined with the timing signal to produce a combination digital modulation signal, as shown in view (C). The square-wave pattern of the digital modulation is filtered to limit the bandwidth of the signal frequencies, as shown in view (D). This system has been used in some high-speed data equipment, but it offers no particular advantage over other systems of modulation, particularly the pulse-modulated systems for high-speed data transmission.

Q-11. What type of modulation depends on the carrier-wave phase shift?

Q-12. What components may be used to build a basic phase modulator?

Q-13. Phase-shift keying is similar to what other two types of modulation?

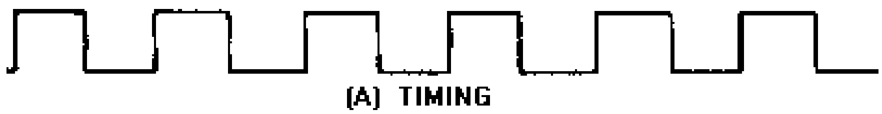


Figure 2-26A.—Simulated phase-shift keying. TIMING.

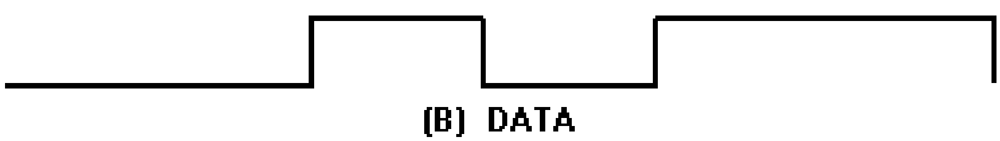


Figure 2-26B.—Simulated phase-shift keying. DATA.

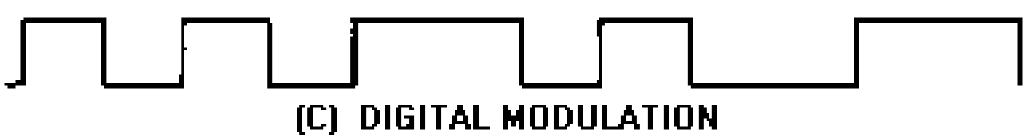


Figure 2-26C.—Simulated phase-shift keying. DIGITAL MODULATION.

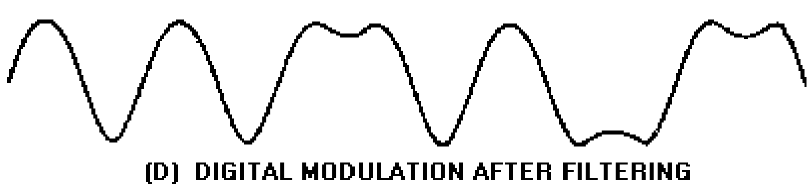


Figure 2-26D.—Simulated phase-shift keying. DIGITAL MODULATION AFTER FILTERING.

PULSE MODULATION

Another type of modulation is PULSE MODULATION. Pulse modulation has many uses, including telegraphy, radar, telemetry, and multiplexing. Far too many applications of pulse modulation exist to elaborate on any one of them, but in this section we will cover the basic principles of pulse modulation.

CHARACTERISTICS

Amplitude modulating a simple rf carrier to a point where it becomes drastically overmodulated could produce a waveform similar to that required in pulse modulation. A modulating signal [view (A) of figure 2-27] that is much larger than the carrier results in the modulation envelope shown in view (B). The modulation envelope would be the same if the modulating wave shape were not sinusoidal; that is, like the one shown in view (C).

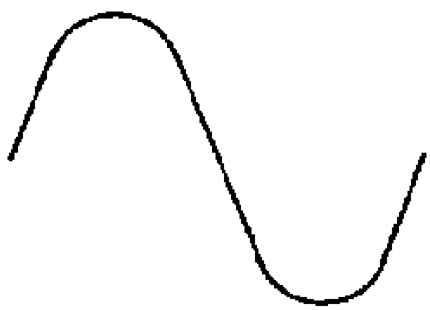


Figure 2-27A.—Overmodulation of a carrier. MODULATING WAVE.

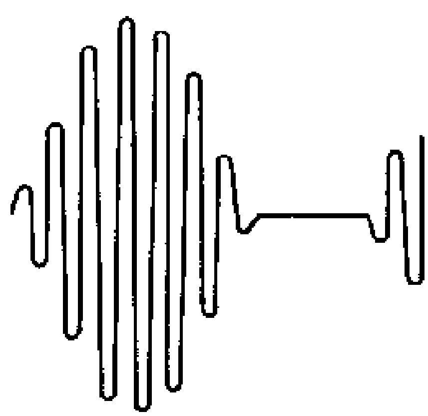


Figure 2-27B.—Overmodulation of a carrier. MODULATION ENVELOPE.

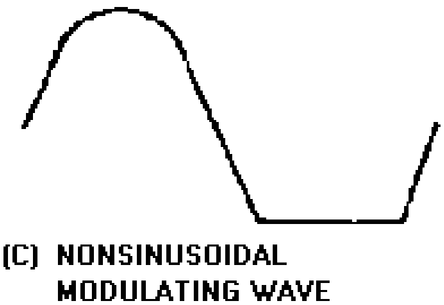


Figure 2-27C.—Overmodulation of a carrier. NONSINUSOIDAL MODULATING WAVE.

Observe the modulating square wave in figure 2-28. Remember that it contains an infinite number of odd harmonics in addition to its fundamental frequency. Assume that a carrier has a frequency of 1 megahertz. The fundamental frequency of the modulating square wave is 1 kilohertz. When these signals heterodyne, two new frequencies will be produced: a sum frequency of 1.001 megahertz and a difference frequency of 0.999 megahertz. The fundamental frequency heterodynes with the carrier. This is also true of all harmonics contained in the square wave. Side frequencies associated with those harmonics will be produced as a result of this process. For example, the third harmonic of the square wave heterodynes with the carrier and produces sideband frequencies at 1.003 and 0.997 megahertz. Another set will be produced by the fifth, seventh, ninth, eleventh, thirteenth, fifteenth, seventeenth, and nineteenth harmonics of the square wave, and so on to infinity.

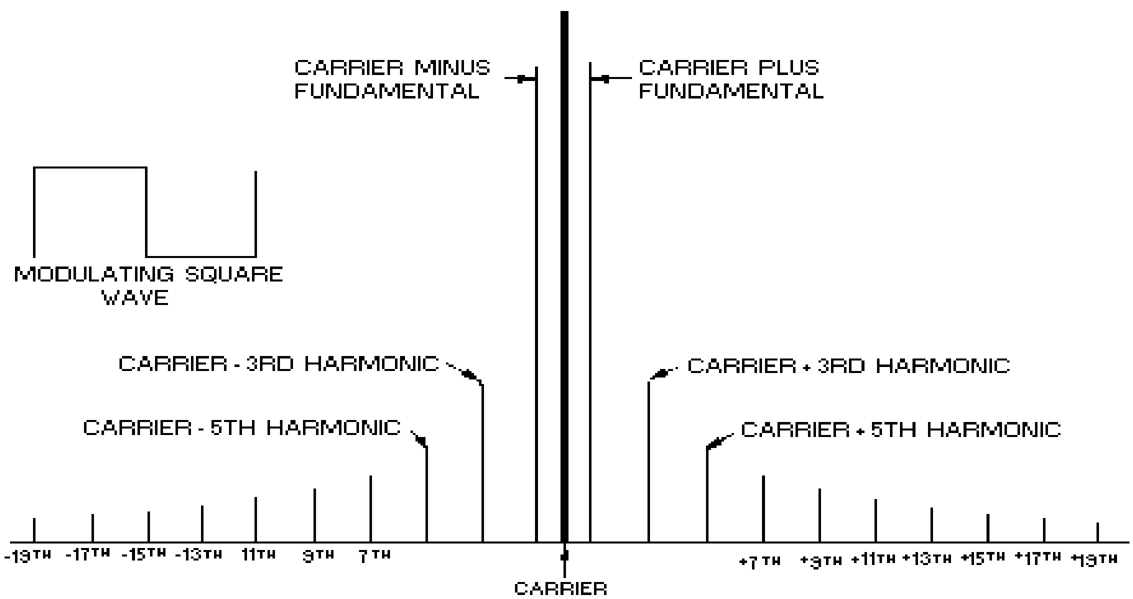


Figure 2-28.—Spectrum distribution when modulating with a square wave.

Look at figure 2-28 and observe the relative amplitudes of the sidebands as they relate to the amplitudes of the harmonics contained in the square wave. Note that the first set of sidebands is directly related to the amplitude of the square wave. The second set of sidebands is related to the third harmonic content of the square wave and is 1/3 the amplitude of the first set. The third set is related to the amplitude of the first set of sidebands and is 115 the amplitude of the first set. This relationship will apply to each additional set of sidebands.

View (A) of figure 2-29 shows the carrier modulated with a square wave. In view (B) the modulating square wave is increased in amplitude; note that the rf peaks increase in amplitude during the positive alternation of the square wave and decrease during the negative half of the square wave. In view (C) the amplitude of the square wave is further increased and the amplitude of the rf wave is almost 0 during the negative alternation of the square wave.

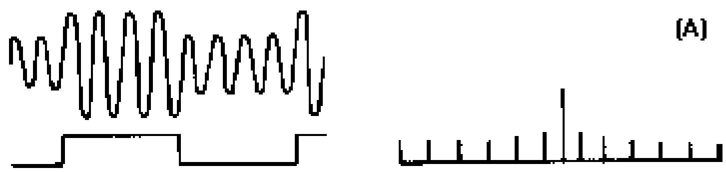


Figure 2-29A.—Various square-wave modulation levels with frequency-spectrum carrier and sidebands.

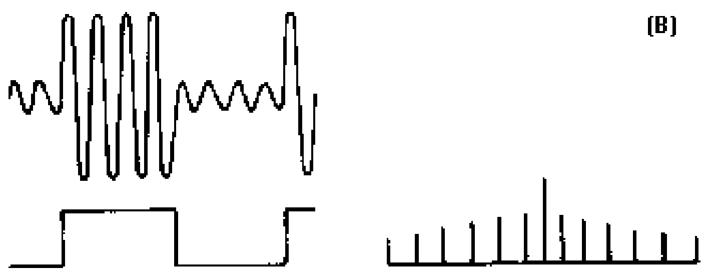


Figure 2-29B.—Various square-wave modulation levels with frequency-spectrum carrier and sidebands.

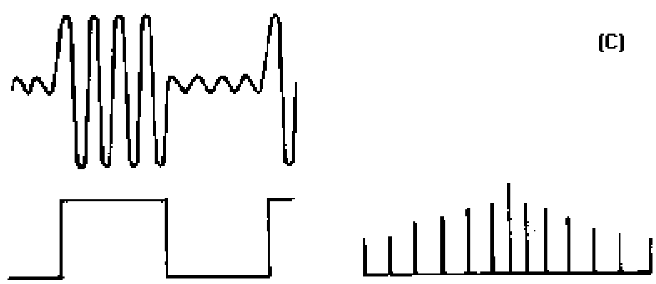


Figure 2-29C.—Various square-wave modulation levels with frequency-spectrum carrier and sidebands.

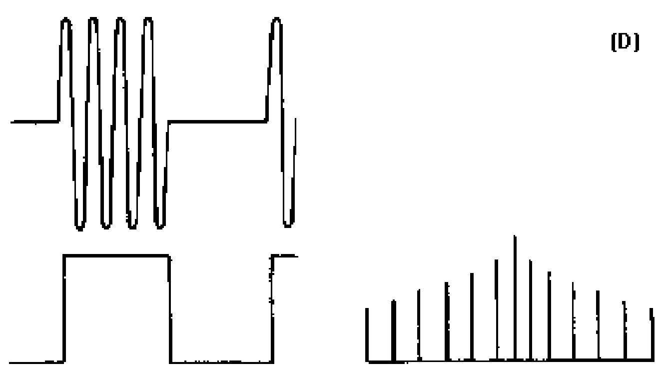


Figure 2-29D.—Various square-wave modulation levels with frequency-spectrum carrier and sidebands.

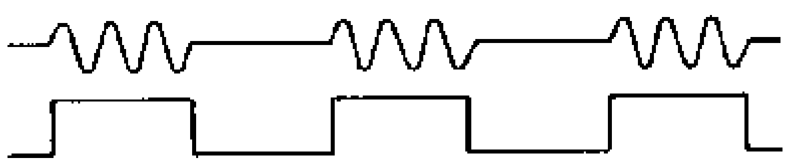
Note the frequency spectrum associated with each of these conditions. The carrier amplitude remains constant, but the sidebands increase in amplitude in accordance with the amplitude of the modulating square wave.

So far in pulse modulation, the same general rules apply as in AM. In view (C) the amplitude of the square wave of voltage is equal to the peak voltage of the unmodulated carrier wave. This is 100-percent modulation, just as in conventional AM. Note in the frequency spectrum that the sideband distribution is also the same as in AM. Keep in mind that the total sideband power is 1/2 of the total power when the modulator signal is a square wave. This is in contrast to 1/3 the total power with sine-wave modulation.

Now refer to view (D). The increase of the square-wave modulating voltage is greater in amplitude than the unmodulated carrier. Notice that the sideband distribution does not change; but, as the sidebands take on more of the transmitted power, so will the carrier.

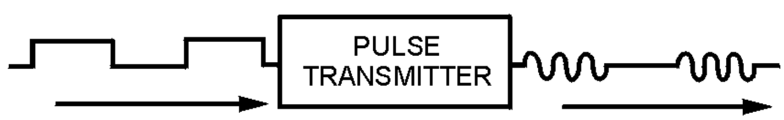
Pulse Timing

Thus far, we have established a carrier and have caused its peaks to increase and decrease as a modulating square wave is applied. Some pulse-modulation systems modulate a carrier in this manner. Others produce no rf until pulsed; that is, rf occurs only during the actual pulse as shown in view (A) of figure 2-30. For example, let's start with an rf carrier frequency of 1 megahertz. Each cycle of the rf requires a certain amount of time to complete. If we allow oscillations to occur for a given period of time only during selected intervals, as in view (B), we are PULSING the system. Note that the pulse transmitter does not produce an rf signal until one of the positive-going modulating pulses is applied. The transmitter then produces the rf carrier until the positive input pulse ends and the input waveform again becomes a negative potential.



(A)

Figure 2-30A.—Pulse transmission.



(B)

Figure 2-30B.—Pulse transmission.

Refer back to figure 1-41 and the over-modulation discussion in chapter 1. You will notice that the overmodulation wave shape of view (D) in figure 2-29 and the pulse-modulation wave shape of figure 2-30, view (B), are very similar to figure 1-41.

Actually, both figure 1-41 and view (D) of figure 2-29 result from overmodulation. Even though the output of the pulse transmitter in figure 2-30 looks like overmodulation, it is not; rather, it is pulsed.

However, the frequency spectrums are similar. Sideband distributions are similar, but not identical, since the pulse transmitter in figure 2-30 is gated on and off instead of being modulated by a square wave as was the case in view (D) of figure 2-29.

Remember, in pulse modulation the sidebands produced to accompany the carrier during transmission are directly related to the harmonic content of the modulating wave shape. In figure 2-31, (view A, view B and view C), observe the square and rectangular wave shapes used to pulse modulate the same carrier frequency in each of the three views.

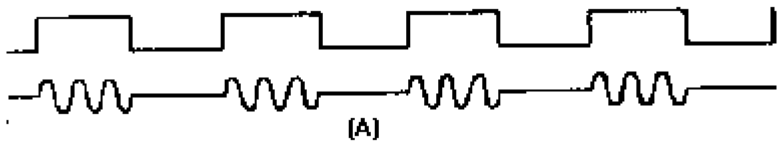


Figure 2-31A.—Varying pulse-modulating waves.

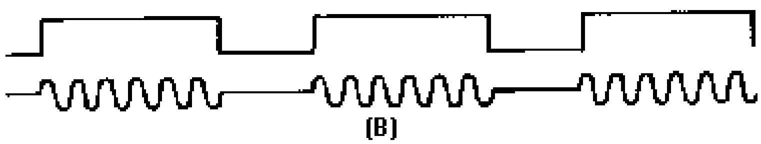


Figure 2-31B.—Varying pulse-modulating waves.

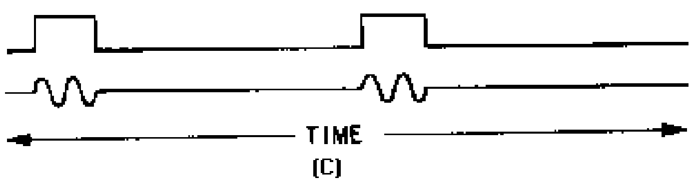


Figure 2-31C.—Varying pulse-modulating waves.

Let's take note of some timing relationships in the three modulating sequences in figure 2-31:

- the time for the rf cycle is the same in each case
- the number of cycles occurring in each group is different
- the ratio between transmitting and non-transmitting time varies
- the transmitter produces an rf wave four times in view (A), three transmission groups in view (B), and only two in view (C)
- rf is generated only during the positive pulses

In figure 2-32, observe the relative time for individual rf cycles. The time for each cycle is the same in views (A) and (B). Since this time is the same, we can assume that the carrier frequency is the same. But in view (C) the time for each cycle is about half that in views (A) and (B). Therefore, the frequency of the carrier in view (C) is nearly twice that of the other two. This illustration shows that carrier frequencies in pulse systems can vary.

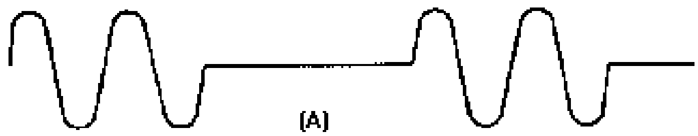


Figure 2-32A.—Carrier frequency.

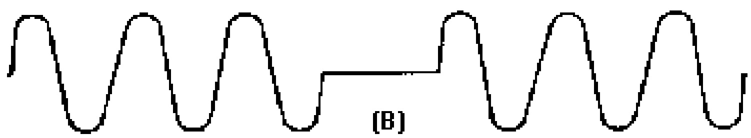


Figure 2-32B.—Carrier frequency.

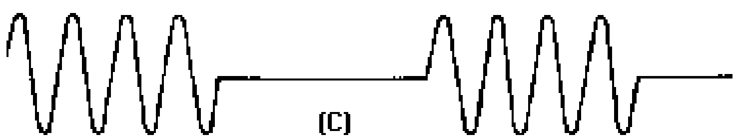


Figure 2-32C.—Carrier frequency.

The carrier frequency is not the only frequency we must concern ourselves with in pulse systems. We must also be concerned with the frequency that is associated with the repetition rate of groups of pulses. Figure 2-33 shows that a specific time period exists between each group of rf pulses. This time is the same for each repetition of the pulse and is called the PULSE-REPETITION TIME (prt). To find out how often these groups of pulses occur, compute PULSE-REPETITION FREQUENCY (prf) using the formula:

$$prf = \frac{1}{prt}$$

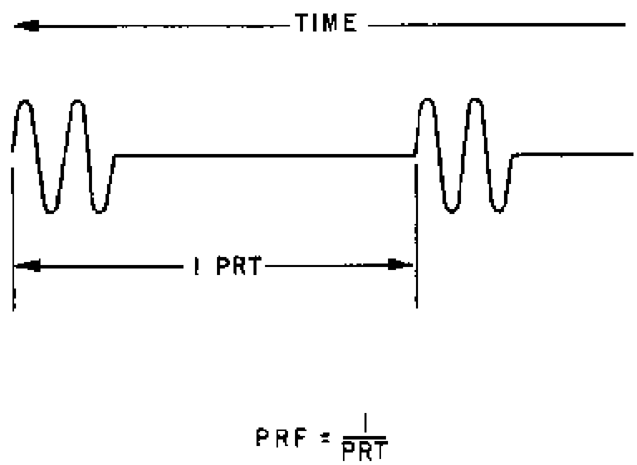


Figure 2-33.—Pulse-repetition time (prt).

Just remember that the pulse-repetition time is the time it takes for a pulse to recur, as shown in figure 2-34. The duration of time of the pulse (**a**) plus the time when no pulse occurs (**b**) equals the total pulse-repetition time.

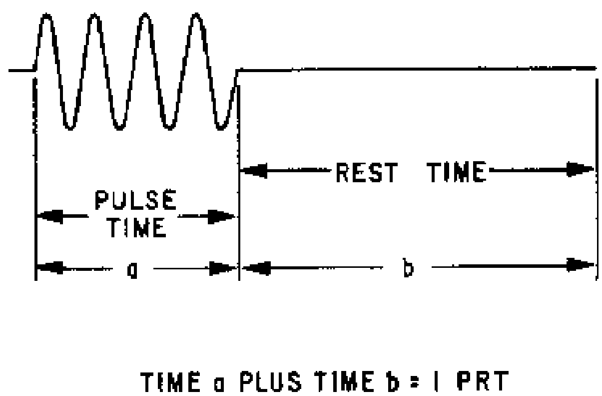


Figure 2-34.—Pulse cycles.

The time during which the pulse is occurring is called PULSE DURATION (pd) or PULSE WIDTH (pw), as shown in figure 2-35. As you will soon see, pulse width is important in pulse modulation.

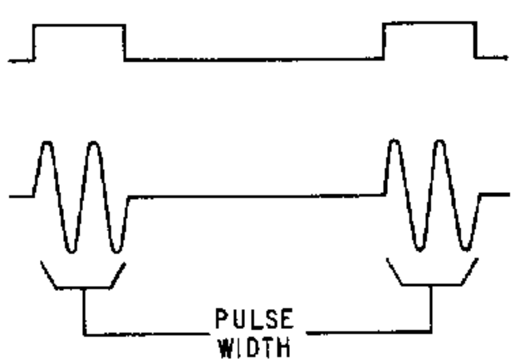


Figure 2-35.—Pulse width (pw).

The time we have been referring to as the time of no pulse, or nonpulse time, is referred to as REST TIME (rt). The duration of this rest time will determine certain capabilities of the pulse-modulation system. The pulse width is the time that the transmitter produces rf oscillations and is the actual pulse transmission time. During the nonpulse time, shown in figure 2-36, the transmitter produces no oscillations and the oscillator is cut off.

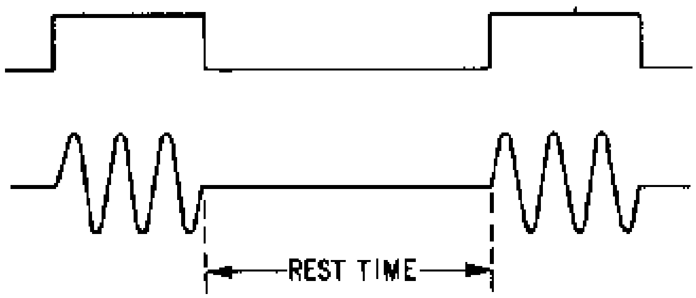


Figure 2-36.—Rest time (rt).

Some pulse transmitter-receiver systems transmit the pulse and then rest, awaiting the return of an echo. Rest time provides the system time for the receive cycle of operation.

Power in a Pulse System

When discussing power in a pulse-modulation system, we have to consider PEAK POWER and AVERAGE POWER. Peak power is the maximum value of the transmitted pulse; average power is the peak power value averaged over the pulse-repetition time. Peak power is very easy to see in a pulse system. In figure 2-37, all pulsed wave shapes have a peak power of 100 watts. Also note that in views (A), (B), and (C) the pulse width is the same, even though the carrier frequency is different. In these three cases average power would be the same. This is because average power is actually equal to the peak power of a pulse averaged over 1 operating cycle. However, the pulse width is increased in view (D) and we have a greater average power with the same prt. In view (E) the decreased pulse width has decreased average power over the same prt.

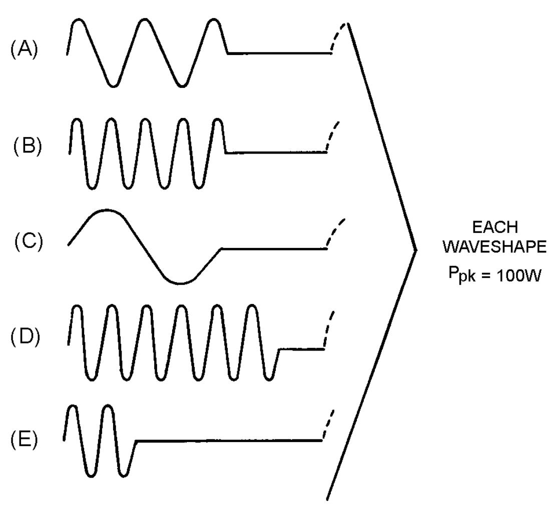


Figure 2-37.—Peak and average power.

Use these simple rules to determine power in a pulsed-wave shape:

- Peak power is the maximum power reached by the transmitter during the pulse.

- Average power equals the peak power averaged over one cycle.

Duty Cycle

In pulse modulation you will need to know the percentage of time the system is actually producing rf. For example, let's say that a pulse system is transmitting 25 percent of the time. This would mean that the pw is 1/4 the prt. For every 60 minutes we operate the pulse system, we actually transmit a total of only 15 minutes.

The DUTY CYCLE is the ratio of working time to total time for intermittently operated devices. Thus, duty cycle represents a ratio of actual transmitting time to transmitting time plus rest time. To establish the duty cycle, divide the pw by the prt of the system. This yields the duty cycle and is expressed as a decimal figure. With this information, we can figure percentage of transmitting time by multiplying the duty cycle by 100.

Applications of Pulse Modulation

Pulse modulation has many applications in the transmission of intelligence information. In telemetry, for example, the width of successive pulses may tell us humidity; the changing of the rest time may tell us pressure. In other applications, as you will see later in this text, the changing of the average power can provide us with intelligence information.

In radar a pulse is transmitted and travels some distance to a target where it is then reflected back to the system. The amount of time it takes provides us with information that can be converted to distance.

Telemetry and radar systems use the principles of pulse modulation described in this section. Let's quickly review what has been presented:

- **Pulse width (pw)** — the duration of time rf frequency is transmitted

- **Rest time (rt)** — the time the transmitter is resting (not transmitting)

- **Carrier frequency** — the frequency of the rf wave generated in the oscillator of the transmitter

- **Pulse-repetition time (prt)** — the total time of 1 complete pulse cycle of operation (rest time plus pulse width)

- **Pulse-repetition frequency (prf)** — the rate, in pulses per second, that the pulse occurs

- **Power peak** — the maximum power contained in the pulse

- **Average power** — the peak power averaged over 1 complete operating cycle

- **Duty cycle** — a decimal number that expresses a ratio in a pulse modulation system of transmit time to total time

Pulse modulation will play a major part in your electronics career. In one way or another, you will encounter it in some form. The function of the particular system may involve many variations of the characteristics presented here. We will now look at some specific applications of pulse modulation in radar and communications systems.

Q-14. Overmodulating an rf carrier in amplitude modulation produces a waveform which is similar to what modulated waveform?

Q-15. What is prt?

Q-16. What is nonpulse time?

Q-17. What is average power in a pulsed system?

RADAR MODULATION

Radio frequency energy in radar is transmitted in short pulses with time durations that may vary from 1 to 50 microseconds or more. If the transmitter is cut off before any reflected energy returns from a target, the receiver can distinguish between the transmitted pulse and the reflected pulse. After all reflections have returned, the transmitter can again be cut on and the process repeated. The receiver output is applied to an indicator which measures the time interval between the transmission of energy and its return as a reflection. Since the energy travels at a constant velocity, the time interval becomes a measure of the distance traveled (RANGE). Since this method does not depend on the relative frequency of the returned signal, or on the motion of the target, difficulties experienced in cw or fm methods are not encountered. The pulse modulation method is used in many military radar applications.

Most radar oscillators operate at pulse voltages between 5 and 20 kilovolts. They require currents of several amperes during the actual pulse which places severe requirements on the modulator. The function of the high-vacuum tube modulator is to act as a switch to turn a pulse ON and OFF at the transmitter in response to a control signal. The best device for this purpose is one which requires the least signal power for control and allows the transfer of power from the transmitter power source to the oscillator with the least loss. The pulse modulator circuits discussed in this section are typical pulse modulators used in radar equipment.

Spark-Gap Modulator

The SPARK-GAP MODULATOR consists of a circuit for storing energy, a circuit for rapidly discharging the storage circuit (spark gap), a pulse transformer, and an ac power source. The circuit for storing energy is essentially a short section of artificial transmission line which is known as the PULSE-FORMING NETWORK (pfn). The pulse-forming network is discharged by a spark gap. Two types of spark gaps are used: FIXED GAPS and ROTARY GAPS. The fixed gap, discussed in this section, uses a trigger pulse to ionize the air between the contacts of the spark gap and to initiate the discharge of the pulse-forming network. The rotary gap is similar to a mechanically driven switch.

A typical fixed, spark-gap modulator circuit is shown in figure 2-38. Between trigger pulses the spark gap is an open circuit. Current flows through the pulse transformer (T1), the pulse-forming network (C1, C2, C3, C4, and L2), the diode (V1), and the inductor (L1) to the plate supply voltage (E_b). These components form the charging circuit for the pulse-forming network.

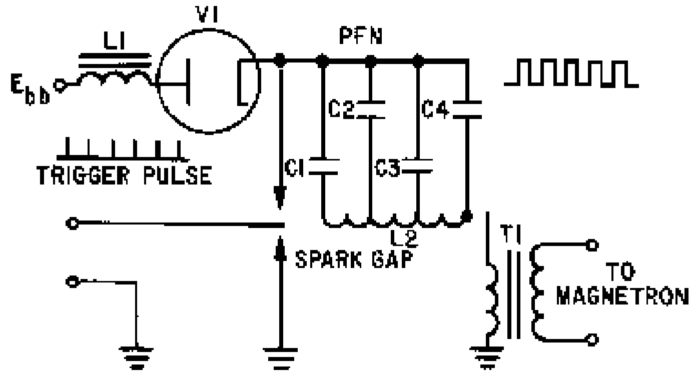


Figure 2-38.—Fixed spark-gap modulator.

The spark gap is actually triggered (ionized) by the combined action of the charging voltage across the pulse-forming network and the trigger pulse. (Ionization was discussed in *NEETS*, Module 6, *Introduction to Electronic Emission, Tubes and Power Supplies*.) The air between the trigger pulse injection point and ground is ionized by the trigger voltage. This, in turn, initiates the ionization of the complete gap by the charging voltage. This ionization allows conduction from the charged pulse-forming network through pulse transformer T1. The output pulse is then applied to an oscillating device, such as a magnetron.

Thyratron Modulator

The hydrogen THYRATRON MODULATOR is an electronic switch which requires a positive trigger of only 150 volts. The trigger potential must rise at the rate of 100 volts per microsecond to cause the modulator to conduct. In contrast to spark gap devices, the hydrogen thyratron (figure 2-39) operates over a wide range of anode voltages and pulse-repetition rates. The grid has complete control over the initiation of cathode emission for a wide range of voltages. The anode is completely shielded from the cathode by the grid. Thus, effective grid action results in very smooth firing over a wide range of anode voltages and repetition frequencies. Unlike most other thyratrons, the positive grid-control characteristic ensures stable operation. In addition, deionization time is reduced by using the hydrogen-filled tube.

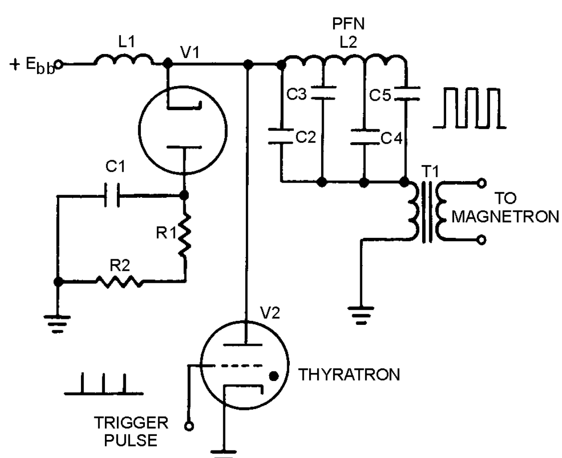


Figure 2-39.—Typical thyratron gas-tube modulator.

The hydrogen thyratron modulator provides improved timing because the synchronized trigger pulse is applied to the control grid of the thyratron (V2) and instantaneous firing is obtained. In addition, only one gas tube is required to discharge the pulse-forming network, and a low amplitude trigger pulse is sufficient to initiate discharge. A damping diode is used to prevent breakdown of the thyratron by reverse-voltage transients. The thyratron requires a sharp leading edge for a trigger pulse and depends on a sudden drop in anode voltage (controlled by the pulse-forming network) to terminate the pulse and cut off the tube.

As shown in figure 2-39, the typical thyratron modulator is very similar to the spark-gap modulator. It consists of a power source (E_b), a circuit for storing energy (L2, C2, C3, C4, and C5), a circuit for discharging the storage circuit (V2), and a pulse transformer (T1). In addition this circuit has a damping diode (V1) to prevent reverse-polarity signals from being applied to the plate of V2 which could cause V2 to breakdown.

With no trigger pulse applied, the pfn charges through T1, the pfn, and the charging coil L1 to the potential of E_b. When a trigger pulse is applied to the grid of V2, the tube ionizes causing the pulse-forming network to discharge through V2 and the primary of T1. As the voltage across the pfn falls below the ionization point of V2, the tube shuts off. Because of the inductive properties of the pfn, the positive discharge voltage has a tendency to swing negative. This negative overshoot is prevented from damaging the thyratron and affecting the output of the circuit by V1, R1, R2, and C1. This is a damping circuit and provides a path for the overshoot transient through V1. It is dissipated by R1 and R2 with C1 acting as a high-frequency bypass to ground, preserving the sharp leading and trailing edges of the pulse. The hydrogen thyratron modulator is the most common radar modulator.

Pulse modulation is also useful in communications systems. The intelligence-carrying capability and power requirements for communications systems differ from those of radar. Therefore, other methods of achieving pulse modulation that are more suitable for communications systems will now be studied.

Q-18. What is the primary component for a spark-gap modulator?

Q-19. What are the basic components of a thyratron modulator?

COMMUNICATIONS PULSE MODULATORS

To transmit intelligence using pulse modulation, you must provide a method to vary some characteristic of the pulse train in accordance with the modulating signal. Figure 2-40 illustrates a simple pulse train. The characteristics of these pulses that can be varied are amplitude, pulse width, pulse-repetition time, and the pulse position as compared to a reference. In addition to these three characteristics, pulses may be transmitted according to a code to represent the different levels of the modulating signal. To ensure maximum fidelity (accuracy in reproducing a modulating wave), the modulating signal has to be represented by enough pulses to restore the original wave shape. Logically, the higher the sampling rate (the more often sampled) of the pulse modulator, the more accurately the original modulating wave can be reproduced. Figure 2-41 illustrates the effectiveness of three pulse-sampling rates. View (A) shows a sampling rate of more than two times the modulating frequency. As you can see, this reproduces the modulating signal very accurately. However, the high sampling rate requires a wide bandwidth and increases the average power required of the transmitter. If less than two samples per cycle are made, you are not able to reproduce the original modulating signal, as shown in view (B). View (C) shows a sampling rate that is two times the highest modulating frequency. This is the minimum sampling rate that will give a sufficiently accurate reproduction of the modulating wave. The standard sampling rate is 2.5 times the highest frequency that is to be transmitted. This ensures the ability to accurately reproduce the modulating waveform. In military voice systems the bandwidth for voice signals is limited to 300 to 3,000 hertz, requiring a sampling frequency of 8 kilohertz. Although the pulse characteristic that is changed may vary for each type of pulse modulation, the sampling frequency will remain constant. We will now briefly discuss common types of pulse modulation.

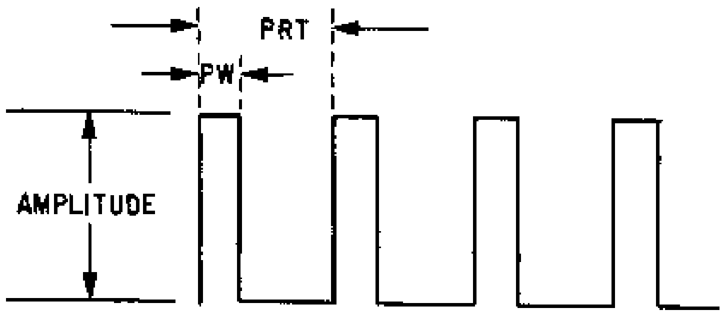


Figure 2-40.—Pulse train.

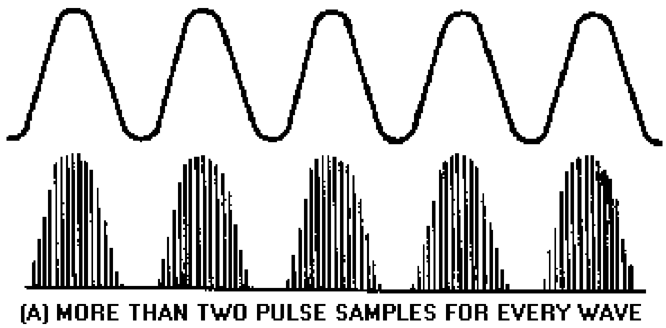


Figure 2-41A.—Pulse sampling rates. MORE THAN TWO PULSE SAMPLES FOR EVERY WAVE.

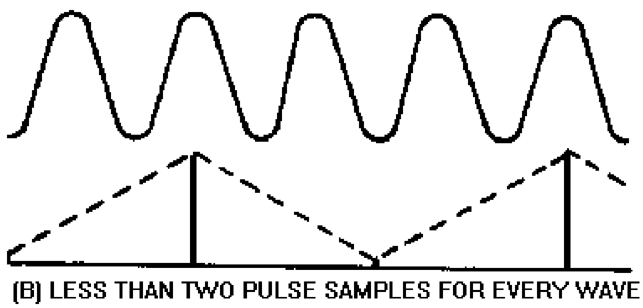


Figure 2-41B.—Pulse sampling rates. LESS THAN TWO PULSE SAMPLES FOR EVERY WAVE.

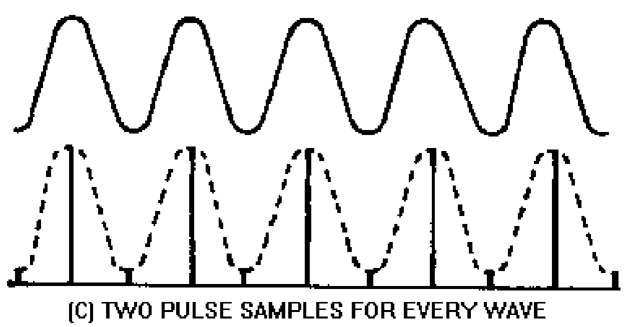


Figure 2-41C.—Pulse sampling rates. TWO PULSE SAMPLES FOR EVERY WAVE.

Pulse-Amplitude Modulation

Some characteristic of the sampling pulses must be varied by the modulating signal for the intelligence of the signal to be present in the pulsed wave. Figure 2-42 shows three typical waveforms in which the pulse amplitude is varied by the amplitude of the modulating signal. View (A) represents a sine wave of intelligence to be modulated on a transmitted carrier wave. View (B) shows the timing pulses which determine the sampling interval. View (C) shows PULSE-AMPLITUDE MODULATION (pam) in which the amplitude of each pulse is controlled by the instantaneous amplitude of the modulating signal at the time of each pulse.

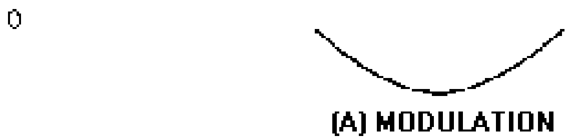


Figure 2-42A.—Pulse-amplitude modulation (pam). MODULATION.

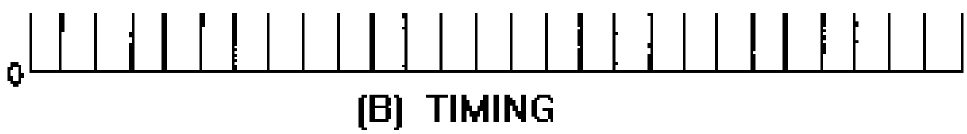


Figure 2-42B.—Pulse-amplitude modulation (pam). TIMING.

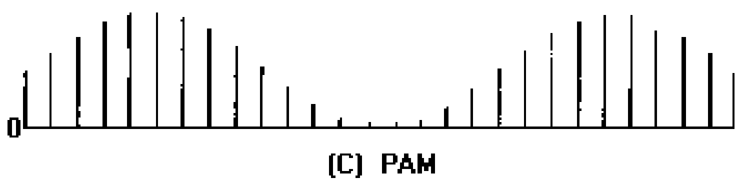


Figure 2-42C.—Pulse-amplitude modulation (pam). PAM.

Pulse-amplitude modulation is the simplest form of pulse modulation. It is generated in much the same manner as analog-amplitude modulation. The timing pulses are applied to a pulse amplifier in which the gain is controlled by the modulating waveform. Since these variations in amplitude actually represent the signal, this type of modulation is basically a form of AM. The only difference is that the signal is now in the form of pulses. This means that pam has the same built-in weaknesses as any other AM signal - high susceptibility to noise and interference. The reason for susceptibility to noise is that any interference in the transmission path will either add to or subtract from any voltage already in the circuit (signal voltage). Thus, the amplitude of the signal will be changed. Since the amplitude of the voltage represents the signal, any unwanted change to the signal is considered a SIGNAL DISTORTION. For this reason, pam is not often used. When pam is used, the pulse train is used to frequency modulate a carrier for transmission. Techniques of pulse modulation other than pam have been developed to overcome problems of noise interference. The following sections will discuss other types of pulse modulation.

Q-20. What action is necessary to impress intelligence on the pulse train in pulse modulation?

Q-21. To ensure the accuracy of a transmission, what is the minimum number of times a modulating wave should be sampled in pulse modulation?

Q-22. What, if any, noise susceptibility advantage exists for pulse-amplitude modulation over analog-amplitude modulation?

Pulse-Time Modulation

In pulse-modulated systems, as in an analog system, the intelligence may be impressed on the carrier by varying any of its characteristics. In the preceding paragraphs the method of modulating a pulse train by varying its amplitude was discussed. Time characteristics of pulses may also be modulated with intelligence information. Two time characteristics may be affected: (1) the *time duration of the pulses*, referred to as PULSE-DURATION MODULATION (pdm) or PULSE-WIDTH MODULATION (pwm); and (2) the *time of occurrence of the pulses*, referred to as PULSE-POSITION MODULATION (ppm), and a special type of PULSE-TIME MODULATION (ptm) referred to as PULSE-FREQUENCY MODULATION (pfm). Figure 2-43 shows these types of ptm in views (C), (D), and (E). Views (A) and (B) show the modulating signal and timing, respectively.

Figure 2-43A.—Pulse-time modulation (ptm). MODULATION.

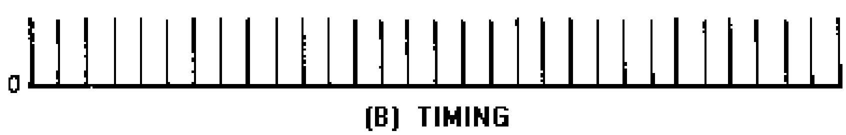


Figure 2-43B.—Pulse-time modulation (ptm). TIMING.

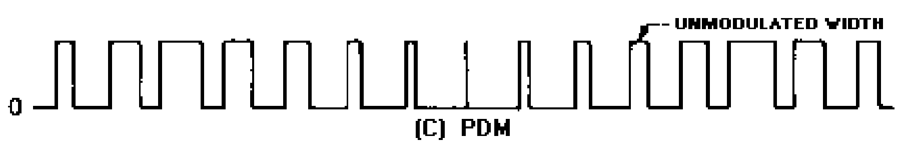


Figure 2-43C.—Pulse-time modulation (ptm). PDM.

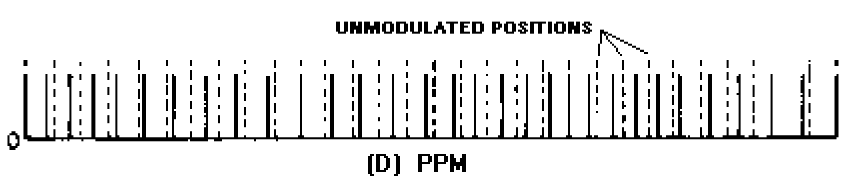


Figure 2-43D.—Pulse-time modulation (ptm). PPM.

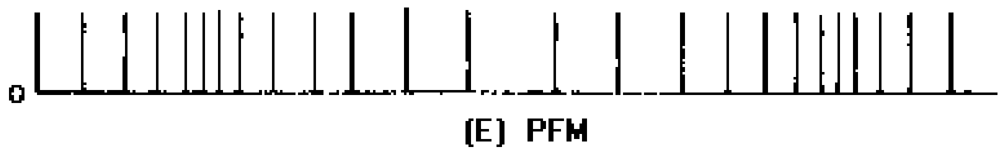


Figure 2-43E.—Pulse-time modulation (ptm). PFM

PULSE-DURATION MODULATION.—Pdm and pwm are designations for a single type of modulation. The width of each pulse in a train is made proportional to the instantaneous value of the modulating signal at the instant of the pulse. Either the leading edges, the trailing edges, or both edges of the pulses may be modulated to produce the variation in pulse width. Pdm can be obtained in a number of ways, one of which is illustrated in views (A) through (D) in figure 2-44. A circuit to produce pdm is shown in figure 2-45. Adding the modulating signal [figure 2-44, view (A)] to a repetitive sawtooth [view (B)] will result in the waveform shown in view (C). This waveform is then applied to a circuit which changes state when the input signal exceeds a specific threshold level. This action produces pulses with widths that are determined by the length of time that the input waveform exceeds the threshold level. The resulting waveform is shown in view (D).

Figure 2-44A.—Pulse-duration modulation (pdm). MODULATING SIGNAL.

Figure 2-44B.—Pulse-duration modulation (pdm). REPETITIVE SAWTOOTH PULSES.

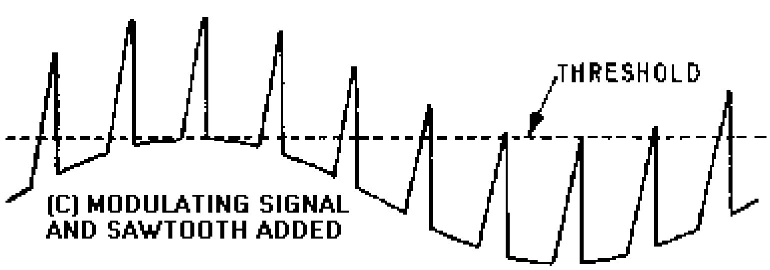


Figure 2-44C.—Pulse-duration modulation (pdm). MODULATING SIGNAL AND SAWTOOTH ADDED.

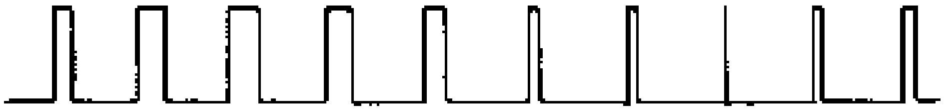


Figure 2-44D.—Pulse-duration modulation (pdm). WIDTH MODULATED PULSES FROM CIRCUIT OF FIGURE 2-45.

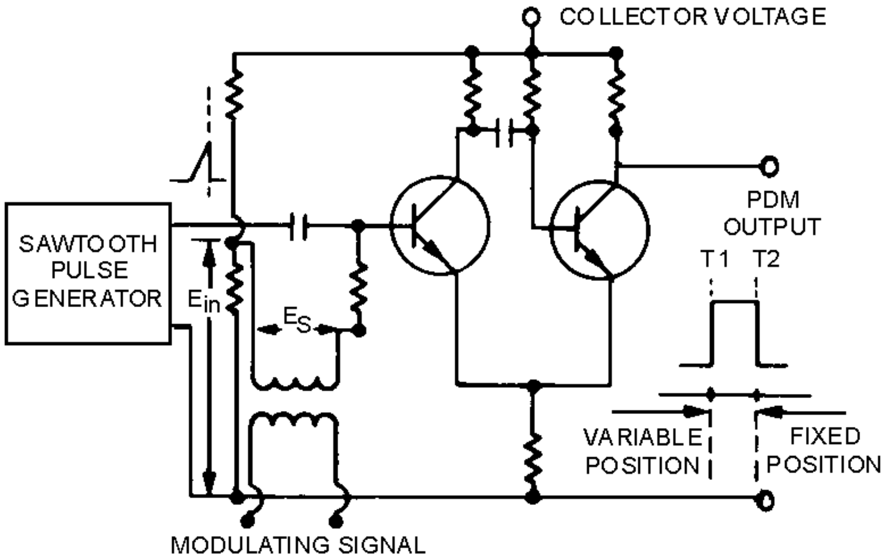


Figure 2-45.—Circuit for producing pdm.

In the circuit of figure 2-45, a series of sawtooth pulses, occurring at the sampling rate, is applied to a one-shot multivibrator. The multivibrator has the signal voltage $E_s$superimposed on the bias voltage E_{in}. Each pulse triggers a cycle of multivibrator operation which terminates after a time interval and varies linearly with the voltage E_s. The pulse of plate voltage produced by the multivibrator will have a leading edge at T1. The leading edge will vary in position with the signal voltage, while the trailing edge at T2 is fixed by the termination of the sawtooth pulse. The length of the output pulse is thus duration or width modulated. If the sawtooth has an instantaneous buildup and a sloping trailing edge, then the leading edge (T1) is fixed and the trailing edge (T2) varies. If the sawtooth generator produces a slope on both leading and trailing edges, both T1 and T2 are variable in position, but the result is still pdm. Pdm is often used because it is of a constant amplitude and is, therefore, less susceptible to noise. When compared with ppm, pdm has the disadvantage of a varying pulse, width and, therefore, of varying power content. This means that the transmitter must be powerful enough to handle the maximum-width pulses, although the average power transmitted is much less than peak power. On the other hand, pdm will still work if the synchronization between the transmitter and receiver fails; in ppm it will not, as will be seen in the next section.

PULSE-POSITION MODULATION.—The amplitude and width of the pulse is kept constant in the system. The position of each pulse, in relation to the position of a recurrent reference pulse, is varied by each instantaneous sampled value of the modulating wave. Ppm has the advantage of requiring constant transmitter power since the pulses are of constant amplitude and duration. It is widely used but has the disadvantage of depending on transmitter-receiver synchronization.

Ppm can be generated in several ways, but we will discuss one of the simplest. Figure 2-46 shows three waveforms associated with developing ppm from pdm. The pdm pulse train is applied to a differentiating circuit. (Differentiation was presented in *NEETS*, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*.) This provides positive- and negative-polarity pulses that correspond to the leading and trailing edges of the pdm pulses. Considering pdm and its generation, you can see that each pulse has a leading and trailing edge. In this case the position of the leading edge is fixed, whereas the trailing edge is not, as shown in view (A) of figure 2-46. The resultant pulses after the differentiation are shown in view (B). The negative pulses are position-modulated in accordance with the modulating waveform. Both the negative and positive pulse are then applied to a rectification circuit. This application eliminates the positive, non-modulated pulses and develops a ppm pulse train, as shown in view (C).

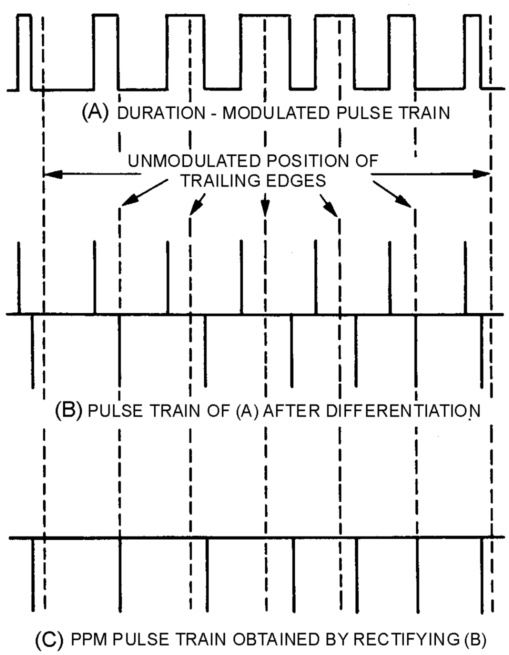


Figure 2-46.—Pulse-position modulation (ppm).

PULSE-FREQUENCY MODULATION.—Pfm is a method of pulse modulation in which the modulating wave is used to frequency modulate a pulse-generating circuit. For example, the pulse rate may be 8,000 pulses per second (pps) when the signal voltage is 0. The pulse rate may step up to 9,000 pps for maximum positive signal voltage, and down to 7,000 pps for maximum negative signal voltage. Figure 2-47, views (A), (B), and (C) show three typical waveforms for pfm. This method of modulation is not used extensively because of complicated pfm generation methods. It requires a stable oscillator that is frequency modulated to drive a pulse generator. Since the other forms of ptm are easier to achieve, they are commonly used.

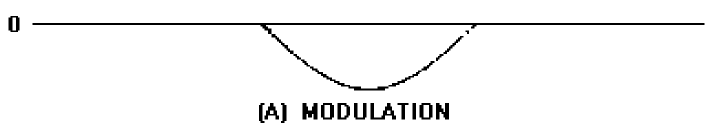


Figure 2-47A.—Pulse-frequency modulation (pfm). MODULATION.

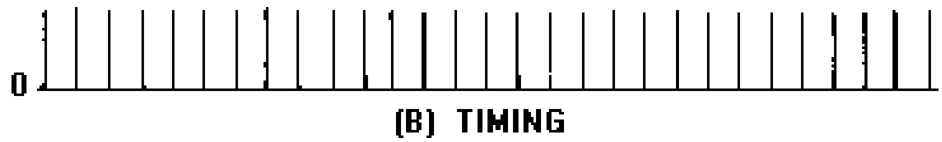


Figure 2-47B.—Pulse-frequency modulation (pfm). TIMING.

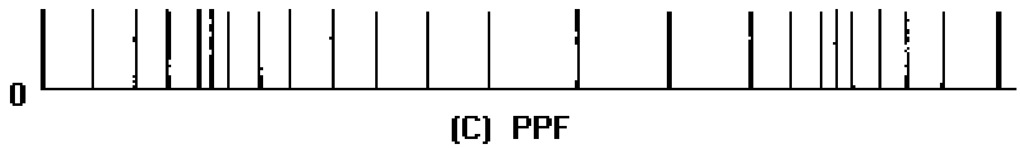


Figure 2-47C.—Pulse-frequency modulation (pfm). PFM.

Q-23. What characteristics of a pulse can be changed in pulse-time modulation?

Q-24. Which edges of the pulse can be modulated in pulse-duration modulation?

Q-25. What is the main disadvantage of pulse-position modulation?

Q-26. What is pulse-frequency modulation?

Pulse-Code Modulation

PULSE-CODE MODULATION (pcm) refers to a system in which the standard values of a QUANTIZED WAVE (explained in the following paragraphs) are indicated by a series of coded pulses. When these pulses are decoded, they indicate the standard values of the original quantized wave. These codes may be *binary*, in which the symbol for each quantized element will consist of pulses and spaces: *ternary*, where the code for each element consists of any one of three distinct kinds of values (such as positive pulses, negative pulses, and spaces); or *n-ary*, in which the code for each element consists of nay number (n) of distinct values. This discussion will be based on the binary pcm system.

All of the pulse-modulation systems discussed previously provide methods of converting analog wave shapes to digital wave shapes (pulses occurring at discrete intervals, some characteristic of which is varied as a continuous function of the analog wave). The entire range of amplitude (frequency or phase) values of the analog wave can be arbitrarily divided into a series of standard values. Each pulse of a pulse train [figure 2-48, view (B)] takes the standard value nearest its actual value when modulated. The modulating wave can be faithfully reproduced, as shown in views (C) and (D). The amplitude range has been divided into 5 standard values in view (C). Each pulse is given whatever standard value is nearest its actual instantaneous value. In view (D), the same amplitude range has been divided into 10 standard levels. The curve of view (D) is a much closer approximation of the modulating wave, view (A), than is the 5-level quantized curve in view (C). From this you should see that the greater the number of standard levels used, the more closely the quantized wave approximates the original. This is also made evident by the fact that an infinite number of standard levels exactly duplicates the conditions of nonquantization (the original analog waveform).

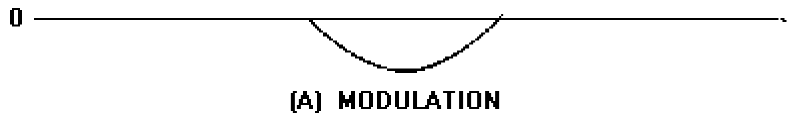


Figure 2-48A.—Quantization levels. MODULATION.

Figure 2-48B.—Quantization levels. TIMING.

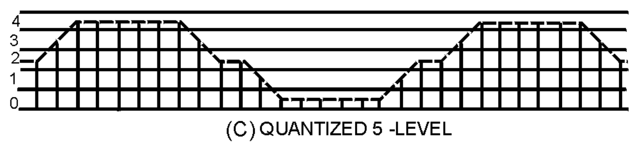


Figure 2-48C.—Quantization levels. QUANTIZED 5-LEVEL.

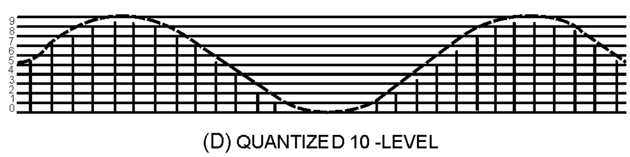


Figure 2-48D.—Quantization levels. QUANTIZED 10-LEVEL.

Although the quantization curves of figure 2-48 are based on 5- and 10-level quantization, in actual practice the levels are usually established at some exponential value of 2, such as $4(2^2)$, $8(2^3)$, $16(2^4)$, $32(2^5)$ . . . $N(2^n)$. The reason for selecting levels at exponential values of 2 will become evident in the discussion of pcm. Quantized fm is similar in every way to quantized AM. That is, the range of frequency deviation is divided into a finite number of standard values of deviation. Each sampling pulse results in a deviation equal to the standard value nearest the actual deviation at the sampling instant. Similarly, for phase modulation, quantization establishes a set of standard values. Quantization is used mostly in amplitude- and frequency-modulated pulse systems.

Figure 2-49 shows the relationship between decimal numbers, binary numbers, and a pulse-code waveform that represents the numbers. The table is for a 16-level code; that is, 16 standard values of a quantized wave could be represented by these pulse groups. Only the presence or absence of the pulses are important. The next step up would be a 32-level code, with each decimal number represented by a series of five binary digits, rather than the four digits of figure 2-49. Six-digit groups would provide a 64-level code, seven digits a 128-level code, and so forth. Figure 2-50 shows the application of pulse-coded groups to the standard values of a quantized wave.

DECIMAL NUMBER	BINARY EQUIVALENT 2^3	2^2	2^1	2^0
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1
10	1	0	1	0
11	1	0	1	1
12	1	1	0	0
13	1	1	0	1
14	1	1	1	0
15	1	1	1	1

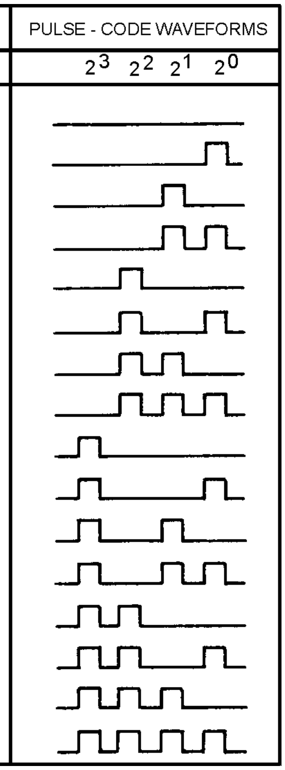


Figure 2-49.—Binary numbers and pulse-code equivalents.

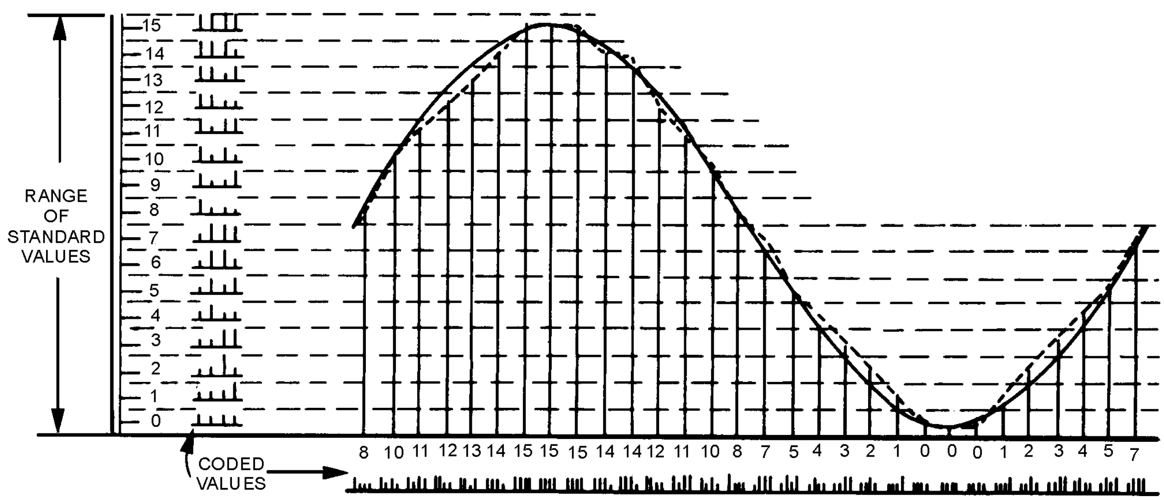


Figure 2-50.—Pulse-code modulation of a quantized wave (128 bits).

In figure 2-50 the solid curve represents the unquantized values of a modulating sinusoid. The dashed curve is reconstructed from the quantized values taken at the sampling interval and shows a very close agreement with the original curve. Figure 2-51 is identical to figure 2-50 except that the sampling interval is four times as great and the reconstructed curve is not faithful to the original. As previously stated, the sampling rate of a pulsed system must be at least twice the highest modulating frequency to get a usable reconstructed modulation curve. At the sampling rate of figure 2-50 and with a 4-element binary code, 128 bits (presence or absence of pulses) must be transmitted for each cycle of the modulating frequency. At the sampling rate of figure 2-51, only 32 bits are required; at the minimum sampling rate, only 8 bits are required.

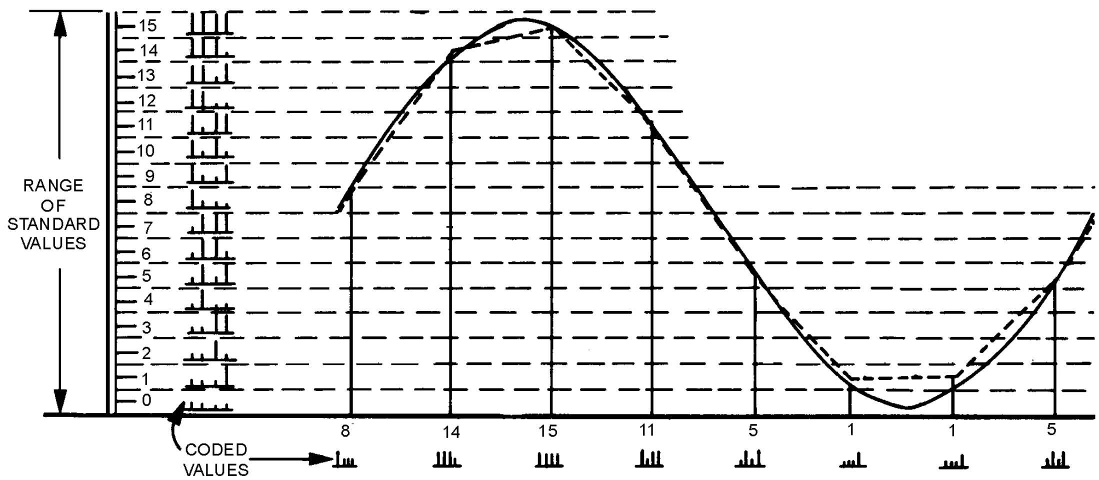


Figure 2-51.—Pulse-code modulation of a quantized wave (32 bits).

As a matter of convenience, especially to simplify the demodulation of pcm, the pulse trains actually transmitted are reversed from those shown in figures 2-49, 2-50, and 2-51; that is, the pulse with the lowest binary value (least significant digit) is transmitted first and the succeeding pulses have increasing binary values up to the code limit (most significant digit). Pulse coding can be performed in a number of ways using conventional circuitry or by means of special cathode ray coding tubes. One form of coding circuit is shown in figure 2-52. In this case, the pulse samples are applied to a holding circuit (a capacitor which stores pulse amplitude information) and the modulator converts pam to pdm. The pdm pulses are then used to gate the output of a precision pulse generator that controls the number of pulses applied to a binary counter. The duration of the gate pulse is not necessarily an integral number of the repetition pulses from the precisely timed clock-pulse generator. Therefore, the clock pulses gated into the binary counter by the pdm pulse may be a number of pulses plus the leading edge of an additional pulse. This "partial" pulse may have sufficient duration to trigger the counter, or it may not. The counter thus responds only to integral numbers, effectively quantizing the signal while, at the same time, encoding it. Each bistable stage of the counter stores ZERO or a ONE for each binary digit it represents (binary 1110 or decimal 14 is shown in figure 2-52). An electronic commutator samples the 2^0, 2^1, 2^2, and 2^3 digit positions in sequence and transmits a mark or space bit (pulse or no pulse) in accordance with the state of each counter stage. The holding circuit is always discharged and reset to zero before initiation of the sequence for the next pulse sample.

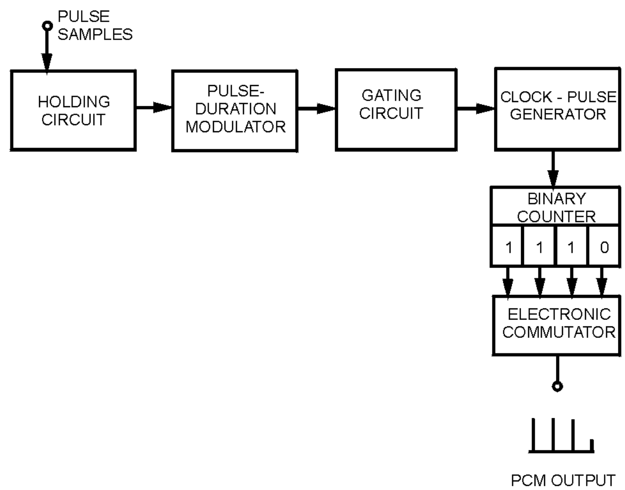


Figure 2-52.—Block diagram of quantizer and pcm coder.

The pcm demodulator will reproduce the correct standard amplitude represented by the pulse-code group. However, it will reproduce the correct standard only if it is able to recognize correctly the presence or absence of pulses in each position. For this reason, noise introduces no error at all if the signal-to-noise ration is such that the largest peaks of noise are not mistaken for pulses. When the noise is random (circuit and tube noise), the probability of the appearance of a noise peak comparable in amplitude to the pulses can be determined. This probability can be determined mathematically for any ration of signal-to-average-noise power. When this is done for 10^5 pulses per second, the approximate error rate for three values of signal power to average noise power is:

17 dB — 10 errors per second

20 dB — 1 error every 20 minutes

22 dB — 1 error every 2,000 hours

Above a threshold of signal-to-noise ration of approximately 20 dB, virtually no errors occur. In all other systems of modulation, even with signal-to-noise ratios as high as 60 dB, the noise will have some effect. Moreover, the pcm signal can be retransmitted, as in a multiple relay link system, as many times as desired, without the introduction of additional noise effects; that is, noise is not cumulative at relay stations as it is with other modulation systems.

The system does, of course, have some distortion introduced by quantizing the signal. Both the standard values selected and the sampling interval tend to make the reconstructed wave depart from the original. This distortion, called QUANTIZING NOISE, is initially introduced at the quantizing and coding modulator and remains fixed throughout the transmission and retransmission processes. Its magnitude can be reduced by making the standard quantizing levels closer together. The relationship of the quantizing noise to the number of digits in the binary code is given by the following standard relationship:

Where:

n is the number of digits in the binary code

$$\frac{\text{peak signal power}}{\text{average quantizing noise power}} = (10.8 + 6n)\ \text{dB}$$

Thus, with the 4-digit code of figure 2-50 and 2-51, the quantizing noise will be about 35 dB weaker than the peak signal which the channel will accommodate.

The advantages of pcm are two-fold. First, noise interference is almost completely eliminated when the pulse signals exceed noise levels by a value of 20 dB or more. Second, the signal may be received and retransmitted as many times as may be desired without introducing distortion into the signal.

Q-27. Pulse-code modulation requires the use of approximations of value that are obtained by what process?

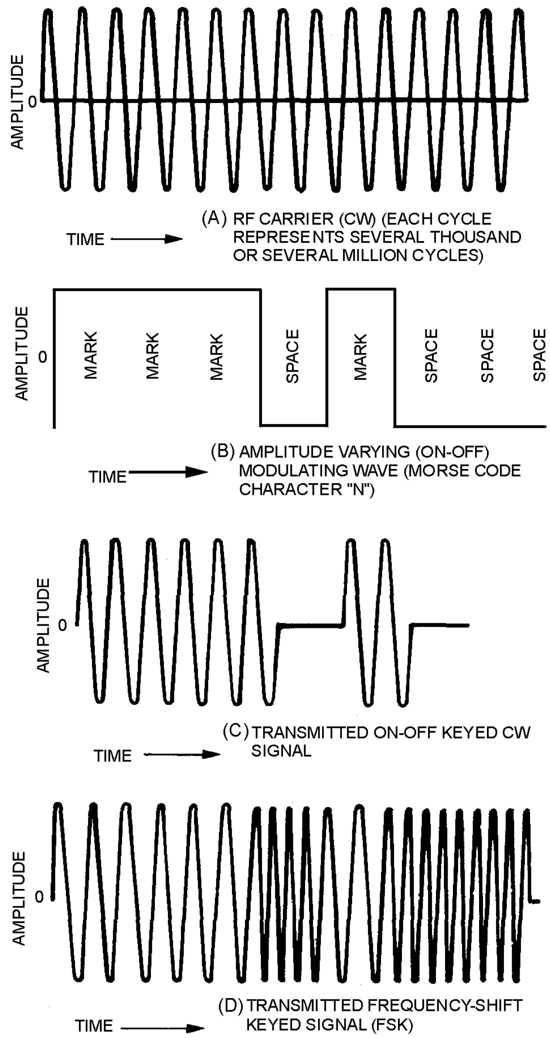
Q-28. If a modulating wave is sampled 10 times per cycle with a 5-element binary code, how many bits of information are required to transmit the signal?

Q-29. What is the primary advantage of pulse-modulation systems?

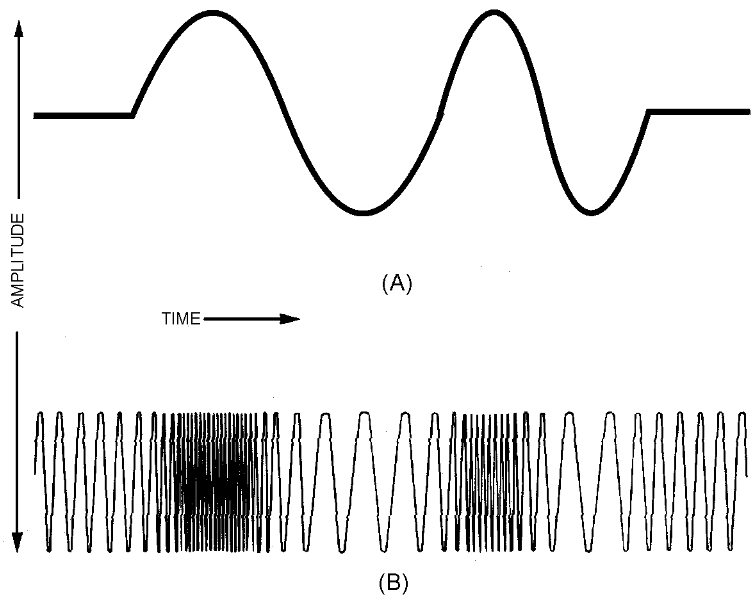
SUMMARY

Now that you have completed this chapter, a short review of what you have learned is in order. The following review will refresh your memory of angle modulation and pulse modulation.

FREQUENCY-SHIFT KEYING (fsk) is similar to cw keying in amplitude modulation and is a form of angle modulation. The carrier frequency is changed between two discrete values by the opening and closing of a key.



In **FREQUENCY MODULATION (fm)** the instantaneous frequency of the radio-frequency wave is varied in accordance with the modulating signal; the amplitude of the radio-frequency wave is kept constant.

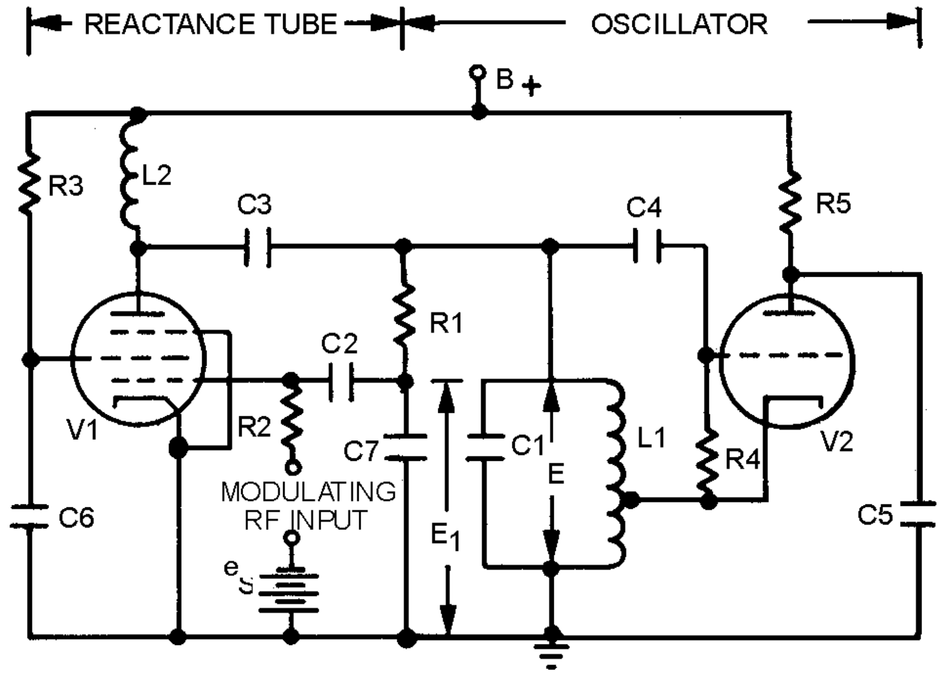


MODULATION INDEX is the ratio of the maximum frequency difference between the modulated and the unmodulated carrier, or between the deviation frequency and the modulation frequency.

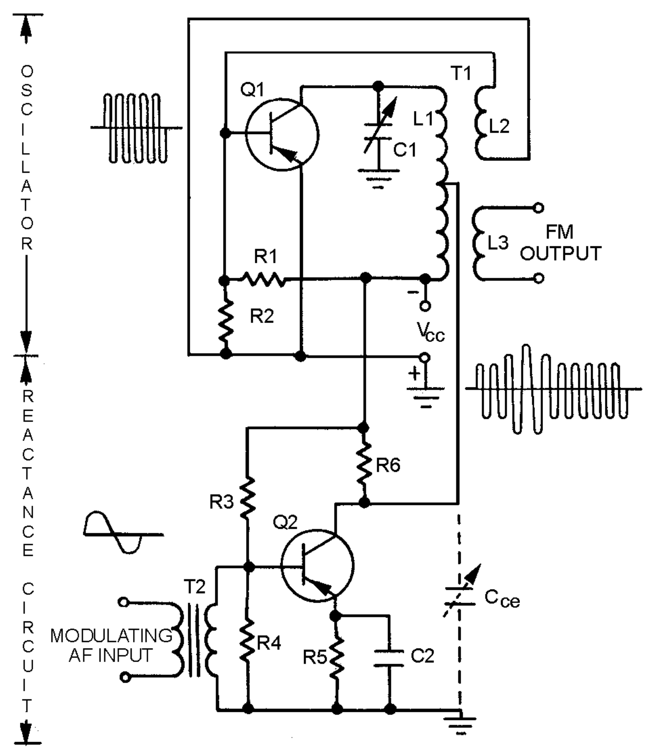
The number of SIGNIFICANT SIDEBANDS and the modulating frequency will determine the bandwidth of the fm wave. The number of significant sidebands can be determined from the modulation index.

MODULATION INDEX	SIGNIFICANT SIDEBANDS
.01	2
.4	2
.5	4
1.0	6
2.0	8
3.0	12
4.0	14
5.0	16
6.0	18
7.0	22
8.0	24
9.0	26
10.0	28
11.0	32
12.0	32
13.0	36
14.0	38
15.0	38

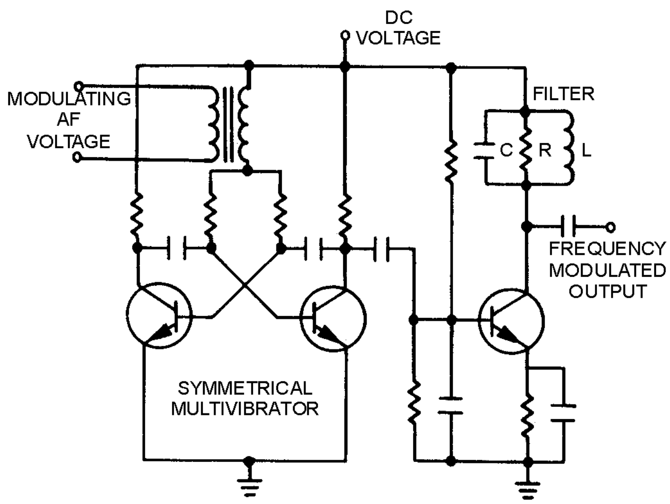
The **REACTANCE-TUBE MODULATOR** is frequency modulated by using a reactance tube in shunt with the tank circuit.



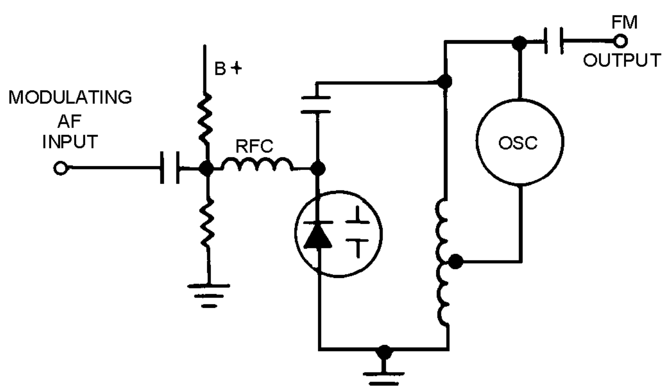
The **SEMICONDUCTOR-REACTANCE MODULATOR** is used to frequency modulate low-power semiconductor transmitters.



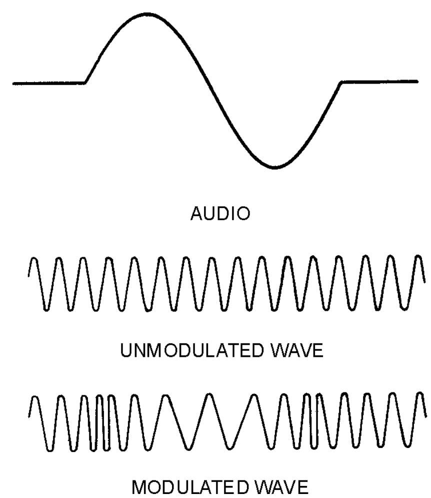
The **MULTIVIBRATOR MODULATOR** uses an astable multivibrator with a modulating voltage inserted in series with the base return of the multivibrator transistors.



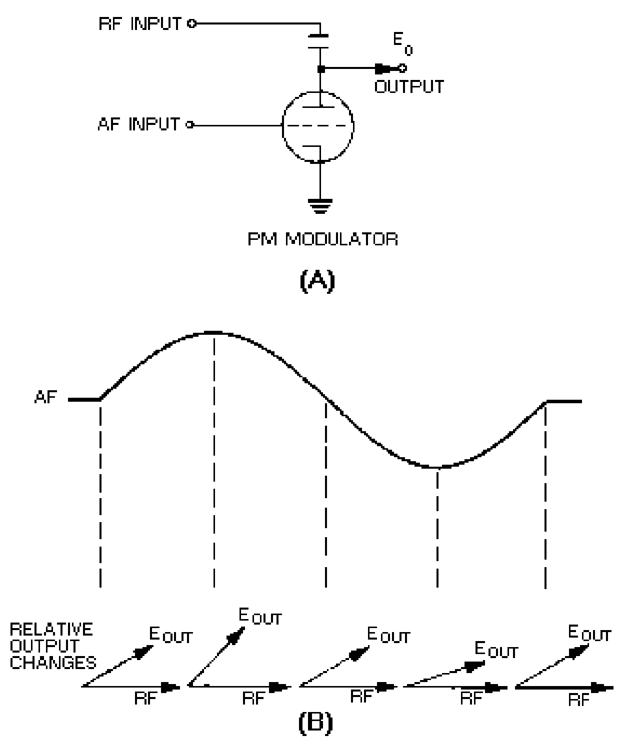
The **VARACTOR FM MODULATOR** uses a VARACTOR. This is a specially designed diode that has a certain amount of capacitance between the junction that can be controlled by reverse biasing.



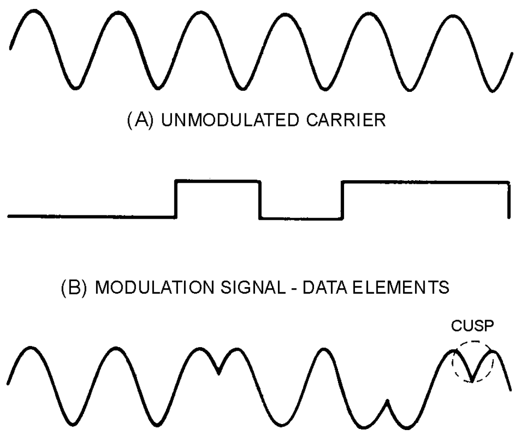
In **PHASE MODULATION** the carrier's phase is caused to shift at the rate of the modulating audio. The amount of phase shift is controlled by the amplitude of the modulating wave.



A **BASIC PHASE MODULATOR** may be a single tube in series with a capacitor to form a phase-shift network. As the impedance of the tube changes, the phase of the output shifts.

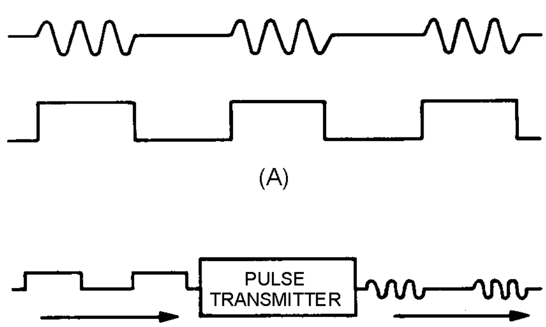


PHASE-SHIFT KEYING (psk) is similar to cw and fsk. It consists of phase reversals of the carrier frequency as modulating signal data elements open and close the modulator key.

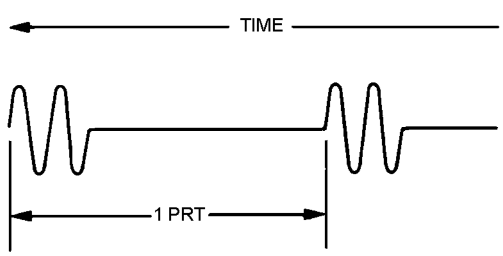


(A) UNMODULATED CARRIER
(B) MODULATION SIGNAL - DATA ELEMENTS
(C) MODULATED CARRIER
(D) MODULATED CARRIER AFTER FILTERING

PULSE MODULATION is modulation in which we allow oscillations to occur for a given period of time only during selected intervals.

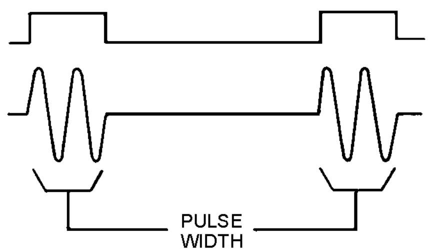


PULSE-REPETITION TIME (prt) is the specific time period between each group of rf pulses.

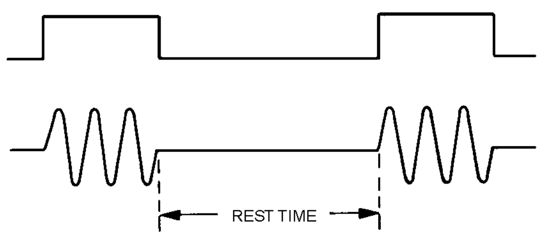


PULSE-REPETITION FREQUENCY (prf) is found by dividing the pulse repetition time into 1. This defines how often the groups of pulses occur.

PULSE WIDTH (pw) or **PULSE DURATION (pd)** is the time that a pulse is occurring.



REST TIME (rt) is the time referred to as nonpulse time.

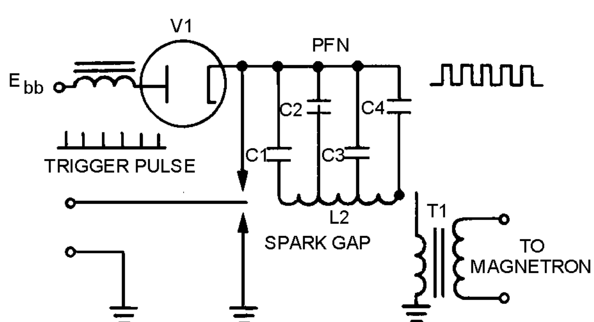


PEAK POWER is the maximum power during a pulse.

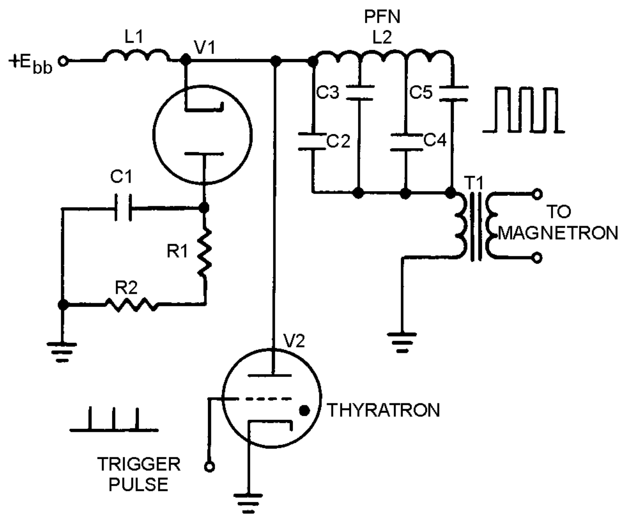
AVERAGE POWER equals the peak power averaged over one complete cycle.

DUTY CYCLE is the ratio of working time to total time, or the ratio of actual transmit time to transmit time plus rest time, for intermittently operated devices.

The **SPARK-GAP MODULATOR** consists of a circuit for storing energy, a circuit for rapidly discharging the storage circuit, a pulse transformer, and a power source.

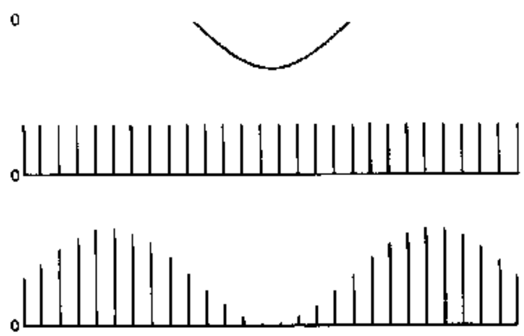


The **THYRATRON MODULATOR** is an electronic switch which requires a positive trigger of only 150 volts. The trigger must rise at the rate of 100 volts per microsecond to fire or cause the modulator to conduct.



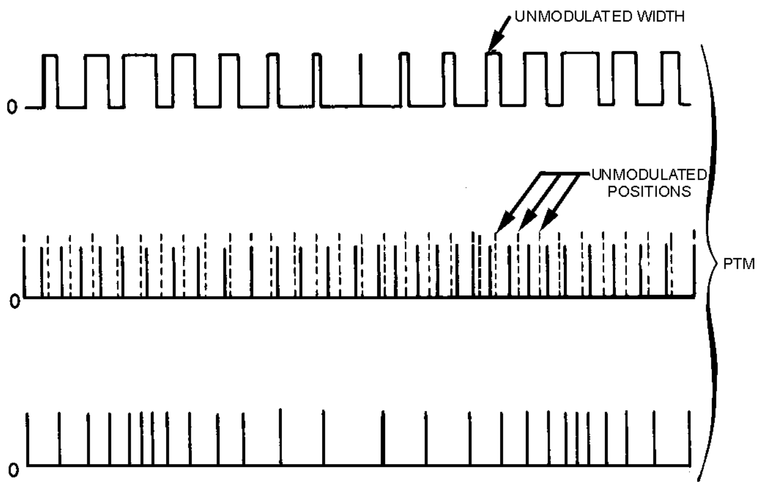
In communications **PULSE-MODULATION SYSTEMS**, the modulating wave must be SAMPLED at 2.5 times the highest modulating frequency to ensure accuracy.

PULSE-AMPLITUDE MODULATION (pam) is modulation in which the amplitude of each pulse is controlled by the instantaneous amplitude of the modulation signal at the time of each pulse.



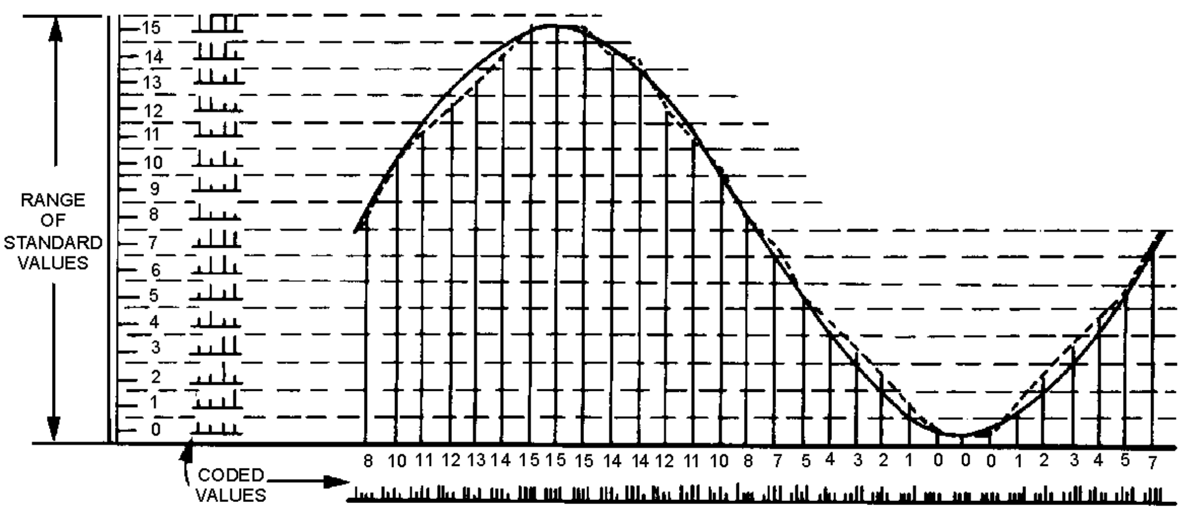
PULSE-DURATION MODULATION (pdm) or **PULSE-WIDTH MODULATION (pwm)** are both designations for a type of modulation. The width of each pulse in a train is made proportional to the instantaneous value of the modulating signal at the instant of the pulse.

PULSE-POSITION MODULATION (ppm) has the advantage of requiring constant transmitter power. The amplitude and width of the pulses are kept constant. At the same time, the position of each pulse, in relation to the position of a recurrent reference pulse, is varied by each instantaneous sampled value of the modulating wave.



PULSE-FREQUENCY MODULATION (pfm) is a method of pulse modulation in which the modulating wave is used to frequency modulate a pulse-generating circuit.

PULSE-CODE MODULATION (pcm) refers to a system in which the standard value of a quantized wave is indicated by a series of coded pulses that give the modulating wave's value at the instant of the sample.



ANSWERS TO QUESTIONS Q1. THROUGH Q29.

A-1. Frequency and phase.

A-2. Frequency-shift keying.

A-3. Resistance to noise interference.

A-4. Instantaneous frequency.

A-5. As the ratio of the frequency deviation to the maximum frequency deviation allowable.

A-6. The number of significant sidebands and the modulating frequency.

A-7. By changing the reactance of an oscillator circuit in consonance with the modulating voltage.

A-8. Collector-to-emitter capacitance.

A-9. An LCR filter.

A-10. Capacitance.

A-11. Phase.

A-12. A phase-shift network such as a variable resistor and capacitor in series.

A-13. Cw and frequency-shift keying.

A-14. Pulse modulation.

A-15. Pulse-repetition time.

A-16. Rest time.

A-17. Peak power during a pulse averaged over pulse time plus rest time.

A-18. Either a fixed spark gap that uses a trigger pulse to ionize the air between the contacts, or a rotary gap that is similar to a mechanical switch.

A-19. Power source, a circuit for storing energy, a circuit for discharging the storage circuit, and a pulse transformer.

A-20. Some characteristic of the pulses has to be varied.

A-21. 2.5 times the highest modulating frequency.

A-22. Both are susceptible to noise and interference.

A-23. The time duration of the pulses or the time of occurrence of the pulses.

A-24. Either, or both at the same time.

A-25. It requires synchronization between the transmitter and receiver.

A-26. A method of pulse modulation in which a modulating wave is used to frequency modulate a pulse-generating circuit.

A-27. Quantization.

A-28. 50.

A-29. Low susceptibility to noise.

CHAPTER 3

DEMODULATION

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Describe cw detector circuit operations for the heterodyne and regenerative detectors.
2. Discuss the requirements for recovery of intelligence from an AM signal and describe the theory of operation of the following AM demodulators: series-diode, shunt-diode, common-emitter, and common-base.
3. Describe fm demodulation circuit operation for the phase-shift and gated-beam discriminators and the ratio-detector demodulator.
4. Describe phase demodulation circuit operation for the peak, low-pass filter, and conversion detectors.

INTRODUCTION

In chapters 1 and 2 you studied how to apply intelligence (modulation) to an rf-carrier wave. Carrier modulation allows the transmission of modulating frequencies without the use of transmission wire as a medium. However, for the communication process to be completed or to be useful, the intelligence must be recovered in its original form at the receiving site. The process of re-creating original modulating frequencies (intelligence) from the rf carrier is referred to as DEMODULATION or DETECTION. Each type of modulation is different and requires different techniques to recover (demodulate) the intelligence. In this chapter we will discuss ways of demodulating AM, cw, fm, phase, and pulse modulation.

The circuit in which restoration is achieved is called the DETECTOR or DEMODULATOR (both of these terms are used in *NEETS*). The term demodulator is used because the demodulation process is considered to be the opposite of modulation. The output of an ideal detector must be an exact reproduction of the modulation existing on the rf wave. Failure to accurately recover this intelligence will result in distortion and degradation of the demodulated signal and intelligence will be lost. The distortion may be in amplitude, frequency, or phase, depending on the nature of the demodulator. A nonlinear device is required for demodulation. This nonlinear device is required to recover the modulating frequencies from the rf envelope. Solid-state detector circuits may be either a pn junction diode or the input junction of a transistor. In electron-tube circuits, either a diode or the grid or plate circuits of a triode electron tube may be used as the nonlinear device.

Q-1. What is demodulation?

Q-2. What is a demodulator?

CONTINUOUS-WAVE DEMODULATION

Continuous-wave (cw) modulation consists of on-off keying of a carrier wave. To recover on-off keyed information, we need a method of detecting the *presence or absence of rf oscillations*. The CW DEMODULATOR detects the presence of rf oscillations and converts them into a recognizable form. Figure 3-1 illustrates the received cw in view (A), the rectified cw from a diode detector in view (B), and the dc output from a filter that can be used to control a relay or light indicator in view (C).

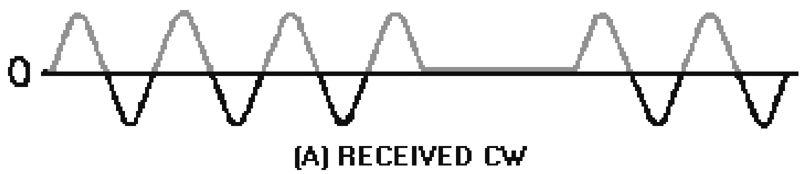


Figure 3-1A.—Cw demodulation. RECEIVED CW.

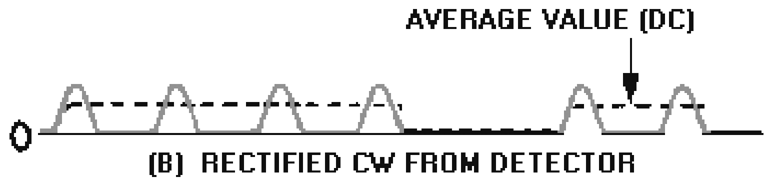


Figure 3-1B.—Cw demodulation. RECTIFIED CW FROM DETECTOR.

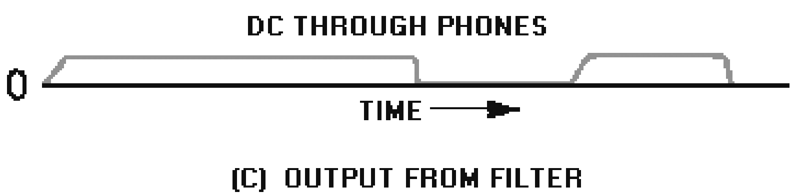


Figure 3-1C.—Cw demodulation. OUTPUT FROM FILTER.

Figure 3-2 is a, simplified circuit that could be used as a cw demodulator. The antenna receives the rf oscillations from the transmitter. The tank circuit, L and C1, acts as a frequency-selective network that is tuned to the desired rf carrier frequency. The diode rectifies the oscillations and C2 provides filtering to provide a constant dc output to control the headset. This demodulator circuit is the equivalent of a wire telegraphy circuit but it has certain disadvantages. For example, if two transmitters are very close in frequency, distinguishing which transmitting station you are receiving is often impossible without a method of fine tuning the desired frequency. Also, if the stations are within the frequency bandpass of the input tank circuit, the tank output will contain a mixture of both signals. Therefore, a method, such as HETERODYNE DETECTION, must be used which provides more than just the information on the presence or absence of a signal.

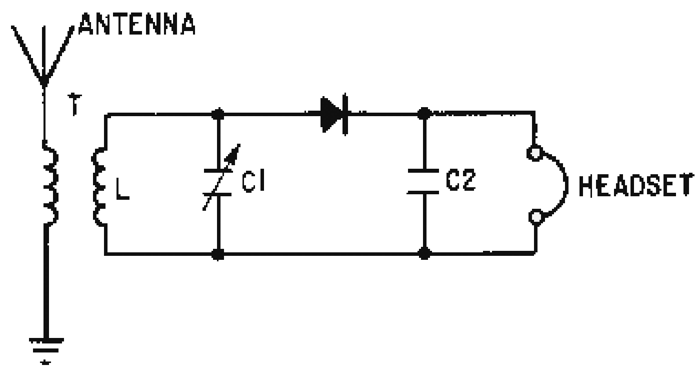


Figure 3-2.—Cw demodulator.

HETERODYNE DETECTION

The use of an af voltage in the detector aids the operator in distinguishing between various signals. Since the carrier is unmodulated, the af voltage can be developed by using the heterodyne procedure discussed in chapter 1. The procedure is to mix the incoming cw signal with locally generated oscillations. This provides a convenient difference frequency in the af range, such as 1,000 hertz. The af difference frequency then is rectified and smoothed by a detector. The af voltage is reproduced by a telephone headset or a loudspeaker.

Consider the heterodyne reception of the code letter **A**, as shown in figure 3-3, view (A). The code consists of a short burst of cw energy (dot) followed by a longer burst (dash). Assume that the frequency of the received cw signal is 500 kilohertz. The locally generated oscillations are adjusted to a frequency which is higher or lower than the incoming rf signal (501 kilohertz in this case), as shown in view (B). The voltage resulting from the heterodyning action between the cw signal [view (A)] and the local oscillator signal [view (B)] is shown in view (C) as the mixed-frequency signal. ENVELOPE (intelligence) amplitude varies at the BEAT (difference) frequency of 1,000 hertz (501,000 – 500,000). The negative half cycles of the mixed frequency are rectified, as shown in view (D). The peaks of the positive half cycles follow the 1,000-hertz beat frequency.

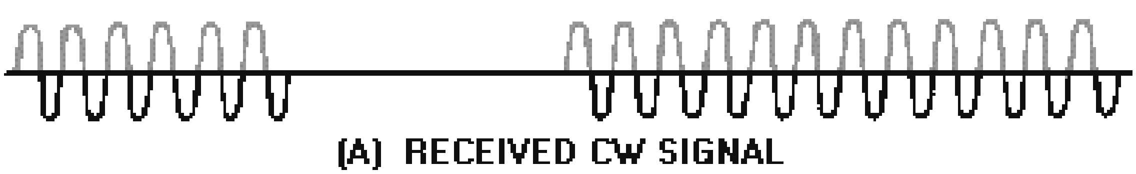


Figure 3-3A.—Heterodyne detection. RECEIVED CW SIGNAL.

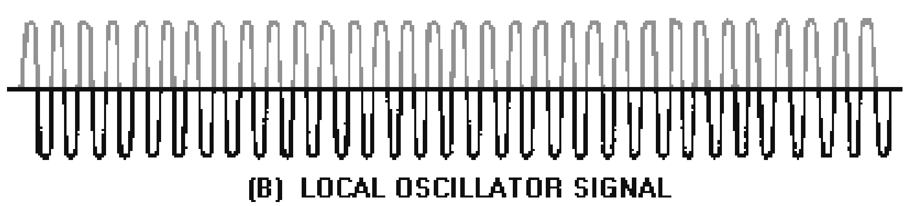


Figure 3-3B.—Heterodyne detection. LOCAL OSCILLATOR SIGNAL.

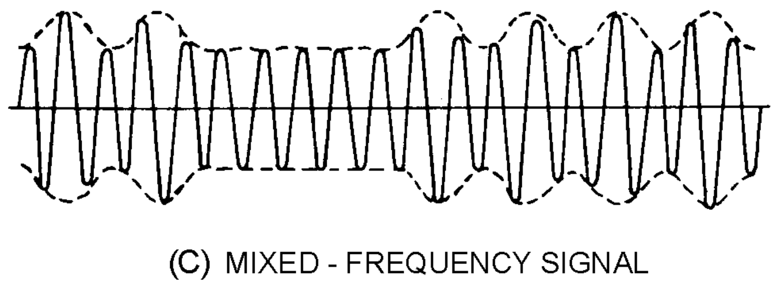


Figure 3-3C.—Heterodyne detection. MIXED-FREQUENCY SIGNAL.

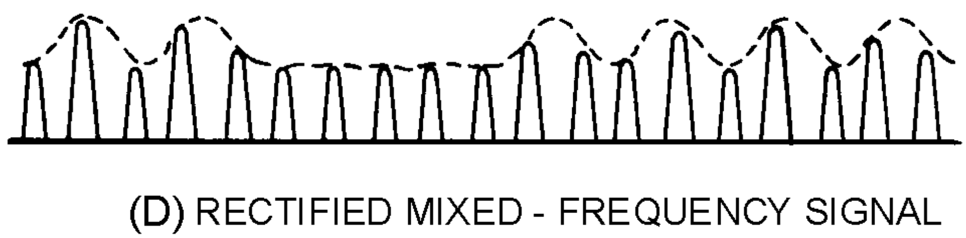


Figure 3-3D.—Heterodyne detection. RECTIFIED MIXED-FREQUENCY SIGNAL.

Figure 3-3E.—Heterodyne detection. AUDIO-BEAT NOTE FROM FILTER.

The cw signal pulsations are removed by the rf filter in the detector output and only the envelope of the rectified pulses remains. The envelope, shown in view (E), is a 1,000-hertz audio-beat note. This 1,000 hertz, dot-dash tone may be heard in a speaker or headphone and identified as the letter **A** by the operator.

The heterodyne method of reception is highly selective and allows little interference from adjacent cw stations. If a cw signal from a radiotelegraph station is operating at 10,000,000 hertz and at the same time an adjacent station is operating at 10,000,300 hertz, a simple detector cannot clearly discriminate between the two stations because the signals are just 300 hertz apart. This is because the bandpass of the tuning circuits is too wide and allows some of the other signal to interfere. The two carrier frequencies differ by only 0.003 percent and a tuned tank circuit cannot easily discriminate between them. However, if a heterodyne detector with a local-oscillator frequency of 10,001,000 hertz is used, then beat notes of 1,000 and 700 hertz are produced by the two signals. These are audio frequencies, which can be distinguished easily by a selective circuit because they differ by 30 percent (compared to the 0.003 percent above).

Even if two stations produce identical beat frequencies, they can be separated by adjusting the local-oscillator or BEAT-FREQUENCY OSCILLATOR (bfo) frequency. For example, if the second station in the previous example had been operating at 10,002,000 hertz, then both stations would have produced a 1,000-hertz beat frequency and interference would have occurred. Adjusting the local-oscillator frequency to 9,999,000 hertz would have caused the desired station at 10,000,000 hertz to produce a 1,000-hertz beat frequency. The other station, at 10,002,000 hertz, would have produced a beat frequency of 3,000 hertz. Either selective circuits or the operator can easily distinguish between these widely differing tones. A trained operator can use the variable local oscillator to distinguish between stations that vary in frequency by only a few hundred hertz.

Q-3. What is the simplest form of cw detector?

Q-4. What are the essential components of a cw receiver system?

Q-5. What principle is used to help distinguish between two cw signals that are close in frequency?

Q-6. How does heterodyning distinguish between cw signals?

REGENERATIVE DETECTOR

A simple, one-transistor REGENERATIVE DETECTOR circuit that uses the heterodyning principle for cw operation is shown in figure 3-4. The circuit can be made to oscillate by increasing the amount of energy fed back to the tank circuit from the collector-output circuit (by physically moving tickler coil L2 closer to L1 using the regeneration control). This feedback overcomes losses in the base-input circuit and causes self-oscillations which are controlled by tuning capacitor C1. The received signal from the antenna and the oscillating frequency are both present at the base of transistor Q1. These two frequencies are heterodyned by the nonlinearity of the transistor. The resulting beat frequencies are then rectified by the emitter-base junction and produce a beat note which is amplified in the collector-output circuit. The af currents in the collector circuit actuate the phones. The REGENERATIVE DETECTOR (figure 3-4) produces its own oscillations, heterodynes them with an incoming signal, and rectifies or detects them.

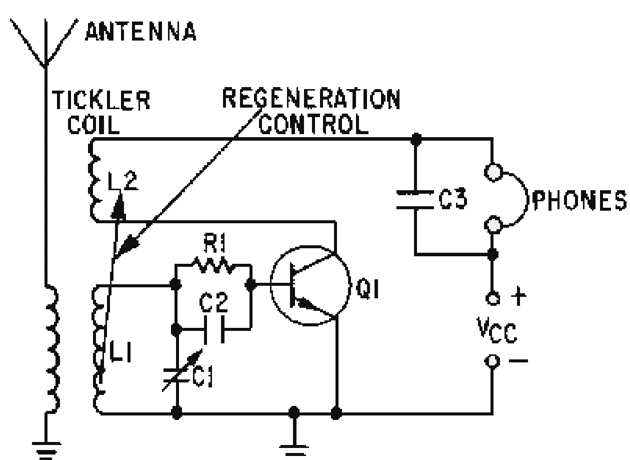


Figure 3-4.—Regenerative detector.

The regenerative detector is used to receive short-wave code signals because it is easy to adjust and has high sensitivity and good selectivity. At high frequencies, the amount of signal detuning necessary to produce an audio-beat note is a small percentage of the signal frequency and causes no trouble. The use of the regenerative detector for low-frequency code reception, however, is usually avoided. At low frequencies the detuning required to produce the proper audio-beat frequency is a considerable percentage of the signal frequency. Although this type detector may be used for AM signals, it has high distortion and is not often used.

Q-7. What simple, one-transistor detector circuit uses the heterodyne principle?

Q-8. What three functions does the transistor in a regenerative detector serve?

AM DEMODULATION

Amplitude modulation refers to any method of modulating an electromagnetic carrier frequency by varying its amplitude in accordance with the message intelligence that is to be transmitted. This is accomplished by heterodyning the intelligence frequency with the carrier frequency. The vector summation of the carrier, sum, and difference frequencies causes the modulation envelope to vary in amplitude at the intelligence frequency, as discussed in chapter 1. In this section we will discuss several circuits that can be used to recover this intelligence from the variations in the modulation envelope.

DIODE DETECTORS

The detection of AM signals ordinarily is accomplished by means of a diode rectifier, which may be either a vacuum tube or a semiconductor diode. The basic detector circuit is shown in its simplest form in view (A) of figure 3-5. Views (B), (C), and (D) show the circuit waveforms. The demodulator must meet three requirements: (1) It must be sensitive to the type of modulation applied at the input, (2) it must be nonlinear, and (3) it must provide filtering. Remember that the AM waveform appears like the diagram of view (B) and the *amplitude* variations of the peaks represent the original audio signal, but *no modulating signal frequencies* exist in this waveform. The waveform contains only three rf frequencies: (1) the *carrier* frequency, (2) the *sum* frequency, and (3) the *difference* frequency. The modulating intelligence is contained in the *difference* between these frequencies. The vector addition of these frequencies provides the modulation envelope which approximates the original modulating waveform. It is this modulation envelope that the DIODE DETECTORS use to reproduce the original modulating frequencies.

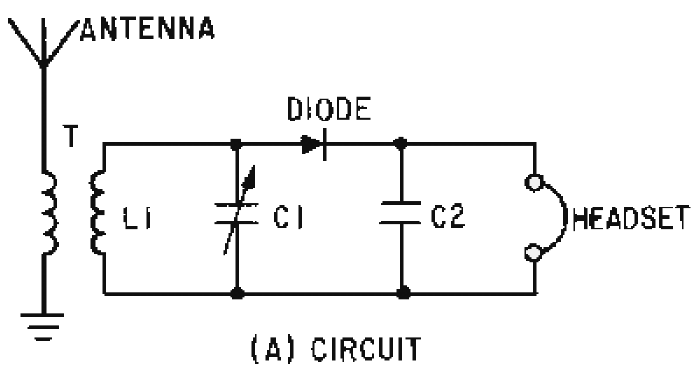


Figure 3-5A.—Series-diode detector and wave shapes. CIRCUIT.

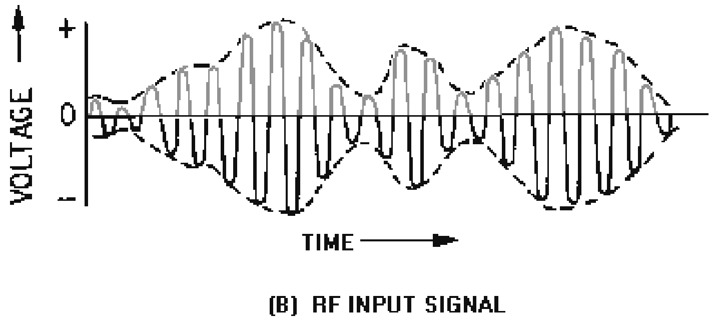


Figure 3-5B.—Series-diode detector and wave shapes. RF INPUT SIGNAL.

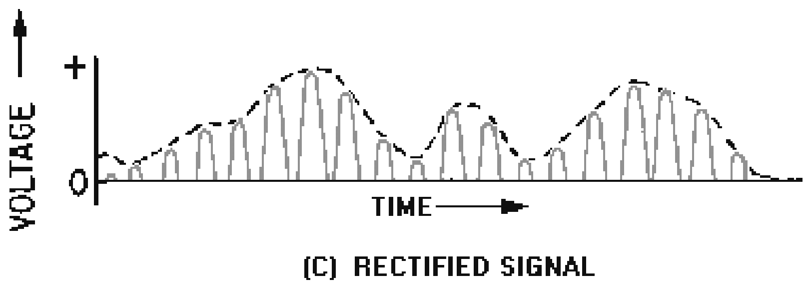


Figure 3-5C.—Series-diode detector and wave shapes. RECTIFIED SIGNAL.

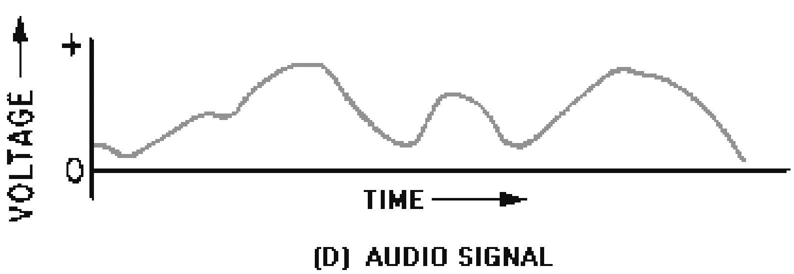


Figure 3-5D.—Series-diode detector and wave shapes. AUDIO SIGNAL.

Series-Diode Detector

Let's analyze the operation of the circuit shown in view (A) of figure 3-5. This circuit is the basic type of diode receiver and is known as a SERIES-DIODE DETECTOR. The circuit consists of an antenna, a tuned LC tank circuit, a semiconductor diode detector, and a headset which is bypassed by capacitor C2. The antenna receives the transmitted rf energy and feeds it to the tuned tank circuit. This tank circuit (L1 and C1) selects which rf signal will be detected. As the tank resonates at the selected frequency, the wave shape in view (B) is developed across the tank circuit. Because the semiconductor is a nonlinear device, it conducts in only one direction. This eliminates the negative portion of the rf carrier and produces the signal shown in view (C). The current in the circuit must be smoothed before the headphones can reproduce the af intelligence. This action is achieved by C2 which acts as a filter to provide an output that is proportional to the peak rf pulses. The filter offers a low impedance to rf and a relatively high impedance to af. (Filters were discussed in *NEETS*, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*.) This action causes C2 to develop the waveform in view (D). This varying af voltage is applied to the headset which then reproduces the original modulating frequency. This circuit is called a series-diode detector (sometimes referred to as a VOLTAGE-DIODE DETECTOR) because the semiconductor diode is in series with both the input voltage and the load impedance. Voltages in the circuit cause an output voltage to develop across the load impedance that is proportional to the input voltage peaks of the modulation envelope.

Q-9. What are the three requirements for an AM demodulator?

Q-10. What does the simplest diode detector use to reproduce the modulating frequency?

Q-11. What is the function of the diode in a series-diode detector?

Q-12. In figure 3-5, what is the function of C2?

Shunt-Diode Detector

The SHUNT-DIODE DETECTOR (figure 3-6) is similar to the series-diode detector except that the output variations are current pulses rather than voltage pulses. Passing this current through a shunt resistor develops the voltage output. The input is an rf modulated envelope. On the negative half cycles of the rf, diode CR1 is forward biased and shunts the signal to ground. On the positive half cycles, current flows from the output through L1 to the input. A field is built up around L1 that tends to keep the current flowing. This action integrates the rf current pulses and causes the output to follow the modulation envelope (intelligence) closely. (Integration was discussed in *NEETS*, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*.) Shunt resistor R1 develops the output voltage from this current flow. Although the shunt detector operates on the principle of current flow, it is the output voltage across the shunt resistor that is used to reproduce the original modulation signal. The shunt-diode detector is easily identified by noting that the detector diode is in parallel with both the input and load impedance. The waveforms associated with this detector are identical to those shown in views (B), (C), and (D) of figure 3-5.

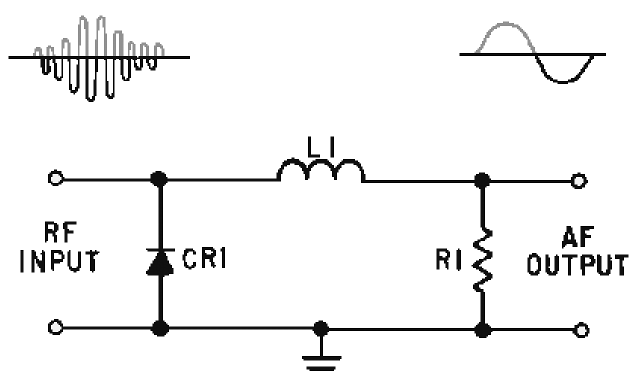


Figure 3-6.—Shunt-diode detector.

The series-diode detector is normally used where large input signals are supplied and a linear output is required. The shunt-diode detector is used where the voltage variations are too small to produce a full output from audio amplifier stages. Additional current amplifiers are required to bring the output to a usable level. Other methods of detection and amplification have been developed which will detect low-level signals. The next sections will discuss two of these circuits, the common-emitter and common-base detectors.

Q-13. How does the current-diode detector differ from the voltage-diode detector?

Q-14. Under what circuit conditions would the shunt detector be used?

COMMON-EMITTER DETECTOR

The COMMON-EMITTER DETECTOR is often used in receivers to supply an amplified detected output. The schematic for a typical transistor common-emitter detector is shown in figure 3-7. Input transformer T1 has a tuned primary that acts as a frequency-selective device. L2 inductively couples the input modulation envelope to the base of transistor Q1. Resistors R1 and R2 are fixed-bias voltage dividers that set the bias levels for Q1. Resistor R1 is bypassed by C2 to eliminate rf. This RC combination also acts as the load for the diode detector (emitter-base junction of Q1). The detected audio is in series with the biasing voltage and controls collector current. The output is developed across R4 which is also bypassed to remove rf by C4. R3 is a temperature stabilization resistor and C3 bypasses it for both rf and af.

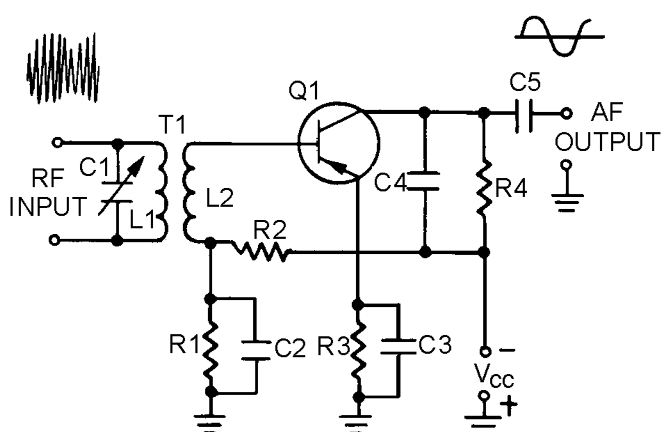


Figure 3-7.—Common-emitter detector.

Q1 is biased for slight conduction with no input signal applied. When an input signal appears on the base of Q1, it is rectified by the emitter-base junction (operating as a diode) and is developed across R1 as a dc bias voltage with a varying af component. This voltage controls bias and collector current for Q1. The output is developed by collector current flow through R4. Any rf ripple in the output is bypassed across the collector load resistor by capacitor C4. The af variations are not bypassed. After the modulation envelope is detected in the base circuit, it is amplified in the output circuit to provide suitable af output. The output of this circuit is higher than is possible with a simple detector. Because of the amplification in this circuit, weaker signals can be detected than with a simple detector. A higher, more usable output is thus developed.

Q-15. Which junction of the transistor in the common-emitter detector detects the modulation envelope?

Q-16. Which component in figure 3-7 develops the af signal at the input?

Q-17. How is the output signal developed in the common-emitter detector?

COMMON-BASE DETECTOR

Another amplifying detector that is used in portable receivers is the COMMON-BASE DETECTOR. In this circuit detection occurs in the emitter-base junction and amplification occurs at the output of the collector junction. The output developed is the equivalent of a diode detector which is followed by a stage of audio amplification, but with more distortion. Figure 3-8 is a schematic of a typical common-base detector. Transformer T1 is tuned by capacitor C3 to the frequency of the incoming modulated envelope. Resistor R1 and capacitor C1 form a self-biasing network which sets the dc operating point of the emitter junction. The af output is taken from the collector circuit through audio transformer T2. The primary of T2 forms the detector output load and is bypassed for rf by capacitor C2.

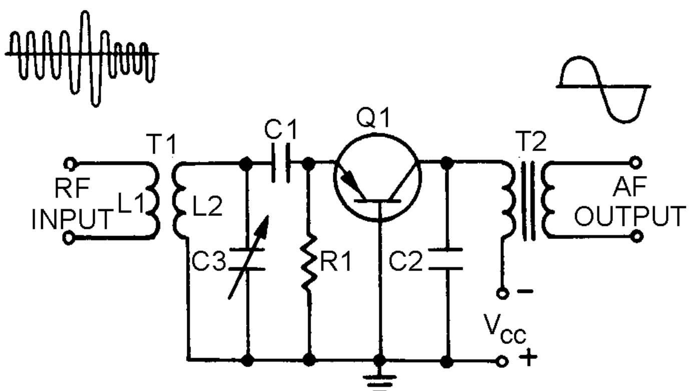


Figure 3-8.—Common-base detector.

The input signal is coupled through T1. When capacitor C3 is tuned to the proper frequency, the signal is passed to the emitter of Q1. When no input signal is present, bias is determined by resistor R1. When the input signal becomes positive, current flows through the emitter-base junction causing it to be forward biased. C1 and R1 establish the dc operating point by acting as a filter network. This action provides a varying dc voltage that follows the peaks of the rf modulated envelope. This action is identical to the diode detector with the emitter-base junction doing the detecting. The varying dc voltage on the emitter changes the bias on Q1 and causes collector current to vary in accordance with the detected voltage. Transformer T2 couples these af current changes to the output. Thus, Q1 detects the AM wave and then provides amplification for the detected waveform.

The four AM detectors just discussed are not the only types that you will encounter. However, they are representative of most AM detectors and the same characteristics will be found in all AM detectors. Now let's study some ways of demodulating frequency-modulated (fm) signals.

Q-18. Which junction acts as the detector in a common-base detector?

Q-19. To what circuit arrangement is a common-base detector equivalent?

Q-20. In figure 3-8, which components act as the filter network in the diode detector?

FM DEMODULATION

In fm demodulators, the intelligence to be recovered is not in amplitude variations; it is in the *variation of the instantaneous frequency* of the carrier, either above or below the center frequency. The detecting device must be constructed so that its output amplitude will vary linearly according to the instantaneous frequency of the incoming signal.

Several types of fm detectors have been developed and are in use, but in this section you will study three of the most common: (1) the phase-shift detector, (2) the ratio detector, and (3) the gated-beam detector.

SLOPE DETECTION

To be able to understand the principles of operation for fm detectors, you need to first study the simplest form of frequency-modulation detector, the SLOPE DETECTOR. The slope detector is essentially a tank circuit which is tuned to a frequency either slightly above or below the fm carrier frequency. View (A) of figure 3-9 is a plot of voltage versus frequency for a tank circuit. The resonant frequency of the tank is the frequency at point 4. Components are selected so that the resonant frequency is higher than the frequency of the fm carrier signal at point 2. The entire frequency deviation for the fm signal falls on the lower slope of the bandpass curve between points 1 and 3. As the fm signal is applied to the tank circuit in view (B), the output amplitude of the signal varies as its frequency swings closer to, or further from, the resonant frequency of the tank. Frequency variations will still be present in this waveform, but it will also develop amplitude variations, as shown in view (B). This is because of the response of the tank circuit as it varies with the input frequency. This signal is then applied to the diode detector in view (C) and the detected waveform is the output. This circuit has the major disadvantage that any amplitude variations in the rf waveform will pass through the tank circuit and be detected. This disadvantage can be eliminated by placing a limiter circuit before the tank input. (Limiter circuits were discussed in *NEETS*, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*.) This circuit is basically the same as an AM detector with the tank tuned to a higher or lower frequency than the received carrier.

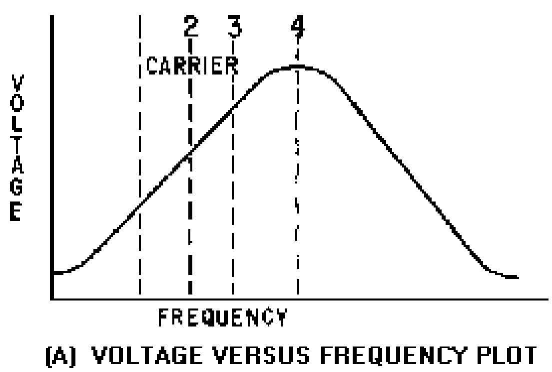


Figure 3-9A.—Slope detector. VOLTAGE VERSUS FREQUENCY PLOT.

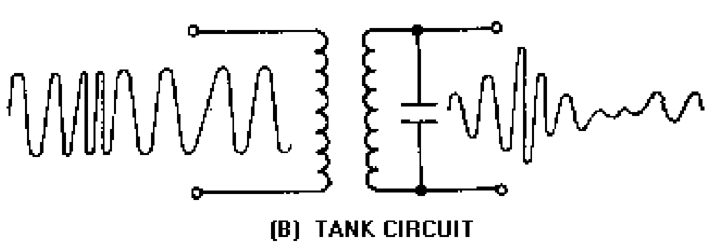


Figure 3-9B.—Slope detector. TANK CIRCUIT.

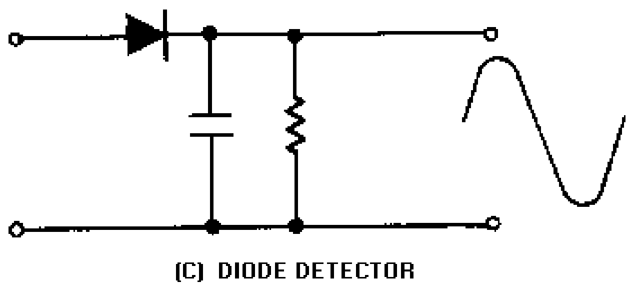


Figure 3-9C.—Slope detector. DIODE DETECTOR.

Q-21. What is the simplest form of fm detector?

Q-22. What is the function of an fm detector?

FOSTER-SEELEY DISCRIMINATOR

The FOSTER-SEELEY DISCRIMINATOR is also known as the PHASE-SHIFT DISCRIMINATOR. It uses a double-tuned rf transformer to convert frequency variations in the received fm signal to amplitude variations. These amplitude variations are then rectified and filtered to provide a dc output voltage. This voltage varies in both amplitude and polarity as the input signal varies in frequency. A typical discriminator response curve is shown in figure 3-10. The output voltage is 0 when the input frequency is equal to the carrier frequency (f_r). When the input frequency rises above the center frequency, the output increases in the positive direction. When the input frequency drops below the center frequency, the output increases in the negative direction.

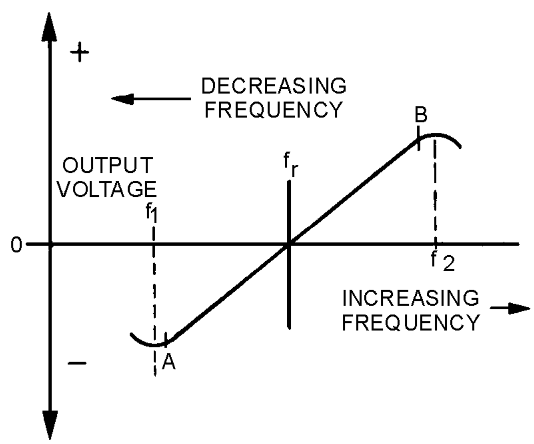


Figure 3-10.—Discriminator response curve.

The output of the Foster-Seeley discriminator is affected not only by the input frequency, but also to a certain extent by the input amplitude. Therefore, using limiter stages before the detector is necessary.

Circuit Operation of a Foster-Seeley Discriminator

View (A) of figure 3-11 shows a typical Foster-Seeley discriminator. The collector circuit of the preceding limiter/amplifier circuit (Q1) is shown. The limiter/amplifier circuit is a special amplifier circuit which limits the amplitude of the signal. This limiting keeps interfering noise low by removing excessive amplitude variations from signals. The collector circuit tank consists of C1 and L1. C2 and L2 form the secondary tank circuit. Both tank circuits are tuned to the center frequency of the incoming fm signal. Choke L3 is the dc return path for diode rectifiers CR1 and CR2. R1 and R2 are not always necessary but are usually used when the back (reverse bias) resistance of the two diodes is different. Resistors R3 and R4 are the load resistors and are bypassed by C3 and C4 to remove rf. C5 is the output coupling capacitor.

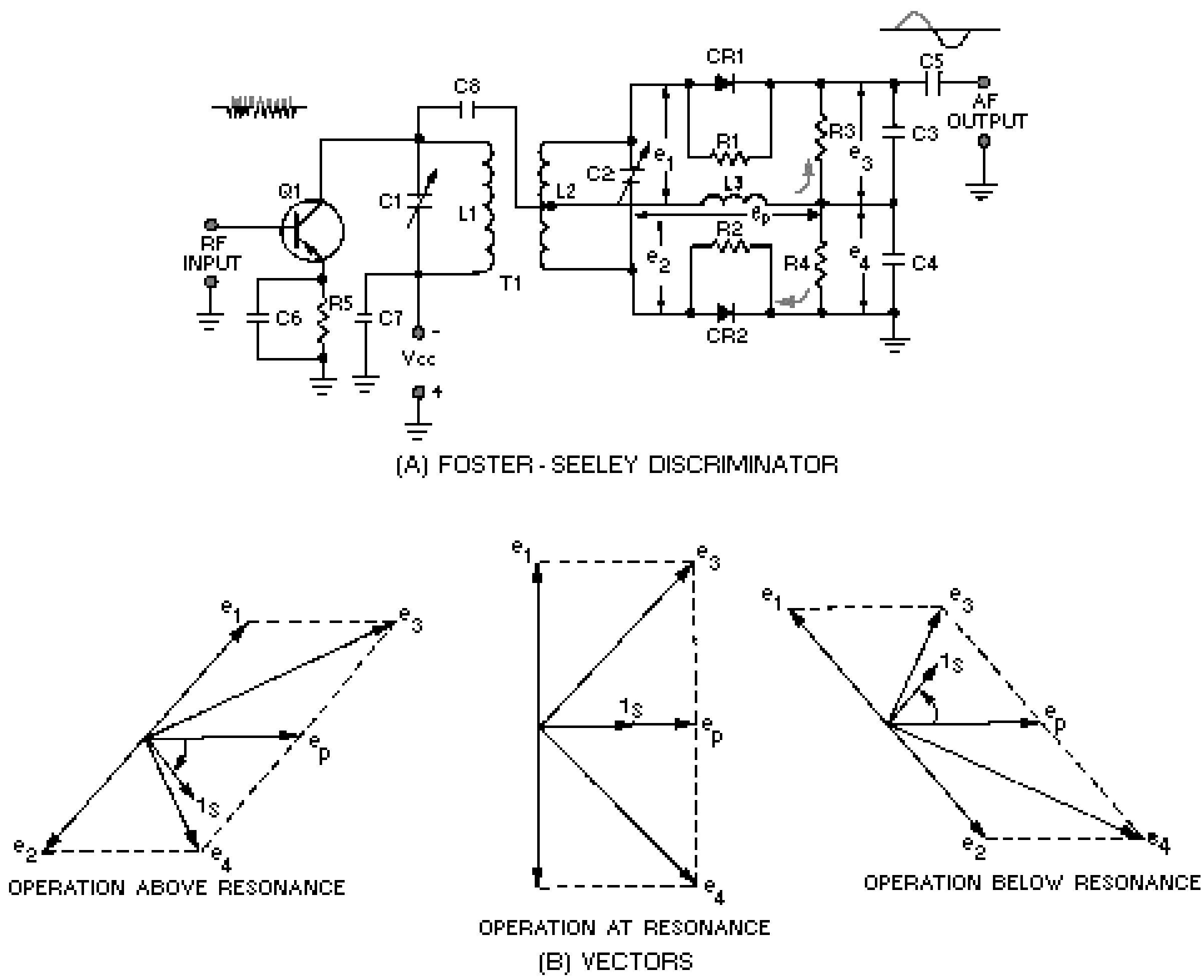


Figure 3-11.—Foster-Seeley discriminator. FOSTER-SEELEY DISCRIMINATOR.

CIRCUIT OPERATION AT RESONANCE.—The operation of the Foster-Seeley discriminator can best be explained using vector diagrams [figure 3-11, view (B)] that show phase relationships between the voltages and currents in the circuit. Let's look at the phase relationships when the *input frequency is equal to the center frequency* of the resonant tank circuit.

The input signal applied to the primary tank circuit is shown as vector e_p. Since coupling capacitor C8 has negligible reactance at the input frequency, rf choke L3 is effectively in parallel with the primary tank circuit. Also, because L3 is effectively in parallel with the primary tank circuit, input voltage e_p also appears across L3. With voltage e_p applied to the primary of T1, a voltage is induced in the secondary which causes current to flow in the secondary tank circuit. When the input frequency is equal to the center frequency, the tank is at resonance and acts resistive. Current and voltage are in phase in a resistance circuit, as shown by is and e_p. The current flowing in the tank causes voltage drops across each half of the balanced secondary winding of transformer T1. These voltage drops are of equal amplitude and opposite polarity with respect to the center tap of the winding. Because the winding is inductive, the voltage across it is 90 degrees out of phase with the current through it. Because of the center-tap arrangement, the voltages at each end of the secondary winding of T1 are 180 degrees out of phase and are shown as e_1 and e_2 on the vector diagram.

The voltage applied to the anode of CR1 is the vector sum of voltages e_p and e_1, shown as e_3 on the diagram. Likewise, the voltage applied to the anode of CR2 is the vector sum of voltages e_p and e_2, shown as e_4 on the diagram. At resonance e_3 and e_4 are equal, as shown by vectors of the same length. Equal anode voltages on diodes CR1 and CR2 produce equal currents and, with equal load resistors, equal and opposite voltages will be developed across R3 and R4. The output is taken across R3 and R4 and will be 0 at resonance since these voltages are equal and of appositive polarity.

The diodes conduct on opposite half cycles of the input waveform and produce a series of dc pulses at the rf rate. This rf ripple is filtered out by capacitors C3 and C4.

OPERATION ABOVE RESONANCE.—A phase shift occurs when an *input frequency higher than the center frequency* is applied to the discriminator circuit and the current and voltage phase relationships change. When a series-tuned circuit operates at a frequency above resonance, the inductive reactance of the coil increases and the capacitive reactance of the capacitor decreases. Above resonance the tank circuit acts like an inductor. Secondary current lags the primary tank voltage, e_p. Notice that secondary voltages e_1 and e_2 are still 180 degrees out of phase with the current (i_S) that produces them. The change to a lagging secondary current rotates the vectors in a clockwise direction. This causes el to become more in phase with e_p while e_2 is shifted further out of phase with e_p. The vector sum of e_p and e_2 is less than that of e_p and e_1. Above the center frequency, diode CR1 conducts more than diode CR2. Because of this heavier conduction, the voltage developed across R3 is greater than the voltage developed across R4; the output voltage is positive.

OPERATION BELOW RESONANCE.—When the *input frequency is lower than the center frequency*, the current and voltage phase relationships change. When the tuned circuit is operated at a frequency lower than resonance, the capacitive reactance increases and the inductive reactance decreases. Below resonance the tank acts like a capacitor and the secondary current leads primary tank voltage e_p. This change to a leading secondary current rotates the vectors in a *counterclockwise* direction. From the vector diagram you should see that e_2 is brought nearer in phase with e_p, while el is shifted further out of phase with e_p. The vector sum of e_p and e_2 is larger than that of e and e_1. Diode CR2 conducts more than diode CR1 below the center frequency. The voltage drop across R4 is larger than that across R3 and the output across both is negative.

Disadvantages

These voltage outputs can be plotted to show the response curve of the discriminator discussed earlier (figure 3-10). When weak AM signals (too small in amplitude to reach the circuit limiting level) pass through the limiter stages, they can appear in the output. These unwanted amplitude variations will cause primary voltage e_p [view (A) of figure 3-11] to fluctuate with the modulation and to induce a similar voltage in the secondary of T1. Since the diodes are connected as half-wave rectifiers, these small AM signals will be detected as they would be in a diode detector and will appear in the output. This unwanted AM interference is cancelled out in the ratio detector (to be studied next in this chapter) and is the main disadvantage of the Foster-Seeley circuit.

Q-23. What type of tank circuit is used in the Foster-Seeley discriminator?

Q-24. What is the purpose of CR1 and CR2 in the Foster-Seeley discriminator?

Q-25. What type of impedance does the tank circuit have above resonance?

RATIO DETECTOR

The RATIO DETECTOR uses a double-tuned transformer to convert the instantaneous frequency variations of the fm input signal to instantaneous amplitude variations. These amplitude variations are then rectified to provide a dc output voltage which varies in amplitude and polarity with the input signal frequency. This detector demodulates fm signals and suppresses amplitude noise without the need of limiter stages.

Circuit Operation

Figure 3-12 shows a typical ratio detector. The input tank capacitor (C1) and the primary of transformer T1 (L1) are tuned to the center frequency of the fm signal to be demodulated. The secondary winding of T1 (L2) and capacitor C2 also form a tank circuit tuned to the center frequency. Tertiary (third) winding L3 provides additional inductive coupling which reduces the loading effect of the secondary on the primary circuit. Diodes CR1 and CR2 rectify the signal from the secondary tank. Capacitor C5 and resistors R1 and R2 set the operating level of the detector. Capacitors C3 and C4 determine the amplitude and polarity of the output. Resistor R3 limits the peak diode current and furnishes a dc return path for the rectified signal. The output of the detector is taken from the common connection between C3 and C4. Resistor R_L is the load resistor. R5, C6, and C7 form a low-pass filter to the output.

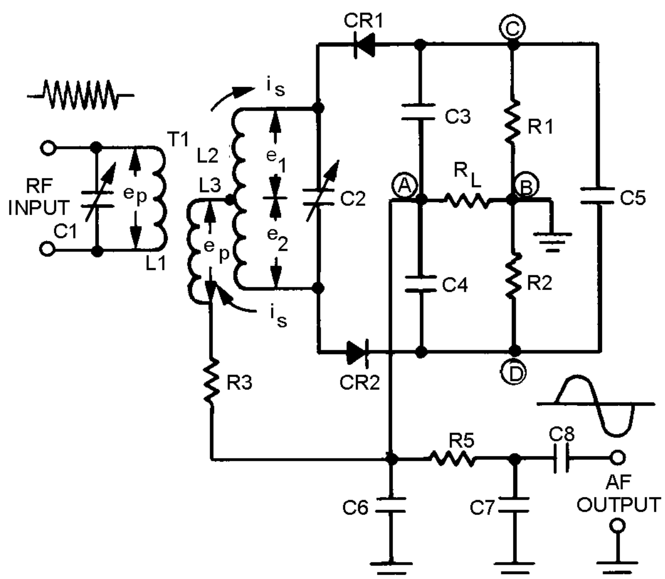


Figure 3-12.—Ratio detector.

This circuit operates on the same principles of phase shifting as did the Foster-Seeley discriminator. In that discussion, vector diagrams were used to illustrate the voltage amplitudes and polarities for conditions at resonance, above resonance, and below resonance. The same vector diagrams apply to the ratio detector but will not be discussed here. Instead, you will study the resulting current flows and polarities on simplified schematic diagrams of the detector circuit.

OPERATION AT RESONANCE.—When the input voltage e_p is applied to the primary in figure 3-12 it also appears across L3 because, by inductive coupling, it is effectively connected in parallel with the primary tank circuit. At the same time, a voltage is induced in the secondary winding and causes current to flow around the secondary tank circuit. At resonance the tank acts like a resistive circuit; that is, the tank current is in phase with the primary voltage e_p. The current flowing in the tank circuit causes voltages e_1 and e_2 to be developed in the secondary winding of T1. These voltages are of equal magnitude and of opposite polarity with respect to the center tap of the winding. Since the winding is inductive, the voltage drop across it is 90 degrees out of phase with the current through it.

Figure 3-13 is a simplified schematic diagram of a ratio detector at resonance. The voltage applied to the cathode of CR1 is the vector sum of e_1 and e_p. Likewise, the voltage applied to the anode of CR2 is the vector sum of e_2 and e_p. No phase shift occurs at resonance and both voltages are equal. Both diodes conduct equally. This equal current flow causes the same voltage drop across both R1 and R2. C3 and C4 will charge to equal voltages with opposite polarities. Let's assume that the voltages across C3 and C4 are equal in amplitude (5 volts) and of opposite polarity and the total charge across C5 is 10 volts. R1 and R2 will each have 5 volts dropped across them because they are of equal values. The output is taken between points **A** and **B**. To find the output voltage, you algebraically add the voltages between points **A** and **B** (loop **ACB** or **ADB**). Point **A** to point **D** is −5 volts. Point **D** to point **B** is + 5 volts. Their algebraic sum is 0 volts and the output voltage is 0 at resonance. If the voltages on branch **ACB** were figured, the same output would be found because the circuit branches are in parallel.

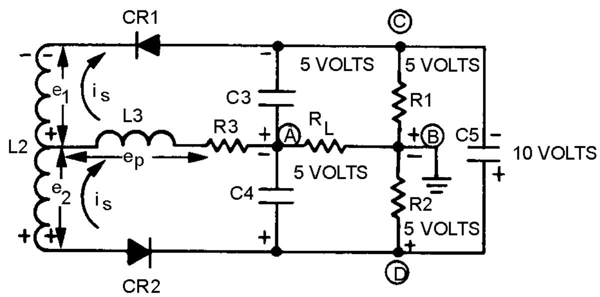


Figure 3-13.—Current flow and polarities at resonance.

When the input signal reverses polarity, the secondary voltage across L2 also reverses. The diodes will be reverse biased and no current will flow. Meanwhile, C5 retains most of its charge because of the long time constant offered in combination with R1 and R2. This slow discharge helps to maintain the output.

OPERATION ABOVE RESONANCE.—When a tuned circuit (figure 3-14) operates at a frequency higher than resonance, the tank is inductive. The secondary current i lags the primary voltage e_p. Secondary voltage e_1 is nearer in phase with primary voltage e, while e_2 is shifted further out of phase with e_p. The vector sum of e_1 and e_p is larger than that of e_2 and e_p. Therefore, the voltage applied to the cathode of CR1 is greater than the voltage applied to the anode of CR2 above resonance.

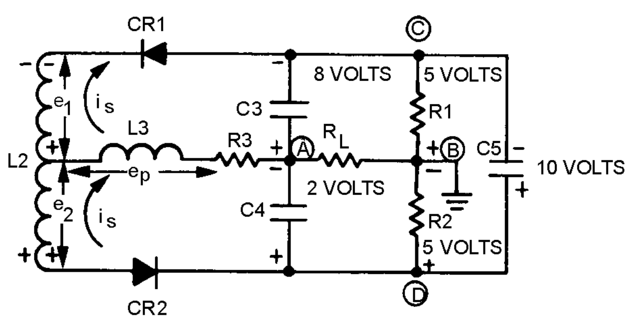


Figure 3-14.—Current flow and polarities above resonance.

Assume that the voltages developed above resonance are such that the higher voltage on the cathode of CR1 causes C3 to charge to 8 volts. The lower voltage on the anode of CR2 causes C4 to charge to 2 volts. Capacitor C5 remains charged to the sum of these two voltages, 10 volts. Again, by adding the voltages in loop **ACB** or **ADB** between points **A** and **B**, you can find the output voltage. Point **A** to point **D** equals -2 volts. Point **D** to point **B** equals +5 volts. Their algebraic sum, and the output, equals +3 volts when tuned above resonance. During the negative half cycle of the input signal, the diodes are reverse biased and C5 helps maintain a constant output.

OPERATION BELOW RESONANCE.—When a tuned circuit operates below resonance (figure 3-15), it is capacitive. Secondary current $_{is}$ leads the primary voltage e_p and secondary voltage e_2 is nearer in phase with primary voltage e_p. The vector sum of e_2 and e_p is larger than the sum of e_1 and e_p. The voltage applied to the anode of CR2 becomes greater than the voltage applied to the cathode of CR1 below resonance.

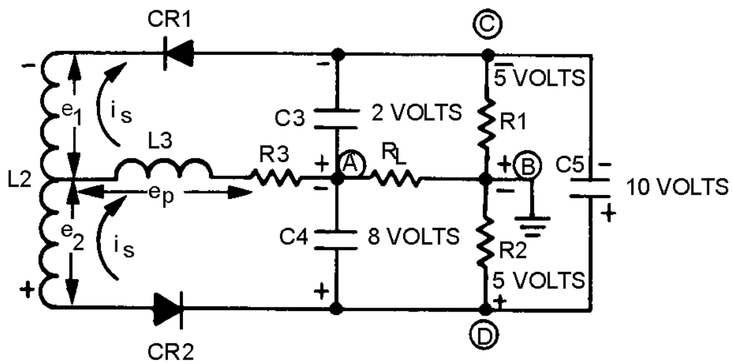


Figure 3-15.—Current flow and polarities below resonance.

Assume that the voltages developed below resonance are such that the higher voltage on the anode of CR2 causes C4 to charge to 8 volts. The lower voltage on the cathode of CR1 causes C3 to charge to 2 volts. Capacitor C5 remains charged to the sum of these two voltages, 10 volts. The output voltage equals –8 volts plus +5 volts, or –3 volts, when tuned below resonance. During the negative half cycle of the input signal, the diodes are reverse biased and C5 helps maintain a constant output.

Advantage of a Ratio Detector

The ratio detector is not affected by amplitude variations on the fm wave. The output of the detector adjusts itself automatically to the average amplitude of the input signal. C5 charges to the sum of the voltages across R1 and R2 and, because of its time constant, tends to filter out any noise impulses. Before C5 can charge or discharge to the higher or lower potential, the noise disappears. The difference in charge across C5 is so slight that it is not discernible in the output. Ratio detectors can operate with as little as 100 millivolts of input. This is much lower than that required for limiter saturation and less gain is required from preceding stages.

Q-26. What is the primary advantage of a ratio detector?

Q-27. What is the purpose of C5 in figure 3-12?

GATED-BEAM DETECTOR

An fm demodulator employing a completely different detection principle is the GATED-BEAM DETECTOR (sometimes referred to as the QUADRATURE DETECTOR). A simplified diagram of a gated-beam detector is shown in figure 3-16. It uses a gated-beam tube to limit, detect, and amplify the received fm signal. The output voltage is 0 when the input frequency is *equal to the center frequency*. When the input frequency rises *above the center frequency*, the output voltage goes positive. When the input frequency drops *below the center frequency*, the output voltage goes negative.

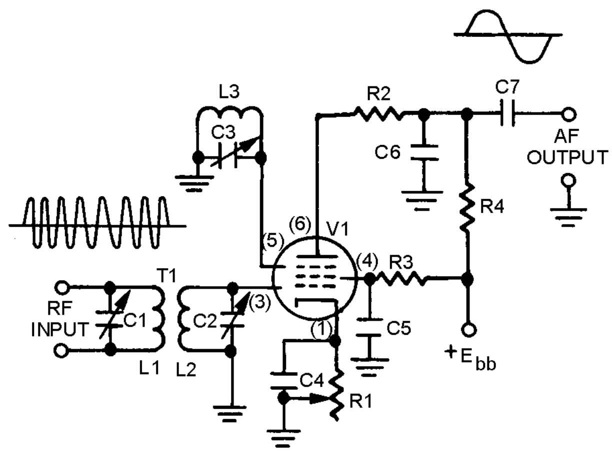


Figure 3-16.—Gated-beam detector.

Circuit Operation

The gated-beam detector employs a specially designed gated-beam tube. The elements of this tube are shown in figure 3-17. The focus electrode forms a shield around the tube cathode except for a narrow slot through which the electron beam flows. The beam of electrons flows toward the limiter grid which acts like a gate. When the gate is open, the electron beam flows through to the next grid. When closed, the gate completely stops the beam.

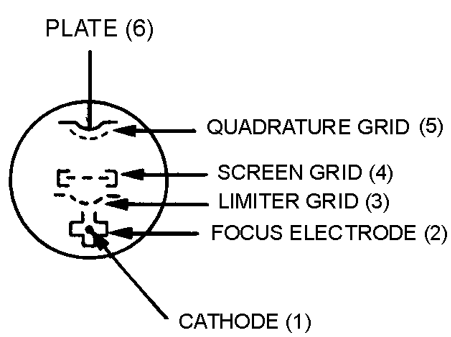


Figure 3-17.—Gated-beam tube physical layout.

After the electron beam passes the limiter grid, the screen grid refocuses the beam toward the quadrature grid. The quadrature grid acts much the same as the limiter grid; it either opens or closes the passage for electrons. These two grids act similar to an AND gate in digital devices; both gates must be open for the passage of electrons to the plate. Either grid can cut off plate current. AND gates were presented in *NEETS*, Module 13, *Introduction to Number Systems, Boolean Algebra, and Logic Circuits*.

Look again at the circuit in figure 3-16. With no signal applied to the limiter grid (3), the tube conducts. The electron beam moving near the quadrature grid (5) induces a current into the grid which develops a voltage across the high-Q tank circuit (L3 and C3). C3 charges until it becomes sufficiently negative to cut off the current flow. L3 tends to keep the current moving and, as its field collapses, discharges C3. When C3 discharges sufficiently, the quadrature grid becomes positive, grid current flows, and the cycle repeats itself. This tank circuit (L3 and C3) is tuned to the center frequency of the received fm signal so that it will oscillate at that frequency.

The waveforms for the circuit are shown in figure 3-18. View (A) is the fm input signal. The limiter-grid gate action creates a wave shape like view (B) because the tube is either cut off or saturated very quickly by the input wave. Note that this is a square wave and is the current waveform passing the limiter grid.

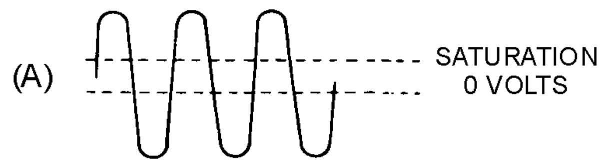


Figure 3-18A.—Gated-beam detector waveforms.

Figure 3-18B.—Gated-beam detector waveforms.

At the quadrature grid the voltage across C3 lags the current which produces it [view (C)]. The result is a series of pulses, shown in view (D), appearing on the quadrature grid at the center frequency, but lagging the limiter-grid voltage by 90 degrees. Because the quadrature grid has the same conduction and cutoff levels as the limiter grid, the resultant current waveform will be transformed into a square wave.

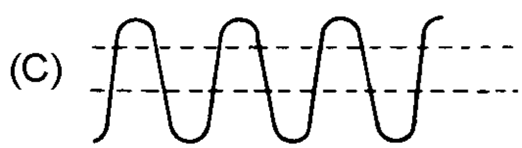


Figure 3-18C.—Gated-beam detector waveforms.

Figure 3-18D.—Gated-beam detector waveforms.

Both the limiter and quadrature grids must be positive at the same time to have plate current. You can see how much conduction time occurs for each cycle of the input by overlaying the current waveforms in views (B) and (D), as shown in view (E). The times when both grids are positive are shown by the shaded area of view (E). These plate current pulses are shown for operation at resonance in view (F).

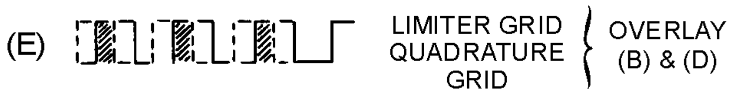


Figure 3-18E.—Gated-beam detector waveforms.

Figure 3-18F.—Gated-beam detector waveforms.

Now consider what happens with a deviation in frequency at the input. If the frequency increases, the frequency across the quadrature tank also increases. Above resonance, the tank appears capacitive to the induced current; voltage then lags the applied voltage by more than 90 degrees, as shown in view (G). Note in view (H) that the two grid signals have moved more out of phase and the average plate current level has decreased.

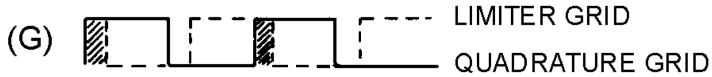


Figure 3-18G.—Gated-beam detector waveforms.

Figure 3-18H.—Gated-beam detector waveforms.

As the input frequency decreases, the opposite action takes place. The two grid signals move more in phase, as shown in view (I), and the average plate current increases, as shown in view (J).

Figure 3-18I.—Gated-beam detector waveforms.

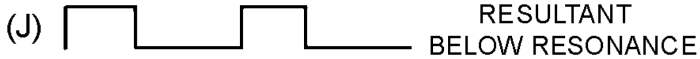


Figure 3-18J.—Gated-beam detector waveforms.

View (K) shows the resultant plate-current pulses when an fm signal is applied to a gated-beam detector. Plate load resistor R4 and capacitor C6 form an integrating network which filters these pulses to form the sine-wave output.

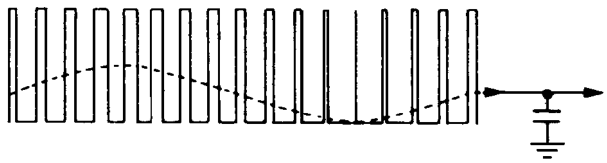


Figure 3-18K.—Gated-beam detector waveforms.

Advantages of the Gated-Beam Detector

The primary advantage of the gated-beam detector lies in its extreme simplicity. It employs only one tube, yet provides a very effective limiter with linear detection. It requires relatively few components and is very easily adjusted.

There are more than the three types of fm demodulators presented in this chapter. However, these are representative of the types with which you will be working. The principles involved in their operation are similar to the other types. You will now briefly study PHASE DEMODULATION which uses the same basic circuitry as fm demodulators.

Q-28. What circuit functions does the tube in a gated-beam detector serve?

Q-29. What condition must exist on both the limiter and quadrature grids for current to flow in a gated-beam detector?

Q-30. Name two advantages of the gated-beam detector.

PHASE DEMODULATION

In phase modulation (pm) the intelligence is contained in the *amount and rate of phase shift* in a carrier wave. You should recall from your study of pm that there is an incidental shift in frequency as the phase of the carrier is shifted. Because of this incidental frequency shift, fm demodulators, such as the Foster-Seeley discriminator and the ratio detector, can also be used to demodulate phase-shift signals.

Another circuit that may be used is the gated-beam (quadrature) detector. Remember that the fm phase detector output was determined by the phase of the signals present at the grids. A QUADRATURE DETECTOR FOR PHASE DEMODULATION works in the same manner.

A basic schematic is shown in figure 3-19. The quadrature-grid signal is excited by a reference from the transmitter. This may be a sample of the unmodulated master oscillator providing a phase reference for the detector.

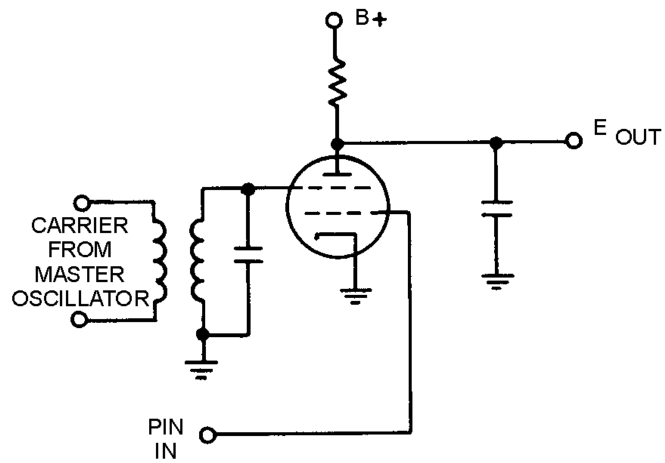


Figure 3-19.—Phase detector.

The modulated waveform is applied to the limiter grid. Gating action in the tube will occur as the phase shifts between the input waveform and the reference. The combined output current from the gated-beam tube will be a series of current pulses. These pulses will vary in width as shown in figure 3-20. The width of these pulses will vary in accordance with the phase difference between the carrier and the modulated wave.

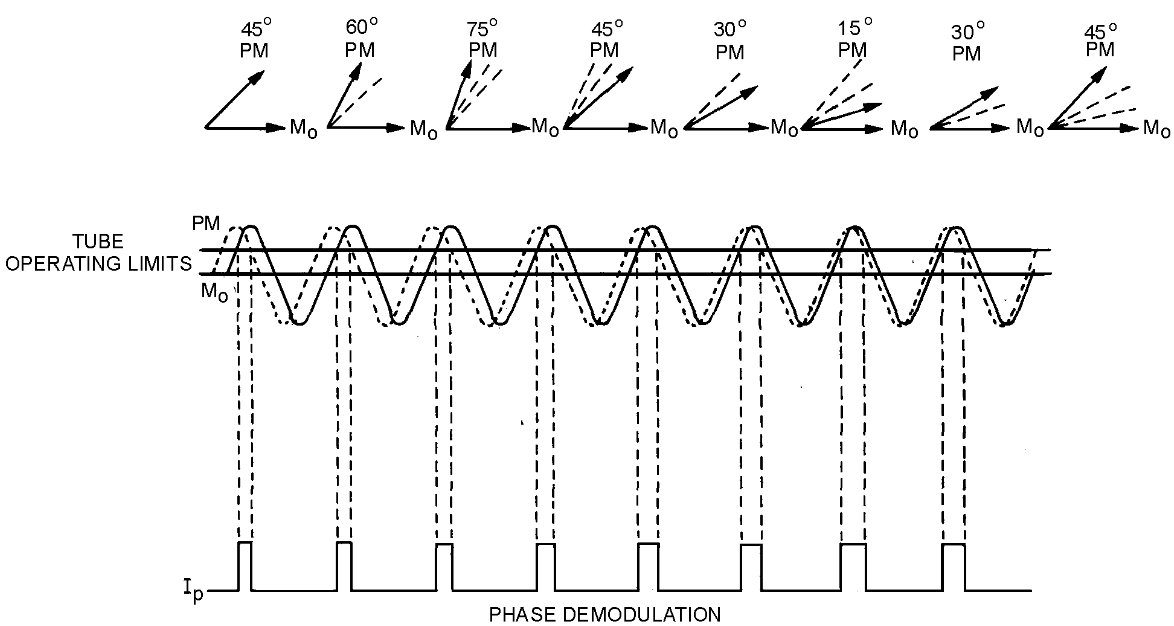


Figure 3-20.—Phase-detector waveforms.

Q-31. Where is the intelligence contained in a phase-modulated signal?

Q-32. Why can phase-modulated signals be detected by fm detectors?

Q-33. How is a quadrature detector changed when used for phase demodulation?

PULSE DEMODULATION

Pulse modulation is used in radar circuits as well as communications circuits, as discussed in chapter 2. A pulse-modulated signal in radar may be detected by a simple circuit that detects the presence of rf energy. Circuits that are capable of this were covered in this chapter in the cw detection discussion; therefore, the information will not be repeated here. A RADAR DETECTOR, in its simplest form, must be capable of producing an output when rf energy (reflected from a target) is present at its input.

In COMMUNICATIONS PULSE DETECTORS the modulated waveform must be restored to its original form. In this chapter you will study three basic methods of pulse demodulation: PEAK, LOW-PASS FILTER, and CONVERSION.

PEAK DETECTION

Peak detection uses the amplitude of a pulse-amplitude modulated (pam) signal or the duration of a pulse-duration modulated (pdm) signal to charge a holding capacitor and restore the original waveform. This demodulated waveform will contain some distortion because the output wave is not a pure sine wave. However, this distortion is not serious enough to prevent the use of peak detection.

Pulse-Amplitude Demodulation

Peak detection is used to detect pam. Figure 3-21 includes a simplified circuit [view (A)] for this demodulator and its waveforms [views (B) and (C)]. CR1 is the input diode which allows capacitor C1 to charge to the peak value of the pam input pulse. Pam input pulses are shown in view (B). CR1 is reverse biased between input pulses to isolate the detector circuit from the input. CR2 and CR3 are biased so that they are normally nonconducting. The discharge path for the capacitor is through the resistor (R1). These components are chosen so that their time constant is at least 10 times the interpulse period (time between pulses). This maintains the charge on C1 between pulses by allowing only a small discharge before the next pulse is applied. The capacitor is discharged just prior to each input pulse to allow the output voltage to follow the peak value of the input pulses. This discharge is through CR2 and CR3. These diodes are turned on by a negative pulse from a source that is time-synchronous with the timing-pulse train at the transmitter. Diode CR3 ensures that the output voltage is near 0 during this discharge period. View (C) shows the output wave shape from this circuit. The peaks of the output signal follow very closely the original modulating wave, as shown by the dotted line. With additional filtering this stepped waveform closely approximates its original shape.

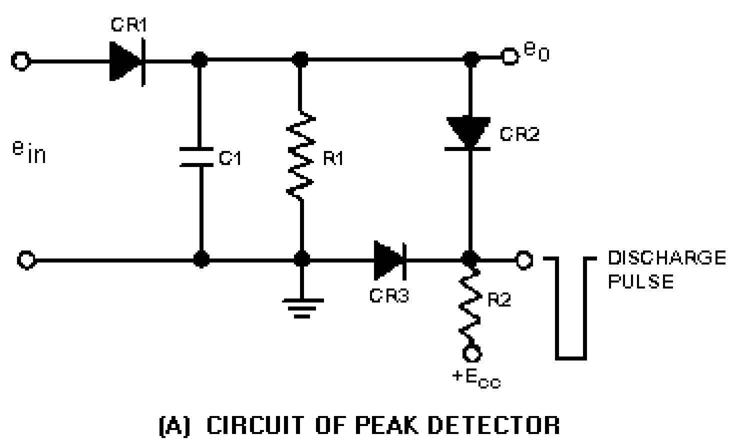


Figure 3-21A.—Peak detector. CIRCUIT OF PEAK DETECTOR.

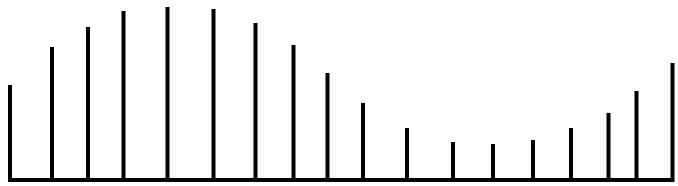


Figure 3-21B.—Peak detector. AMPLITUDE MODULATED PULSES.

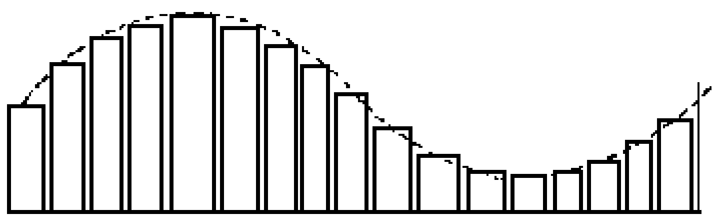


Figure 3-21C.—Peak detector. PEAK DETECTION.

Pulse-Duration Modulation

The peak detector circuit may also be used for pdm. To detect pdm, you must modify view (A) of figure 3-21 so that the time constant for charging C1 through CR1 is at least 10 times the maximum received pulse width. This may be done by adding a resistor in series with the cathode or anode circuit of CR1. The amplitude of the voltage to which C1 charges, before being discharged by the negative pulse, will be directly proportional to the input pulse width. A longer pulse width allows C1 to charge to a higher potential than a short pulse. This charge is held, because of the long time constant of R1 and C1, until the discharge pulse is applied to diodes CR2 and CR3 just prior to the next incoming pulse. These charges across C1 result in a wave shape similar to the output shown for pam detection in view (C) of figure 3-21.

Q-34. In its simplest form, what functions must a radar detector be capable of performing?

Q-35. What characteristic of a pulse does a peak detector sample?

Q-36. What is the time constant of the resistor and capacitor in a peak detector for pam?

Q-37. How can a peak detector for pam be modified to detect pdm?

LOW-PASS FILTER

Another method of demodulating pdm is by the use of a low-pass filter. If the voltage of a pulse waveform is averaged over both the pulse and no-pulse time, average voltage is the result. Since the amplitude of pdm pulses is constant, average voltage is directly proportional to pulse width. The pulse width varies with the modulation (intelligence) in pdm. Because the average value of the pulse train varies in accordance with the modulation, the intelligence may be extracted by passing the width-modulated pulses through a low-pass filter. The components of such a filter must be selected so that the filter passes only the desired modulation frequencies. As the varying-width pulses are applied to the low-pass filter, the average voltage across the filter will vary in the same way as the original modulating voltage. This varying voltage will closely approximate the original modulating voltage.

CONVERSION

Pulse-position modulation (ppm), pulse-frequency modulation (pfm), and pulse-code modulation (pcm) are most easily demodulated by first converting them to either pdm or pam. After conversion these pulses are demodulated using either peak detection or a low-pass filter. This conversion may be done in many ways, but your study will be limited to the simpler methods.

Pulse-Position Modulation

Ppm can be converted to pdm by using a flip-flop circuit. (Flip flops were discussed in *NEETS*, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*.) Figure 3-22 shows the waveforms for conversion of ppm to pdm. View (A) is the pulse-modulated pulse train and view (B) is a series of reset trigger pulses. The trigger pulses must be synchronized with the unmodulated position of the ppm pulses, but with a fixed time delay from these pulses. As the position-modulated pulse is applied to the flip-flop, the output is driven positive, as shown in view (C). After a period of time, the trigger pulse is again generated and drives the flip-flop output negative and the pulse ends. Because the ppm pulses are constantly varying in position with reference to the unmodulated pulses, the output of the flip-flop also varies in duration or width. This pdm signal can now be applied to one of the circuits that has already been discussed for demodulation.

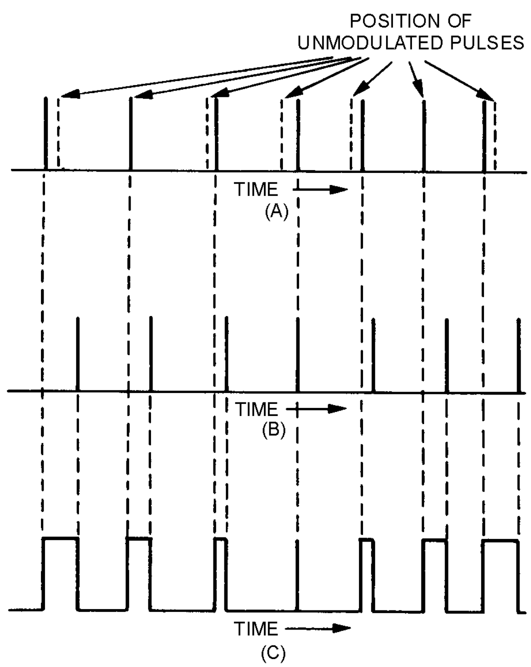


Figure 3-22.—Conversion of ppm to pdm.

Pulse-frequency modulation is a variation of ppm and may be converted by the same method.

Pulse-Code Modulation

Pulse-code modulation can easily be decoded, provided the pulse-code groups have been transmitted in reverse order; that is, if the pulse with the lowest value is transmitted first, the pulse with the highest value is transmitted last. A circuit that will provide a constant value of current without regard to its load is known as a current source. A current source is used to apply the pcm pulses to an RC circuit, such as that shown in figure 3-23, view (A). The current source must be capable of supplying a linear charge to C1 that will increase each time a pulse is applied if C1 is not allowed to discharge between pulses. In other words, if C1 charges to 16 volts during the period of one pulse, then each additional pulse increases the charge by 16 volts. Thus, the cumulative value increases by 16 volts for each received pulse. This does not provide a usable output unless a resistor is chosen that allows C1 to discharge to one-half its value between pulses. If only one pulse is received at T1, C1 charges to 16 volts and then begins to discharge. At T2 the charge has decayed to 8 volts and continues to decay unless another pulse is received. At T3 it has a 4-volt charge and at T4 it only has a 2-volt charge. At the sampling time, a 1-volt charge remains; this charge corresponds to the binary-weighted pulse train of 0001. Now we will apply a pcm signal which corresponds to the binary-coded equivalent of 7 volts (0111) in figure 3-23, view (A). View (B) is the pulse code that is received. Remember that the pulses are transmitted in reverse order. View (C) is the response curve of the circuit. At T1 the pulse corresponding to the least significant digit is applied and C1 charges to 16 volts. C1 discharges between pulses until it reaches 8 volts at T2. At T2 another pulse charges it to 24 volts. At T3, C1 has discharged through R1 to a value of 12 volts. The pulse at T3 increases the charge on C1 by 16 volts to a total charge of 28 volts. At T4, C1 has discharged to one-half its value and is at 14 volts. No pulse is present at T4 so C1 will not receive an additional charge. C1 continues to discharge until T5 when it has reached 7 volts and is sampled to provide a pam pulse which can be peak detected. This sampled output corresponds to the original sampling of the analog voltage in the modulation.

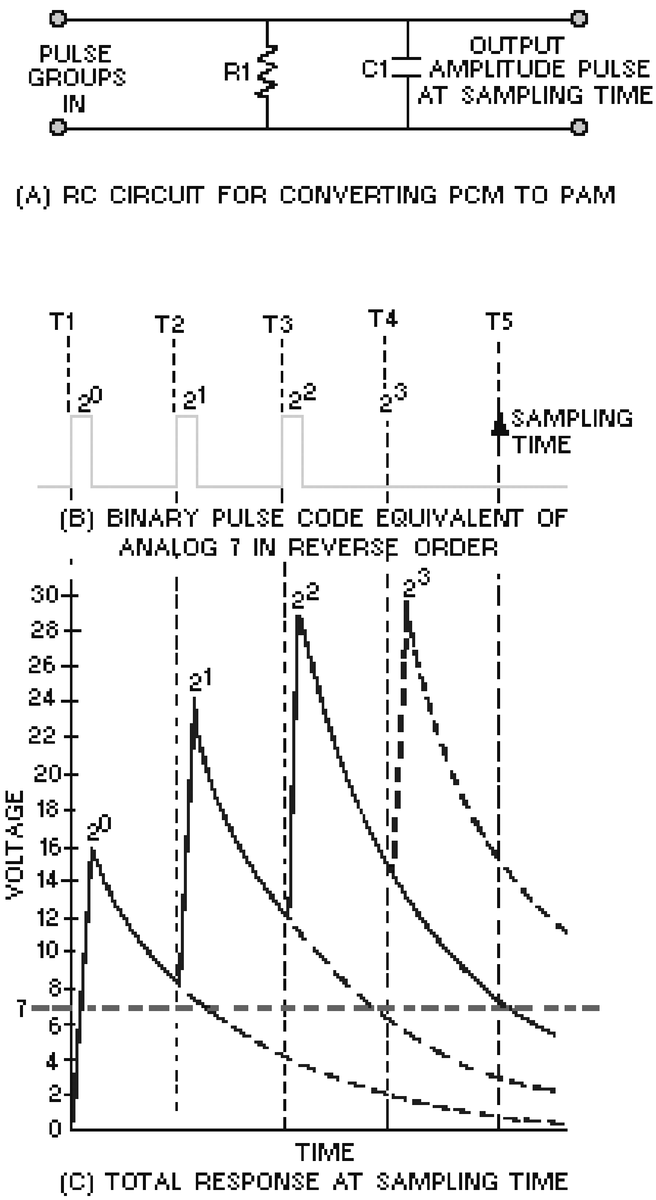


Figure 3-23.—Pcm conversion.

When the pcm demodulator recognizes the presence or absence of pulses in each position, it reproduces the correct standard amplitude represented by the pulse code group. For this reason, noise introduces no error if the largest peaks of noise are not mistaken for pulses. The pcm signal can be retransmitted as many times as desired without the introduction of additional noise effects so long as the signal-to-noise ratio is maintained at a level where noise pulses are not mistaken for a signal pulse. This is not the only method for demodulating pcm, but it is one of the simplest.

This completes your study of demodulation. You should remember that this module has been a basic introduction to the principles of modulation and demodulation. With the advent of solid-state electronics, integrated circuits have replaced discrete components. Although you cannot trace the signal flow through these circuits, the end result of the electronic action within the integrated circuit is the same as it would be with discrete components.

Q-38. How does a low-pass filter detect pdm?

Q-39. How is conversion used in pulse demodulation?

Q-40. What is the discharge rate for the capacitor in a pcm converter?

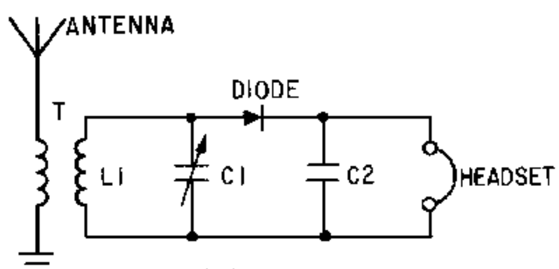
SUMMARY

Now that you have completed this chapter, a short review of what you have learned is in order. The following summary will refresh your memory of demodulation, its basic principles, and typical circuitry required to accomplish this task.

DEMODULATION, also called DETECTION, is the process of re-creating original modulating frequencies (intelligence) from radio frequencies.

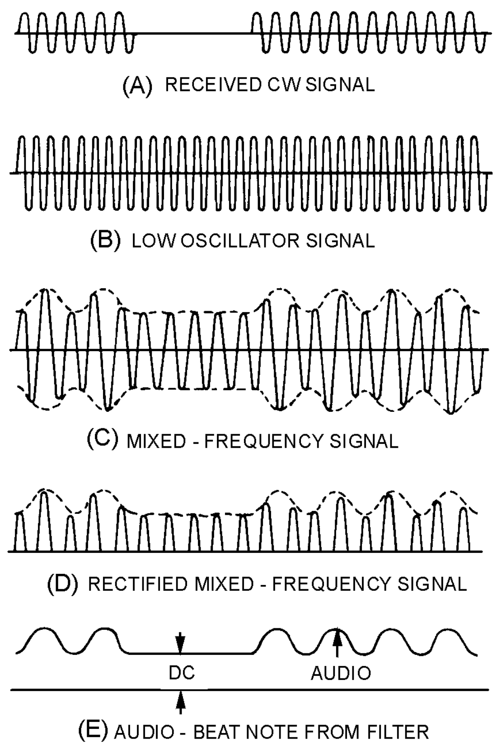
The **DEMODULATOR**, or DETECTOR, is the circuit in which the original modulating frequencies are restored.

A **CW DEMODULATOR** is a circuit that is capable of detecting the presence of rf energy.

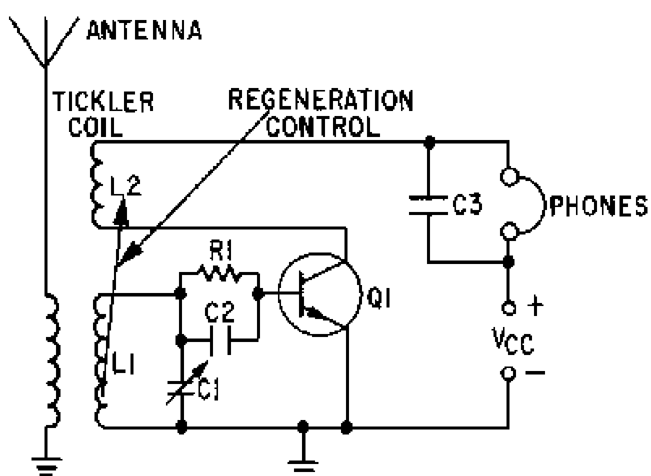


(A) CIRCUIT

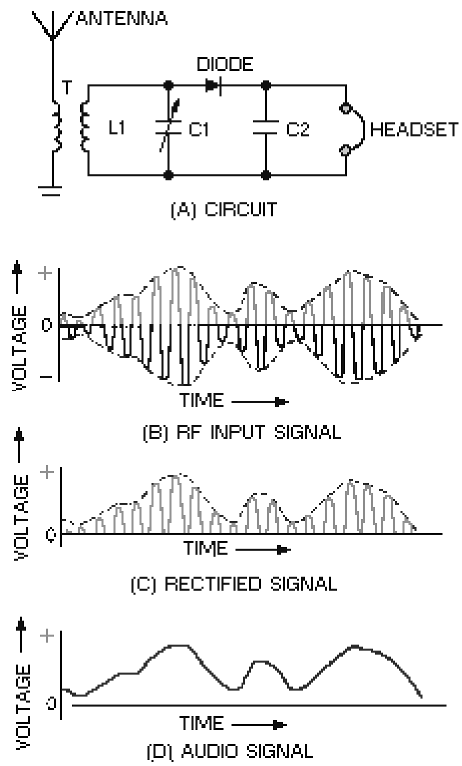
HETERODYNE DETECTION uses a locally generated frequency to beat with the cw carrier frequency to provide an audio output.



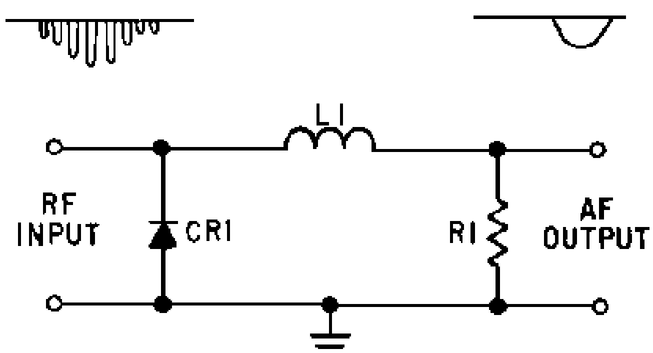
The **REGENERATIVE DETECTOR** produces its own oscillations, heterodynes them with an incoming signal, and detects them.



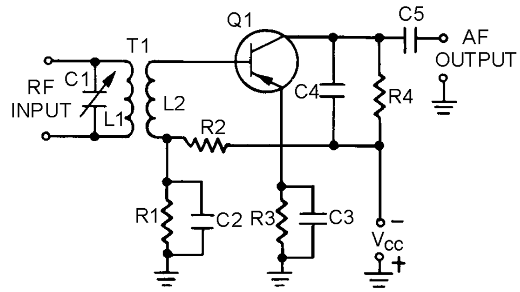
The **SERIES- (VOLTAGE-) DIODE DETECTOR** has a rectifier diode that is in series with the input voltage and the load impedance.



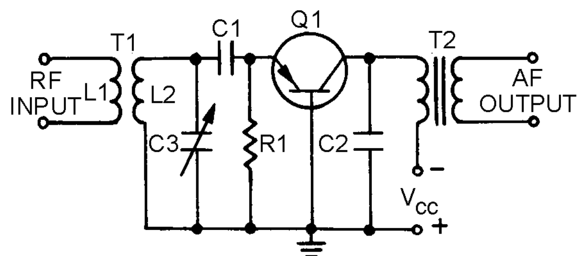
SHUNT- (CURRENT-) DIODE DETECTOR is characterized by a rectifier diode in parallel with the input and load impedance.



The **COMMON-EMITTER DETECTOR** is usually used in receivers to supply a detected and amplified output.

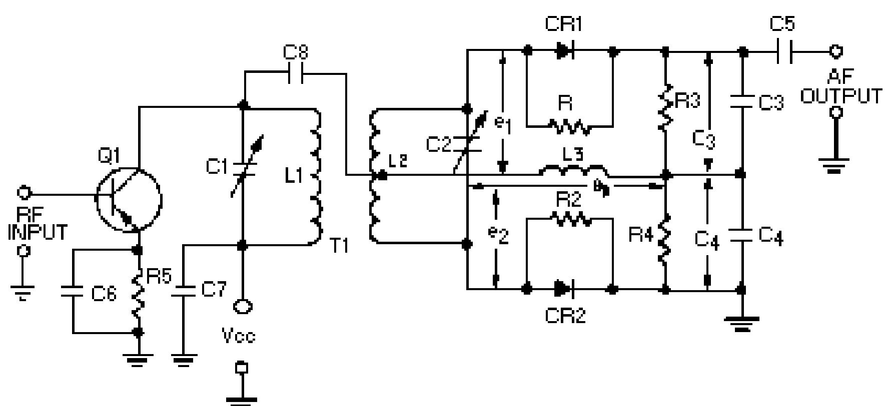


The **COMMON-BASE DETECTOR** is an amplifying detector that is used in portable receivers.

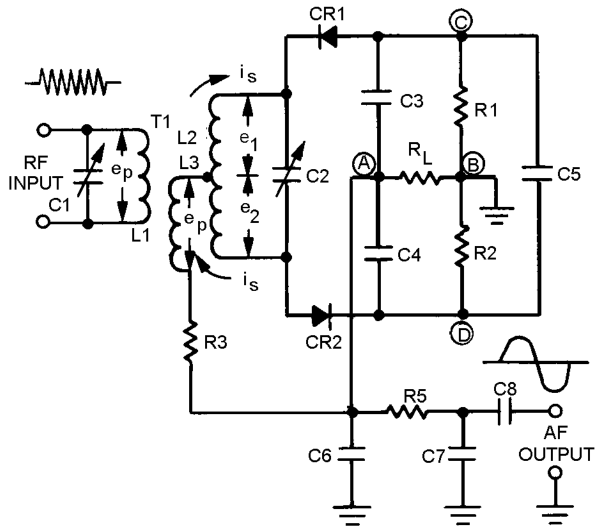


The **SLOPE DETECTOR** is the simplest form of frequency detector. It is essentially a tank circuit tuned slightly away from the desired fm carrier.

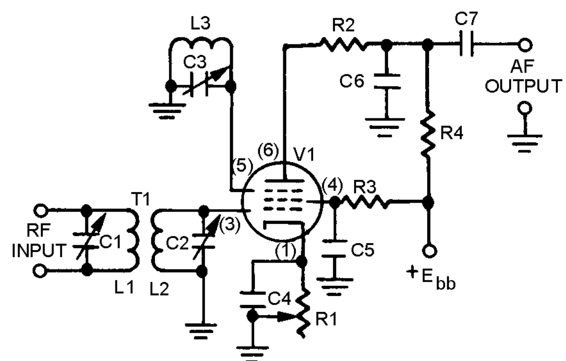
The **FOSTER-SEELEY DISCRIMINATOR** uses a double tuned rf transformer to convert frequency changes of the received fm signal into amplitude variations of the rf wave.



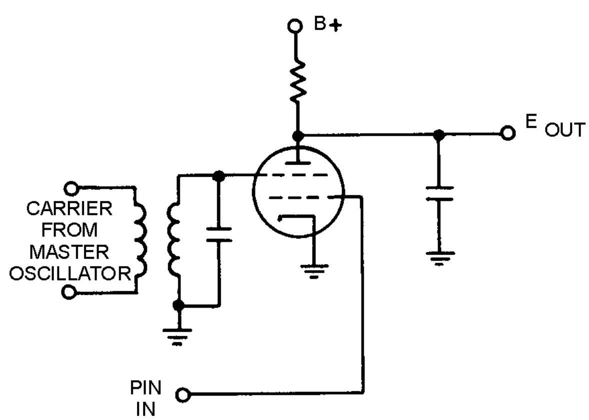
The **RATIO DETECTOR** uses a double-tuned transformer connected so that the instantaneous frequency variations of the fm input signal are converted into instantaneous amplitude variations.



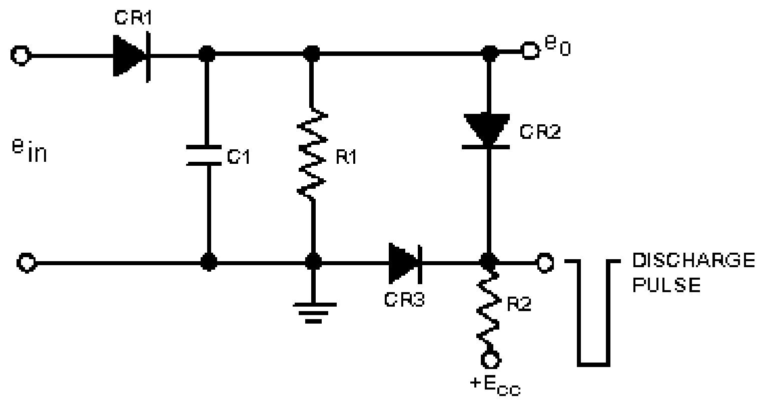
The **GATED-BEAM DETECTOR** uses a specially-designed tube to limit, detect, and amplify the received fm signal.



PHASE DEMODULATION may be accomplished using a frequency discriminator or a quadrature detector.



PEAK DETECTION uses the amplitude, or duration, of a pulse to charge a holding capacitor and restore the modulating waveform.



(A) CIRCUIT OF PEAK DETECTOR

A **LOW-PASS FILTER** is used to demodulate pdm by averaging the pulse amplitude over the entire period between pulses.

PULSE CONVERSION is used to convert ppm, pdm, or pcm to pdm or pam for demodulation.

ANSWERS TO QUESTIONS Q1. THROUGH Q40.

A-1. Re-creating original modulating frequencies (intelligence) from radio frequencies.

A-2. Circuit in which intelligence restoration is achieved.

A-3. A circuit that can detect the presence or absence of rf energy.

A-4. An antenna, tank circuit for tuning, rectifier for detection, filter to give constant output, and an indicator device.

A-5. Heterodyning.

A-6. By giving a different beat frequency for each signal.

A-7. Regenerative detector.

A-8. Oscillator, mixer, and detector.

A-9. (1) Sensitive to the type of modulation applied, (2) nonlinear, and (3) provide filtering.

A-10. The modulation envelope.

A-11. Rectifies the rf pulses in the received signal.

A-12. To filter the rf pulses and develop the modulating wave (intelligence) from the modulation envelope.

A-13. The current-diode detector is in parallel with the input and load.

A-14. When the input voltage variations are too small to give a usable output from a series detector.

A-15. Emitter-base junction.

A-16. R1.

A-17. By the collector current flow through R4.

A-18. Emitter-base junction.

A-19. A diode detector followed by a stage of audio amplification.

A-20. C1 and R1.

A-21. Slope detector.

A-22. Converting frequency variations of received fm signals to amplitude variations.

A-23. A double-tuned tank circuit.

A-24. Rectify the rf voltage from the discriminator.

A-25. Inductive.

A-26. Suppresses amplitude noise without limiter stages.

A-27. It helps to maintain a constant circuit voltage to prevent noise fluctuations from interfering with the output.

A-28. Limits, detects, and amplifies.

A-29. Both grids must be positively biased.

A-30. Extreme simplicity, few components, and ease of adjustment.

A-31. In the amount and rate of phase shift of the carrier wave.

A-32. Because of the incidental frequency shift that is caused while phase-shifting a carrier wave that is similar to fm modulation.

A-33. The quadrature grid signal is excited by a reference from the transmitter.

A-34. Detecting the presence of rf energy.

A-35. Pulse amplitude or pulse duration.

A-36. At least 10 times the interpulse period.

A-37. By making the time constant for charging the capacitor at least 10 times the maximum received pulse width.

A-38. By averaging the value of the pulses over the period of the pulse-repetition rate.

A-39. Ppm, pfm, and pcm are converted to either pdm or pam for demodulation.

A-40. It will discharge to one-half its value between pulses.

APPENDIX I

GLOSSARY

AMPLITUDE—Used to represent values of electrical current or voltage. The greater its height, the greater the value it represents.

AMPLITUDE MODULATION—Any method of varying the amplitude of an electromagnetic carrier frequency in accordance with the intelligence to be transmitted

ANGLE MODULATION—Modulation in which the angle of a sine-wave carrier is varied by a modulating wave.

AVERAGE POWER—The peak power value averaged over the pulse-repetition time.

BANDWIDTH—The section of the frequency spectrum that specific signals occupy.

BASE-INJECTION MODULATOR—Similar to control-grid modulator. Gain of a transistor is varied by changing the bias on its base.

BLOCKED-GRID KEYING—A method of keying in which the bias is varied to turn plate current on and off.

BUFFER—A voltage amplifier used between the oscillator and power amplifier.

CARBON MICROPHONE—Microphone in which sound waves vary the resistance of a pile of carbon granules. May be single-button or double-button.

CARRIER FREQUENCY—The assigned transmitter frequency.

CARRIER—Radio-frequency sine wave.

CATHODE KEYING—A system in which the cathode circuit is interrupted so that neither grid current nor plate current can flow.

CATHODE MODULATOR—Voltage on the cathode is varied to produce the modulation envelope.

CHANNEL—Carrier frequency assignment usually with a fixed bandwidth.

COLLECTOR-INJECTION MODULATOR—Transistor equivalent of plate modulator. Modulating voltage is applied to collector circuit.

COMMON-BASE DETECTOR—An amplifying detector where detection occurs in emitter-base junction and amplification occurs at the output of the collector junction.

COMMON-EMITTER DETECTOR—Often used in receivers to supply detected and amplified output. The emitter-base junction acts as the detector.

COMPLEX WAVE—A wave composed of two or more parts.

CONTINUOUS-WAVE KEYING—The "on-off" keying of a carrier.

CONTROL-GRID MODULATOR—Uses a variation of grid bias to vary the instantaneous plate voltage and current. The modulating signal is applied to the control grid.

CONVERSION—The process of changing ppm or pcm to pdm or pam to make them easier to demodulate.

CRYSTAL MICROPHONE—Uses the piezo-electric effect of crystalline materials to generate a voltage from sound waves.

CUSPS—Sharp phase reversals.

CW DEMODULATOR—A circuit that detects the presence of rf oscillations and converts them into a useful form.

CYCLE—360 degree rotation of a vector generating a sine wave.

DEMODULATION—The removal of intelligence from a transmission medium.

DEMODULATION or DETECTION—The process of re-creating original modulating-frequency intelligence from the rf carrier.

DEMODULATOR or DETECTOR—A circuit in which demodulation or restoration of the original intelligence is achieved.

DIODE DETECTOR—A simple type of crystal receiver.

DUTY CYCLE—The ratio of working time to total time for intermittently operated devices.

DYNAMIC MICROPHONE—A device in which sound waves move a coil of fine wire that is mounted on the back of a diaphragm and located in the magnetic field of a permanent magnet.

EMITTER-INJECTION MODULATOR—The transistor equivalent of the cathode modulator. The gain is varied by changing the voltage on the emitter.

FIDELITY—The ability to faithfully re-produce the input in the output.

FINAL POWER AMPLIFIER (fpa)—The final stage of amplification in a transmitter.

FIXED SPARK GAP—A device used to discharge the pulse-forming network. A trigger pulse ionizes the air between two contacts to initiate the discharge.

FOSTER-SEELEY DISCRIMINATOR—A circuit that uses a double-tuned rf transformer to convert frequency variations in the received fm signal to amplitude variations. Also known as a phase-shift discriminator.

FREQUENCY DEVIATION—The amount the frequency departs from the carrier frequency.

FREQUENCY MODULATION (fm)—Angle modulation in which the modulating signal causes the carrier frequency to vary. The *amplitude* of the modulating signal determines how far the frequency changes and the *frequency* of the modulating signal determines how fast the frequency changes.

FREQUENCY MULTIPLIERS—Special rf power amplifiers that multiply the input frequency.

FREQUENCY—The rate at which the vector that generates a sine wave rotates.

FREQUENCY-SHIFT KEYING (fsk)—Frequency modulation somewhat similar to continuous-wave (cw) keying in AM transmitters. The carrier is shifted between two differing frequencies by opening and closing a key.

GATED-BEAM DETECTOR—An fm demodulator that uses a special gated-beam tube to limit, detect, and amplify the received fm signal. Also known as a quadrature detector.

HARMONIC FREQUENCIES—Integral multiples of a fundamental frequency

HETERODYNE DETECTION—The use of an af voltage to distinguish between available signals. The incoming cw signal is mixed with locally generated oscillations to give an af output.

HETERODYNING—Mixing two frequencies across a nonlinear impedance.

HIGH-LEVEL MODULATION—Modulation produced in the plate circuit of the last radio stage of the system.

INSTANTANEOUS AMPLITUDE—The amplitude at any given point along a sine wave at a specific instant in time.

INTERMEDIATE POWER AMPLIFIER (ipa)—The amplifier between the oscillator and final power amplifier.

KEY CLICKS—Interference in the form of "clicks" or "thumps" caused by the sudden application or removal of power.

KEY-CLICK FILTERS—Filters used in keying systems to prevent key-click interference.

KEYING RELAYS—Relays used in high- power transmitters where the ordinary hand key cannot accommodate the plate current without excessive arcing.

LINEAR IMPEDANCE—An impedance in which a change in current through a device changes in direct proportion to the voltage applied to the device.

LOW-LEVEL MODULATION—Modulation produced in an earlier stage than the final.

LOW-PASS FILTER—A method of demodulating pdm by averaging the voltage over pulse and no-pulse time.

LOWER SIDEBAND—All difference frequencies below that of the carrier.

MACHINE KEYING—A method of cw keying using punched tape or other mechanical means to key a transmitter.

MAGNETIC MICROPHONE—A microphone in which the sound waves vibrate a moving armature. The armature consists of a coil wound on the armature and located between the pole pieces of a permanent magnet. The armature is mechanically linked to the diaphragm.

MARK—An interval during which a signal is present. Also the presence of an rf signal in cw keying. The key-closed condition (presence of data) in communications systems.

MASTER OSCILLATOR POWER AMPLIFIER (MOPA)—A transmitter in which the oscillator is isolated from the antenna by a power amplifier.

MICROPHONE—An energy converter that changes sound energy into electrical energy.

MODULATED WAVE—A complex wave consisting of a carrier and a modulating wave that is transmitted through space.

MODULATING WAVE—An information wave representing intelligence.

MODULATION FACTOR (M)—An indication of relative magnitudes of the rf carrier and the audio-modulating signal.

MODULATION INDEX—The ratio of frequency deviation to the frequency of the modulating signal.

MODULATION—The ability to impress intelligence upon a transmission medium, such as radio waves.

MODULATOR—The last audio stage in which intelligence is applied to the rf stage to modulate the carrier.

MULTIPLICATION FACTOR—The number of times an input frequency is multiplied.

MULTIVIBRATOR MODULATOR—An astable multivibrator used to provide frequency modulation by inserting the modulating af voltage in series with the base-return of the multivibrator transistors.

NEGATIVE ALTERNATION—That part of a sine wave that is below the reference level.

NONLINEAR DEVICE—A device in which the output does not rise and fall directly with the input.

NONLINEAR IMPEDANCE—An impedance in which the resulting current through the device is not proportional to the applied voltage.

OVERMODULATION—A condition that exists when the peaks of the modulating signal are limited.

PEAK AMPLITUDE—The maximum value above or below the reference line.

PEAK DETECTION—Detection that uses the amplitude of pam or the duration of pdm to charge a holding capacitor and restore the original waveform.

PEAK POWER—The maximum value of the transmitted pulse.

PERCENT OF MODULATION—The degree of modulation defined in terms of the maximum permissible amount of modulation.

PERIOD—The duration of a waveform.

PHASE MODULATION (pm)—Angle modulation in which the phase of the carrier is controlled by the modulating waveform. The *amplitude* of the modulating wave determines the amount of phase shift and the *frequency* of the modulation determines how often the phase shifts.

PHASE or PHASE ANGLE—The angle that exists between the starting point of a vector and its position at that instant. An indication of how much of a cycle has been completed at any given instant in time.

PHASE-SHIFT DISCRIMINATOR—See Foster-Seeley discriminator.

PHASE-SHIFT KEYING—Similar to ON-OFF cw keying in AM systems and frequency-shift keying in fm systems. Each time a *mark* is received, the phase is reversed. No phase reversal takes place when a space is received.

PLATE KEYING—A keying system in which the plate supply is interrupted.

PLATE MODULATOR—An electron-tube modulator in which the modulating voltage is applied to the plate circuit of the tube.

POSITIVE ALTERNATION—That part of a sine wave that is above the reference line.

PULSE DURATION (pd)—The period of time during which a pulse is present.

PULSE MODULATION—A form of modulation in which one of the characteristics of a pulse train is varied.

PULSE WIDTH (pw)—The period of time during which a pulse occurs.

PULSE—A surge of plate current that occurs when a tube is momentarily saturated.

PULSE-AMPLITUDE MODULATION (pam)—Pulse modulation in which the *amplitude* of the pulses is varied by the modulating signal.

PULSE-CODE MODULATION (pcm)—A modulation system in which the standard values of a quantized wave are indicated by a series of coded pulses.

PULSE-DURATION MODULATION (pdm)—Pulse modulation in which the *time duration* of the pulses is changed by the modulating signal.

PULSE-FORMING NETWORK (pfn)—A circuit used for storing energy. Essentially a short section of artificial transmission line.

PULSE-FREQUENCY MODULATION (pfm)—Pulse modulation in which the modulating voltage varies the *repetition rate* of a pulse train.

PULSE-POSITION MODULATION (ppm)—Pulse modulation in which the *position* of the pulses is varied by the modulating voltage.

PULSE-REPETITION FREQUENCY—The rate, in pulses per second, at which the pulses occur.

PULSE-REPETITION TIME (prt)—The total time for one complete pulse cycle of operation (*rest time* plus *pulse width*).

PULSE-TIME MODULATION (ptm)—Pulse modulation that varies one of the *time characteristics* of a pulse train (*pwm, pdm, ppm,* and *pfm*).

PULSE-WIDTH MODULATION (pwm)—Pulse modulation in which the *duration* of the pulses is varied by the modulating voltage.

PULSING—Allowing oscillations to occur for a specific period of time only during selected intervals.

QUANTIZED WAVE—A wave created by arbitrarily dividing the entire range of amplitude (or frequency, or phase) values of an analog wave into a series of standard values. Each sample takes the standard value nearest its actual value when modulated.

QUANTIZING NOISE—A distortion introduced by quantizing the signal.

RADAR DETECTOR—A detector which, in its simplest form, only needs to be capable of producing an output when rf energy (reflected from a target) is present at its input.

RATIO DETECTOR—A detector that uses a double-tuned transformer to convert the instantaneous frequency variations of the fm input signal to instantaneous amplitude variations.

RATIO OF TRANSMITTED POWERS—The power ratio (*fsk* verses *AM*) that expresses the overall improvement of fsk transmission when compared to AM under rapid-fading and high-noise conditions.

REACTANCE TUBE—A tube connected in parallel with the tank circuit of an oscillator. Provides a signal that will either lag or lead the signal produced by the tank.

REACTANCE-TUBE MODULATOR—An fm modulator that uses a reactance tube in parallel with the oscillator tank circuit.

REGENERATIVE DETECTOR—A detector circuit that produces its own oscillations, heterodynes them with an incoming signal, and detects them.

REST FREQUENCY—The carrier frequency during the constant-amplitude portions of a phase modulation signal.

REST TIME (rt)—The time when there is no pulse or nonpulse.

ROTARY GAP—Similar to a mechanically driven switch. Used to discharge a pulse-forming network.

SENSITIVITY OF A MICROPHONE—Efficiency of a microphone. Describes micro- phone power delivered to a matched-impedance load as compared to the sound level being converted. Usually expressed in terms of the electrical power level.

SERIES-DIODE DETECTOR—The semiconductor diode in series with the input voltage and the load impedance. Sometimes called a voltage-diode detector.

SHUNT—Means the same as parallel or to place in parallel with other components.

SHUNT-DIODE DETECTOR—A diode detector in which the diode is in parallel with the input voltage and the load impedance. Also known as a current detector because it operates with smaller input levels.

SIGNAL DISTORTION—Any unwanted change to the signal.

SIGNIFICANT SIDEBANDS—Those sidebands with significantly large amplitude.

SINE WAVE—The basic synchronous alternating waveform for all complex waveforms.

SLOPE DETECTOR—A tank circuit tuned to a frequency, either slightly above or below an fm carrier frequency, that is used to detect intelligence.

SPACE—Absence of an rf signal in cw keying. Key-open condition or lack of data in communications systems. Also a period of no signal.

SPARK-GAP MODULATOR—Modulator consists of a circuit for storing energy, a circuit for rapidly discharging the storage circuit (spark gap), a pulse transformer, and a power source.

SPECTRUM ANALYSIS—The display of electromagnetic energy arranged according to wavelength or frequency.

SPLATTER—Unwanted sideband frequencies that are generated from overmodulation.

THYRATRON MODULATOR—An electronic switch that requires a low potential to turn it on.

TRANSMISSION MEDIUM—A means of transferring intelligence from point to point. Can be described as light, smoke, sound, wirelines, or radio-frequency waves.

UPPER SIDEBAND—All of the sum frequencies above the carrier.

VARACTOR—A diode, or pn junction, that is designed to have a certain amount of capacitance between junctions.

VARACTOR FM MODULATOR—An fm modulator which uses avoltage-variable capacitordiode (varactor).

VOLTAGE-DIODE DETECTOR—Series detector in which the crystal is in series with the input voltage and the load impedance.

VECTOR—Mathematical method of showing both magnitude and direction.

WAVELENGTH—The physical dimension of a sine wave.

MODULE 12 INDEX

ASSIGNMENT 1

Textbook assignment: Chapter 1, “Amplitude Modulation,” pages 1-1 through 1-75.

1-1. The action of impressing intelligence upon a transmission medium is referred to as

1. modulating
2. demodulating
3. heterodyning
4. wave generating

1-2. You can communicate with others using which of the following transmissions mediums?

1. Light
2. Wire lines
3. Radio waves
4. Each of the above

1-3. When you use a vector to indicate force in a diagram, what do (a) length and (b) arrowhead position indicate?

1. (a) Magnitude (b) direction
2. (a) Magnitude (b) frequency
3. (a) Phase (b) frequency
4. (a) Phase (b) direction

1-4. Vectors are used to show which of the following characteristics of a sine wave?

1. Fidelity
2. Amplitude
3. Resonance
4. Distortion

1-5. A rotating coil in the uniform magnetic field between two magnets produces a sine wave. It is called a sine wave because the voltage depends on which of the following factors?

1. The number of turns in the coil
2. The speed at which the coil is rotating
3. The angular position of the coil in the magnetic field
4. Each of the above

1-6. The trigonometric relationship for the sine of an angle in a right triangle is figured using which of the following ratios?

1. Opposite side ÷ hypotenuse
2. Adjacent side ÷ hypotenuse
3. Hypotenuse ÷ opposite side
4. Hypotenuse ÷ adjacent side

1-7. The part of a sine wave that is above the voltage reference line is referred to as the

1. peak amplitude
2. positive alternation
3. negative alternation
4. instantaneous amplitude

1-8. The degree to which a cycle has been completed at any given instant is referred to as the

1. phase
2. period
3. frequency
4. amplitude

1-9. The frequency of the sine wave is determined by which of the following sine-wave factors?

1. The maximum voltage
2. The rate at which the vector rotates
3. The number of degrees of vector rotation
4. Each of the above

1-10. Which of the following mathematical relationships do you use to figure the period of a sine wave?

1. $\frac{1}{\text{phase}}$
2. $\frac{1}{\text{duration}}$
3. $\frac{1}{\text{frequency}}$
4. $\frac{1}{\text{amplitude}}$

1-11. Which of the following Greek letters is the symbol for wavelength?

1. θ
2. ϑ
3. λ
4. ω

1-12. Which of the following waveform characteristics determines the wavelength of a sine wave?

1. Phase
2. Period
3. Amplitude
4. Phase Angle

1-13. An electromagnetic wavefront moves through free space at approximately what speed in meters per second?

1. 3,000,000
2. 30,000,000
3. 300,000,000
4. 3,000,000,000

1-14. What is the wavelength of a 1.5 MHz frequency?

1. 100 meters
2. 200 meters
3. 300 meters
4. 400 meters

1-15. As the frequency of a signal is increased, what change can be noted about its wavelength?

1. It decreases
2. It increases
3. It remains the same

1-16. The ability of a circuit to faithfully reproduce the input signal in the output is known by what term?

1. Fidelity
2. Fluctuation
3. Directivity
4. Discrimination

1-17. In rf communications, modulation impresses information on which of the following types of waves?

1. Carrier wave
2. Complex wave
3. Modulated wave
4. Modulating wave

1-18. Which of the following types of modulation is a form of amplitude modulation?

1. Angle
2. Phase
3. Frequency
4. Continuous-wave

1-19. With a sine-wave input, how will the output compare to the input in (a) a linear circuit and (b) a nonlinear circuit?

1. (a) Proportional
   (b) proportional
2. (a) Proportional
   (b) not proportional
3. (a) Not proportional
   (b) not proportional
4. (a) Not proportional
   (b) proportional

1-20. What effect, if any, does a nonlinear device have on a sine wave?

1. It amplifies without distortion
2. It attenuates without distortion
3. It generates harmonic frequencies
4. None

1-21. For the heterodyning action to occur in a circuit, (a) what number of frequencies must be present and (b) to what type of circuit must they be applied?

1. (a) Two (b) linear
2. (a) Two (b) nonlinear
3. (a) Three (b) nonlinear
4. (a) Three (b) linear

1-22. Spectrum analysis is used to view which of the following characteristics of an rf signal?

1. Phase
2. Bandwidth
3. Modulating wave
4. Modulation envelope

1-23. The method of rf communication that uses either the presence or absence of a carrier in a prearranged code is what type of modulation?

1. Pulse modulation
2. Amplitude modulation
3. Continuous-wave modulation
4. Pulse-time modulation

1-24. What is the purpose of the key in a cw transmitter?

1. It generates the rf oscillations
2. It heterodynes the rf oscillations
3. It controls the rf output
4. It amplifies the rf signal

1-25. To ensure frequency stability in a cw transmitter, you should NOT key what circuit?

1. The mixer
2. The detector
3. The oscillator
4. The rf amplifier

1-26. When keying a high-power transmitter, what component should you use to reduce the shock hazard?

1. A coil
2. A relay
3. A resistor
4. A capacitor

1-27. Interference detected by a receiver is often caused by the application and removal of power in nearby transmitters. This interference can be prevented by using what type of circuit in such transmitters?

1. Power filter
2. On-off filter
3. Key-click filter
4. Rf detector filter

1-28. Transmitter machine keying was developed for which of the following purposes?

1. To increase the speed of communications
2. To make communications more intelligible
3. To reduce interference
4. Each of the above

1-29. Which of the following advantages is a benefit of cw communications?

1. Wide bandwidth
2. Fast transmission
3. Long-range operation
4. Each of the above

1-30. To prevent a transmitter from being loaded unnecessarily, where should you connect the antenna?

1. At the oscillator input
2. At the oscillator output
3. At the power-amplifier input
4. At the power-amplifier output

1-31. Amplifier tubes are added to the output of a transmitter for which of the following reasons?

1. To increase power
2. To increase frequency
3. To increase stability
4. To increase selectivity

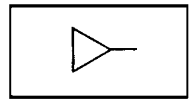
1-32. Which of the following combinations of frequency multiplier stages will produce a total multiplication factor of 72?

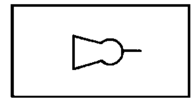
1. 36, 36
2. 4, 3, 3, 2
3. 4, 4, 3, 2
4. 18, 18, 18, 18

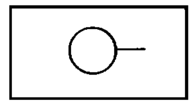
1-33. To change sound energy into electrical energy, which of the following devices should you use?

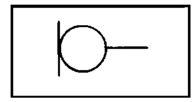
1. A speaker
2. A microphone
3. An amplifier
4. An oscillator

1-34. Which of the following is the schematic symbol for a microphone?

1. 

3. 

2. 

4. 

1-35. What component in a carbon microphone converts a dc voltage into a varying current?

1. Button
2. Diaphragm
3. Transformer
4. Carbon granules

1-36. The action of the double-button carbon microphone is similar to which of the following electronic circuits?

1. A limiter
2. An oscillator
3. A voltage doulber
4. A push-pull amlifier

1-37. A carbon microphone has which of the following advantages over other types of microphones?

1. Ruggedness
2. Sensitivity
3. Low output voltage
4. Frequency response

1-38. The voltage produced by mechanical stress placed on certain crystals is a result of which of the following effects?

1. Hall
2. Acoustic
3. Electrostatic
4. Piezoelectric

1-39. If you require a microphone that is lightweight, has high sensitivity, is rugged, requires no external voltage, can withstand temperature, vibration, and moisture extremes, and has a uniform frequency response of 40 to 15,000 hertz, which of the following types of microphones should you select?

1. Carbon
2. Crystal
3. Dynamic
4. Electrostatic

1-40. What component in a magnetic microphone causes the lines of flux to alternate?

1. The coil
2. The magnet
3. The diaphragm
4. The armature

1-41. What are the two major sections of an AM transmitter?

1. Audio frequency unit and radio frequency unit
2. Audio frequency unit and master oscillator
3. Audio frequency unit and final power amplifier
4. Audio frequency unit and intermediate power amplifier

1-42. The intermediate power amplifier serves what function in a transmitter?

1. It generates the carrier
2. It modulates the carrier
3. It increases the frequency of the signal
4. It increases the power level of the signal

1-43. The final audio stage in an AM transmitter is the

1. mixer
2. modulator
3. multipler
4. multiplexer

1-44. The vertical axis on a frequency-spectrum graph represents which of the following waveform characteristics?

1. Phase
2. Duration
3. Frequency
4. Amplitude

1-45. When a 500-hertz signal modulates a 1-megahertz carrier, the 1-megahertz carrier and what two other frequencies are transmitted?

1. 500 and 999,500 hertz
2. 500 and 1,000,500 hertz
3. 999,500 and 1,500,000 hertz
4. 999,500 and 1,000,500 hertz

1-46. If 750 hertz modulates a 100-kilohertz carrier, what would the upper-sideband frequency be?

1. 99,250 hertz
2. 100,000 hertz
3. 100,500 hertz
4. 100,750 hertz

1-47. In an AM wave, where is the audio intelligence located?

1. In the carrier frequency
2. In the spacing between the sideband frequencies
3. In the spacing between the carrier and sideband frequencies
4. In the sideband frequencies

1-48. What determines the bandwidth of an AM wave?

1. The carrier frequency
2. The number of sideband frequencies
3. The lowest modulating frequency
4. The highest modulating frequency

1-49. If an 860-kilohertz AM signal is modulated by frequencies of 5 and 10 kilohertz, what is the bandwidth?

1. 5 kilohertz
2. 10 kilohertz
3. 15 kilohertz
4. 20 kilohertz

1-50. If a 1-megahertz signal is modulated by frequencies of 50 and 75 kilohertz, what is the resulting maximum frequency range?

1. 925,000 to 1,000,000 hertz
2. 925,000 to 1,075,000 hertz
3. 975,000 to 1,025,000 hertz
4. 1,000,000 to 1,075,000 hertz

1-51. If an rf carrier is 100 percent AM-modulated, what will be the rf output when the modulating signal is (a) at its negative peak and (b) at its positive peak?

1. (a) 0
   (b) 2 times the amplitude of the unmodulated carrier
2. (a) 0
   (b) 1/2 the amplitude of the unmodulated carrier
3. (a) 1/2 the amplitude of the unmodulated carrier
   (b) 1/2 the amplitude of the unmodulated carrier
4. (a) 1/2 the amplitude of the unmodulated carrier
   (b) 2 times the amplitude of the unmodulated carrier

1-52. In an AM signal that is 100 percent modulated, what maximum voltage value is present in each sideband?

1. 1/4 the carrier voltage
2. 1/2 the carrier voltage
3. 3/4 the carrier voltage
4. Same as the carrier voltage

1-53. Overmodulation of an AM signal will have which, if any, of the following effects on the bandwidth?

1. It will increase
2. It will decrease
3. It will remain the same

1-54. In a carrier wave with a peak amplitude of 400 volts and a peak modulating voltage of 100 volts, what is the modulation factor?

1. 0.15
2. 0.25
3. 0.45
4. 0.55

1-55. The percent of modulation for a modulated carrier wave is figured using which of the following formulas?

1. $\frac{E_m}{E_c}$
2. $\frac{E_c}{E_m}$
3. $\frac{E_m}{E_c} \times 100$
4. $\frac{E_c}{E_m} \times 100$

1-56. Modulation produced in the plate circuit of the last radio stage of a system is known by what term?

1. Low-level modulation
2. High-level modulation
3. Final-amplifier modulation
4. Radio frequency modulation

1-57. Which, if any, of the following advantages is a primary benefit of plate modulation?

1. It operates at low efficiency
2. It operates at low power levels
3. It operates with high efficiency
4. None of the above

1-58. A final rf power amplifier biased for plate modulation operates in what class of operation?

1. A
2. B
3. AB
4. C

1-59. Heterodyning action in a plate modulator takes place in what circuit?

1. Grid
2. Plate
3. Screen
4. Cathode

1-60. A plate modulator produces a modulated rf output by controlling which of the following voltages?

1. Plate voltage
2. Cathode voltage
3. Grid-bias voltage
4. Grid-input voltage

1-61. To achieve 100-percent modulation in a plate modulator, what maximum voltage must the modulator tube be capable of providing to the final power amplifier (fpa)?

1. Twice the fpa plate voltage
2. The same as the fpa plate voltage
3. Three times the fpa plate voltage
4. Half the fpa plate voltage

1-62. In a plate modulator, with no modulation, how will the plate current of the final rf amplifier appear on a scope?

1. A series of pulses at the carrier frequency
2. A series of pulses at twice the carrier frequency
3. A series of pulses at 1/4 the carrier frequency
4. A series of pulses at 1/2 the carrier frequency

1-63. In the collector-injection modulator, af and rf are heterodyned by injecting the rf into (a) what circuit and the af into (b) what circuit?

1. (a) Base (b) collector
2. (a) Base (b) emitter
3. (a) Emitter (b) collector
4. (a) Emitter (b) base

1-64. Plate- and collector-injection modulators are the most commonly used modulators for which of the following reasons?

1. The rf amplifier stages can be operated class C for linearity
2. The rf amplifier stages can be operated class C for maximum efficiency
3. They require small amounts of audio power
4. They require large amounts of audio power

1-65. A control-grid modulator would be used in which of the following situations?

1. In extremely high-power, wideband equipment where high-level modulation is difficult to achieve
2. In cases where the use of a minimum of audio power is desired
3. In portable and mobile equipment to reduce size and power requirements
4. Each of the above

1-66. Which of the following inputs is/are applied to the grid of a control-grid modulator?

1. Rf
2. Af
3. Dc bias
4. Each of the above

1-67. Excessive modulating signal levels have which, if any, of the following effects on a control-grid modulator?

1. They increase output. amplitude
2. They decrease output amplitude
3. They create distortion
4. None

1-68. Compared to a plate modulator, the control-grid modulator has which of the following advantages?

1. It is more efficient
2. It has less distortion
3. It requires less power from the modulator
4. It requires less power from the amplifier

1-69. The control-grid modulator is similar to which of the following modulator circuits?

1. Plate
2. Cathode
3. Base-injection
4. Emitter-injection

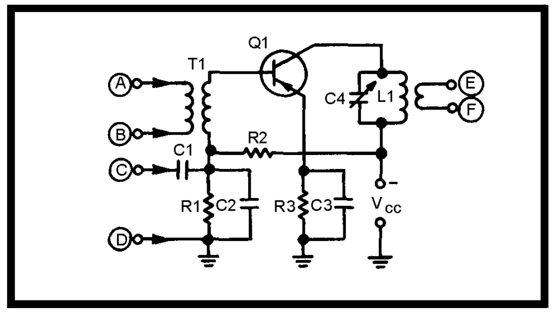


Figure 1A.—Modulator circuit.

IN ANSWERING QUESTIONS 1-70 THROUGH 1-72, REFER TO FIGURE 1A.

1-70. What components in the circuit establish the bias for Q1?

1. R1 and R2
2. R2 and R3
3. R1 and R3

1-71. The rf voltage in the circuit is applied at (a) what points and the af voltage is applied at (b) what points?

1. (a) A and B (b) C and D
2. (a) C and D (b) A and B
3. (a) C and D (b) E and F
4. (a) E and F (b) C and D

1-72. What components develop the rf modulation envelope?

1. C1 and R1
2. C2 and R1
3. C3 and R3
4. C4 and L1

1-73. A cathode modulator is used in which of the following situations?

1. When rf power is unlimited and distortion can be tolerated
2. When rf power is limited and distortion cannot be tolerated
3. When af power is unlimited and distortion can be tolerated
4. When af power is limited and distortion cannot be tolerated

1-74. In a cathode modulator, the modulating voltage is in series with which of the following voltages?

1. The grid voltage only
2. The plate voltage only
3. Both the grid and plate voltages
4. The cathode voltage only

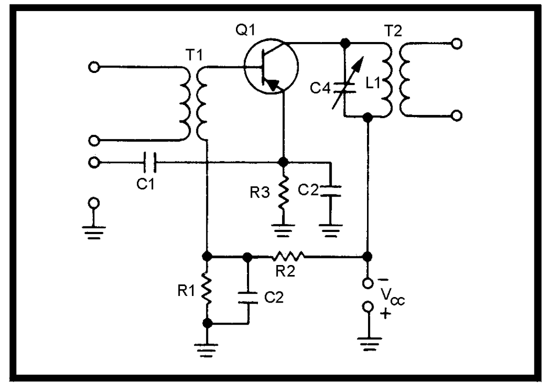


Figure 1B.—Emitter-injection modulator.

IN ANSWERING QUESTION 1-75, REFER TO FIGURE 1B.

1-75. In the circuit, what components develop the modulation envelope?

1. Q1
2. C2 and R1
3. C3 and R3
4. C4 and L1

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Angle and Pulse Modulation," pages 2-1 through 2-64.

2-1. Frequency-shift keying resembles what type of AM modulation?

1. CW modulation
2. Analog AM modulation
3. Plate modulation
4. Collector-injection modulation

2-2. Frequency-shift keying is generated using which of the following methods?

1. By shifting the frequency of an oscillator at an af rate
2. By shifting the frequency of an oscillator at an rf rate
3. By keying an af oscillator at an rf rate
4. By keying an af oscillator at an af rate

2-3. In a frequency-shift keyed signal, where is the intelligence contained?

1. In the duration of the rf energy
2. In the frequency of the rf energy
3. In the amplitude of the rf energy
4. In the spacing between bursts of rf energy

2-4. If an fsk transmitter has a MARK frequency of 49.575 kilohertz and a SPACE frequency of 50.425 kilohertz, what is the assigned channel frequency?

1. 49 kilohertz
2. 49.575 kilohertz
3. 50 kilohertz
4. 50.425 kilohertz

2-5. Fsk is NOT affected by noise interference for which of the following reasons?

1. Noise is outside the bandwidth of an fsk signal
2. Fsk does not rely on the amplitude of the transmitted signal to carry intelligence
3. The wide bandwidth of an fsk signal prevents noise interference
4. Each of the above

2-6. In an fsk transmitter, what stage is keyed?

1. The oscillator
2. The power supply
3. The power amplifier
4. The buffer amplifier

2-7. When the amount of oscillator frequency shift in an fsk transmitter is determined, which of the following factors must be considered?

1. The number of buffer amplifiers
2. The transmitter power output
3. The frequency multiplication factor for the transmitter amplifiers
4. The oscillator rest frequency

2-8. In an fsk transmitter with a doubler and a tripler stage, the desired frequency shift is 1,200 hertz. To what maximum amount is the oscillator frequency shift limited?

1. 60 hertz
2. 100 hertz
3. 120 hertz
4. 200 hertz

2-9. Fsk has which of the following advantages over cw?

1. Fsk has a more stable oscillator
2. Fsk is easier to generate
3. Fsk rejects unwanted weak signals
4. Fsk does not have noise in its output

2-10. The "ratio of transmitted powers" provides what information?

1. Transmitter power out in a cw system
2. Transmitter power out in an fsk system
3. Improvement shown using cw instead of fsk transmission
4. Improvement shown using fsk instead of cw transmission methods

2-11. In an fm signal, (a) the RATE of shift is proportional to what characteristic of the modulating signal, and (b) the AMOUNT of shift is proportional to what characteristic?

1. (a) Amplitude (b) amplitude
2. (a) Amplitude (b) frequency
3. (a) Frequency (b) frequency
4. (a) Frequency (b) amplitude

THIS SPACE LEFT BLANK INTENTIONALLY.

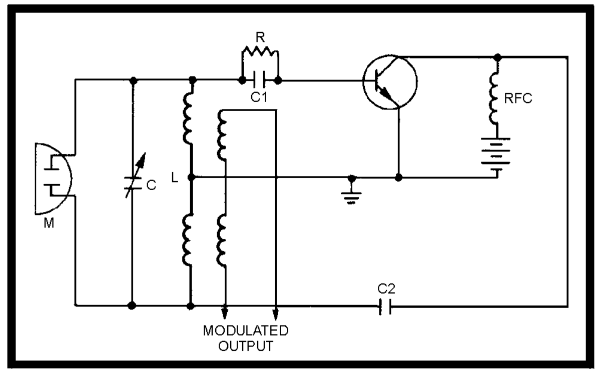


Figure 2A.—Oscillator circuit.

IN ANSWERING QUESTIONS 2-12 THROUGH 2-14, REFER TO FIGURE 2A.

2-12. When a sound wave strikes the condenser microphone (M), it has which, if any, of the following effects on the oscillator circuit?

1. It changes output phase
2. It changes output voltage
3. It changes output frequency
4. It has no effect

2-13. What is the purpose of capacitor C in the circuit?

1. It helps set the carrier frequency of the oscillator
2. It prevents amplitude variations in the oscillator output
3. It sets the maximum frequency deviation of the oscillator
4. It varies the output frequency in accordance with the modulating voltage

2-14. A 1,000-hertz tone of a certain loudness causes the frequency-modulated carrier for the circuit to vary ±1,000 hertz at a rate of 1,000 times per second. If the AMPLITUDE of the modulating tone is doubled, what will be the maximum carrier variation?

1. ±1,000 hertz at 1,000 times per second
2. ±1,000 hertz at 2,000 times per second
3. ±2,000 hertz at 1,000 times per second
4. ±2,000 hertz at 2,000 times per second

2-15. The maximum deviation for a 1.5 MHz carrier is set at ±50 kHz. If the carrier varies between 1.5125 MHz and 1.4875 MHz (±12.5 kHz), what is the percentage of modulation?

1. 25 %
2. 50 %
3. 75 %
4. 100 %

2-16. An fm transmitter has a 50-watt carrier with no modulation. What maximum amount of output power will it have when it is 50-percent modulated?

1. 25 watts
2. 50 watts
3. 75 watts
4. 100 watts

2-17. Frequencies that are located between adjacent channels to prevent interference are referred to as

1. sidebands
2. bandwidths
3. guard bands
4. blank channels

2-18. Modulation index may be figured by using which of the following formulas?

1. $2f/f_m$
2. $f_m/2f$
3. $f_m/\Delta f$
4. $\Delta f/f_m$

2-19. A 50-megahertz fm carrier varies between 49.925 megahertz and 50.075 megahertz 10,000 times per second. What is its modulation index?

1. 5
2. 10
3. 15
4. 20

MODULATION INDEX	SIGNIFICANT SIDEBANDS
.01	2
.4	2
.5	4
1.0	6
2.0	8
3.0	12
4.0	14
5.0	16
6.0	18
7.0	22
8.0	24
9.0	26
10.0	28
11.0	32
12.0	32
13.0	36
14.0	38
15.0	38

Figure 2B.—Modulation index table.

IN ANSWERING QUESTIONS 2-20 AND 2-21, REFER TO FIGURE 2B.

2-20. An fm-modulated carrier varies between 925 kilohertz and 1,075 kilohertz 15,000 times per second. What is the bandwidth, in kilohertz, of the transmitted signal? (HINT: You will need to figure MI to be able to find the sidebands.)

1. 120
2. 240
3. 340
4. 420

2-21. The spectrum of a 500 kilohertz fm-modulated carrier has a 60-kilohertz bandwidth and contains 12 significant sidebands. How much, in kilohertz, is the carrier deviated?

1. ±5
2. ±7.5
3. ±10
4. ±15

2-22. In a reactance-tube modulator, the reactance tube shunts what part of the oscillator circuitry?

1. The amplifier
2. The tank circuit
3. The biasing network
4. The feedback network

2-23. With no modulating signal applied, a reactance tube has which, if any, of the following effects on the output of an oscillator?

1. It will decrease amplitude
2. It will increase amplitude
3. It will change resonant frequency
4. It will have no effect

2-24. The reactance-tube frequency modulates the oscillator by which of the following actions?

1. By shunting the tank circuit with a variable resistance
2. By shunting the tank circuit with a variable reactance
3. By shunting the tank circuit with a variable capacitance
4. By causing a resultant current flow in the tank circuit which either leads or lags resonant current

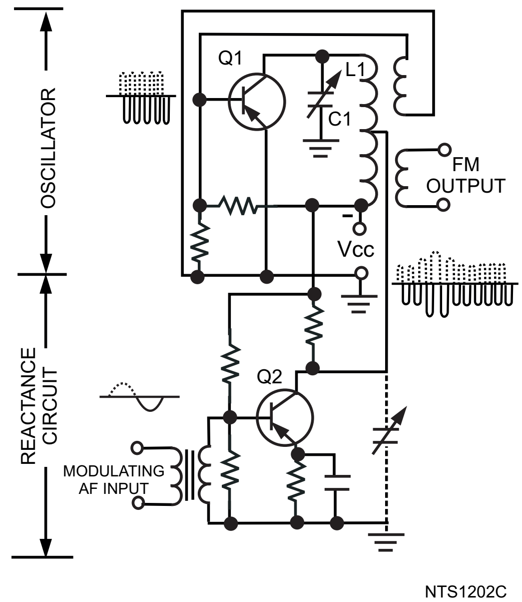


Figure 2C.—Semiconductor reactance modulator.

IN ANSWERING QUESTIONS 2-25 AND 2-26, REFER TO FIGURE 2C.

2-25. The semiconductor reactance modulator in the circuit is in parallel with a portion of the oscillator tank circuit coil. Modulation results because of interaction with which of the following transistor characteristics?

1. Collector-to-emitter resistance
2. Collector-to-emitter capacitance
3. Base-to-emitter resistance
4. Base-to-emitter capacitance

2-26. With a positive-going modulating signal applied to the base of Q2, (a) what will circuit capacitance do and (b) what will the output frequency do?

1. (a) Decrease (b) decrease
2. (a) Decrease (b) increase
3. (a) Increase (b) increase
4. (a) Increase (a) decrease

2-27. What type of circuit is used to remove the AM component in the output of a semiconductor reactance modulator?

1. A mixer
2. A filter
3. A limiter
4. A buffer amplifier

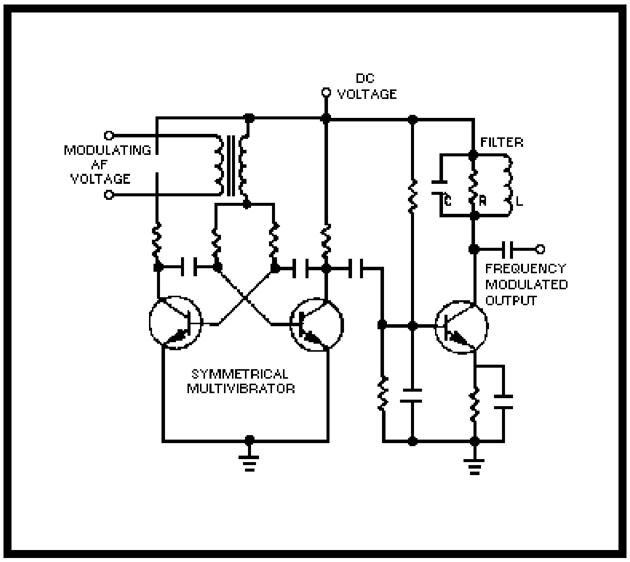


Figure 2D.—Multivibrator modulator.

IN ANSWERING QUESTIONS 2-28 AND 2-29, REFER TO FIGURE 2D.

2-28. The multivibrator modulator produces fm modulation by which of the following actions?

1. By modulating the collector voltages
2. By modulating the base-return voltages
3. By modulating the value of the base value of the base capacitors
4. By modulating the value of the base resistors

2-29. What is the purpose of the filter on the output of the multivibrator modulator?

1. To establish the fundamental operating frequency
2. To eliminate unwanted frequency variations
3. To eliminate unwanted odd harmonics
4. To eliminate unwanted even harmonics

2-30. A multivibrator frequency modulator is limited to frequencies below what maximum frequency?

1. 1 megahertz
2. 2 megahertz
3. 5 megahertz
4. 10 megahertz

2-31. To ensure the frequency stability of an fm transmitter, which, if any, of the following actions could be taken?

1. Modulate a crystal-controlled oscillator at the desired frequency
2. Modulate a low-frequency oscillator, and use frequency multipliers to achieve the operating frequency
3. Modulate a low-frequency oscillator, and heterodyne it with a higher-frequency oscillator to achieve the desired frequency
4. None of the above

2-32. A varactor is a variable device that acts as which of the following components?

1. Resistor
2. Inductor
3. Capacitor
4. Transistor

2-33. As the positive potential is increased on the cathode of a varactor, (a) what happens to reverse bias and (b) how is dielectric width affected?

1. (a) Increases (b) increases
2. (a) Increases (b) decreases
3. (a) Decreases (b) decreases
4. (a) Decreases (b) increases

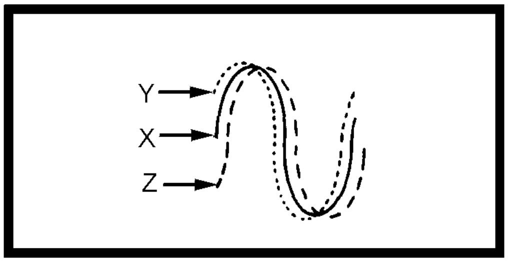


Figure 2E.—Phase relationships.

IN ANSWERING QUESTION 2-34, REFER TO FIGURE 2E.

2-34. In the figure, (a) waveform X has what phase relationship to waveform Y, and (b) waveform Y has what relationship to waveform Z?

1. (a) Lags (b) leads
2. (a) Lags (b) lags
3. (a) Leads (b) lags
4. (a) Leads (b) leads

2-35. A 10 kilohertz, 10-volt square wave is applied as the phase-modulating signal to a transmitter with a carrier frequency of 60 megahertz. What is the output frequency during the constant-amplitude portions of the modulating signal?

1. 10 kilohertz
2. 59,990 kilohertz
3. 60,000 kilohertz
4. 60,010 kilohertz

2-36. In a phase modulator, the frequency during the constant-amplitude portion of the modulating wave is the

1. peak frequency
2. rest frequency
3. deviation frequency
4. modulating frequency

2-37. In phase modulation, (a) the AMPLITUDE of the modulating signal determines what characteristic of the phase shift, and (b) the FREQUENCY of the modulating signal determines what characteristic of the phase shift?

1. (a) Rate (b) rate
2. (a) Rate (b) amount
3. (a) Amount (b) amount
4. (a) Amount (b) rate

2-38. The frequency spectrums of a phase-modulated signal resemble the spectrum of which, if any, of the following types of modulation?

1. Amplitude modulated
2. Frequency modulated
3. Continuous-wave modulated
4. None of the above

2-39. Compared to fm, increasing the modulating frequency in phase modulation has what effect, if any, on the bandwidth of the phase-modulated signal?

1. It increases
2. It decreases
3. None

2-40. A simple phase modulator consists of a capacitor in series with a variable resistance. What total amount of carrier shift will occur when X_C is 10 times the resistance?

1. 0 degrees
2. 45 degrees
3. 60 degrees
4. 90 degrees

2-41. The primary advantage of phase modulation over frequency modulation is that phase modulation has better carrier

1. power stability
2. amplitude stability
3. frequency stability
4. directional stability

2-42. Phase-shift keying is most useful under which of the following code element conditions?

1. When mark elements are longer than space elements
2. When mark elements are shorter than space elements
3. When mark and space elements are the same length
4. When mark and space elements are longer than synchronizing elements

2-43. When a carrier is phase-shift keying modulated, (a) a data bit ONE will normally cause the carrier to shift its phase what total number of degrees, and (b) a data bit ZERO will cause the carrier to shift its phase what total number of degrees?

1. (a) 60 (b) 0
2. (a) 0 (b) 180
3. (a) 180 (b) 180
4. (a) 180 (b) 0

2-44. Which of the following circuits is used to generate a phase-shift keyed signal?

1. Logic circuit
2. Phasor circuit
3. Phasitron circuit
4. Longitudinal circuit

2-45. When a carrier is modulated by a square wave, what maximum number of sideband pairs will be generated?

1. 1
2. 9
3. 3
4. An infinite number

2-46. As the square wave modulating voltage is increased to the same amplitude as that of the carrier, what will be the effect on (a) the carrier amplitude and (b) amplitude of the sidebands?

1. (a) Remains constant
   (b) Increases
2. (a) Decreases
   (b) Increases
3. (a) Increases
   (b) Remains constant
4. (a) Increases
   (b) Decreases

2-47. In a square-wave modulated signal, total sideband power is what percentage of the total power?

1. 0 percent
2. 25 percent
3. 33 percent
4. 50 percent

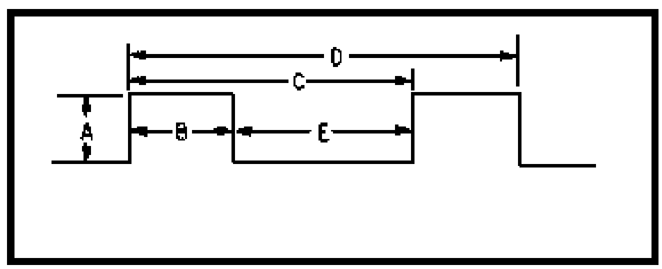


Figure 2F.—Waveform.

IN ANSWERING QUESTIONS 2-48 THROUGH 2-51, REFER TO FIGURE 2F. SELECT THE FIGURE LETTER THAT CORRESPONDS WITH THE WAVEFORM LISTED IN THE QUESTIONS. LETTERS MAY BE USED ONCE, MORE THAN ONCE, OR NOT AT ALL.

2-48. Pulse width.

1. A
2. B
3. C
4. D

2-49. Rest time.

1. B
2. C
3. D
4. E

2-50. Pulse duration.

1. A
2. B
3. D
4. E

2-51. Pulse-repetition time.

1. B
2. C
3. D
4. E

2-52. Which of the following ratios is used to determine pulse-repetition frequency (prf)?

1. $\text{Prf} = \frac{1}{\text{prt}}$
2. $\text{Prf} = \frac{1}{\text{pw}}$
3. $\text{Prf} = \frac{1}{\text{pd}}$
4. $\text{Prf} = \frac{1}{\text{rt}}$

2-53. Average power in a pulse-modulation system is defined as the

1. power during rest time
2. power during each pulse
3. power during each pulse averaged over rest time
4. power during each pulse averaged over one operating cycle

2-54. In pulse modulation, what term is used to indicate the ratio of time the system is actually producing rf?

1. Rest cycle
2. Duty cycle
3. Average cycle
4. Transmit cycle

2-55. In a pulse-modulation system, which of the following formulas is used to figure the percentage of transmitting time?

1. $\frac{\text{pw}}{\text{prt}} \times 100$
2. $\frac{\text{prt}}{\text{pw}} \times 100$
3. $\frac{\text{pw}}{\text{rt}} \times 100$
4. $\frac{\text{rt}}{\text{pw}} \div 100$

2-56. When pulse modulation is used for range finding in a radar application, which of the following types of pulse information is used?

1. Reflected pulse return interval
2. Reflected pulse duration
3. Reflected pulse amplitude
4. Reflected pulse frequency

2-57. In a spark-gap modulator, what is the function of the pulse-forming network?

1. To store energy
2. To increase the level of stored energy
3. To act as a power bleeder
4. To rapidly discharge stored energy

2-58. The damping diode in a thyratron modulator serves which of the following purposes?

1. It discharges the pulse-forming network
2. It limits the input signal
3. It prevents the breakdown of the thyratron by reverse-voltage transients
4. It rectifies the input signal

2-59. Compared to a spark-gap modulator, the thyratron modulator exhibits which of the following advantages?

1. Improved timing
2. Higher output pulses
3. Higher trigger voltage
4. Operates over a narrower range of anode voltages and pulse-repetition rates

2-60. To transmit intelligence using pulse modulation, which of the following pulse characteristics may be varied?

1. Pulse duration
2. Pulse amplitude
3. Pulse-repetition time
4. Each of the above

2-61. To accurately reproduce a modulating signal in a pulse-modulated system, what minimum number of samples must be taken per cycle?

1. One
2. Two
3. Three
4. Four

2-62. What is the simplest form of pulse modulation?

1. Pulse-code modulation
2. Pulse-duration modulation
3. Pulse-frequency modulation
4. Pulse-amplitude modulation

2-63. The same pulse characteristic is varied in which of the following types of pulse modulations?

1. Pam and pdm
2. Pdm and pwm
3. Pwm and ppm
4. Ppm and pam

THIS SPACE LEFT BLANK INTENTIONALLY.

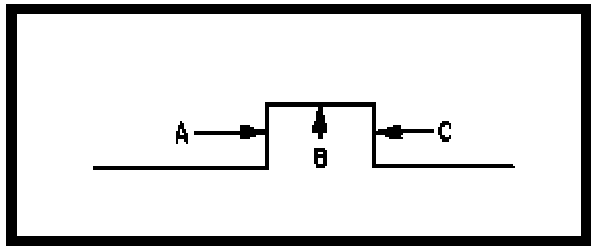


Figure 2G.—Waveform.

IN ANSWERING QUESTION 2-64, REFER TO FIGURE 2G.

2-64. Which of the points shown in the waveform may be varied in pulse-duration modulation?

1. A only
2. B only
3. C only
4. A and/or C

2-65. Which, if any, of the following is the primary disadvantage of pulse-position modulation?

1. It depends on transmitter-receiver synchronization
2. It is susceptible to noise interference
3. Transmitter power varies
4. None of the above

2-66. A pfm transmitter transmits 10,000 pulses per second without a modulating signal applied. How, if at all, will a modulating signal affect the transmitted pulse rate?

1. It will decrease the transmitted pulse rate
2. It will increase the transmitted pulse rate
3. Both 1 and 2 above
4. It will not affect the transmitted pulse rate

2-67. The process of arbitrarily dividing a wave into a series of standard values is referred to as

1. arbitration
2. quantization
3. interposition
4. approximation

2-68. A pcm system is capable of transmitting 32 standard levels that ate sampled 2.5 times per cycle of a 3-kilohertz modulating signal. What maximum number of bits per second are transmitted?

1. 18,750
2. 37,500
3. 75,000
4. 240,000

2-69. Which of the following is a characteristic of a pcm system that makes it advantageous for use in multiple-relay link systems?

1. Average power is constant
2. Average power decreases with each relay
3. Noise is not cumulative at relay stations
4. Quantization noise decreases with each relay

ASSIGNMENT 3

Textbook assignment: Chapter 3, "Demodulation," pages 3-1 through 3-35.

3-1. The process of recreating the original modulating frequencies (intelligence) from an rf carrier is known by which of the following terms?

1. Detection
2. Demodulation
3. Both 1 and 2 above
4. Distribution

3-2. When a demodulator fails to accurately recover intelligence from a modulated carrier, which of the following types of distortion result?

1. Phase
2. Frequency
3. Amplitude
4. Each of the above

3-3. In a demodulator circuit, which of the following components is required for demodulation to occur?

1. A linear device
2. A nonlinear device
3. A variable resistor

3-4. In cw demodulation, the first requirement of the circuit is the ability to detect

1. the presence or absence of the carrier
2. amplitude variations in the carrier
3. frequency variations in the carrier
4. phase variations in the carrier

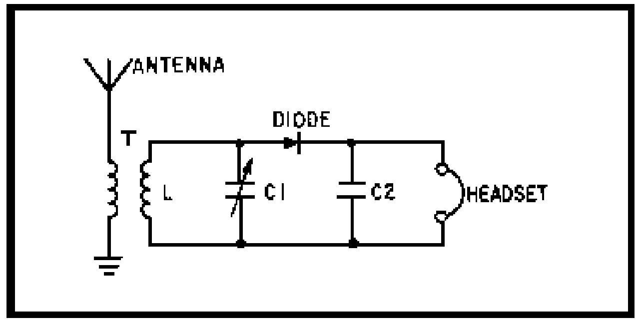


Figure 3A.—Cw demodulation.

IN ANSWERING QUESTION 3-5, REFER TO FIGURE 3A.

3-5. In the figure, L and C1 form a frequency-selective network that serves what purpose?

1. It removes the carrier
2. It rectifies the oscillations
3. It tunes the circuit to the desired rf carrier
4. It provides filtering to maintain a constant dc output

3-6. To aid in distinguishing between two or more cw signals that are close to the same frequency, which of the following detectors is used?

1. Diode
2. Crystal
3. Heterodyne
4. Transistor

3-7. Assume that two signals are received, one at 500 kHz and the other at 501 kHz. What frequency, in kHz, should be mixed with them to distinguish the 501 kHz signal by producing a 1 kHz output?

1. 499
2. 500
3. 501
4. 502

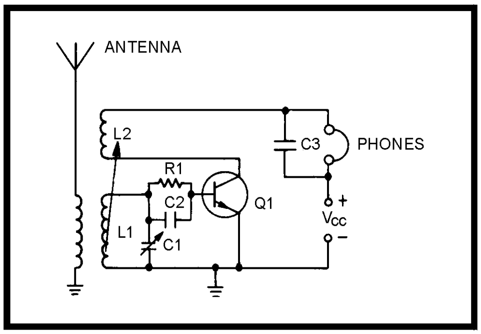


Figure 3B.—Detector.

IN ANSWERING QUESTIONS 3-8 THROUGH 3-11, REFER TO FIGURE 3B.

3-8. The detector circuit in the figure uses the heterodyning principle to detect the incoming signal. What type of detector is it?

1. Hartley
2. Colpitts
3. Armstrong
4. Regenerative

3-9. What component controls the operating frequency of the detector?

1. C1
2. C2
3. L2
4. R1

3-10. What component provides the feedback necessary for oscillations to occur?

1. R1
2. C2
3. C3
4. L2

3-11. Which of the following circuit functions does Q1 perform?

1. Mixer
2. Detector
3. Oscillator
4. Each of the above

3-12. A circuit that is nonlinear, provides filtering, and is sensitive to the type of modulation applied to it fulfills the requirements of which, if any, of the following circuits?

1. Mixer
2. Modulator
3. Demodulator
4. None of the above

3-13. A detector uses which of the following signals to approximate the original waveform?

1. The sum frequency
2. The carrier frequency
3. The modulation envelope

THIS SPACE LEFT BLANK INTENTIONALLY.

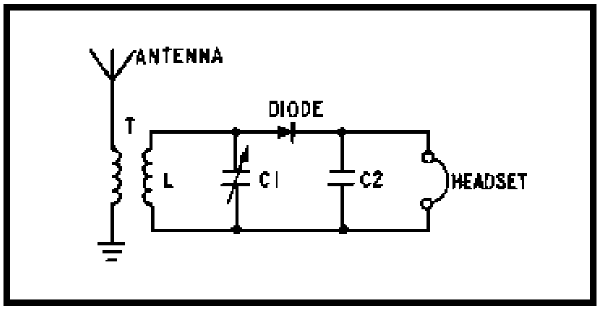


Figure 3C.—Detector.

IN ANSWERING QUESTIONS 3-14 THROUGH 3-16, REFER TO FIGURE 3C.

3-14. What type of detector is shown in the figure?

1. Series-diode
2. Parallel-diode
3. Inductive-diode
4. Capacitive-diode

3-15. What is the purpose of C1 and L?

1. To smooth the incoming rf
2. To select the desired af signal
3. To select the desired rf signal
4. To smooth the detected af signal

3-16. What is the purpose of C2?

1. To smooth the incoming rf signal
2. To select the desired af signal
3. To select the desired rf signal
4. To smooth the detected af signal

3-17. A shunt diode circuit is used as a detector in which of the following instances?

1. When a large input signal is supplied
2. When a large output current is required
3. When the input signal variations overdrive the audio amplifier stages
4. When the input signal variations are too small to produce a full output from audio amplifier stages

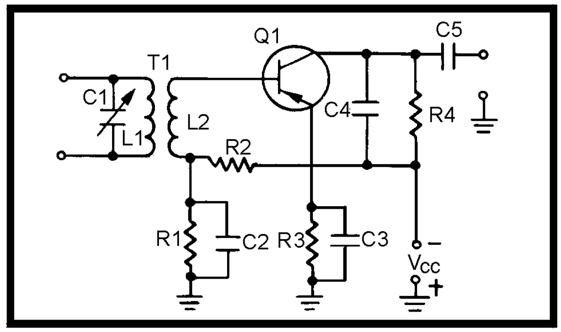


Figure 3D.—Detector.

IN ANSWERING QUESTIONS 3-18 THROUGH 3-21, REFER TO FIGURE 3D.

3-18. What type of detector is shown in the figure?

1. Ratio
2. Common-base
3. Regenerative
4. Common-emitter

3-19. What circuit component acts as the load for the detected audio?

1. R1
2. R2
3. R3
4. R4

3-20. What is the purpose of C4?

1. Tp bypass af
2. To bypass rf
3. To remove power supply voltage variations
4. To determine the operating frequency of the circuit

3-21. This detector circuit is used under which of the following circuit conditions?

1. When higher frequencies are used
2. When the best possible frequency selection is required
3. When weak signals need to be detected
4. When strong signals need to be detected

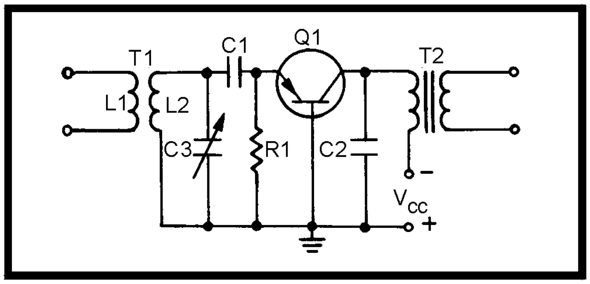


Figure 3E.—Circuit.

IN ANSWERING QUESTIONS 3-22 THROUGH 3-24, REFER TO FIGURE 3E.

3-22. What type of Circuits is/are shown in the figure?

1. A detector
2. An amplifier
3. Both 1 and 2 above
4. An oscillator

3-23. What is the purpose of T2?

1. To filter rf
2. To filter af
3. To couple the af output
4. To couple the rf output

3-24. What is the function of C1 and R1?

1. To act as an integrator
2. To act as a frequency-selective network
3. To act as a filter network
4. To act as a differentiator

THIS SPACE LEFT BLANK INTENTIONALLY.

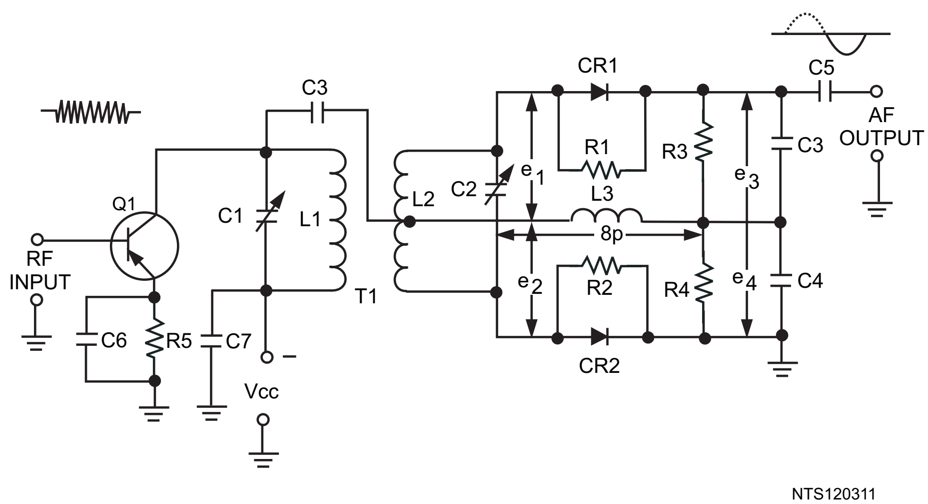


Figure 3F.—Foster-Seeley discriminator.

IN ANSWERING QUESTIONS 3-25 THROUGH 3-30, REFER TO FIGURE 3F.

3-25. In the figure, what is the purpose of Q1?

1. To act as a limiter only
2. To act as an amplifier only
3. To act as a limiter and an amplifier
4. To act as an oscillator

3-26. To what frequency are C1/L1 and C2/L2 tuned?

1. The af input
2. The center frequency of fm signal
3. The lowest fm deviation frequency
4. The highest fm deviation frequency

3-27. What is the function of L3?

1. To couple af to the output
2. To couple rf from the tank circuits to CR1 and CR2
3. To prevent af from being coupled to the power supply
4. To provide the dc return path for CR1 and CR2

3-28. At resonance, what is the amplitude of e_3 compared to e_4?

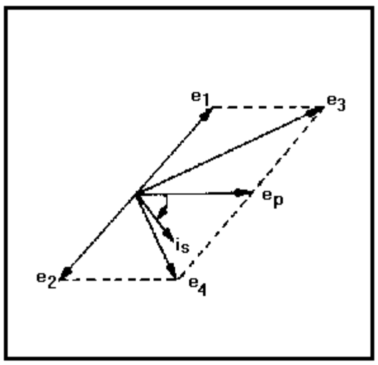
1. e_3 is less than e_4
2. e_3 is equal to e_4
3. e_3 is greater than e_4

3-29. When the circuit is operating ABOVE resonance, (a) does inductive reactance increase or decrease, and (b) does capacitive reactance increase or decrease?

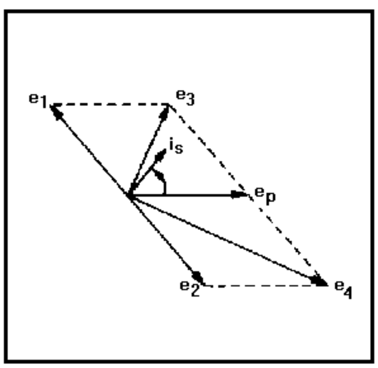
1. (a) Increases (b) increases
2. (a) Increases (b) decreases
3. (a) Decreases (b) decreases
4. (a) Decreases (b) increases

3-30. Circuit operation BELOW resonance is represented by which of the following vector diagrams?

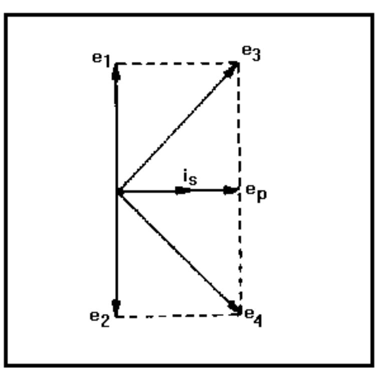
1.



2.



3.



THIS SPACE LEFT BLANK INTENTIONALLY.

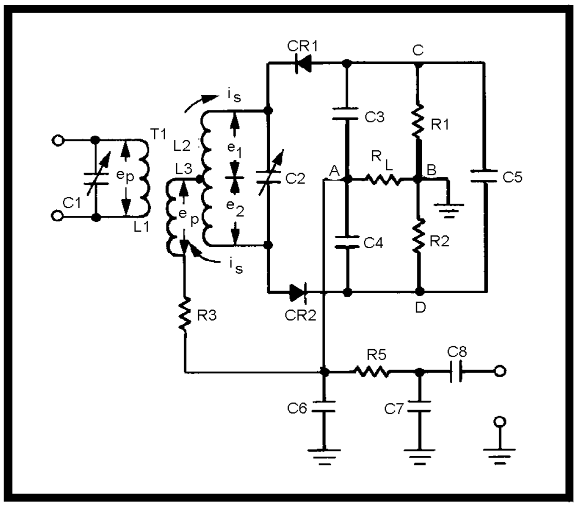


Figure 3G.—Ratio detector.

IN ANSWERING QUESTIONS 3-31 THROUGH 3-40, REFER TO FIGURE 3G.

3-31. To what frequency(ies) are (a) L1 and C1 and (b) L2 and C2 tuned?

1. (a) Center frequency
   (b) lower frequency limit
2. (a) Center frequency
   (b) center frequency
3. (a) Lower frequency limit
   (b) center frequency
4. (a) Lower frequency limit
   (b) lower frequency limit

3-32. What circuit filtering function do R5, C6, and C7 provide?

1. Low-pass
2. High-pass
3. Band-pass
4. Band-reject

3-33. At resonance, what type of circuit does the tank circuit appear to be?

1. Reactive
2. Resistive
3. Inductive
4. Capacitive

3-34. At resonance, what is the phase relationship between tank current and primary voltage?

1. Tank current leads primary voltage by 90 degrees
2. Tank current lags primary voltage by 90 degrees
3. Tank current and primary voltage are in phase
4. Tank current and primary voltage are out of phase

3-35. At resonance, what relative amount of conduction takes place through CR1 compared to that for CR2?

1. CR1 conducts more than CR2
2. CR1 conducts less than CR2
3. CR1 and CR2 conduct the same amount

3-36. At resonance, (a) will the charges on C3 and C4 be equal or unequal, and (b) will their polarities be the same or opposite?

1. (a) Equal (b) same
2. (a) Equal (b) opposite
3. (a) Unequal (b) opposite
4. (a) Unequal (b) same

3-37. ABOVE resonance, both voltages e_1 and e_2 have specific phase shift relationships to voltage e in that they either shift nearer to or farther from the phase of e_p. What are the phase relationships between (a) e_1 and e_p and (b) e_2 and e_p?

1. (a) e_1 is nearer to e_p
   (b) e_2 is nearer to e_p
2. (a) e_1 is nearer to e_p
   (b) e_2 is farther from e_p
3. (a) e_1 is farther from e_p
   (b) e_2 is nearer to e_p
4. (a) e_1 is farther from e_p
   (b) e_2 is farther from e_p

3-38. If C3 is charged to 6 volts and C4 is charged to 4 volts, what is the output voltage?

1. 1 volt
2. 2 volts
3. 3 volts
4. 4 volts

3-39. When operating BELOW resonance, what is the relationship of the vector sum of e_1 and e_p to the vector sum of e_2 and e_p?

1. The sum of e_1 and e_p is larger than the sum of e_2 and e_p
2. The sum of e_1 and e_p is smaller than the sum of e_2 and e_p
3. The slam of e_1 and e_p is equal to the sum of e_2 an e_p

3-40. What components help to reduce the effects of amplitude variations at the input of the circuit?

1. R1, R2, and C5
2. R5, C6, and C7
3. R1, R2, C3, and C4
4. L1, L2, L3, and C2

3-41. What is the minimum input voltage, in millivolts, required for a ratio detector?

1. 100
2. 200
3. 300
4. 400

3-42. Which of the following circuit functions is performed by the gated-beam detector?

1. Limiter
2. Detector
3. Amplifier
4. Each of the above

THIS SPACE LEFT BLANK INTENTIONALLY.

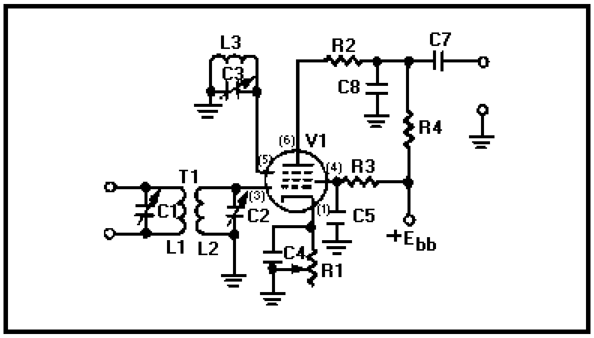


Figure 3H.—Gated-beam detector.

IN ANSWERING QUESTIONS 3-43 THROUGH 3-47, REFER TO FIGURE 3H.

3-43. What components in the circuit are used to set the reference frequency for a gated-beam detector?

1. C1 and L1
2. C2 and L2
3. C3 and L3
4. C4 and R1

3-44. What tube pins connect to elements that perform in a manner similar to an AND gate in a digital device?

1. Pins 1 and 3
2. Pins 3 and 4
3. Pins 3 and 5
4. Pins 4 and 5

3-45. What type of tank circuit is the quadrature tank (L3 and C3)?

1. Low-Q
2. High-Q
3. Nonresonant
4. Series-resonant

3-46. For plate current to flow, what must be the polarities of (a) the quadrature grid and (b) the limiter grid?

1. (a) Negative (b) negative
2. (a) Negative (b) positive
3. (a) Positive (b) positive
4. (a) Positive (b) negative

3-47. ABOVE the center frequency of the received fm signal, (a) will the tank appear capacitive or inductive, and (b) will the average plate current increase or decrease?

1. (a) Inductive (b) increase
2. (a) Inductive (b) decrease
3. (a) Capacitive (b) decrease
4. (a) Capacitive (b) increase

3-48. To demodulate a phase-modulated signal, which, if any, of the following types of demodulators may be used?

1. Peak
2. Quadrature
3. Series-diode
4. None of the above

3-49. Which of the following circuits can be used as a communications pulse demodulator?

1. Conversion
2. Peak detector
3. Low-pass filter
4. Each of the above

THIS SPACE LEFT BLANK INTENTIONALLY.

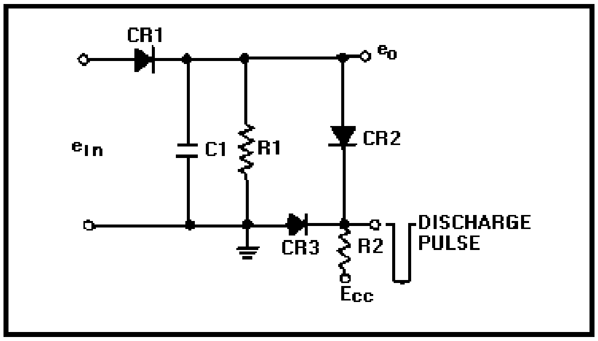


Figure 3I.—Detector.

IN ANSWERING QUESTIONS 3-50 THROUGH 3-52, REFER TO FIGURE 3I.

3-50. To detect pulse-amplitude modulation, what value must the RC time constant of R1 and C1 in the circuit be?

1. Five times the pulse width
2. Ten times the pulse width
3. Five times the interpulse period
4. Ten times the interpulse period

3-51. Which, if any, of the following functions is the purpose of CR2?

1. To quickly discharge C1 between received pulses
2. To rectify input pulses
3. To clamp the output to a positive level
4. None of the above

3-52. What change must be made to the circuit to detect pulse-duration modulation?

1. Remove R1
2. Increase the value of R1
3. Decrease the value of R1
4. Add a resistor in series with CR1

3-53. When a pulse-duration modulated signal is determined by using a low-pass filter, what characteristic of the signal is used?

1. Width
2. Amplitude
3. Frequency
4. Pulse position

3-54. To detect pulse-duration modulation, the low-pass filter components must be selected so that they pass only the

1. carrier frequency
2. intermediate frequency
3. pulse-repetition frequency
4. desired modulating frequency

3-55. What type(s) of modulation is/are normally detected by first converting it/them to another type of modulation?

1. Ppm only
2. Pfm only
3. Pcm only
4. Ppm, pfm, and pcm

3-56. What type of circuit can be used to convert from ppm to pdm for demodulation?

1. An amplifier
2. A flip-flop
3. An oscillator
4. A transformer

THIS SPACE LEFT BLANK INTENTIONALLY.

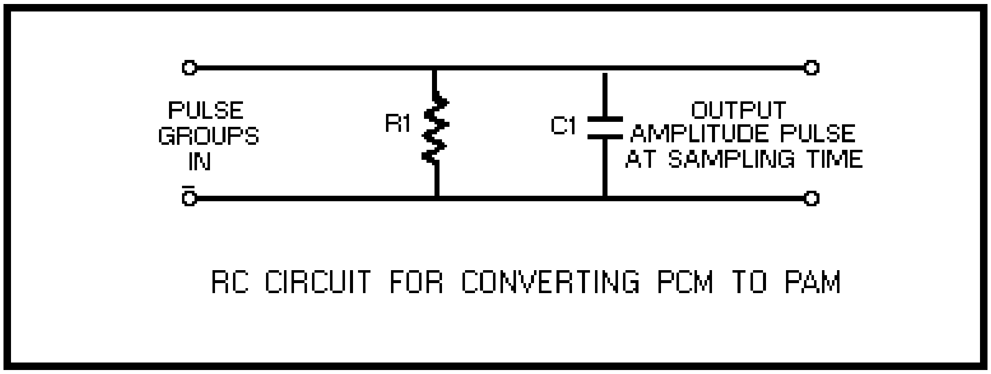


Figure 3J.—Pcm conversion.

IN ANSWERING QUESTIONS 3-57 THROUGH 3-59, REFER TO FIGURE 3J.

3-57. To convert from pcm to pam, what type of circuit is used to apply the pcm to the input of the circuit shown?

1. A constant-current source
2. A constant-voltage source
3. A limiter-amplifier source
4. An oscillator-amplifier source

3-58. If C1 is allowed to charge 16 volts during the period of one pulse, each additional pulse increases the charge by 16 volts. With the binary-coded equivalent of an analog 12 applied to the input, what will be the output of the circuit at sampling time?

1. 10 volts
2. 12 volts
3. 14 volts
4. 16 volts

3-59. Between pulses, R1 must allow C1 to discharge to what voltage?

1. 0 volts
2. One fourth of the charge on C1
3. One half of the charge on C1
4. Three fourths of the charge on C1

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 13—Introduction to Number Systems and Logic Circuits

NAVEDTRA 14185

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 13 of a series.

History of the course:

Sep 1998: *Original edition released. Prepared by FTCS(SS) Steven F. Reith.*
Jan 2004: *Administrative update. Reviewed by ETC(SW/AW) Scott Collie. Graphics updated No change in technical content.*

NAVSUP Logistics Tracking Number
0504-LP-026-8380

TABLE OF CONTENTS

ASSIGNMENT QUESTIONS follow Index.

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

CHAPTER 1

NUMBER SYSTEMS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the NRTC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completion of this chapter, you should be able to do the following:

1. Recognize different types of number systems as they relate to computers.

2. Identify and define unit, number, base/radix, positional notation, and most and least significant digits as they relate to decimal, binary, octal, and hexadecimal number systems.

3. Add and subtract in binary, octal, and hexadecimal number systems.

4. Convert values from decimal, binary, octal, hexadecimal, and binary-coded decimal number systems to each other and back to the other systems.

5. Add in binary-coded decimal.

INTRODUCTION

How many days' leave do you have on the books? How much money do you have to last until payday? It doesn't matter what the question is—if the answer is in dollars or days or cows, it will be represented by numbers.

Just try to imagine going through one day without using numbers. Some things can be easily described without using numbers, but others prove to be difficult. Look at the following examples:

I am stationed on the aircraft carrier *Nimitz*.

He owns a green Chevrolet.

The use of numbers wasn't necessary in the preceding statements, but the following examples depend on the use of numbers:

I have $25 to last until payday.

I want to take 14 days' leave.

You can see by these statements that numbers play an important part in our lives.

BACKGROUND AND HISTORY

Man’s earliest number or counting system was probably developed to help determine how many possessions a person had. As daily activities became more complex, numbers became more important in trade, time, distance, and all other phases of human life.

As you have seen already, numbers are extremely important in your military and personal life. You realize that you need more than your fingers and toes to keep track of the numbers in your daily routine.

Ever since people discovered that it was necessary to count objects, they have been looking for easier ways to count them. The abacus, developed by the Chinese, is one of the earliest known calculators. It is still in use in some parts of the world.

Blaise Pascal (French) invented the first adding machine in 1642. Twenty years later, an Englishman, Sir Samuel Moreland, developed a more compact device that could multiply, add, and subtract. About 1672, Gottfried Wilhelm von Leibniz (German) perfected a machine that could perform all the basic operations (add, subtract, multiply, divide), as well as extract the square root. Modern electronic digital computers still use von Liebniz's principles.

MODERN USE

Computers are now employed wherever repeated calculations or the processing of huge amounts of data is needed. The greatest applications are found in the military, scientific, and commercial fields. They have applications that range from mail sorting, through engineering design, to the identification and destruction of enemy targets. The advantages of digital computers include speed, accuracy, and man-power savings. Often computers are able to take over routine jobs and release personnel for more important work—work that cannot be handled by a computer.

People and computers do not normally speak the same language. Methods of translating information into forms that are understandable and usable to both are necessary. Humans generally speak in words and numbers expressed in the decimal number system, while computers only understand coded electronic pulses that represent digital information.

In this chapter you will learn about number systems in general and about binary, octal, and hexadecimal (which we will refer to as hex) number systems specifically. Methods for converting numbers in the binary, octal, and hex systems to equivalent numbers in the decimal system (and vice versa) will also be described. You will see that these number systems can be easily converted to the electronic signals necessary for digital equipment.

TYPES OF NUMBER SYSTEMS

Until now, you have probably used only one number system, the decimal system. You may also be familiar with the Roman numeral system, even though you seldom use it.

THE DECIMAL NUMBER SYSTEM

In this module you will be studying modern number systems. You should realize that these systems have certain things in common. These common terms will be defined using the decimal system as our base. Each term will be related to each number system as that number system is introduced.

Each of the number systems you will study is built around the following components: the UNIT, NUMBER, and BASE (RADIX).

Unit and Number

The terms *unit* and *number* when used with the decimal system are almost self-explanatory. By definition the unit is a single object; that is, an apple, a dollar, a day. A number is a symbol representing a unit or a quantity. The figures 0, 1, 2, and 3 through 9 are the symbols used in the decimal system. These symbols are called Arabic numerals or figures. Other symbols may be used for different number systems. For example, the symbols used with the Roman numeral system are letters — V is the symbol for 5, X for 10, M for 1,000, and so forth. We will use Arabic numerals and letters in the number system discussions in this chapter.

Base (Radix)

The base, or radix, of a number system tells you the number of symbols used in that system. The base of any system is always expressed in decimal numbers. The base, or radix, of the decimal system is 10. This means there are 10 symbols — 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 — used in the system. A number system using three symbols — 0, 1, and 2 — would be base 3; four symbols would be base 4; and so forth. Remember to count the zero or the symbol used for zero when determining the number of symbols used in a number system.

The base of a number system is indicated by a subscript (decimal number) following the value of the number. The following are examples of numerical values in different bases with the subscript to indicate the base:

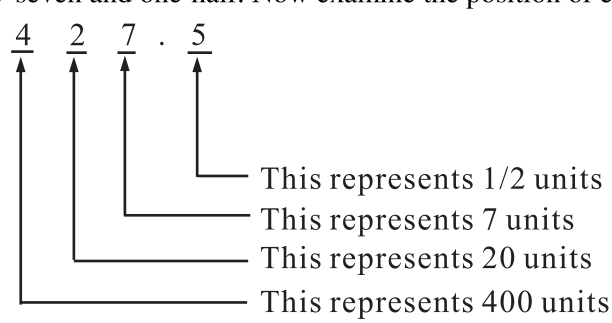
$$7592_{10} \qquad 214_5 \quad 123_4 \quad 656_7$$

You should notice the highest value symbol used in a number system is always one less than the base of the system. In base 10 the largest value symbol possible is 9; in base 5 it is 4; in base 3 it is 2.

Positional Notation and Zero

You must observe two principles when counting or writing quantities or numerical values. They are the POSITIONAL NOTATION and the ZERO principles.

Positional notation is a system where the value of a number is defined not only by the symbol but by the symbol's position. Let's examine the decimal (base 10) value of 427.5. You know from experience that this value is four hundred twenty-seven and one-half. Now examine the position of each number:



NTS130101

If 427.5 is the quantity you wish to express, then each number must be in the position shown. If you exchange the positions of the 2 and the 7, then you change the value.

Each position in the positional notation system represents a power of the base, or radix. A POWER is the number of times a base is multiplied by itself. The power is written above and to the right of the base and is called an EXPONENT. Examine the following base 10 line graph:

Radix Point

$10^3 \quad 10^2 \quad 10^1 \quad 10^0 \quad . \quad 10^{-1} \quad 10^{-2} \quad 10^{-3}$

10^3	= 10 x 100, or 1000
10^2	= 10 x 10, or 100
10^1	= 10 x 1, or 10
10^0	= (any number raised to the power of 0 equals 1)
10^{-1}	= 1 + 10, or .1
10^{-2}	= 1 + 100. or .01
10^{-3}	= 1 + 1000, or .001

NTS130102

Now let's look at the value of the base 10 number 427.5 with the positional notation line graph:

Radix Point

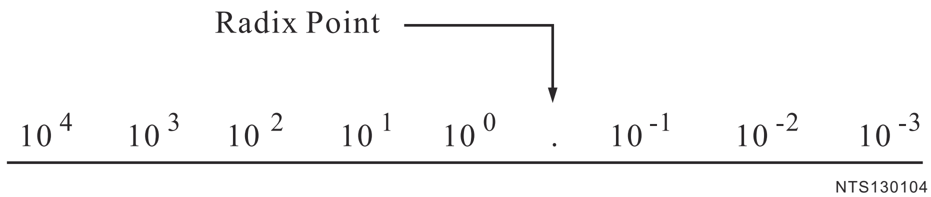
10^2	10^1	10^0	.	10^{-1}
4	2	7	.	5

10^2	= 4 x 100, or 400
10^1	= 2 x 10, or 20
10^0	= 7 x 1, or 7
10^{-1}	= 5 x .1, or .5

NTS130103

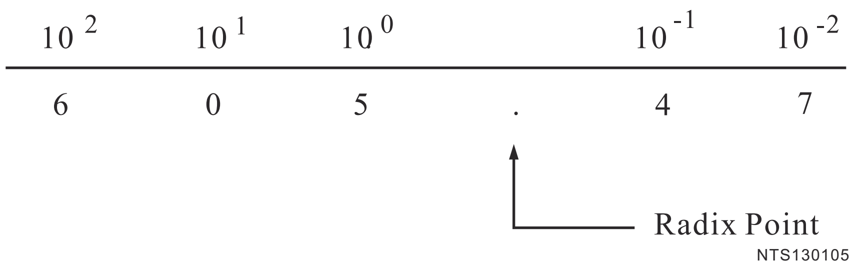
You can see that the power of the base is multiplied by the number in that position to determine the value for that position.

The following graph illustrates the progression of powers of 10:



All numbers to the left of the decimal point are whole numbers, and all numbers to the right of the decimal point are fractional numbers. A whole number is a symbol that represents one, or more, complete objects, such as one apple or $5. A fractional number is a symbol that represents a portion of an object, such as half of an apple (.5 apples) or a quarter of a dollar ($0.25). A mixed number represents one, or more, complete objects, and some portion of an object, such as one and a half apples (1.5 apples). When you use any base other than the decimal system, the division between whole numbers and fractional numbers is referred to as the RADIX POINT. The decimal point is actually the radix point of the decimal system, but the term radix point is normally not used with the base 10 number system.

Just as important as positional notation is the use of the zero. The placement of the zero in a number can have quite an effect on the value being represented. Sometimes a position in a number does not have a value between 1 and 9. Consider how this would affect your next paycheck. If you were expecting a check for $605.47, you wouldn't want it to be $65.47. Leaving out the zero in this case means a difference of $540.00. In the number 605.47, the zero indicates that there are no tens. If you place this value on a bar graph, you will see that there are no multiples of 10^1.



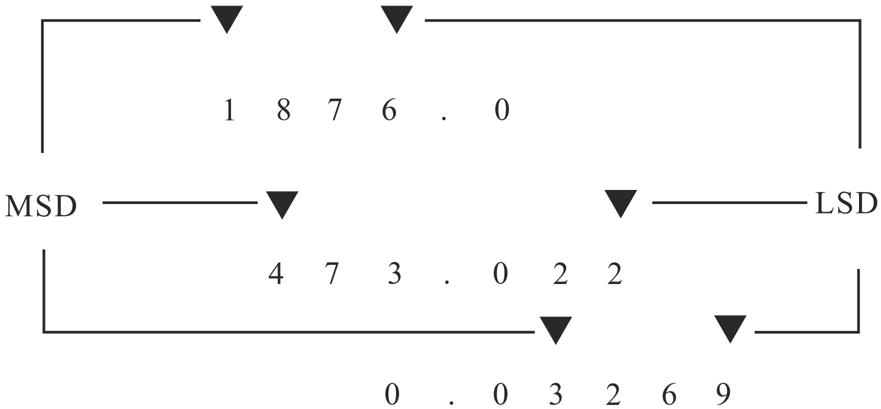
Most Significant Digit and Least Significant Digit (MSD and LSD)

Other important factors of number systems that you should recognize are the MOST SIGNIFICANT DIGIT (MSD) and the LEAST SIGNIFICANT DIGIT (LSD).

The MSD in a number is the digit that has the *greatest* effect on that number.

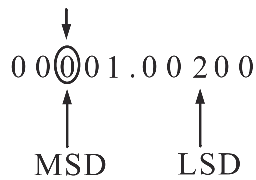
The LSD in a number is the digit that has the *least* effect on that number.

Look at the following examples:



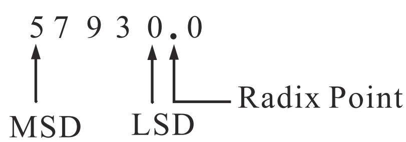
NTS130106

You can easily see that a change in the MSD will increase or decrease the value of the number the greatest amount. Changes in the LSD will have the smallest effect on the value. The nonzero digit of a number that is the farthest LEFT is the MSD, and the nonzero digit farthest RIGHT is the LSD, as in the following example:



NTS130107

In a whole number the LSD will always be the digit immediately to the left of the radix point.



NTS130108

Q1. What term describes a single object?

Q2. A symbol that represents one or more objects is called a _________.

Q3. The symbols 0, 1, 2, and 3 through 9 are what type of numerals?

Q4. What does the base, or radix, of a number system tell you about the system?

Q5. How would you write one hundred seventy-three base 10?

Q6. What power of 10 is equal to 1,000? 100? 10? 1?

Q7. The decimal point of the base 10 number system is also known as the _________.

Q8. What is the MSD and LSD of the following numbers

(a) 420.

(b) 1045.06

(c) 0.0024

(d) 247.0001

Carry and Borrow Principles

Soon after you learned how to count, you were taught how to add and subtract. At that time, you learned some concepts that you use almost everyday. Those concepts will be reviewed using the decimal system. They will also be applied to the other number systems you will study.

ADDITION—Addition is a form of counting in which one quantity is added to another. The following definitions identify the basic terms of addition:

AUGEND—The quantity to which an addend is added

ADDEND—A number to be added to a preceding number

SUM —The result of an addition (the sum of 5 and 7 is 12)

CARRY —A carry is produced when the sum of two or more digits in a vertical column equals or exceeds the base of the number system in use

How do we handle the carry; that is, the two-digit number generated when a carry is produced? The lower order digit becomes the sum of the column being added; the higher order digit (the carry) is added to the next higher order column. For example, let's add 15 and 7 in the decimal system:

1	Carry
15	Augend
+7	Addend
22	Sum

NTS130109

Starting with the first column, we find the sum of 5 and 7 is 12. The 2 becomes the sum of the lower order column and the 1 (the carry) is added to the upper order column. The sum of the upper order column is 2. The sum of 15 and 7 is, therefore, 22.

The rules for addition are basically the same regardless of the number system being used. Each number system, because it has a different number of digits, will have a unique digit addition table. These addition tables will be described during the discussion of the adding process for each number system.

A decimal addition table is shown in table 1-1. The numbers in row X and column Y may represent either the addend or the augend. If the numbers in X represent the augend, then the numbers in Y must represent the addend and vice versa. The sum of X + Y is located at the point in array Z where the selected X row and Y column intersect.

Table 1-1. —Decimal Addition Table

+	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9	10
2	2	3	4	5	6	7	8	9	10	11
3	3	4	5	6	7	8	9	10	11	12
4	4	5	6	7	8	9	10	11	12	13
5	5	6	7	8	9	10	11	12	13	14
6	6	7	8	9	10	11	12	13	14	15
7	7	8	9	10	11	12	13	14	15	16
8	8	9	10	11	12	13	14	15	16	17
9	9	10	11	12	13	14	15	16	17	18

X

Z

Y

To add 5 and 7 using the table, first locate one number in the X row and the other in the Y column. The point in field Z where the row and column intersect is the sum. In this case the sum is 12.

SUBTRACTION.—The following definitions identify the basic terms you will need to know to understand subtraction operations:

- SUBTRACT—To take away, as a part from the whole or one number from another
- MINUEND—The number from which another number is to be subtracted
- SUBTRAHEND—The quantity to be subtracted
- REMAINDER, or DIFFERENCE—That which is left after subtraction
- BORROW—To transfer a digit (equal to the base number) from the next higher order column for the purpose of subtraction.

Use the rules of subtraction and subtract 8 from 25. The form of this problem is probably familiar to you:

$$\begin{array}{rrl} & {}^{1}\,{}^{15} & \text{Carry} \\ \longrightarrow & \cancel{2}\cancel{5} & \text{Minuend} \\ \longrightarrow & \underline{-8} & \text{Subtrahend} \\ & 17 & \text{Difference} \end{array}$$

NTS130110

It requires the use of the *borrow*; that is, you cannot subtract 8 from 5 and have a positive difference. You must borrow a 1, which is really one group of 10. Then, one group of 10 plus five groups of 1 equal 15, and 15 minus 8 leaves a difference of 7. The 2 was reduced by 1 by the borrow; and since nothing is to be subtracted from it, it is brought down to the difference.

Since the process of subtraction is the opposite of addition, the addition table 1-1 may be used to illustrate subtraction facts for any number system we may discuss.

In addition,

$$X + Y = Z$$

In subtraction, the reverse is true; that is,

$$Z - Y = X$$

OR

$$Z - X = Y$$

Thus, in subtraction the minuend is always found in array Z and the subtrahend in either row X or column Y. If the subtrahend is in row X, then the remainder will be in column Y. Conversely, if the subtrahend is in column Y, then the difference will be in row X. For example, to subtract 8 from 15, find 8 in either the X row or Y column. Find where this row or column intersects with a value of 15 for Z; then move to the remaining row or column to find the difference.

THE BINARY NUMBER SYSTEM

The simplest possible number system is the BINARY, or base 2, system. You will be able to use the information just covered about the decimal system to easily relate the same terms to the binary system.

Unit and Number

The base, or radix—you should remember from our decimal section—is the number of symbols used in the number system. Since this is the base 2 system, only two symbols, 0 and 1, are used. The base is indicated by a subscript, as shown in the following example:

$$1_2$$

When you are working with the decimal system, you normally don't use the subscript. Now that you will be working with number systems other than the decimal system, it is important that you use the subscript so that you are sure of the system being referred to. Consider the following two numbers:

$$11 \qquad 11$$

With no subscript you would assume both values were the same. If you add subscripts to indicate their base system, as shown below, then their values are quite different:

$$11_{10} \qquad 11_2$$

The base ten number 11_{10} is eleven, but the base two number 11_2 is only equal to three in base ten. There will be occasions when more than one number system will be discussed at the same time, so you MUST use the proper Subscript.

Positional Notation

As in the decimal number system, the principle of positional notation applies to the binary number system. You should recall that the decimal system uses powers of 10 to determine the value of a position. The binary system uses powers of 2 to determine the value of a position. A bar graph showing the positions and the powers of the base is shown below:

$2^4 \; 2^3 \; 2^2 \; 2^1 \; 2^0 \,.\; 2^{-1} \; 2^{-2} \; 2^{-3}$
2^4 is equal to 2 x 2 x 2 x 2, or 16_{10}
2^3 is equal to 2 x 2 x 2, or 8_{10}
2^2 is equal to 2 x 2, or 4_{10}
2^1 is equal to 2 x 1, or 2_{10}
2^0 is equal to 1_{10}
2^{-1} is equal to 1/2, or $.5_{10}$
2^{-2} is equal to 1/4, or $.25_{10}$
2^{-3} is equal to 1/8, or $.125_{10}$

NTS130111

All numbers or values to the left of the radix point are whole numbers, and all numbers to the right of the radix point are fractional numbers.

Let's look at the binary number 101.1 on a bar graph:

$$\begin{array}{cccccc} 2^2 & 2^1 & 2^0 & . & 2^{-2} \\ \hline 1 & 0 & 1 & . & 1 \end{array}$$

NTS130112

Working from the radix point to the right and left, you can determine the decimal equivalent:

$$\begin{aligned} 1 \text{ x } 2^{-1} &= .5_{10} \\ 1 \text{ x } 2^{0} &= 1.0_{10} \\ 1 \text{ x } 2^{1} &= 0.0_{10} \\ 1 \text{ x } 2^{2} &= 4.0_{10} \\ \hline & 5.5_{10} \end{aligned}$$

NTS130113

Table 1-2 provides a comparison of decimal and binary numbers. Notice that each time the total number of binary symbol positions increase, the binary number indicates the next higher power of 2. By this example, you can also see that more symbol positions are needed in the binary system to represent the equivalent value in the decimal system.

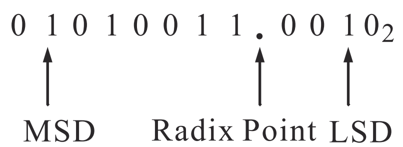
Table 1-2. —Decimal and Binary Comparison

	DECIMAL	BINARY	
10^0	0	0	2^0
	1	1	
	2	10	2^1
	3	11	
	4	100	2^2
	5	101	
	6	110	
	7	111	
	8	1000	2^3
	9	1001	
10^1	10	1010	
	11	1011	
	12	1100	
	13	1101	
	14	1110	
	15	1111	
	16	10000	2^4
	17	10001	
	18	10010	
	19	10011	
	20	10100	

NTS1301T2

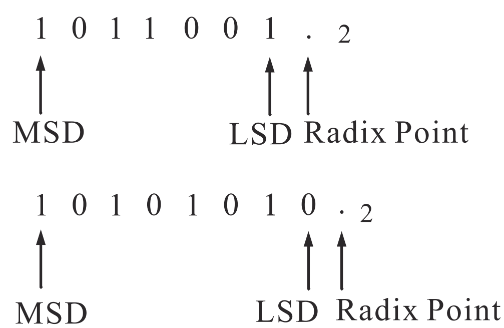
MSD and LSD

When you're determining the MSD and LSD for binary numbers, use the same guidelines you used with the decimal system. As you read from left to right, the first nonzero digit you encounter is the MSD, and the last nonzero digit is the LSD.



NTS130114

If the number is a whole number, then the first digit to the left of the radix point is the LSD.



NTS130115

Here, as in the decimal system, the MSD is the digit that will have the most effect on the number; the LSD is the digit that will have the least effect on the number.

The two numerals of the binary system (1 and 0) can easily be represented by many electrical or electronic devices. For example, 1_2 may be indicated when a device is active (on), and 0_2 may be indicated when a device is nonactive (off).

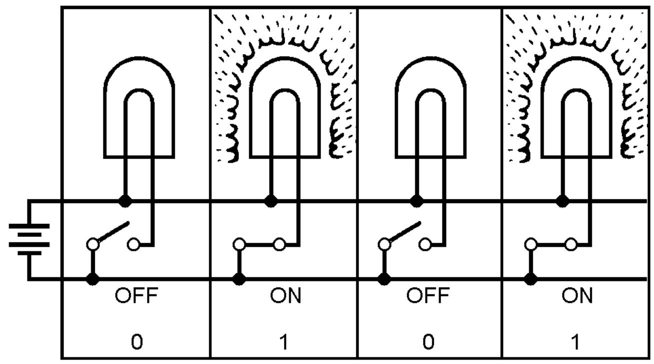


Figure 1-1. —Binary Example

Look at the preceding figure. It illustrates a very simple binary counting device. Notice that 1_2 is indicated by a lighted lamp and 0_2 is indicated by an unlighted lamp. The reverse will work equally well. The unlighted state of the lamp can be used to represent a binary 1 condition, and the lighted state can represent the binary 0 condition. Both methods are used in digital computer applications. Many other devices are used to represent binary conditions. They include switches, relays, diodes, transistors, and integrated circuits (ICs).

Addition of Binary Numbers

Addition of binary numbers is basically the same as addition of decimal numbers. Each system has an augend, an addend, a sum, and carries. The following example will refresh your memory:

1	Carry
15	Augend
+7	Addend
22	Sum

NTS130116

Since only two symbols, 0 and 1, are used with the binary system, only four combinations of addition are possible.

0 + 0

1 + 0

0 + 1

1 + 1

The sum of each of the first three combinations is obvious:

$$0 + 0 = 0_2$$

$$0 + 1 = 1_2$$

$$1 + 0 = 1_2$$

The fourth combination presents a different situation. The sum of 1 and 1 in any other number system is 2, but the numeral 2 does not exist in the binary system. Therefore, the sum of 1_2 and 1_2 is 10_2 (spoken as one zero base two), which is equal to 2_{10}.

$$\begin{array}{rl} {}^{1} & \text{Carry} \\ 1_2 & \text{Augend} \\ + \underline{1_2} & \text{Addend} \\ 10_2 & \text{Sum} \end{array}$$

NTS130117

Study the following examples using the four combinations mentioned above:

$$\begin{array}{rl} 101_2 & \text{Augend} \\ + \underline{010_2} & \text{Addend} \\ 111_2 & \text{Sum} \end{array}$$

NTS130118

$$\begin{array}{rl} {}^{1} & \text{Carry} \\ 101_2 & \text{Augend} \\ + \underline{101_2} & \text{Addend} \\ 1010_2 & \text{Sum} \end{array}$$

NTS130119

When a carry is produced, it is noted in the column of the next higher value or in the column immediately to the left of the one that produced the carry.

Example: Add 1011_2 and 1101_2.

Solution: Write out the problem as shown:

$$\begin{array}{rl} 1011_2 & \text{Augend} \\ + \underline{1101_2} & \text{Addend} \end{array}$$

NTS130120

As we noted previously, the sum of 1 and 1 is 2, which cannot be expressed as a single digit in the binary system. Therefore, the sum of 1 and 1 produces a carry:

$$\begin{array}{rl} {}^{1} & \text{Carry} \\ 1011_2 & \text{Augend} \\ + \underline{1101_2} & \text{Addend} \\ 0_2 & \text{Sum} \end{array}$$

NTS130121

The following steps, with the carry indicated, show the completion of the addition:

Previous Carry Used

$$\begin{array}{rl} 1\cancel{1} & \text{Carry} \\ 1011_2 & \text{Augend} \\ +\ \underline{1101_2} & \text{Addend} \\ 00_2 & \end{array}$$

NTS130122

When the carry is added, it is marked through to prevent adding it twice.

Previous Carry Used

$$\begin{array}{rl} 1\cancel{1}\cancel{1} & \text{Carry} \\ 1011_2 & \text{Augend} \\ +\ \underline{1101_2} & \text{Addend} \\ 000_2 & \end{array}$$

NTS130123

Previous Carry Used

$$\begin{array}{rl} 1\cancel{1}\cancel{1}\cancel{1} & \text{Carry} \\ 1011_2 & \text{Augend} \\ +\ \underline{1101_2} & \text{Addend} \\ 11000_2 & \end{array}$$

NTS130124

In the final step the remaining carry is brought down to the sum.

In the following example you will see that more than one carry may be produced by a single column. This is something that does not occur in the decimal system.

Example: Add 1_2, 1_2, 1_2, and 1_2

$$\begin{array}{rl} 1_2 & \text{Augend} \\ 1_2 & \text{1st Addend} \\ 1_2 & \text{2nd Addend} \\ +\ \underline{1_2} & \text{3rd Addend} \end{array}$$

The sum of the augend and the first addend is 0 with a carry. The sum of the second and third addends is also 0 with a carry. At this point the solution resembles the following example:

$$\begin{array}{rl} 1 & \text{Carry} \\ 1 & \text{Carry} \\ 1_2 & \text{Augend} \\ 1_2 & \text{1st Addend} \\ 1_2 & \text{2nd Addend} \\ +\ \underline{1_2} & \text{3rd Addend} \\ 0_2 & \end{array}$$

The sum of the carries is 0 with a carry, so the sum of the problem is as follows:

1	Carry
11	Carry
1_2	Augend
1_2	1st Addend
1_2	2nd Addend
$+\ 1_2$	3rd Addend
100_2	

The same situation occurs in the following example:

Add 100_2, 101_2, and 111_2

100_2	Augend
101_2	Addend
$+\ 111_2$	Addend

1	Carry
100_2	Augend
101_2	Addend
$+\ 111_2$	Addend
0_2	Sum

11	Carry
100_2	Augend
101_2	Addend
$+\ 111_2$	Addend
00_2	Sum

As in the previous example, the sum of the four 1s is 0 with two carries, and the sum of the two carries is 0 with one carry. The final solution will look like this:

1	Carry
1111	Carry
100_2	Augend
101_2	Addend
$+\ 111_2$	Addend
10000_2	Sum

In the addition of binary numbers, you should remember the following binary addition rules:

$$\text{Rule 1: } 0_2 + 0_2 = 0_2$$

$$\text{Rule 2: } 1_2 + 0_2 = 1_2$$

$$\text{Rule 3: } 0_2 + 1_2 = 1_2$$

$$\text{Rule 4: } 1_2 + 1_2 = 10_2$$

Now practice what you've learned by solving the following problems:

Q9.

Add:

$$\begin{array}{r} 10101_2 \\ +\ 1010_2 \\ \hline \end{array}$$

Q10.

Add:

$$\begin{array}{r} 10011_2 \\ +\ 1010_2 \\ \hline \end{array}$$

Q11.

Add:

$$\begin{array}{r} 11101_2 \\ +\ 100_2 \\ \hline \end{array}$$

Q12.

Add:

$$\begin{array}{r} 10110_2 \\ +\ 11001_2 \\ \hline \end{array}$$

Q13.

Add:

$$\begin{array}{r} 111_2 \\ +\ 1_2 \\ \hline \end{array}$$

Q14.

Add:

$$
\begin{array}{r}
1010010_2 \\
1110111_2 \\
+\quad 10101_2 \\
\hline
\end{array}
$$

Subtraction of Binary Numbers

Now that you are familiar with the addition of binary numbers, subtraction will be easy. The following are the four rules that you must observe when subtracting:

Rule 1: $0_2 - 0_2 = 0_2$

Rule 2: $1_2 - 0_2 = 1_2$

Rule 3: $1_2 - 1_2 = 0_2$

Rule 4: $0_2 - 1_2 = 1_2$ with a borrow

The following example ($10110_2 - 1100_2$) demonstrates the four rules of binary subtraction:

$$
\begin{array}{rl}
10110_2 & \text{Minuend} \\
-\ 1100_2 & \text{Subtrahend} \\
\hline
?010_2 & \text{Difference}
\end{array}
$$

Rule 1: $0_2 - 0_2 = 0_2$

Rule 2: $1_2 - 0_2 = 1_2$

Rule 3: $1_2 - 1_2 = 0_2$

Rule 4: (Explained below)

Rule 4 presents a different situation because you cannot subtract 1 from 0. Since you cannot subtract 1 from 0 and have a positive difference, you must borrow the 1 from the next higher order column of the minuend. The borrow may be indicated as shown below:

$$
\begin{array}{rl}
10 & \text{Borrow} \\
0 & \text{After borrow} \\
\cancel{1}0110_2 & \text{Minuend} \\
-\ 1100_2 & \text{Subtrahend} \\
\hline
1010_2 & \text{Difference}
\end{array}
$$

Rule 4: $(0_2 - 1_2)$

Now observe the following method of borrowing across more than one column in the example, $1000_2 - 1_2$:

$$
\begin{array}{rl}
1\ 1 & \text{Borrow} \\
0\ \cancel{10}\cancel{10}\cancel{10} & \text{After borrow (base 2)} \\
\cancel{1}000_2 & \text{Minuend} \\
-\ \qquad 1_2 & \text{Subtrahend} \\
\hline
0111_2 & \text{Difference}
\end{array}
$$

Let’s practice some subtraction by solving the following problems:

Q15. *Subtract:*

$$
\begin{array}{r}
11001_2 \\
-\ \ 1001_2 \\
\hline
\end{array}
$$

Q16. *Subtract:*

$$
\begin{array}{r}
10101_2 \\
-\ \ 1010_2 \\
\hline
\end{array}
$$

Q17. *Subtract:*

$$
\begin{array}{r}
11111_2 \\
-\ \ \ 10_2 \\
\hline
\end{array}
$$

Q18. Subtract:

$$\begin{array}{r} 111_2 \\ -\ \ 100_2 \\ \hline \end{array}$$

Q19. Subtract:

$$\begin{array}{r} 10001_2 \\ -\ \ 11_2 \\ \hline \end{array}$$

Q20. Subtract:

$$\begin{array}{r} 100000_2 \\ -\ \ 1_2 \\ \hline \end{array}$$

Complementary Subtraction

If you do any work with computers, you will soon find out that most digital systems cannot subtract—they can only add. You are going to need a method of adding that gives the results of subtraction. Does that sound confusing? Really, it is quite simple. A COMPLEMENT is used for our subtractions. A complement is something used to complete something else.

In most number systems you will find two types of complements. The first is the amount necessary to complete a number up to the highest number in the number system. In the decimal system, this would be the difference between a given number and all 9s. This is called the nines complement or the radix-1 or R's-1 complement. As an example, the nines complement of 254 is 999 minus 254, or 745.

The second type of complement is the difference between a number and the next higher power of the number base. As an example, the next higher power of 10 above 999 is 1,000. The difference between 1,000 and 254 is 746. This is called the tens complement in the decimal number system. It is also called the radix or R's complement. We will use complements to subtract. Let's look at the *magic* of this process. There are three important points we should mention before we start: (1) Never complement the minuend in a problem, (2) always disregard any carry beyond the number of positions of the largest of the original numbers, and (3) add the R's complement of the original subtrahend to the original minuend. This will have the same effect as subtracting the original number. Let's look at a base ten example in which we subtract 38 from 59:

R's
Complement

$$\begin{array}{r} 59 \\ -\ 38 \\ \hline 21 \end{array} \rightarrow \boxed{\begin{array}{r} 100 \\ -\ 38 \\ \hline 62 \end{array}} \rightarrow \begin{array}{r} 59 \\ +62 \\ \hline 121 \end{array}$$

Add the R's complement of 38

Disregard any carry

Now let's look at the number system that most computers use, the binary system. Just as the decimal system, had the nines (R's-1) and tens (R's) complement, the binary system has two types of complement methods. These two types are the ones (R's-1) complement and the twos (R's) complement. The binary system R's-1 complement is the difference between the binary number and all 1s. The R's complement is the difference between the binary number and the next higher power of 2.

Let's look at a quick and easy way to form the R's-1 complement. To do this, change each 1 in the original number to 0 and each 0 in the original number to 1 as has been done in the example below.

1011011_2

100100_2 R's-1 complement

There are two methods of achieving the R's complement. In the first method we perform the R's-1 complement and then add 1. This is much easier than subtracting the original number from the next higher power of 2. If you had subtracted, you would have had to borrow.

Saying it another way, to reach the R's complement of any binary number, change all 1s to 0s and all 0s to 1s, and then add 1.

As an example let's determine the R's complement of 10101101_2:

Step 1 -- R's - 1 complement	01010010_2
Step 2 -- Add 1:	$+\ \ 1_2$
	01010011_2

The second method of obtaining the R's complement will be demonstrated on the binary number 0010110100_2.

Step 1—Start with the LSD, working to the MSD, writing the digits as they are up to and including the first one.

first one

0010110100_2

100_2

Step 2—Now R's-1 complement the remaining digits:

$$\underbrace{1101001}_{\text{R's-1 complement of remaining digits}}\overset{\substack{\text{first one}\\ \downarrow}}{1}00_2$$

Now let's R's complement the same number using both methods:

Method 1

$$\begin{array}{rl}
1001100_2 & \\
0110011_2 & \text{R's-1 complement} \\
+ \quad\quad 1_2 & \text{Add 1} \\
\hline
0110100_2 & \text{R's complement answer}
\end{array}$$

Method 2

$$\begin{array}{ll}
1001100_2 & \\
\underbrace{0110}_{\text{R's-1 complement of remaining digits}}\underbrace{100}_{\text{Unchanged digits}}{}_2 & \text{R's-1 complement}
\end{array}$$

Now let's do some subtracting by using the R's complement method. We will go through the subtraction of 3_{10} from 9_{10} ($0011_2$from 1001_2):

$$\begin{array}{r}
9_{10} \\
-\ 3_{10} \\
\hline
\end{array}
\qquad
\begin{array}{rl}
1001_2 & \text{Minuend} \\
-\ 0011_2 & \text{Subtrahend} \\
\hline
\end{array}$$

Step 1—Leave the minuend alone:

1001_2 remains 1001_2

Step 2—Using either method, R's complement the subtrahend:

1101_2 R's complement of subtrahend

Step 3—Add the R's complement found in step 2 to the minuend of the original problem:

$$
\begin{array}{rl}
1001_2 & \text{Original minuend} \\
+\ 1101_2 & \text{R's complement of subtrahend} \\
\hline
10110_2 & \text{Difference of original problem}
\end{array}
$$

Step 4—Remember to discard any carry beyond the size of the original number. Our original problem had four digits, so we discard the carry that expanded the difference to five digits. This carry we disregard is significant to the computer. It indicates that the difference is positive. Because we have a carry, we can read the difference directly without any further computations. Let's check our answer:

$$
\begin{array}{rcr}
1001_2 & = & 9_{10} \\
-\ 0011_2 & = & -3_{10} \\
\hline
1\ 0110_2 & = & 6_{10}
\end{array}
$$

$\uparrow$

Discard

If we do *not* have a carry, it indicates the difference is a negative number. In that case, the difference must be R's complemented to produce the correct answer.

Let's look at an example that will explain this for you.

Subtract 9_{10} from 5_{10} (1001_2 from 0101_2):

$$
\begin{array}{rrl}
5_{10} & 0101_2 & \text{Minuend} \\
\underline{-9}_{10} & \underline{-\ 1001}_2 & \text{Subtrahend} \\
-4_{10} & &
\end{array}
$$

Step 1—Leave the minuend alone:

0101_2 remains 0101_2

Step 2—R's complement the subtrahend:

0111_2 R's complement of subtrahend

Step 3—Add the R's complement found in step 2 to the minuend of the original problem:

$$
\begin{array}{rl}
0101_2 & \text{Original minuend} \\
+\ 0111_2 & \text{Twos complement} \\
\hline
1100_2 & \text{Difference of original problem}
\end{array}
$$

Step 4—We do *not* have a carry; and this tells us, and any computer, that our difference (answer) is negative. With no carry, we must R's complement the difference in step 3. We will then have arrived at the answer (difference) to our original problem. Let's do this R's complement step and then check our answer:

0100_2 R's complement of difference in step 3

Remember, we had no carry in step 3. That showed us our answer was going to be negative. Make sure you indicate the difference is negative. Let's check the answer to our problem:

$$
\begin{array}{r}
0101_2 = 5_{10} \\
\underline{-\ 1001}_2 = \underline{-9}_{10} \\
-\ 0100_2 = -4_{10}
\end{array}
$$

Try solving a few subtraction problems by using the complement method:

Q21. Subtract:

$$
\begin{array}{r}
325_{10} \\
\underline{-\ 104}_{10}
\end{array}
$$

Q22. Subtract:

$$
\begin{array}{r}
10010111_2 \\
\underline{-\ 00110100}_2
\end{array}
$$

Q23. Subtract:

$$
\begin{array}{r}
1011_2 \\
\underline{-\ 1100}_2
\end{array}
$$

OCTAL NUMBER SYSTEM

The octal, or base 8, number system is a common system used with computers. Because of its relationship with the binary system, it is useful in programming some types of computers.

Look closely at the comparison of binary and octal number systems in table 1-3. You can see that one octal digit is the equivalent value of three binary digits. The following examples of the conversion of octal 225_8 to binary and back again further illustrate this comparison:

Octal to binary

2	2	5_8
010	010	101_2

Binary to Octal

010	010	101_2
2	2	5_8

Table 1-3. —Binary and Octal Comparison

	BINARY	OCTAL	
2^0	0	0	8^0
	1	1	
2^1	10	2	
	11	3	
2^2	100	4	
	101	5	
	110	6	
	111	7	
2^3	1000	10	8^1
	1001	11	
	1010	12	
	1011	13	
	1100	14	
	1101	15	
	1110	16	
	1111	17	
2^4	10000	20	
	10001	21	
	10010	22	
	10011	23	
	10100	24	
	10101	25	
	10110	26	
	10111	27	
	11000	30	

NTS1301T3

Unit and Number

The terms that you learned in the decimal and binary sections are also used with the octal system.

The unit remains a single object, and the number is still a symbol used to represent one or more units.

Base (Radix)

As with the other systems, the radix, or base, is the number of symbols used in the system. The octal system uses eight symbols — 0 through 7. The base, or radix, is indicated by the subscript 8.

Positional Notation

The octal number system is a positional notation number system. Just as the decimal system uses powers of 10 and the binary system uses powers of 2, the octal system uses power of 8 to determine the value of a number's position. The following bar graph shows the positions and the power of the base:

$$8^3 \; 8^2 \; 8^1 \; 8^0 \bullet 8^{-1} \; 8^{-2} \; 8^{-3}$$

Remember, that the power, or exponent, indicates the number of times the base is multiplied by itself. The value of this multiplication is expressed in base 10 as shown below:

$$8^3 = 8 \text{ x } 8 \text{ x } 8, \text{ or } 512_{10}$$

$$8^2 = 8 \text{ x } 8, \text{ or } 64_{10}$$

$$8^1 = 8_{10}$$

$$8^0 = 1_{10}$$

$$8^{-1} = \frac{1}{8}, \text{ or } .125_{10}$$

$$8^{-2} = \frac{1}{8 \text{ x } 8}, \text{ or } \frac{1}{64}, \text{ or } .015625_{10}$$

$$8^{-3} = \frac{1}{8 \text{ x } 8 \text{ x } 8}, \text{ or } \frac{1}{512}, \text{ or } .0019531_{10}$$

All numbers to the left of the radix point are whole numbers, and those to the right are fractional numbers.

MSD and LSD

When determining the most and least significant digits in an octal number, use the same rules that you used with the other number systems. The digit farthest to the left of the radix point is the MSD, and the one farthest right of the radix point is the LSD.

Example:

$$\begin{array}{cccccccc} 4 & 7 & 3 & 2 & \bullet & 2 & 6 & 1_8 \\ \uparrow & & & & \uparrow & & & \uparrow \\ \text{MSD} & & & & \text{Radix Point} & & & \text{LSD} \end{array}$$

If the number is a whole number, the MSD is the nonzero digit farthest to the left of the radix point and the LSD is the digit immediately to the left of the radix point. Conversely, if the number is a fraction only, the nonzero digit closest to the radix point is the MSD and the LSD is the nonzero digit farthest to the right of the radix point.

Addition of Octal Numbers

The addition of octal numbers is not difficult provided you remember that anytime the sum of two digits exceeds 7, a carry is produced. Compare the two examples shown below:

$$\begin{array}{r} 4_8 \\ +2_8 \\ \hline 6_8 \end{array} \qquad \begin{array}{r} 4_8 \\ +4_8 \\ \hline 10_8 \end{array}$$

The octal addition table in table 1-4 will be of benefit to you until you are accustomed to adding octal numbers. To use the table, simply follow the directions used in this example:

Add: 6_8 and 5_8

Table 1-4. —Octal Addition Table

+	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7	10
2	2	3	4	5	6	7	10	11
3	3	4	5	6	7	10	11	12
4	4	5	6	7	10	11	12	13
5	5	6	7	10	11	12	13	14
6	6	7	10	11	12	13	14	15
7	7	10	11	12	13	14	15	16

X

Z

Y

Locate the 6 in the X column of the figure. Next locate the 5 in the Y column. The point in area Z where these two columns intersect is the sum. Therefore,

$$\begin{array}{r} 6_8 \\ +5_8 \\ \hline 13_8 \end{array}$$ (spoken, "one three, base eight")

If you use the concepts of addition you have already learned, you are ready to add octal numbers.

Work through the solutions to the following problems:

$$\begin{array}{rl} 11 & \text{Carry} \\ 456_8 & \text{Augend} \\ +\ 123_8 & \text{Addend} \\ \hline 601_8 & \text{Sum} \end{array}$$

$$\begin{array}{rl} 11111 & \text{Carry} \\ 77714_8 & \text{Augend} \\ +\quad 76_8 & \text{Addend} \\ \hline 100012_8 & \text{Sum} \end{array}$$

As was mentioned earlier in this section, each time the sum of a column of numbers exceeds 7, a carry is produced. More than one carry may be produced if there are three or more numbers to be added, as in this example:

$$\begin{array}{rl} 7_8 & \text{Augend} \\ 7_8 & \text{Addend} \\ \underline{+\ \ 7_8} & \text{Addend} \end{array}$$

The sum of the augend and the first addend is 6_8 with a carry. The sum of 6_8 and the second addend is 5_8 with a carry. You should write down the 5_8 and add the two carries and bring them down to the sum, as shown below:

$$\begin{array}{rl} 1 & \text{Carry} \\ 1 & \text{Carry} \\ 7_8 & \text{Augend} \\ \underline{+\ \ 7_8} & \text{First addend} \\ 6_8 & \text{Subsum} \\ \underline{+\ \ 7_8} & \text{Second addend} \\ 25_8 & \text{Sum} \end{array}$$

$$\begin{array}{rl} 1 & \text{Carry} \\ 1 & \text{Carry} \\ 7_8 & \text{Augend} \\ 7_8 & \text{Addend} \\ \underline{+\ \ 7_8} & \text{Addend} \\ 25_8 & \text{Sum} \end{array}$$

Now let's try some practice problems:

Q24. Add:

$$\begin{array}{r} 3_8 \\ \underline{+\ \ 5_8} \end{array}$$

Q25. Add:

$$\begin{array}{r} 22_8 \\ \underline{+\ \ 36_8} \end{array}$$

Q26. Add:

$$\begin{array}{r} 621_8 \\ +\ \underline{174_8} \end{array}$$

Q27. Add:

$$\begin{array}{r} 13255_8 \\ +\ \underline{7031_8} \end{array}$$

Q28. Add

$$\begin{array}{r} 24_8 \\ 42_8 \\ \underline{+\ \ 63_8} \end{array}$$

Q29. Add:

$$\begin{array}{r} 3_8 \\ 5_8 \\ 2_8 \\ 6_8 \\ \underline{+\ \ 4_8} \end{array}$$

Subtraction of Octal Numbers

The subtraction of octal numbers follows the same rules as the subtraction of numbers in any other number system. The only variation is in the quantity of the borrow. In the decimal system, you had to borrow a group of 10_{10}. In the binary system, you borrowed a group of 2_{10}. In the octal system you will borrow a group of 8_{10}.

Consider the subtraction of 1 from 10 in decimal, binary, and octal number systems:

DECIMAL	BINARY	OCTAL
$\begin{array}{r} 10_{10} \\ \underline{-\ \ 1_{10}} \\ 9_{10} \end{array}$	$\begin{array}{r} 10_{2} \\ \underline{-\ \ 1_{2}} \\ 1_{2} \end{array}$	$\begin{array}{r} 10_{8} \\ \underline{-\ \ 1_{8}} \\ 7_{8} \end{array}$

In each example, you cannot subtract 1 from 0 and have a positive difference. You must use a borrow from the next column of numbers. Let's examine the above problems and show the borrow as a *decimal* quantity for clarity:

$$\begin{array}{rcrcrl} {}^{10} & & {}^{2} & & {}^{8} & \text{Borrow} \\ \cancel{1}0_{10} & & \cancel{1}0_{2} & & \cancel{1}0_{8} & \\ -\ 1_{10} & & -\ 1_{2} & & -\ 1_{8} & \\ \hline 9_{10} & & 1_{2} & & 7_{8} & \end{array}$$

When you use the borrow, the column you borrow from is reduced by 1, and the amount of the borrow is added to the column of the minuend being subtracted. The following examples show this procedure:

$$\begin{array}{rl} {}^{10} & \text{Borrow (Base 10)} \\ {}^{2}\ \ & \text{After borrow} \\ \cancel{3}4_{10} & \text{Minuend} \\ -\ 9_{10} & \text{Subtrahend} \\ \hline 25_{10} & \text{Difference} \end{array}$$

$$\begin{array}{rl} {}^{10} & \text{Borrow (Base 8)} \\ {}^{3}\ \ & \text{After borrow} \\ \cancel{4}6_{8} & \text{Minuend} \\ -\ 7_{8} & \text{Subtrahend} \\ \hline 37_{8} & \text{Difference} \end{array}$$

In the octal example 7_8 cannot be subtracted from 6_8, so you must borrow from the 4. Reduce the 4 by 1 and add 10_8 (the borrow) to the 6_8 in the minuend. By subtracting 7_8 from 16_8, you get a difference of 7_8. Write this number in the difference line and bring down the 3. You may need to refer to table 1-4, the octal addition table, until you are familiar with octal numbers. To use the table for subtraction, follow these directions. Locate the subtrahend in column Y. Now find where this line intersects with the minuend in area Z. The remainder, or difference, will be in row X directly above this point.

Do the following problems to practice your octal subtraction:

Q30. Subtract:

$$\begin{array}{r} 765_8 \\ -\ 444_8 \\ \hline \end{array}$$

Q31. Subtract:

$$\begin{array}{r} 44_8 \\ -\ 6_8 \\ \hline \end{array}$$

Q32. Subtract:

$$\begin{array}{r} 532_8 \\ -\ \underline{174}_8 \end{array}$$

Q33. Subtract:

$$\begin{array}{r} 1023_8 \\ -\ \underline{424}_8 \end{array}$$

Q34. Subtract:

$$\begin{array}{r} 423_8 \\ -\ \underline{326}_8 \end{array}$$

Q35. Subtract:

$$\begin{array}{r} 7776_8 \\ -\ \underline{7}_8 \end{array}$$

Check your answers by adding the subtrahend and difference for each problem.

HEXADECIMAL (HEX) NUMBER SYSTEM

The hex number system is a more complex system in use with computers. The name is derived from the fact the system uses 16 symbols. It is beneficial in computer programming because of its relationship to the binary system. Since 16 in the decimal system is the fourth power of 2 (or 2^4); one hex digit has a value equal to four binary digits. Table 1-5 shows the relationship between the two systems.

Table 1-5. —Binary and Hexadecimal Comparison

	BINARY	OCTAL	
2^0	0	0	16^0
	1	1	
2^1	10	2	
	11	3	
2^2	100	4	
	101	5	
	110	6	
	111	7	
2^3	1000	8	
	1001	9	
	1010	A	
	1011	B	
	1100	C	
	1101	D	
	1110	E	
	1111	F	
2^4	10000	10	16^1
	10001	11	
	10010	12	
	10011	13	
	10100	14	
	10101	15	
	10110	16	
	10111	17	
	11000	18	
	11001	19	
	11010	1A	
	11011	1B	
	11100	1C	

NTS1301T5

Unit and Number

As in each of the previous number systems, a unit stands for a single object.

A number in the hex system is the symbol used to represent a unit or quantity. The Arabic numerals 0 through 9 are used along with the first six letters of the alphabet. You have probably used letters in math problems to represent unknown quantities, but in the hex system A, B, C, D, E, and F, each have a definite value as shown below:

$$A_{16} = 10_{10}$$

$$B_{16} = 11_{10}$$

$$C_{16} = 12_{10}$$

$$D_{16} = 13_{10}$$

$$E_{16} = 14_{10}$$

$$F_{16} = 15_{10}$$

Base (Radix)

The base, or radix, of this system is 16, which represents the number of symbols used in the system. A quantity expressed in hex will be annotated by the subscript 16, as shown below:

$$A3EF_{16}$$

Positional Notation

Like the binary, octal, and decimal systems, the hex system is a positional notation system. Powers of 16 are used for the positional values of a number. The following bar graph shows the positions:

$$16^3\ 16^2\ 16^1\ 16^0 \bullet\ 16^{-1}\ 16^{-2}\ 16^{-3}$$

Multiplying the base times itself the number of times indicated by the exponent will show the equivalent decimal value:

$$16^3 = 16 \text{ x } 16 \text{ x } 16, \text{ or } 4096_{10}$$

$$16^3 = 16 \text{ x } 16, \text{ or } 256_{10}$$

$$16^1 = 16_{10}$$

$$16^0 = 1_{10}$$

$$16^{-1} = \frac{1}{16}, \text{ or } .0625_{10}$$

$$16^{-2} = \frac{1}{16 \text{ x } 16}, \text{ or } .0039062_{10}$$

$$16^{-3} = \frac{1}{16 \text{ x } 16 \text{ x } 16}, \text{ or } .0002441_{10}$$

You can see from the positional values that usually fewer symbol positions are required to express a number in hex than in decimal. The following example shows this comparison:

625_{16} is equal to 1573_{10}

MSD and LSD

The most significant and least significant digits will be determined in the same manner as the other number systems. The following examples show the MSD and LSD of whole, fractional, and mixed hex numbers:

```
 7   9   E   4   .16
 ↑           ↑   ↑
MSD         LSD  Radix Point
```

```
   .  1  8  2  A16
   ↑  ↑        ↑
Radix MSD     LSD
Point
```

```
3   B   C . E   4   2   F16
↑         ↑             ↑
MSD       Radix         LSD
          Point
```

Addition of Hex Numbers

The addition of hex numbers may seem intimidating at first glance, but it is no different than addition in any other number system. The same rules apply. Certain combinations of symbols produce a carry while others do not. Some numerals combine to produce a sum represented by a letter. After a little practice you will be as confident adding hex numbers as you are adding decimal numbers.

Study the hex addition table in table 1-6. Using the table, add 7 and 7. Locate the number 7 in both columns X and Y. The point in area Z where these two columns intersect is the sum; in this case 7 + 7 = E. As long as the sum of two numbers is 15_{10} or less, only one symbol is used for the sum. A carry will be produced when the sum of two numbers is 16_{10} or greater, as in the following examples:

$$\begin{array}{r} 8_{16} \\ +\ 8_{16} \\ \hline 10_{16} \end{array} \qquad \begin{array}{r} A_{16} \\ +\ D_{16} \\ \hline 17_{16} \end{array} \qquad \begin{array}{r} D_{16} \\ +\ 9_{16} \\ \hline 16_{16} \end{array}$$

Table 1-6. —Hexadecimal Addition Table

+	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
1	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	10
2	2	3	4	5	6	7	8	9	A	B	C	D	E	F	10	11
3	3	4	5	6	7	8	9	A	B	C	D	E	F	10	11	12
4	4	5	6	7	8	9	A	B	C	D	E	F	10	11	12	13
5	5	6	7	8	9	A	B	C	D	E	F	10	11	12	13	14
6	6	7	8	9	A	B	C	D	E	F	10	11	12	13	14	15
7	7	8	9	A	B	C	D	E	F	10	11	12	13	14	15	16
8	8	9	A	B	C	D	E	F	10	11	12	13	14	15	16	17
9	9	A	B	C	D	E	F	10	11	12	13	14	15	16	17	18
A	A	B	C	D	E	F	10	11	12	13	14	15	16	17	18	19
B	B	C	D	E	F	10	11	12	13	14	15	16	17	18	19	1A
C	C	D	E	F	10	11	12	13	14	15	16	17	18	19	1A	1B
D	D	E	F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C
E	E	F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D
F	F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E

X

Z

Y

Use the addition table and follow the solution of the following problems:

$$\begin{array}{rl} 456_{16} & \text{Augend} \\ +\ \underline{784}_{16} & \text{Addend} \\ \text{BDA}_{16} & \text{Sum} \end{array}$$

In this example each column is straight addition with no carry.

Now add the addend (784_{16}) and the sum (BDA_{16}) of the previous problem:

$$\begin{array}{rl} 1\ \ 1\quad\ \ & \text{Carry} \\ 784_{16} & \text{Augend} \\ +\downarrow \text{BDA}_{16} & \text{Addend} \\ \hline 1\,3\,5\,\text{E}_{16} & \text{Sum} \end{array}$$

Here the sum of 4 and A is E. Adding 8 and D is 15_{16}; write down 5 and carry a 1. Add the first carry to the 7 in the next column and add the sum, 8, to B. The result is 13_{16}; write down 3 and carry a 1. Since only the last carry is left to add, bring it down to complete the problem.

Now observe the procedures for a more complex addition problem. You may find it easier to add the Arabic numerals in each column first:

$$
\begin{array}{rl}
{}^{1\,1\,1} & \text{Carry} \\
C14_{16} & \text{Augend} \\
19E_{16} & \text{Addend} \\
571_{16} & \text{Addend} \\
+\quad BB3_{16} & \text{Addend} \\
\hline
1ED6_{16} & \text{Sum}
\end{array}
$$

The sum of 4, E, 1, and 3 in the first column is 16_{16}. Write down the 6 and the carry. In the second column, 1, 1, 9, and 7 equals 12_{16}. Write the carry over the next column. Add B and 2 — the sum is D. Write this in the sum line. Now add the final column, 1, 1, 5, and C. The sum is 13_{16}. Write down the carry; then add 3 and B — the sum is E. Write down the E and bring down the final carry to complete the problem.

Now solve the following addition problems:

Q36. Add:

$$
\begin{array}{r}
4A3C_{16} \\
+\quad 93\,5\,1_{16} \\
\hline
\end{array}
$$

Q37. Add:

$$
\begin{array}{r}
4321_{16} \\
+\quad DCBA_{16} \\
\hline
\end{array}
$$

Q38. Add:

$$
\begin{array}{r}
274_{16} \\
+\quad FEB_{16} \\
\hline
\end{array}
$$

Q39. Add:

$$
\begin{array}{r}
79DF_{16} \\
+\quad A641_{16} \\
\hline
\end{array}
$$

Q40. Add:

$$
\begin{array}{r}
ECFD_{16} \\
+\quad A4AE_{16} \\
\hline
\end{array}
$$

Q41. Add:

$$\begin{array}{r} BC_{16} \\ A23_{16} \\ +\ \ FC9_{16} \\ \hline \end{array}$$

Subtraction of Hex Numbers

The subtraction of hex numbers looks more difficult than it really is. In the preceding sections you learned all the rules for subtraction. Now you need only to apply those rules to a new number system. The symbols may be different and the amount of the borrow is different, but the rules remain the same.

Use the hex addition table (table 1-6) to follow the solution of the following problems:

$$\begin{array}{rl} ABC_{16} & \text{Minuend} \\ -\ \ 642_{16} & \text{Subtrahend} \\ \hline \end{array}$$

Working from left to right, first locate the subtrahend (2) in column Y. Follow this line across area Z until you reach C. The difference is located in column X directly above the C — in this case A. Use this same procedure to reach the solution:

$$\begin{array}{rl} ABC_{16} & \text{Minuend} \\ -\ \ 642_{16} & \text{Subtrahend} \\ \hline 47A_{16} & \text{Difference} \end{array}$$

Now examine the following solutions:

$$\begin{array}{rl} 7E5E_{16} & \text{Minuend} \\ -\ \ 471_{16} & \text{Subtrahend} \\ \hline 3744_{16} & \text{Difference} \end{array}$$

$$\begin{array}{rl} 1E9C4_{16} & \text{Minuend} \\ -\ \ F4A1_{16} & \text{Subtrahend} \\ \hline F523_{16} & \text{Difference} \end{array}$$

In the previous example, when F was subtracted from 1E, a borrow was used. Since you cannot subtract F from E and have a positive difference, a borrow of 10_{16} was taken from the next higher value column. The borrow was added to E, and the higher value column was reduced by 1.

The following example shows the use of the borrow in a more difficult problem:

$$
\begin{array}{rccccl}
 & & & & 10_{16} & \text{Borrow} \\
 & & & 2 & & \text{Minuend reduced by 1} \\
 & 4 & A & \cancel{3} & 7_{16} & \text{Minuend} \\
- & 2 & C & 4 & B_{16} & \text{Subtrahend} \\
\hline
 & & & & C_{16} & \text{Difference}
\end{array}
$$

In this first step, B cannot be subtracted from 7, so you take a borrow of 10_{16} from the next higher value column. Add the borrow to the 7 in the minuend; then subtract (17_{16} minus B_{16} equals C_{16}). Reduce the number from which the borrow was taken (3) by 1.

To subtract 4_{16} from 2_{16} also requires a borrow, as shown below:

$$
\begin{array}{rccccl}
 & & & 10_{16} & 10_{16} & \text{Borrow} \\
 & & 9 & 2 & & \text{Minuend reduced by 1} \\
 & 4 & \cancel{A} & \cancel{3} & 7_{16} & \text{Minuend} \\
- & 2 & C & 4 & B_{16} & \text{Subtrahend} \\
\hline
 & & & E & C_{16} & \text{Difference}
\end{array}
$$

Borrow 10_{16} from the A and reduce the minuend by 1. Add the borrow to the 2 and subtract 4_{16} from 12_{16}. The difference is E.

When solved the problem looks like this:

$$
\begin{array}{rccccl}
 & & 10 & 10 & 10 & \text{Borrow (Base 16)} \\
 & 3 & 9 & 2 & & \text{Minuend reduced by 1} \\
 & \cancel{4} & \cancel{A} & \cancel{3} & 7_{16} & \text{Minuend} \\
- & 2 & C & 4 & B_{16} & \text{Subtrahend} \\
\hline
 & 1 & D & E & C_{16} & \text{Difference}
\end{array}
$$

Remember that the borrow is 10_{16} not 10_{10}.

There may be times when you need to borrow from a column that has a 0 in the minuend. In that case, you borrow from the next highest value column, which will provide you with a value in the 0 column that you can borrow from.

$$
\begin{array}{rcccl}
 & & F & & \text{Borrow reduced by 1} \\
 & & \cancel{10} & 10 & \text{Borrow (Base 16)} \\
 & 1 & & & \text{Minuend reduced by 1} \\
 & \cancel{2} & 0 & 7_{16} & \text{Minuend} \\
- & & & A_{16} & \text{Subtrahend} \\
\hline
 & 1 & F & D_{16} & \text{Difference}
\end{array}
$$

To subtract A from 7, you must borrow. To borrow you must first borrow from the 2. The 0 becomes 10_{16}, which can give up a borrow. Reduce the 10_{16} by 1 to provide a borrow for the 7. Reducing 10_{16} by 1 equals F. Subtracting A_{16} from 17_{16} gives you D_{16}. Bring down the 1 and F for a difference of $1FD_{16}$.

Now let's practice what we've learned by solving the following hex subtraction problems:

Q42. *Subtract:*

$$\begin{array}{r} 758_{16} \\ \underline{-\ 423}_{16} \end{array}$$

Q43. *Subtract:*

$$\begin{array}{r} D9F_{16} \\ \underline{-\ 46A}_{16} \end{array}$$

Q44. *Subtract:*

$$\begin{array}{r} A1C6_{16} \\ \underline{+\ \ C95_{16}} \end{array}$$

Q45. *Subtract:*

$$\begin{array}{r} 4057_{16} \\ \underline{-\ 9A4}_{16} \end{array}$$

Q46. *Subtract:*

$$\begin{array}{r} 13579_{16} \\ \underline{-\ 2ABD}_{16} \end{array}$$

Q47. *Subtract:*

$$\begin{array}{r} EFACD_{16} \\ \underline{-\ ACBBE}_{16} \end{array}$$

CONVERSION OF BASES

We mentioned in the introduction to this chapter that digital computers operate on electrical pulses. These pulses or the absence of, are easily represented by binary numbers. A pulse can represent a binary 1, and the lack of a pulse can represent a binary 0 or vice versa.

The sections of this chapter that discussed octal and hex numbers both mentioned that their number systems were beneficial to programmers. You will see later in this section that octal and hex numbers are easily converted to binary numbers and vice versa..

If you are going to work with computers, there will be many times when it will be necessary to convert decimal numbers to binary, octal, and hex numbers. You will also have to be able to convert binary, octal, and hex numbers to decimal numbers. Converting each number system to each of the others will be explained. This will prepare you for converting from any base to any other base when needed.

DECIMAL CONVERSION

Some computer systems have the capability to convert decimal numbers to binary numbers. They do this by using additional circuitry. Many of these systems require that the decimal numbers be converted to another form before entry.

Decimal to Binary

Conversion of a decimal number to any other base is accomplished by dividing the decimal number by the radix of the system you are converting to. The following definitions identify the basic terms used in division:

- DIVIDEND—The number to be divided
- DIVISOR—The number by which a dividend is divided
- QUOTIENT—The number resulting from the division of one number by another
- REMAINDER—The final undivided part after division that is less or of a lower degree than the divisor

To convert a base 10 whole number to its binary equivalent, first set up the problem for division:

$$2\overline{)5_{10}}$$

Step 1—Divide the base 10 number by the radix (2) of the binary system and extract the remainder (this becomes the binary number's LSD).

$$\begin{array}{rrl} & 2 & \text{Quotient} \\ \text{Divisor} & 2\overline{)5_{10}} & \\ & \underline{4} & \\ & 1 & \text{Remainder} \longrightarrow 1 \end{array}$$

Step 2—Continue the division by dividing the quotient of step 1 by the radix (2 x 2).

$$\begin{array}{rl} 1 & \text{(Quotient from step 1)} \\ 2\overline{)2} & \\ \underline{2} & \\ 0 & \text{Remainder} \longrightarrow 0 \end{array}$$

Step 3—Continue dividing quotients by the radix until the quotient becomes smaller that the divisor; then do one more division. The remainder is our MSD.

$$\begin{array}{r} 0 \\ 2\overline{)1} \\ \underline{0} \\ 1 \end{array} \quad \begin{array}{l} \text{(Quotient from step 2)} \\ \\ \\ \text{Remainder} \longrightarrow 1 \end{array}$$

The remainder in step 1 is our LSD. Now rewrite the solution, and you will see that 5_{10} equals 101_2. Now follow the conversion of 23_{10} to binary:

Step 1—Set up the problem for division:

$$2\overline{)23_{10}}$$

Step 2—Divide the number and extract the remainder:

$$\begin{array}{r} 11 \\ 2\overline{)23} \\ \underline{2} \\ 03 \\ \underline{2} \\ 1 \end{array} \quad \text{Remainder} \longrightarrow 1 \text{ (LSD)}$$

$$\begin{array}{r} 5 \\ 2\overline{)11} \\ \underline{10} \\ 1 \end{array} \quad \begin{array}{l} \text{(Quotient from previous step)} \\ \\ \text{Remainder} \longrightarrow 1 \end{array}$$

$$\begin{array}{r} 2 \\ 2\overline{)5} \\ \underline{4} \\ 1 \end{array} \quad \begin{array}{l} \text{(Quotient from previous step)} \\ \\ \text{Remainder} \longrightarrow 1 \end{array}$$

$$\begin{array}{r} 1 \\ 2\overline{)2} \\ \underline{2} \\ 0 \end{array} \quad \begin{array}{l} \text{(Quotient from previous step)} \\ \\ \text{Remainder} \longrightarrow 0 \end{array}$$

$$\begin{array}{r} 0 \\ 2\overline{)1} \\ \underline{0} \\ 1 \end{array} \quad \begin{array}{l} \text{(Quotient from previous step)} \\ \\ \text{Remainder} \longrightarrow 1 \text{ (MSD)} \end{array}$$

Step 3—Rewrite the solution from MSD to LSD:

$$10111_2$$

No matter how large the decimal number may be, we use the same procedure. Let's try the problem below. It has a larger dividend:

$$\begin{array}{r} 52 \\ 2\overline{)105} \\ \underline{10} \\ 05 \\ \underline{4} \\ 1 \end{array} \longrightarrow 1 \text{ (LSD)}$$

$$\begin{array}{r} 26 \\ 2\overline{)52} \\ \underline{4} \\ 12 \\ \underline{12} \\ 0 \end{array} \longrightarrow 0$$

$$\begin{array}{r} 13 \\ 2\overline{)26} \\ \underline{2} \\ 06 \\ \underline{6} \\ 0 \end{array} \longrightarrow 0$$

$$\begin{array}{r} 6 \\ 2\overline{)13} \\ \underline{12} \\ 1 \end{array} \longrightarrow 1$$

$$\begin{array}{r} 3 \\ 2\overline{)6} \\ \underline{6} \\ 0 \end{array} \longrightarrow 0$$

$$\begin{array}{r} 1 \\ 2\overline{)3} \\ \underline{2} \\ 1 \end{array} \longrightarrow 1$$

$$\begin{array}{r} 0 \\ 2\overline{)1} \\ \underline{0} \\ 1 \end{array} \longrightarrow 1 \text{ (MSD)}$$

105_{10} equals 1101001_2

We can convert fractional decimal numbers by multiplying the fraction by the radix and extracting the portion of the product to the *left* of the radix point. Continue to multiply the fractional portion of the previous product until the desired degree of accuracy is attained.

Let's go through this process and convert 0.25_{10} to its binary equivalent:

$$
\begin{array}{rcccr}
 & & & & .25_{10} \\
 & & & & \underline{\times \quad 2} \\
\text{MSD} & \longleftarrow & 0 & \longleftarrow & 0.50 \\
 & & & & \underline{\times \quad 2} \\
\text{LSD} & \longleftarrow & 1 & \longleftarrow & 1.00
\end{array}
$$

The *first* figure to the left of the radix point is the MSD, and the last figure of the computation is the LSD. Rewrite the solution from MSD to LSD preceded by the radix point as shown:

$$.01_2$$

Now try converting $.625_{10}$ to binary:

$$
\begin{array}{rcccr}
 & & & & .625 \\
 & & & & \underline{\times \quad 2} \\
\text{MSD} & \longleftarrow & 1 & \longleftarrow & 1.250 \\
 & & & & \underline{\times \quad 2} \\
 & & 0 & \longleftarrow & 0.500 \\
 & & & & \underline{\times \quad 2} \\
\text{LSD} & \longleftarrow & 1 & \longleftarrow & 1.000 \\
 & & & & \underline{\times \quad 2} \\
 & & & & 0.000
\end{array}
$$

$$.625_{10} \text{ equals } .101_2$$

As we mentioned before, you should continue the operations until you reach the desired accuracy. For example, convert $.425_{10}$ to five places in the binary system:

$$
\begin{array}{rcccr}
 & & & & .425 \\
 & & & & \underline{\times \quad 2} \\
\text{MSD} & \longleftarrow & 0 & \longleftarrow & 0.850 \\
 & & & & \underline{\times \quad 2} \\
 & & 1 & \longleftarrow & 1.700 \\
 & & & & \underline{\times \quad 2} \\
 & & 1 & \longleftarrow & 1.400 \\
 & & & & \underline{\times \quad 2} \\
 & & 0 & \longleftarrow & 0.800 \\
 & & & & \underline{\times \quad 2} \\
 & & 1 & \longleftarrow & 1.600 \\
 & & & & \underline{\times \quad 2} \\
 & & 1 & \longleftarrow & 1.200 \\
 & & & & \underline{\times \quad 2} \\
\text{LSD} & \longleftarrow & 0 & \longleftarrow & 0.400
\end{array}
$$

Although the multiplication was carried out for seven places, you would only use what is required. Write out the solution as shown:

$$.01101_2$$

To convert a mixed number such as 37.625_{10} to binary, split the number into its whole and fractional components and solve each one separately. In this problem carry the fractional part to four places. When the conversion of each is completed, recombine it with the radix point as shown below:

$$37_{10} = 100101_2$$

$$.625_{10} = .1010_2$$

$$37.625_{10} = 100101.1010_2$$

Convert the following decimal numbers to binary:

Q48. 72_{10}.

Q49. 97_{10}.

Q50. 243_{10}.

Q51. 0.875_{10} *(four places).*

Q52. 0.33_{10} *(four places).*

Q53. 17.42_{10} *(five places)*

Decimal to Octal

The conversion of a decimal number to its base 8 equivalent is done by the repeated division method. You simply divide the base 10 number by 8 and extract the remainders. The first remainder will be the LSD, and the last remainder will be the MSD.

Look at the following example. To convert 15_{10} to octal, set up the problem for division:

$$8\overline{)15_{10}}$$

Since 8 goes into 15 one time with a 7 remainder, 7 then is the LSD. Next divide 8 into the quotient (1). The result is a 0 quotient with a 1 remainder. The 1 is the MSD:

$$\begin{array}{l} \quad 1 \\ 8\overline{)15_{10}} \\ \quad \underline{8} \\ \quad 7 \longrightarrow 7\ \text{(LSD)} \\ \quad 0 \\ 8\overline{)1} \\ \quad \underline{0} \\ \quad 1 \longrightarrow 1\ \text{(MSD)} \end{array}$$

Now write out the number from MSD to LSD as shown:

$$17_8$$

The same process is used regardless of the size of the decimal number. Naturally, more divisions are needed for larger numbers, as in the following example:

Convert 264_{10} to octal:

$$\begin{array}{l} \quad\ \ 33 \\ 8\overline{)264_{10}} \\ \quad \underline{24} \\ \quad\ \ 24 \\ \quad\ \ \underline{24} \\ \quad\quad 0 \longrightarrow 0 \text{ (LSD)} \end{array}$$

$$\begin{array}{l} \quad\ \ 4 \\ 8\overline{)33} \\ \quad \underline{32} \\ \quad\ \ 1 \longrightarrow 1 \end{array}$$

$$\begin{array}{l} \quad 0 \\ 8\overline{)4} \\ \quad \underline{0} \\ \quad 4 \longrightarrow 4 \text{ (MSD)} \end{array}$$

By rewriting the solution, you find that the octal equivalent of 264_{10} is as follows:

$$410_8$$

To convert a decimal fraction to octal, *multiply* the fraction by 8. Extract everything that appears to the left of the radix point. The first number extracted will be the MSD and will follow the radix point. The last number extracted will be the LSD.

Convert 0.05_{10} to octal:

$$\begin{array}{rcrcr} & & & & .05 \\ & & & & \underline{\times \quad 8} \\ \text{MSD} & \longleftarrow & 0 & \longleftarrow & 0.40 \\ & & & & \underline{\times \quad 8} \\ & & 3 & \longleftarrow & 3.20 \\ & & & & \underline{\times \quad 8} \\ & & 1 & \longleftarrow & 1.60 \\ & & & & \underline{\times \quad 8} \\ & & 4 & \longleftarrow & 4.80 \\ & & & & \underline{\times \quad 8} \\ \text{LSD} & \longleftarrow & 6 & \longleftarrow & 6.40 \end{array}$$

Write the solution from MSD to LSD:

$$.03146_8$$

You can carry the conversion out to as many places as needed, but usually four or five places are enough.

To convert a mixed decimal number to its octal equivalent, split the number into whole and fractional portions and solve as shown below:

Convert 105.589_{10} to octal:

$$\begin{array}{r} 13 \\ 8\overline{)105} \\ \underline{8} \\ 25 \\ \underline{24} \\ 1 \end{array} \longrightarrow 1 \text{ (LSD)}$$

$$\begin{array}{r} 1 \\ 8\overline{)13} \\ \underline{8} \\ 5 \end{array} \longrightarrow 5$$

$$\begin{array}{r} 0 \\ 8\overline{)1} \\ \underline{0} \\ 1 \end{array} \longrightarrow 1 \text{ (MSD)}$$

$$\begin{array}{rcrcr} & & & & 0.589 \\ & & & & \underline{\times \quad 8} \\ \text{MSD} & \longleftarrow & 4 & \longleftarrow & 4.712 \\ & & & & \underline{\times \quad 8} \\ & & 5 & \longleftarrow & 5.696 \\ & & & & \underline{\times \quad 8} \\ & & 5 & \longleftarrow & 5.568 \\ & & & & \underline{\times \quad 8} \\ \text{LSD} & \longleftarrow & 4 & \longleftarrow & 4.544 \end{array}$$

Combine the portions into a mixed number:

$$151.4554_8$$

Convert the following decimal numbers to octal:

Q54. 7_{10}

Q55. 43_{10}

Q56. 499_{10}

Q57. 0.951_{10} *(four places).*

Q58. 0.004_{10} *(five places).*

Q59. 252.17_{10} *(three places).*

Decimal to Hex

To convert a decimal number to base 16, follow the repeated division procedures you used to convert to binary and octal, only divide by 16. Let's look at an example:

Convert 63_{10} to hex:

$$\begin{array}{r l}
 & 3 \\
16 & \overline{)63_{10}} \\
 & \underline{48} \\
 & 15_{10} \longrightarrow F_{16} \longrightarrow \text{LSD}
\end{array}$$

$$\begin{array}{r l}
 & 0 \\
16 & \overline{)3} \\
 & \underline{0} \\
 & 3 \longrightarrow 3_{16} \longrightarrow \text{MSD}
\end{array}$$

Therefore, the hex equivalent of 63_{10} is $3F_{16}$.

You have to remember that the remainder is in base 10 and must be converted to hex if it exceeds 9. Let's work through another example:

Convert 174_{10} to hex:

$$\begin{array}{r l}
 & 10 \\
16 & \overline{)174} \\
 & \underline{16} \\
 & 14 \\
 & \underline{0} \\
 & 14_{10} \longrightarrow E_{16} \longrightarrow \text{LSD}
\end{array}$$

$$\begin{array}{r l}
 & 0 \\
16 & \overline{)10} \\
 & \underline{0} \\
 & 10_{10} \longrightarrow A_{16} \longrightarrow \text{MSD}
\end{array}$$

Write the solution from MSD to LSD:

$$AE_{16}$$

There will probably be very few times when you will have to convert a decimal fraction to a hex fraction. If the occasion should arise, the conversion is done in the same manner as binary or octal. Use the following example as a pattern:

Convert 0.695_{10} to hex:

$$
\begin{array}{rcccr}
 & & & & .695 \\
 & & & & \underline{\times \quad 16} \\
 & & & & 4.170 \\
 & & & & \underline{6.950} \\
\text{MSD} & \longleftarrow & B_{16} & \longleftarrow & 11.120 \\
 & & & & \underline{\times \quad 16} \\
 & & & & .720 \\
 & & & & \underline{1.200} \\
 & & 1_{16} & \longleftarrow & 1.920 \\
 & & & & \underline{\times \quad 16} \\
 & & & & 5.520 \\
 & & & & \underline{9.200} \\
 & & E_{16} & \longleftarrow & 14.720 \\
 & & & & \underline{\times \quad 16} \\
 & & & & 4.320 \\
 & & & & \underline{7.200} \\
\text{LSD} & \longleftarrow & B_{16} & \longleftarrow & 11.520
\end{array}
$$

The solution: $.B1EB_{16}$

Should you have the need to convert a decimal mixed number to hex, convert the whole number and the fraction separately; then recombine for the solution.

Convert the following decimal numbers to hex:

Q60. 42_{10}.

Q61. 83_{10}.

Q62. 176_{10}.

Q63. 491_{10}.

Q64. 0.721_{10} *(four places).*

The converting of binary, octal, and hex numbers to their decimal equivalents is covered as a group later in this section.

BINARY CONVERSION

Earlier in this chapter, we mentioned that the octal and hex number systems are useful to computer programmers. It is much easier to provide data to a computer in one or the other of these systems. Likewise, it is important to be able to convert data from the computer into one or the other number systems for ease of understanding the data.

Binary to Octal

Look at the following numbers:

$$10111001001101_2$$

$$27115_8$$

You can easily see that the octal number is much easier to say. Although the two numbers look completely different, they are equal.

Since 8 is equal to 2^3, then one octal digit can represent three binary digits, as shown below:

$$0_8 = 000_2$$
$$1_8 = 001_2$$
$$2_8 = 010_2$$
$$3_8 = 011_2$$
$$4_8 = 100_2$$
$$5_8 = 101_2$$
$$6_8 = 110_2$$
$$7_8 = 111_2$$

With the use of this principle, the conversion of a binary number is quite simple. As an example, follow the conversion of the binary number at the beginning of this section.

Write out the binary number to be converted. Starting at the radix point and moving left, break the binary number into groups of three as shown. This grouping of binary numbers into groups of three is called binary-coded octal (BCO). Add 0s to the left of any MSD that will fill a group of three:

$$010\ 111\ 001\ 001\ 101 \bullet_2$$

Radix Point

Next, write down the octal equivalent of each group:

010	111	001	001	$101._2$
2	7	1	1	$5._8$

To convert a binary fraction to its octal equivalent, starting at the radix point and moving right, expand each digit into a group of three:

$$\bullet\ 100\ 001\ 110\ 011_2$$

Radix Point

Add 0s to the right of the LSD if necessary to form a group of three. Now write the octal digit for each group of three, as shown below:

.100	001	110	$011._2$
.4	1	6	3_8

To convert a mixed binary number, starting at the radix point, form groups of three both right and left:

101	101	100.	001	$110._2$
5	5	4.	1	6_8

Radix Point

Convert the following binary numbers to octal:

Q65. 10_2.

Q66. 1010_2.

Q67. 101111_2.

Q68. 0.0011_2.

Q69. 0.110011_2.

Q70. 110111.010101_2.

Binary to Hex

The table below shows the relationship between binary and hex numbers. You can see that four binary digits may be represented by one hex digit. This is because 16 is equal to 2^4.

HEX		BINARY
0	=	0000
1	=	0001
2	=	0010
3	=	0011
4	=	0100
5	=	0101
6	=	0110
7	=	0111
8	=	1000
9	=	1001
A	=	1010
B	=	1011
C	=	1100
D	=	1101
E	=	1110
F	=	1111

Using this relationship, you can easily convert binary numbers to hex. Starting at the radix point and moving either right or left, break the number into groups of four. The grouping of binary into four bit groups is called binary-coded hexadecimal (BCH).

Convert 111010011_2 to hex:

0001	1101	$0011._2$
1	D	$3._{16}$

Radix Point

Add 0s to the left of the MSD of the whole portion of the number and to the right of the LSD of the fractional part to form a group of four.

Convert $.111_2$ to hex:

$$\frac{.1110_2}{.E_{16}}$$

In this case, if a 0 had not been added, the conversion would have been $.7_{16}$, which is incorrect.

Convert the following binary numbers to hex:

Q71. 10_2.

Q72. 1011_2.

Q73. 101111_2.

Q74. 0.0011_2.

Q75. 0.110011_2.

Q76. 110111.010101$_2$.

OCTAL CONVERSION

The conversion of one number system to another, as we explained earlier, is done to simplify computer programming or interpreting of data.

Octal to Binary

For some computers to accept octal data, the octal digits must be converted to binary. This process is the reverse of binary to octal conversion.

To convert a given octal number to binary, write out the octal number in the following format. We will convert octal 567_8:

5	6	7_8

Next, below each octal digit write the corresponding three-digit binary-coded octal equivalent:

5	6	7_8
101	110	111_2

Solution: 567_8 equals 101 110 111_2

Remove the conversion from the format:

101110111_2

As you gain experience, it may not be necessary to use the block format.

An octal fraction ($.123_8$) is converted in the same manner, as shown below:

.1	2	3
.001	010	011_2

Solution: $.123_8$ equals $.001010011_2$

Apply these principles to convert mixed numbers as well.

Convert 32.25_8 to binary:

3	2.	2	5_8
011	010.	010	101_2

Solution: 32.25_8 equals 011010.010101_2

Convert the following numbers to binary:

Q77. 73_8

Q78. 512_8

Q79. 403_8

Q80. 0.456_8

Q81. 0.73_8

Q82. 36.5_8

Octal to Hex

You will probably not run into many occasions that call for the conversion of octal numbers to hex. Should the need arise, conversion is a two-step procedure. Convert the octal number to binary; then convert the binary number to hex. The steps to convert 53.7_8 to hex are shown below:

5	3 .	7_8
101	011.	111_2

Regroup the binary digits into groups of four and add zeros where needed to complete groups; then convert the binary to hex.

0010	1011.	1110_2
2	B .	E_{16}

Solution: 53.7_8 equals $2B.E_{16}$

Convert the following numbers to hex:

Q83. 74_8

Q84. 512_8

Q85. 0.03_8

Q86. 14.42_8

HEX CONVERSION

The procedures for converting hex numbers to binary and octal are the reverse of the binary and octal conversions to hex.

Hex to Binary

To convert a hex number to binary, set up the number in the block format you used in earlier conversions. Below each hex digit, write the four-digit binary equivalent. Observe the following example:

Convert ABC_{16} to binary:

A	B	C_{16}
1010	1011	1100_2

Solution: $ABC_{16} = 101010111100_2$

Hex to Octal

Just like the conversion of octal to hex, conversion of hex to octal is a two-step procedure. First, convert the hex number to binary; and second, convert the binary number to octal. Let's use the same example we used above in the hex to binary conversion and convert it to octal:

A	B	C_{16}
1010	1011	1100_2

101	010	111	100_2
5	2	7	4_8

Convert these base 16 numbers to their equivalent base 2 and base 8 numbers:

Q87. 23_{16}

Q88. $1B_{16}$

Q89. $0.E4_{16}$

Q90. $45.A_{16}$

CONVERSION TO DECIMAL

Computer data will have little meaning to you if you are not familiar with the various number systems. It is often necessary to convert those binary, octal, or hex numbers to decimal numbers. The need for understanding is better illustrated by showing you a paycheck printed in binary. A check in the amount of $\$10,010,101.00_2$ looks impressive but in reality only amounts to $\$149.00_{10}$

Binary to Decimal

The computer that calculates your pay probably operates with binary numbers, so a conversion takes place in the computer before the amount is printed on your check. Some computers, however, don't automatically convert from binary to decimal. There may be times when you must convert mathematically.

To convert a base 2 number to base 10, you must know the decimal equivalent of each power of 2. The decimal value of a power of 2 is obtained by multiplying 2 by itself the number of times indicated by the exponent for whole numbers; for example, $2^4 = 2 \times 2 \times 2 \times 2$ or 16_{10}.

For fractional numbers, the decimal value is equal to 1 divided by 2 multiplied by itself the number of times indicated by the exponent. Look at this example:

$$2^{-3} = \frac{1}{2 \times 2 \times 2} \text{ or } .125_{10}$$

The table below shows a portion of the positions and decimal values of the binary system:

Radix Point

2^5	2^4	2^3	2^2	2^1	2^0	•	2^{-1}	2^{-2}	2^{-3}
32	16	8	4	2	1	•	.5	.25	.125

Remember, earlier in this chapter you learned that any number to the 0 power is equal to 1_{10}.

Another method of determining the decimal value of a position is to multiply the preceding value by 2 for whole numbers and to divide the preceding value by 2 for fractional numbers, as shown below:

Radix Point

2^5	2^4	2^3	2^2	2^1	2^0	•	2^{-1}	2^{-2}	2^{-3}	2^{-4}
32	16	8	4	2	1	•	.5	.25	.125	.0625

MULTIPLY BY 2 DIVIDE BY 2

Let's convert a binary number to decimal by using the positional notation method. First, write out the number to be converted; then, write in the decimal equivalent for each position with a 1 indicated. Add these values to determine the decimal equivalent of the binary number. Look at our example:

Problem: $\underline{1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1_2}$
Decimal: 32 8 1
Value:

$$\begin{array}{r} 1 \\ 8 \\ +\,32 \\ \hline 41_{10} \end{array}$$

You may want to write the decimal equivalent for each position as we did in the following example. Add only the values indicated by a 1.

Problem: $\underline{1 \quad 0 \quad 1 \quad 1 \quad 0. \quad 0 \quad 1_2}$
Decimal: 16 8 4 2 1. .5 .25
Value:

$$\begin{array}{r} .25 \\ 2. \\ 4. \\ +\,16. \\ \hline 22.25_{10} \end{array}$$

You should make sure that the decimal values for each position are properly aligned before adding. For practice let's convert these binary numbers to decimal:

Q91. 10010_2

Q92. 1111100_2

Q93. 1010101_2

Q94. 0.0101_2

Q95. 0.1010_2

Q96. 1101101.1111_2

Octal to Decimal

Conversion of octal numbers to decimal is best done by the positional notation method. This process is the one we used to convert binary numbers to decimal.

First, determine the decimal equivalent for each position by multiplying 8 by itself the number of times indicated by the exponent. Set up a bar graph of the positions and values as shown below:

Positional Notation:	8^4	8^3	8^2	8^1	$8^0_.$	8^{-1}	8^{-2}	8^{-3}
Decimal Equivalent:	4096	512	64	8	1.	.125	.015625	.0019531

To convert an octal number to decimal, write out the number to be converted, placing each digit under the proper position.

Example:

4096	512	64	8	1
		7	4	3_8

Next, multiply the decimal equivalent by the corresponding digit of the octal number; then, add this column of figures for the final solution:

4096	512	64	8	1
		×7	×4	×3

$$\begin{array}{r} 3 \\ 32 \\ +\ 448 \\ \hline 483_{10} \end{array}$$

Solution: 743_8 is equal to 483_{10}

Now follow the conversion of 26525_8 to decimal:

4096	512	64	8	1
×2	×6	×5	×2	×5

$$\begin{array}{r} 5 \\ 16 \\ 320 \\ 3072 \\ 8192 \\ \hline 11605_{10} \end{array}$$

Solution: $11{,}605_{10}$ is the decimal equivalent of $26{,}525_8$

To convert a fraction or a mixed number, simply use the same procedure.

Example: Change $.5_8$ to decimal:

$$\begin{array}{r} .125 \\ \hline \times\ 5_8 \end{array} \longrightarrow .625_{10}$$

Example: Convert 24.36_8 to decimal:

$$\begin{array}{ccccc} 64 & 8 & 1. & .125 & .015625 \\ \hline & \times 2 & \times 4. & \times 3 & \times 6 \end{array}$$

$$\begin{array}{r} .09375 \\ .37500 \\ 4.00000 \\ +16.00000 \\ \hline 20.46875_{10} \end{array}$$

Solution: 24.36_8 equals 20.46875_{10}

If your prefer or find it easier, you may want to convert the octal number to binary and then to decimal.

Convert the following numbers to decimal:

Q97. 17_8

Q98. 64_8

Q99. 375_8

Q100. 0.4_8

Q101. 0.61_8

Q102. 10.22_8

Hex to Decimal

It is difficult to comprehend the magnitude of a base 16 number until it is presented in base 10; for instance, $E0_{16}$ is equal to 224_{10}. You must remember that usually fewer digits are necessary to represent a decimal value in base 16.

When you convert from base 16 to decimal, you may use the positional notation system for the powers of 16 (a bar graph). You can also convert the base 16 number to binary and then convert to base 10.

Note in the bar graph below that each power of 16 results in a tremendous increase in the decimal equivalent. Only one negative power (16^{-1}) is shown for demonstration purposes:

Radix Point

Positional Notation:	16^4	16^3	16^2	16^1	$16^0.$	16^{-1}
Decimal Equivalent:	65.536	4.096	256	16	1.	.0625

Just as you did with octal conversion, write out the hex number, placing each digit under the appropriate decimal value for that position. Multiply the decimal value by the base 16 digit and add the values. (Convert A through F to their decimal equivalent before multiplying). Let's take a look at an example.

Convert $2C_{16}$ to decimal:

4096 256 16 1.

×2 ×C

(12)

12
+32
44_{10}

The decimal equivalent of $2C_{16}$ is 44_{10}.

Use the same procedure we used with binary and octal to convert base 16 fractions to decimal.

If you choose to convert the hex number to binary and then to decimal, the solution will look like this:

2 C_{16}

× 0010 × 1100_2

4
8
+32
44_{10}

Convert these base 16 numbers to base 10:

Q103. 24_{16}

Q104. $A5_{16}$

Q105. DB_{16}

Q106. $3E6.5_{16}$

BINARY-CODED DECIMAL

In today's technology, you hear a great deal about microprocessors. A microprocessor is an integrated circuit designed for two purposes: data processing and control.

Computers and microprocessors both operate on a series of electrical pulses called words. A word can be represented by a binary number such as 10110011_2. The word length is described by the number of digits or BITS in the series. A series of four digits would be called a 4-bit word and so forth. The most common are 4-, 8-, and 16-bit words. Quite often, these words must use binary-coded decimal inputs.

Binary-coded decimal, or BCD, is a method of using binary digits to represent the decimal digits 0 through 9. A decimal digit is represented by four binary digits, as shown below:

BCD		Decimal
0000	=	0
0001	=	1
0010	=	2
0011	=	3
0100	=	4
0101	=	5
0110	=	6
0111	=	7
1000	=	8
1001	=	9

You should note in the table above that the BCD coding is the binary equivalent of the decimal digit.

Since many devices use BCD, knowing how to handle this system is important. You must realize that BCD and binary are not the same. For example, 49_{10} in binary is 110001_2, but 49_{10} in BCD is 01001001_{BCD}. Each decimal digit is converted to its binary equivalent.

BCD Conversion

You can see by the above table, conversion of decimal to BCD or BCD to decimal is similar to the conversion of hexadecimal to binary and vice versa.

For example, let's go through the conversion of 264_{10} to BCD. We'll use the block format that you used in earlier conversions. First, write out the decimal number to be converted; then, below each digit write the BCD equivalent of that digit:

2	6	4_{10}
0010	0110	0100_{BCD}

The BCD equivalent of 264_{10} is 001001100100_{BCD}. To convert from BCD to decimal, simply reverse the process as shown:

1001	1000	0011_{BCD}
9	8	3_{10}

BCD Addition

The procedures followed in adding BCD are the same as those used in binary. There is, however, the possibility that addition of BCD values will result in invalid totals. The following example shows this:

Add 9 and 6 in BCD:

$$\begin{array}{lrcr} & 1001_{BCD} & = & 9_{10} \\ & \underline{+\ 0110_{BCD}} & = & 6_{10} \\ \text{INVALID BCD} \longrightarrow & 1111 & & 15_{10} \end{array}$$

The sum 1111_2 is the binary equivalent of 15_{10}; however, 1111 is not a valid BCD number. You cannot exceed 1001 in BCD, so a correction factor must be made. To do this, you add 6_{10} (0110_{BCD}) to the sum of the two numbers. The "add 6" correction factor is added to any BCD group larger than 1001_2. Remember, there is no 1010_2, 1011_2, 1100_2, 1101_2, 1110_2, or 1111_2 in BCD:

$$\begin{array}{rl} 1111 & \longleftarrow \text{INVALID BCD} \\ \underline{0110_{BCD}} & \quad \text{Add } 6_{10} \\ 0001\ 0101 & \longleftarrow \text{New BCD} \end{array}$$

The sum plus the add 6 correction factor can then be converted back to decimal to check the answer. Put any carries that were developed in the add 6 process into a new 4-bit word:

0001	0101_{BCD}
1	5_{10}

Now observe the addition of 60_{10} and 55_{10} in BCD:

$$
\begin{array}{rcl}
60_{10} & = & 0110\ 0000_{BCD} \\
55_{10} & = & \underline{0101\ 0101}_{BCD} \\
 & & 1011\ 0101 \longleftarrow \text{INVALID BCD}
\end{array}
$$

In this case, the higher order group is invalid, but the lower order group is valid. Therefore, the correction factor is added only to the higher order group as shown:

$$
\begin{array}{r}
1011\ 0101 \\
\underline{+\ 0110\ 0000} \\
0001\ 0001\ 0101_{BCD}
\end{array}
\qquad \text{Add } 6_{10}
$$

Convert this total to decimal to check your answer:

$$
\begin{array}{ccc}
0001 & 0001 & 0101_{BCD} \\
\hline
1 & 1 & 5_{10}
\end{array}
$$

Remember that the correction factor is added only to groups that exceed 9_{10} (1001_{BCD}).

Convert the following numbers to BCD and add:

Q107.

$$
\begin{array}{r}
3_{10} \\
\underline{+\ 5_{10}}
\end{array}
$$

Q108.

$$
\begin{array}{r}
1_{10} \\
\underline{+\ 8_{10}}
\end{array}
$$

Q109.

$$
\begin{array}{r}
7_{10} \\
\underline{+\ 4_{10}}
\end{array}
$$

Q110.

$$\begin{array}{r} 14_{10} \\ +\ 8_{10} \\ \hline \end{array}$$

SUMMARY

Now that you've completed this chapter, you should have a basic understanding of number systems. The number systems that were dealt with are used extensively in the microprocessor and computer fields. The following is a summary of the emphasized terms and points found in the "Number Systems" chapter.

The **UNIT** represents a single object.

A **NUMBER** is a symbol used to represent one or more units.

The **RADIX** is the base of a positional number system. It is equal to the number of symbols used in that number system.

A **POSITIONAL NOTATION** is a system in which the value or magnitude of a number is defined not only by its digits or symbol value, but also by its position. Each position represents a power of the radix, or base, and is ranked in ascending or descending order.

$$\begin{array}{ccccccc} 2^2 & 2^1 & 2^0 & & 2^{-1} & 2^{-2} & 2^{-3} \\ 1 & 0 & 1 & \bullet & 0 & 1 & 1_2 \\ & & & \uparrow & & & \\ & & & \text{Radix Point} & & & \end{array}$$

The **MOST SIGNIFICANT DIGIT (MSD)** is a digit within a number (whole or fractional) that has the largest effect (weighing power) on that number.

$$\begin{array}{cccccc} 4 & 7 & 9 & 3 & .0 & {}_{10} \\ \uparrow & & & \uparrow & & \\ \text{MSD} & & & \text{LSD} & & \end{array}$$

$$\begin{array}{cccccc} 0 & .1 & 0 & 6 & 6 & {}_{8} \\ & \uparrow & & & \uparrow & \\ & \text{MSD} & & & \text{LSD} & \end{array}$$

The **LEAST SIGNIFICANT DIGIT (LSD)** is a digit within a number (whole or fractional) that has the least effect (weighting power) on that number.

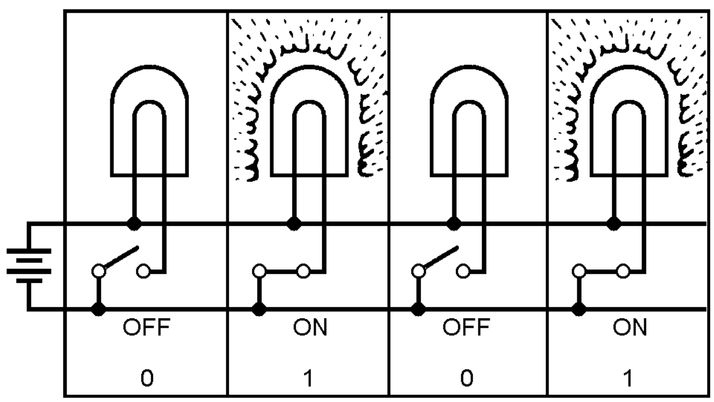
$4\ 3\ A\ F_{16}$

MSD LSD

$.0\ 1\ 0\ 0\ 1\ 1_2$

MSD LSD

The **BINARY NUMBER SYSTEM** is a base 2 system. The symbols 1 and 0 can be used to represent the state of electrical/electronic devices. A binary 1 may indicate the device is active; a 0 may indicate the device is inactive.



The **OCTAL NUMBER SYSTEM** is a base 8 system and is quite useful as a tool in the conversion of binary numbers. This system works because 8 is an integral power of 2; that is, $2^3 = 8$. The use of octal numbers reduces the number of digits required to represent the binary equivalent of a decimal number.

The **HEX NUMBER SYSTEM** is a base 16 system and is sometimes used in computer systems. A binary number can be converted directly to a base 16 number if the binary number is first broken into groups of four digits.

The basic rules of **ADDITION** apply to each of the number systems. Each system becomes unique when carries are produced.

SUBTRACTION in each system is based on certain rules of that number system. The borrow varies in magnitude according to the number system in use. In most computers, subtraction is accomplished by using the complement (R's or R's-1) of the subtrahend and adding it to the minuend.

To **CONVERT A WHOLE BASE 10 NUMBER** to another system, divide the decimal number by the base of the number system to which you are converting. Continue dividing the quotient of the previous division until it can no longer be done. Extract the remainders — the remainder from the first computation will yield the LSD; the last will provide the MSD.

$$\begin{array}{r} 31 \\ 8\overline{)248} \\ \underline{24} \\ 08 \\ \underline{8} \\ 0 \end{array} \longrightarrow 0 \text{ (LSD)}$$

$$\begin{array}{r} 3 \\ 8\overline{)31} \\ \underline{24} \\ 7 \end{array} \longrightarrow 7$$

$$\begin{array}{r} 0 \\ 8\overline{)3} \\ \underline{0} \\ 3 \end{array} \longrightarrow 3 \text{ (MSD)}$$

To **CONVERT DECIMAL FRACTIONS**, multiply the fraction by the base of the desired number system. Extract those digits that move to the left of the radix point. Continue to multiply the fractional product for as many places as needed. The first digit left of the radix point will be the MSD, and the last will be the LSD. The example to the right shows the process of converting 248.32_{10} to the octal equivalent (370.243_8).

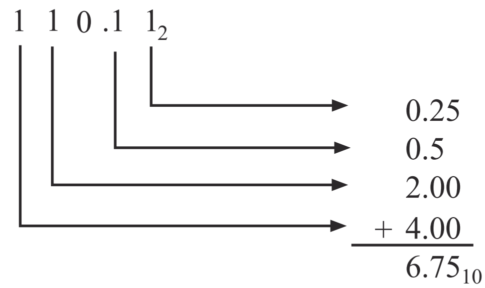
$$\begin{array}{rcr} & & .32 \\ & & \underline{\times \quad 8} \\ \text{(MSD)}\ 2 & \longleftarrow & 2.56 \\ & & \underline{\times \quad 8} \\ 4 & \longleftarrow & 4.48 \\ & & \underline{\times \quad 8} \\ \text{(LSD)}\ 3 & \longleftarrow & 3.84 \end{array}$$

BINARY numbers are converted to **OCTAL** and **HEX** by the grouping method. Three binary digits equal one octal digit; four binary digits equal one hex digit.

$$\underbrace{1\ 0\ 1}_{5} \qquad \underbrace{0\ 1\ 1_2}_{3_8}$$

$$\overbrace{1\ 0\ 1\ 0}^{A} \qquad \overbrace{0\ 1\ 0\ 0_2}^{4_{16}}$$

To **CONVERT** binary, octal, and hex numbers to DECIMAL use the POWERS of the base being converted.



BINARY-CODED DECIMAL (BCD) is a coding system used with some microprocessors. A correction factor is needed to correct invalid numbers

ANSWERS TO QUESTIONS Q1. THROUGH Q110.

A1. *Unit*

A2. *Number*

A3. *Arabic*

A4. *The number of symbols used in the system*

A5. 173_{10}

A6. 10^3, 10^2, 10^1, 10^0,

A7. *Radix point*

A8.

(a) MSD – 4, LSD – 0

(b) MSD – 1, LSD – 6

(c) MSD – 2, LSD – 4

(d) MSD – 2, LSD – 1

A9. 11111_2

A10. 11101_2

A11. 100001_2

A12. 101111_2

A13. 1000_2

A14. 11011110_2

A15. 10000_2

A16. 1011_2

A17. 11101_2

A18. 11_2

A19. 1110_2

A20. 11111_2

A21. 221_{10}

A22. 01100011_2

A23. -0001_2

A24. 10_8

A25. 60_8

A26. 1015_8

A27. 22306_8

A28. 151_8

A29. 24_8

A30. 321_8

A31. 36_8

A32. 336_8

A33. 377_8

A34. 104_8

A35. 7767_8

A36. $DD8D_{16}$

A37. $11FDB_{16}$

A38. $125F_{16}$

A39. 12020_{16}

A40. $191AB_{16}$

A41. $1AA8_{16}$

A42. 335_{16}

A43. 935_{16}

A44. 9531_{16}

A45. $36B3_{16}$

A46. $10ABC_{16}$

A47. $42F0F_{16}$

A48. 1001000_2

A49. 1100001_2

A50. 11110011_2

A51. 0.1110_2

A52. 0.0101_2

A53. 10001.01101_2

A54. 7_8

A55. 53_8

A56. 763_8

A57. 0.7467_8

A58. 0.00203_8

A59. 374.127_8

A60. $2A_{16}$

A61. 53_{16}

A62. $B0_{16}$

A63. $1EB_{16}$

A64. $0.B893_{16}$

A65. 2_8

A66. 12_8

A67. 57_8

A68. 0.14_8

A69. 0.63_8

A70. 67.25_8

A71. 2_{16}

A72. B_{16}

A73. $2F_{16}$

A74. 0.3_{16}

A75. $0.CC_{16}$

A76. 37.54_{16}

A77. 111011_2

A78. 101001010_2

A79. 100000011_2

A80. 0.100101110_2

A81. 0.111011_2

A82. 11110.101_2

A83. $3C_{16}$

A84. $14A_{16}$

A85. $0.0C_{16}$

A86. $C.88_{16}$

A87. 100011_2; 43_8

A88. 11011_2; 33_8

A89. 0.111001_2; 0.71_8

A90. 1000101.101_2; 105.5_8

A91. 18_{10}

A92. 124_{10}

A93. 85_{10}

A94. 0.3125_{10}

A95. 0.625_{10}

A96. 109.9375_{10}

A97. 15_{10}

A98. 52_{10}

A99. 253_{10}

A100. 0.5_{10}

A101. 0.765625_{10}

A102. 8.28125_{10}

A103. 36_{10}

A104. 165_{10}

A105. 219_{10}

A106. 998.3125_{10}

A107. 1000_{BCD}

A108. 1001_{BCD}

A109. $0001\ 0001_{BCD}$

A110. $0010\ 0010_{BCD}$

CHAPTER 2

FUNDAMENTAL LOGIC CIRCUITS

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to do the following:

1. Identify general logic conditions, logic states, logic levels, and positive and negative logic as these terms and characteristics apply to the inputs and outputs of fundamental logic circuits.
2. Identify the following logic circuit gates and interpret and solve the associated Truth Tables:
   a. AND
   b. OR
   c. Inverters (NOT circuits)
   d. NAND
   e. NOR
3. Identify variations of the fundamental logic gates and interpret the associated Truth Tables.
4. Determine the output expressions of logic gates in combination.
5. Recognize the laws, theorems, and purposes of Boolean algebra.

INTRODUCTION

In chapter 1 you learned that the two digits of the binary number system can be represented by the state or condition of electrical or electronic devices. A binary 1 can be represented by a switch that is closed, a lamp that is lit, or a transistor that is conducting. Conversely, a binary 0 would be represented by the same devices in the opposite state: the switch open, the lamp off, or the transistor in cut-off.

In this chapter you will study the four basic logic gates that make up the foundation for digital equipment. You will see the types of logic that are used in equipment to accomplish the desired results. This chapter includes an introduction to Boolean algebra, the logic mathematics system used with digital equipment. Certain Boolean expressions are used in explanation of the basic logic gates, and their expressions will be used as each logic gate is introduced.

COMPUTER LOGIC

Logic is defined as the science of reasoning. In other words, it is the development of a reasonable or logical conclusion based on known information.

GENERAL LOGIC

Consider the following example: If it is true that all Navy ships are gray and the USS *Lincoln* is a Navy ship, then you would reach the logical conclusion that the USS *Lincoln* is gray.

To reach a logical conclusion, you must assume the qualifying statement is a condition of truth. For each statement there is also a corresponding false condition. The statement "USS Lincoln is a Navy ship" is true; therefore, the statement "USS *Lincoln* is not a Navy ship" is false. There are no *in-between* conditions.

Computers operate on the principle of logic and use the **TRUE** and **FALSE** logic conditions of a logical statement to make a programmed decision.

The conditions of a statement can be represented by symbols (variables); for instance, the statement "Today is payday" might be represented by the symbol P. If today actually is payday, then P is TRUE. If today is not payday, then P is FALSE. As you can see, a statement has two conditions. In computers, these two conditions are represented by electronic circuits operating in two **LOGIC STATES**. These logic states are 0 (*zero*) and 1 (*one*). Respectively, 0 and 1 represent the FALSE and TRUE conditions of a statement.

When the TRUE and FALSE conditions are converted to electrical signals, they are referred to as **LOGIC LEVELS** called *HIGH* and *LOW*. The 1 state might be represented by the presence of an electrical signal (HIGH), while the 0 state might be represented by the absence of an electrical signal (LOW).

If the statement "Today is payday" is FALSE, then the statement "Today is NOT payday" must be TRUE. This is called the COMPLEMENT of the original statement. In the case of computer math, complement is defined as the opposite or negative form of the original statement or variable. If today were payday, then the statement "Today is not payday" would be FALSE. The complement is shown by placing a bar, or VINCULUM, over the statement symbol (in this case, $\overline{P}$). This variable is spoken as NOT P. Table 2-1 shows this concept and the relationship with logic states and logic levels.

Table 2-1. —Relationship of Digital Logic Concepts and Terms

Example 1: Assume today is payday				
STATEMENT	**SYMBOL**	**CONDITION**	**LOGIC STATE**	**LOGIC LEVEL**
Original: TODAY IS PAYDAY	**P**	**TRUE**	**1**	**HIGH**
Complement: TODAY IS NOT PAYDAY	$\overline{P}$	**FALSE**	**0**	**LOW**
Example 2: Assume today is <u>not</u> payday				
Original: TODAY IS NOT PAYDAY	**P**	**FALSE**	**0**	**LOW**
Complement: TODAY IS NOT PAYDAY	$\overline{P}$	**TRUE**	**1**	**HIGH**

In some cases, more than one variable is used in a single expression. For example, the expression $AB\overline{C}D$ is spoken "A AND B AND NOT C AND D."

POSITIVE AND NEGATIVE LOGIC

To this point, we have been dealing with one type of **LOGIC POLARITY**, positive. Let's further define logic polarity and expand to cover in more detail the differences between positive and negative logic.

Logic polarity is the type of voltage used to represent the logic 1 state of a statement. We have determined that the two logic states can be represented by electrical signals. Any two distinct voltages may be used. For instance, a positive voltage can represent the 1 state, and a negative voltage can represent the 0 state. The opposite is also true.

Logic circuits are generally divided into two broad classes according to their polarity — positive logic and negative logic. The voltage levels used and a statement indicating the use of positive or negative logic will usually be specified on logic diagrams supplied by manufacturers.

In practice, many variations of logic polarity are used; for example, from a high-positive to a low-positive voltage, or from positive to ground; or from a high-negative to a low-negative voltage, or from negative to ground. A brief discussion of the two general classes of logic polarity is presented in the following paragraphs.

Positive Logic

Positive logic is defined as follows: If the signal that activates the circuit (the 1 state) has a voltage level that is more POSITIVE than the 0 state, then the logic polarity is considered to be POSITIVE. Table 2-2 shows the manner in which positive logic may be used.

Table 2-2. —Examples of Positive Logic

EXAMPLE 1	Active signal — TRUE, 1, HIGH = +10 volts Complement — FALSE, 0, LOW = 0 volts
EXAMPLE 2	Active signal — TRUE, 1, HIGH = 0 volts Complement — FALSE, 0, LOW = -10 volts

NTST130202

As you can see, in positive logic the 1 state is at a more positive voltage level than the 0 state.

Negative Logic

As you might suspect, negative logic is the opposite of positive logic and is defined as follows: If the signal that activates the circuit (the 1 state) has a voltage level that is more NEGATIVE than the 0 state, then the logic polarity is considered to be NEGATIVE. Table 2-3 shows the manner in which negative logic may be used.

Table 2-3.—Examples of Negative Logic

EXAMPLE 1	Active signal —TRUE, 1, LOW = +5 volts Complement —FALSE, 0, HIGH = +10 volts
EXAMPLE 2	Active signal —TRUE, 1, LOW = -10 volts Complement — FALSE, 0, HIGH = -5 volts

NTST130203

NOTE: The logic level LOW now represents the 1 state. This is because the 1 state voltage is more negative than the 0 state.

In the examples shown for negative logic, you notice that the voltage for the logic 1 state is more negative with respect to the logic 0 state voltage. This holds true in example 1 where both voltages are positive. In this case, it may be easier for you to think of the TRUE condition as being less positive than the FALSE condition. Either way, the end result is negative logic.

The use of positive or negative logic for digital equipment is a choice to be made by design engineers. The difficulty for the technician in this area is limited to understanding the type of logic being used and keeping it in mind when troubleshooting.

NOTE:

UNLESS OTHERWISE NOTED, THE REMAINDER OF THIS BOOK WILL DEAL ONLY WITH POSITIVE LOGIC.

LOGIC INPUTS AND OUTPUTS

As you study logic circuits, you will see a variety of symbols (variables) used to represent the inputs and outputs. The purpose of these symbols is to let you know what inputs are required for the desired output.

If the symbol A is shown as an input to a logic device, then the logic level that represents A must be **HIGH** to activate the logic device. That is, it must satisfy the input requirements of the logic device before the logic device will issue the TRUE output.

Look at view A of figure 2-1. The symbol X represents the input. As long as the switch is open, the lamp is not lit. The open switch represents the logic 0 state of variable X.

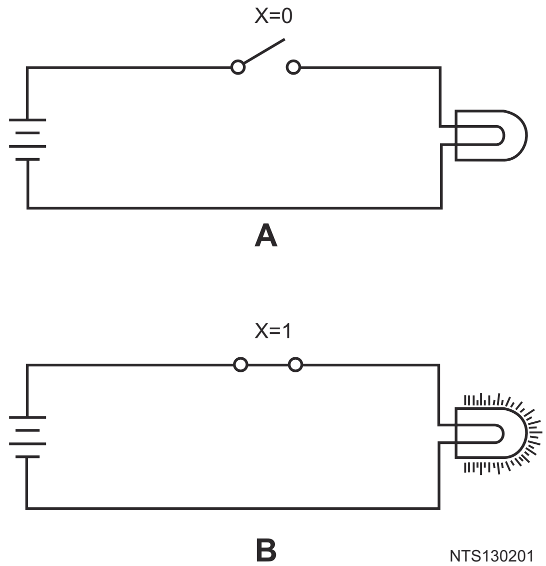


Figure 2-1. —Logic switch: A. Logic 0 state; B. Logic 1 state.

Closing the switch (view B), represents the logic 1 state of X. Closing the switch completes the circuit causing the lamp to light. The 1 state of X satisfied the input requirement and the circuit therefore produced the desired output (logic HIGH); current was applied to the lamp causing it to light.

If you consider the lamp as the output of a logic device, then the same conditions exist. The TRUE (1 state) output of the logic device is to have the lamp lit. If the lamp is not lit, then the output of the logic device is FALSE (0 state).

As you study logic circuits, it is important that you remember the state (1 or 0) of the inputs and outputs.

So far in this chapter, we have discussed the two conditions of logical statements, the logic states representing these two conditions, logic levels and associated electrical signals and positive and negative logic. We are now ready to proceed with individual logic device operations. These make up the majority of computer circuitry.

As each of the logic devices are presented, a chart called a TRUTH TABLE will be used to illustrate all possible input and corresponding output combinations. Truth Tables are particularly helpful in understanding a logic device and for showing the differences between devices.

The logic operations you will study in this chapter are the AND, OR, NOT, NAND, and NOR. The devices that accomplish these operations are called *logic gates*, or more informally, *gates*. These gates are the foundation for all digital equipment. They are the "decision-making" circuits of computers and other types of digital equipment. By making decisions, we mean that certain conditions must exist to produce the desired output.

In studying each gate, we will introduce various mathematical SYMBOLS known as **BOOLEAN ALGEBRA** expressions. These expressions are nothing more than descriptions of the input requirements necessary to activate the circuit and the resultant circuit output.

THE AND GATE

The **AND** gate is a logic circuit that requires all inputs to be TRUE at the same time in order for the output to be TRUE.

LOGIC SYMBOL

The standard symbol for the AND gate is shown in figure 2-2. Variations of this standard symbol may be encountered. These variations become necessary to illustrate that an AND gate may have more than one input.

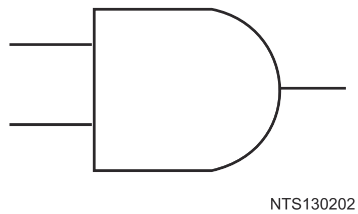


Figure 2-2. —AND gate.

If we apply two variables, A and B, to the inputs of the AND gate, then both A and B would have to be TRUE at the same time to produce the desired TRUE output. . The symbol f designates the output function. The *Boolean expression* for this operation is $f = A \cdot B$ or $f = AB$. The expression is spoken, "f = A AND B." The dot, or lack of, indicates the AND function.

AND GATE OPERATION

We can demonstrate the operation of the AND gate with a simple circuit that has two switches in series as shown in figure 2-3. You can see that both switches would have to be closed at the same time to light the lamp (view A). Any other combination of switch positions (view B) would result in an open circuit and the lamp would not light (logic 0).

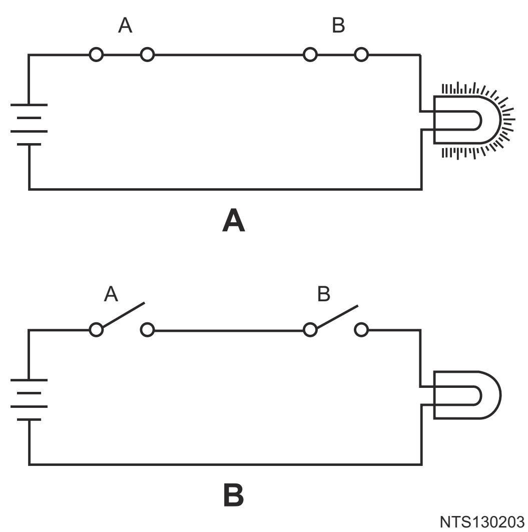


Figure 2-3.—AND gate equivalent circuit: A. Logic 1 state; B. Logic 0 state.

Now look at figure 2-4. Signal A is applied to one input of the AND gate and signal B to the other. At time T_0, both inputs are LOW (logic 0) and f is LOW. At T_1, A goes HIGH (logic 1); B remains LOW; and as a result, f remains LOW. At T_2, A goes LOW and B goes HIGH; f, however, is still LOW, because the proper input conditions have not been satisfied (A and B both HIGH at the same time). At T_4, both A and B are HIGH. As a result, f is HIGH. The input requirements have been satisfied, so the output is HIGH (logic 1).

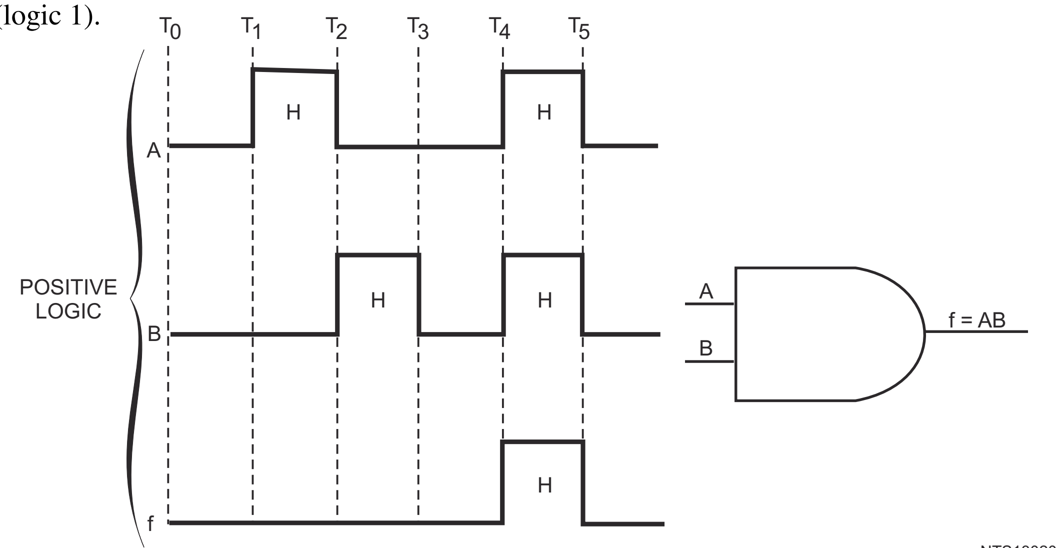
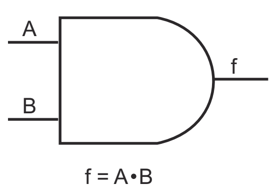


Figure 2-4. —AND gate input and output signals.

TRUTH TABLE

Now let’s refer to figure 2-5. As you can see, a Truth Table and a Table of Combinations are shown. The latter is a deviation of the Truth Table. It uses the HIGH and LOW logic levels to depict the gate’s inputs and resultant output combinations rather than the 1 and 0 logic states. By comparing the inputs and outputs of the two tables, you see how one can easily be converted to the other (remember, 1 = HIGH and 0 = LOW). The Table of Combinations is shown here only to familiarize you with its existence, it will not be seen again in this book. As we mentioned earlier, the Truth Table is a chart that shows all possible combinations of inputs and the resulting outputs. Compare the AND gate Truth Table (figure 2-5) with the input signals shown in figure 2-4.



A	B	f
0	0	0
0	1	0
1	0	0
1	1	1

TRUTH TABLE

A	B	f
L	L	L
L	H	L
H	L	L
H	H	H

TABLE OF COMBINATIONS

NTS130205

Figure 2-5. —AND gate logic symbol, Truth Table, and Table of Combinations.

The first combination (A = 0, B = 0) corresponds to T_0 in figure 2-4; the second to T_1; the third to T_2; and the last to T_4. When constructing a Truth Table, you must include all possible combinations of the inputs, including the all 0s combination.

A Truth Table representing an AND gate with three inputs (X, Y, and Z) is shown below. Remember that the two-input AND gate has four possible combinations, with only one of those combinations providing a HIGH output. An AND gate with three inputs has eight possible combinations, again with only one combination providing a HIGH output. Make sure you include all possible combinations. To check if you have all combinations, raise 2 to the power equal to the number of input variables. This will give you the total number of possible combinations. For example:

EXAMPLE 1-AB = 2^2 = 4 combinations

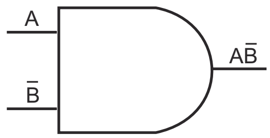
EXAMPLE 2-XYZ = 2^3 = 8 combinations

X	**Y**	**Z**	**f**
0	**0**	**0**	**0**
0	**0**	**1**	**0**
0	**1**	**0**	**0**
0	**1**	**1**	**0**
1	**0**	**0**	**0**
1	**0**	**1**	**0**
1	**1**	**0**	**0**
1	**1**	**1**	**1**
f = XYZ			

As with all AND gates, all the inputs must be HIGH at the same time to produce a HIGH output. Don't be confused if the complement of a variable is used as an input. When a complement is indicated as an input to an AND gate, it must also be HIGH to satisfy the input requirements of the gate. The Boolean expression for the output is formulated based on the TRUE inputs that give a TRUE output. Here is an adage that might help you better understand the AND gate:

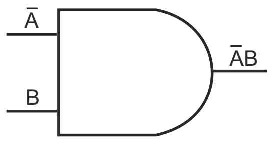
In order to produce a 1 output, all the inputs must be 1. If any or all of the inputs is/are 0, then the output will be 0.

Referring to the following examples should help you cement this concept in your mind. Remember, the inputs, whether the original variable or the complement must be high in order for the output to be high. The three examples given are all AND gates with two inputs. Keep in mind the Boolean expression for the output is the result of all the inputs being HIGH.



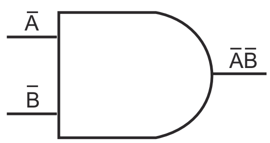
A	$\overline{B}$	f
0	0	0
0	1	0
1	0	0
1	1	1

$f = A\,\overline{B}$



$\overline{A}$	B	f
0	0	0
0	1	0
1	0	0
1	1	1

$f = \overline{A}\,B$



$\overline{A}$	$\overline{B}$	f
0	0	0
0	1	0
1	0	0
1	1	1

$f = \overline{A}\,\overline{B}$

NTS130205a

Figure 2-5a. —AND gate logic with two inputs, Truth Table.

You will soon be able to recognize the Truth Table for the other types of logic gates without having to look at the logic symbol.

Q1. What is defined as "the science of reasoning?"

Q2. With regard to computer logic circuits, what is meant by "complement?"

Q3. What are the complements of the following terms?

a. Q

b. R

c. V

d. Z

Q4. If logic 1 = –5 vdc and logic 0 = –10 vdc, what logic polarity is being used?

Q5. If logic 1 = +2 vdc and logic 0 = –2 vdc, what logic polarity is being used?

Q6. If logic 1 = –5 vdc and logic 0 = 0 vdc, what logic polarity is being used?

Q7. What is the Boolean expression for the output of an AND gate that has R and S as inputs?

Q8. What must be the logic state of R and S to produce the TRUE output?

Q9. How many input combinations exist for a four-input AND gate?

THE OR GATE

The OR gate differs from the AND gate in that only ONE input has to be HIGH to produce a HIGH output. An easy way to remember the OR gate is that any HIGH input will yield a HIGH output.

LOGIC SYMBOL

Figure 2-6shows the standard symbol for the OR gate. The number of inputs will vary according to the needs of the designer.

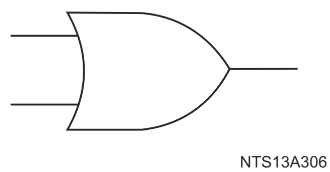


Figure 2-6. —OR gate.

The OR gate may also be represented by a simple circuit as shown in figure 2-7. In the OR gate, two switches are placed in parallel. If either or both of the switches are closed (view A), the lamp will light. The only time the lamp will not be lit is when both switches are open (view B).

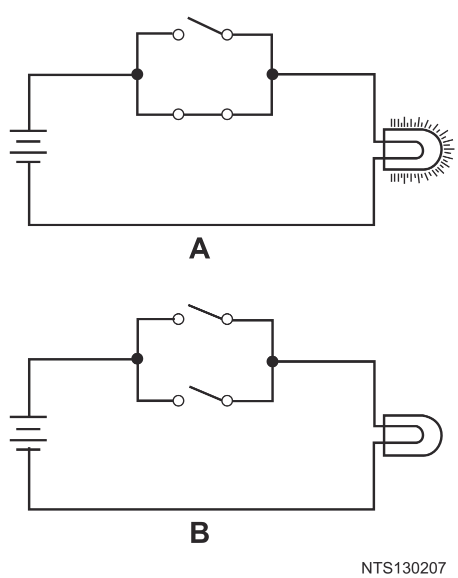


Figure 2-7. —OR gate equivalent circuit: A. Logic 1 state; B. Logic 0 state.

Let's assume we are applying two variables, X and Y, to the inputs of an OR gate. For the circuit to produce a HIGH output, either variable X, variable Y, or both must be HIGH. The Boolean expression for this operation is f = X+Y and is spoken "f equals X OR Y." The plus sign indicates the OR function and should not be confused with addition.

OR GATE OPERATION

Look at figure 2-8. At time T_0, both X and Y are LOW and f is LOW. At T_1, X goes HIGH producing a HIGH output. At T_2 when both inputs go LOW, f goes LOW. When Y goes HIGH at T_3, f also goes HIGH and remains HIGH until both inputs are again LOW. At T_5, both X and Y go HIGH causing f to go HIGH.

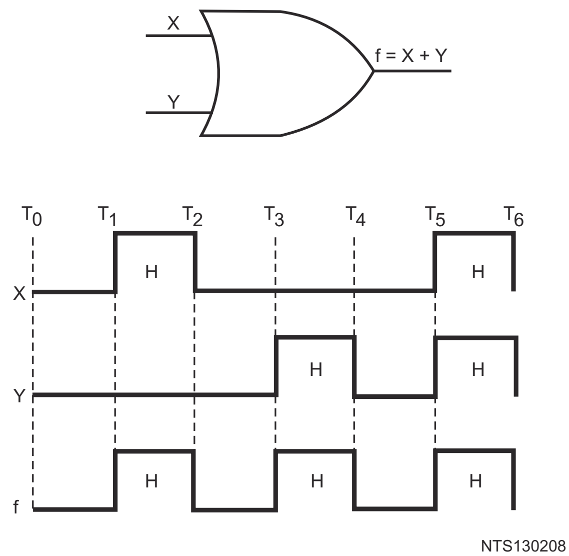


Figure 2-8. —OR gate input and output signals.

TRUTH TABLE

Using the inputs X and Y, let's construct a Truth Table for the OR gate. You can see from the discussion of figure 2-8 that there are four combinations of inputs. List each of these combinations of inputs and the respective outputs and you have the Truth Table for the OR gate.

X	Y	f
0	0	0
0	1	1
1	0	1
1	1	1
f = X + Y		

When writing or stating the Boolean expression for an OR gate with more than two inputs, simply place the OR sign (+) between each input and read or state the sign as OR. For example, the Boolean expression for an OR gate with the inputs of A, B, C, and D would be:

f = A+B+C+D

This expression is spoken "f equals A OR B OR C OR D."

You can substitute the complements for the original statements as we did with the AND gate or use negative logic; but for an output from an OR gate, at least one of the inputs must be TRUE.

Q10. Write the Boolean expression for an OR gate having G, K, and L as inputs.

Q11. How many input combinations are possible using G, K, and L?

Q12. How many of those combinations will produce a HIGH output?

THE INVERTER

The INVERTER, often referred to as a NOT gate, is a logic device that has an output opposite of the input. It is sometimes called a NEGATOR. It may be used alone or in combination with other logic devices to fulfill equipment requirements.

When an inverter is used alone, it is represented by the symbol shown in figure 2-9 (view A). It will more often be seen in conjunction with the symbol for an amplifier (view B). Symbols for inverters used in combination with other devices will be shown later in the chapter.

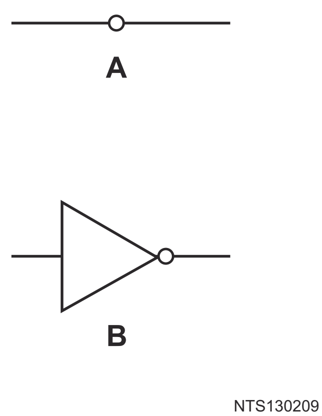


Figure 2-9. —Inverter: A. Symbol for inverter used alone; B. Symbol for an amplifier/inverter.

Let's go back to the statement "Today is payday." We stated that P represents the TRUE state. If we apply P to the input of the inverter as shown in figure 2-10, then the output will be the opposite of the input. The output, in this case, is $\overline{P}$. At times T_0 through T_2, P is LOW. Consequently, the output ($\overline{P}$) is HIGH. At T_2, P goes HIGH and as a result $\overline{P}$ goes LOW. $\overline{P}$ remains LOW as long as P is HIGH and vice versa. The Boolean expression for the output of this gate is $f = \overline{P}$.

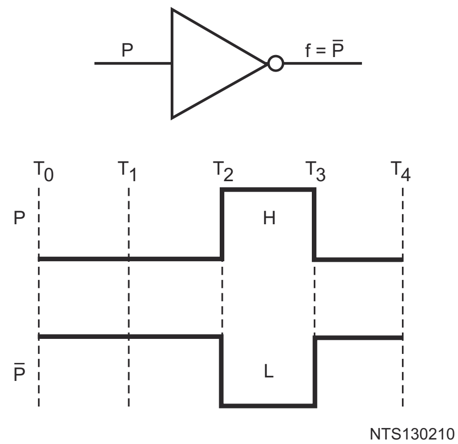


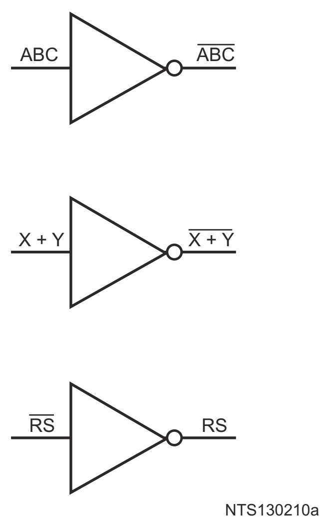
Figure 2-10. —Inverter input and resultant output.

You will recall that $\overline{P}$ is the complement of P.

The Truth Table for an inverter is shown below.

P	**f**
0	**1**
1	**0**

The output of an inverter will be the complement of the input. The following examples show various inputs to inverters and the resulting outputs:



The vinculum, or NOT sign, is placed over the entire output or removed from the output, depending on the input. If we applied $A\overline{B}C$ to an inverter, the output would be $\overline{A\overline{B}C}$. And if we ran that output through another inverter, the output would be $A\overline{B}C$

Q13. What is the complement of XYZ?

Q14. The input to an inverter is $\overline{\overline{X} + (YZ)}$. *What is the output*

Q15. In a properly functioning circuit, can both the input and output of an inverter be HIGH at the same time?

THE NAND GATE

The NAND gate is another logic device commonly found in digital equipment. This gate is simply an AND gate with an inverter (NOT gate) at the output.

LOGIC SYMBOL

The logic symbol for the NAND gate is shown in figure 2-11.

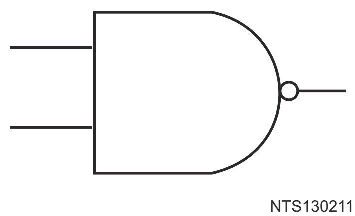


Figure 2-11. —NAND gate.

The NAND gate can have two or more inputs. The output will be LOW only when all the inputs are HIGH. Conversely, the output will be HIGH when any or all of the inputs are LOW.

The NAND gate performs two functions, AND and NOT. Separating the NAND symbol to show these two functions would reveal the equivalent circuits depicted in figure 2-12. This should help you better understand how the NAND gate functions.

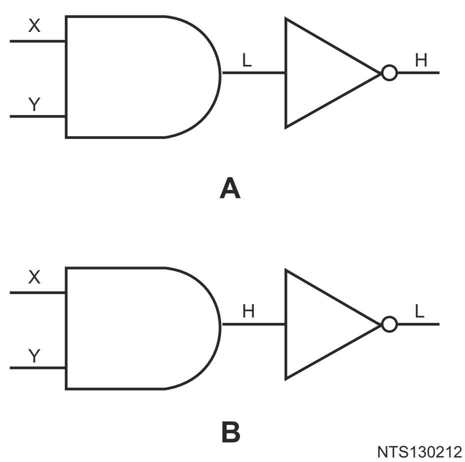


Figure 2-12. —NAND gate equivalent circuit: A. Either X or Y or both are LOW; B. Both X and Y are HIGH.

Inputs X and Y are applied to the AND gate. If either X or Y or both are LOW (view A), then the output of the AND gate is LOW. A LOW (logic 0) on the input of the inverter results in a HIGH (logic 1) output. When both X and Y are HIGH (view B), the output of the AND gate is HIGH; thus the output of the inverter is LOW. The Boolean expression for the output of a NAND gate with these inputs is $f = \overline{XY}$. The expression is spoken "X AND Y quantity NOT." The output of any NAND gate is the negation of the input. For example, if our inputs are X and $\overline{Y}$, the output will be $\overline{X\overline{Y}}$.

NAND GATE OPERATION

Now, let's observe the logic level inputs and corresponding outputs as shown in figure 2-13. At time T_0, X and Y are both LOW. The output is HIGH; the opposite of an AND gate with the same inputs. At T_1, X goes HIGH and Y remains LOW. As a result, the output remains HIGH. At T_2, X goes LOW and Y goes HIGH. Again, the output remains HIGH. When both X and Y are HIGH at T_4, the output goes LOW. The output will remain LOW only as long as both X and Y are HIGH.

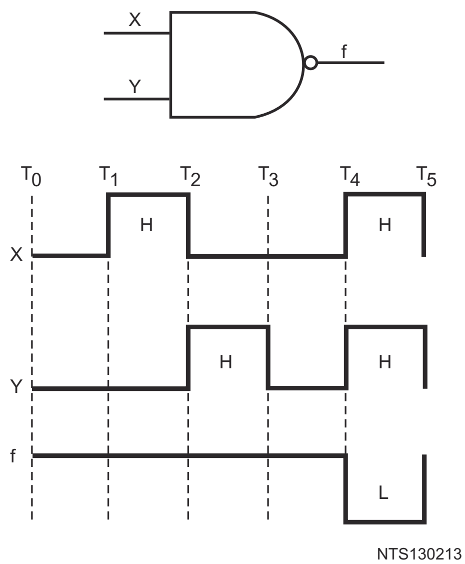


Figure 2-13. —NAND gate input and output signals.

TRUTH TABLE

The Truth Table for a NAND gate with X and Y as inputs is shown below.

X	Y	f
0	0	1
0	1	1
1	0	1
1	1	0
$F = \overline{XY}$		

Q16. A NAND gate has Z and X as inputs. What will be the output logic level if Z is HIGH and X is LOW?

Q17. What must be the state of the inputs to a NAND gate in order to produce a LOW output?

Q18. What is the output Boolean expression for a NAND gate with inputs A, $\overline{B}$*, and C?*

Q19. A NAND gate has inputs labeled as A, $\overline{B}$*, and C. If A and* $\overline{B}$ *are HIGH, C must be at what logic level to produce a HIGH output?*

THE NOR GATE

As you might expect, the NOR gate is an OR gate with an inverter on the output.

LOGIC SYMBOL

The standard logic symbol for this gate is shown in figure 2-14. More than just the two inputs may be shown.

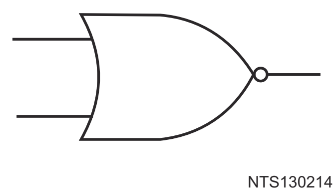


Figure 2-14. —NOR gate.

The NOR gate will have a HIGH output only when all the inputs are LOW.

When broken down, the two functions performed by the NOR gate can be represented by the equivalent circuit depicted in figure 2-15. When both inputs to the OR gate are LOW, the output is LOW. A LOW applied to an inverter gives a HIGH output. If either or both of the inputs to the OR gate are HIGH, the output will be HIGH. When this HIGH output is applied to the inverter, the resulting output is LOW. The Boolean expression for the output of this NOR gate is $f = \overline{K + L}$. The expression is spoken, "K OR L quantity NOT."

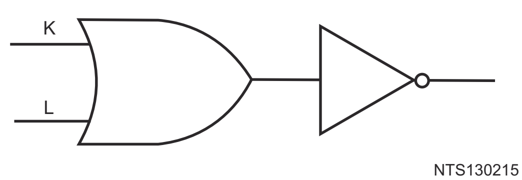


Figure 2-15. —NOR gate equivalent circuit.

NOR GATE OPERATION

The logic level inputs and corresponding outputs for a NOR gate are shown in figure 2-16. At time T_0, both K and L are LOW; as a result, f is HIGH. At T_1, K goes HIGH, L remains LOW, and f goes LOW. At T_2, K goes LOW, L goes HIGH, and the output remains LOW. The output goes HIGH again at T_3 when both inputs are LOW. At T_4 when both inputs are HIGH, the output goes LOW and remains LOW until T_5 when both inputs go LOW. Remember the output is just opposite of what it would be for an OR gate.

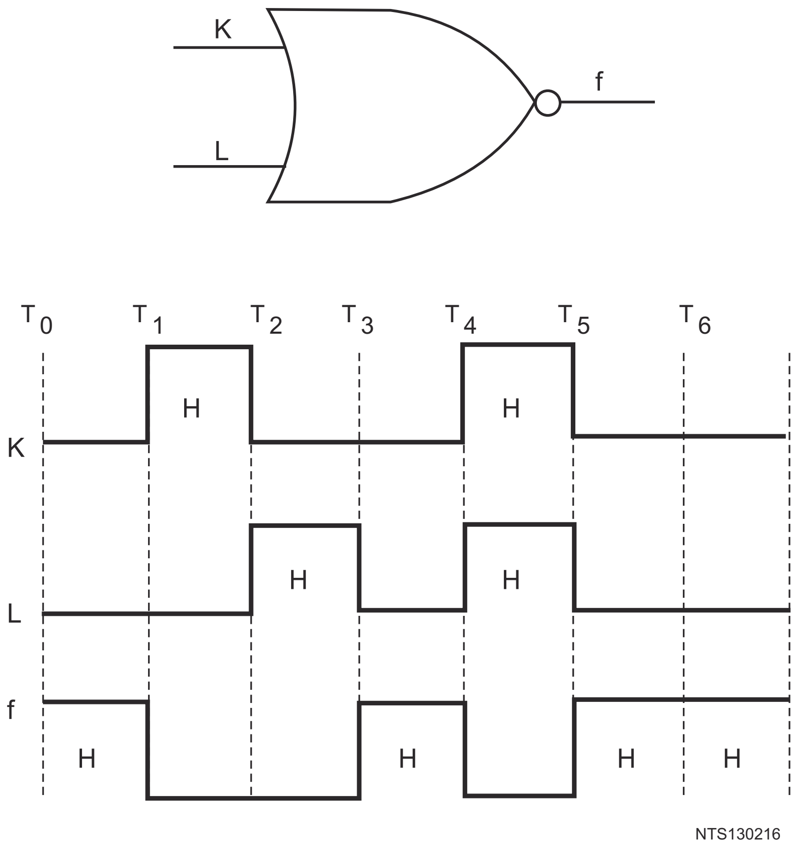


Figure 2-16.—NOR gate input and output signals.

TRUTH TABLE

The Truth Table for a NOR gate with K and L as inputs is shown below.

K	L	f
0	0	1
0	1	0
1	0	0
1	1	0
f = $\overline{K + L}$		

Q20. How does a NOR gate differ from an OR gate?

Q21. What will be the output of a NOR gate when both inputs are HIGH?

Q22. What is the output Boolean expression for a NOR gate with R and T as inputs?

Q23. In what state must the inputs to a NOR gate be in order to produce a logic 1 output?

VARIATIONS OF FUNDAMENTAL GATES

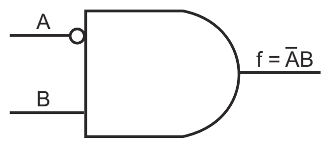
Now that you are familiar with fundamental logic gates, let's look at some variations of these gates that you may encounter.

Up to now you have seen inverters used alone or on the output of AND and OR gates. Inverters may also be used on one or more of the inputs to the logic gates. Take a look at the examples as discussed in the following paragraphs.

AND/NAND GATE VARIATIONS

If we place an inverter on one input of a two-input AND gate, the output will be quite different from that of the standard AND gate.

In figure 2-17, we have placed an inverter on the A input. When A is HIGH, the inverter makes it a LOW going into the AND gate. In order for the output to be HIGH, A would have to be LOW while B is HIGH, as shown in the Truth Table. If the inverter were on the B input, the output expression would then be $f = A\overline{B}$.



A	B	f
0	0	0
0	1	1
1	0	0
1	1	0

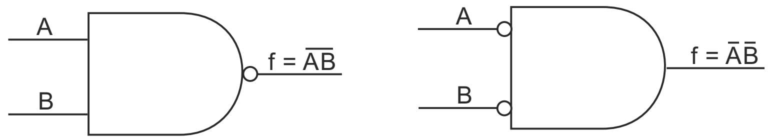
$f = \overline{A}B$

NTS130217

Figure 2-17. —AND gate with one inverted input.

Now let's compare a NAND gate to an AND gate with an inverter on each input. Figure 2-18 shows these gates and the associated Truth Tables. With the NAND gate (view A), the output is HIGH when either or both inputs is/are LOW. The AND gate with inverters on each input (view B), produces a HIGH output only when both inputs are LOW. This comparison also points out the differences between the expressions $f = A\overline{B}$ (A AND B quantity NOT) and $f = \longrightarrow \overline{A}\ \overline{B}$ (NOT A AND NOT B).

Now, look over the Truth Tables for figures 2-17, 2-18, and 2-19; look at how the outputs vary with inverters in different positions.



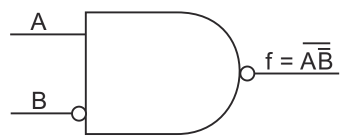
A	B	f
0	0	1
0	1	1
1	0	1
1	1	0

A

A	B	f
0	0	1
0	1	0
1	0	0
1	1	0

B NTS130218

Figure 2-18. —Comparison of NAND gate and AND gate with inverted inputs: A. NAND gate; B. AND gate with inverters on each input.



A	B	f
0	0	1
0	1	1
1	0	0
1	1	0

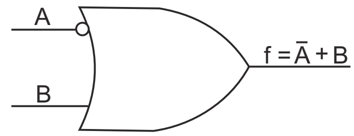
NTS130219

Figure 2-19. —NAND gate with one inverted input.

OR/NOR GATE VARIATIONS

The outputs of OR and NOR gates may also be changed with the use of inverters.

An OR gate with one input inverted is shown in figure 2-20. The output of this OR gate requires that A be LOW, B be HIGH, or both of these conditions existing at the same time in order to have a HIGH output. Since the A input is inverted, it must be LOW if B is LOW in order to produce a HIGH output. Therefore the output is f = $\overline{A}$ +B.

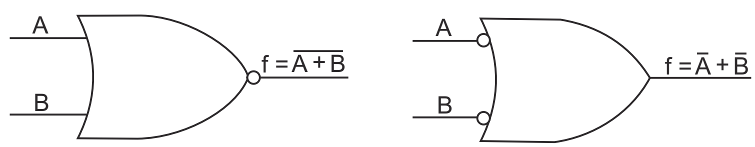


A	B	f
0	0	1
0	1	1
1	0	0
1	1	1

NTS130220

Figure 2-20. —OR gate with one inverted input.

Figure 2-21compares a NOR gate (view A), to an OR gate with inverters on both inputs (view B), and shows the respective Truth Tables. The NOR gate will produce a HIGH output only when both inputs are LOW. The OR gate with inverted inputs produces a HIGH output with all input combinations EXCEPT when both inputs are HIGH. This figure also illustrates the differences between the expressions f = $\overline{A + B}$ (A OR B quantity NOT) and f = $\overline{A} + \overline{B}$ (NOT A OR NOT B).



A	B	f
0	0	1
0	1	0
1	0	0
1	1	0

$f = \overline{A + B}$

A

A	B	f
0	0	1
0	1	1
1	0	1
1	1	0

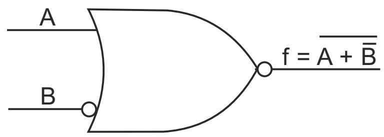
$f = \overline{A} + \overline{B}$

B

NTS130221

Figure 2-21. —Comparison of NOR gate and OR gate with inverted inputs: A. NOR gate; B. OR gate with inverters on both inputs.

As with the NAND gate, one or more inputs to NOR gates may be inverted. Figure 2-22 shows the result of inverting a NOR gate input. In this case, because of the inversion of the B input and the inversion of the output, the only time this gate will produce a HIGH output is when A is LOW and B is HIGH. The output Boolean expression for this gate is f = $\overline{A + \overline{B}}$, spoken “A OR NOT B quantity NOT.”



A	B	f
0	0	0
0	1	1
1	0	0
1	1	0

f = $\overline{A + \overline{B}}$

NTS130222

Figure 2-22. —NOR gate with one inverted input.

Table 2-4 illustrates AND, NOR, NAND, and OR gate combinations that produce the same output. You can see by the table that there is more than one way to achieve a desired output. Although the gates have only two inputs, the table can be extended to more than two inputs.

Table 2-4.—Equivalent AND and NOR, NAND and OR Gates

TRUTH TABLES

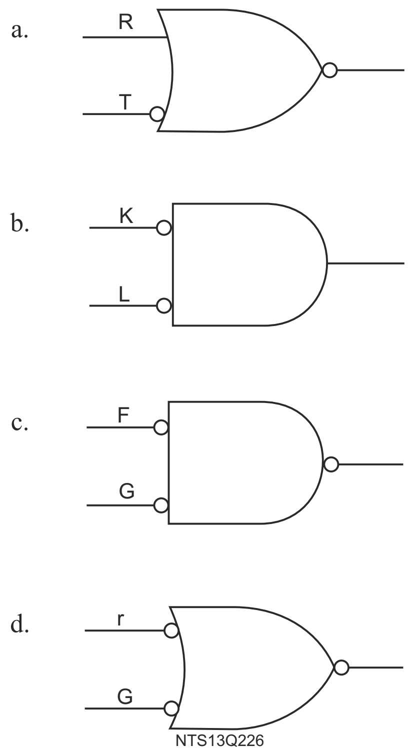
	AND GATES	NOR GATES	A	B	f
1	A, B → AB_ → AB	A, B → $\overline{\bar{A}+\bar{B}}$	0 0 1 1	0 1 0 1	0 0 0 1
2	A, B → $\bar{A}B$	A, B → $\overline{A+\bar{B}}$	0 0 1 1	0 1 0 1	0 1 0 0
3	A, B → $A\bar{B}$	A, B → $\overline{\bar{A}+B}$	0 0 1 1	0 1 0 1	0 0 1 0
4	A, B → $\bar{A}\bar{B}$	A, B → $\overline{A+B}$	0 0 1 1	0 1 0 1	1 0 0 0
	NAND GATES	**OR GATES**	**A**	**B**	**f**
5	A, B → $\overline{\bar{A}\bar{B}}$	A, B → $A+B$	0 0 1 1	0 1 0 1	0 1 1 1
6	A, B → $\overline{A\bar{B}}$	A, B → $\bar{A}+B$	0 0 1 1	0 1 0 1	1 1 0 1
7	A, B → $\overline{\bar{A}B}$	A, B → $A+\bar{B}$	0 0 1 1	0 1 0 1	1 0 1 1
8	A, B → $\overline{AB}$	A, B → $\bar{A}+\bar{B}$	0 0 1 1	0 1 0 1	1 1 1 0

NTS13T0204

Q24. What is the output Boolean expression for an AND gate with A and B as inputs when the B input is inverted?

Q25. What is the equivalent logic gate of a two-input NAND gate with both inputs inverted?

Q26. What is the output Boolean expression for the following gates?



LOGIC GATES IN COMBINATION

When you look at logic circuit diagrams for digital equipment, you are not going to see just a single gate, but many combinations of gates. At first it may seem confusing and complex. If you interpret one gate at a time, you can work your way through any network. In this section, we will analyze several combinations of gates and then provide you with some practice problems.

Figure 2-23 (view A) shows a simple combination of AND gates. The outputs of gates 1 and 2 are the inputs to gate 3. You already know that both inputs to an AND gate must be HIGH at the same time in order to produce a HIGH output.

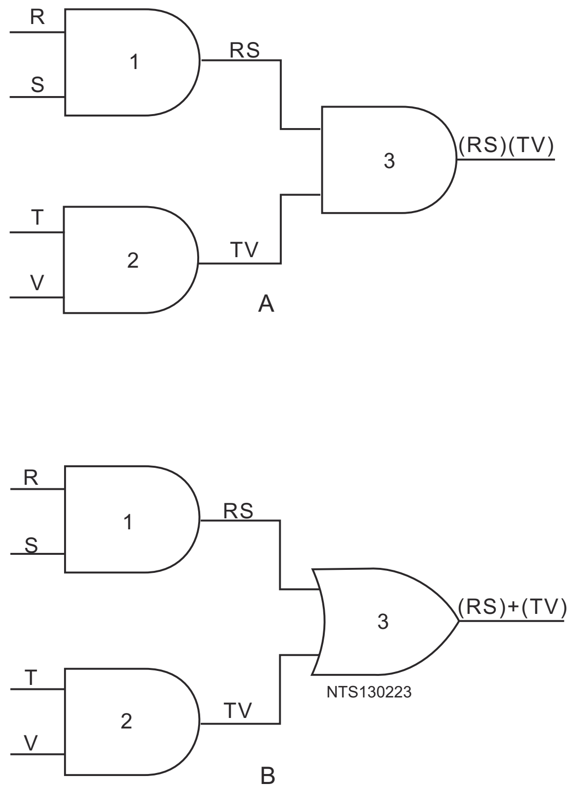


Figure 2-23. —Logic gate combinations: A. Simple combination of AND gates; B. Simple combination of AND gates and OR gate.

The output Boolean expression of gate 1 is RS, and the output expression of gate 2 is TV. These two output expressions become the inputs to gate 3. Remember, the output Boolean expression is the result of the inputs, in this case (RS)(TV); spoken "quantity R AND S AND quantity T AND V."

In view B we have changed gate 3 to an OR gate. The outputs of gates 1 and 2 remain the same but the output of gate 3 changes as you would expect. The output of gate 3 is now (RS)+(TV); spoken "quantity R AND S OR quantity T AND V."

In figure 2-24 (view A), the outputs of two OR gates are being applied as the input to third OR gate. The output for gate 1 is R+S, and the output for gate 2 is T+V. With these inputs, the output expression of gate 3 is (R+S)+(T+V).

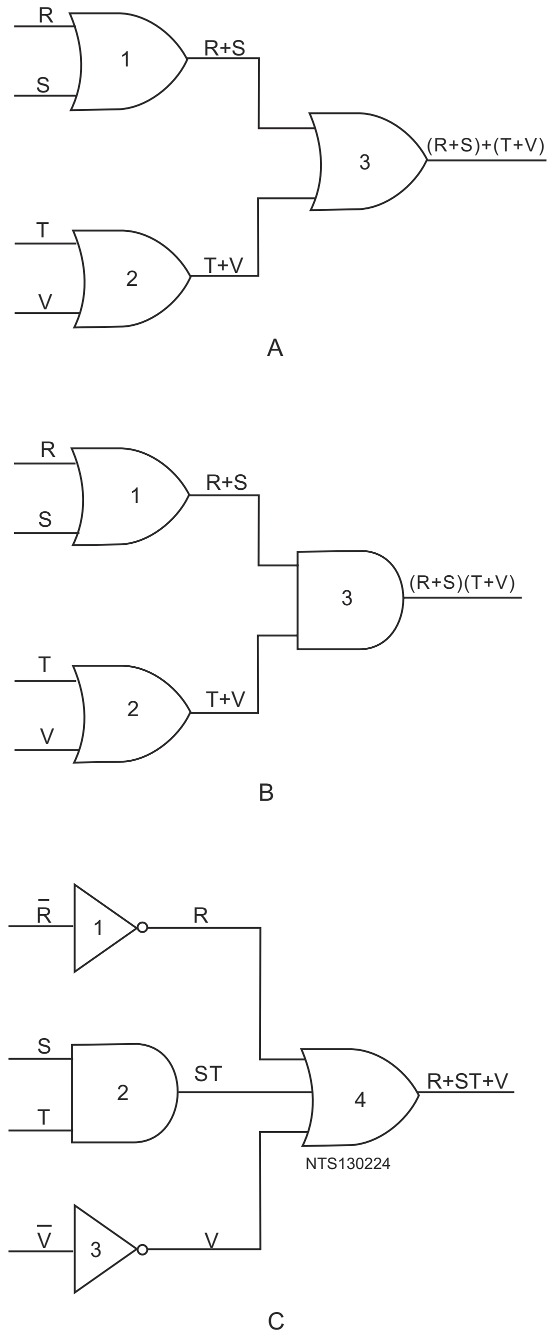


Figure 2-24. —Logic gate combinations: A. Simple combination of OR gates; B. Simple combination of OR gates and AND gate; C. Output expression without the parentheses.

In view B, gate 3 has been changed to an AND gate. The outputs of gates 1 and 2 do not change, but the output expression of gate 3 does. In this case, the gate 3 output expression is (R+S)(T+V). This expression is spoken, "quantity R OR S AND quantity T OR V." The parentheses are used to separate the input terms and to indicate the AND function. Without the parentheses the output expression would read R+ST+V, which is representative of the circuit in view C. As you can see, this is not the same circuit as the one depicted in view B. It is very important that the Boolean expressions be written and spoken correctly.

The Truth Table for the output expression of gate 3 (view B) will help you better understand the output. When studying this Truth Table, notice that the only time f is HIGH (logic 1) is when either or both R and S AND either or both T and V are HIGH (logic 1).

R	S	T	V	f
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	0
0	1	0	0	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	1
1	0	0	0	0
1	0	0	1	1
1	0	1	0	1
1	0	1	1	1
1	1	0	0	0
1	1	0	1	1
1	1	1	0	1
1	1	1	1	1
f = (R+S) (T+V)				

Now let's determine the output expression for the NOR gate in figure 2-25. First write the outputs of gates 1, 2, and 3:

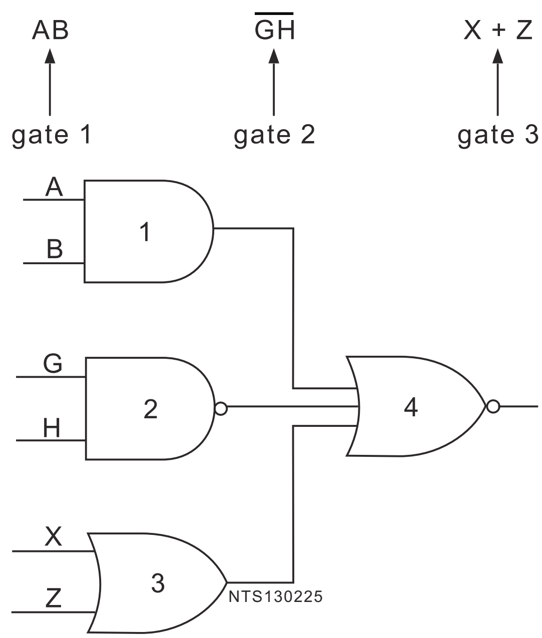


Figure 2-25. —Logic gate combinations.

Since all three outputs are applied to gate 4, proceed as you would for any NOR gate. We separate each input to gate 4 with an OR sign (+) and then place a vinculum over the entire expression. The output expression of gate 4 is:

$$\overline{(AB) + (\overline{GH}) + (X + Z)}$$

The Truth Table shown below is only for gate 4.

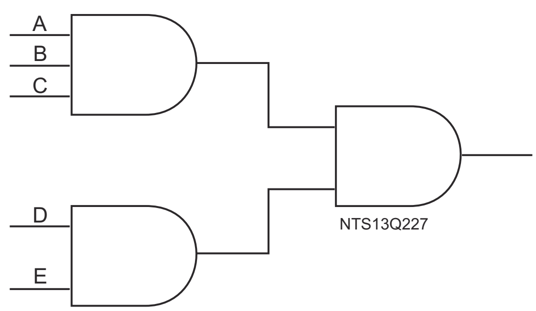
(AB)	(GH)	(X+Z)	f
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	0
$f = \overline{(AB) + (\overline{GH}) + (X + Z)}$			

NTS130225A

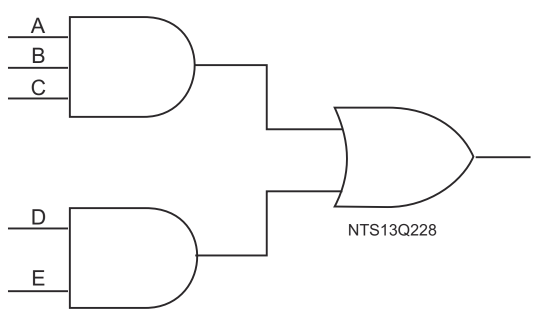
When you are trying to determine the outputs of logic gates in combination, take them <u>one gate at a time</u>!

Now write the output expressions for the following logic gate combinations:

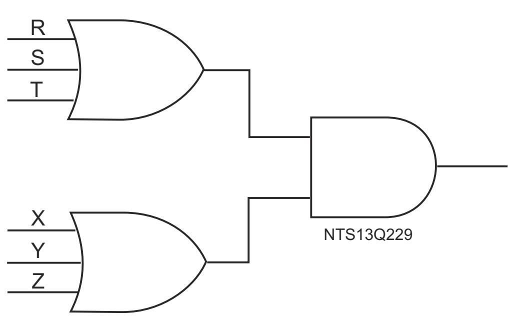
Q27.



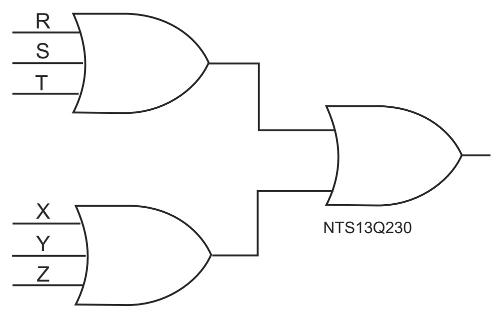
Q28.



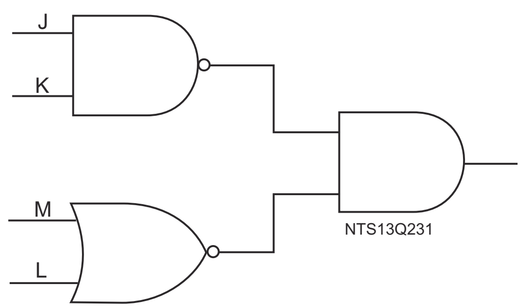
Q29.



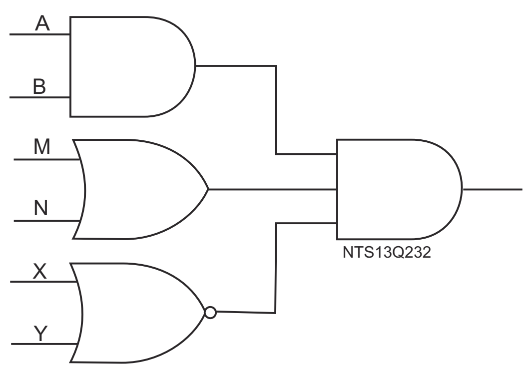
Q30.



Q31.



Q32.



BOOLEAN ALGEBRA

Boolean logic, or Boolean algebra as it is called today, was developed by an English mathematician, George Boole, in the 19th century. He based his concepts on the assumption that most quantities have two possible conditions — TRUE and FALSE. This is the same theory you were introduced to at the beginning of this chapter.

Throughout our discussions of fundamental logic gates, we have mentioned Boolean expressions. A Boolean expression is nothing more than a description of the input conditions necessary to get the desired output. These expressions are based on Boole's laws and theorems.

PURPOSE

Boolean algebra is used primarily by design engineers. Using this system, they are able to arrange logic gates to accomplish desired tasks. Boolean algebra also enables the engineers to achieve the desired output by using the fewest number of logic gates. Since space, weight, and cost are important factors in the design of equipment, you would usually want to use as few parts as possible.

Figure 2-26 (view A), shows a rather complex series of gates. Through proper application of Boolean algebra, the circuit can be simplified to the single OR gate shown in view B. Figure 2-27 shows the simplification process and the Boolean laws and theorem used to accomplish it.

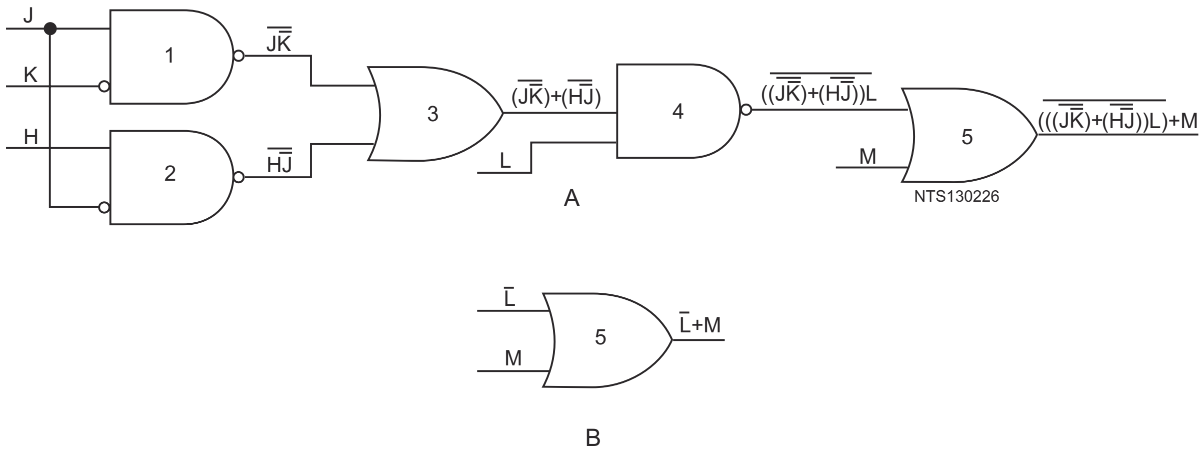


Figure 2-26. —Logic simplification: A. Complex series of gates; B. Simplified single OR gate.

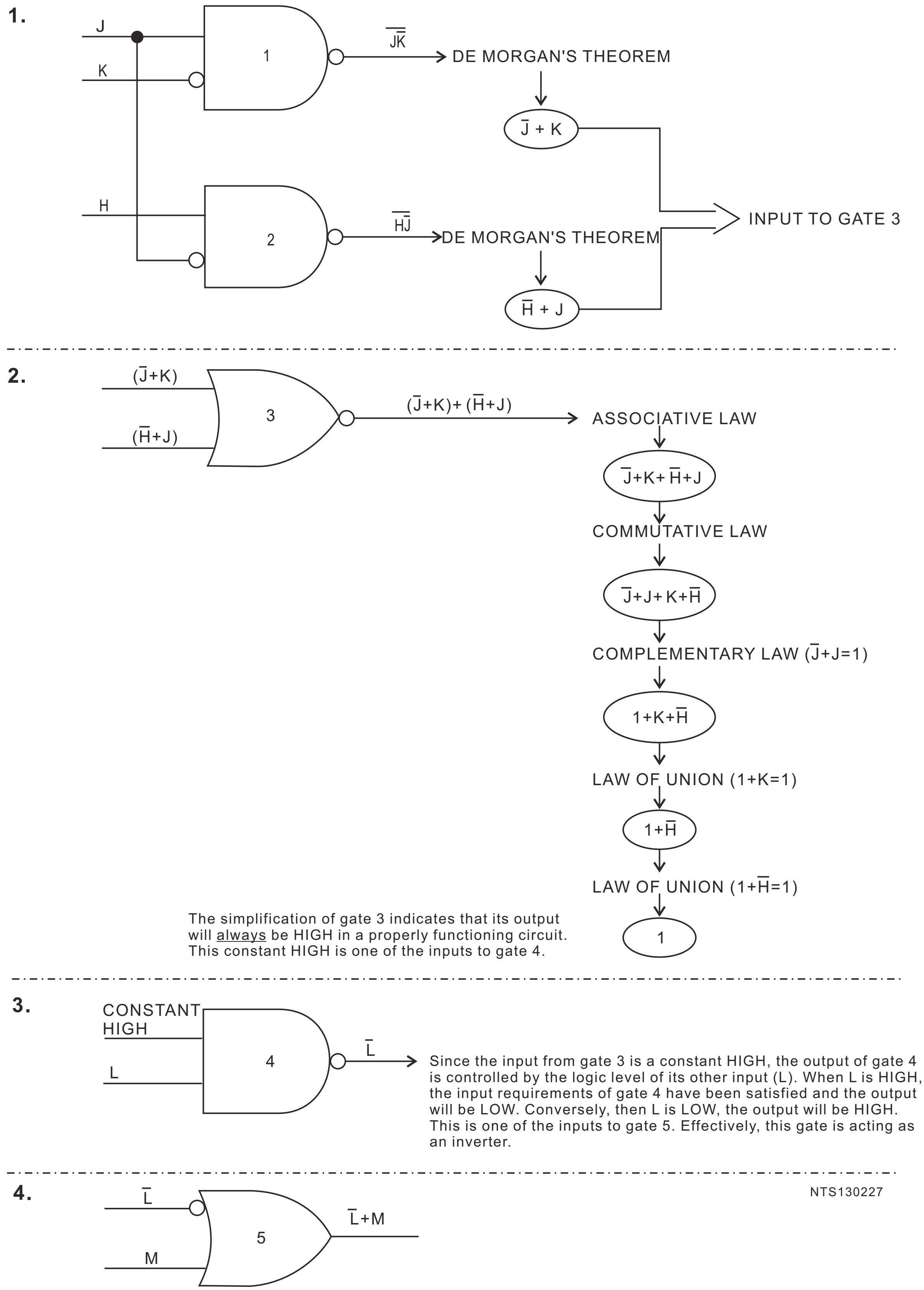


Figure 2-27. —Logic circuit simplification process.

LAWS AND THEOREMS

Each of the laws and theorems of Boolean algebra, along with a simple explanation, is listed below.

LAW OF IDENTITY —a term that is TRUE in one part of an expression will be TRUE in all parts of the expression (A = A or A = A).

COMMUTATIVE LAW —the order in which terms are written does not affect their value (AB = BA, A+B = B+A).

ASSOCIATIVE LAW —a simple equality statement A(BC) = ABC or A+(B+C) = A+B+C.

IDEMPOTENT LAW —a term ANDed with itself or ORed with itself is equal to that term (AA = A, A+A = A).

DOUBLE NEGATIVE LAW —a term that is inverted twice is equal to the term $\overline{\overline{A}}$ = **A.**

COMPLEMENTARY LAW —a term ANDed with its complement equals 0, and a term ORed with its complement equals 1 ($A\overline{A} = 0$, $A+\overline{A} = 1$).

LAW OF INTERSECTION —a term ANDed with 1 equals that term and a term ANDed with 0 equals 0 ($A\cdot1 = A$, $A\cdot0 = 0$).

LAW OF UNION —a term ORed with 1 equals 1 and a term ORed with 0 equals that term (A+1 = 1, A+0 = A).

DeMORGAN'S THEOREM —this theorem consists of two parts: (1) $\overline{AB} = \overline{A} + \overline{B}$ and (2) $\overline{A + B} = \overline{A} \cdot \overline{B}$ (Look at the fourth and eighth sets of gates in table 2-4).

DISTRIBUTIVE LAW —(1) a term (A) ANDed with an parenthetical expression (B+C) equals that term ANDed with each term within the parenthesis: $A\cdot(B+C) = AB+AC$; (2) a term (A) ORed with a parenthetical expression (B ·C) equals that term ORed with each term within the parenthesis: $A+(BC) = (A+B)\cdot(A+C)$.

LAW OF ABSORPTION —this law is the result of the application of several other laws: $A\cdot(A+B) = A$ or $A+(AB) = A$.

LAW OF COMMON IDENTITIES —the two statements $A\cdot(\overline{A}+B) = AB$ and $A+\overline{A}B = A+B$ are based on the complementary law.

Table 2-5. —Boolean Laws and Theorems

1.	Law of Identity	$A = A$ $\overline{A} = \overline{A}$
2.	Commutative Law	$A \cdot B = B \cdot A$ $A + B = B + A$
3.	Associative Law	$A \cdot (B \cdot C) = A \cdot B \cdot C$ $A + (B + C) = A + B + C$
4.	Idempotent Law	$A \cdot A = A$ $A + A = A$
5.	Double Negative Law	$\overline{\overline{A}} = A$
6.	Complementary Law	$A \cdot \overline{A} = 0$ $A + \overline{A} = 1$
7.	Law of Intersection	$A \cdot 1 = A$ $A \cdot 0 = 0$
8.	Law of Union	$A + 1 = 1$ $A + 0 = A$
9.	DeMorgan's Theorem	$\overline{AB} = \overline{A} + \overline{B}$ $\overline{A + B} = \overline{A}\,\overline{B}$
10.	Distributive Law	$A \cdot (B + C) = (A \cdot B) + (A \cdot C)$ $A + (BC) = (A + B) \cdot (A + C)$
11.	Law of Absorption	$A \cdot (A + B) = A$ $A + (AB) = A$
12.	Law of Common Identities	$A \cdot (\overline{A} + B) = AB$ $A + (\overline{A}B) = A + B$

NTS13T205

If you wish a more detailed study of Boolean algebra, we suggest you obtain *Mathematics, Volume 3*, NAVEDTRA 10073-A1.

Q33. Boolean algebra is based on the assumption that most quantities have _______ conditions.

Q34. Boolean algebra is used primarily by _______ to simplify circuits.

SUMMARY

This chapter has presented information on logic, fundamental logic gates, and Boolean laws and theorems. The information that follows summarizes the important points of this chapter.

LOGIC is the development of a logical conclusion based on known information.

Computers operate on the assumption that statements have two conditions — **TRUE** and **FALSE.**

POSITIVE LOGIC is defined as follows: If the signal that activates the circuit (the 1 state) has a voltage level that is more POSITIVE than the 0 state, then the logic polarity is considered to be POSITIVE.

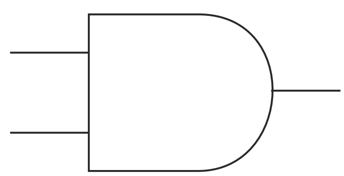
NEGATIVE LOGIC is defined as follows: If the signal that activates the circuit (the 1 state) has a voltage level that is more NEGATIVE than the 0 state, then the logic polarity is considered to be NEGATIVE.

In **DIGITAL LOGIC** (positive or negative), the TRUE condition of a statement is represented by the logic 1 state and the FALSE condition is represented by the logic 0 state.

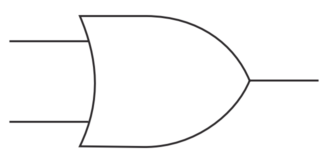
LOGIC LEVELS High and LOW represent the voltage levels of the two logic states. Logic level HIGH represents the more positive voltage while logic level LOW represents the less positive (more negative) voltage. In positive logic, the HIGH level corresponds to the TRUE or 1 state and the LOW level corresponds to the FALSE or 0 state. In negative logic, the HIGH level corresponds to the FALSE or 0 state and the LOW level corresponds to the TRUE or 1 state.

A **BOOLEAN EXPRESSION** is a statement that represents the inputs and outputs of logic gates.

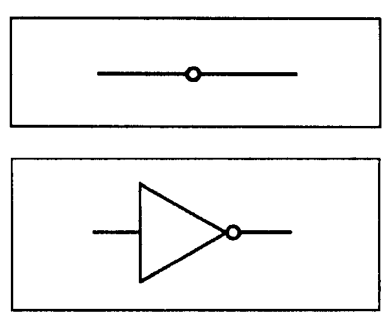
The **AND GATE** requires all inputs to be HIGH at the same time in order to produce a HIGH output.



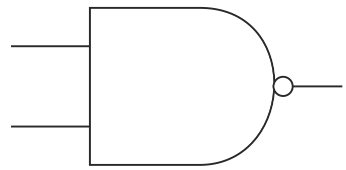
The **OR GATE** requires one or both inputs to be HIGH in order to produce a HIGH output.



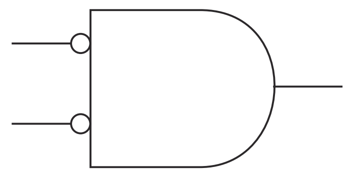
INVERTER (NOT function or negator) is a logic gate used to complement the state of the input variable; that is, a 1 becomes a 0 or a 0 becomes a 1. It may be used on any input or output of any gate to obtain the desired result.

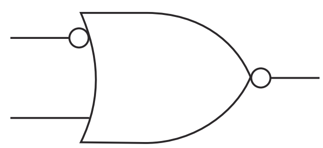


The NAND GATE functions as an AND gate with an inverted output.



The NOR GATE functions as an OR gate with an inverted output.





When deriving the output Boolean expression of a combination of gates, solve one gate at a time.

Boolean algebra is used primarily for the design and simplification of circuits.

ANSWERS TO QUESTIONS Q1. THROUGH Q34.

A1. Logic.

A2. The opposite of the original statement.

A3.

a. $\overline{Q}$,

b. $\overline{R}$,

c. $\overline{V}$,

d. $\overline{Z}$

A4. Positive.

A5. Positive.

A6. Negative.

A7. $f = RS$.

A8. Both must be 1s (HIGH) at the same time.

A9. *16.*

A10. *$f = G+K+L$.*

A11. *Eight.*

A12. *Seven.*

A13. $\overline{XYZ}$

A14. $\overline{X}$ *+ (YZ).*

A15. *No.*

A16. *HIGH.*

A17. *All inputs must be HIGH.*

A18. $\overline{A\overline{B}C}$

A19. *Low.*

A20. *It has an inverter on the output.*

A21. *Low.*

A22. $\overline{R + T}$

A23. *All inputs must be low.*

A24. $A\overline{B}$.

A25. *OR gate.*

A26.

- *a.* $\overline{R + \overline{T}}$
- *b.* $\overline{K}\,\overline{L}$
- *c.* $\overline{\overline{F}}\,\overline{\overline{G}}$
- *d.* $\overline{\overline{F} + \overline{G}}$

A27. *(ABC)(DE).*

A28. *(ABC)+(DE).*

A29. *(R+S+T) (X+Y+Z).*

A30. *(R+S+T)+(X+Y+Z).*

A31. $(\overline{JK})(\overline{M + L})$.

A32. $(AB)(M + N)(\overline{X + Y})$.

A33. Two.

A34. Design engineers.

CHAPTER 3

SPECIAL LOGIC CIRCUITS

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to do the following:

1. Recognize the types of special logic circuits used in digital equipment.
2. Identify exclusive OR and exclusive NOR circuits and interpret their respective Truth Tables.
3. Identify adder and subtracter circuits.
4. Identify the types of flip-flops used in digital equipment and their uses.
5. Identify counters, registers, and clock circuits.
6. Describe the elements that make up logic families — RTL, DTL, TTL, CMOS.

INTRODUCTION

Figure 3-1 is a portion of a typical logic diagram. It is similar to the diagrams you will encounter as your study of digital circuitry progresses.

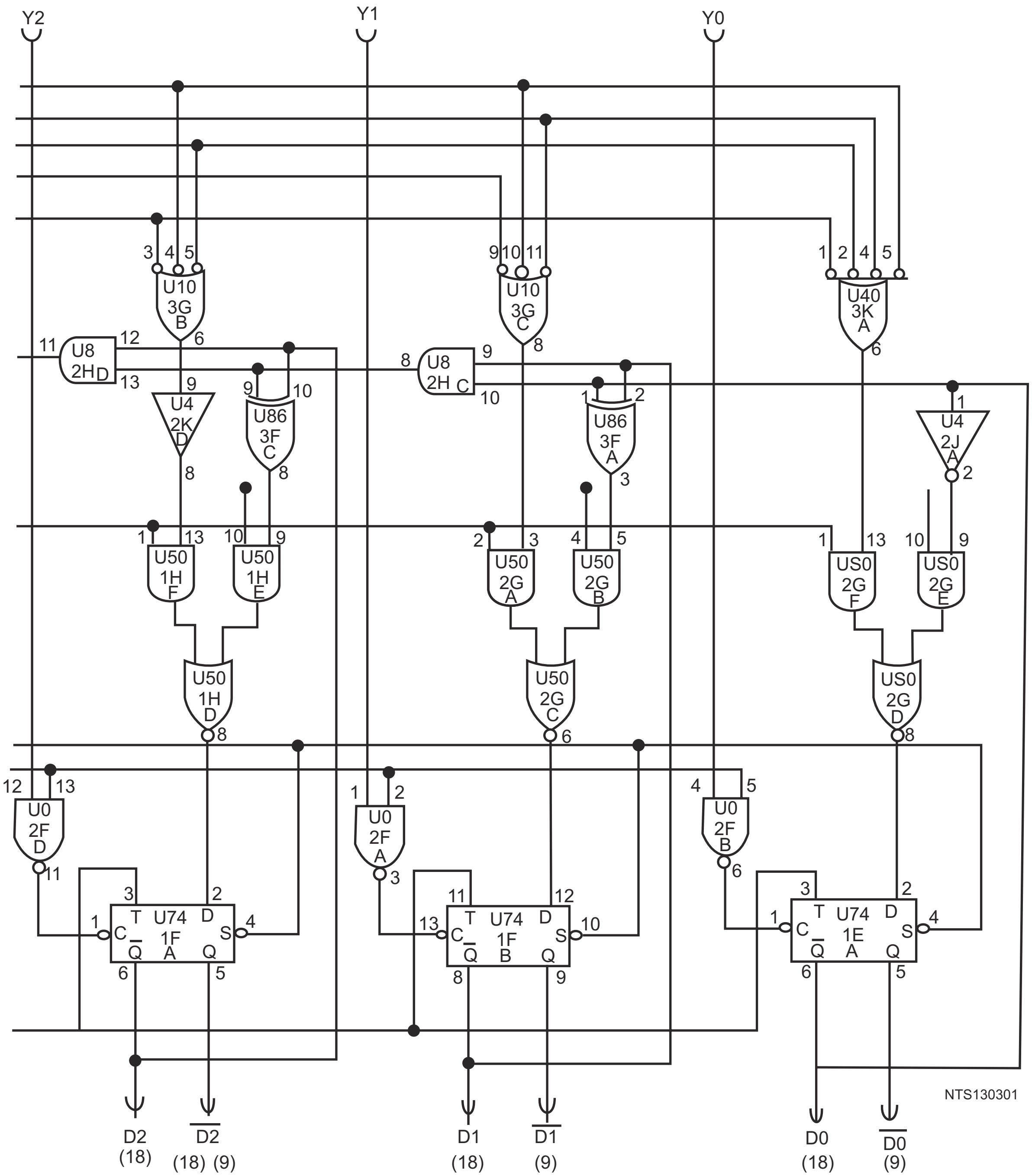


Figure 3-1. —Typical logic diagram.

Look closely at the figure. You will see many familiar logic gates. You will also see several that you may not recognize.

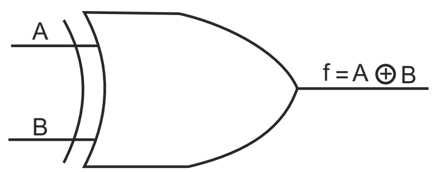
Digital equipment must be capable of many more operations than those described in chapter 2. Provisions must be made for accepting information; performing arithmetic or logic operations; and transferring, storing, and outputting information. Timing circuits are included to ensure that all operations occur at the proper time.

In this chapter you will become acquainted with the logic circuits used to perform the operations mentioned above.

THE EXCLUSIVE OR GATE

The exclusive OR gate is a modified OR gate that produces a HIGH output when only one of the inputs is HIGH. You will often see the abbreviation X-OR used to identify this gate. When both inputs are HIGH or when both inputs are LOW, the output is LOW.

The standard symbol for an exclusive OR gate is shown in figure 3-2 along with the associated Truth Table. The operation function sign for the exclusive OR gate is ⊕.



A	B	f
0	0	0
0	1	1
1	0	1
1	1	0

NTS130302

Figure 3-2. —Exclusive OR gate and Truth Table.

If you were to observe the input and output signals of an X-OR gate, the results would be similar to those shown in figure 3-3. At T_0, both inputs are LOW and the output is LOW. At T_1, A goes to HIGH and remains HIGH until T_2. During this time the output is HIGH. At T_3, B goes HIGH and remains HIGH through T_5. At T_4, A again goes HIGH and remains HIGH through T_5. Between $T_3$and T_4 , the output is HIGH. At T_4, when both A and B are HIGH, the output goes LOW.

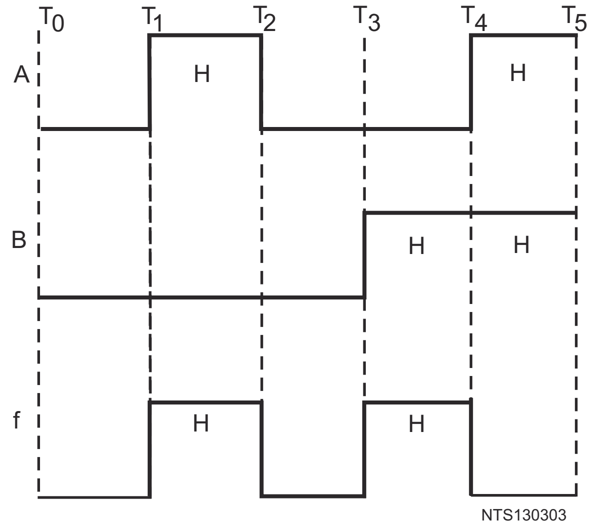
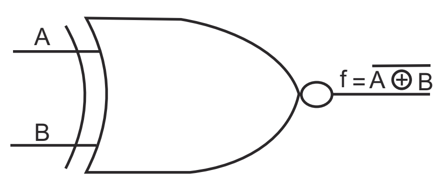


Figure 3-3. —Exclusive OR gate timing diagram.

THE EXCLUSIVE NOR GATE

The exclusive NOR (X-NOR) gate is nothing more than an X-OR gate with an inverted output. It produces a HIGH output when the inputs are either all HIGH or all LOW. The standard symbol and the Truth Table are shown in figure 3-4. The operation function sign is ⊕ with a vinculum over the entire expression.



A	B	f
0	0	1
0	1	0
1	0	0
1	1	1

NTS130304

Figure 3-4. —Exclusive NOR gate and Truth Table.

A timing diagram for the X-NOR gate is shown in figure 3-5. You can see that from T_0 to T_1, when both inputs are LOW, the output is HIGH. The output goes LOW when the inputs are opposite; one HIGH and the other LOW. At time T_3, both inputs go HIGH causing the output to go HIGH.

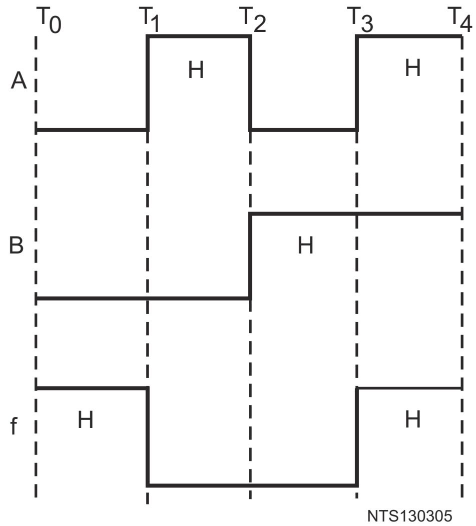


Figure 3-5. —Exclusive NOR gate timing diagram.

Q1. What is the sign of operation for the X-OR gate?

Q2. What will be the output of an X-OR gate when both inputs are HIGH?

Q3. A two-input X-OR gate will produce a HIGH output when the inputs are at what logic levels?

Q4. What type of gate is represented by the output Boolean expression $\overline{T \oplus R}$ *?*

Q5. What will be the output of an X-NOR gate when both inputs are LOW?

ADDERS

Adders are combinations of logic gates that combine binary values to obtain a sum. They are classified according to their ability to accept and combine the digits. In this section we will discuss quarter adders, half adders, and full adders.

QUARTER ADDER

A quarter adder is a circuit that can add two binary digits but will not produce a carry. This circuit will produce the following results:

0 plus 0 = 0

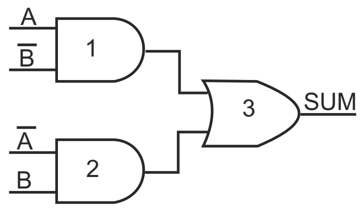
0 plus 1 = 1

1 plus 0 = 1

1 plus 1 = 0 (no carry)

You will notice that the output produced is the same as the output for the Truth Table of an X-OR. Therefore, an X-OR gate can be used as a quarter adder.

The combination of gates in figure 3-6 will also produce the desired results. When A and B are both LOW (0), the output of each AND gate is LOW (0); therefore, the output of the OR gate is LOW (0). When A is HIGH and B is LOW, then $\overline{B}$ is HIGH and AND gate 1 produces a HIGH output, resulting in a sum of 1 at gate 3. With A LOW and B HIGH, gate 2 output is HIGH, and the sum is 1. When both A and B are HIGH, neither AND gate has an output, and the output of gate 3 is LOW (0); no carry is produced.



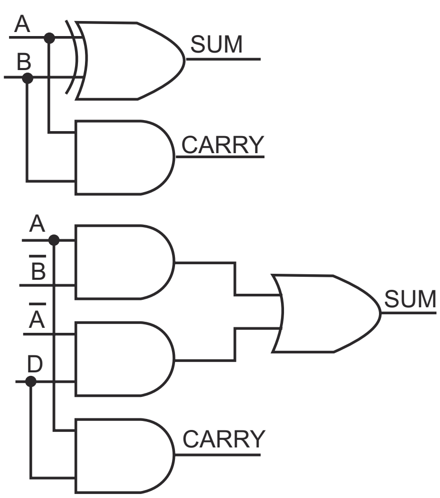
NTS130306

Figure 3-6. —Quarter adder.

HALF ADDER

A half adder is designed to combine two binary digits and produce a carry.

Figure 3-7 shows two ways of constructing a half adder. An AND gate is added in parallel to the quarter adder to generate the carry. The SUM column of the Truth Table represents the output of the quarter adder, and the CARRY column represents the output of the AND gate.



A	D	SUM	CARRY
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

NTS130307

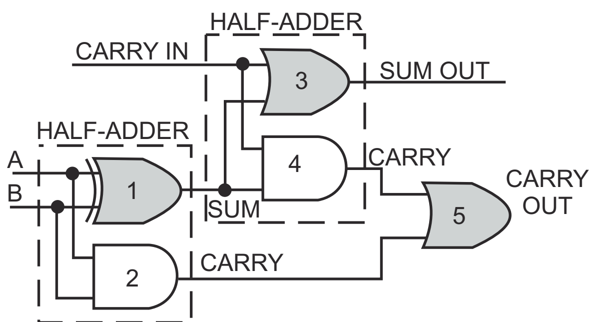
Figure 3-7. —Half adders and Truth Table.

We have seen that the output of the quarter adder is HIGH when either input, but not both, is HIGH. It is only when both inputs are HIGH that the AND gate is activated and a carry is produced. The largest sum that can be obtained from a half adder is 10_2 (1_2 plus 1_2).

FULL ADDER

The full adder becomes necessary when a carry input must be added to the two binary digits to obtain the correct sum. A half adder has no input for carries from previous circuits.

One method of constructing a full adder is to use two half adders and an OR gate as shown in figure 3-8. The inputs A and B are applied to gates 1 and 2. These make up one half adder. The sum output of this half adder and the carry-from a previous circuit become the inputs to the second half adder. The carry from each half adder is applied to gate 5 to produce the carry-out for the circuit.



A	B	CARRY IN	SUM OUT	CARRY OUT
0	0	0	0	0
0	1	0	1	0
1	0	0	1	0
1	1	0	0	1
0	0	1	1	0
0	1	1	0	1
1	0	1	0	1
1	1	1	1	1

NTS130308

Figure 3-8. —Full adder and Truth Table.

Now let's add a series of numbers and see how the circuit operates.

First, let's add 1 and 0. When either A or B is HIGH, gate 1 has an output. This output is applied to gates 3 and 4. Since the carry-in is 0, only gate 3 will produce an output. The sum of 1_2 and 0 is 1_2.

Now let's add 1_2 and 1_2. If A and B are both HIGH, the output of gate 1 is LOW. When the carry-in is 0 (LOW), the output of gate 3 is LOW. Gate 2 produces an output that is applied to gate 5, which produces the carry-out. The sum of 1_2 and 1_2 is 10_2, just as it was for the half adder.

When A and B are both LOW and the carry-in is 1, only gate 3 has an output and produces a sum of 1_2 with no carry-out.

Now, let's add A or B and a carry-in. For example, let's assume that A is HIGH and B is LOW. With these conditions, gate 1 will have an output. This output and the carry-in applied to gates 3 and 4 will produce a sum out of 0 and a carry of 1. This carry from gate 4 will cause gate 5 to produce a carry-out. The sum of A and a carry (1_2 plus 1_2) is 10_2.

When A, B, and the carry-in are all HIGH, a sum of 1 and a carry-out are produced. First, consider A and B. When both are HIGH, the output of gate 1 is LOW, and the output of gate 2 is HIGH, giving us a carry-out at gate 5. The carry-in produces a 1 output at gate 3, giving us a sum of 1. The output of the full adder is 11_2. The sum of 1_2 plus 1_2 plus 1_2 is 11_2.

PARALLEL ADDERS

The adders discussed in the previous section have been limited to adding single-digit binary numbers and carries. The largest sum that can be obtained using a full adder is 11_2.

Parallel adders let us add multiple-digit numbers. If we place full adders in parallel, we can add two- or four-digit numbers or any other size desired.

Figure 3-9 uses STANDARD SYMBOLS to show a parallel adder capable of adding two, two-digit binary numbers. In previous discussions we have depicted circuits with individual logic gates shown. Standard symbols (blocks) allow us to analyze circuits with inputs and outputs only. One standard symbol may actually contain many and various types of gates and circuits. The addend would be input on the A inputs (A_2 = MSD, A_1 = LSD), and the augend input on the B inputs (B_2 = MSD, B_1 = LSD). For this explanation we will assume there is no input to C_0 (carry from a previous circuit).

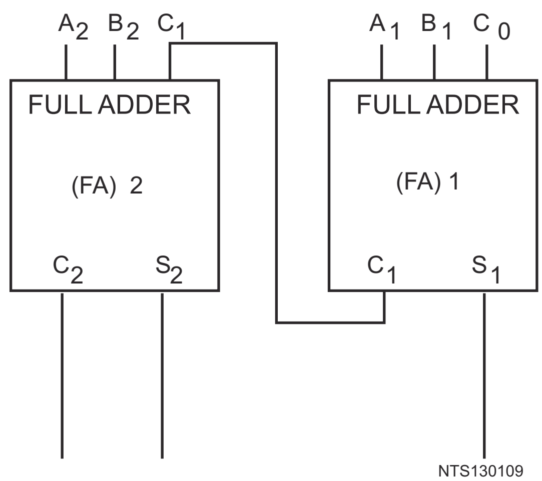


Figure 3-9. —Parallel binary adder.

Now let's add some two-digit numbers. To add 10_2 (addend) and 01_2 (augend), assume there are numbers at the appropriate inputs. The addend inputs will be 1 on A_2 and 0 on A_1. The augend inputs will be 0 on B_2 and 1 on B_1. Working from right to left, as we do in normal addition, let's calculate the outputs of each full adder.

With A_1 at 0 and B_1 at 1, the output of adder 1 will be a sum (S_1) of 1 with no carry (C_1). Since A_2 is 1 and B_2 is 0, we have a sum (S_2) of 1 with no carry (C_2) from adder 1. To determine the sum, read the outputs (C_2, S_2, and S_1) from left to right. In this case, $C_2 = 0$, $S_2 = 1$, and $S_1 = 1$. The sum, then, of 10_2 and 01_2 is 011_2 or 11_2.

To add 11_2 and 01_2, assume one number is applied to A_1 and A_2, and the other to B_1 and B_2, as shown in figure 3-10. Adder 1 produces a sum (S_1) of 0 and a carry (C_1) of 1. Adder 2 gives us a sum (S_2)

of 0 and a carry (C_2) of 1. By reading the outputs (C_2, S_2, and S_1), we see that the sum of 11_2 and 01_2 is 100_2.

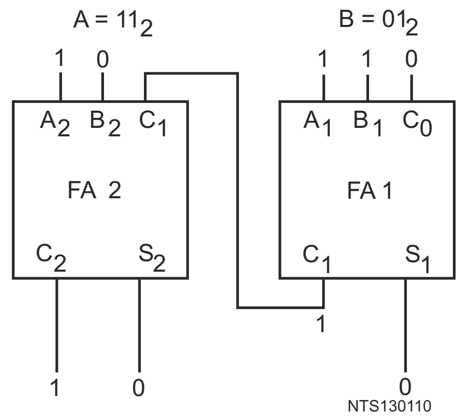


Figure 3-10. —Parallel addition.

As you know, the highest binary number with two digits is 11_2. Using the parallel adder, let's add 11_2 and 11_2.

First, apply the addend and augend to the A and B inputs. Calculate the output of each full adder beginning with full adder 1. With A_1 and B_1 at 1, S_1 is 0 and C_1 is 1. Since all three inputs (A_2, B_2, and C_1) to full adder 2 are 1, the output will be 1 at S_2 and 1 at C_2. The output of the circuit, as you read left to right, is 110_2, the sum of 11_2 and 11_2.

Parallel adders may be expanded by combining more full adders to accommodate the number of digits in the numbers to be added. There must be one full adder for each digit.

Q6. What advantage does a half adder have over a quarter adder?

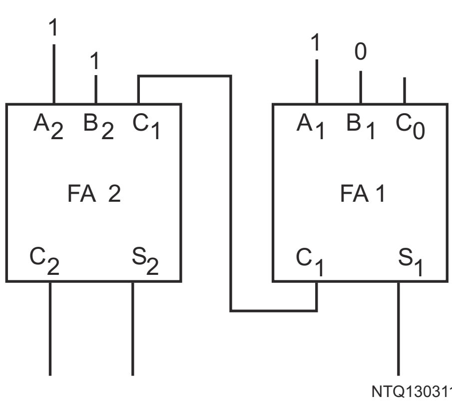
Q7. An X-OR gate may be used as what type of adder?

Q8. What will be the output of a half adder when both inputs are 1s?

Q9. What type of adder is used to handle a carry from a previous circuit?

Q10. How many full adders are required to add four-digit numbers?

Q11. With the inputs shown below, what will be the output of S_1, S_2, and C_2?



Q12. What is the output of C_1?

SUBTRACTION

Subtraction is accomplished in computers by the R's complement and add method. This is the same method you used in chapter 1 to subtract binary numbers.

R's complement subtraction allows us to use fewer circuits than would be required for separate add and subtract functions. Adding X-OR gates to full adders, as shown in figure 3-11, enables the circuit to perform R's complement subtraction as well as addition.

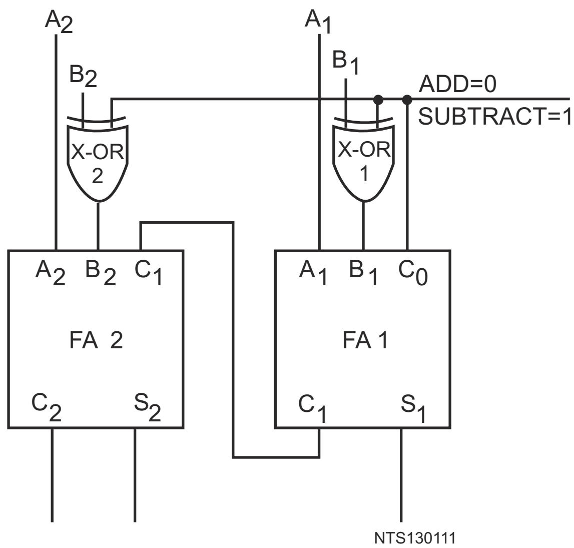


Figure 3-11. —R's complement adder/subtracter.

To add two numbers using this circuit, the addend and augend are applied to the A and B inputs. The B inputs are applied to one input of the X-OR gates. A control signal is applied to the other input of the X-OR gates. When the control signal is LOW, the circuit will add; and when it is HIGH, the circuit will subtract.

In the add mode, the outputs of the X-OR gates will be the same as the B inputs. Addition takes place in the same manner as described in parallel addition.

Before we attempt to show subtraction, let's review R's complement subtraction. To subtract 10_2 from 11_2, write down the minuend (11_2). Perform the R's complement on the subtrahend. Now add the minuend and the complemented subtrahend.

11_2	minuend
$+ \quad 10_2$	R's complement
101	Difference

Disregard the most significant 1, and the difference between 11_2 and 10_2 is 01_2. The most significant 1 will not be used in the example shown in the following paragraph.

Now let's subtract 10_2 from 11_2 using the adder/subtracter circuit. The minuend (11_2) is input on the A terminals, and the subtrahend (10_2) is input on the B terminals. In the subtract mode, a 1 from the control circuit is input to each of the X-OR gates and to the C_0 carry input. By applying a 1 to each of the X-OR gates, you find the output will be the complement of the subtrahend input at B_1 and B_2. Since B_1 is a 0, the output of X-OR 1 will be 1. The input B_2 to X-OR 2 will be inverted to a 0. The HIGH input to C_0 acts as a carry from a previous circuit. The combination of the X-OR gates and the HIGH at C_0 produces the R's complement of the subtrahend. The full adders add the minuend and the R's complement of the subtrahend and produce the difference. The output of C_2 is not used. The outputs of S_2 and S_1 are 0 and 1, respectively, indicating a difference of 01_2. Therefore, 11_2 minus 10_2 equals 01_2.

Q13. What type of logic gates are added to a parallel adder to enable it to subtract?

Q14. How many of these gates would be needed to add a four-digit number?

Q15. In the add mode, what does the output of C_2 indicate?

Q16. In the subtract mode, a 1 at C_0 performs what portion of the R's complement?

Q17. In the subtract mode, which portion of the problem is complemented?

FLIP-FLOPS

Flip-flops (FFs) are devices used in the digital field for a variety of purposes. When properly connected, flip-flops may be used to store data temporarily, to multiply or divide, to count operations, or to receive and transfer information.

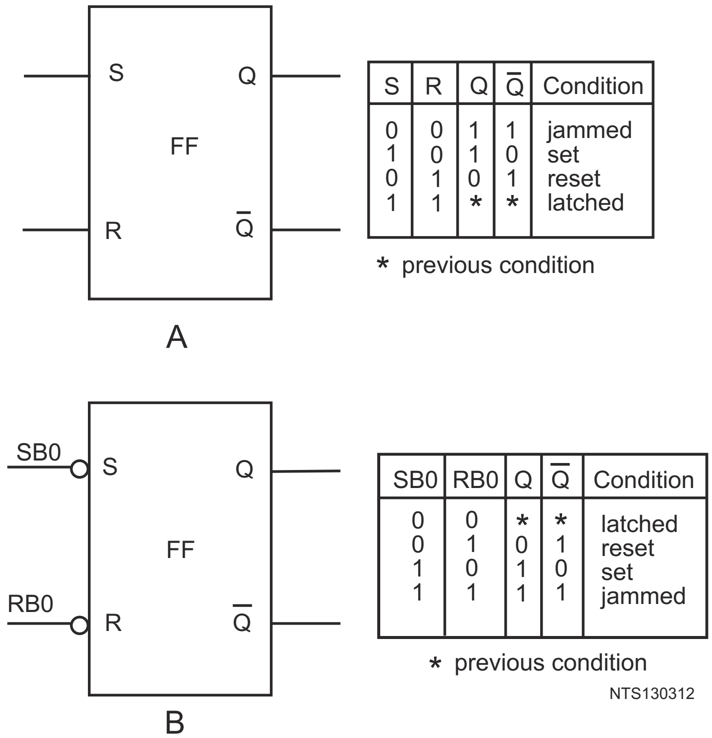
Flip-flops are bistable multivibrators. The types used in digital equipment are identified by the inputs. They may have from two up to five inputs depending on the type. They are all common in one respect. They have two, and only two, distinct output states. The outputs are normally labeled Q and $\overline{Q}$ and should always be complementary. When Q = 1, then $\overline{Q}$ = 0 and vice versa.

In this section we will discuss four types of FFs that are common to digital equipment. They are the R-S, D, T, and J-K FFs.

R-S FLIP-FLOP

The R-S FF is used to temporarily hold or store information until it is needed. A single R-S FF will store one binary digit, either a 1 or a 0. Storing a four-digit binary number would require four R-S FFs.

The standard symbol for the R-S FF is shown in figure 3-12, view A. The name is derived from the inputs, R for reset and S for set. It is often referred to as an R-S LATCH. The outputs Q and $\overline{Q}$ are complements, as mentioned earlier.



S	R	Q	$\overline{Q}$	Condition
0	0	1	1	jammed
1	0	1	0	set
0	1	0	1	reset
1	1	*	*	latched

SB0	RB0	Q	$\overline{Q}$	Condition
0	0	*	*	latched
0	1	0	1	reset
1	0	1	0	set
1	1	1	1	jammed

Figure 3-12. —R-S flip-flop: A. Standard symbol; B. R-S FF with inverted inputs.

The R-S FF has two output conditions. When the Q output is HIGH and $\overline{Q}$ is LOW, the FF is set. When Q is LOW and $\overline{Q}$ is HIGH, the FF is reset. When the R and S inputs are both LOW, the Q and $\overline{Q}$ outputs will both be HIGH. When this condition exists, the FF is considered to be JAMMED and the outputs cannot be used. The jammed condition is corrected when either S or R goes HIGH.

To set the flip-flop requires a HIGH on the S input and a LOW on the R input. To reset, the opposite is required; S input LOW and R input HIGH. When both R and S are HIGH, the FF will hold or "latch" the condition that existed before both inputs went HIGH.

Because the S input of this FF requires a logic LOW to set, a more easily understood symbol is shown in figure 3-12, view B. Refer to this view while reading the following paragraph.

In our description of R-S FF operation, let's assume that the signals applied to the S and R inputs are the LSDs of two different binary numbers. Let's also assume that these two binary numbers represent the speed and range of a target ship. The LSDs will be called SB0 (Speed Bit 0) and RB0 (Range Bit 0) and will be applied to the S and R inputs respectively. Refer to figure 3-12, view B, and figure 3-13. At time T_0, both SB0 and RB0 are HIGH, as a result, both Q and $\overline{Q}$ are HIGH. This is the jammed state and as mentioned earlier, cannot be used in logic circuitry. At T_1, SB0 goes LOW and RB0 remains HIGH; Q goes LOW and $\overline{Q}$ remains HIGH; the FF is reset. At T_2 RB0 goes LOW and SB0 remains LOW; the FF is latched in the reset condition. At T_3, SB0 goes HIGH and RB0 remains LOW; the FF sets. At T_4 SB0 goes LOW and RB0 goes HIGH; the FF resets. When SB0 and RB0 input conditions reverse at T_5, the FF sets. The circuit is put in the latch condition at T_6 when SB0 goes LOW. Notice that the output changes states ONLY when the inputs are in opposite states.

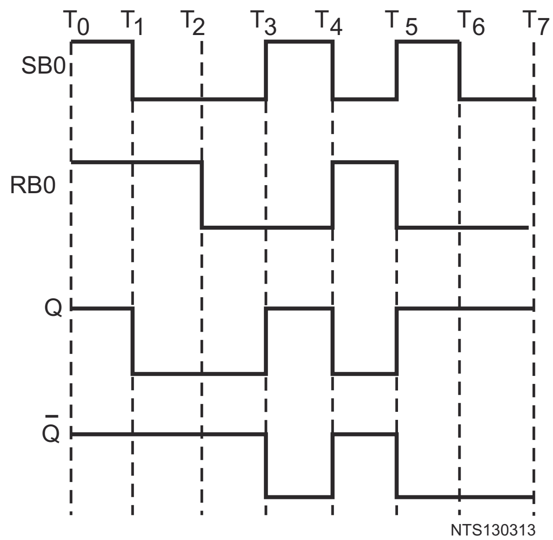


Figure 3-13. —R-S flip-flop with inverted inputs timing diagram.

Figure 3-14 shows two methods of constructing an R-S FF. We can use these diagrams to prove the Truth Table for the R-S FF.

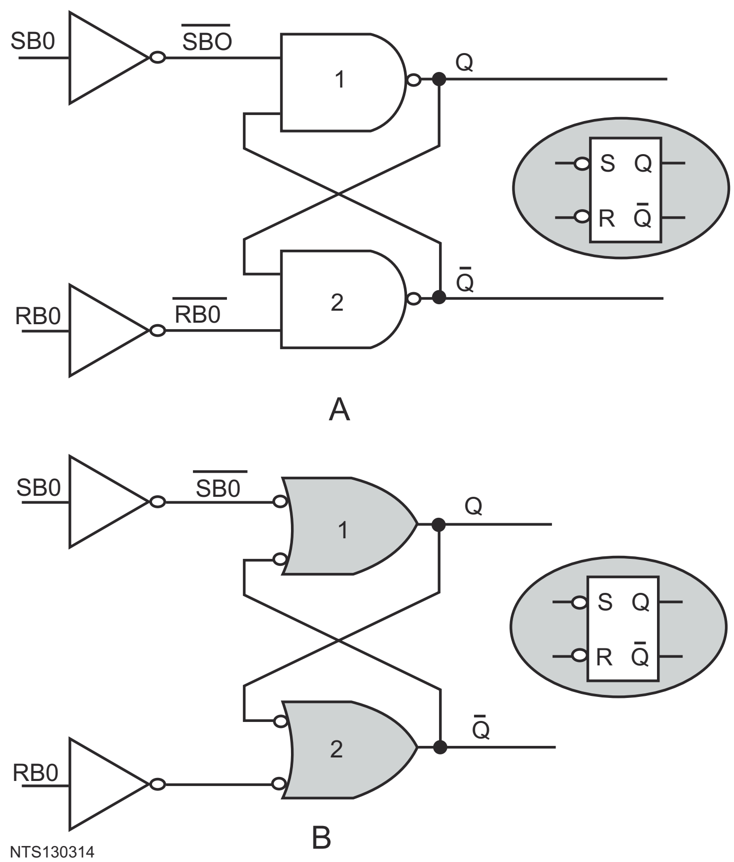


Figure 3-14. —R-S FF construction: A. Using cross-coupled NAND gates; B. Using cross-coupled OR gates.

Look at figure 3-14, view A. Let's assume SB0 is HIGH and RB0 is LOW. You should remember from chapter 2 that the output of an inverter is the complement of the input. In this case, since SB0 is HIGH, $\overline{\text{SBO}}$ will be LOW. The LOW input to NAND gate 1 causes the Q output to go HIGH. This HIGH Q output is also fed to the input of NAND gate 2. The other input to NAND gate 2, $\overline{\text{RBO}}$, is HIGH. With both inputs to gate 2 HIGH, the output goes LOW. The LOW $\overline{\text{Q}}$ output is also fed to NAND gate 1 to be used as the "latch" signal. If SB0 goes LOW while this condition exists, there will be no change to the outputs because the FF would be in the latched condition; both SB0 and RB0 LOW.

When RB0 is HIGH and SB0 is LOW, $\overline{\text{RBO}}$ being LOW drives the output, $\overline{\text{Q}}$, to a HIGH condition. The HIGH $\overline{\text{Q}}$ and HIGH $\overline{\text{SBO}}$ inputs to gate 1 cause the output, Q, to go LOW. This LOW is also fed to NAND gate 2 to be used as the latch signal. Since SB0 is LOW, the FF will again go into the latched mode if RB0 goes LOW.

The cross-coupled OR gates in figure 3-14, view B, perform the same functions as the NAND gate configuration of view A. A HIGH input at SB0 produces a HIGH Q output, and a LOW at RB0 produces a LOW $\overline{\text{Q}}$ output. The cross-coupled signals ($\overline{\text{Q}}$ to gate 1 and Q to gate 2) are used as the latch signals just as in view A. You can trace other changes of the inputs using your knowledge of basic logic gates.

Q18. What are R-S FFs used for?

Q19. How many R-S FFs are required to store the number 100101_2?

Q20. For an R-S FF to change output conditions, the inputs must be in what states?

Q21. How may R-S FFs be constructed?

TOGGLE FLIP-FLOP

The toggle, or T, flip-flop is a bistable device that changes state on command from a common input terminal.

The standard symbol for a T FF is illustrated in figure 3-15, view A. The T input may be preceded by an inverter. An inverter indicates a FF will toggle on a HIGH-to-LOW transition of the input pulse. The absence of an inverter indicates the FF will toggle on a LOW-to-HIGH transition of the pulse.

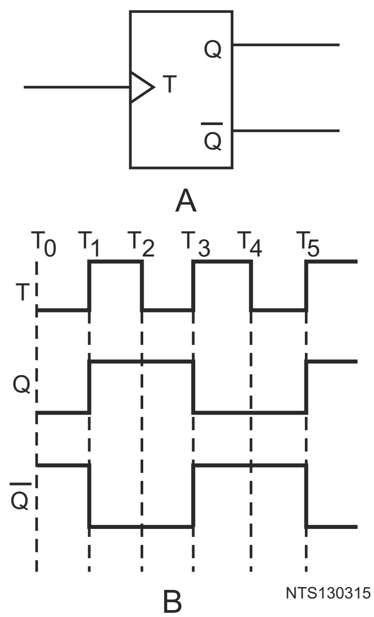


Figure 3-15. —Toggle (T) flip-flop: A. Standard symbol; B. Timing diagram.

The timing diagram in figure 3-15, view B, shows the toggle input and the resulting outputs. We will assume an initial condition (T_0) of Q being LOW and $\overline{Q}$ being HIGH. At T_1, the toggle changes from a LOW to a HIGH and the device changes state; Q goes HIGH and $\overline{Q}$ goes LOW. The outputs remain the same at T_2 since the device is switched only by a LOW-to-HIGH transition. At T_3, when the toggle goes HIGH, Q goes LOW and $\overline{Q}$ goes HIGH; they remain that way until T_5.

Between T_1 and T5, two complete cycles of T occur. During the same time period, only one cycle is observed for Q or $\overline{Q}$. Since the output cycle is one-half the input cycle, this device can be used to divide the input by 2.

The most commonly used T FFs are J-K FFs wired to perform a toggle function. This use will be demonstrated later in this section.

Q22. How many inputs does a T FF have?

Q23. What is the purpose of using T FFs?

D FLIP-FLOP

The D FF is a two-input FF. The inputs are the data (D) input and a clock (CLK) input. The clock is a timing pulse generated by the equipment to control operations. The D FF is used to store data at a predetermined time and hold it until it is needed. This circuit is sometimes called a delay FF. In other words, the data input is delayed up to one clock pulse before it is seen in the output.

The simplest form of a D FF is shown in figure 3-16, view A. Now, follow the explanation of the circuit using the Truth Table and the timing diagram shown in figure 3-16, views B and C.

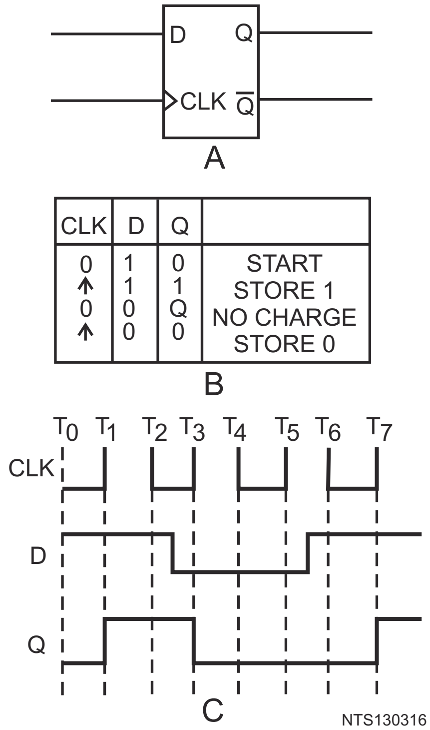


Figure 3-16. —D flip-flop: A. Standard symbol; B. Truth Table; C. Timing diagram.

Depending on the circuit design, the clock (CLK) can be a square wave, a constant frequency, or asymmetrical pulses. In this example the clock (CLK) input will be a constant input at a given frequency. This frequency is determined by the control unit of the equipment. The data (D) input will be present when there is a need to store information. Notice in the Truth Table that output Q reflects the D input only when the clock transitions from 0 to 1 (LOW to HIGH).

Let's assume that at T_0, CLK is 0, D is 1, and Q is 0. Input D remains at 1 for approximately 2 1/2 clock pulses. At T_1, when the clock goes to 1, Q also goes to 1 and remains at 1 even though D goes to 0 between T_2 and T_3. At T_3, the positive-going pulse of the clock causes Q to go to 0, reflecting the condition of D. The positive-going clock pulse at T_5 causes no change in the output because D is still LOW. Between T_5 and T_6, D goes HIGH, but Q remains LOW until T_7 when the clock goes HIGH.

The key to understanding the output of the D FF is to remember that the data (D) input is seen in the output only after the clock has gone HIGH.

You may see D FF symbols with two additional inputs — CLR (clear) and PR (preset). These inputs are used to set the start condition of the FF — CLR sets Q to 0; PR sets Q to 1. Figure 3-17 shows the standard symbol with the CLR and PR inputs. Since these inputs are preceded by inverters (part of the FF), a LOW-going signal is necessary to activate the FF. These signals (CLR and PR) override any existing condition of the output.

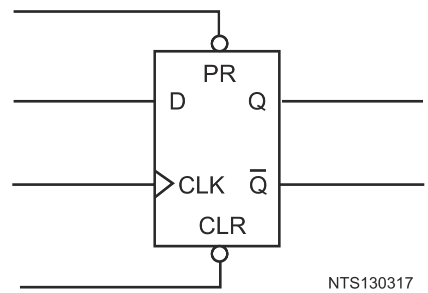


Figure 3-17. —D flip-flop with PR and CLR inputs.

You may also see an inverter at the clock input. In this case, the output will change on the negative-going transition of the clock pulse.

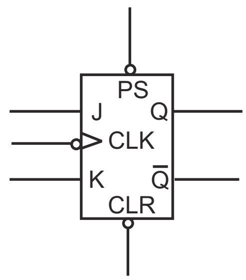
Q24. What are the inputs to a D FF?

Q25. How long is data delayed by a D FF?

Q26. What condition must occur to have a change in the output of a D FF?

J-K FLIP-FLOP

The J-K FF is the most widely used FF because of its versatility. When properly used it may perform the function of an R-S, T, or D FF. The standard symbol for the J-K FF is shown in view A of figure 3-18.

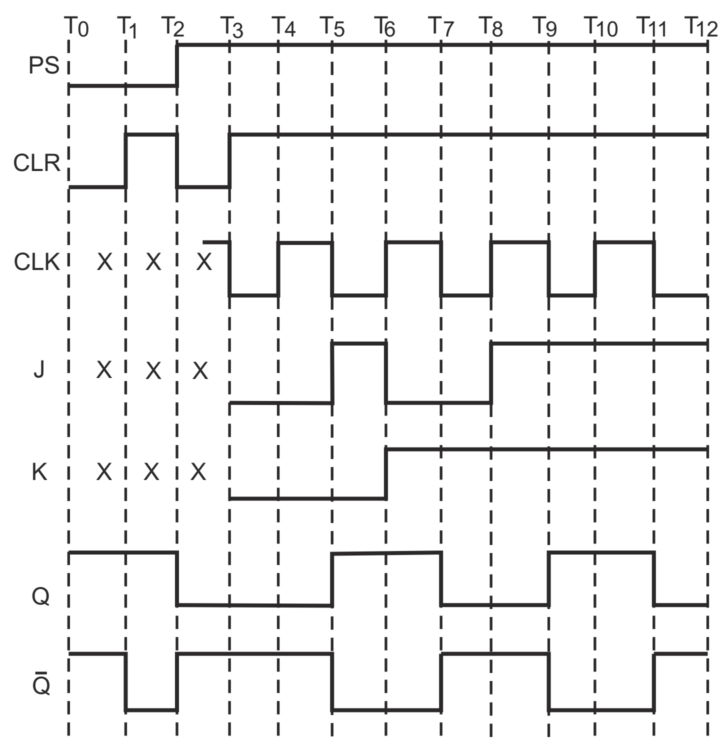


A

	INPUTS					OUTPUTS	
	PRESET	CLEAR	CLOCK	J	K	Q	$\overline{Q}$
1	0	0	X	X	X	1	1
2	0	1	X	X	X	1	0
3	1	0	X	X	X	0	1
4	1	1	↓	0	0	0	$\overline{Q}$
5	1	1	↓	1	0	1	0
6	1	1	↓	0	1	0	1
7	1	1	↓	1	1	TOGGLE	
8	1	1	0	X	X	Q	$\overline{Q}$

X = IRRELEVANT

B



C

Figure 3-18. —J-K flip-flop: A. Standard symbol; B. Truth Table; C. Timing diagram.

The J-K is a five-input device. The J and K inputs are for data. The CLK input is for the clock; and the PS and CLR inputs are the preset and clear inputs, respectively. The outputs Q and $\overline{Q}$ are the normal complementary outputs.

Observe the Truth Table and timing diagram in figure 3-18, views B and C, as the circuit is explained.

Line 1 of the Truth Table corresponds to T_0 in the timing diagram. The PS and CLR inputs are both LOW. The CLK, J, and K inputs are irrelevant. At this point the FF is jammed, and both Q and $\overline{Q}$ are HIGH. As with the R-S FF, this state cannot be used.

At T_1, PS remains LOW while CLR goes HIGH. The Q output remains HIGH and $\overline{Q}$ goes LOW. The FF is in the PRESET condition (line 2 of the Truth Table).

At T_2, PS goes HIGH, CLR goes LOW, Q goes LOW, and $\overline{Q}$ goes HIGH. At this point the FF is CLEARED (line 3 of the Truth Table). The condition of the CLK, J, and K inputs have no effect on the PS and CLR actions since these inputs override the other inputs. Starting at T_3, PS and CLR will be held at HIGHs so as not to override the other actions of the FF. Using the PS and CLR inputs only, the circuit will function as an R-S FF.

Between T_2 and T_3, the CLK input is applied to the device. Since the CLK input has an inverter, all actions will take place on the negative-going transition of the clock pulse.

Line 4 of the Truth Table shows both PS and CLR HIGH, a negative-going CLK, and J and K at 0, or LOW. This corresponds to T_3 on the timing diagram. In this condition the FF holds the previous condition of the output. In this case the FF is reset. If the circuit were set when these inputs occurred, it would remain set.

At time T_5, we have a negative-going clock pulse and a HIGH on the J input. This causes the circuit to set, Q to go HIGH, and $\overline{Q}$ to go LOW. See line 5 of the Truth Table.

At T_6, J goes LOW, K goes HIGH, and the clock is in a positive-going transition. There is no change in the output because all actions take place on the negative clock transition.

At T_7, when J is LOW, K is HIGH; the clock is going negative, the FF resets, Q goes LOW, and $\overline{Q}$ goes HIGH (line 6).

With both J and K HIGH and a negative-going clock (as at T_9 and line 7), the FF will toggle or change state with each clock pulse. It will continue to toggle as long as J and K both remain HIGH.

Line 8 of the Truth Table indicates that as long as the clock is in any condition other than a negative-going transition, there will be no change in the output regardless of the state of J or K.

As mentioned at the beginning of this section, J-K FFs may be used as R-S, T, or D FFs.

Figure 3-19 shows how a J-K can be made to perform the other functions.

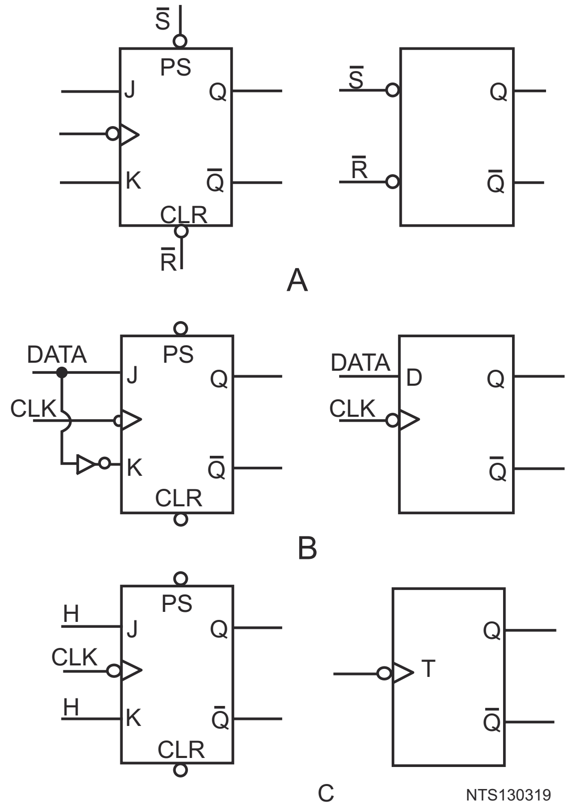


Figure 3-19. —J-K versatility: A. Using just the PS and CLR inputs; B. Data applied to the J input; C. Both J and K inputs held HIGH.

In view A, using just the PS and CLR inputs of the J-K will cause it to react like an R-S FF.

In view B, data is applied to the J input. This same data is applied to the K input through an inverter to ensure that the K input is in the opposite state. In this configuration, the J-K performs the same function as a D FF.

View C shows both the J and K inputs held at 1, or HIGH. The FF will change state or toggle with each negative-going transition of the clock just as a T FF will.

Now you can see the versatility of the J-K FF.

Q27. What type of FF can be used as an R-S, a T, or a D FF?

Q28. What will be the output of Q if J is HIGH, PS and CLR are HIGH, and the clock is going negative?

Q29. Assume that K goes HIGH and J goes LOW; when will the FF reset?

Q30. What logic levels must exist for the FF to be toggled by the clock?

Q31. What two inputs to a J-K FF will override the other inputs?

Q32. How is the J-K FF affected if PS and CLR are both LOW?

CLOCKS AND COUNTERS

Clocks and counters are found in all types of digital equipment. Although they provide different functions, they are all constructed of circuits with which you are familiar. By changing the way the circuits are interconnected, we can build timing circuits, multipliers and dividers, and storage units. In this section we will discuss the purpose, construction, and operation of these important digital circuits.

CLOCKS

Clocks have been mentioned in the preceding section with regard to their action with FFs. You will recall that the clock is a timing signal generated by the equipment to control operations. This control feature is demonstrated in both the D and J-K FFs. Remember that the clock output had to be in a certain condition for the FFs to perform their functions.

The simplest form of a clock is the astable or free-running multivibrator. A schematic diagram of a typical free-running multivibrator is shown in figure 3-20 along with its output waveforms. This multivibrator circuit is called free running because it alternates between two different output voltages during the time it is active. Outputs 1 and 2 will be equal and opposite since Q_1 and Q_2 conduct alternately. The frequency of the outputs may be altered within certain limits by varying the values of R_2C_1 and R_3C_2. You may want to review the operation of the astable multivibrator in NEETS, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*. Although the astable multivibrator circuit seems to produce a good, balanced square wave, it lacks the frequency stability necessary for some types of equipment.

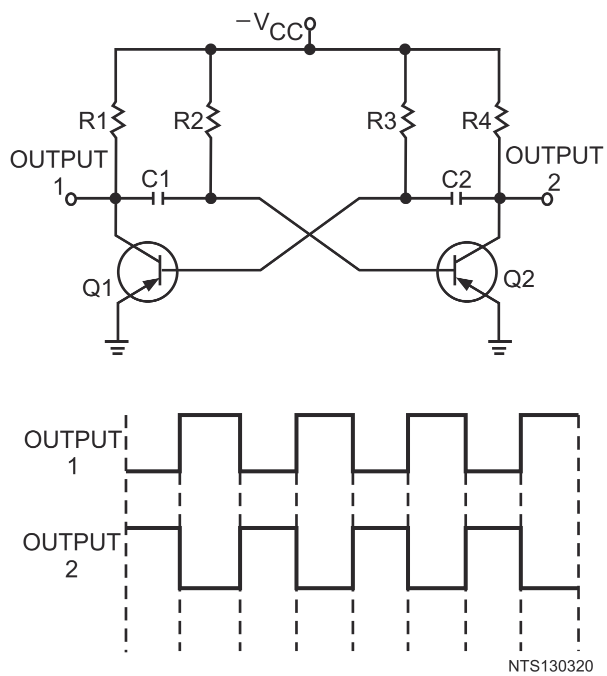


Figure 3-20. —Free-running multivibrator.

The frequency stability of the astable multivibrator can be increased by applying a trigger pulse to the circuit. The frequency of the trigger must be higher than the free-running frequency of the multivibrator. The output frequency will match the trigger frequency and produce a more stable output.

Another method of producing a stable clock pulse is to use a triggered monostable or one-shot multivibrator. You will recall from NEETS, Module 9, that a one-shot multivibrator has one stable state and will only change states when acted on by an outside source (the trigger). A block diagram of a monostable multivibrator with input and output signals is shown in figure 3-21. The duration of the output pulse is dependent on the charge time of an RC network in the multivibrator. Each trigger input results in a complete cycle in the output, as shown in figure 3-21. Trigger pulses are supplied by an oscillator.

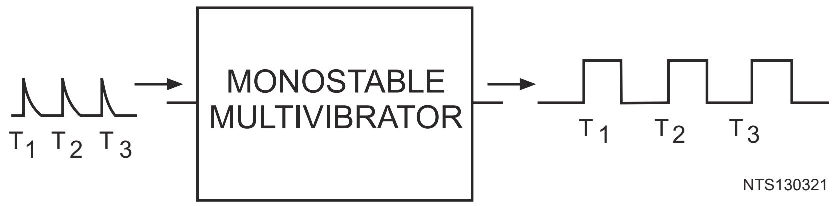


Figure 3-21. —Monostable multivibrator block diagram.

The circuits described previously are very simple clocks. However, as the complexity of the system increases, so do the timing requirements. Complex systems have multiphase clocks to control a variety of operations. Multiphase clocks allow functions involving more than one operation to be completed during a single clock cycle. They also permit an operation to extend over more than one clock cycle.

A block diagram of a two-phase clock system is shown in figure 3-22, view A. The astable multivibrator provides the basic timing for the circuit, while the one-shot multivibrators are used to shape the pulses. Outputs Q and $\overline{Q}$ are input to one-shot multivibrators 1 and 2, respectively. The resulting outputs are in phase with the inputs, but the duration of the pulse is greatly reduced as shown in view B.

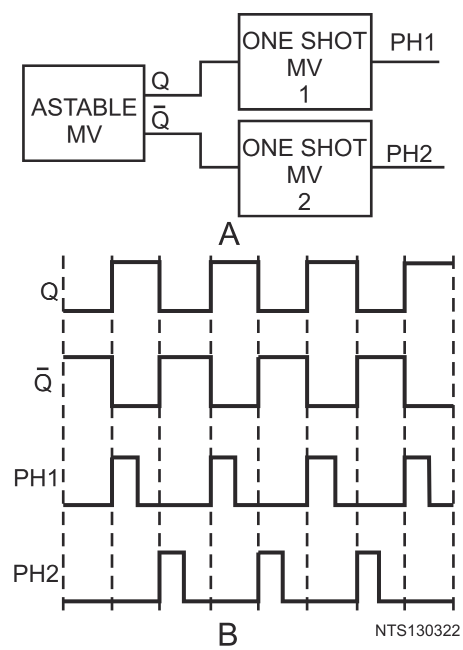


Figure 3-22. —Two-phase clock: A. Block diagram; B. Timing diagram.

Clocks are designed to provide the most efficient operation of the equipment. During the design phase, the frequency, pulse width, and the number of phases required is determined; and the clock circuit is built to meet those requirements.

Most modern high-speed equipment uses crystal-controlled oscillators as the basis for their timing networks. Crystals are stable even at extremely high frequencies.

Q33. What is a clock with regard to digital equipment?

Q34. What is the simplest type of clock circuit?

Q35. What is needed to use a monostable or one-shot multivibrator for a clock circuit?

Q36. What type of clock is used when more than one operation is to be completed during one clock cycle?

COUNTERS

A counter is simply a device that counts. Counters may be used to count operations, quantities, or periods of time. They may also be used for dividing frequencies, for addressing information in storage, or for temporary storage.

Counters are a series of FFs wired together to perform the type of counting desired. They will count up or down by ones, twos, or more.

The total number of counts or stable states a counter can indicate is called MODULUS. For instance, the modulus of a four-stage counter would be 16_{10}, since it is capable of indicating 0000_2 to 1111_2. The term *modulo* is used to describe the count capability of counters; that is, modulo-16 for a four-stage binary counter, modulo-11 for a decade counter, modulo-8 for a three-stage binary counter, and so forth.

Ripple Counters

Ripple counters are so named because the count is like a chain reaction that ripples through the counter because of the time involved. This effect will become more evident with the explanation of the following circuit.

Figure 3-23, view A, shows a basic four-stage, or modulo-16, ripple counter. The inputs and outputs are shown in view B. The four J-K FFs are connected to perform a toggle function; which, you will recall, divides the input by 2. The HIGHs on the J and K inputs enable the FFs to toggle. The inverters on the clock inputs indicate that the FFs change state on the negative-going pulse.

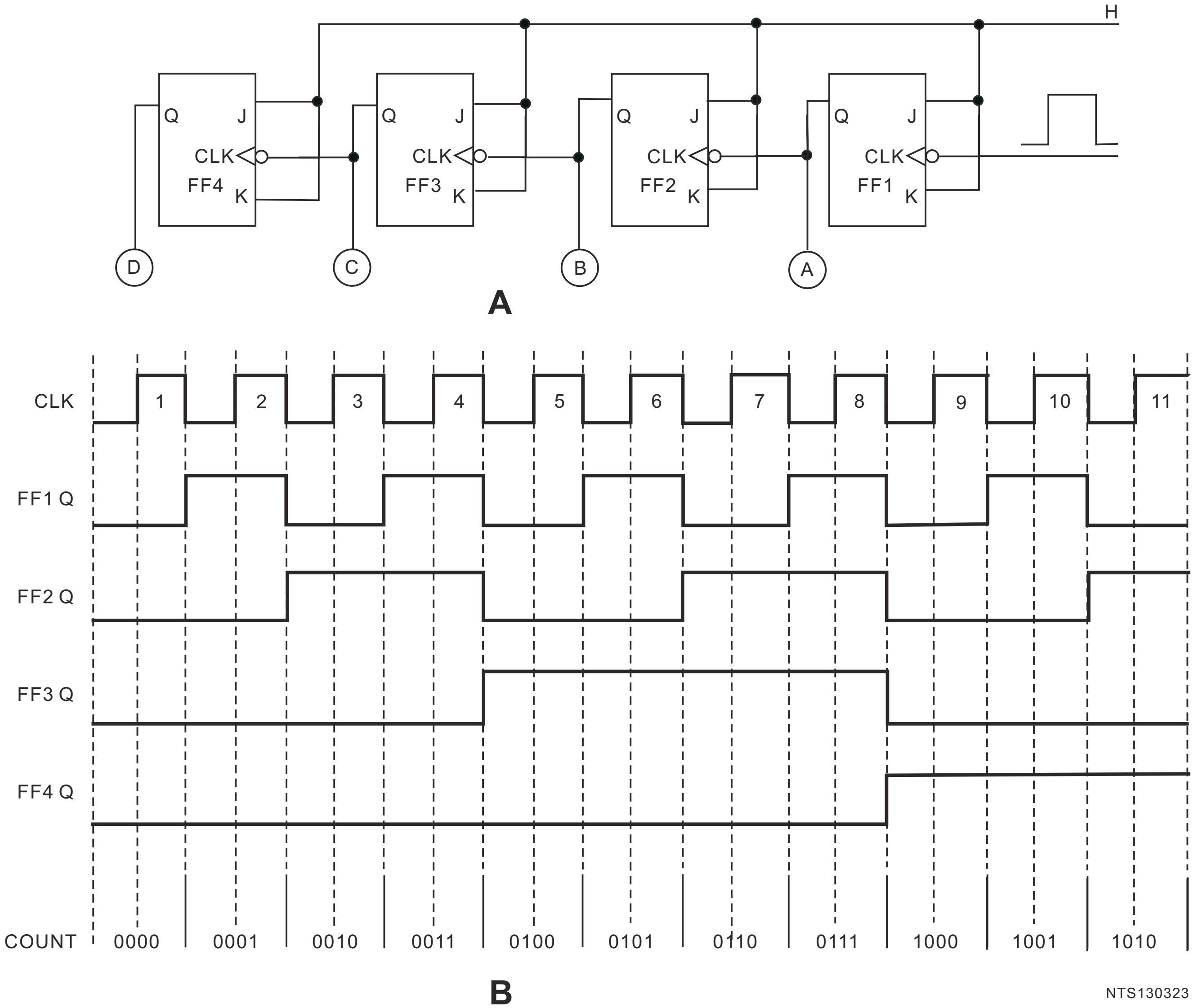


Figure 3-23. —Four-stage ripple counter: A. Logic diagram; B. Timing diagram.

Assume that A, B, C, and D are lamps and that all the FFs are reset. The lamps will all be out, and the count indicated will be 0000_2. The negative-going pulse of clock pulse 1 causes FF1 to set. This lights lamp A, and we have a count of 0001_2. The negative-going pulse of clock pulse 2 toggles FF1, causing it to reset. This negative-going input to FF2 causes it to set and causes B to light. The count after two clock pulses is 0010_2, or 2_{10}. Clock pulse 3 causes FF1 to set and lights lamp A. The setting of FF1 does not affect FF2, and lamp B stays lit. After three clock pulses, the indicated count is 0011_2.

Clock pulse 4 causes FF1 to reset, which causes FF2 to reset, which causes FF3 to set, giving us a count of 0100_2. This step shows the ripple effect.

This setting and resetting of the FFs will continue until all the FFs are set and all the lamps are lit. At that time the count will be 1111_2 or 15_{10}. Clock pulse 16 will cause FF1 to reset and lamp A to go out. This will cause FF2 through FF4 to reset, in order, and will extinguish lamps B, C, and D. The counter would then start at 0001_2 on clock pulse 17. To display a count of 16_{10} or 10000_2, we would need to add another FF.

The ripple counter is also called an ASYNCHRONOUS counter. Asynchronous means that the events (setting and resetting of FFs) occur one after the other rather than all at once. Because the ripple count is asynchronous, it can produce erroneous indications when the clock speed is high. A high-speed clock can cause the lower stage FFs to change state before the upper stages have reacted to the previous clock pulse. The errors are produced by the FFs' inability to keep up with the clock.

Synchronous Counter

High-frequency operations require that all the FFs of a counter be triggered at the same time to prevent errors. We use a SYNCHRONOUS counter for this type of operation.

The synchronous counter is similar to a ripple counter with two exceptions: The clock pulses are applied to each FF, and additional gates are added to ensure that the FFs toggle in the proper sequence.

A logic diagram of a three-state (modulo-8) synchronous counter is shown in figure 3-24, view A. The clock input is wired to each of the FFs to prevent possible errors in the count. A HIGH is wired to the J and K inputs of FF1 to make the FF toggle. The output of FF1 is wired to the J and K inputs of FF2, one input of the AND gate, and indicator A. The output of FF2 is wired to the other input of the AND gate and indicator B. The AND output is connected to the J and K inputs of FF3. The C indicator is the only output of FF3.

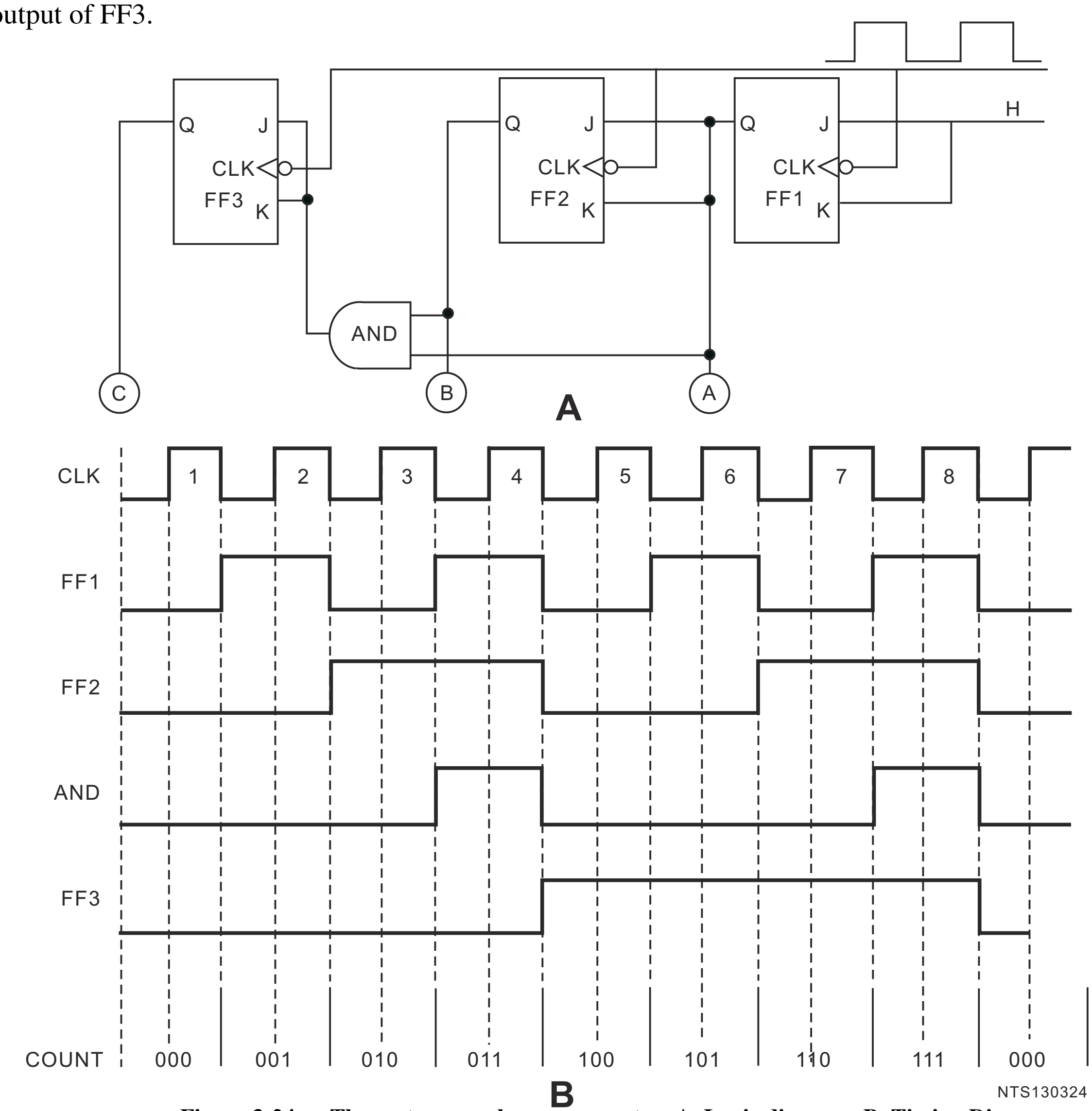


Figure 3-24. —Three-stage synchronous counter: A. Logic diagram; B. Timing Diagram.

During the explanation of this circuit, you should follow the logic diagram, view A, and the pulse sequences, view B.

Assume the following initial conditions: The outputs of all FFs, the clock, and the AND gate are 0; the J and K inputs to FF1 are HIGH. The negative-going portion of the clock pulse will be used throughout the explanation.

Clock pulse 1 causes FF1 to set. This HIGH lights lamp A, indicating a binary count of 001. The HIGH is also applied to the J and K inputs of FF2 and one input of the AND gate. Notice that FF2 and

FF3 are unaffected by the first clock pulse because the J and K inputs were LOW when the clock pulse was applied.

As clock pulse 2 goes LOW, FF1 resets, turning off lamp A. In turn, FF2 will set, lighting lamp B and showing a count of 010_2. The HIGH from FF2 is also felt by the AND gate. The AND gate is not activated at this time because the signal from FF1 is now a LOW. A LOW is present on the J and K inputs of FF3, so it is not toggled by the clock.

Clock pulse 3 toggles FF1 again and lights lamp A. Since the J and K inputs to FF2 were LOW when pulse 3 occurred, FF2 does not toggle but remains set. Lamps A and B are lit, indicating a count of 011_2. With both FF1 and FF2 set, HIGHs are input to both inputs of the AND gate, resulting in HIGHs to J and K of FF3. No change occurred in the output of FF3 on clock pulse 3 because the J and K inputs were LOW at the time.

Just before clock pulse 4 occurs, we have the following conditions: FF1 and FF2 are set, and the AND gate is outputting a HIGH to the J and K inputs of FF3. With these conditions all of the FFs will toggle with the next clock pulse.

At clock pulse 4, FF1 and FF2 are reset, and FF3 sets. The output of the AND gate goes to 0, and we have a count of 100_2.

It appears that the clock pulse and the AND output both go to 0 at the same time, but the clock pulse arrives at FF3 before the AND gate goes LOW because of the transit time of the signal through FF1, FF2, and the AND gate.

Between pulses 4 and 8, FF3 remains set because the J and K inputs are LOW. FF1 and FF2 toggle in the same sequence as they did on clock pulses 1, 2, and 3.

Clock pulse 7 results in all of the FFs being set and the AND gate output being HIGH. Clock pulse 8 causes all the FFs to reset and all the lamps to turn off, indicating a count of 000_2 . The next clock pulse (9) will restart the count sequence.

Q37. What is the modulus of a five-stage binary counter?

Q38. An asynchronous counter is also called a ___________ counter.

Q39. J-K FFs used in counters are wired to perform what function?

Q40. What type of counter has clock pulses applied to all FFs?

Q41. In figure 3-24, view A, what logic element enables FF3 to toggle with the clock?

Q42. What is the largest count that can be indicated by a four-stage counter?

Decade Counter

A decade counter is a binary counter that is designed to count to 10_{10}, or 1010_2. An ordinary four-stage counter can be easily modified to a decade counter by adding a NAND gate as shown in figure 3-25. Notice that FF2 and FF4 provide the inputs to the NAND gate. The NAND gate outputs are connected to the CLR input of each of the FFs.

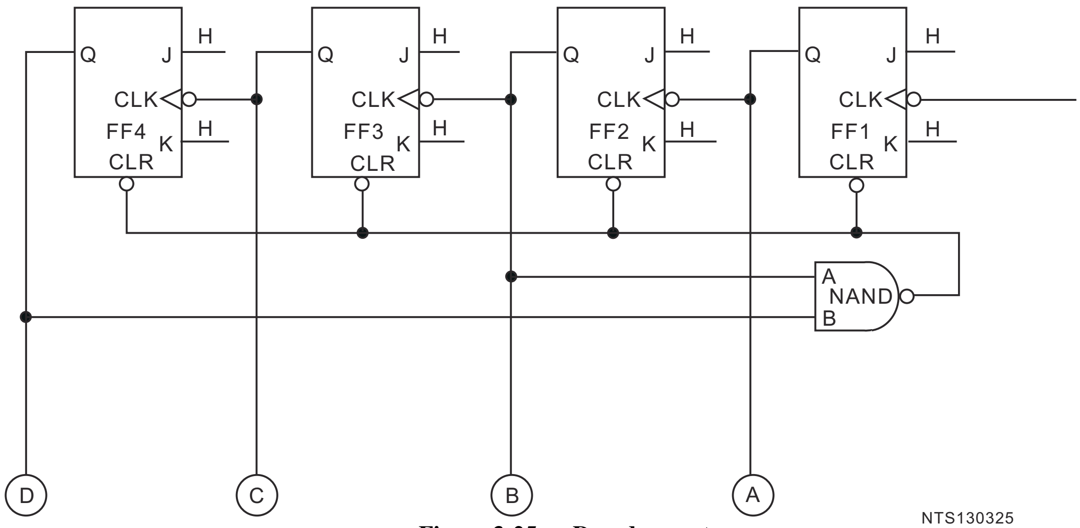


Figure 3-25. —Decade counter.

The counter operates as a normal counter until it reaches a count of 1010_2, or 10_{10}. At that time, both inputs to the NAND gate are HIGH, and the output goes LOW. This LOW applied to the CLR input of the FFs causes them to reset to 0. Remember from the discussion of J-K FFs that CLR and PS or PR override any existing condition of the FF. Once the FFs are reset, the count may begin again. The following table shows the binary count and the inputs and outputs of the NAND gate for each count of the decade counter:

BINARY COUNT	NAND GATE INPUTS		NAND GATE OUTPUT
*******	A	B	*******
0000	0	0	1
0001	0	0	1
0010	1	0	1
0011	1	0	1
0100	0	0	1
0101	0	0	1
0110	1	0	1
0111	1	0	1
1000	0	1	1
1001	0	1	1
1010	1	1	0

Changing the inputs to the NAND gate can cause the maximum count to be changed. For instance, if FF4 and FF3 were wired to the NAND gate, the counter would count to 1100_2 (12_{10}), and then reset.

Q43. How many stages are required for a decade counter?

Q44. In figure 3-25, which two FFs must be HIGH to reset the counter?

Ring Counter

A ring counter is defined as a loop of bistable devices (flip-flops) interconnected in such a manner that only one of the devices may be in a specified state at one time. If the specified condition is HIGH, then only one device may be HIGH at one time. As the clock, or input, signal is received, the specified state will shift to the next device at a rate of 1 shift per clock, or input, pulse.

Figure 3-26, view A, shows a typical four-stage ring counter. This particular counter is composed of R-S FFs. J-K FFs may be used as well. Notice that the output of each AND gate is input to the R, or reset side, of the nearest FF and to the S, or set side, of the next FF. The Q output of each FF is applied to the B input of the AND gate that is connected to its own R input.

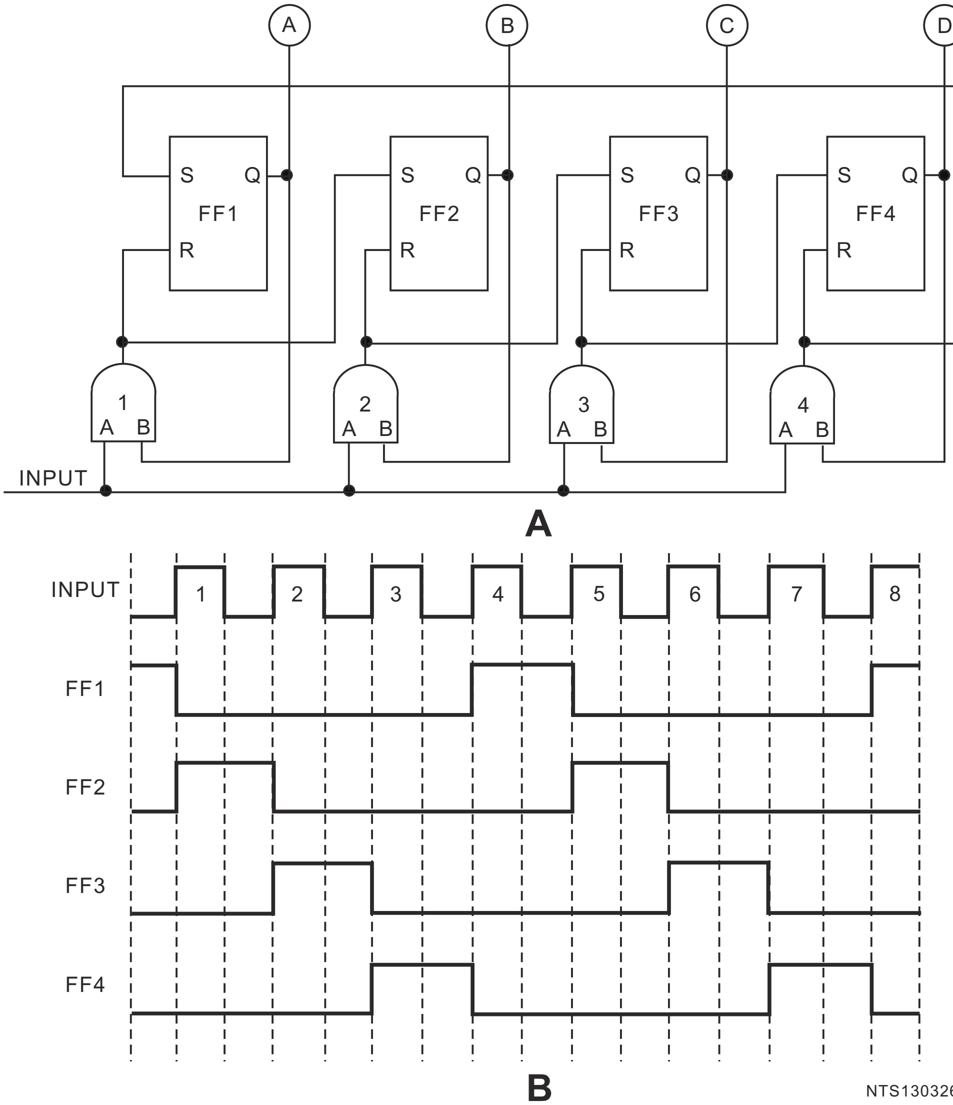


Figure 3-26. —Ring counter: A. Logic diagram; B. Timing diagram.

The circuit input may be normal CLK pulses or pulses from elsewhere in the equipment that would indicate some operation has been completed.

Now, let's look at the circuit operation and observe the signal flow as shown in figure 3-26, view B.

For an initial condition, let's assume that the output of FF1 is HIGH and that the input and FF2, FF3, and FF4 are LOW. Under these conditions, lamp A will be lit; and lamps B, C, and D will be extinguished. The HIGH from FF1 is also applied to the B input of AND gate 1.

The first input pulse is applied to the A input of each of the AND gates. The B inputs to AND gates 2, 3, and 4 are LOW since the outputs of FF2, FF3, and FF4 are LOW. AND gate 1 now has HIGHs on both inputs and produces a HIGH output. This HIGH simultaneously resets FF1 and sets FF2. Lamp A then goes out, and lamp B goes on. We now have a HIGH on AND gate 2 at the B input. We also have a LOW on AND gate 1 at input B.

Input pulse 2 will produce a HIGH output from AND gate 2 since AND gate 2 is the only one with HIGHs on both inputs. The HIGH from AND gate 2 causes FF2 to reset and FF3 to set. Indicator B goes out and C goes on.

Pulse 3 will cause AND gate 3 to go HIGH. This results in FF3 being reset and FF4 being set. Pulse 4 causes FF4 to reset and FF1 to set, bringing the counter full circle to the initial conditions. As long as the counter is operational, it will continue to light the lamps in sequence — 1, 2, 3, 4; 1, 2, 3, 4, etc.

As we stated at the beginning of this section, only one FF may be in the specified condition at one time. The specified condition shifts one position with each input pulse.

Q45. In figure 3-26, view A, which AND gate causes FF3 to set?

Q46. Which AND gate causes FF3 to reset?

Q47. What causes the specified condition to shift position?

Q48. If the specified state is OFF, how many FFs may be off at one time?

Down Counters

Up to this point the counters that you have learned about have been up counters (with the exception of the ring counter). An up counter starts at 0 and counts to a given number. This section will discuss DOWN counters, which start at a given number and count down to 0.

Up counters are sometimes called INCREMENT counters. Increment means to increase. Down counters are called DECREMENT counters. Decrement means to decrease.

A three-stage, ripple down counter is shown in figure 3-27, view A. Notice that the PS (preset) input of the J-K FFs is used in this circuit. HIGHs are applied to all the J and K inputs. This enables the FFs to toggle on the input pulses.

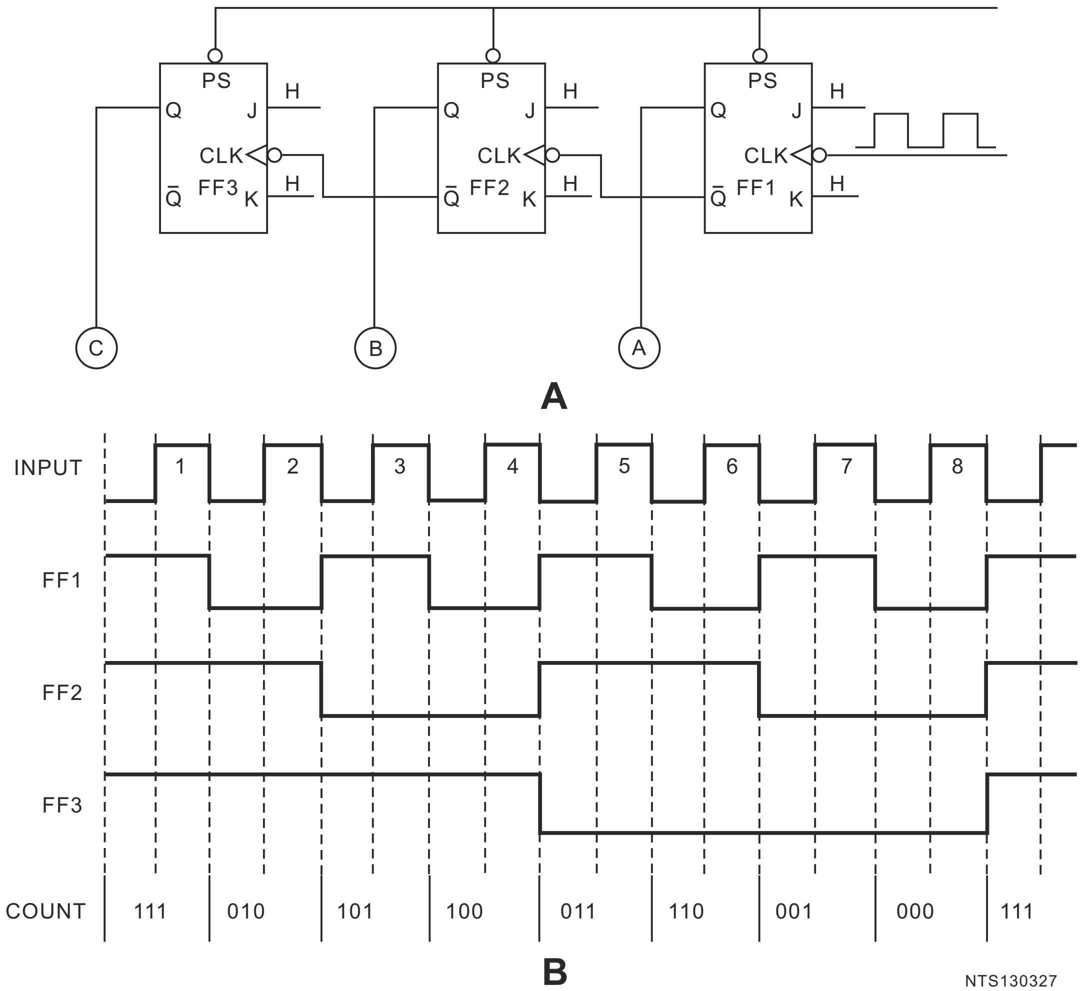


Figure 3-27. —Down counter: A. Logic diagram; B. Timing diagram.

A negative-going pulse is applied to all PS terminals to start the countdown. This causes all the FFs to set and also lights indicators A, B, and C. The beginning count is 111_2 (7_{10}). At the same time, LOWs are applied to the CLK inputs of FF2 and FF3, but they do not toggle because the PS overrides any change. All actions in the counter will take place on the negative-going portion of the input pulse. Let's go through the pulse sequences in figure 3-27, view B.

CP1 causes FF1 to toggle and output Q to go LOW. Lamp A is turned off. Notice that $\overline{Q}$ goes HIGH but no change occurs in FF2 or FF3. Lamps B and C are now on, A is off, and the indicated count is 110_2 (6_{10}).

CP2 toggles FF1 again and lights lamp A. When Q goes HIGH, $\overline{Q}$ goes LOW. This negative-going signal causes FF2 to toggle and reset. Lamp B is turned off, and a HIGH is felt at the CLK input of FF3. The indicated count is 101_2 (5_{10}); lamps A and C are on, and B is off.

At CP3, FF1 toggles and resets. Lamp A is turned off. A positive-going signal is applied to the CLK input of FF2. Lamp B remains off and C remains on. The count at this point is 100_2 (4_{10}).

CP4 toggles FF1 and causes it to set, lighting lamp A. Now FF1, output $\overline{Q}$, goes LOW causing FF2 to toggle. This causes FF2 to set and lights lamp B. Output of FF2, $\overline{Q}$, then goes LOW, which causes FF3 to reset and turn off lamp C. The indicated count is now 011_2 (3_{10}).

The next pulse, CP5, turns off lamp A but leaves B on. The count is now 010_2. CP6 turns on lamp A and turns off lamp B, for a count of 001_2. CP7 turns off lamp A. Now all the lamps are off, and the counter indicates 000.

On the negative-going signal of CP8, all FFs are set, and all the lamps are lighted. The CLK pulse toggles FF1, making output Q go HIGH. As output $\overline{Q}$ goes LOW, the negative-going signal causes FF2 to toggle. As FF2, output Q, goes HIGH, output $\overline{Q}$ goes LOW, causing FF3 to toggle and set. As each FF sets, its indicator lamp lights. The counter is now ready to again start counting down from 111_2 with the next CLK pulse.

Q49. How many FFs are required to count down from 15_{10}?

Q50. What signal causes FF2 to toggle?

REGISTERS

A register is a temporary storage device. Registers are used to store data, memory addresses, and operation codes. Registers are normally referred to by the number of stages they contain or by the number of bits they will store. For instance, an eight-stage register would be called an 8-bit register. The contents of the register is also called a WORD. The contents of an 8-bit register is an 8-bit word. The contents of a 4-bit register is a 4-bit word and so forth.

Registers are also used in the transfer of data to and from input and output devices such as teletypes, printers, and cathode-ray tubes.

Most registers are constructed of FFs and associated circuitry. They permit us to load or store data and to transfer the data at the proper time.

PARALLEL REGISTERS

Parallel registers are designed to receive or transfer all bits of data or information simultaneously.

A 4-bit parallel register is shown in figure 3-28. The data inputs are A, B, C, and D. The FFs store the data until it is needed. AND gates 5, 6, 7, and 8 are the transfer gates.

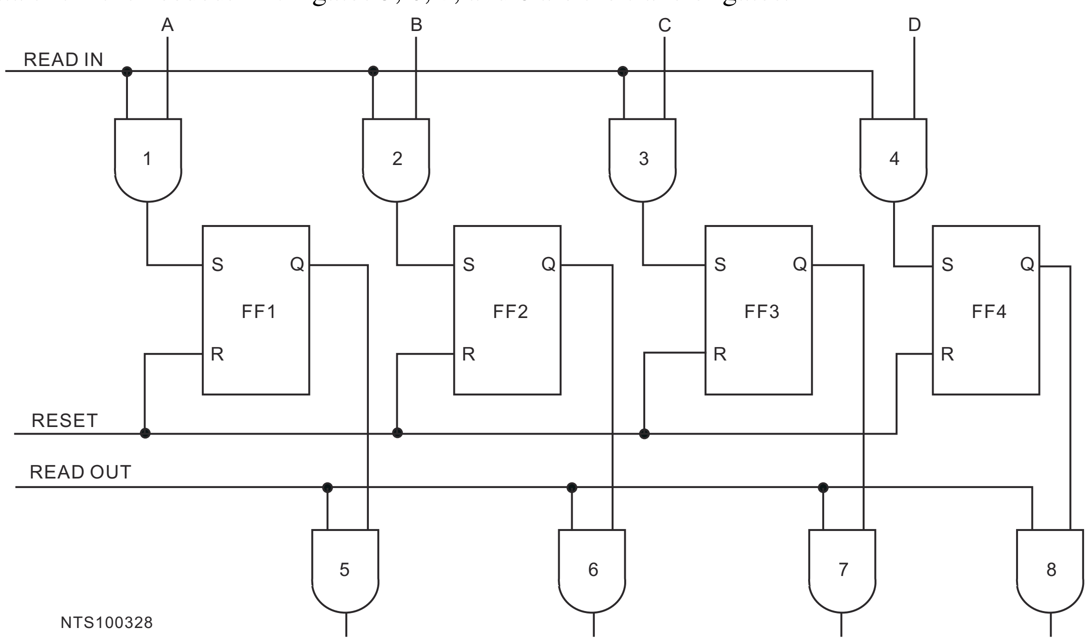


Figure 3-28. —Four-bit parallel register.

Before we go through the operation of the register, let's set some initial conditions. Assume that inputs A, B, and D are HIGH and that FF2 and FF4 are set from a previous operation. The READ IN, READ OUT, and RESET inputs are all LOW.

To begin the operation, we apply a reset pulse to the RESET input of all the FFs, clearing the Q outputs to LOWs. This step ensures against any erroneous data transfer that would occur because of the states of FF2 and FF4.

Inputs A, B, C, and D are input to gates 1, 2, 3, and 4, respectively. When the READ IN input goes HIGH, AND gates 1, 2, and 4 go HIGH, causing FF1, FF2, and FF4 to set. The output of AND gate 3 does not change since the C input is LOW. The 4-bit word, 1101, is now stored in the register. The outputs of FFs 1, 2, 3, and 4 are applied to AND gates 5, 6, 7, and 8, respectively.

When the data is required for some other operation, a positive-going pulse is applied to the READ OUT inputs of the AND gates. This HIGH, along with the HIGHs from the FFs, causes the outputs of AND gates 5, 6, and 8 to go HIGH. Since the Q output of FF3 is LOW, the output of gate 7 will be LOW. The 4-bit word, 1101, is transferred to where it is needed.

Q51. How many stages are required to store a 16-bit word?

Q52. Simultaneous transfer of data may be accomplished with what type of register?

Q53. How are erroneous transfers of data prevented?

SHIFT REGISTERS

A shift register is a register in which the contents may be shifted one or more places to the left or right. This type of register is capable of performing a variety of functions. It may be used for serial-to-parallel conversion and for scaling binary numbers.

Before we get into the operation of the shift register, let's discuss serial-to-parallel conversion, parallel-to-serial conversion, and scaling.

Serial and Parallel Transfers and Conversion

Serial and parallel are terms used to describe the method in which data or information is moved from one place to another. **SERIAL TRANSFER** means that the data is moved along a single line one bit at a time. A control pulse is required to move each bit. **PARALLEL TRANSFER** means that each bit of data is moved on its own line and that all bits transfer simultaneously as they did in the parallel register. A single control pulse is required to move all bits.

Figure 3-29 shows how both of these transfers occur. In each case, the four-bit word 1101 is being transferred to a storage device. In view A, the data moves along a single line. Each bit of the data will be stored by an individual control pulse. In view B, each bit has a separate input line. One control pulse will cause the entire word to be stored.

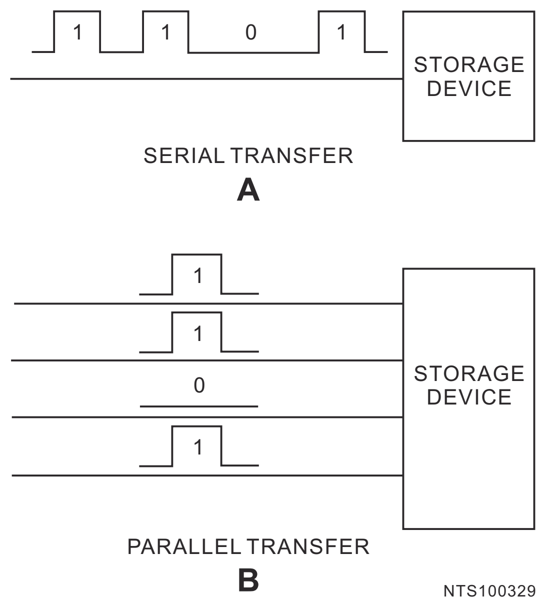


Figure 3-29. —Data transfer methods: A. Serial transfer; B. Parallel transfer.

Serial-to-parallel conversion or parallel-to-serial conversion describes the manner in which data is stored in a storage device and the manner in which that data is removed from the storage device.

Serial-to-parallel conversion means that data is transferred into the storage device or register in serial fashion and removed in parallel fashion, as in figure 3-30, view A. Parallel-to-serial conversion means the data is transferred into the storage device in parallel and removed as serial data, as shown in view B.

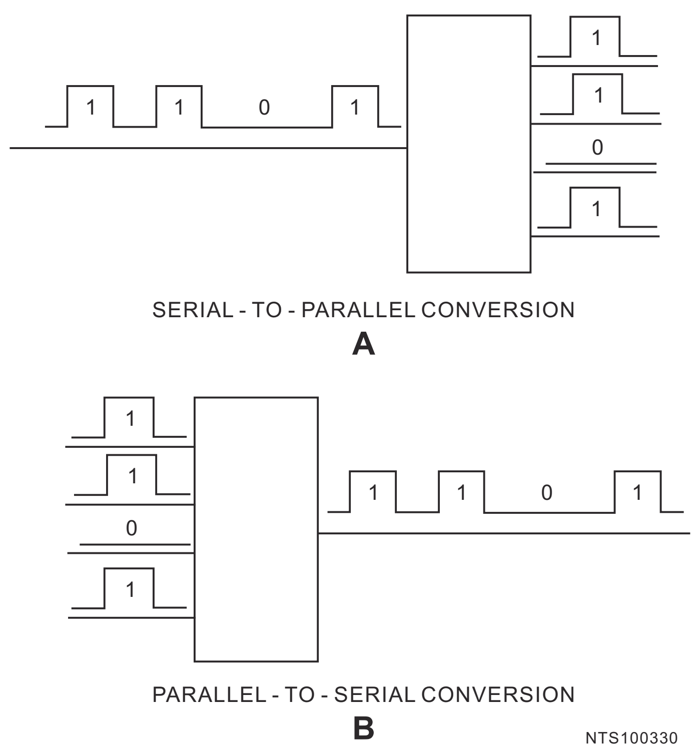


Figure 3-30. —Data conversion methods: A. Serial-to-parallel; B. Parallel-to-serial.

Serial transfer takes time. The longer the word length, the longer the transfer will take. Although parallel transfer is much faster, it requires more circuitry to transfer the data.

Scaling

SCALING means to change the magnitude of a number. Shifting binary numbers to the left increases their value, and shifting to the right decreases their value. The increase or decrease in value is based on powers of 2.

A shift of one place to the left increases the value by a power of 2, which in effect is multiplying the number by 2. To demonstrate this, let's assume that the following block diagram is a 5-bit shift register containing the binary number 01100.

0	1	1	0	0

Shifting the entire number one place to the left will put the register in the following condition:

1	1	0	0	0

The binary number 01100 has a decimal equivalent of 12. If we convert 11000_2 to decimal, we find it has a value of 24_{10}. By shifting the number one place to the left, we have multiplied it by 2. A shift of two places to the left would be the equivalent of multiplying the number by 2^2, or 4; three places by 2^3, or 8; and so forth.

Shifting a binary number to the right decreases the value of the number by a power of 2 for each place. Let's look at the same 5-bit register containing 01100_2 and shift the number to the right.

0	1	1	0	0

A shift of one place to the right will result in the register being in the following condition:

0	0	1	1	0

By comparing decimal equivalents you can see that we have decreased the value from 12_{10} to 6_{10}. We have effectively divided the number by 2. A shift of two places to the right is the equivalent of dividing the number by 2^2, or 4; three places by 2^3; or 8; and so forth.

Shift Register Operations

Figure 3-31shows a typical 4-bit shift register. This particular register is capable of left shifts only. There are provisions for serial and parallel input and serial and parallel output. Additional circuitry would be required to make right shifts possible.

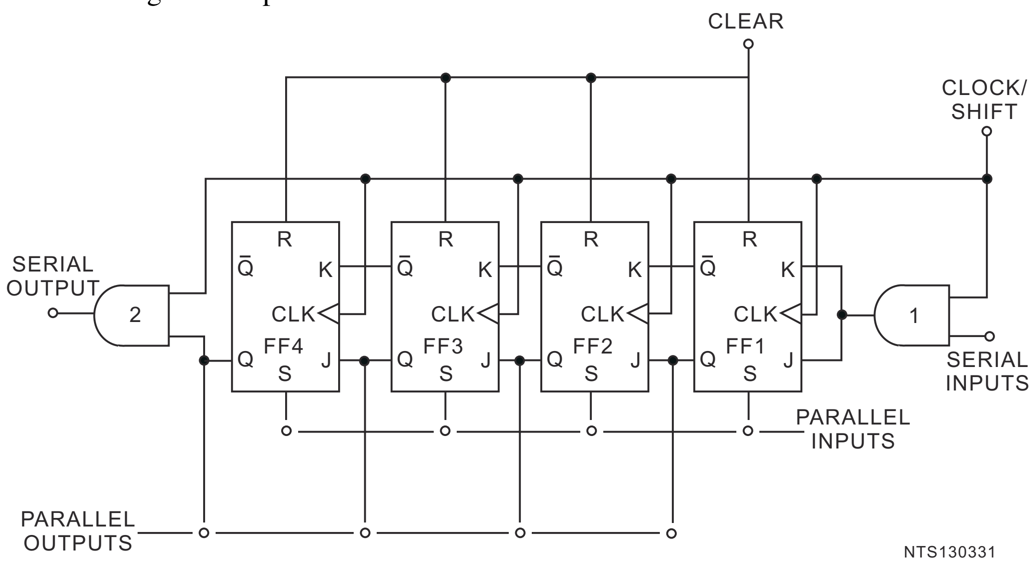


Figure 3-31. —Shift register.

Before any operation takes place, a CLEAR pulse is applied to the RESET terminal of each FF to ensure that the Q output is LOW.

The simplest modes of operation for this register are the parallel inputs and outputs. Parallel data is applied to the SET inputs of the FFs and results in either a 1 or 0 output, depending on the input. The outputs of the FFs may be sampled for parallel output. In this mode, the register functions just like the parallel register covered earlier in this section.

Parallel-to-Serial Conversion

Now let's look at parallel-to-serial conversion. We will use the 4-bit shift register in figure 3-31 and the timing sequence in figure 3-32 to aid you in understanding the operations.

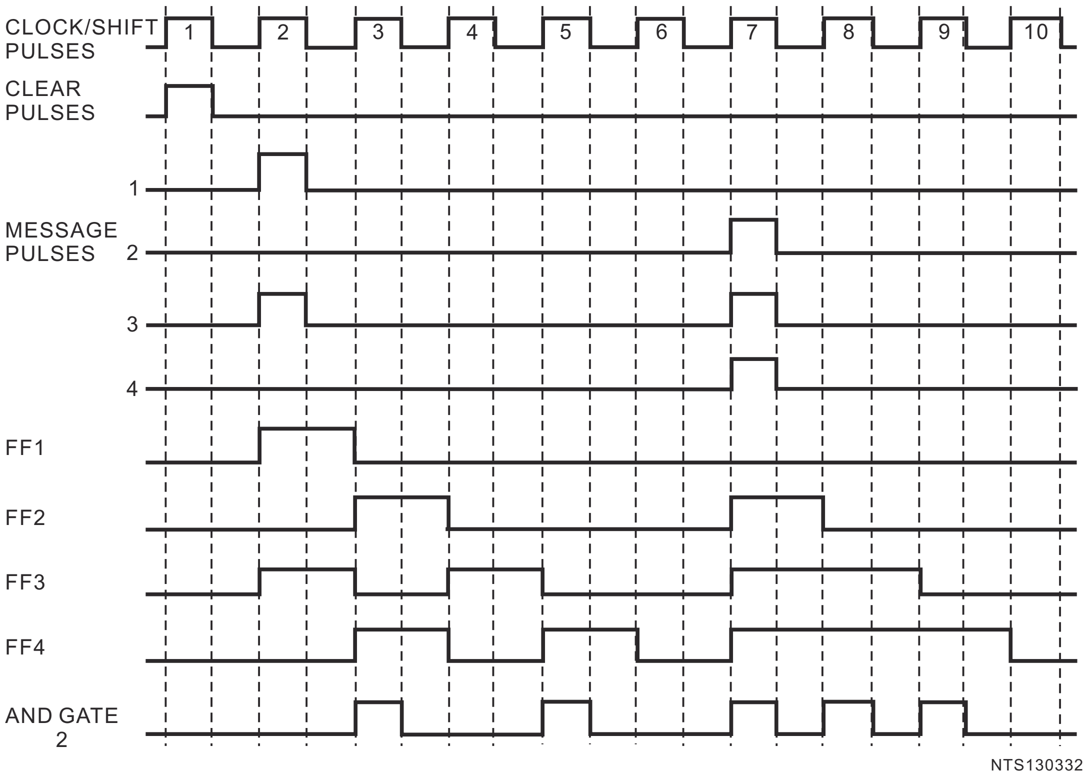


Figure 3-32. —Parallel-to-serial conversion timing diagram.

At CP1, a CLEAR pulse is applied to all the FFs, resetting the register to a count of 0. The number 0101_2 is applied to the parallel inputs at CP2, causing FF1 and FF3 to set. At this point, the J inputs of FF2 and FF4 are HIGH. AND gate 2 has a LOW output since the FF4 output is LOW. This LOW output represents the first digit of the number 0101_2 to be output in serial form. At the same time we have HIGHs on the K inputs of FF1 and FF3. (Notice the NOT symbol on FF1 at input K. With no serial input to AND gate 1, the output is LOW; therefore, the K input to FF1 is held HIGH). With these conditions CP3 causes FF1 and FF3 to reset and FF2 and FF4 to set. The HIGH output of FF4, along with CP3, causes AND gate 2 to output a HIGH. This represents the second digit of the number 0101_2.

At CP4, FF2 and FF4 reset, and FF3 sets. FF1 remains reset because of the HIGH at the K input. The output of AND gate 2 goes LOW because the output of FF4 is LOW and the third digit of the number is output on the serial line. CP5 causes FF4 to set and FF3 to reset. CP5 and the HIGH from FF4 cause AND gate 2 to output the last digit of the number on the serial line. It took a total of four CLK pulses to input the number in parallel and output it in serial.

CP6 causes FF4 to reset and effectively clears the register for the next parallel input.

Between CP7 and CP10, the number 1110_2 is input as parallel data and output as serial data.

Serial-to-Parallel Conversion

Serial input is accomplished much in the same manner as serial output. Instead of shifting the data out one bit at a time, we shift the data in one bit at a time.

To understand this conversion, you should again use figure 3-31and also the timing diagram shown in figure 3-33. In this example we will convert the number 1011_2 from serial data to parallel data.

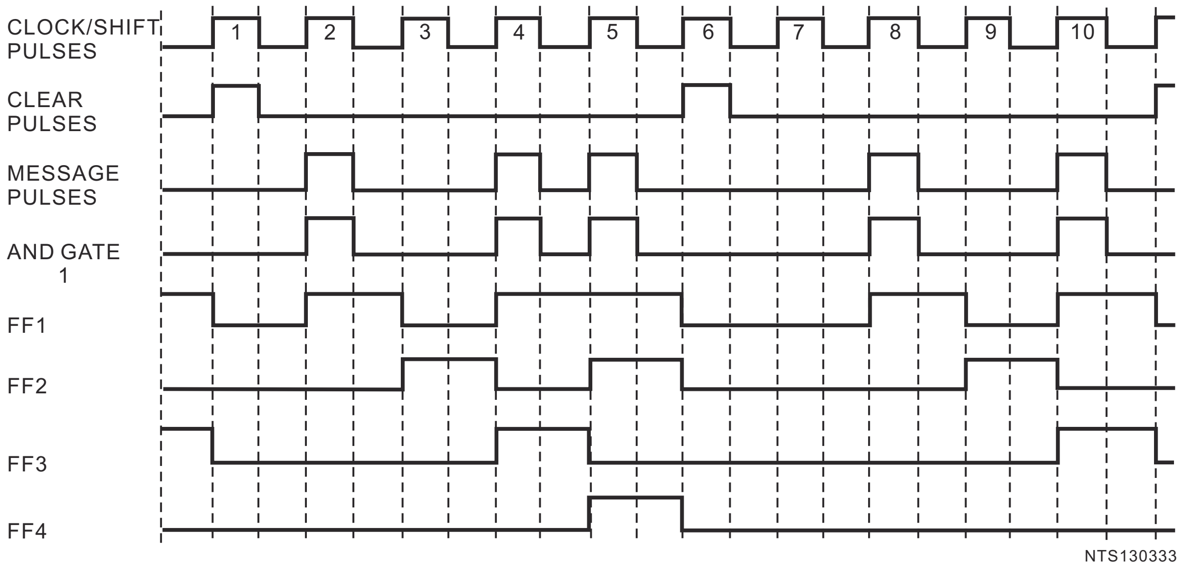


Figure 3-33. —Serial-to-parallel conversion timing diagram.

A CLEAR pulse resets all the FFs at CP1. At CP2, the most significant bit of the data is input to AND gate 1. This HIGH along with the clock pulse causes AND gate 1 to output a HIGH. The HIGH from the AND gate and the clock pulse applied to FF1 cause the FF to set. FFs 2, 3, and 4 are held reset. At this point, the MSD of the data has been shifted into the register.

The next bit of data is a 0. The output of AND gate 1 is LOW. Because of the inverter on the K input of FF1, the FF senses a HIGH at that input and resets. At the same time this is occurring, the HIGH on the J input of FF2 (from FF1) and the CLK cause FF2 to set. The two MSDs, 1 and 0, are now in the register.

CP4 causes FF3 to set and FF2 to reset. FF1 is set by the CLK pulse and the third bit of the number. The register now contains 0101_2, as a result of shifting the first three bits of data.

The remaining bit is shifted into the register by CP5. FF1 remains set, FF2 sets, FF3 resets, and FF4 sets. At this point, the serial transfer is complete. The binary word can be sampled on the parallel output lines. Once the parallel data is transferred, a CLEAR pulse resets the FFs (CP6), and the register is ready to input the next word.

Scaling Operation

Using the shift register shown in figure 3-31 for scaling a number is quite simple. The number to be scaled is loaded into the register either in serial or parallel form. Once the data is in the register, the scaling takes place in the same manner as that for shifting the data for serial output. A single clock pulse will cause each bit of data to shift one place to the left. Remember that each shift is the equivalent of increasing the value by a power of 2. The scaled data is read from the parallel outputs. Care must be taken not to over shift the data to the point that the MSDs are shifted out of the register.

Q54. Serial-to-parallel and parallel-to-serial conversions are accomplished by what type of circuit?

Q55. What type of data transfer requires the most time?

Q56. What is the main disadvantage of parallel transfer?

Q57. How many FFs would be required for an 8-bit shift register?

Q58. How many clock pulses are required to output a 4-bit number in serial form?

Q59. Two shifts to the left are equal to increasing the magnitude of a number by how much?

Q60. To increase the magnitude of a number by 2^3, you must shift the number how many times and in what direction?

LOGIC FAMILIES

Logic families are groups of logic circuits that are based on particular types of elements (resistors, transistors, and so forth). Families are identified by the manner in which the elements are connected, and, in some cases, by the types of elements used.

Logic circuits of a particular family can be interconnected without having to use additional circuitry. In other words, the output of one logic circuit can be used as the input to another logic circuit. This feature is known as compatibility. All circuits within a logic family will be compatible with the other circuits within that family.

As a technician, your responsibility will be to identify defective parts and repair or replace them as required. It will be beneficial for you to have a basic knowledge of the types of logic that are used in digital equipment.

Logic circuits are usually manufactured as integrated circuits and packaged in dual-inline packages (DIP), modified transistor outlines (TO), or flat packs. These packaging techniques are described in NEETS, Module 7, *Introduction to Solid-State Devices and Power Supplies.*

Circuitry in a package is normally shown using standard logic symbols instead of individual components such as transistors, diodes, and so forth. Figure 3-34 shows four examples of this type of packaging. The numbered blocks (1-14 and 1-16) are the pins on the package. Circuit packages are also identified by a manufacturer's part number. Similar circuits produced by different manufacturers will not carry the same identification numbers in all cases.

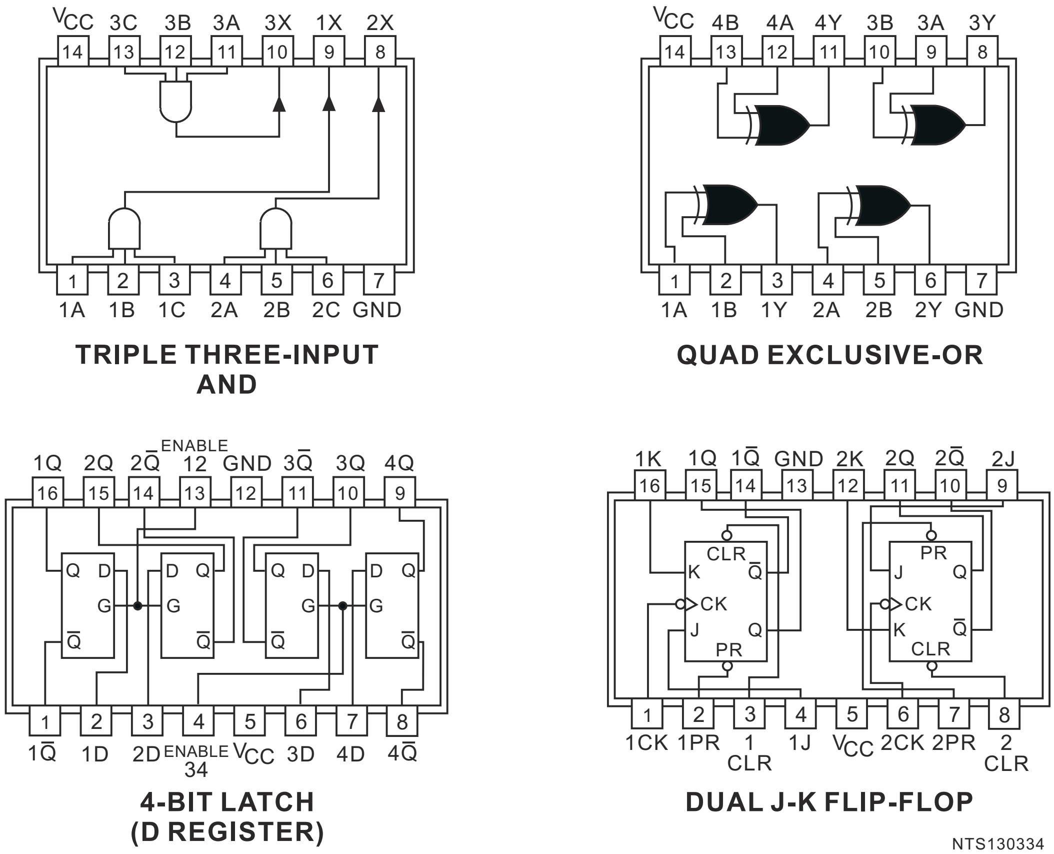


Figure 3-34. —Logic packages.

As mentioned before, logic families are identified by the elements used and the manner in which the elements are used. A brief description of some of the more common logic families follows.

RTL (RESISTOR-TRANSISTOR LOGIC)

In this type of logic, inputs are applied to resistors, and the output is produced by a transistor. RTL is normally constructed from discrete components (individual resistors and transistors). Some circuits are manufactured as integrated circuits and packaged in modified transistor outline (TO) packages, as shown in figure 3-35. An in-depth coverage of circuit packaging can be found in NEETS, Module 14, *Introduction to Microelectronics*.

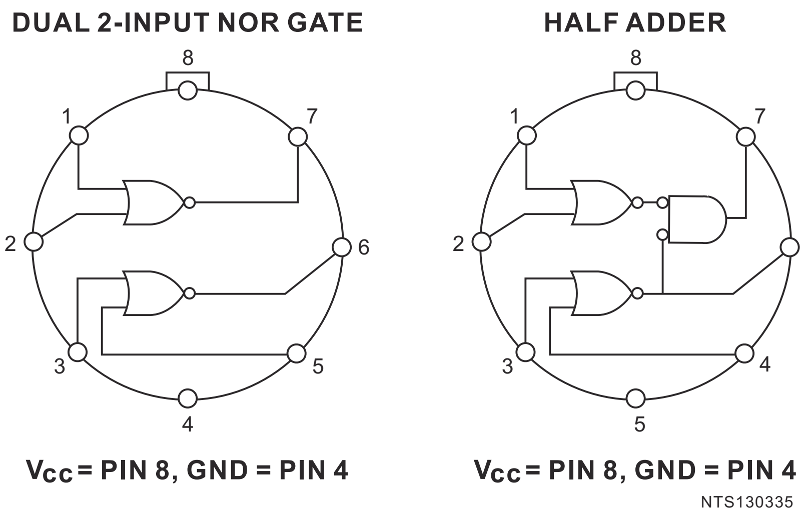


Figure 3-35. —RTL integrated circuits.

DTL (DIODE TRANSISTOR LOGIC)

Input signals are applied to diodes in this logic family. The diodes either conduct or cut off and produce the desired output from the transistor. DTL is normally found in dual-inline packages (DIP) as well as older discrete component logic.

TTL (TRANSISTOR-TRANSISTOR LOGIC)

In TTL, transistors with multiple emitters are used for the logic inputs. Additional transistors are used to produce the desired output. TTL is normally packed in DIPs and is quite common in military equipment.

CMOS (COMPLEMENTARY METAL OXIDE SEMICONDUCTORS)

The CMOS logic circuits use metal oxide semiconductors similar to field-effect transistors (FETs).

LOGIC FAMILY USE

The logic family used in a piece of equipment is determined by the design engineers. The type of logic used will be based on the requirements of the equipment and on what family best fulfills the requirements.

The use of integrated circuits enables designers to produce equipment that is very small and highly efficient when compared to other methods of construction. The block diagram shown in figure 3-36, view A, represents an 8-bit, serial-input and parallel-output shift register. This circuit is contained in a standard 14-pin DIP measuring about 0.75 inch long and 0.25 inch wide. View B shows this circuit package.

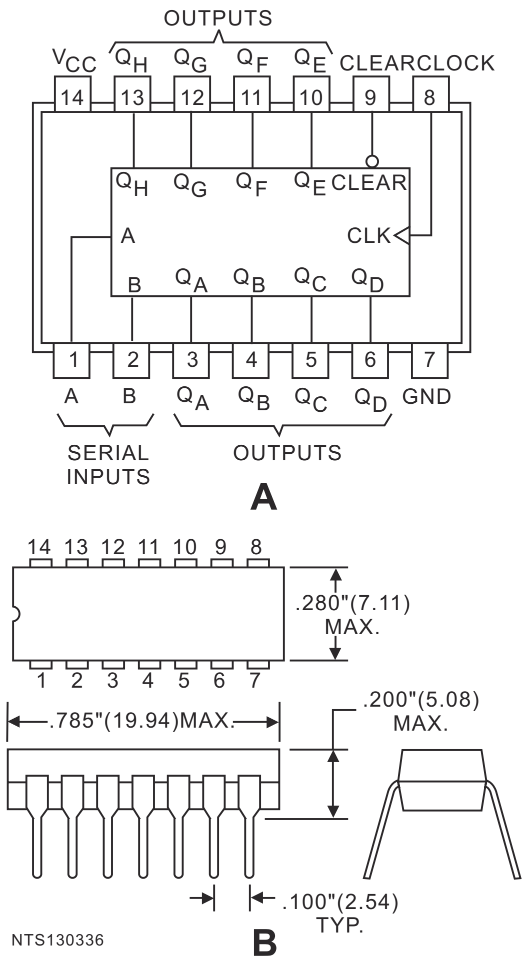


Figure 3-36. —Integrated logic circuits: A. Shift register; B. Logic package.

Q61. What are RTL, DTL, and TTL examples of?

Q62. What type of logic family uses diodes in the input?

Q63. What is the most common type of integrated circuit packaging found in military equipment?

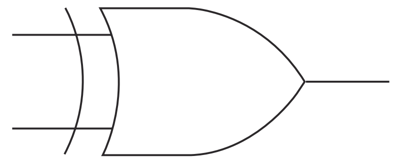
Q64. Circuits that can be interconnected without additional circuitry are known as ____________ circuits.

SUMMARY

Now that you have completed this chapter, you should have a basic understanding of the more common special logic circuits. The following is a summary of the emphasized terms and points found in the "Special Logic Circuits" chapter.

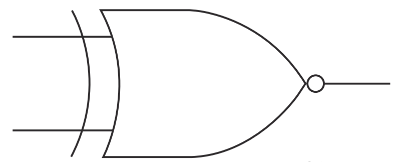
SPECIAL LOGIC CIRCUITS perform arithmetic and logic operations; input, output, store and transfer information; and provide proper timing for these operations.

EXCLUSIVE OR (X-OR) circuits produce a 1 output when ONLY one input is HIGH. Can be used as a quarter adder.



NTS130337

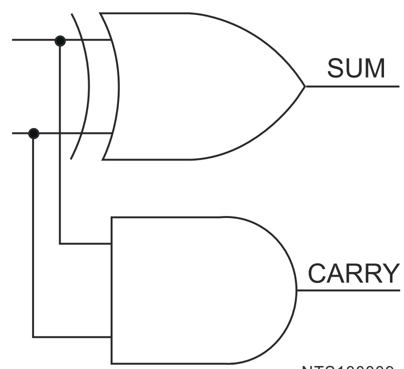
EXCLUSIVE NOR (X-NOR) circuits produce a 1 output when all inputs are 0 and when more than 1 input is 1.



NTS130338

QUARTER ADDER circuits produce the sum of two numbers but do not generate a carry.

HALF ADDER circuits produce the sum of two numbers and generate a carry.



NTS130339

FULL ADDER circuits add a carry to obtain the correct sum.

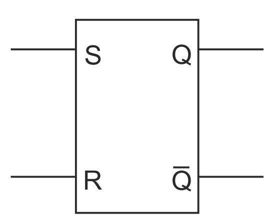
PARALLEL ADDER circuits use full adders connected in parallel to accommodate the addition of multiple-digit numbers.

STANDARD SYMBOLS depict logic circuitry with blocks, showing only inputs and outputs. One block may contain many types of gates and circuits.

SUBTRACTION in binary is accomplished by complementing and adding.

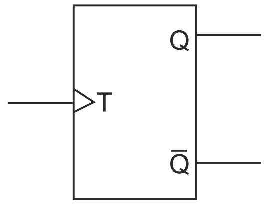
FLIP-FLOP are bistable multivibrators used for storage, timing, arithmetic operations, and transfer of information.

R-S FFs have the Q output of the FF HIGH in the set mode and LOW in the reset mode.



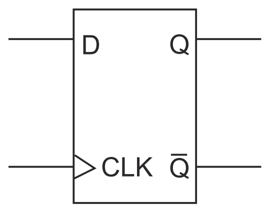
NTS130340A

T (TOGGLE) FF$_S$ change state with each pulse applied to the input. Each T FF will divide the input by 2.



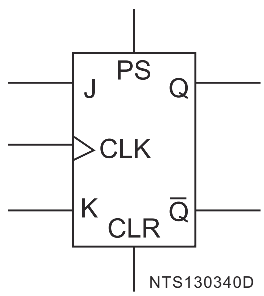
NTS130340B

D FF is used to store data at a predetermined time.



NTS130340C

J-K FF is the most versatile FF. J-Ks can perform the same functions as all the other FFs.



NTS130340D

CLOCKS are circuits that generate the timing and control signals for other operations.

COUNTERS are used to count operations, quantities, or periods of time. They can be used to divide frequencies, to address information in storage, or as temporary storage.

MODULUS of a counter is the total number of counts or stable states a counter may indicate.

UP COUNTERS count from 0 to a predetermined number.

DOWN COUNTERS count from a predetermined number to 0.

RING COUNTERS are loop counters that may be used for timing operations.

REGISTERS are used as temporary storage devices as well as for transfer of information.

PARALLEL REGISTERS receive or transfer all bits of data simultaneously.

SHIFT REGISTERS are used to perform serial-to-parallel and parallel-to-serial conversion and for scaling binary numbers.

SERIAL TRANSFER causes all bits of data to be transferred on a single line.

PARALLEL TRANSFER has each data bit on its own line.

SCALING of binary numbers means to increase or decrease the magnitude of a number by a power of 2.

LOGIC FAMILIES are composed of logic circuits based on particular types of elements.

ANSWERS TO QUESTIONS Q1. THROUGH Q64.

A1. ⊕.

A2. Low (0).

A3. One or the other of the inputs must be HIGH, but not both at the same time.

A4. Exclusive NOR (X-NOR).

A5. HIGH.

A6. The half adder generates a carry.

A7. Quarter adder.

A8. Sum equals 0 with a carry of 1.

A9. Full adder.

A10. Four.

A11. $S_1 = 1$, $S_2 = 0$ *and* $C_2 = 1$.

A12. $C_1 = 0$.

A13. X-OR gates.

A14. Four.

A15. MSD of the sum.

A16. Add 1 portion.

A17. Subtrahend.

A18. Storing information.

A19. Six.

A20. 1 and 0, or opposite states.

A21. By cross-coupling NAND or OR gates.

A22. One.

A23. To divide the input by 2.

A24. Clock and data.

A25. Up to one clock pulse.

A26. A positive-going clock pulse.

A27. J-K flip-flop.

A28. Set, or HIGH (1).

A29. When the clock pulse goes LOW.

A30. Both J and K must be HIGH.

A31. Clear (CLR) and preset (PS or PR).

A32. The flip-flop is jammed.

A33. A timing signal.

A34. An astable or free-running multivibrator.

A35. Triggers.

A36. A multiphase clock.

A37. 32.

A38. Ripple.

A39. Toggle.

A40. Synchronous.

A41. The AND gate.

A42. 1111_2, *or* 15_{10}.

A43. Four.

A44. FFs 2 and 4.

A45. Two.

A46. Three.

A47. The input, or clock pulse.

A48. One.

A49. Four.

A50. $\overline{Q}$ *output of FF 1 going LOW.*

A51. 16.

A52. Parallel.

A53. By clearing the register.

A54. Shift register.

A55. Serial.

A56. Requires more circuitry.

A57. Eight.

A58. Four.

A59. 2^2*, or four times.*

A60. Three to the left.

A61. Logic families.

A62. DTL (diode transistor logic).

A63. DIPs (dual inline packages).

A64. Compatible.

APPENDIX I

GLOSSARY

ADDEND —A number to be added to an augend.

ADDITION —A form of counting where on quantity is added to another.

AND GATE —A logic circuit in which all inputs must be HIGH to produce a HIGH output.

ASSOCIATIVE LAW —A simple equality statement A(BC) = ABC or A+(B+C) = A+B+C.

AUGEND —A number to which another number is to be added.

BASE —The number of symbols used in the particular number system.

BCD (BINARY CODED DECIMAL) —A method of using binary digits to represent the decimal digits 0 through 9.

BINARY SYSTEM —The base 2 number system using 0 and 1 as the symbols.

BOOLEAN ALGEBRA —A mathematical concept based on the assumption that most quantities have two possible conditions —TRUE and FALSE.

BOOLEAN EXPRESSION —A description of the input or output conditions of a logic gate.

BORROW —To transfer a digit (equal to the base of the number system) from the next higher order column for the purpose of subtraction.

CARRY —A carry is produced when the sum of two or more numbers in a vertical column equals or exceeds the base of the number system in use.

CLOCK —A circuit that generates timing control signals in a computer or other type of digital equipment.

COMMUTATIVE LAW —The order in which terms are written does not affect their value; AB = BA, A+B = B+A.

COMPATIBILITY —The feature of logic families that allows interconnection of circuits without the need for additional circuitry.

COMPLEMENT —Something used to complete something else.

COMPLEMENTARY LAW —A term ANDed with its complement is 0, and a term ORed with its complement is 1; $A\overline{A} = 0$, $A + \overline{A} = 1$.

CONVERSION —To change a number in one base to its equivalent in another base.

COUNTER —A device that counts.

D FLIP-FLOP —Stores the data bit (D) in conjunction with the clock input.

DECADE COUNTER —Counter from 0 to 10_{10} in base 2, then resets.

DECIMAL POINT —The radix point for the decimal system.

DECIMAL SYSTEM —A number system with a base or radix of 10.

DEMORGAN'S THEOREM —This theorem has two parts: the first states that $\overline{AB} = \overline{A} + \overline{B}$; the second states that $\overline{A + B} = \overline{A}\ \overline{B}$.

DIFFERENCE —That which is left after subtraction.

DISTRIBUTIVE LAW —(1) a term (A) ANDed with a parenthetical expression (B+C) equals that term ANDed with each term within the parenthesis: A(B+C) = AB+AC; (2) a term (A) ORed with a parenthetical expression (BC) equals that term ORed with each term within the parenthesis: A+(BC) = (A+B)(A+C).

DIVIDEND —A number to be divided.

DIVISOR —A number by which a dividend is divided.

DOUBLE NEGATIVE LAW —A term that is inverted twice is equal to the term; $\overline{\overline{A}} = A$.

DOWN COUNTER—A circuit that counts from a predetermined number down to 0.

EXCLUSIVE-NOR (X-NOR) —A logic circuit that produces a HIGH output when all inputs are LOW or all inputs are HIGH.

EXCLUSIVE-OR (X-OR) GATE —A logic circuit that produces a HIGH output when one and only one input is HIGH.

EXPONENT —A number above and to the right of a base indicating the number of time the base is multiplied by itself; $2^4 = 2 \times 2 \times 2 \times 2$.

FLIP-FLOP —A bistable multivibrator.

FRACTIONAL NUMBER —A symbol to the right of the radix point that represents a portion of a complete object.

HEXADECIMAL (HEX) SYSTEM —The base 16 number system using 0 through 9 and A, B, C, D, E, and F as symbols.

IDEMPOTENT LAW —States that a term ANDed with itself or ORed with itself is equal to the term; AA = A, A+A = A.

INVERTER —A logic gate that outputs the complement of its input.

J-K FLIP-FLOP —Can perform the functions of the RS, T, and D flip-flops.

LAW OF ABSORPTION —This law is the result of the application of several other laws. It states that A(A+B) = A or A+(AB) = A.

LAW OF COMMON IDENTITIES —The two statements $A(\overline{A}+B) = AB$ and $A+\overline{A}B = A+B$ are based on the complementary law.

LAW OF IDENTITY —States that a term TRUE in one part of an expression will be TRUE in all parts of the expression; A = A, $\overline{A} = \overline{A}$.

LAW OF INTERSECTION —A term ANDed with 1 equals that term, and a term ANDed with 0 equals 0; $A \cdot 1 = A$, $A \cdot 0 = 0$.

LAW OF UNION —A term ORed with 1 equals 1; a term ORed with 0 equals that term; A+1 = 1, A + 0 = 0.

LEAST SIGNIFICANT (LSD) —The digit which has the least effect on the value of a number.

LOGIC —The science of reasoning; the development of a reasonable or logical conclusion based on known information.

LOGIC FAMILY —A group of logic circuits based on specific types of circuit elements (DTL, TTL, CMOS, and so forth).

LOGIC GATES —Decision-making circuits in computers and other types of equipment.

LOGIC POLARITY —The polarity of a voltage used to represent the logic 1 state.

LOGIC SYMBOL —Standard symbol used to indicate a particular logic function.

MINUEND —The number from which another number is subtracted.

MIXED NUMBER —Represents one or more complete units and a portion of a single unit.

MODULUS —The number of different values that a counter can contain or display.

MOST SIGNIFICANT DIGIT (MSD) —The digit which if changed will have the greatest effect on the value of a number.

NAND GATE —An AND gate with an inverted output. The output is LOW when all inputs are HIGH, and HIGH when any or all inputs are LOW.

NEGATIVE LOGIC —The voltage representing logic state 1 is more negative that the voltage representing a logic state 0.

NEGATOR —See inverter.

NOR GATE —An OR gate with an inverted output. The output is LOW when any or all inputs are HIGH, and HIGH when all inputs are LOW.

NOT CIRCUIT —See inverter.

NUMBER —A symbol used to represent a unit or a quantity.

OCTAL SYSTEM —The base 8 number system using 0 through 7 as the symbols.

OR GATE —A logic circuit which produces a HIGH output when one or more inputs is/are HIGH.

PARALLEL DATA —Each bit of data has a separate line and all bits are moved simultaneously.

PARALLEL REGISTER —A register that receives, stores, and transfers data in a parallel mode.

POSITIONAL NOTATION —A method where the value of the number is defined by the symbol and the symbol's position.

POSITIVE LOGIC —The voltage representing logic state 1 is more positive than the voltage representing a logic state 0.

POWER OF A NUMBER —The number of times a base is multiplied by itself. The power of a base is indicated by the exponent; that is, $10^3 = 10 \times 10 \times 10$.

QUOTIENT —The result in division.

RADIX POINT —The symbol that separates whole numbers and fractional numbers.

RADIX —The total number of symbols used in a particular number system.

REGISTER —A circuit of flip-flops designed to receive, store, and transfer data.

REMAINDER —The final undivided part that is less than the divisor.

RING COUNTER —A loop in which only one flip-flop will be set at any given time; used in timing.

RIPPLE (ASYNCHRONOUS) COUNTER —A circuit that counts from 0 to a specified value. Subject to error at high frequency.

R's (RADIX) COMPLEMENT —The difference between a given number and the next higher power of the number system (1000_8 minus 254_8 equals 524_8).

R's-1 (RADIX-1) COMPLEMENT —The difference between a given number and the highest value symbol in the number system (777_8 minus 254_8 equals 524_8).

R-S FLIP-FLOP —A flip-flop with two inputs —S (set) and R (reset). The Q output is HIGH in the set mode and LOW in the reset mode.

SERIAL DATA —All data bits are transferred one bit at a time along a single conductor.

SHIFT REGISTER —A register capable of serial-to-parallel and parallel-to-serial conversion and scaling.

SHIFTING —Moving the contents of a register right or left to scale the number or to input or output serial data.

SUBSCRIPT —A number written below and to the right of a value indicating the base or radix of the number system in use (35_8).

SUBTRACTION —Taking away one number from another.

SUBTRAHEND —The quantity to be subtracted from the minuend.

SUM —The result in addition.

SYNCHRONOUS COUNTER —Performs the same function as a ripple counter but error free at high frequency.

T FLIP-FLOP —A single input flip-flop that changes state with each positive pulse or each negative pulse. Divides input frequency by two.

TRUTH TABLE —A chart showing all possible input combinations and the resultant outputs.

UNIT —A single object.

UP/DOWN COUNTER —A counter circuit that can count up or down on command.

VINCULUM —A bar over a logic statement indicating the FALSE condition of the statement.

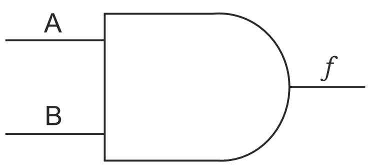
WHOLE NUMBER —A symbol that represents one or more complete objects.

ZERO —A symbol that indicates no numerical value for a position in positional notation.

APPENDIX II

LOGIC SYMBOLS

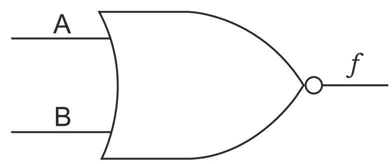
AND



A	B	f
0	0	0
0	1	0
1	0	0
1	1	1

$f = AB$

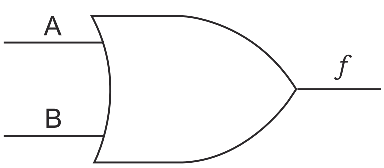
NOR



A	B	f
0	0	1
0	1	0
1	0	0
1	1	0

$f = \overline{A + B}$

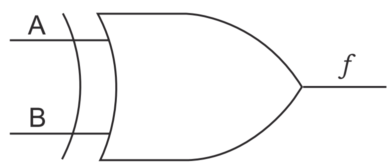
OR



A	B	f
0	0	0
0	1	1
1	0	1
1	1	1

$f = A + B$

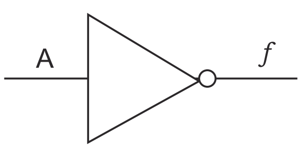
EXCLUSIVE-OR



A	B	f
0	0	0
0	1	1
1	0	1
1	1	0

$f = A \oplus B = A\overline{B} + \overline{A}B$

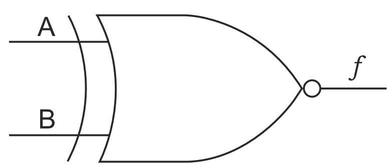
INVERTER



A	f
0	1
1	0

$f = \overline{A}$

EXCLUSIVE-NOR

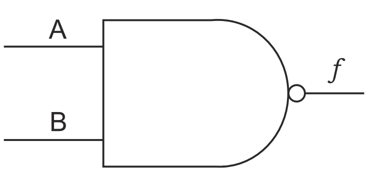


A	B	f
0	0	1
0	1	0
1	0	0
1	1	1

$f = \overline{A \oplus B}$

NTS13AP201

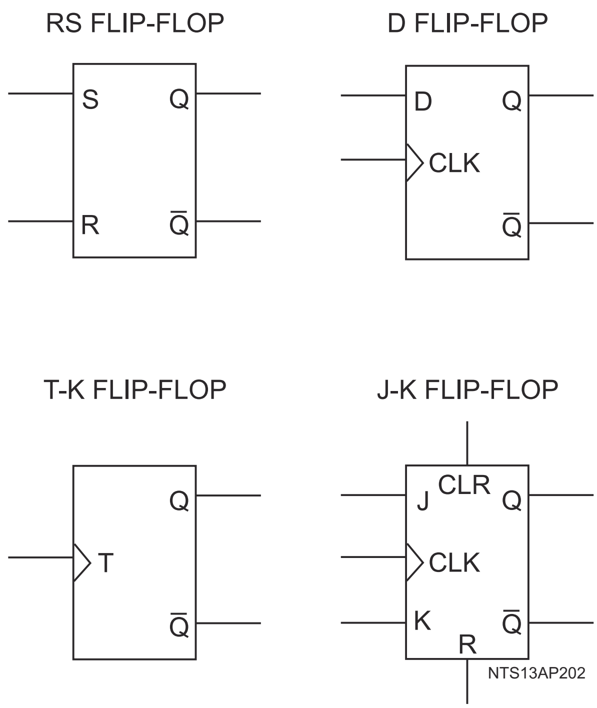
NAND



A	B	f
0	0	1
0	1	1
1	0	1
1	1	0

$f = \overline{AB}$

Logic Symbols (Cont'd)



MODULE 13

INDEX

ASSIGNMENT 1

Textbook Assignment: "Number Systems," chapter 1, pages 1-1 through 1-69.

1-1. Modern number systems are built around which of the following components?

1. Unit, number, and radix
2. Number, base, and radix
3. Position, power, and unit
4. Digit, power, and position

1-2. What term describes a single object in a modern number system?

1. Unit
2. Base
3. Digit
4. Number

1-3. What is a number?

1. A quantity of objects
2. A counting system based on symbols
3. The decimal system
4. A symbol representing a unit or a quantity

1-4. Which of the following symbols is NOT an Arabic figure?

1. C
2. 2
3. 7
4. 9

1-5. What term describes the number of symbols used in a number system?

1. Power
2. Radix
3. Exponent
4. Subscript

1-6. What is the base of a number system using all the letters of the alphabet —A=0, B=1, C=2, and so forth?

1. 29
2. 25
3. 28
4. 26

1-7. A number system uses the symbols 0 through 4. What is its base?

1. 6
2. 5
3. 3
4. 4

1-8. Using positional notation, what two factors determine the value of a number?

1. Symbol and position
2. Position and base
3. Radix and symbol
4. Unit and symbol

1-9. How many decimal units are represented by the 5 in the number 1572_{10}?

1. 100
2. 50
3. 500
4. 5000

1-10. The 1 in 1572_{10} is equal to what power of ten?

1. 10^1
2. 10^2
3. 10^3
4. 10^4

1-11. The power of a number is indicated by the

1. subscript
2. exponent
3. radical
4. radix

1-12. What is the value of 5 times 10^0?

1. 1
2. 0.5
3. 5
4. 50

1-13. Which of the following numbers is a mixed number?

1. 14.03
2. 156
3. 1,257
4. .0004

1-14. What term describes the symbol that separates the whole and fractional numbers in any number system?

1. Exponent
2. Radix point
3. Decimal point
4. Position point

1-15. What is the MSD of (a) 0.4201, (b) 13, and (c) 32.06?

1. (a) 4 (b) 1 (c) 3
2. (a) 1 (b) 3 (c) 2
3. (a) 2 (b) 1 (c) 2
4. (a) 4 (b) 3 (c) 3

1-16. What term is defined as a number to be added to a preceding number?

1. Addend
2. Augend
3. Carry
4. Sum

1-17. Identify A through D in the following example.

$$\begin{array}{rl} 1 & -\text{A} \\ 25 & -\text{B} \\ \underline{+\ 17} & -\text{C} \\ 42 & -\text{D} \end{array}$$

1. Addend, sum, carry, augend
2. Carry, augend, addend, sum
3. Augend, carry, sum, addend
4. Carry, augend, sum, addend

IN ANSWERING QUESTION 1-18, PERFORM THE INDICATED OPERATION.

1-18. Add:

$$\begin{array}{r} 326_{10} \\ \underline{+\ 192_{10}} \end{array}$$

1. 418
2. 508
3. 528
4. 518

1-19. In subtraction, the (a) is subtracted from the (b).

1. (a) Minuend (b) subtrahend
2. (a) Remainder (b) subtrahend
3. (a) Subtrahend (b) difference
4. (a) Subtrahend (b) minuend

1-20. A borrow is required in which of the following examples?

1. $\begin{array}{r} 64 \\ \underline{-\ 59} \end{array}$

2. $\begin{array}{r} 32 \\ \underline{-\ 12} \end{array}$

3. $\begin{array}{r} 59 \\ \underline{-\ 17} \end{array}$

4. $\begin{array}{r} 29 \\ \underline{-\ 11} \end{array}$

1-21. The result of subtraction is known as the

1. addend
2. quotient
3. difference
4. subtrahend

1-22. What are the (a) radix and (b) the symbols used in the binary number system?

1. (a) 1 (b) 0, 1
2. (a) 2 (b) 1, 0
3. (a) 3 (b) 0, 1, 2
4. (a) 0 (b) 1, 2

1-23. Positional notation for the binary system is based on powers of

1. 1
2. 2
3. 3
4. 0

1-24. What is the decimal equivalent of 2^3?

1. 6
2. 2
3. 8
4. 10

1-25. The decimal number 1 is equal to what power of two?

1. 2^1
2. 2^2
3. 2^3
4. 2^0

1-26. Which of the following powers of two indicates a fraction?

1. 2^1
2. 2^0
3. 2^{-2}
4. 2^2

1-27. Which digit is the MSD in the binary number 1011001?

1. The 0 farthest to the left
2. The 1 farthest to the left
3. The 0 farthest to the right
4. The 1 farthest to the right

1-28. Which of the following combinations of binary addition is INCORRECT?

1. 1 + 1 = 1 with a carry
2. 1 + 0 = 1
3. 0 + 0 = 0
4. 0 + 1 = 1

IN ANSWERING QUESTIONS 1-29 THROUGH 1-33, PERFORM THE INDICATED OPERATION.

1-29. Add:

$$\begin{array}{r} 10010_2 \\ \underline{+\quad 1010_2} \end{array}$$

1. 10000_2
2. 11100_2
3. 10101_2
4. 11010_2

1-30. Add:

$$\begin{array}{r} 11101_2 \\ \underline{+\ 01001_2} \end{array}$$

1. 11111_2
2. 100110_2
3. 111010_2
4. 100000_2

1-31. Add:

$$\begin{array}{r} 111_2 \\ \underline{+\ 001_2} \end{array}$$

1. 100_2
2. 1010_2
3. 1001_2
4. 1000_2

1-32. Add:

$$\begin{array}{r} 100_2 \\ 010_2 \\ \underline{+\ 001_2} \end{array}$$

1. 1000_2
2. 1001_2
3. 101_2
4. 111_2

1-33. Add:

$$\begin{array}{r} 11010_2 \\ 100_2 \\ \underline{+\ \ 111_2} \end{array}$$

1. 011110_2
2. 101010_2
3. 100101
4. 111001_2

1-34. Which of the following rules of binary subtraction is correct?

1. 0 – 0 = 0 with a borrow
2. 1 – 0 = 0
3. 2 – 1 = 1
4. 1 – 0 = 1

1-35. In the following example, which number is the minuend?

$$\begin{array}{r} 10110_2 \\ \underline{-\ \ 1100_2} \end{array}$$

1. 10110_2
2. 1100_2
3. 1010_2
4. 10101_2

1-36. Which of the following rules of binary subtraction requires the use of a borrow?

1. 0 – 0
2. 1 – 0
3. 1 – 1
4. 0 – 1

1-37. In binary subtraction, what is the value of a borrow when it is moved to the next lower order column?

1. 1_2
2. 2_2
3. 10_2
4. 10_{10}

IN ANSWERING QUESTIONS 1-38 THROUGH 1-40, PERFORM THE INDICATED OPERATION.

1-38. Subtract:

$$\begin{array}{r} 11111_2 \\ \underline{-\ \ 10101_2} \end{array}$$

1. 10100_2
2. 11010_2
3. 00101_2
4. 01010_2

1-39. Subtract:

$$\begin{array}{r} 10101_2 \\ \underline{-\ \ 1111_2} \end{array}$$

1. 00110_2
2. 01001_2
3. 01100_2
4. 00010_2

1-40. Subtract:

$$\begin{array}{r} 10001_2 \\ -\ \ 0110_2 \\ \hline \end{array}$$

1. 1010_2
2. 1011_2
3. 1110_2
4. 1001_2

1-41. Subtraction is accomplished by which of the following methods in a computer that can only add?

1. Binary subtraction
2. Decimal complement
3. Complementary subtraction
4. Minuend complement

1-42. What is the R's-1 complement of 633_{10}?

1. 477_{10}
2. 466_{10}
3. 366_{10}
4. 377_{10}

1-43. What is the R's-1 complement of 1011101_2?

1. 0100010_2
2. 1011110_2
3. 8988898_{10}
4. 8988899_{10}

1-44. What is the R's complement of 395_{10}?

1. 604_{10}
2. 605_{10}
3. 715_{10}
4. 714_{10}

1-45. Which of the following parts of a subtraction problem must be complemented to perform complementary subtraction?

1. Subtrahend
2. Difference
3. Remainder
4. Minuend

1-46. Which of the following examples is the correct step when using the R's complement to subtract 123_{10} from 264_{10}?

1. $\begin{array}{r} 735 \\ +\ 123 \\ \hline \end{array}$

2. $\begin{array}{r} 264 \\ +\ 876 \\ \hline \end{array}$

3. $\begin{array}{r} 264 \\ +\ 321 \\ \hline \end{array}$

4. $\begin{array}{r} 264 \\ +\ 877 \\ \hline \end{array}$

1-47. Which of the following solutions is correct when using the R's complement method of subtracting 516_{10} from 845_{10}?

1. $\begin{array}{r} 845 \\ +\ 483 \\ \hline 328 \end{array}$

2. $\begin{array}{r} 845 \\ +\ 484 \\ \hline 329 \end{array}$

3. $\begin{array}{r} 845 \\ +\ 615 \\ \hline 460 \end{array}$

4. $\begin{array}{r} 155 \\ +\ 516 \\ \hline 671 \end{array}$

1-48. Which of the following numbers is the R's-1 complement of the binary number 10101?

1. 10100_2
2. 01011_2
3. 01100_2
4. 01010_2

1-49. Perform the R's-1 complement of the following binary numbers.

(a) 1000, (b) 1101, (c) 0100

1. (a) 0111_2 (b) 0010_2 (c) 1011_2
2. (a) 0110_2 (b) 0011_2 (c) 1010_2
3. (a) 0111_2 (b) 0011_2 (c) 1011_2
4. (a) 0110_2 (b) 0010_2 (c) 1010_2

1-50. Which of the following statements regarding the forming of the R's complement of a binary number is correct?

1. Retain the MSD and complement all other digits
2. Complement all digits and subtract 1
3. Complement all digits
4. Retain the least significant 1 and complement all other digits to the left

1-51. Which of the following numbers is the R's complement of 100101_2?

1. 101010_2
2. 011011_2
3. 110110_2
4. 101001_2

IN ANSWERING QUESTIONS 1-52 AND 1-53, USE THE R'S COMPLEMENT METHOD TO SOLVE THE PROBLEMS.

1-52. Subtract:

$$\begin{array}{r} 1001_2 \\ \underline{-\ 101_2} \end{array}$$

1. 110_2
2. 010_2
3. 100_2
4. 1100_2

1-53. Subtract:

$$\begin{array}{r} 1101_2 \\ \underline{-\ 1010_2} \end{array}$$

1. 0001_2
2. 0100_2
3. 0110_2
4. 0011_2

1-54. In the previous problems, what is indicated by the carry that expands the difference by one place?

1. The answer is incorrect
2. The answer is a positive number
3. The answer is correct
4. The answer is a negative number

IN ANSWERING QUESTIONS 1-55 THROUGH 1-57, SOLVE THE SUBTRACTION PROBLEMS AND INDICATE THE SIGN OF THE DIFFERENCE.

1-55. Subtract:

$$\begin{array}{r} 10010_2 \\ \underline{-\ 10001_2} \end{array}$$

1. 00001_2 positive
2. 01111_2 negative
3. 01100_2 positive
4. 00001_2 negative

1-56. Subtract:

$$\begin{array}{r} 0001_2 \\ \underline{-\ 1111_2} \end{array}$$

1. 0010_2 negative
2. 0111_2 positive
3. 1110_2 negative
4. 1111_2 positive

1-57. Subtract:

$$\begin{array}{r} 01111_2 \\ \underline{-\ 10000_2} \end{array}$$

1. 11111_2 negative
2. 00001_2 negative
3. 01111_2 positive
4. 10000_2 positive

1-58. What is the radix of the octal number system?

1. 10_{10}
2. 0 to 7
3. 8
4. 7_8

1-59. Which of the following is NOT a valid octal number?

1. 604_8
2. 591_8
3. 743_8
4. 477_8

1-60. Which of the following equations is correct?

1. $8^0 = 8$
2. $8^1 = 1$
3. $8^4 = 8 \times 8 \times 8$
4. $8^2 = 8 \times 8$

1-61. One octal digit is represented by how many binary digits?

1. One
2. Two
3. Three
4. Four

1-62. What is the decimal value of 8^3?

1. 64_{10}
2. 128_{10}
3. 256_{10}
4. 512_{10}

1-63. Which of the following symbols is the least significant digit of the octal number 1622.374?

1. 1
2. 2
3. 6
4. 4

1-64. What is the sum of 4_8 and 4_8?

1. 10_{10}
2. 10_8
3. 7_8
4. 16_8

1-65. What is the sum of 77_8 and 3_8?

1. 127_8
2. 80_8
3. 100_8
4. 102_8

IN ANSWERING QUESTION 1-66, PERFORM THE INDICATED OPERATION.

1-66. Add:

$$\begin{array}{r} 374_8 \\ \underline{+\ \ 165_8} \end{array}$$

1. 465_8
2. 561_8
3. 437_8
4. 531_8

1-67. Find the sum of 7741_8 and 67_8.

1. 10000_8
2. 10030_8
3. 7030_8
4. 10730_8

1-68. What is the sum of 7_8, 6_8, 5_8, and 4_8?

1. 22_8
2. 17_8
3. 26_8
4. 24_8

IN ANSWERING QUESTIONS 1-69 AND 1-70, PERFORM THE INDICATED OPERATION.

1-69. Add:

$$\begin{array}{r} 2601_8 \\ +\ 5035_8 \\ \hline \end{array}$$

1. 7636_8
2. 7640_8
3. 10636_8
4. 10700_8

1-70. Add:

$$\begin{array}{r} 2345_8 \\ +\ 7654_8 \\ \hline \end{array}$$

1. 10110_8
2. 10741_8
3. 10571_8
4. 12221_8

1-71. Which, if any, of the following is a difference between subtracting octal numbers and subtracting decimal numbers?

1. The amount of the borrow
2. The octal minuend is converted to decimal
3. The octal subtrahend is converted to decimal
4. None; there is no difference

IN ANSWERING QUESTIONS 1-72 THROUGH 1-74, PERFORM THE INDICATED OPERATION.

1-72. Subtract:

$$\begin{array}{r} 646_8 \\ -\ 421_8 \\ \hline \end{array}$$

1. 125_8
2. 265_8
3. 225_8
4. 225_{10}

1-73. Subtract:

$$\begin{array}{r} 421_8 \\ -\ 267_8 \\ \hline \end{array}$$

1. 144_8
2. 232_8
3. 132_8
4. 142_8

1-74. Subtract:

$$\begin{array}{r} 3000_8 \\ -\ 777_8 \\ \hline \end{array}$$

1. 2001_8
2. 2011_8
3. 2111_8
4. 2000_8

ASSIGNMENT 2

Textbook assignment: Chapter 1, "Number Systems," page 1-30 through 1-62.

2-1. Which of the following numbers does NOT represent a hexadecimal value?

1. 2DF4
2. A32B
3. 47CE
4. 9FGF

2-2. What is the decimal value of the highest symbol in the hex system?

1. 15_{10}
2. 16_{10}
3. F_{16}
4. 10_{16}

2-3. The decimal number 256 is equal to what power of 16?

1. 16^3
2. 16^2
3. 16^4
4. 16^1

2-4. List the MSD and LSD of the hex number F24.ECB.

1. MSD = 4, LSD = E
2. MSD = 4, LSD = B
3. MSD = F, LSD = 4
4. MSD = F, LSD = B

IN ANSWERING QUESTIONS 2-5 THROUGH 2-7, PERFORM THE INDICATED OPERATION.

2-5. Find the sum.

$$\begin{array}{r} A_{16} \\ \underline{+\ 4_{16}} \end{array}$$

1. C_{16}
2. D_{16}
3. E_{16}
4. F_{16}

2-6. Add:

$$\begin{array}{r} 1E_{16} \\ \underline{+\ 19_{16}} \end{array}$$

1. 33_{16}
2. 31_{16}
3. 37_{16}
4. 47_{16}

2-7. Add:

$$\begin{array}{r} 478_{16} \\ \underline{+792_{16}} \end{array}$$

1. $C11_{16}$
2. $C0A_{16}$
3. $B00_{16}$
4. BFA_{16}

2-8. When a borrow is taken from a hex number, that number is reduced by how much?

1. 1
2. 10_{16}
3. 16_{10}
4. 10_{10}

IN ANSWERING QUESTION 2-9, FIND THE DIFFERENCE.

2-9. Subtract:

$$\begin{array}{r} 10_{16} \\ \underline{-\ 8_{16}} \end{array}$$

1. A_{16}
2. 7_{16}
3. 8_{16}
4. 9_{16}

2-10. To begin conversion of a decimal number to a different base, divide the (a) by (b).

1. (a) new base
   (b) 10
2. (a) decimal number
   (b) 2
3. (a) decimal number
   (b) the new base
4. (a) new base
   (b) the decimal equivalent

2-11. Which of the following terms describes the first remainder when converting decimal numbers to other bases?

1. MSD
2. LSD
3. Radix of the new base
4. Exponent of the new base

IN ANSWERING QUESTIONS 2-12 THROUGH 2-14, USE THE DIVISION METHOD TO CONVERT DECIMAL NUMBERS TO BINARY.

2-12. 43_{10}

1. 101011_2
2. 110101_2
3. 100011_2
4. 101101_2

2-13. 63_{10}

1. 101011_2
2. 110011_2
3. 101101_2
4. 111111_2

2-14. 130_{10}

1. 1111000_2
2. 1111010_2
3. 10000010_2
4. 10001010_2

2-15. To convert fractional decimal numbers to binary, multiply the number by (a) and extract the portion of the product to the (b) of the radix point.

1. (a) 2 (b) left
2. (a) 2 (b) right
3. (a) 10 (b) left
4. (a) 10 (b) right

IN ANSWERING QUESTIONS 2-16 THROUGH 2-19, PERFORM THE INDICATED OPERATION.

2-16. Convert 0.75_{10} to binary.

1. 0.10_2
2. 0.01_2
3. 0.11_2
4. 1.00_2

2-17. Convert 0.625_{10} to binary.

1. 0.011_2
2. 0.101_2
3. 0.110_2
4. 0.100_2

2-18. Convert 12.5_{10} to base 2.

1. 1100.10_2
2. 1010.01_2
3. 1001.10_2
4. 1101.01_2

2-19. Convert 33.34_{10} to base 2 (four places).

1. 11110.1010_2
2. 100001.1010_2
3. 100001.0101_2
4. 100010.1011_2

IN ANSWERING QUESTIONS 2-20 THROUGH 2-22, USE THE DIVISION METHOD TO CONVERT DECIMAL NUMBERS TO OCTAL.

2-20. 193_{10}

1. 62_8
2. 142_8
3. 301_8
4. 403_8

2-21. 746_{10}

1. 1352_8
2. 2531_8
3. 1476_8
4. 2312_8

2-22. 3007_{10}

1. 5000_8
2. 5677_8
3. 4771_8
4. 4115_8

IN ANSWERING QUESTIONS 2-23 THROUGH 2-25, PERFORM THE INDICATED OPERATION.

2-23. Convert 0.305_{10} to octal (four places).

1. 0.5765_8
2. 0.1471_8
3. 0.3050_8
4. 0.2341_8

2-24. Convert 78.9_{10} to octal (three places).

1. 100.417_8
2. 103.714_8
3. 116.714_8
4. 116.147_8

2-25. Convert 506.66_{10} to octal (four places).

1. 677.5063_8
2. 521.4401_8
3. 653.3774_8
4. 772.5217_8

2-26. What is the hex equivalent of 45_{10}?

1. $1F_{16}$
2. 24_{16}
3. $2A_{16}$
4. $2D_{16}$

2-27. What is the hex equivalent of 255_{10}?

1. AE_{16}
2. BC_{16}
3. CG_{16}
4. FF_{16}

IN ANSWERING QUESTIONS 2-28 THROUGH 2-30, PERFORM THE INDICATED OPERATION.

2-28. Convert 1609_{10} to hex.

1. $5A5_{16}$
2. 649_{16}
3. $C41_{16}$
4. $A95_{16}$

2-29. Convert 0.84 to hex (three places).

1. $0.D70_{16}$
2. $0.1F3_{16}$
3. $0.AAC_{16}$
4. $0.C3E_{16}$

2-30. Convert 0.109_{10} to hex (four places).

1. 0.1114_{16}
2. $0.101F_{16}$
3. $0.1BE7_{16}$
4. $0.09D4_{16}$

2-31. What is the hex equivalent of 174.95_{10}? Carry out two places.

1. $9F.C4_{16}$
2. $AE.F3_{16}$
3. $AE.9C_{16}$
4. $BA.EC_{16}$

2-32. What is the hex equivalent of 7023.869_{10}? Carry out to four places.

1. $1C5E.A9F6_{16}$
2. $1B6F.DE76_{16}$
3. $1D7C.EC87_{16}$
4. $1AA9.DB1A_{16}$

IN ANSWERING QUESTIONS 2-33 THROUGH 2-38, CONVERT THE BINARY NUMBERS TO THEIR OCTAL EQUIVALENT.

2-33. 0011010_2

1. 150_8
2. 032_8
3. 042_8
4. 062_8

2-34. 001010011100_2

1. 516_8
2. 147_8
3. 667_8
4. 1234_8

2-35. 0101011011111101_2

1. 53375_8
2. 155776_8
3. 255771_8
4. 126771_8

2-36. 0.1110101000_2

1. 0.1560_8
2. 0.1650_8
3. 0.750_8
4. 0.724_8

2-37. 1001000.00011110_2

1. 110.074_8
2. 440.036_8
3. 410.070_8
4. 220.074_8

2-38. 1111111011.11110011_2

1. 1773.746_8
2. 7751.363_8
3. 1773.633_8
4. 7751.473_8

IN ANSWERING QUESTIONS 2-39 THROUGH 2-42, CONVERT THE BINARY NUMBERS TO THEIR HEXADECIMAL EQUIVALENTS.

2-39. 101101_2

1. $B1_{16}$
2. 55_{16}
3. $1B_{16}$
4. $2D_{16}$

2-40. 111010110010_2

1. $EB2_{16}$
2. $EC4_{16}$
3. $7B2_{16}$
4. 7262_{16}

2-41. 0.0100111100_2

1. $0.4\ F_{16}$
2. $0.9\ E_{16}$
3. $0.13C_{16}$
4. 0.236_{16}

2-42. 11011100.110010101011_2

1. 670.6253_{16}
2. $9A.BAC_{16}$
3. $DC.CAB_{16}$
4. $AB.CDE_{16}$

IN ANSWERING QUESTIONS 2-43 THROUGH 2-45, CONVERT THE OCTAL NUMBERS TO THEIR BINARY EQUIVALENTS.

2-43. 571_8

1. 101111001_2
2. 1011111_2
3. 101011101_2
4. 110111010_2

2-44. 1312_8

1. 0101101010_2
2. 1011001010_2
3. 00111000101_2
4. 01010101010_2

2-45. 136.52_8

1. 1110110.101100_2
2. 11011111.101010_2
3. 01011011.010101_2
4. 01011110.101010_2

IN ANSWERING QUESTIONS 2-46 AND 2-47, PERFORM THE INDICATED OPERATIONS.

2-46. Convert 24.73_8 to hex.

1. 11.76_{16}
2. $14.EC_{16}$
3. $24.7D_{16}$
4. $20.A6_{16}$

2-47. Convert 657.13_8 to hex.

1. 328.065_{16}
2. $D37.26_{16}$
3. $1AF.2C_{16}$
4. $20A.B1_{16}$

IN ANSWERING QUESTIONS 2-48 THROUGH 2-50, CONVERT THE HEX NUMBERS TO BINARY.

2-48. $2A_{16}$

1. 01001010_2
2. 00101010_2
3. 01001100_2
4. 00011100_2

2-49. $E47_{16}$

1. 111001000111_2
2. 111101000111_2
3. 110010001110_2
4. 101010100111_2

2-50. $8C.1F_{16}$

1. 100111.00111110_2
2. 10001100.00011111_2
3. 1001100.00011110_2
4. 10011101.00011111_2

IN ANSWERING QUESTIONS 2-51 AND 2-52, CONVERT THE HEX NUMBERS TO OCTAL.

2-51. $74E_{16}$

1. 7416_8
2. 7217_8
3. 3516_8
4. 4636_8

2-52. $F1.C8_{16}$

1. 741.620_8
2. 661.304_8
3. 331.64_8
4. 361.62_8

2-53. What is the decimal equivalent of 1011_2

1. 13_{10}
2. 11_{10}
3. 9_{10}
4. 10_{10}

IN ANSWERING QUESTIONS 2-54 AND 2-55, PERFORM THE INDICATED OPERATION.

2-54. Convert 101101_2 to decimal.

1. 55_{10}
2. 29_{10}
3. 36_{10}
4. 45_{10}

2-55. Convert 1110111011_2 to decimal.

1. 773_{10}
2. 867_{10}
3. 955_{10}
4. 1673_{10}

IN ANSWERING QUESTIONS 2-56 THROUGH 2-60, CONVERT THE OCTAL NUMBERS TO THEIR DECIMAL EQUIVALENTS.

2-56. 133_8

1. 100_{10}
2. 107_{10}
3. 63_{10}
4. 91_{10}

2-57. 2737_8

1. 1899_{10}
2. 1503_{10}
3. 2105_{10}
4. 1511_{10}

2-58. 777.7_8

1. 511.875_{10}
2. 614.750_{10}
3. 614.875_{10}
4. 511.750_{10}

2-59. 1603.75_8 (three places)

1. 1219.840_{10}
2. 907.650_{10}
3. 899.953_{10}
4. 1143.150_{10}

2-60. 2000.1_8

1. 1250.25_{10}
2. 1024.125_{10}
3. 969.5_{10}
4. 1000.05_{10}

IN ANSWERING QUESTIONS 2-61 THROUGH 2-64, CONVERT THE HEX NUMBERS TO THEIR DECIMAL EQUIVALENTS.

2-61. $1B3_{16}$

1. 435_{10}
2. 1185_{10}
3. 2390_{10}
4. 4275_{10}

2-62. $10AF_{16}$

1. 4271_{10}
2. 2985_{10}
3. 3417_{10}
4. 4003_{10}

2-63. $C3.6_{16}$

1. 46.5_{10}
2. 84.4_{10}
3. 150.505_{10}
4. 195.375_{10}

2-64. $4DD.E_{16}$

1. 1245.7505_{10}
2. 1245.8750_{10}
3. 733.7505_{10}
4. 733.9375_{10}

2-65. What term describes the use of four binary digits to represent one decimal digit?

1. Decimal-coded binary
2. Octal-coded binary
3. Binary-coded decimal
4. Hexadecimal notation

2-66. How many binary digits are required to represent the decimal number 243 in BCD?

1. 12
2. 8
3. 3
4. 4

IN ANSWERING QUESTIONS 2-67 AND 2-68, PERFORM THE INDICATED OPERATION.

2-67. Convert 389_{10} to BCD.

1. $0000\ 1110\ 0101_{BCD}$
2. $0011\ 0001\ 1101_{BCD}$
3. $0011\ 1000\ 1001_{BCD}$
4. $0011\ 0100\ 0101_{BCD}$

2-68. Convert 010001010111_{BCD} to decimal.

1. 897_{10}
2. 857_{10}
3. 497_{10}
4. 457_{10}

2-69. Which of the following numbers has the highest decimal value?

1. 10011001_2
2. 11100001_2
3. $1001\ 1001_{BCD}$
4. $1000\ 0110_{BCD}$

2-70. Which of the following numbers is NOT a valid BCD number?

1. 0110
2. 1001
3. 1000
4. 1010

IN ANSWERING QUESTIONS 2-71 THROUGH 2-75, PERFORM THE INDICATED OPERATION.

2-71. Add:

$$\begin{array}{r} 0101_{BCD} \\ \underline{0010}_{BCD} \end{array}$$

1. 0111_{BCD}
2. 1001_{BCD}
3. 1010_{BCD}
4. 1000_{BCD}

2-72. Add:

$$\begin{array}{r} 1001_{BCD} \\ \underline{0111}_{BCD} \end{array}$$

1. $0001\ 0000_{BCD}$
2. $0001\ 0110_{BCD}$
3. $0111\ 0110_{BCD}$
4. $0001\ 1111_{BCD}$

2-73. Add:

$$\begin{array}{r} 1001_{BCD} \\ \underline{+100}_{BCD} \end{array}$$

1. 1101_{BCD}
2. 0011_{BCD}
3. $0001\ 0011_{BCD}$
4. $0001\ 0110_{BCD}$

2-74. Add:

$$\begin{array}{r} 0010\ \ 0011_{BCD} \\ \underline{+\ 0001\ \ 1000}_{BCD} \end{array}$$

1. $0011\ 1011_{BCD}$
2. $0011\ 0001_{BCD}$
3. $0100\ 1011_{BCD}$
4. $0100\ 0001_{BCD}$

2-75. Add:

$$\begin{array}{r} 0110\ \ 0100_{BCD} \\ \underline{+\ 1001\ \ 0010}_{BCD} \end{array}$$

1. $0000\ 1111\ 0110_{BCD}$
2. $0001\ 0101\ 1100_{BCD}$
3. $0001\ 0101\ 0110_{BCD}$
4. $0001\ 0111\ 0100_{BCD}$

ASSIGNMENT 3

Textbook assignment: Chapter 2, "Fundamental Logic Circuits," pages 2-1 through 2-36; and "Special Logic Circuits," Chapter 3, pages 3-1 through 3-3.

NOTE: UNLESS OTHERWISE INDICATED, ALL QUESTIONS AND ANSWERS REFER TO POSITIVE LOGIC.

3-1. Logic is the development of a reasonable conclusion based on known information.

1. T
2. F

3-2. Which of the following methods is used to represent the FALSE condition of the logic symbol, F?

1. (F)
2. [F]
3. $\overline{F}$
4. F

3-3. Which of the following statements describes logic polarity?

1. Negative logic is indicated by a vinculum
2. Positive logic is always represented by a positive voltage
3. A logic 1 is a positive voltage; a logic 0 is a negative voltage
4. The change in voltage polarity to represent a logic 1

3-4. Which of the following examples represents positive logic?

1. −5v equals 0, −10v equals 1
2. +5v equals 0, +10v equals 1
3. +5v equals 1, +10v equals 0
4. −5v equals 1, +5v equals 0

3-5. Of the following examples, choose the one that represents negative logic.

1. −15v equals 0, −10v equals 1
2. −10v equals 1, −15v equals 0
3. 0v equals 0, −10v equals 1
4. −5v equals 0, 0v equals 1

3-6. If the letter X represents an input to a logic device, what logic state of X must exist to activate the device or contribute to the activation of the device?

1. 0
2. 1
3. Positive logic
4. Negative logic

3-7. A chart that lists all possible input combinations and resultant output is called a

1. Truth Table
2. Decision Table
3. logic symbol listing
4. polarity magnitude listing

3-8. What is a logic gate?

1. A block diagram
2. An astable multivibrator
3. A bistable multivibrator
4. A decision-making circuit

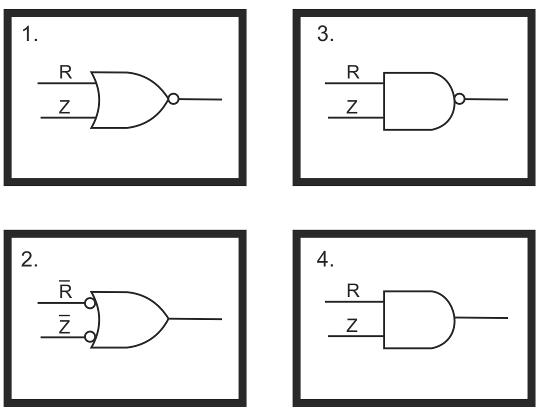
3-9. Which of the following logic gates requires all inputs to be TRUE at the same time to produce a TRUE output?

1. OR
2. NOT
3. AND
4. NAND

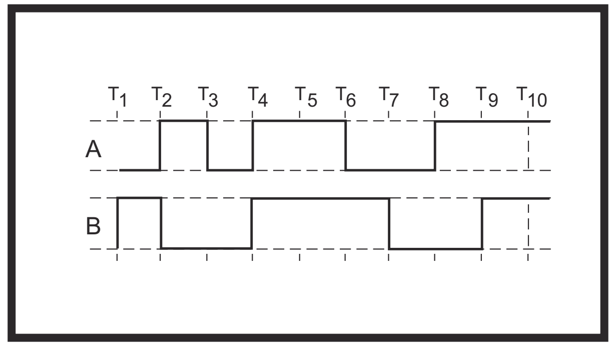
3-10. Which of the following output Boolean expressions is/are correct for an AND gate?

1. $f = AB$
2. $f = A \cdot B$
3. Both 1 and 2 above
4. $A=B$

3-11. Which of the following symbols represents the output Boolean expression RZ?



NTS13A301



NTS13A302

Figure 3A. —AND gate timing diagram.

IN ANSWERING QUESTIONS 3-12 THROUGH 3-14, REFER TO FIGURE 3A.

3-12. At which of the following times will the output of a two input AND gate go HIGH?

1. T_2, T_5, and T_8
2. T_4 only
3. T_2, T_6, and T_{10}
4. T_4 and T_9

3-13. At which of the following times will the output of the AND gate be LOW?

1. T_1 to T_4 and T_5 to T_8
2. T_1 to T_4 and T_6 to T_9
3. T_4 to T_6 and T_8 to T_{10}
4. T_1 to T_3 and T_6 to T_{10}

3-14. If input $\overline{B}$ were used instead of input B, how would the output be affected, if at all?

1. It would be HIGH from T_2 to T_4
2. It would be LOW from T_4 to T_6
3. It would be HIGH from T_2 to $T_3$and T_8 to T_9
4. It would not be affected

3-15. What is the output Boolean expression for an AND gate having F, $\overline{G}$, $\overline{K}$, and $\overline{L}$ as inputs?

1. $f = F\,\overline{G}\,\overline{K}\,\overline{L}$
2. $f = F + \overline{G}\,\overline{K}\,\overline{L}$
3. $f = \overline{F}\,G\,K\,L$
4. $f = F\,\overline{G} + \overline{K}\,\overline{L}$

THIS SPACE LEFT BLANK INTENTIONALLY.

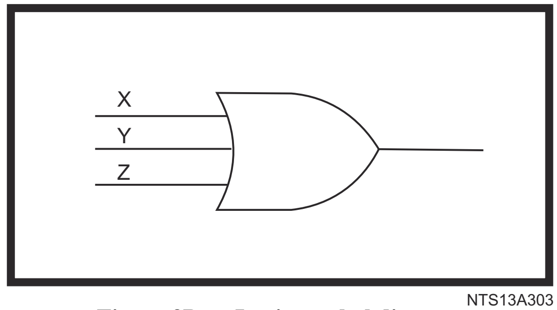


Figure 3B. —Logic symbol diagram.

IN ANSWERING QUESTIONS 3-16 THROUGH 3-18, REFER TO FIGURE 3B.

3-16. Which of the following gates is represented by the symbol in the figure?

1. AND
2. OR
3. NOR
4. X-OR

3-17. What is the output Boolean expression for the gate?

1. X, Y+Z
2. $X+Y \oplus X$
3. X+Y+Z
4. $X \oplus Y \oplus Z$

THIS SPACE LEFT BLANK INTENTIONALLY.

3-18. Which of the following Truth Tables correspond to the gate in the figure?

1.

X	Y	Z	f
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

2.

X	Y	Z	f
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

3.

X	Y	Z	f
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

4.

X	Y	Z	f
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

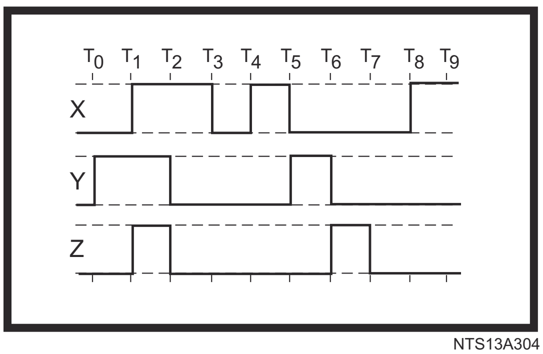


Figure 3C. — OR gate timing diagram.

IN ANSWERING QUESTIONS 3-19 THROUGH 3-21, REFER TO FIGURE 3C.

3-19. The gate will have a HIGH output at which of the following times?

1. T_0 to T_3, T_4 to T_7, and T_8 to T_9
2. T_1 to T_2 and T_4 to T_9
3. T_3 to T_4 and T_7 to T_8
4. T_1 to T_2 only

3-20. What are the input logic states between T_5 and T_6?

1. X = 1, Y = 0, Z = 0
2. X = 1, Y = 0, Z = 1
3. X = 0, Y = 1, Z = 0
4. X = 0, Y = 1, Z = 1

3-21. Between T_0 and T_6, at what times will the output of the gate be LOW?

1. T_3 to T_4
2. T_2 to T_5
3. T_3 to T_6
4. T_0 to T_1

3-22. What is the Boolean expression for an OR gate having the following inputs?

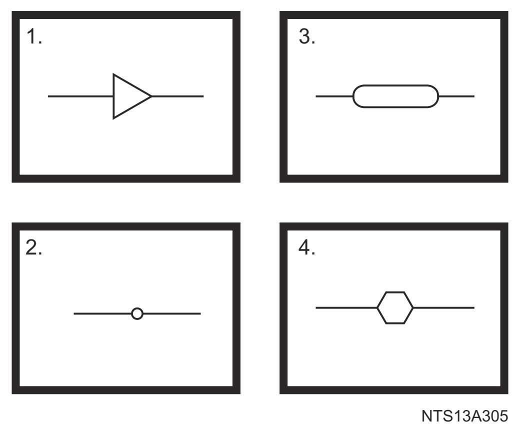
T (LOW)
$\overline{R}$ (LOW)
$\overline{P}$ (HIGH)

1. $T \overline{R} \overline{P}$
2. $\overline{T} R \overline{P}$
3. $\overline{T} + R + \overline{P}$
4. $T + \overline{R} + \overline{P}$

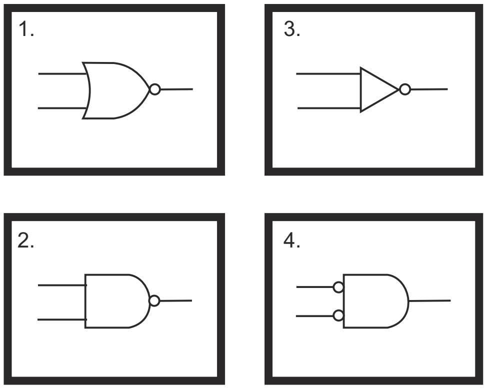
3-23. What is the purpose of an inverter?

1. To change logic polarity
2. To change voltage levels
3. To amplify the input
4. To complement the input

3-24. Which of the following symbols represents an inverter?



3-25. Which of the following is the standard logic symbol for a NAND gate?



NTS13A306

3-26. What is the output expression for an inverter with the following input?

$(R\overline{Q}) + (\overline{S}T)$

1. $(\overline{R}Q) + (S\overline{T})$
2. $R\overline{Q}\,\overline{S}T$
3. $\overline{RQST}$
4. $\overline{(R\overline{Q}) + (\overline{S}T)}$

3-27. Which of the following gates produces a HIGH output when any or all of the inputs are LOW?

1. AND
2. OR
3. NAND
4. NOR

X	Y	f
0	0	1
1	0	1
1	1	0

NTS13A307

Figure 3D. —Incomplete NAND gate Truth Table.

IN ANSWERING QUESTION 3-28, REFER TO FIGURE 3D.

3-28. Which of the following NAND gate input combinations and output function is missing?

1. X = 0, Y = 1, f = 0
2. X = 0, Y = 1, f = 1
3. X = 0, Y = 0, f = 0
4. X = 1, Y = 1, f = 1

3-29. What is the output Boolean expression for a NAND gate with inputs G, K, and $\overline{P}$?

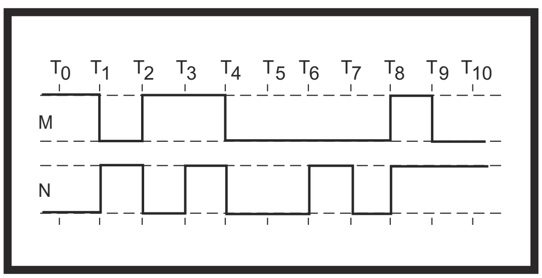
1. $\overline{GK}P$
2. $\overline{GKP}$
3. $\overline{GK\overline{P}}$
4. $\overline{G}\,\overline{K}\,\overline{P}$

3-30. The output of a NOR gate will be HIGH under which of the following conditions?

1. When all inputs are HIGH
2. When all inputs are LOW
3. When one input is HIGH
4. When one input is LOW

3-31. What is the output Boolean expression for a NOR gate with $\overline{P}$, Q, and R as inputs?

1. $\overline{\overline{P}+Q+R}$
2. $\overline{\overline{P}}+\overline{Q}+\overline{R}$
3. $P+\overline{Q}+\overline{R}$
4. $\overline{P}+Q+R$



NTS13A308

Figure 3E. —Input signal timing diagram.

IN ANSWERING QUESTIONS 3-32 THROUGH 3-34, REFER TO FIGURE 3E.

3-32. Figure 3E represents the input signals to a NOR gate. What is the output expression?

1. $M+N$
2. $\overline{M}+\overline{N}$
3. $\overline{\overline{M}+\overline{N}}$
4. $\overline{M+N}$

3-33. At which of the following times will the output be HIGH?

1. T_3 to T_4
2. T_8 to T_9
3. T_4 to T_6 and T_7 to T_8
4. T_3 to T_4 and T_8 to T_9

3-34. What should the output be between times T_3 and T_4?

1. LOW
2. HIGH

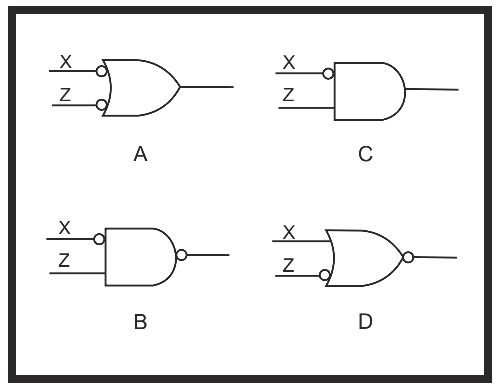


Figure 3F. —Logic gates. NTS13A309

IN ANSWERING QUESTIONS 3-35 THROUGH 3-37, REFER TO FIGURE 3F.

3-35. What is the output expression for gate D?

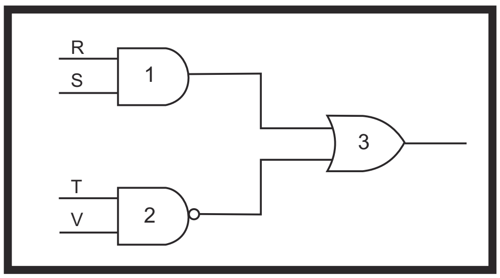
1. $\overline{X}+\overline{Z}$
2. $\overline{XZ}$
3. $\overline{\overline{X}+\overline{Z}}$
4. $\overline{X+\overline{Z}}$

3-36. The Truth Tables are identical for which two gates?

1. C and D
2. A and D
3. B and C
4. A and B

3-37. Which gate represents the output expression $\overrightarrow{\overline{X}Z}$?

1. A
2. B
3. C
4. D



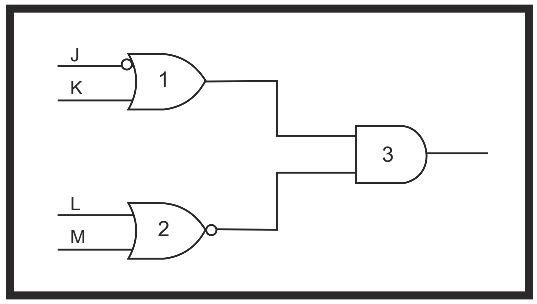
NTS13A310

Figure 3G. —Logic circuit.

IN ANSWERING QUESTION 3-38, REFER TO FIGURE 3G.

3-38. Which of the following output expressions represents the output of gate 3?

1. (RS) (TV)
2. RS + $\overline{TV}$
3. (RS) + ($\overline{TV}$)
4. ($\overline{RS}$) +(TV)



NTS13A311

Figure 3H. —Logic circuit.

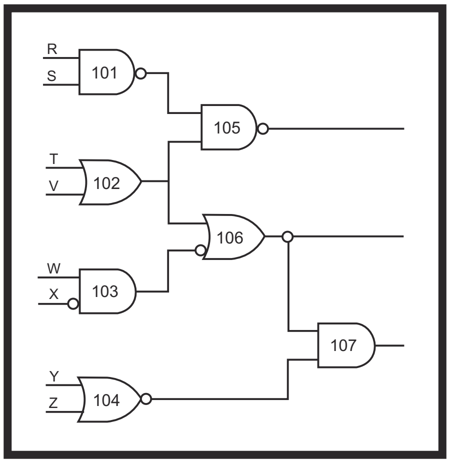
IN ANSWERING QUESTIONS 3-39 AND 3-40, REFER TO FIGURE 3H.

3-39. The output expression JK represents the output of which, if any, of the following gates?

1. 1
2. 2
3. 3
4. None of the above

3-40. What is the output expression for gate 3?

1. $\overline{J}K(\overline{L}+\overline{M})$
2. $(\overline{J}+K)\overline{(L+M)}$
3. $(\overline{J}+K)+(\overline{L}+\overline{M})$
4. $K(\overline{J}+1)(L\overline{M})$



NTS13A312

Figure 3I. —Logic circuit.

IN ANSWERING QUESTIONS 3-41 THROUGH 3-45, REFER TO FIGURE 3I.

3-41. What is the output expression for gate 105?

1. $(\overline{RS}) + (T + V)$
2. RSTV
3. $\overrightarrow{(\overline{RS})(T + V)}$
4. $(\overline{RS}) + (\overline{T + V})$

3-42. Which of the following gates provides a common output to two other gates?

1. 101
2. 102
3. 104
4. 106

3-43. Which of the following expressions represents the output of gate 106?

1. $W\overline{X}TV$
2. $W + \overline{X} + T + V$
3. $(TV) + (W\overline{X})$
4. $(T + V) + (\overline{W\overline{X}})$

3-44. Which of the following conditions will cause gate 107 output to be HIGH?

1. Gate 106 is LOW; Y and Z are HIGH
2. Gate 106 is LOW; Y is HIGH, Z is LOW
3. Gate 106 is HIGH; Y and Z are LOW
4. Gate 106 is HIGH; Y is LOW, Z is HIGH

3-45. What is the output expression for gate 107?

1. $(TV + W\overline{X})(Y + Z)$
2. $(\overrightarrow{T + V + W + \overline{X}})(YZ)$
3. $((T+V)+(\overline{W}X))(\overline{Y+Z})$
4. $(T + V)(\overline{W\overline{X}})(Y + Z)$

3-46. Boolean algebra is used primarily by which of the following groups?

1. Fabricators
2. Technicians
3. Design engineers
4. Repair personnel

3-47. Which of the following Boolean laws states, "a term that is TRUE in one part of an expression will be TRUE in all parts of the expression"?

1. Identity
2. Commutative
3. Complementary
4. Double negative

3-48. The examples AB = BA and A+B = B+A represent which of the following Boolean laws?

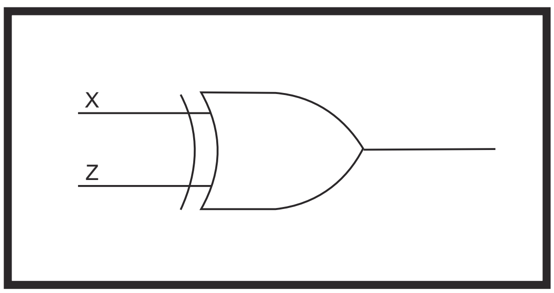
1. Associative
2. Commutative
3. Intersection
4. Union

3-49. What type of logic gate is modified to produce an exclusive OR gate?

1. AND
2. NAND
3. OR
4. NOR

3-50. Which of the following symbols represents the operation function of an exclusive OR gate?

1. Θ
2. $\oplus$
3. O
4. $\otimes$



NTS13A313

Figure 3J. —Logic gate.

IN ANSWERING QUESTIONS 3-51 AND 3-52, REFER TO FIGURE 3J.

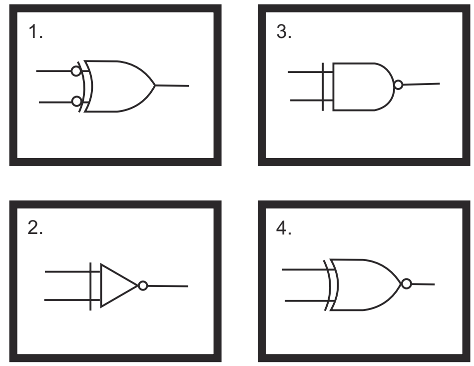
3-51. What will be the output of the gate when (a) X is HIGH and Z is LOW, and (b) X and Z are both HIGH?

1. (a) LOW (b) LOW
2. (a) LOW (b) HIGH
3. (a) HIGH (b) LOW
4. (a) HIGH (b) HIGH

3-52. What will be the output of the gate when (a) X and Z are both LOW and (b) X is LOW and Z is HIGH?

1. (a) LOW (b) LOW
2. (a) LOW (b) HIGH
3. (a) HIGH (b) LOW
4. (a) HIGH (b) HIGH

3-53. Which of the following symbols represents an exclusive NOR gate?



NTS13A314

3-54. What is the output of an exclusive NOR gate when (a) all inputs are LOW and (b) all inputs are HIGH?

1. HIGH (b) HIGH
2. LOW (b) HIGH
3. HIGH (b) LOW
4. LOW (b) LOW

3-55. What is the output expression for an exclusive NOR gate with R and T as inputs?

1. $\overline{R \oplus T}$
2. $\overline{R} + \overline{T}$
3. $\overline{R + T}$
4. $\overline{R} \oplus \overline{T}$

ASSIGNMENT 4

Textbook assignment: Chapter 4, "Special Logic Circuits," pages 3-3 through 3-67.

4-1. Which of the following circuits will add two binary digits but not produce a carry?

1. AND gate
2. Half adder
3. Quarter adder
4. Summation amplifier

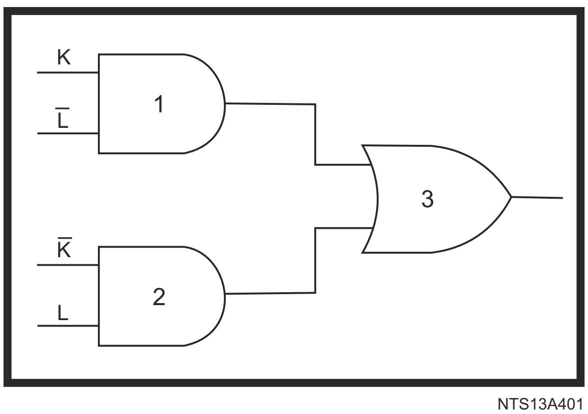


Figure 4A.—Logic circuit.

IN ANSWERING QUESTIONS 4-2 THROUGH 4-5, REFER TO FIGURE 4A.

4-2. What type of circuit is depicted in the figure?

1. Inverter amplifier
2. Half subtracter
3. Half adder
4. Quarter adder

4-3. Which of the following gates will have HIGH outputs when K is HIGH and L is LOW?

1. 1 only
2. 3 only
3. 1 and 2
4. 1 and 3

4-4. What will be the output of gate 3 when K and L are both HIGH?

1. LOW
2. HIGH

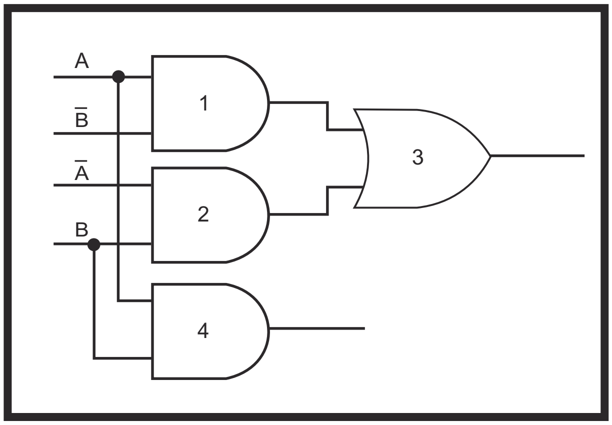
4-5. Assume that K and L are both HIGH. Which of the following statements represents the output of gate 3?

1. 1 plus 1 = 0 (No carry)
2. 1 plus 0 = 1
3. 0 plus 1 = 1
4. 0 plus 0 = 0

4-6. Which of the following gates performs the same function as a quarter adder?

1. Half adder
2. Exclusive OR
3. NOR
4. Exclusive NOR

THIS SPACE LEFT BLANK INTENTIONALLY.



NTS13A402

Figure 4B.—Logic circuit.

IN ANSWERING QUESTIONS 4-7 THROUGH 4-9, REFER TO FIGURE 4B.

4-7. Which of the following circuits is shown in the figure?

1. Quarter adder
2. Half adder
3. Full adder
4. Subtracter

4-8. Which of the following gate combinations may be replaced with an exclusive OR gate?

1. 1, 2, and 3
2. 1, 2, and 4
3. 1, 3, and 4
4. 2, 3, and 4

4-9. The output of gate 3 is the (a) and the output of gate 4 is the (b).

1. (a) Carry (b) sum
2. (a) Sum (b) difference
3. (a) Sum (b) carry
4. (a) Carry (b) difference

4-10. What is the largest sum that can be obtained from a half adder?

1. 01_2
2. 10_2
3. 11_2
4. 100_2

4-11. Which of the following statements describes the difference between a half adder and a full adder?

1. A half adder produces a carry
2. A full adder produces a carry
3. A half adder will add a carry from another circuit
4. A full adder will add a carry from another circuit

THIS SPACE LEFT BLANK INTENTIONALLY.

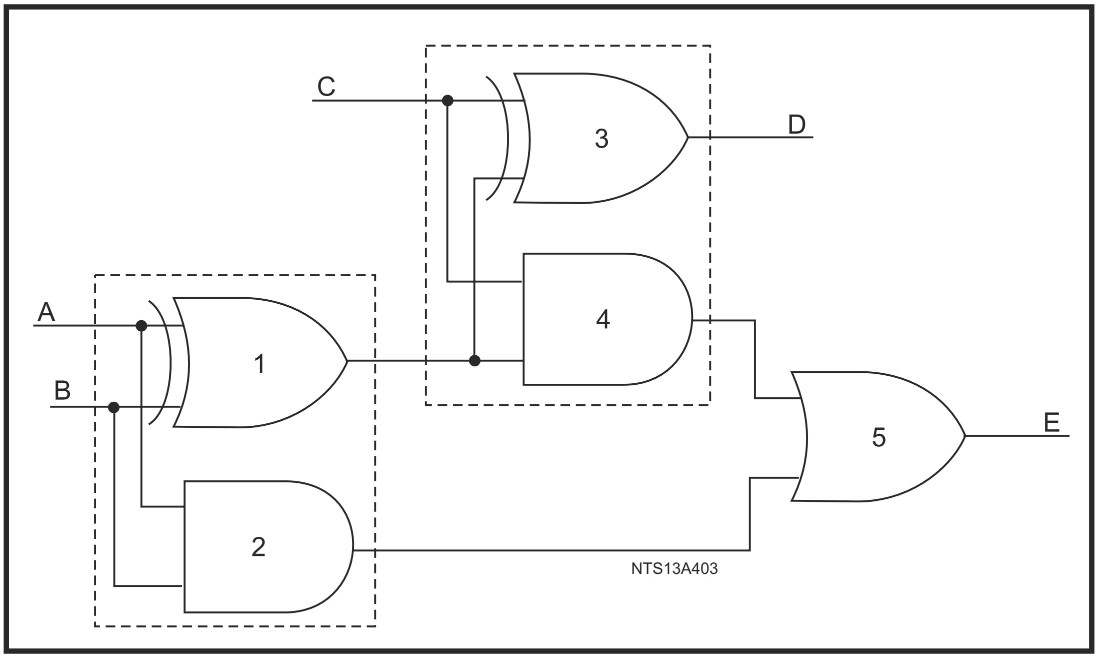


Figure 4C.—Logic circuit.

IN ANSWERING QUESTIONS 4-12 THROUGH 4-14, REFER TO FIGURE 4C.

4-12. What type of circuit is shown in the figure?

1. Full adder
2. Subtracter adder
3. Double half adder
4. Double exclusive OR

4-13. If A and B are 0 and C is 1, what will be the outputs at D and E?

1. D = 0, E = 0
2. D = 0, E = 1
3. D = 1, E = 0
4. D = 1, E = 1

4-14. Under which of the following conditions will outputs D and E both be HIGH?

1. A, B, and C are HIGH
2. A and B are HIGH, C is LOW
3. A and B are LOW, C is HIGH
4. A, B, and C are LOW

4-15. What is the largest sum that can be obtained from a full adder?

1. 10_2
2. 11_2
3. 100_2
4. 111_2

4-16. How many full adders are required to form a parallel adder capable of adding 10001_2 and 1000_2?

1. Five
2. Six
3. Three
4. Four

4-17. Which of the following statements describes the method used in computers to subtract binary numbers?

1. R's complement the minuend and add to the subtrahend
2. Add the minuend and subtrahend and complement the sum
3. R's complement the subtrahend and add to the minuend
4. Subtract the subtrahend from the minuend and complement the difference

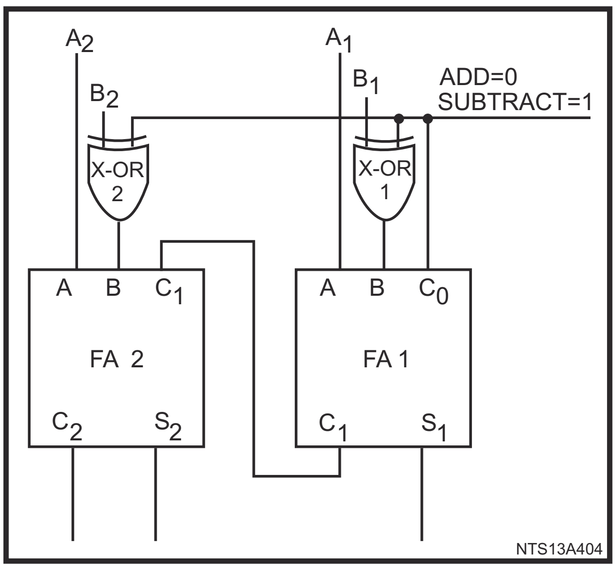


Figure 4D.—Adder/subtracter circuit.

IN ANSWERING QUESTIONS 4-18 THROUGH 4-20, REFER TO FIGURE 4D.

4-18. What type of complement is performed by X-OR gates 1 and 2?

1. R's complement
2. Minuend complement
3. R's + 1 complement
4. Difference complement

4-19. Which of the following inputs is used for the least significant digit of the minuend?

1. A_1
2. A_2
3. B_1
4. B_2

4-20. What will be the output of (a) X-OR 2 and (b) X-OR 1 in the subtract mode with a subtrahend of 10_2?

1. (a) 0 (b) 0
2. (a) 0 (b) 1
3. (a) 1 (b) 0
4. (a) 1 (b) 1

4-21. Flip-flops are what type of multivibrators?

1. Astable
2. Monostable
3. Free running
4. Bistable

4-22. Flip-flops may NOT be used for which of the following operations?

1. Temporary storage
2. Subtraction
3. Division
4. Transfer of information

4-23. When, if ever, will the outputs Q and $\overline{Q}$ of an R-S flip-flop be the same?

1. When R and S are both LOW
2. When R is LOW and S is HIGH
3. When R is HIGH and S is LOW
4. Never

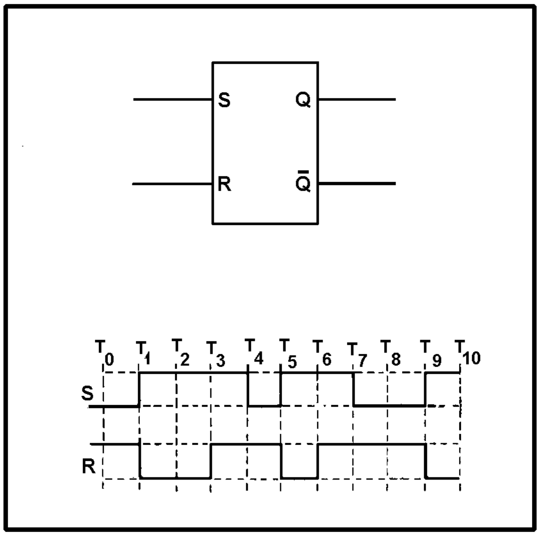


Figure 4E.—R-S flip-flop with timing diagram.

WHEN ANSWERING QUESTIONS 4-24 AND 4-25, REFER TO FIGURE 4E.

4-24. Assume the flip-flop is set at T_0. At which of the following times will the flip-flop be reset?

1. T_1 to T_3, T_5 to T_6, and T_9 to T_{10}
2. T_0 to T_1, T_3 to T_5, and T_6 to T_9
3. T_1 to T_3, T_5 to T_7, and T_9 to T_{10}
4. T_1 to T_4, T_5 to T_7, and T_9 to T_{10}

4-25. What happens to the flip-flop at T_6?

1. It sets
2. It resets
3. It sets and immediately resets
4. It remains reset

4-26. Which of the following statements describes a toggle flip-flop?

1. A monostable device
2. An astable device that changes state only on a set pulse
3. A two input bistable device
4. A bistable device with a single input

4-27. A T flip-flop is used primarily for which of the following functions?

1. To divide the input frequency by two
2. To double the input frequency
3. To amplify the input frequency
4. To invert the input frequency

4-28. What are the inputs to a D flip-flop?

1. Set and reset
2. Set and clock
3. Data and clock
4. Reset and data

4-29. What is the purpose of a D flip-flop?

1. To eliminate the output of the equipment
2. To divide the data input by the clock frequency
3. To store data until it is needed
4. To toggle the data input

4-30. An inverter on the clock input has which of the following effects on the D flip-flop?

1. The output will change on the negative-going transition of the clock pulse
2. The output will change on the positive-going transition of the clock pulse
3. The data input will change at the clock frequency
4. The output will change at the clock frequency

4-31. Which of the following statements is true concerning CLR and PR pulses to the D flip-flop?

1. CLR causes Q to go high, PR causes Q to go low
2. CLR and PR override any existing output condition
3. Other inputs override CLR and PR
4. CLR and PR have no effect on the output

4-32. Which of the following statements is correct concerning D flip-flops?

1. The output is delayed up to one clock pulse
2. Input data is delayed until it coincides with the clock
3. The clock is delayed until it coincides with the input data
4. The output is always a square wave

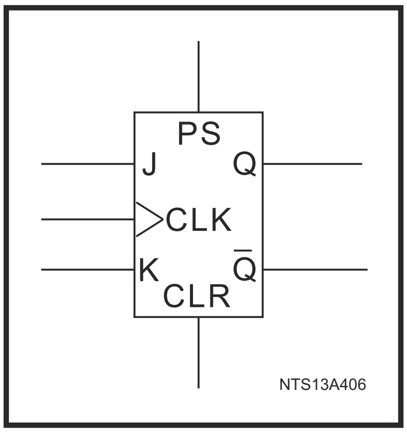


Figure 4F.—Standard symbol for a J-K flip-flop.

WHEN ANSWERING QUESTIONS 4-33 THROUGH 4-36, REFER TO FIGURE 4F.

4-33. The flip-flop shown in the figure may be used in place of which of following flip-flops?

1. R-S
2. T
3. D
4. Each of the above

4-34. With the clock applied and J and K inputs held HIGH, what is the output at Q?

1. Constant HIGH
2. Constant LOW
3. Toggle at one half the clock frequency
4. Toggle at twice the clock frequency

4-35. What will be the $\overline{Q}$ output if K is HIGH and CLK goes HIGH?

1. HIGH
2. LOW

4-36. A pulse on which of the following inputs will cause the flip-flop to set regardless of the other inputs?

1. CLK
2. CLR
3. J or K
4. PR or PS

4-37. The circuit which generates a timing signal to control operations is called a/an

1. clock
2. counter
3. oscillator
4. bistable multivibrator

4-38. Which of the following statements is true regarding astable multivibrators used as clocks?

1. As multivibrator frequency increases, stability decreases
2. Output 2 will have a higher voltage than output 1
3. The frequency stability may be increased by applying a higher frequency trigger
4. A trigger of lower frequency will stabilize the output frequency

4-39. Which of the following types of circuits will produce a stable clock when triggered by an outside source?

1. R-S flip-flop
2. Bistable multivibrator
3. One-shot multivibrator
4. D flip-flop

4-40. Which of the following types of clocks would probably be used in a complex piece of equipment with a variety of timing requirements?

1. Single triggered-monostable
2. Single free-running
3. Triggered-astable for each section of the equipment
4. Multiphase

4-41. What is the modulus of a 3-stage binary counter?

1. 7
2. 8
3. 3
4. 4

4-42. Counters may be used for which of the following purposes?

1. Counting operations, quantities and time
2. Dividing frequency
3. Addressing information in storage
4. Each of the above

THIS SPACE LEFT BLANK INTENTIONALLY

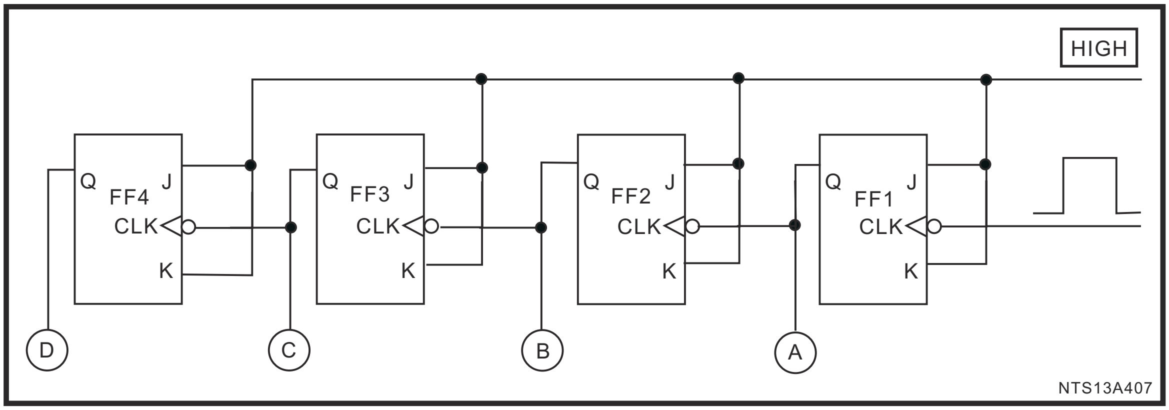


Figure 4G.—Standard symbol circuit.

WHEN ANSWERING QUESTIONS 4-43 THROUGH 4-46, REFER TO FIGURE 4G. FOR EACH QUESTION, ASSUME THAT ALL FFs ARE INITIALLY RESET.

4-43. What type of circuit is shown?

1. J-K flip-flops
2. Clock
3. Shift register
4. Ripple counter

4-44. Assume all lights are out. After 6 input pulses, which two lamps will be lit?

1. A and B
2. A and C
3. B and C
4. B and D

4-45. Assume all lamps are out. After 16 input pulses, which lamps, if any, will be lit?

1. A, B, C, and D
2. B, C, and D
3. A, B, and D
4. None

4-46. What is the main disadvantage of using this circuit with a high frequency input?

1. The circuit will burn up
2. Possible errors in the output
3. The flip-flops will act as T flip-flops
4. Above a certain frequency the circuit will count in the opposite direction

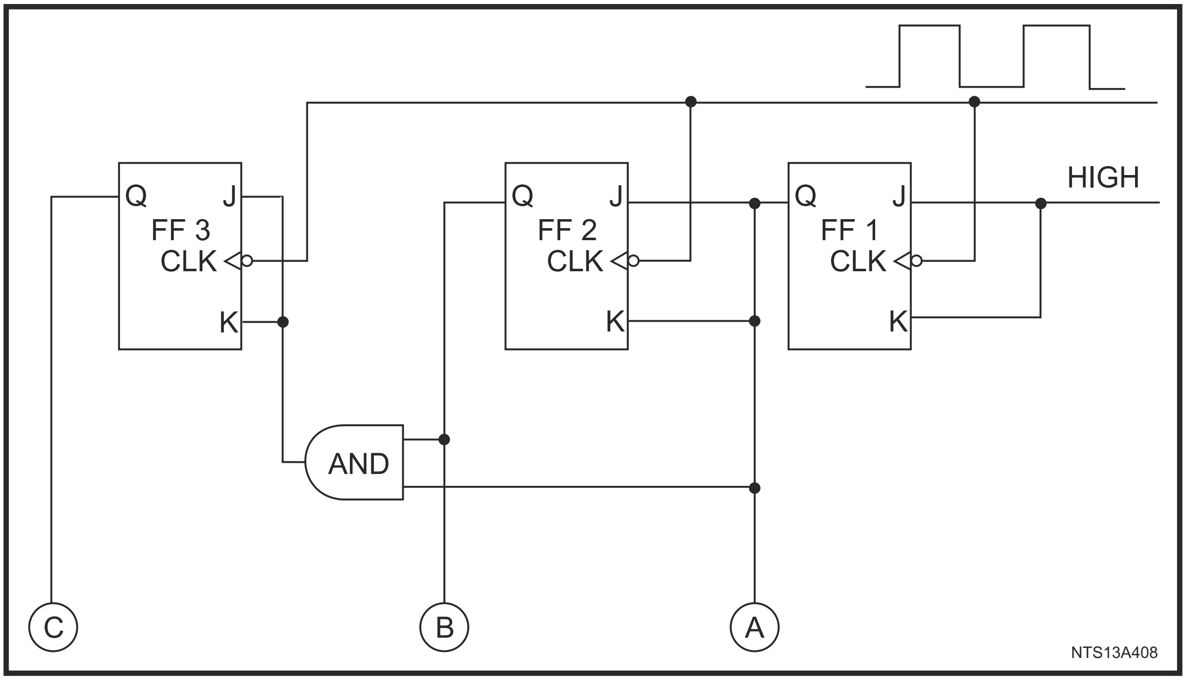


Figure 4H.—Counter circuit.

WHEN ANSWERING QUESTIONS 4-47 THROUGH 4-50, REFER TO FIGURE 4H.

4-47. What is the maximum count that this counter is capable of holding?

1. 5_8
2. 7_8
3. 3_8
4. 4_8

4-48. What type of counter is shown?

1. Ring
2. Asynchronous
3. Synchronous
4. Decade

4-49. Under which of the following conditions will the output of the AND gate be HIGH?

1. FF 1 and FF 2 are set
2. FF 1 and FF 2 are reset
3. FF 2 and FF 3 are set
4. FF 1 and FF 3 are reset

4-50. With all the FFs initially reset, the AND gate output will be HIGH after how many input pulses?

1. 3 only
2. 4 only
3. 7 only
4. 3 and 7

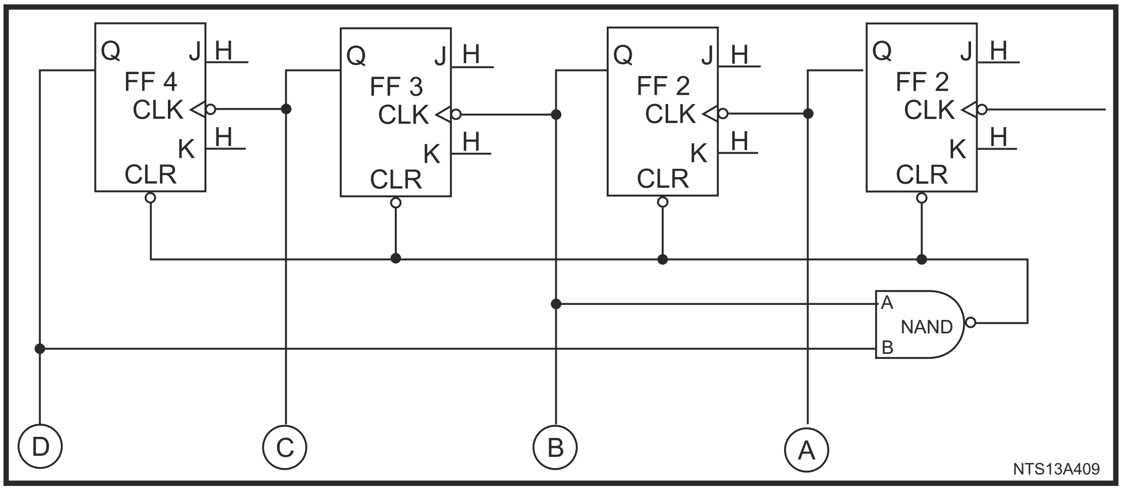


Figure 4I.—Counter circuit.

WHEN ANSWERING QUESTIONS 4-51 AND 4-52, REFER TO FIGURE 4I.

4-51. What is the maximum binary count that will be shown before all the flip-flops reset?

1. 111_2
2. 1010_2
3. 1111_2
4. 1000_2

4-52. To change the maximum count from 10_{10} to 9_{10}, which two flip-flops would be wired as NAND gate inputs?

1. FF 1 and FF 2
2. FF 2 and FF 3
3. FF 1 and FF 4
4. FF 1 and FF 3

THIS SPACE LEFT BLANK INTENTIONALLY.

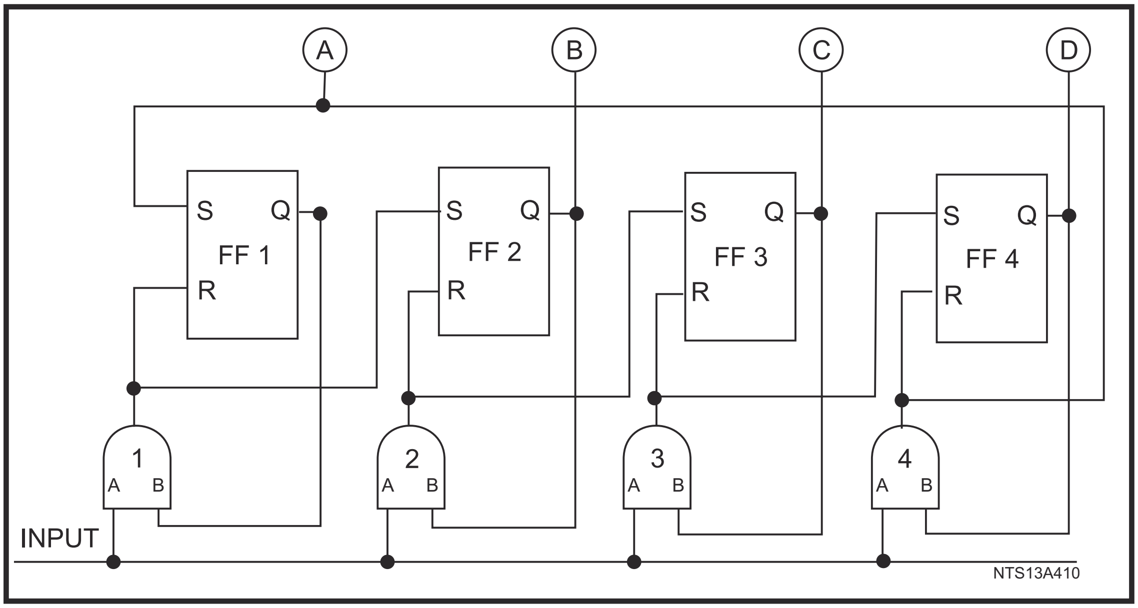


Figure 4J.—Counter circuit.

WHEN ANSWERING QUESTIONS 4-53 THROUGH 4-55, REFER TO FIGURE 4J.

4-53. Which of the following types of counters is shown in the figure?

1. Ring
2. Asynchronous
3. Synchronous
4. Decade

4-54. At any given time, how many flip-flops may be set?

1. Only one
2. Any two
3. Any three
4. All

4-55. Which of the following conditions must exist to set FF 3?

1. AND gate 1 output HIGH
2. AND gate 2 output HIGH
3. AND gate 3 output LOW
4. Clock input to AND gate 3 LOW and AND gate 4 output HIGH

THIS SPACE LEFT BLANK INTENTIONALLY

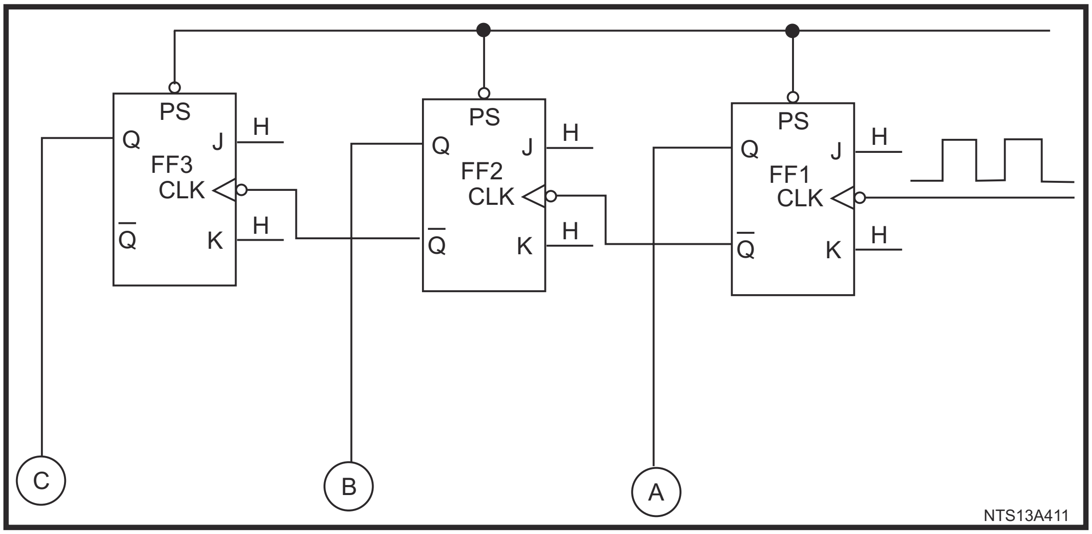


Figure 4K.—Down counter.

WHEN ANSWERING QUESTIONS 4-56 AND 4-57, REFER TO FIGURE 4K.

4-56. Assume only FF 1 and FF 3 are set. Which, if any, of the flip-flops will be set after the next clock pulse?

1. FF 2 only
2. FF 3 only
3. FF 2 and FF 3
4. None

4-57. Assume that all FFs are set, which of the following actions will take place at clock pulse 4?

1. FF 1, FF 2 and FF 3 will set
2. FF 1 will set, FF 2 and FF 3 will reset
3. FF 1, FF 2, and FF 3 will reset
4. FF 1 and FF 2 will set, FF 3 will reset

4-58. What term identifies a series of FFs designed to temporarily store information?

1. Data word
2. Counter
3. Register
4. DIP package

4-59. Which of the following statements describes parallel transfer?

1. Data is transferred one bit at a time
2. All data bits are transferred simultaneously
3. Data is received in serial form and transferred in parallel form
4. Data is transferred on a single line

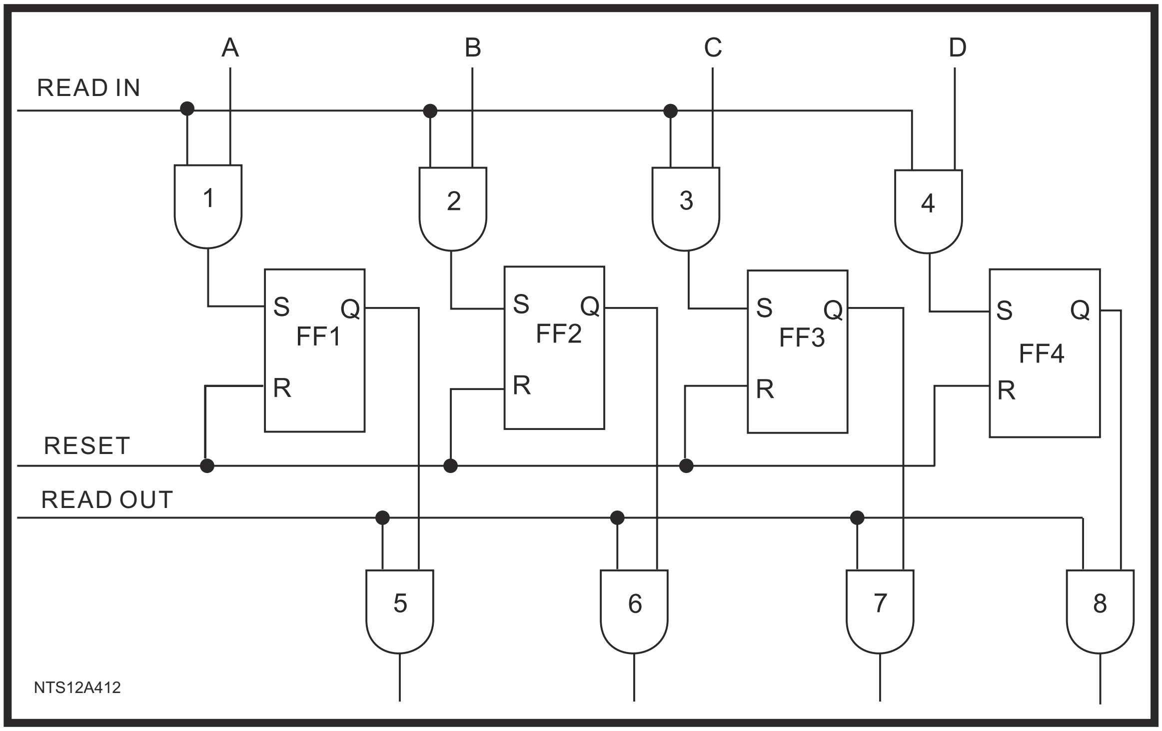


Figure 4L.—Parallel register.

WHEN ANSWERING QUESTIONS 4-60 AND 4-61, REFER TO FIGURE 4L.

4-60. Which of the following methods will clear the register of old information no longer needed?

1. A HIGH applied to the READ IN line
2. A HIGH applied to the RESET line
3. A LOW applied on the RESET and a HIGH on the READ OUT lines
4. HIGHs are applied on the A, B, C, D, and READ IN lines

4-61. Under which of the following conditions will the output of gate 7 be HIGH?

1. When FF 3 and RESET are HIGH
2. When gate 3 and FF 3 are HIGH
3. When FF 3 and READ OUT are HIGH
4. When the output of gate 3 is HIGH

THIS SPACE LEFT BLANK INTENTIONALLY

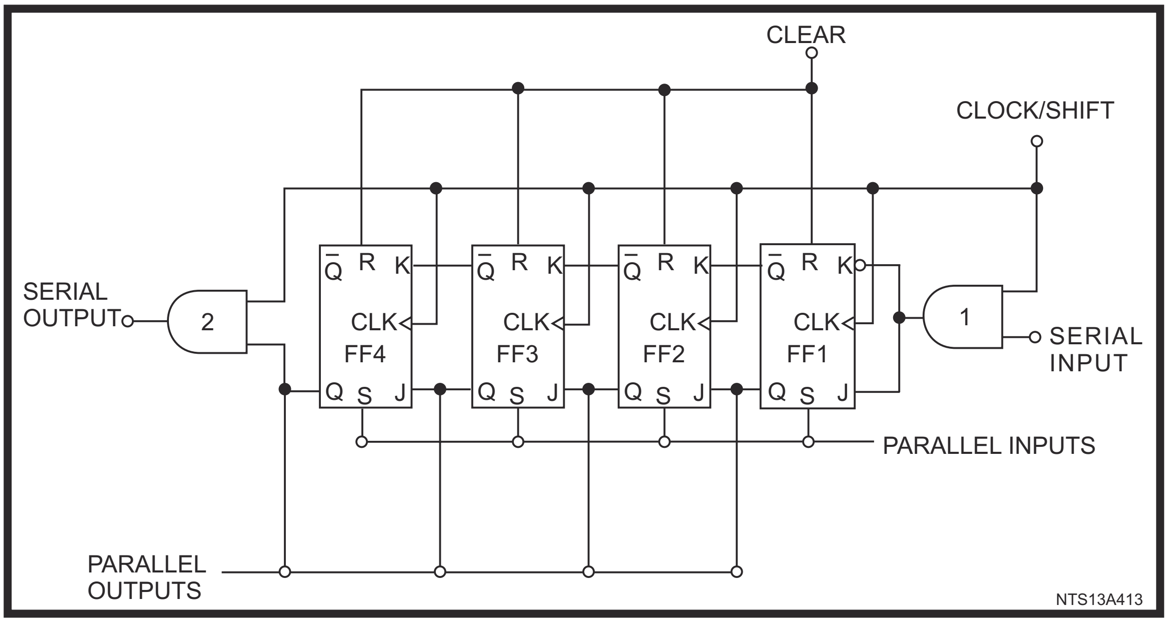


Figure 4M.—Shift register.

WHEN ANSWERING QUESTIONS 4-62 THROUGH 4-67, REFER TO FIGURE 4M.

4-62. Which of the following operations is the circuit capable of performing?

1. Serial-to-parallel conversion
2. Parallel-to-serial conversion
3. Left shifts
4. Each of the above

4-63. What is the maximum word length the register is capable of handling?

1. 6 bit
2. 5 bit
3. 3 bit
4. 4 bit

4-64. How many clock/shift pulses are required to serially input a word into the register?

1. 1
2. 5
3. 3
4. 4

4-65. To output a word in parallel form, how many output lines are required?

1. 1
2. 2
3. 4
4. 8

4-66. To increase the value of a word by one power of 2, how many shifts are required?

1. 1
2. 2
3. 3
4. 0

4-67. Which of the following operations takes the longest time?

1. Serial in, parallel out
2. Serial in, serial out
3. Parallel in, serial out
4. Parallel in, parallel out

4-68. Shifting a word in a shift register 3 places to the left is equal to multiplying the number by how much?

1. 10_{10}
2. 2_{10}
3. 8_{10}
4. 4_{10}

4-69. Logic families are identified by which of the following means?

1. Logic polarity required
2. The types of elements used
3. Size and cost of manufacture
4. Packaging (DIP, TO, flat packs)

4-70. A TTL logic circuit would use which of the following types of elements?

1. Diode-transistor
2. Resistor-transistor
3. Transistor-transistor
4. Complementary metal oxide semiconductors

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 14—Introduction to Microelectronics

NAVEDTRA 14186

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

CREDITS

The illustrations listed below are included through the courtesy of the designated source. Permission to use these illustrations is gratefully acknowledged. Permission to reproduce illustrations and other materials in this publication must be obtained from the source.

SOURCE	FIGURE
Siliconix, Inc.	1-33
Pana Vise Products, Inc. (former company name: Harris Semiconductors)	1-34

Permission to use the trademark “PANAVISE” by Pana Vise Products, Inc. is gratefully acknowledged.

The 2M section of the Education and Training Division, Naval Air Reward Facility, Pensacola, Florida, and the Naval Undersea Warfare Engineering Center, Keyport, Washington, provided many of the figures included in this NRTC. Their assistance is gratefully acknowledged.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 14 of a series.

History of the course:

Sep 1998: Original edition released.

Sep 2003: Administrative update released. Technical content reviewed and revised by ETC(SW) *Jack Weatherford.*

NAVSUP Logistics Tracking Number
0504-LP-026-8390

TABLE OF CONTENTS

ASSIGNMENT QUESTIONS follow Index.

CHAPTER 1

MICROELECTRONICS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each topic. These learning objectives serve as a preview of the information you are expected to learn in the topic. The comprehensive check questions are based on the objectives. By successfully completing the OCC-ECC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completion of this topic, you will be able to:

1. Outline the progress made in the history of microelectronics.
2. Describe the evolution of microelectronics from point-to-point wiring through high element density state-of-the-art microelectronics.
3. List the advantages and disadvantages of point-to-point wiring and high element density state-of-the-art microelectronics.
4. Identify printed circuit boards, diodes, transistors, and the various types of integrated circuits. Describe the fabrication techniques of these components.
5. Define the terminology used in microelectronic technology including the following terms used by the Naval Systems Commands:
    a. microelectronics
    b. microcircuit
    c. microcircuit module
    d. miniature electronics
    e. system packaging
    f. levels of packaging (0 to IV)
    g. modular assemblies
    h. cordwood modules
    i. micromodules
6. Describe typical packaging levels presently used for microelectronic systems.
7. Describe typical interconnections used in microelectronic systems.
8. Describe environmental considerations for microelectronics.

INTRODUCTION

In *NEETS, Module 6*, *Introduction to Electronic Emission, Tubes, and Power Supplies*, you learned that Thomas Edison's discovery of thermionic emission opened the door to electronic technology. Progress was slow in the beginning, but each year brought new and more amazing discoveries. The development of vacuum tubes soon led to the simple radio. Then came more complex systems of communications. Modern systems now allow us to communicate with other parts of the world via satellite. Data is now collected from space by probes without the presence of man because of microelectronic technology.

Sophisticated control systems allow us to operate equipment by remote control in hazardous situations, such as the handling of radioactive materials. We can remotely pilot aircraft from takeoff to landing. We can make course corrections to spacecraft millions of miles from Earth. Space flight, computers, and even video games would not be possible except for the advances made in microelectronics.

The most significant step in modern electronics was the development of the transistor by Bell Laboratories in 1948. This development was to solid-state electronics what the Edison Effect was to the vacuum tube. The solid-state diode and the transistor opened the door to microelectronics.

MICROELECTRONICS is defined as that area of technology associated with and applied to the realization of electronic systems made of extremely small electronic parts or elements. As discussed in topic 2 of *NEETS, Module 7*, *Introduction to Solid-State Devices and Power Supplies*, the term *microelectronics* is normally associated with integrated circuits (IC). Microelectronics is often thought to include only integrated circuits. However, many other types of circuits also fall into the microelectronics category. These will be discussed in greater detail under solid-state devices later in this topic.

During World War II, the need to reduce the size, weight, and power of military electronic systems became important because of the increased use of these systems. As systems became more complex, their size, weight, and power requirements rapidly increased. The increases finally reached a point that was unacceptable, especially in aircraft and for infantry personnel who carried equipment in combat. These unacceptable factors were the driving force in the development of smaller, lighter, and more efficient electronic circuit components. Such requirements continue to be important factors in the development of new systems, both for military and commercial markets. Military electronic systems, for example, continue to become more highly developed as their capability, reliability, and maintainability is increased. Progress in the development of military systems and steady advances in technology point to an ever-increasing need for increased technical knowledge of microelectronics in your Navy job.

Q1. What problems were evident about military electronic systems during World War II?

Q2. What discovery opened the door to solid-state electronics?

Q3. *What is microelectronics?*

EVOLUTION OF MICROELECTRONICS

The earliest electronic circuits were fairly simple. They were composed of a few tubes, transformers, resistors, capacitors, and wiring. As more was learned by designers, they began to increase both the size and complexity of circuits. Component limitations were soon identified as this technology developed.

VACUUM-TUBE EQUIPMENT

Vacuum tubes were found to have several built-in problems. Although the tubes were lightweight, associated components and chassis were quite heavy. It was not uncommon for such chassis to weigh 40 to 50 pounds. In addition, the tubes generated a lot of heat, required a warm-up time from 1 to 2 minutes, and required hefty power supply voltages of 300 volts dc and more.

No two tubes of the same type were exactly alike in output characteristics. Therefore, designers were required to produce circuits that could work with any tube of a particular type. This meant that additional components were often required to tune the circuit to the output characteristics required for the tube used.

Figure 1-1 shows a typical vacuum-tube circuit. The actual size of the transformer is approximately $4 \times 4 \times 3$ inches. Capacitors are approximately 1×3 inches. The components in the figure are very large when compared to modern microelectronics.

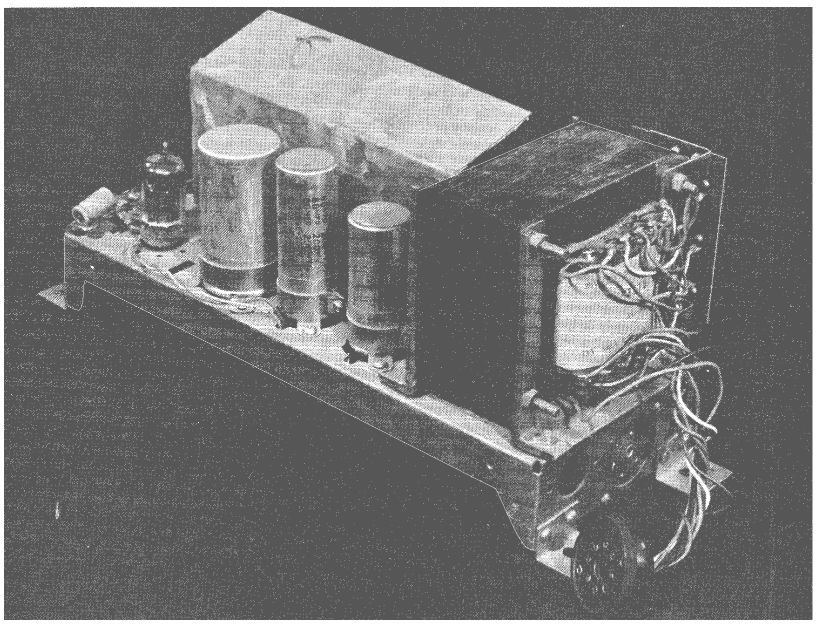


Figure 1-1.—Typical vacuum-tube circuit.

A circuit could be designed either as a complete system or as a functional part of a larger system. In complex systems, such as radar, many separate circuits were needed to accomplish the desired tasks. Multiple-function tubes, such as dual diodes, dual triodes, tetrodes, and others helped considerably to reduce the size of circuits. However, weight, heat, and power consumption continued to be problems that plagued designers.

Another major problem with vacuum-tube circuits was the method of wiring components referred to as POINT-TO-POINT WIRING. Figure 1-2 is an excellent example of point-to-point wiring. Not only did this wiring look like a rat's nest, but it often caused unwanted interactions between components. For example, it was not at all unusual to have inductive or capacitive effects between wires. Also, point-to-point wiring posed a safety hazard when troubleshooting was performed on energized circuits because of exposed wiring and test points. Point-to-point wiring was usually repaired with general-purpose test equipment and common hand tools.

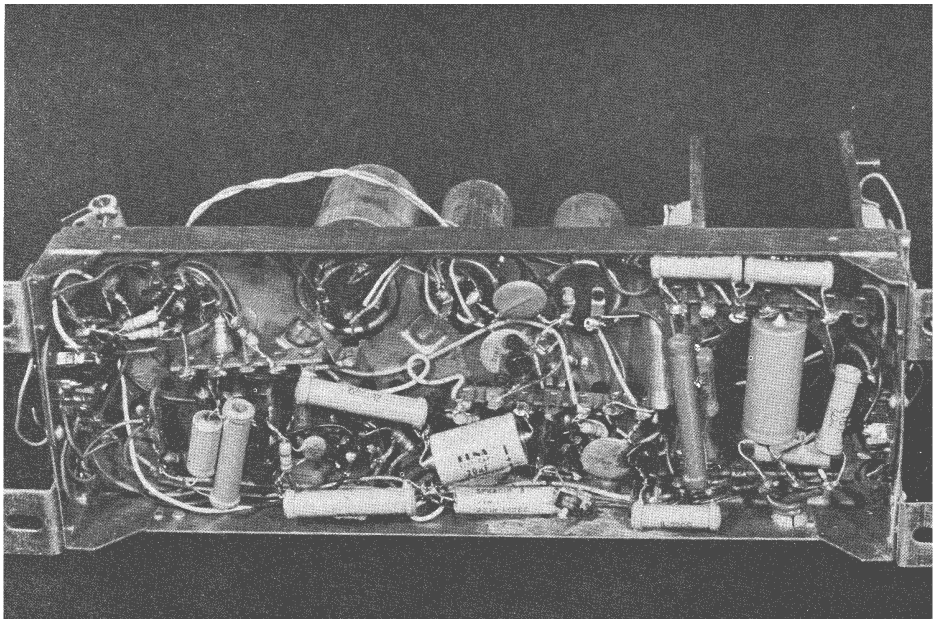


Figure 1-2.—Point-to-point wiring.

Vacuum-tube circuits proved to be reliable under many conditions. Still, the drawbacks of large size, heavy weight, and significant power consumption made them undesirable in most situations. For example, computer systems using tubes were extremely large and difficult to maintain. ENIAC, a completely electronic computer built in 1945, contained 18,000 tubes. It often required a full day just to locate and replace faulty tubes.

In some applications, we are still limited to vacuum tubes. Cathode-ray tubes used in radar, television, and oscilloscopes do not, as yet, have solid-state counterparts.

One concept that eased the technician's job was that of MODULAR PACKAGING. Instead of building a system on one large chassis, it was built of MODULES or blocks. Each module performed a necessary function of the system. Modules could easily be removed and replaced during troubleshooting and repair. For instance, a faulty power supply could be exchanged with a good one to keep the system operational. The faulty unit could then be repaired while out of the system. This is an example of how the module concept improved the efficiency of electronic systems. Even with these advantages, vacuum tube modules still had faults. Tubes and point-to-point wiring were still used and excessive size, weight, and power consumption remained as problems to be overcome.

Vacuum tubes were the basis for electronic technology for many years and some are still with us. Still, emphasis in vacuum-tube technology is rapidly becoming a thing of the past. The emphasis of technology is in the field of microelectronics.

Q4. What discovery proved to be the foundation for the development of the vacuum tube?

Q5. Name the components which greatly increase the weight of vacuum-tube circuitry.

Q6. What are the disadvantages of point-to-point wiring?

Q7. What is a major advantage of modular construction?

Q8. When designing vacuum-tube circuits, you must take what characteristics of tubes into consideration?

SOLID-STATE DEVICES

Now would be a good time for you to review the first few pages of *NEETS, Module 7, Introduction to Solid-State Devices and Power Supplies*, as a refresher for solid-state devices.

The transition from vacuum tubes to solid-state devices took place rapidly. As new types of transistors and diodes were created, they were adapted to circuits. The reductions in size, weight, and power use were impressive. Circuits that earlier weighed as much as 50 pounds were reduced in weight to just a few ounces by replacing bulky components with the much lighter solid-state devices.

The earliest solid-state circuits still relied on point-to-point wiring which caused many of the disadvantages mentioned earlier. A metal chassis, similar to the type used with tubes, was required to provide physical support for the components. The solid-state chassis was still considerably smaller and lighter than the older, tube chassis. Still greater improvements in component mounting methods were yet to come.

One of the most significant developments in circuit packaging has been the PRINTED CIRCUIT BOARD (pcb), as shown in figure 1-3. The pcb is usually an epoxy board on which the circuit leads have been added by the PHOTOETCHING process. This process is similar to photography in that copper-clad boards are exposed to controlled light in the desired circuit pattern and then etched to remove the unwanted copper. This process leaves copper strips (LANDS) that are used to connect the components. In general, printed circuit boards eliminate both the heavy, metal chassis and the point-to-point wiring.

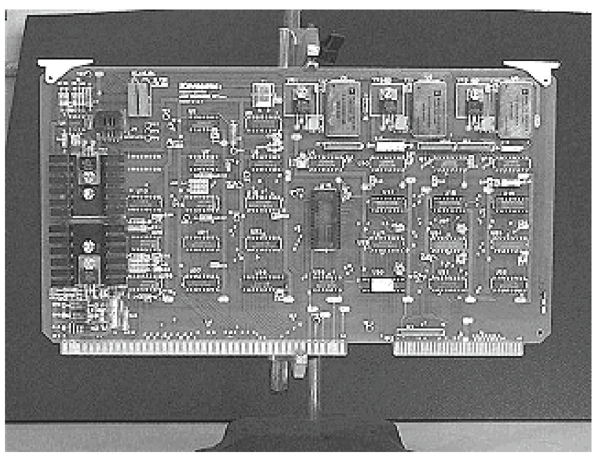


Figure 1-3.—Printed circuit board (pcb).

Although printed circuit boards represent a major improvement over tube technology, they are not without fault. For example, the number of components on each board is limited by the sizes and shapes of components. Also, while vacuum tubes are easily removed for testing or replacement, pcb components are soldered into place and are not as easily removed.

Normally, each pcb contains a single circuit or a subassembly of a system. All printed circuit boards within the system are routinely interconnected through CABLING HARNESSES (groups of wiring or ribbons of wiring). You may be confronted with problems in faulty harness connections that affect system reliability. Such problems are often caused by wiring errors, because of the large numbers of wires in a harness, and by damage to those wires and connectors.

Another mounting form that has been used to increase the number of components in a given space is the CORDWOOD MODULE, as shown in figure 1-4. You can see that the components are placed perpendicular to the end plates. The components are packed very closely together, appearing to be stacked like cordwood for a fireplace. The end plates are usually small printed circuit boards, but may be insulators and solid wire, as shown in the figure. Cordwood modules may or may not be ENCAPSULATED (totally imbedded in solid material) but in either case they are difficult to repair.

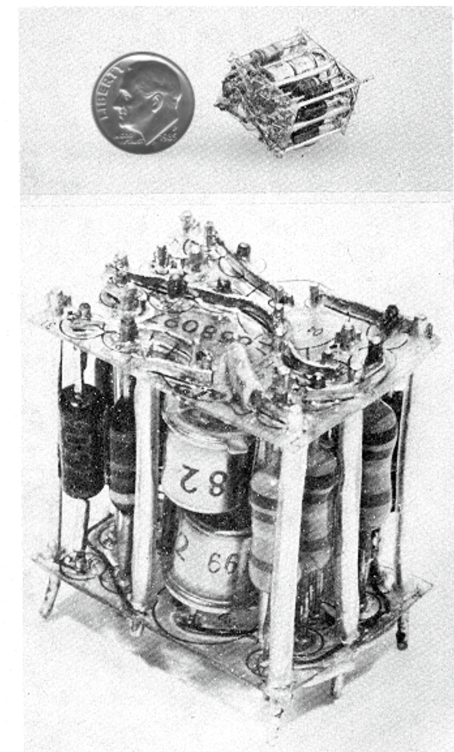


Figure 1-4.—Cordwood module.

Q9. List the major advantages of printed circuit boards.

Q10. What is the major disadvantage of printed circuit boards?

Q11. The ability to place more components in a given space is an advantage of the **_______**.

INTEGRATED CIRCUITS

Many advertisements for electronic equipment refer to integrated circuits or solid-state technology. You know the meaning of the term *solid-state*, but what is an INTEGRATED CIRCUIT? The accepted Navy definition for an integrated circuit is that it consists of elements inseparably associated and formed on or within a single SUBSTRATE (mounting surface). In other words, the circuit components and all interconnections are formed as a unit. You will be concerned with three types of integrated circuits: MONOLITHIC, FILM, and HYBRID.

MONOLITHIC INTEGRATED CIRCUITS are those that are formed completely within a semiconductor substrate. These integrated circuits are commonly referred to as SILICON CHIPS.

FILM INTEGRATED CIRCUITS are broken down into two categories, THIN FILM and THICK FILM. Film components are made of either conductive or nonconductive material that is deposited in desired patterns on a ceramic or glass substrate. Film can only be used as passive circuit components, such as resistors and capacitors. Transistors and/or diodes are added to the substrate to complete the circuit. Differences in thin and thick film will be discussed later in this topic.

HYBRID INTEGRATED CIRCUITS combine two or more integrated circuit types or combine one or more integrated circuit types and DISCRETE (separate) components. Figure 1-5 is an example of a hybrid integrated circuit consisting of silicon chips and film circuitry. The two small squares are chips and the irregularly shaped gray areas are film components.

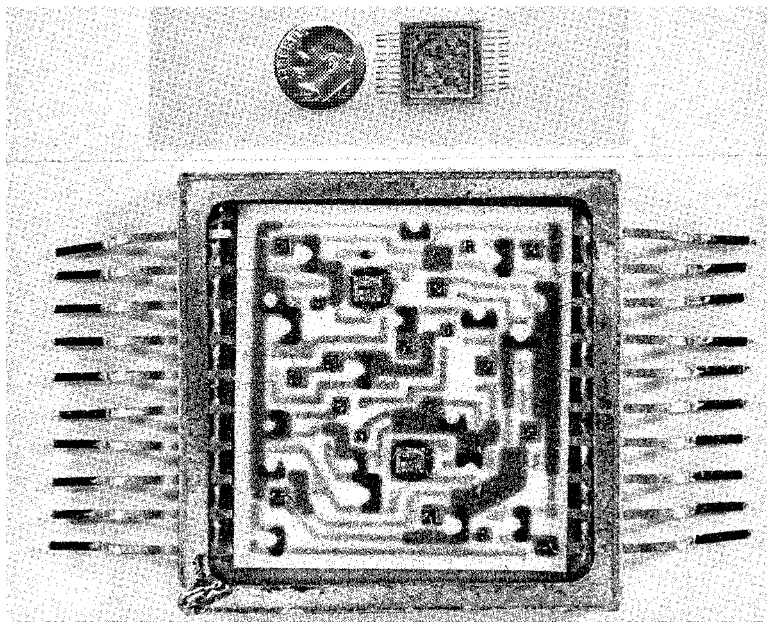


Figure 1-5.—Hybrid integrated circuit.

STATE-OF-THE-ART MICROELECTRONICS

Microelectronic technology today includes thin film, thick film, hybrid, and integrated circuits and combinations of these. Such circuits are applied in DIGITAL, SWITCHING, and LINEAR (analog) circuits. Because of the current trend of producing a number of circuits on a single chip, you may look for further increases in the packaging density of electronic circuits. At the same time you may expect a reduction in the size, weight, and number of connections in individual systems. Improvements in reliability and system capability are also to be expected.

Thus, even as existing capabilities are being improved, new areas of microelectronic use are being explored. To predict where all this use of technology will lead is impossible. However, as the demand for increasingly effective electronic systems continues, improvements will continue to be made in state-of-the-art microelectronics to meet the demands.

LARGE-SCALE INTEGRATION (lsi) and VERY LARGE-SCALE INTEGRATION (vlsi) are the results of improvements in microelectronics production technology. Figure 1-6 is representative of lsi. As shown in the figure, the entire SUBSTRATE WAFER (slice of semiconductor or insulator material) is used instead of one that has been separated into individual circuits. In lsi and vlsi, a variety of circuits can be implanted on a wafer resulting in further size and weight reduction. ICs in modern computers, such as home computers, may contain the entire memory and processing circuits on a single substrate.

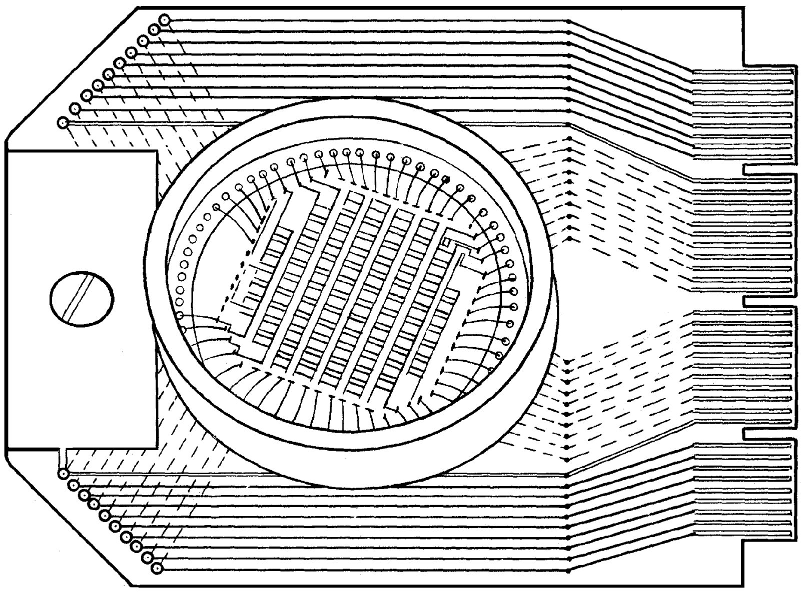


Figure 1-6.—Large-scale integration device (lsi).

Large-scale integration is generally applied to integrated circuits consisting of from 1,000 to 2,000 logic gates or from 1,000 to 64,000 bits of memory. A logic gate, as you should recall from *NEETS, Module 13, Introduction to Number Systems, Boolean Algebra, and Logic Circuits*, is an electronic switching network consisting of combinations of transistors, diodes, and resistors. Very large-scale integration is used in integrated circuits containing over 2,000 logic gates or greater than 64,000 bits of memory.

Q12. Define integrated circuit.

Q13. What are the three major types of integrated circuits?

Q14. How do monolithic ICs differ from film ICs?

Q15. What is a hybrid IC?

Q16. How many logic gates could be contained in lsi?

FABRICATION OF MICROELECTRONIC DEVICES

The purpose of this section is to give you a simplified overview of the manufacture of microelectronic devices. The process is far more complex than will be described here. Still, you will be able to see that microelectronics is not magic, but a highly developed technology.

Development of a microelectronic device begins with a demand from industry or as the result of research. A device that is needed by industry may be a simple diode network or a complex circuit consisting of thousands of components. No matter how complex the device, the basic steps of production are similar. Each type of device requires circuit design, component arrangement, preparation of a substrate, and the depositing of proper materials on the substrate.

The first consideration in the development of a new device is to determine what the device is to accomplish. Once this has been decided, engineers can design the device. During the design phase, the engineers will determine the numbers and types of components and the interconnections, needed to complete the planned circuit.

COMPONENT ARRANGEMENT

Planning the component arrangement for a microelectronic device is a very critical phase of production. Care must be taken to ensure the most efficient use of space available. With simple devices, this can be accomplished by hand. In other words, the engineers can prepare drawings of component placement. However, a computer is used to prepare the layout for complex devices. The computer is able to store the characteristics of thousands of components and can provide a printout of the most efficient component placement. Component placement is then transferred to extremely large drawings. During this step, care is taken to maintain the patterns as they will appear on the substrate. Figure 1-7 shows a fairly simple IC MASK PATTERN. If this pattern were being prepared for production, it would be drawn several hundred times the size shown and then photographed. The photo would then be reduced in size until it was the actual desired size. At that time, the pattern would be used to produce several hundred patterns that would be used on one substrate. Figure 1-8 illustrates how the patterns would be distributed to act as a WAFER MASK for manufacturing.

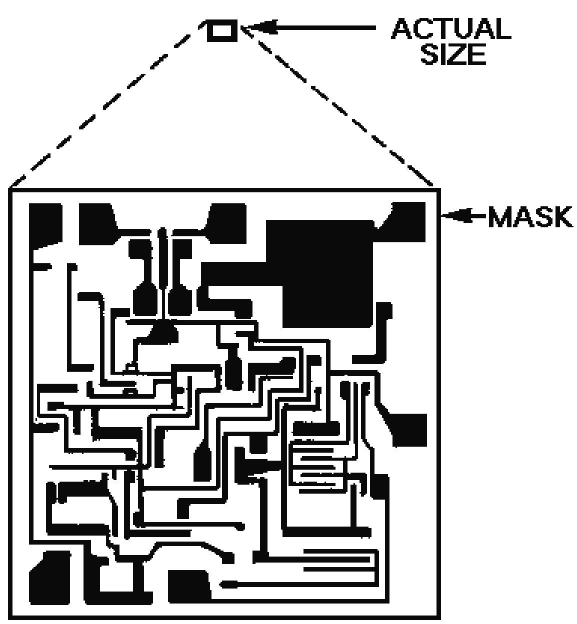


Figure 1-7.—IC mask pattern.

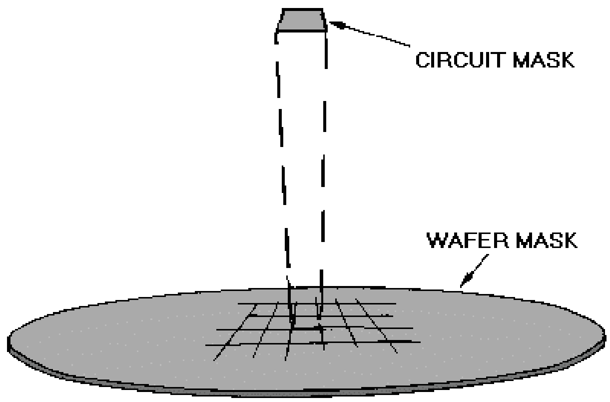


Figure 1-8.—Wafer mask distribution.

A wafer mask is a device used to deposit materials on a substrate. It allows material to be deposited in certain areas, but not in others. By changing the pattern of the mask, we can change the component arrangement of the circuit. Several different masks may be used to produce a simple microelectronic device. When used in proper sequence, conductor, semiconductor, or insulator materials may be applied to the substrate to form transistors, resistors, capacitors, and interconnecting leads.

SUBSTRATE PRODUCTION

As was mentioned earlier in this topic, microelectronic devices are produced on a substrate. This substrate will be of either insulator or semiconductor material, depending on the type of device. Film and hybrid ICs are normally constructed on a glass or ceramic substrate. Ceramic is usually the preferred material because of its durability.

Substrates used in monolithic ICs are of semiconductor material, usually silicon. In this type of IC, the substrate can be an active part of the IC. Glass or ceramic substrates are used only to provide support for the components.

Semiconductor substrates are produced by ARTIFICIALLY GROWING cylindrical CRYSTALS of pure silicon or germanium. Crystals are "grown" on a SEED CRYSTAL from molten material by slowly lifting and cooling the material repeatedly. This process takes place under rigidly controlled atmospheric and temperature conditions.

Figure 1-9 shows a typical CRYSTAL FURNACE. The seed crystal is lowered until it comes in contact with the molten material-silicon in this case. It is then rotated and raised very slowly. The seed crystal is at a lower temperature than the molten material. When the molten material is in contact with the seed, it solidifies around the seed as the seed is lifted. This process continues until the grown crystal is of the desired length. A typical crystal is about 2 inches in diameter and 10 to 12 inches long. Larger diameter crystals can be grown to meet the needs of the industry. The purity of the material is strictly controlled to maintain specific semiconductor properties. Depending on the need, n or p impurities are added to produce the desired characteristics. Several other methods of growing crystals exist, but the basic concept of crystal production is the same.

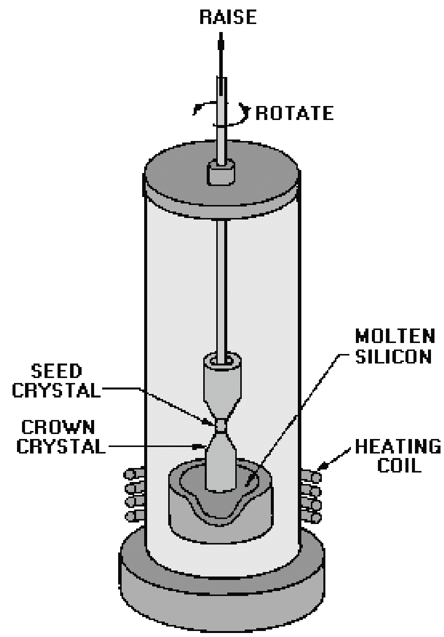


Figure 1-9.—Crystal furnace.

The cylinder of semiconductor material that is grown is sliced into thicknesses of .010 to .020 inch in the first step of preparation, as shown in figure 1-10. These wafers are ground and polished to remove any irregularities and to provide the smoothest surface possible. Although both sides are polished, only the side that will receive the components must have a perfect finish.

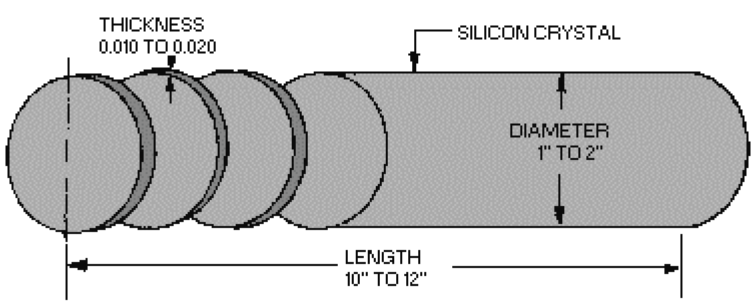


Figure 1-10.—Silicon crystal and wafers.

Q17. What are the basic steps in manufacturing an IC?

Q18. Computer-aided layout is used to prepare _______ devices.

Q19. What purpose do masks serve?

Q20. What type of substrates are used for film and hybrid ICs?

Q21. Describe the preparation of a silicon substrate.

FABRICATION OF IC DEVICES

Fabrication of monolithic ICs is the most complex aspect of microelectronic devices we will discuss. Therefore, in this introductory module, we will try to simplify this process as much as possible. Even though the discussion is very basic, the intent is still to increase your appreciation of the progress in microelectronics. You should, as a result of this discussion, come to realize that advances in manufacturing techniques are so rapid that staying abreast of them is extremely difficult.

Monolithic Fabrication.

Two types of monolithic fabrication will be discussed. These are the DIFFUSION METHOD and the EPITAXIAL METHOD.

DIFFUSION METHOD.—The DIFFUSION process begins with the highly polished silicon wafer being placed in an oven (figure 1-11). The oven contains a concentration impurity made up of impurity atoms which yield the desired electrical characteristics. The concentration of impurity atoms is diffused into the wafer and is controlled by controlling the temperature of the oven and the time that the silicon wafer is allowed to remain in the oven. This is called DOPING. When the wafer has been uniformly doped, the fabrication of semiconductor devices may begin. Several hundred circuits are produced simultaneously on the wafer.

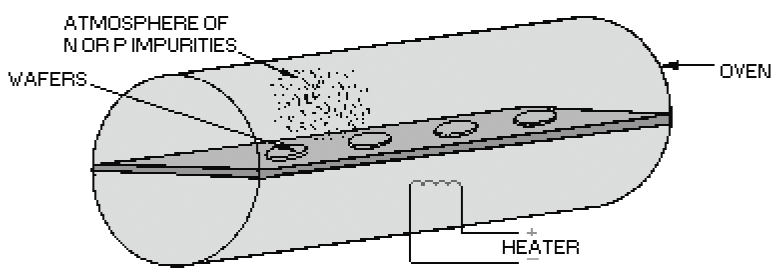


Figure 1-11.—Wafers in a diffusion oven.

The steps in the fabrication process described here, and illustrated in figure 1-12, would produce an npn, planar-diffused transistor. But, with slight variations, the technique may also be applied to the production of a complete circuit, including diodes, resistors, and capacitors. The steps are performed in the following order:

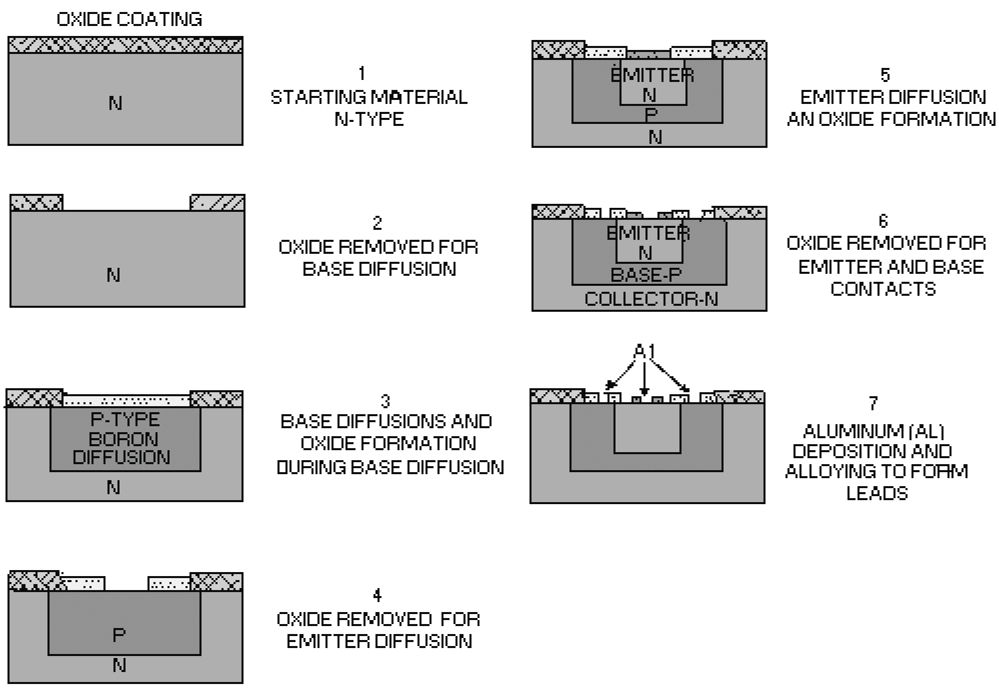


Figure 1-12.—Planar-diffused transistor.

1. An oxide coating is thermally grown over the n-type silicon starting material.

2. By means of the photolithographic process, a window is opened through the oxide layer. This is done through the use of masks, as discussed earlier.

3. The base of the transistor is formed by placing the wafer in a diffusion furnace containing a p-type impurity, such as boron. By controlling the temperature of the oven and the length of time that the wafer is in the oven, you can control the amount of boron diffused through the window (the boron will actually spread slightly beyond the window opening). A new oxide layer is then allowed to form over the area exposed by the window.

4. A new window, using a different mask much smaller than the first, is opened through the new oxide layer.

5. An n-type impurity, such as phosphorous, is diffused through the new window to form the emitter portion of the transistor. Again, the diffused material will spread slightly beyond the window opening. Still another oxide layer is then allowed to form over the window.

6. By means of precision-masking techniques, very small windows (about 0.005 inch in diameter) are opened in both the base and emitter regions of the transistor to provide access for electrical currents.

7. Aluminum is then deposited in these windows and alloyed to form the leads of the transistor or the IC.

(Note that the pn junctions are covered throughout the fabrication process by an oxide layer that prevents contamination.)

EPITAXIAL METHOD.—The EPITAXIAL process involves depositing a very thin layer of silicon to form a uniformly doped crystalline region (epitaxial layer) on the substrate. Components are produced by diffusing appropriate materials into the epitaxial layer in the same way as the planar-diffusion method. When planar-diffusion and epitaxial techniques are combined, the component characteristics are improved because of the uniformity of doping in the epitaxial layer. A cross section of a typical planar-epitaxial transistor is shown in figure 1-13. Note that the component parts do not penetrate the substrate as they did in the planar-diffused transistor.

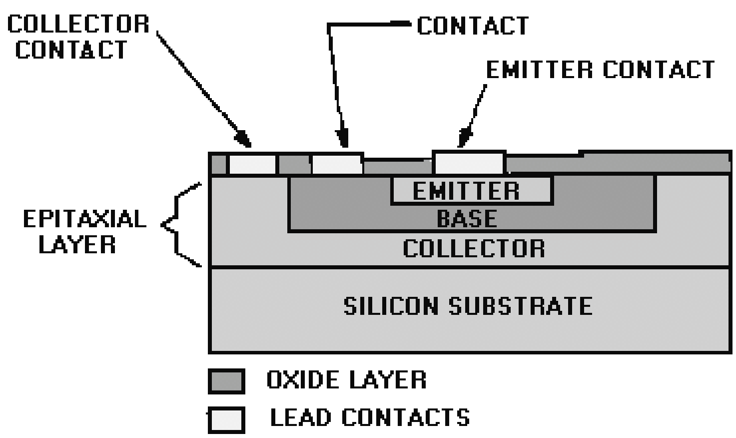


Figure 1-13.—Planar-epitaxial transistor.

ISOLATION.—Because of the closeness of components in ICs, ISOLATION from each other becomes a very important factor. Isolation is the prevention of unwanted interaction or leakage between components. This leakage could cause improper operation of a circuit.

Techniques are being developed to improve isolation. The most prominent is the use of silicon oxide, which is an excellent insulator. Some manufacturers are experimenting with single-crystal silicon grown on an insulating substrate. Other processes are also used which are far too complex to go into here. With progress in isolation techniques, the reliability and efficiency of ICs will increase rapidly.

Thin Film

Thin film is the term used to describe a technique for depositing passive circuit elements on an insulating substrate with coating to a thickness of 0.0001 centimeter. Many methods of thin-film deposition exist, but two of the most widely used are VACUUM EVAPORATION and CATHODE SPUTTERING.

VACUUM EVAPORATION.—Vacuum evaporation is a method used to deposit many types of materials in a highly evacuated chamber in which the material is heated by electricity, as shown in figure 1-14. The material is radiated in straight lines in all directions from the source and is shadowed by any objects in its path.

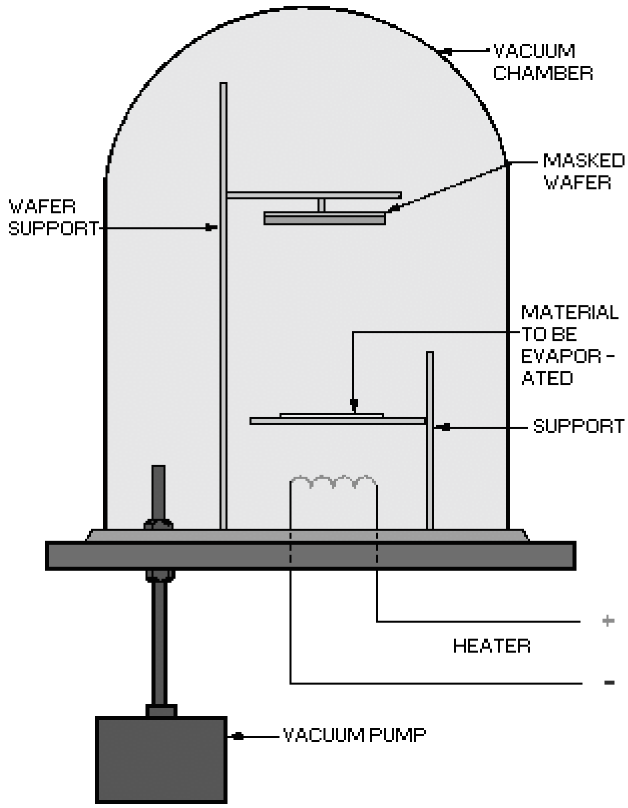


Figure 1-14.—Vacuum evaporation oven.

The wafers, with appropriate masks (figure 1-15), are placed above and at some distance from the material being evaporated. When the process is completed, the vacuum is released and the masks are removed from the wafers. This process leaves a thin, uniform film of the deposition material on all parts of the wafers exposed by the open portions of the mask. This process is also used to deposit interconnections (leads) between components of an IC.

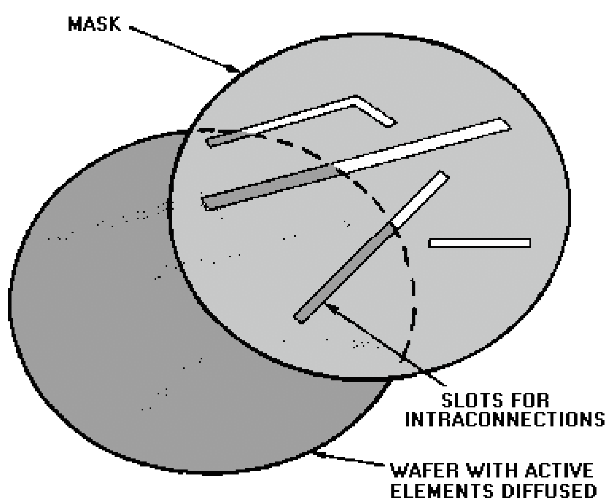


Figure 1-15.—Evaporation mask.

The vacuum evaporation technique is most suitable for deposition of highly reactive materials, such as aluminum, that are difficult to work with in air. The method is clean and allows a better contact between the layer of deposited material and the surface upon which it has been deposited. In addition, because evaporation beams travel in straight lines, very precise patterns may be produced.

CATHODE-SPUTTERING.—A typical cathode-sputtering system is illustrated in figure 1-16. This process is also performed in a vacuum. A potential of 2 to 5 kilovolts is applied between the anode and cathode (source material). This produces a GLOW DISCHARGE in the space between the electrodes. The rate at which atoms are SPUTTERED off the source material depends on the number of ions that strike it and the number of atoms ejected for each ion bombardment. The ejected atoms are deposited uniformly over all objects within the chamber. When the sputtering cycle is completed, the vacuum in the chamber is released and the wafers are removed. The masks are then removed from the wafers, leaving a deposit that forms the passive elements of the circuit, as shown in figure 1-17.

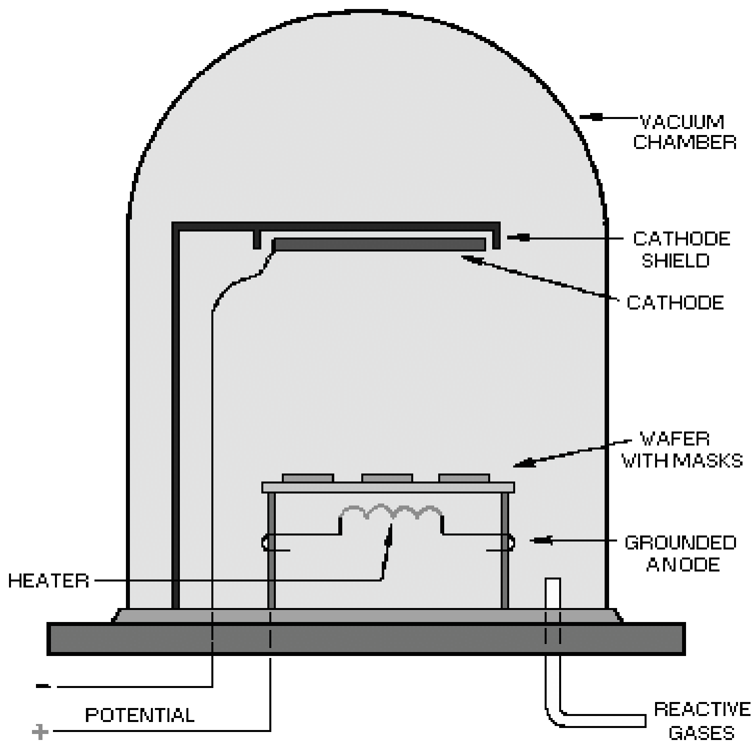


Figure 1-16.—Cathode-sputtering system.

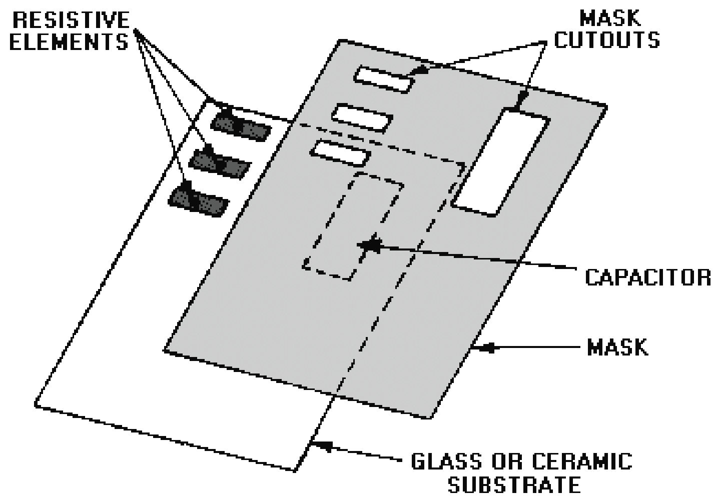


Figure 1-17.—Cathode-sputtering mask.

Finely polished glass, glazed ceramic, and oxidized silicon have been used as substrate materials for thin films. A number of materials, including nichrome, a compound of silicon oxide and chromium cermets, tantalum, and titanium, have been used for thin-film resistors. Nichrome is the most widely used.

The process for producing thin-film capacitors involves deposition of a bottom electrode, a dielectric, and finally a top electrode. The most commonly used dielectric materials are silicon monoxide and silicon dioxide.

Thick Film

Thick films are produced by screening patterns of conducting and insulating materials on ceramic substrates. A thick film is a film of material with a thickness that is at least 10 times greater than the mean free path of an electron in that material, or approximately 0.001 centimeter. The technique is used to produce only passive elements, such as resistors and capacitors.

PROCEDURES.—One procedure used in fabricating a thick film is to produce a series of stencils called SCREENS. The screens are placed on the substrate and appropriate conducting or insulating materials are wiped across the screen. Once the conducting or insulating material has been applied, the screens are removed and the formulations are fired at temperatures above 600 degrees Celsius. This process forms alloys that are permanently bonded to the insulating substrate. To a limited extent, the characteristics of the film can be controlled by the firing temperature and length of firing time.

RESISTORS.—Thick-film resistance values can be held to a tolerance of ±10 percent. Closer tolerances are obtained by trimming each resistor after fabrication. Hundreds of different cermet formulations are used to produce a wide range of component parameters. For example, the material used for a 10-ohm-per-square resistor is quite different from that used for a 100-kilohm-per-square resistor.

CAPACITORS AND RESISTOR-CAPACITOR NETWORKS.—Capacitors are formed by a sequence of screenings and firings. Capacitors, in this case, consist of a bottom plate, intraconnections, a dielectric, and a top plate. For resistor-capacitor networks, the next step would be to deposit the resistor material through the screen. The final step is screening and firing of a glass enclosure to seal the unit.

Hybrid Microcircuit

A hybrid microcircuit is one that is fabricated by combining two or more circuit types, such as film and semiconductor circuits, or a combination of one or more circuit types and discrete elements. The primary advantage of hybrid microcircuits is design flexibility; that is, hybrid microcircuits can be designed to provide wide use in specialized applications, such as low-volume and high-frequency circuits.

Several elements and circuits are available for hybrid applications. These include discrete components that are electrically and mechanically compatible with ICs. Such components may be used to perform functions that are supplementary to those of ICs. They can be handled, tested, and assembled with essentially the same technology and tools. A hybrid IC showing an enlarged chip is shown in figure 1-18.

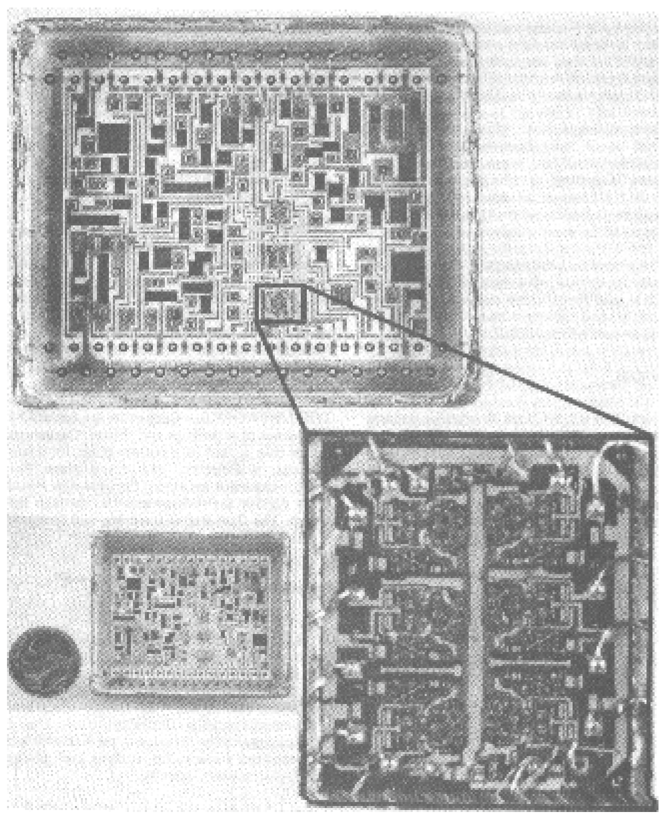


Figure 1-18.—Hybrid IC showing an enlarged chip.

Complete circuits are available in the form of UNCASED CHIPS (UNENCAPSULATED IC DICE). These chips are usually identical to those sold as part of the manufacturer's regular production line. They must be properly packaged and connected by the user if a high-quality final assembly is to be obtained. The circuits are usually sealed in a package to protect them from mechanical and environmental stresses. One-mil (0.001-inch), gold-wire leads are connected to the appropriate pins which are brought out of the package to allow external connections.

Q22. Name the two types of monolithic IC construction discussed.

Q23. How do the two types of monolithic IC construction differ?

Q24. What is isolation?

Q25. What methods are used to deposit thin-film components on a substrate?

Q26. How are thick-film components produced?

Q27. What is a hybrid IC?

Q28. What is the primary advantage of hybrid circuits?

PACKAGING TECHNIQUES

Once the IC has been produced, it requires a housing that will protect it from damage. This damage could result from moisture, dirt, heat, radiation, or other sources. The housing protects the device and aids in its handling and connection into the system in which the IC is used. The three most common types of packages are the modified TRANSISTOR-OUTLINE (TO) PACKAGE, the FLAT PACK, and the DUAL INLINE PACKAGE (DIP).

Transitor-Outline Package

Transistor-Outline Package. The transistor-outline (TO) package was developed from early experience with transistors. It was a reliable package that only required increasing the number of leads to make it useful for ICs. Leads normally number between 2 and 12, with 10 being the most common for IC applications. Figure 1-19 is an exploded view of a TO-5 package. Once the IC has been attached to the header, bonding wires are used to attach the IC to the leads. The cover provides the necessary protection for the device. Figure 1-20 is an enlarged photo of an actual TO-5 with the cover removed. You can easily see that the handling of an IC without packaging would be difficult for a technician.

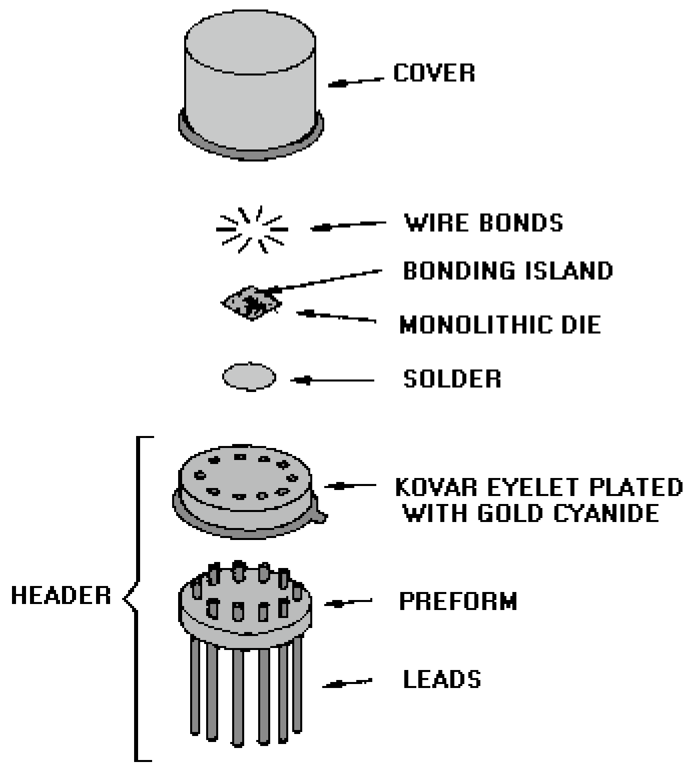


Figure 1-19.—Exploded TO-5.

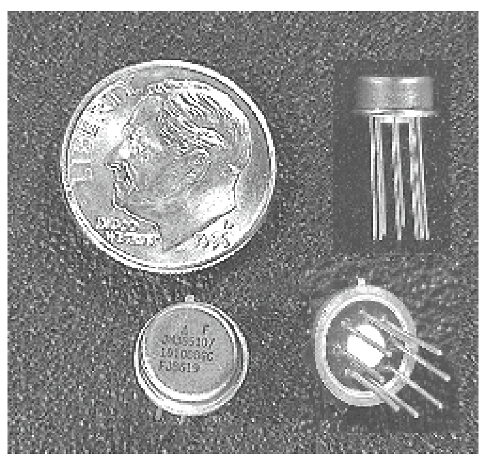


Figure 1-20.—TO-5 package.

The modified TO-5 package (figure 1-21) can be either plugged into [view (A)] or embedded in [view (B)] a board. The embedding method is preferred. Whether the package is plugged in or embedded, the interconnection area of the package leads must have sufficient clearance on both sides of the board. The plug-in method does not provide sufficient clearance between pads to route additional circuitry. When the packages are embedded, sufficient space exists between the pads [because of the increased diameter of the interconnection pattern, shown at the right in view (B)] for additional conductors.

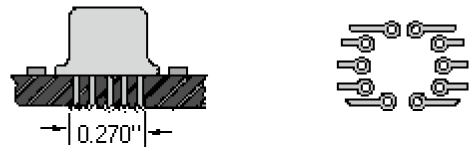


Figure 1-21A.—TO-5 mounting PLUG-IN MOUNTING

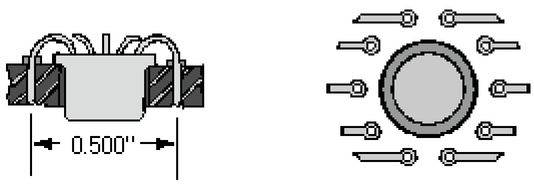


Figure 1-21B.—TO-5 mounting EMBEDDED CAN(LEADS PLUGGED IN)

Flat Pack

Many types of IC flat packs are being produced in various sizes and materials. These packages are available in square, rectangular, oval, and circular configurations with 10 to 60 external leads.

They may be made of metal, ceramic, epoxy, glass, or combinations of those materials. Only the ceramic flat pack will be discussed here. It is representative of all flat packs with respect to general package requirements (see figure 1-22).

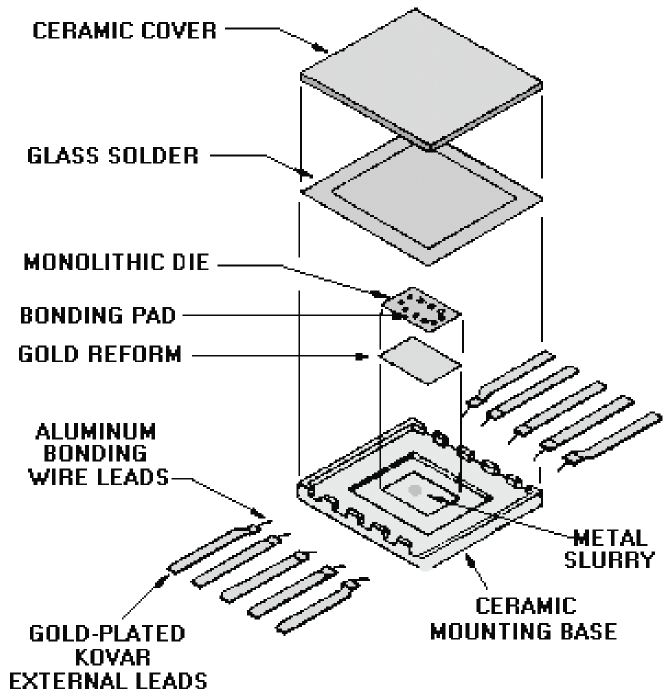


Figure 1-22.—Enlarged flat pack exploded view.

After the external leads are sealed to the mounting base, the rectangular area on the inside bottom of the base is treated with metal slurry to provide a surface suitable for bonding the monolithic die to the base. The lead and the metalized area in the bottom of the package are plated with gold. The die is then attached by gold-silicon bonding.

The die-bonding step is followed by bonding gold or aluminum wires between the bonding islands on the IC die and on the inner portions of the package leads. Next, a glass-soldered preformed frame is placed on top of the mounting base. One surface of the ceramic cover is coated with Pyroceram glass, and the cover is placed on top of the mounting base. The entire assembly is placed in an oven at 450 degrees Celsius. This causes the glass solder and Pyroceram to fuse and seal the cover to the mounting base. A ceramic flat pack is shown in figure 1-23. It has been opened so that you can see the chip and bonding wires.

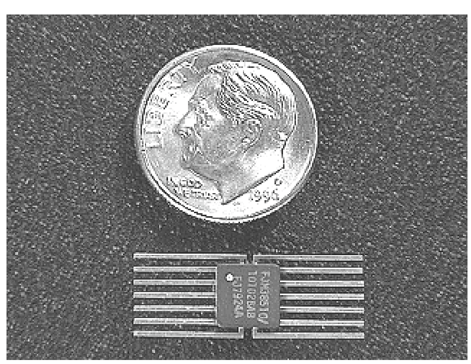


Figure 1-23.—Typical flat pack.

Dual Inline Package

The dual inline package (DIP) was designed primarily to overcome the difficulties associated with handling and inserting packages into mounting boards. DIPs are easily inserted by hand or machine and require no spreaders, spacers, insulators, or lead-forming tools. Standard hand tools and soldering irons can be used to field-service the devices. Plastic DIPs are finding wide use in commercial applications, and a number of military systems are incorporating ceramic DIPS.

The progressive stages in the assembly of a ceramic DIP are illustrated in figure 1-24, views (A) through (E). The integrated-circuit die is sandwiched between the two ceramic elements, as shown in view (A). The element on the left of view (A) is the bottom half of the sandwich and will hold the integrated-circuit die. The ceramic section on the right is the top of the sandwich. The large well in view (B) protects the IC die from mechanical stress during sealing operations. Each of the ceramic elements is coated with glass which has a low melting temperature for subsequent joining and sealing. View (B) shows the Kovar lead frame stamped and bent into its final shape. The excess material is intended to preserve pin alignment. The holes at each end are for the keying jig used in the final sealing operation. The lower half of the ceramic package is inserted into the lead frame shown in view (C). The die is mounted in the well and leads are attached. The top ceramic elements are bonded to the bottom element shown in view (D) and the excess material is removed from the package. View (E) is the final product.

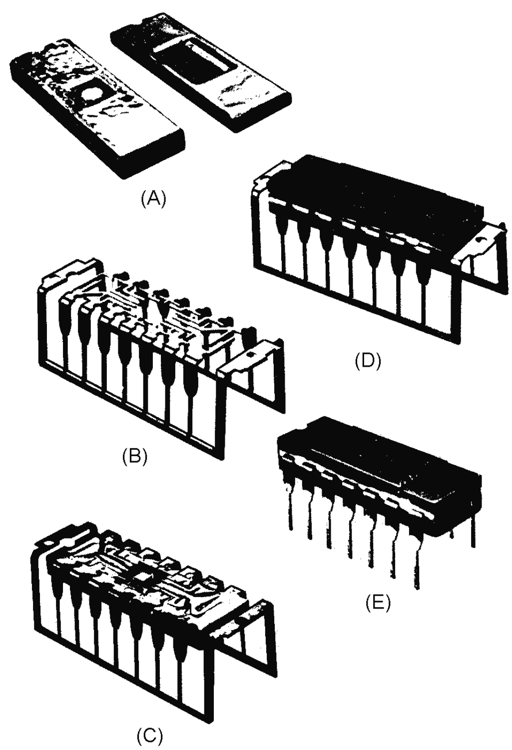


Figure 1-24.—DIP packaging steps.

Ceramic DIPs are processed individually while plastic DIPs are processed in quantities of two or more (in chain fashion). After processing, the packages are sawed apart. The plastic package also uses a Kovar lead frame, but the leads are not bent until the package is completed. Because molded plastic is used to encapsulate the IC die, no void will exist between the cover and die, as is the case with ceramic packaging.

At present, ceramic DIPs are the most common of the two package types to be found in Navy microelectronic systems. Figure 1-25 shows a DIP which has been opened.

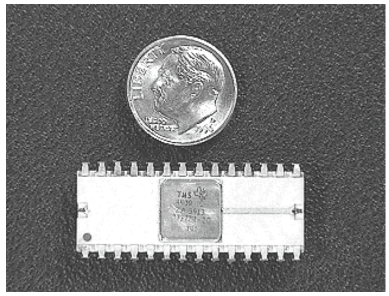


Figure 1-25.—Dual inline package (DIP).

RECENT DEVELOPMENTS IN PACKAGING

Considerable effort has been devoted to eliminating the fine wires used to connect ICs to Kovar leads. The omission of these wires reduces the cost of integrated circuits by eliminating the costs associated with the bonding process. Further, omission of the wires improves reliability by eliminating a common cause of circuit failure.

A promising packaging technique is the face-down (FLIP-CHIP) mounting method by which conductive patterns are evaporated inside the package before the die is attached. These patterns connect the external leads to bonding pads on the inside surface of the die. The pads are then bonded to appropriate pedestals on the package that correspond to those of the bonding pads on the die (figure 1-26).

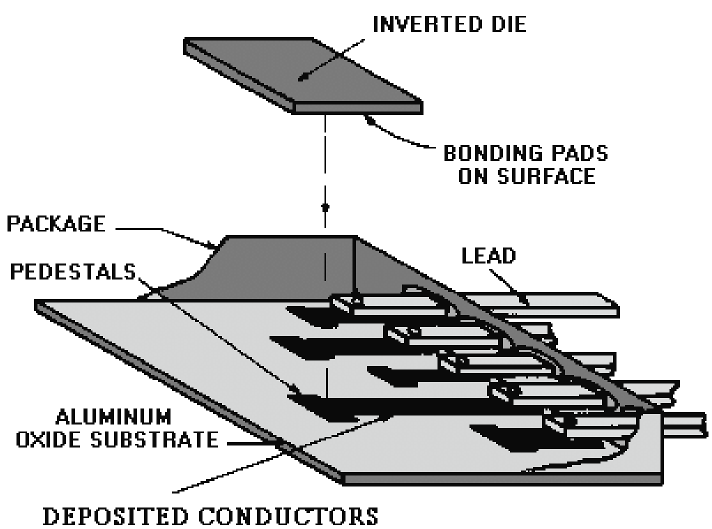


Figure 1-26.—Flip-chip package.

The BEAM-LEAD technique is a process developed to batch-fabricate (fabricate many at once) semiconductor circuit elements and integrated circuits with electrodes extended beyond the edges of the wafer, as shown in figure 1-27. This type of structure imposes no electrical difficulty, and parasitic capacitance (under 0.05 picofarad per lead) is equivalent to that of a wire-bonded and brazed-chip assembly. In addition, the electrodes may be tapered to allow for lower inductance, impedance matching, and better heat conductance. The beam-lead technique is easily accomplished and does not have the disadvantages of chip brazing and wire bonding. The feasibility of this technique has been demonstrated in a variety of digital, linear, and thin-film circuits.

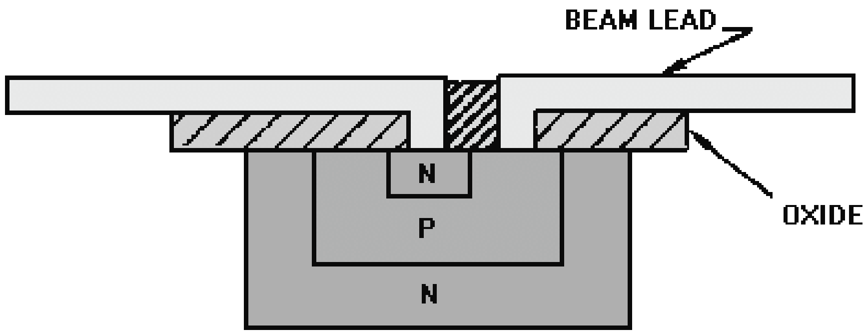


Figure 1-27.—Beam-lead technique.

Another advance in packaging is that of increasing the size of DIPs. General-purpose DIPs have from 4 to 16 pins. Because of lsi and vlsi, manufacturers are producing DIPs with up to 64 pins. Although size is increased considerably, all the advantages of the DIP are retained. DIPs are normally designed to a particular specification set by the user.

Q29. What is the purpose of the IC package?

Q30. What are the three most common types of packages?

Q31. What two methods of manufacture are being used to eliminate bonding wires?

EQUIVALENT CIRCUITS

At the beginning of this topic, we discussed many applications of microelectronics. You should understand that these applications cover all areas of modern electronics technology. Microelectronic ICs are produced that can be used in many of these varying circuit applications to satisfy the needs of modern technology. This section will introduce you to some of these applications and will show you some EQUIVALENT CIRCUIT comparisons of discrete components and integrated circuits.

J-K FLIP-FLOP AND IC SIZES

Integrated circuits can be produced that combine all the elements of a complete electronic circuit. This can be done with either a single chip of silicon or a single chip of silicon in combination with film components. The importance of this new production method in the evolution of microelectronics can be demonstrated by comparing a conventional J-K flip-flop circuit incorporating solid-state discrete devices and the same type of circuit employing integrated circuitry. (A J-K flip-flop is a circuit used primarily in computers.)

You should recall from *NEETS, Module 13*, *Introduction to Number Systems, Boolean Algebra, and Logic Circuits*, that a basic flip-flop is a device having two stable states and two input terminals (or types of input signals), each of which corresponds to one of the two states. The flip-flop remains in one state until caused to change to the other state by application of an input voltage pulse.

A J-K flip-flop differs from the basic flip-flop because it has a third input terminal. A clock pulse, or trigger, is usually applied to this input to ensure proper timing in the circuit. An input signal must occur at the same time as the clock pulse to change the state of the flip-flop. The conventional J-K flip-flop circuit in figure 1-28 requires approximately 40 discrete components, 200 connections, and 300 processing operations. Each of these 300 operations (seals and connections) represents a possible source of failure. If all the elements of this circuit are integrated into one chip of silicon, the number of connections drops to approximately 14. This is because all circuit elements are intraconnected inside the package and the 300 processing operations are reduced to approximately 30. Figure 1-29 represents a size comparison of a discrete J-K circuit and an integrated circuit of the same type.

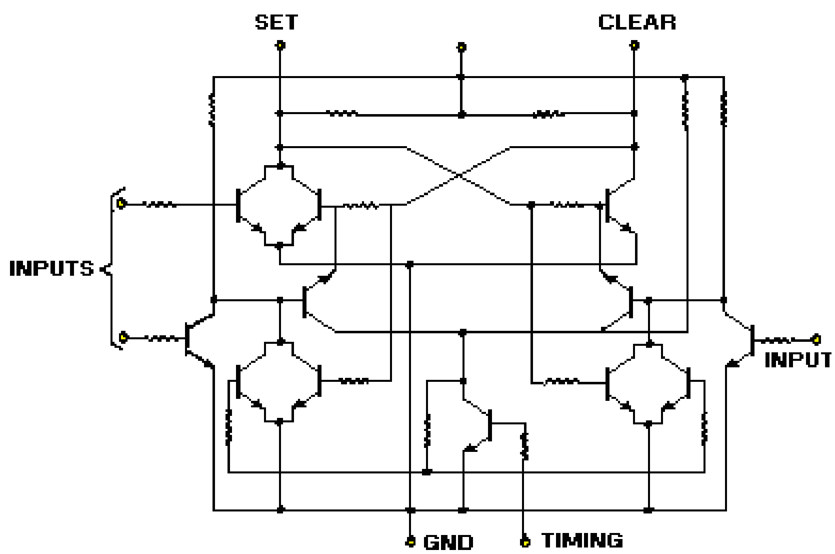


Figure 1-28.—Schematic diagram of a J-K flip-flop.

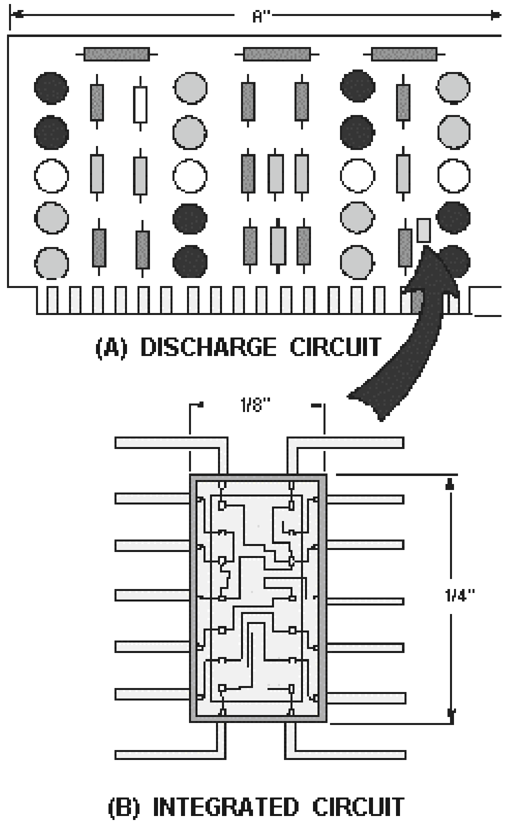


Figure 1-29.—J-K flip-flop discrete component and an IC.

IC PACKAGE LEAD IDENTIFICATION (NUMBERING)

When you look at an IC package you should notice that the IC could be connected incorrectly into a circuit. Such improper replacement of a component would likely result in damage to the equipment. For this reason, each IC has a REFERENCE MARK to align the component for placement. The dual inline package (both plastic and ceramic) and the flat pack have a notch, dot, or impression on the package. When the package is viewed from the top, pin 1 will be the first pin in the *counterclockwise* direction next to the reference mark. Pin 1 may also be marked directly by a hole or notch or by a tab on it (in this case pin 1 is the counting reference). When the package is viewed from the top, all other pins are numbered consecutively in a *counterclockwise* direction from pin 1, as shown in figure 1-30, views (A) and (B).

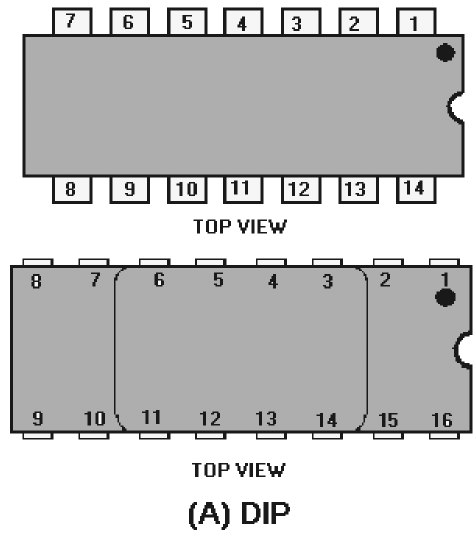


Figure 1-30A.—DIP and flat-pack lead numbering. DIP

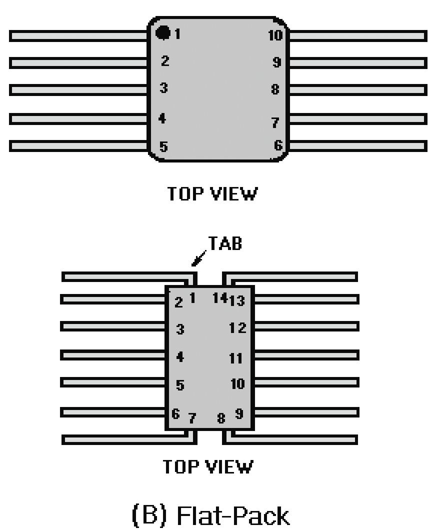


Figure 1-30B.—DIP and flat-pack lead numbering. Flat-Pack

The TO-5 can has a tab for the reference mark. When numbering the leads, you must view the TO-5 can from the *bottom*. Pin 1 will be the first pin in a *clockwise* direction from the tab. All other pins will be numbered consecutively in a clockwise direction from pin 1, as shown in figure 1-31.

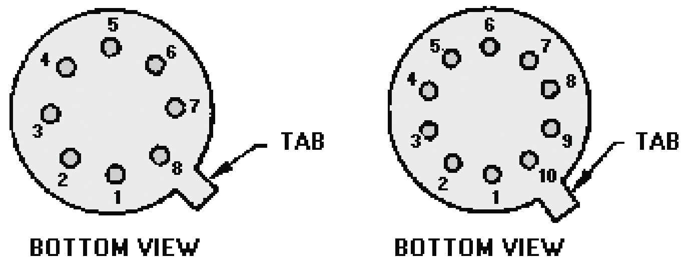


Figure 1-31.—Lead numbering for a TO-5.

IC IDENTIFICATION

As mentioned earlier, integrated circuits are designed and manufactured for hundreds of different uses. Logic circuits, clock circuits, amplifiers, television games, transmitters, receivers, and musical instruments are just a few of these applications.

In schematic drawings, ICs are usually represented by one of the schematic symbols shown in figure 1-32. The IC is identified *according to its use* by the numbers printed on or near the symbol. That series of numbers and letters is also stamped on the case of the device and can be used along with the data sheet, as shown in the data sheet in figure 1-33, by circuit designers and maintenance personnel. This data sheet is provided by the manufacturer. It provides a schematic diagram and describes the type of device, its electrical characteristics, and typical applications. The data sheet may also show the pin configurations with all pins labeled. If the pin configurations are not shown, there may be a schematic diagram showing pin functions. Some data sheets give both pin configurations and schematic diagrams, as shown in figure 1-34. This figure illustrates a manufacturer's data sheet with all of the pin functions shown.

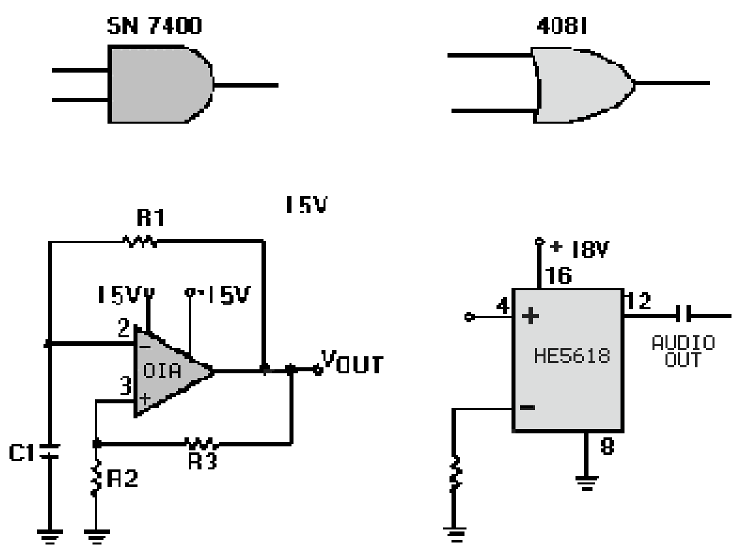


Figure 1-32.—Some schematic symbols for ICs.

LH101 LH201 OPERATIONAL AMPLIFIER

FOR AMPLIFIERS, VOLTAGE COMPARATORS, LOW DRIFT SAMPLE-AND-HOLD

Features

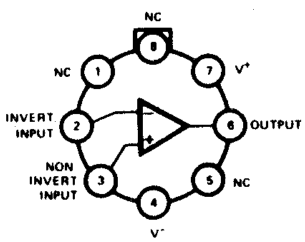
- Low Offsets and Temperature Drift
- Internal 30 pF Capacitor for Frequency Compensation
- Operation from ±5 to ±20 Volt Power Supplies
- Low Current Drain, 1.8 mA at ±20 Volts Typical
- Continuous Short-Circuit Protection
- No Latch Up When Common Mode Range Is Exceeded
- Same Pin Configuration as 709 Amplifier

Description

The LH101/LH201 is stable for all feedback configurations, even with capacitive loads, with no external compensation capacitors. Low power dissipation permits high voltage operation across the full temperature range.

PIN CONFIGURATIONS

Metal Can Package

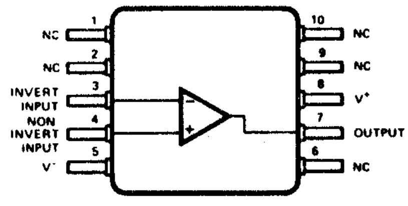


TOP VIEW

NOTE PIN 4 CONNECTED TO CASE

ORDER NUMBERS: LH101H OR LH201H
SEE PACKAGE 1

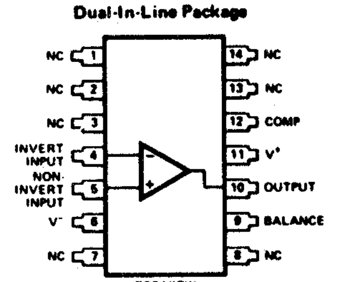
Flat Package



TOP VIEW

ORDER NUMBERS: LH101F OR LH201F
SEE PACKAGE 4

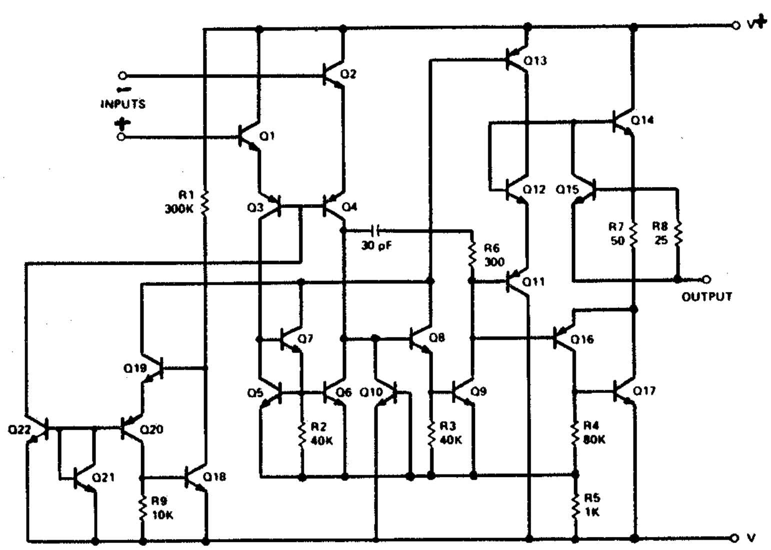
Dual-In-Line Package



TOP VIEW

ORDER NUMBERS: LH101D OR LH201D
SEE PACKAGE 11

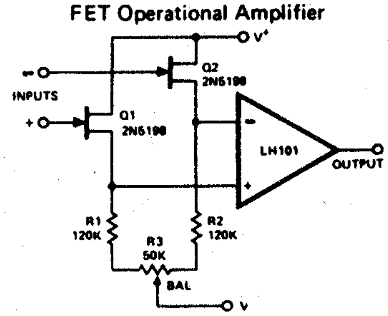
SCHEMATIC DIAGRAM



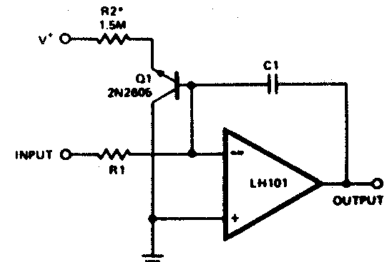
✱ NC – NOT CONNECTED INTERNALLY

TYPICAL APPLICATIONS

FET Operational Amplifier



Integrator with Bias Current Compensation



*SELECT FOR ZERO INTEGRATOR DRIFT.

Courtesy of Siliconix Incorporated

Figure 1-33.—Manufacturer's Data Sheet.

FEATURES

- MONOLITHIC CONSTRUCTION
- INITIAL ACCURACY +10V ± 0.010V
- OUTPUT VOLTAGE ERROR, TOTAL ±1/4 LSB
- LOW NOISE $20\mu V_{p-p}$
- WIDE INPUT RANGE 12V TO 30V
- LOW POWER DISSIPATION 30mW
- OUTPUT SHORT CIRCUIT PROTECTION
- ADJUSTABLE OUTPUT

APPLICATIONS

- AN ECONOMICAL EXTERNAL REFERENCE FOR: HI-5608; DAC 08; AD1408; AD559
- VOLTAGE REGULATOR REFERENCE
- PORTABLE BATTERY OPERATED EQUIPMENT
- NEGATIVE 10V REFERENCE

DESCRIPTION

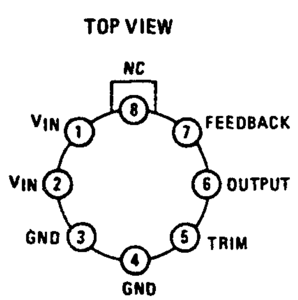
HA-1608 is a monolithic +10V adjustable voltage reference featuring accuracy and temperature stability specifications detailed exclusively for 8 bit data conversion systems. A stable +10V output is provided by a reference zener and buffer amplifier coupled with laser trimmed feedback and zener bias resistors. Long term stability is ensured through integration of all reference components into a monolithic design. Flexibility of HA-1608 is provided through an external trim control which allows the user to adjust the output voltage for binary or BCD applications without affecting overall performance.

These devices provide a total output voltage error of ± 1/4 LSB for 8 bit D/A or A/D converters. Low standby power (0.3mW) makes HA-1608 a natural selection for portable battery operated equipment, comparator references, and reference stacking circuits. These devices can also be used on -10V references.

HA-1608 is packaged in 8 pin metal cans (TO-99) and 8 pin DIPs. The pinout is arranged for convenient replacement of other less accurate regulators in applications demanding minimal change with temperature and time. HA-1608-2 is specified for -55°C to +125°C operation while the HA-1608-5 operates from 0°C to +75°C.

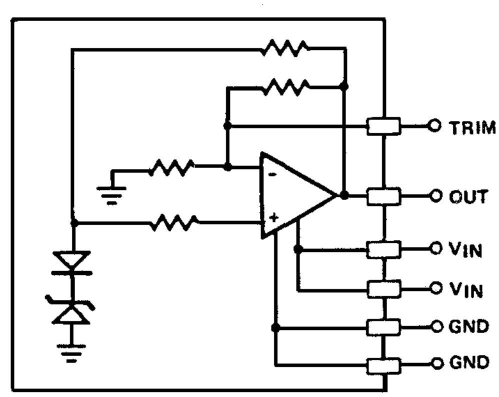
PINOUT

Section 11 for Packaging



* NC – NOT CONNECTED INTERNALLY

FUNCTIONAL SCHEMATIC



Courtesy of Harris Semiconductors

Figure 1-34.—Manufacturer's Data Sheet.

Q32. On DIP and flat-pack ICs viewed from the top, pin 1 is located on which side of the reference mark?

Q33. DIP and flat-pack pins are numbered consecutively in what direction?

Q34. Viewed from the bottom, TO-5 pins are coumted on what direction?

Q35. VThe numbers and letters on ICs and schematics serve what purpose?

MICROELECTRONIC SYSTEM DESIGN CONCEPTS

You should understand the terminology used in microelectronics to become an effective and knowledgeable technician. You should be familiar with packaging concepts from a maintenance standpoint and be able to recognize the different types of assemblies. You should also know the electrical and environmental factors that can affect microelectronic circuits. In the next section of this topic we will define and discuss each of these areas.

TERMINOLOGY

As in any special electronics field, microelectronics terms and definitions are used to clarify communications. This is done so that everyone involved in microelectronics work has the same knowledge of the field. You can imagine how much trouble you would have remembering 10 or more different names and definitions for a resistor. If standardization didn't exist for the new terminology, you would have far more trouble understanding microelectronics. To standardize terminology in microelectronics, the Navy has adopted several definitions with which you should become familiar. These definitions will be presented in this section.

Microelectronics

Microelectronics is that area of electronics technology associated with electronics systems built from extremely small electronic parts or elements. Most of today's computers, weapons systems, navigation systems, communications systems, and radar systems make extensive use of microelectronics technology.

Microcircuit

A microcircuit is not what the old-time technician would recognize as an electronic circuit. The old-timer would no longer see the familiar discrete parts (individual resistors, capacitors, inductors, transistors, and so forth). Microelectronic circuits, as discussed earlier, are complete circuits mounted on a substrate (integrated circuit). The process of fabricating microelectronic circuits is essentially one of building discrete component characteristics either into or onto a single substrate. This is far different from soldering resistors, capacitors, transistors, inductors, and other discrete components into place with wires and lugs. The component characteristics built into microcircuits are referred to as ELEMENTS, rather than discrete components. Microcircuits have a high number of these elements per substrate compared to a circuit with discrete components of the same relative size. As a matter of fact, microelectronic circuits often contain thousands of times the number of discrete components. The term *HIGH EQUIVALENT CIRCUIT DENSITYS* is a description of this element-to-discrete part relationship. For example, suppose you have a circuit with 1,000 discrete components mounted on a chassis which is 8 × 10 × 2 inches. The equivalent circuit in microelectronics might be built into or onto a single substrate which is only 3/8 × 1 × 1/4 inch. The 1,000 elements of the microcircuit would be very close to each other (high density) by comparison to the distance between discrete components mounted on the large chassis. The elements within the substrate are interconnected on the single substrate itself to perform an electronic function. A microcircuit does not have any discrete components mounted on it as do printed circuit boards, circuit card assemblies, and modules composed exclusively of discrete component parts.

Microcircuit Module

Microcircuits may be used in combination with discrete components. An assembly of microcircuits or a combination of microcircuits and discrete conventional electronic components that performs one or more distinct functions is a microcircuit module. The module is constructed as an independently packaged, replaceable unit. Examples of microcircuit modules are printed circuit boards and circuit card assemblies. Figure 1-35 is a photograph of a typical microcircuit module.

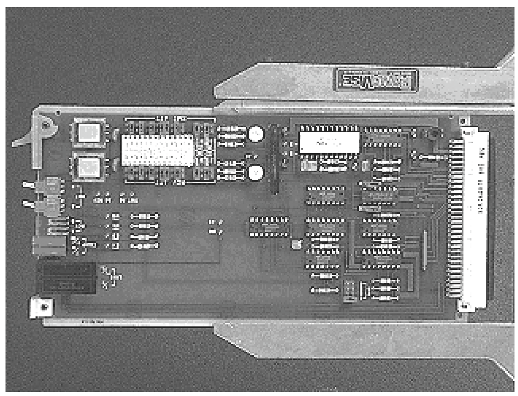


Figure 1-35.—Microcircuit module.

Miniature Electronics

Miniature electronics includes miniature electronic components and packages. Some examples are printed circuit boards, printed wiring boards, circuit card assemblies, and modules composed *exclusively* of discrete electronic parts and components (excluding microelectronic packages) mounted on boards, assemblies, or modules. MOTHER BOARDS, large printed circuit boards with plug-in modules, are considered miniature electronics. Cordwood modules also fall into this category. Miniature motors, synchros, switches, relays, timers, and so forth, are also classified as miniature electronics.

Recall that microelectronic components contain integrated circuits. Miniature electronics contain discrete elements or parts. You will notice that printed circuit boards and circuit card assemblies are mentioned in more than one definition. To identify the class (microminiature or miniature) of the unit, you must first determine the types of components used.

Q36. Standardized terms improve what action between individuals?

Q37. Microcircuit refers to any component containing what types of elements?

Q38. Components made up exclusively of discrete elements are classified as what type of electronics?

SYSTEM PACKAGING

When a new electronics system is developed, several areas of planning require special attention. An area of great concern is that of ensuring that the system performs properly. The designer must take into account all environmental and electrical factors that may affect the system. This includes temperature, humidity, vibration, and electrical interference. The design factor that has the greatest impact on you, as the technician, is the MAINTAINABILITY of the system. The designer must take into account how well you will be able to locate problems, identify the faulty components, and make the necessary repairs. If a system cannot be maintained easily, then it is not an efficient system. PACKAGING, the method of enclosing and mounting components, is of primary importance in system maintainability.

Levels of Packaging

For the benefit of the technician, system packaging is usually broken down to five levels (0 to IV). These levels are shown in figure 1-36.

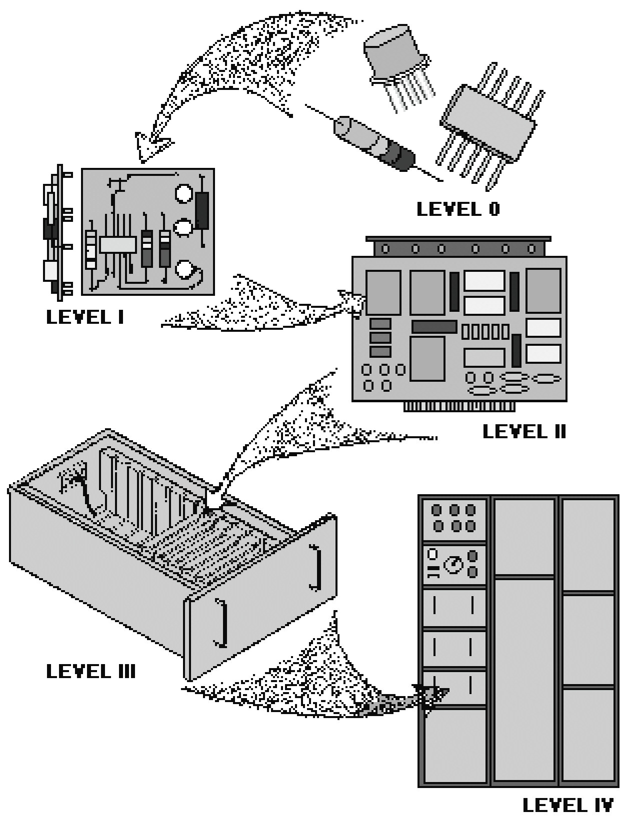


Figure 1-36.—Packaging levels.

LEVEL 0.—Level 0 packaging identifies nonrepairable parts, such as integrated circuits, transistors, resistors, and so forth. This is the lowest level at which you can perform maintenance. You are limited to simply replacing the faulty element or part. Depending on the type of part, repair might be as simple as plugging in a new relay. If the faulty part is an IC, special training and equipment will be required to accomplish the repair. This will be discussed in topic 2.

LEVEL I.—This level is normally associated with small modules or submodules that are attached to circuit cards or mother boards. The analog-to-digital (A/D) converter module is a device that converts a signal that is a function of a continuous variable (like a sine wave) into a representative number sequence in digital form. The A/D converter in figure 1-37 is a typical level I component. At this level of maintenance you can replace the faulty module with a good one. The faulty module can then be repaired at a later time or discarded. This concept significantly reduces the time equipment is inoperable.

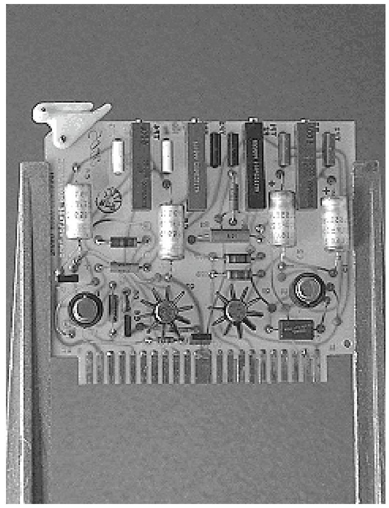


Figure 1-37.—Printed circuit board (pcb).

LEVEL II.—Level II packaging is composed of large printed circuit boards and/or cards (mother boards). Typical units of this level are shown in figures 1-37 and 1-38. In figure 1-38 the card measures 15×5.25 inches. The large dual inline packages (DIPs) are 2.25 inches x 0.75 inch. Other DIPs on the pcb are much smaller. Interconnections are shown between DIPs. You should also be able to locate a few discrete components. Repair consists of removing the faulty DIP or discrete component from the pcb and replacing it with a new part. Then the pcb is placed back into service. The removed part may be a level 0 or I part and would be handled as described in those sections. In some cases, the entire pcb should be replaced.

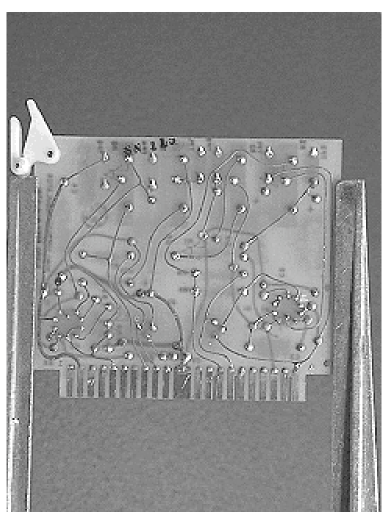


Figure 1-38.—Printed circuit board (pcb).

LEVEL III.—Drawers or pull-out chassis are level III units, as shown in figure 1-36. These are designed for accessibility and ease of maintenance. Normally, circuit cards associated with a particular subsystem will be grouped together in a drawer. This not only makes for an orderly arrangement of subsystems but also eliminates many long wiring harnesses. Defective cards are removed from such drawers and defective components are repaired as described in level II.

LEVEL IV.—Level IV is the highest level of packaging. It includes the cabinets, racks, and wiring harnesses necessary to interconnect all of the other levels. Other pieces of equipment of the same system classified as level IV, such as radar antennas, are broken down into levels 0 to III in the same manner.

During component troubleshooting procedures, you progress from level IV to III to II and on to level 0 where you identify the faulty component. As you become more familiar with a system, you should be able to go right to the drawer or module causing the problem.

Q39. Resistors, capacitors, transistors, and the like, are what level of packaging?

Q40. Modules or submodules attached to a mother board are what packaging level?

Q41. What is the packaging level of a pcb?

INTERCONNECTIONS IN PRINTED CIRCUIT BOARDS

As electronic systems become more complex, interconnections between components also becomes more complex. As more components are added to a given space, the requirements for interconnections become extremely complicated. The selection of conductor materials, insulator materials, and component physical size can greatly affect the performance of the circuit. Poor choices of these materials can contribute to poor signals, circuit noise, and unwanted electrical interaction between components. The three most common methods of interconnection are the conventional pcb, the multilayer pcb, and the modular assembly. Each of these will be discussed in the following sections.

Conventional Printed Circuit Board

Printed circuit boards were discussed earlier in topic 1. You should recall that a conventional pcb consists of glass-epoxy insulating base on which the interconnecting pattern has been etched. The board may be single- or double-sided, depending on the number of components mounted on it. Figures 1-37 and 1-38 are examples of conventional printed circuit boards.

Multilayer Printed Circuit Board.

The multilayer printed circuit board is emerging as the solution is interconnection problems associated with high-density packaging. Multilayer boards are used to:

- reduce weight
- conserve space in interconnecting circuit modules
- eliminate costly and complicated wiring harnesses
- provide shielding for a large number of conductors
- provide uniformity in conductor impedance for high-speed switching systems

- allow greater wiring density on boards

Figure 1-39 illustrates how individual boards are mated to form the multilayer unit. Although all multilayer boards are similarly constructed, various methods can be used to interconnect the circuitry from layer to layer. Three proven processes are the clearance-hole, plated-through hole, and layer build-up methods.

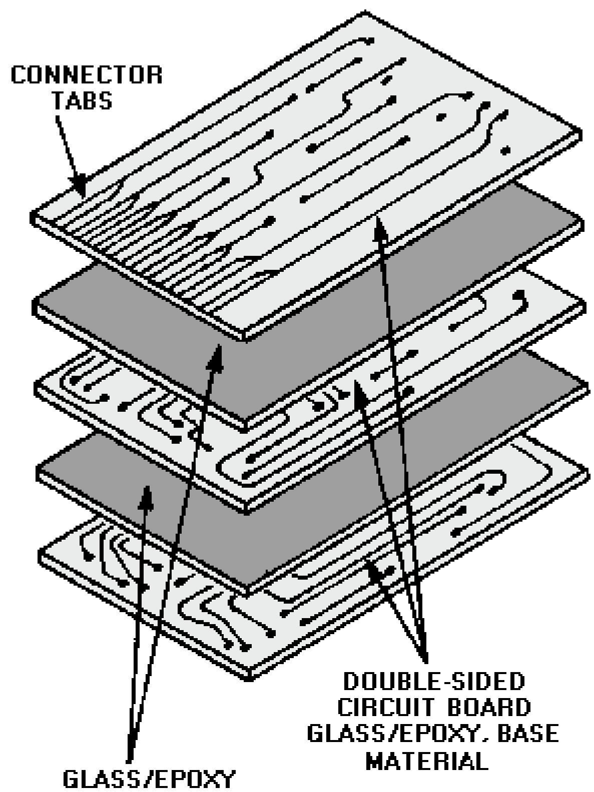


Figure 1-39.—Multilayer pcb.

CLEARANCE-HOLE METHOD.—In the CLEARANCE-HOLE method, a hole is drilled in the copper island (terminating end) of the appropriate conductor on the top layer. This provides access to a conductor on the second layer as shown by hole A in figure 1-40. The clearance hole is filled with solder to complete the connection. Usually, the hole is drilled through the entire assembly at the connection site. This small hole is necessary for the SOLDER-FLOW PROCESS used with this interconnection method.

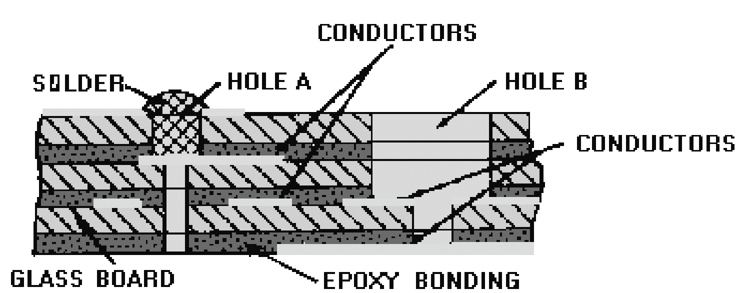


Figure 1-40.—Clearance-hole interconnection.

Conductors, located several layers below the top, are connected by using a STEPPED-DOWN HOLE PROCESS. Before assembly of a three-level board, a clearance hole is drilled down to the first layer to be interconnected. The first layer to be interconnected is predrilled with a hole smaller than those drilled in layers 1 and 2; succeeding layers to be connected have progressively smaller clearance holes. After assembly, the exposed portion of the conductors are interconnected by filling the stepped-down holes with solder, as shown by hole B in figure 1-40. The larger the number of interconnections required at one point, the larger must be the diameter of the clearance holes on the top layer. Large clearance holes on the top layer allow less space for components and reduce packaging density.

PLATED-THROUGH-HOLE METHOD.—The PLATED-THROUGH-HOLE method of interconnecting conductors is illustrated in figure 1-41. The first step is to assemble temporarily all the layers into their final form. Holes, corresponding to required connections, are drilled through the entire assembly and then the unit is disassembled. The internal walls of those holes to be interconnected are plated with metal which is 0.001 inch thick. This, in effect, connects the conductor on the board surface through the hole itself. This process is identical to that used for standard printed circuit boards. The boards are then reassembled and permanently bonded together with heat and pressure. All the holes are plated through with metal.

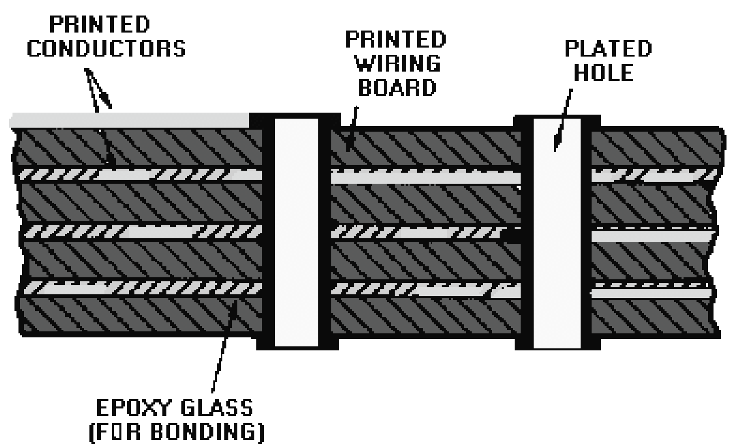


Figure 1-41.—Plated through-hole interconnection.

LAYER BUILD-UP METHOD.—With the LAYER BUILD-UP method, conductors and insulation layers are alternately deposited on a backing material, as shown in figure 1-42. This method produces copper interconnections between layers and minimizes the thermal expansion effects of dissimilar materials. However, reworking the internal connections in built-up layers is usually difficult, if not impossible.

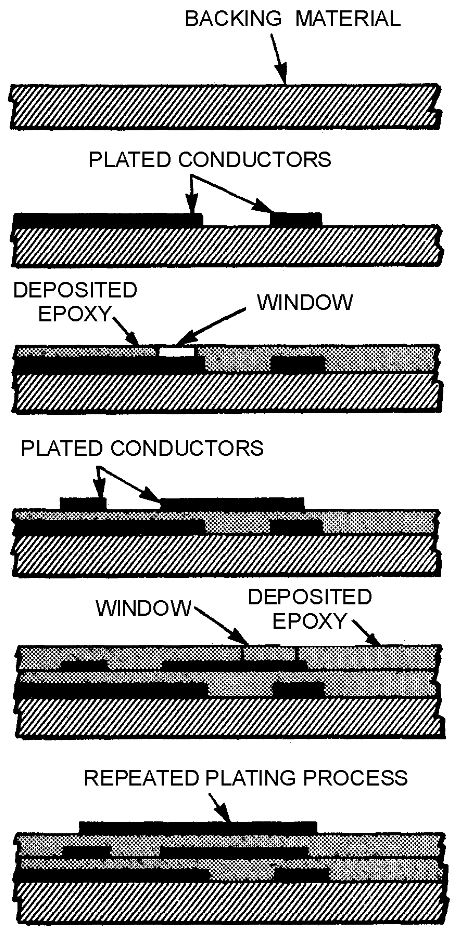


Figure 1-42.—Layer build-up technique.

Advantages and Disadvantages of Printed Circuit Boards

Some of the advantages and disadvantages of printed circuit boards were discussed earlier in this topic. They are strong, lightweight, and eliminate point-to-point wiring. Multilayer printed circuit boards allow more components per card. Entire circuits or even subsystems may be placed on the same card. However, these cards do have some drawbacks. For example, all components are wired into place, repair of cards requires special training and/or special equipment, and some cards cannot be economically repaired because of their complexity (these are referred to as THROWAWAYS).

Modular Assemblies

The MODULAR-ASSEMBLY (nonrepairable item) approach was devised to achieve ultra-high density packaging. The evolution of this concept, from discrete components to microelectronics, has progressed through various stages. These stages began with cordwood assemblies and functional blocks and led to complete subsystems in a single package. Examples of these configurations are shown in figure 1-43, view (A), view (B), and view (C).

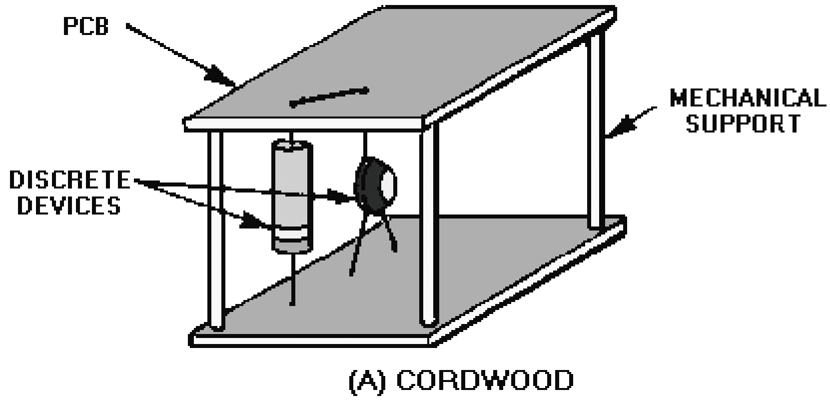


Figure 1-43A.—Evolution of modular assemblies. CORDWOOD.

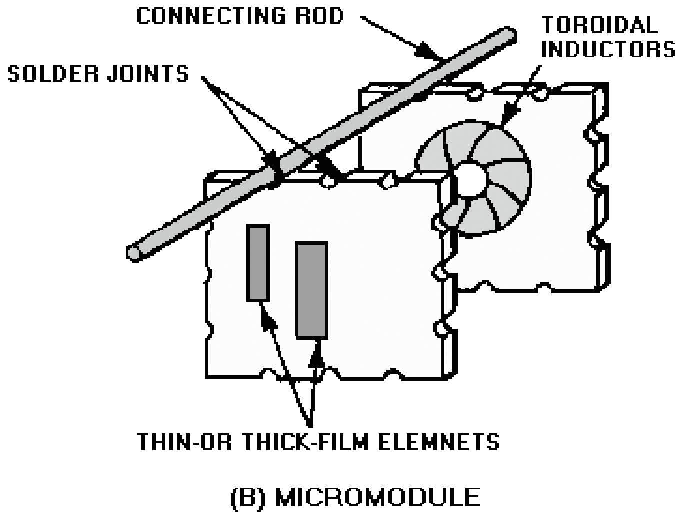


Figure 1-43B.—Evolution of modular assemblies. MICROMODULE.

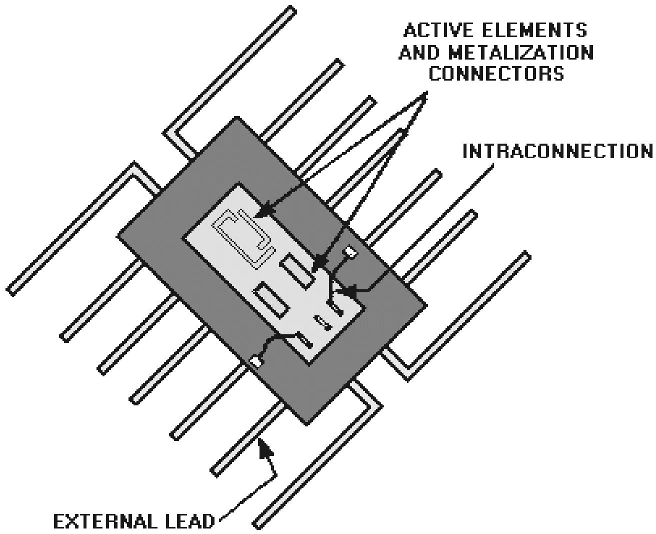


Figure 1-43C.—Evolution of modular assemblies. INTEGRATED-CIRCUIT.

Cordwood Modules

The cordwood assembly, shown in view (A) of figure 1-43, was designed and fabricated in various forms and sizes, depending on user requirements. This design was used to reduce the physical size and increase the component density and complexity of circuits through the use of discrete devices. However, the use of the technique was somewhat limited by the size of available discrete components used.

Micromodules

The next generation assembly was the micromodule. Designers tried to achieve maximum density in this design by using discrete components, thick- and thin-film technologies, and the insulator substrate principle. The method used in this construction technique allowed for the efficient use of space and also provided the mechanical strength necessary to withstand shock and vibration.

Semiconductor technology was then improved further with the introduction of the integrated circuit. The flat-pack IC form, shown in view (C), emphasizes the density and complexity that exists with this technique. This technology provides the means of reducing the size of circuits. It also allows the reduction of the size of systems through the advent of the lsi circuits that are now available and vlsi circuits that are being developed by various IC manufacturers.

Continuation of this trend toward microminiaturization will result in system forms that will require maintenance personnel to be specially trained in maintenance techniques to perform testing, fault isolation, and repair of systems containing complex miniature and microminiature circuits.

Q42. What are the three most common methods of interconnections?

Q43. Name the three methods of interconnecting components in multilayer printed circuit boards.

Q44. What is one of the major disadvantages of multilayer printed circuit boards?

Q45. What was the earliest form of micromodule?

ENVIRONMENTAL CONSIDERATIONS

The environmental requirements of each system design are defined in the PROCUREMENT SPECIFICATION. Typical environmental requirements for an IC, for example, are shown in table 1-1. After these system requirements have been established, components, applications, and packaging forms are considered. This then leads to the most effective system form.

Table 1-1.—Environmental Requirements

Temperature Operating Nonoperating	–28º C to +65º C –62º C to +75º C (MIL-E-16400E)
Humidity	95 percent plus condensation (MIL-E-16400E)
Shock	250 to 600 g (MIL-S-901C)
Vibration	5 to 15 Hz, 0.060 DA 16 to 25 Hz, 0.040 DA 26 to 33 Hz, 0.020 DA Resonance test in three mutual perpendicular planes. (MIL-STD-167)
RF Interference	30 Hz to 40 GHz

In the example in table 1-1, the environmental requirements are set forth as MILITARY STANDARDS for performance. The actual standard for a particular factor is in parentheses. To meet each of these standards, the equipment or component must perform adequately within the test guidelines. For example, to pass the shock test, the component must withstand a shock of 250 to 600 $_{Gs}$ (force of gravity). During vibration testing, the component must withstand vibrations of 5 to 15 cycles per second for 0.06 day, or about 1 1/2 hours; 16 to 25 cycles for 1 hour; and 26 to 33 cycles for 1/2 hour. Rf interference between 30 hertz and 40 gigahertz must not affect the performance of the component. Temperature and humidity factors are self-explanatory.

When selecting the most useful packaging technique, the system designer must consider not only the environmental and electrical performance requirements of the system, but the maintainability aspects as well. The system design will, therefore, reflect performance requirements of maintenance and repair personnel.

ELECTRICAL CONSIDERATIONS

The electrical characteristics of a component can sometimes be adversely affected when it is placed in a given system. This effect can show up as signal distortion, an improper timing sequence, a frequency shift, or numerous other types of unwanted interactions. Techniques designed to minimize the effects of system packaging on component performance are incorporated into system design by planners. These techniques should not be altered during your maintenance. Two of the techniques used by planners are discussed in the following sections.

Ground Planes and Shielding.

At packaging levels I and II, COPPER PLANES with voids, where feed-through is required, can be placed anywhere within the multilayer board. These planes tend to minimize interference between circuits and from external sources.

At other system levels, CROSS TALK (one signal interfering with another), rf generation within the system, and external interference are suppressed through the use of various techniques. These techniques are shown in figure 1-44. As shown in the figure, rf shielding is used on the mating surfaces of the package, cabling is shielded, and heat sinks are provided.

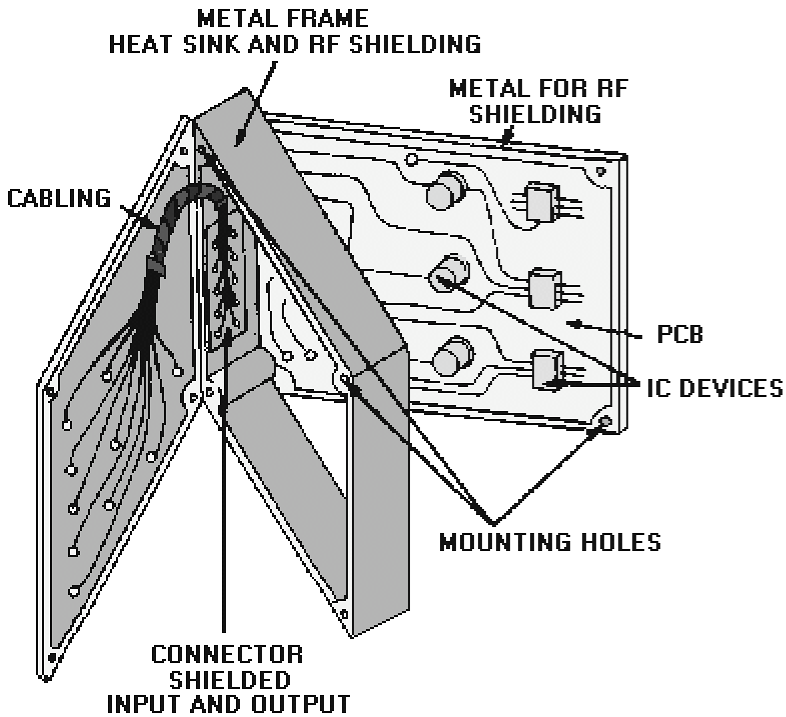


Figure 1-44.—Ground planes and shielding.

Interconnection and Intraconnections

To meet the high-frequency characteristics and propagation timing required by present and future systems, the device package must not have excessive distributed capacitance and/or inductance. This type of packaging is accomplished in the design of systems using ICs and other microelectronic devices by using shorter leads internal to the package and by careful spacing of complex circuits on printed circuit boards. To take advantage of the inherent speed of the integrated circuit, you must keep the signal propagation time between circuits to a minimum. The signal is delayed approximately 1 nanosecond per foot, so reducing the distance between circuits as much as possible is necessary. This requires the use of structures, such as high-density digital systems with an emphasis on large-scale integration, for systems in the future. Also, maintenance personnel should be especially concerned with the spacing of circuits, lead dress, and surface cleanliness. These factors affect the performance of high-speed digital and analog circuits.

Q46. In what publication are environmental requirements for equipment defined?

Q47. In what publication would you find guidelines for performance of military electronic parts?

Q48. Who is responsible for meeting environmental and electrical requirements of a system?

Q49. What methods are used to prevent unwanted component interaction?

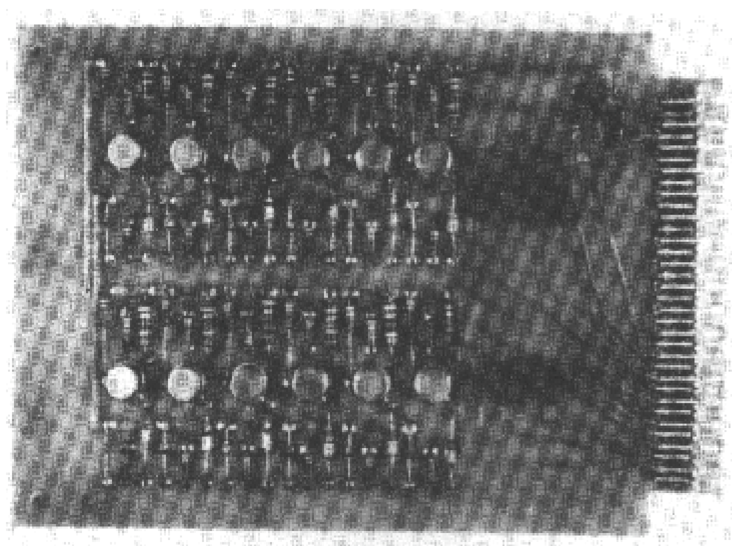
SUMMARY

This topic has presented information on the development and manufacture of microelectronic devices. The information that follows summarizes the important points of this topic.

VACUUM-TUBE CIRCUITS in most modern military equipment are unacceptable because of size, weight, and power use.

Discovery of the transistor in 1948 marked the beginning of **MICROELECTRONICS**.

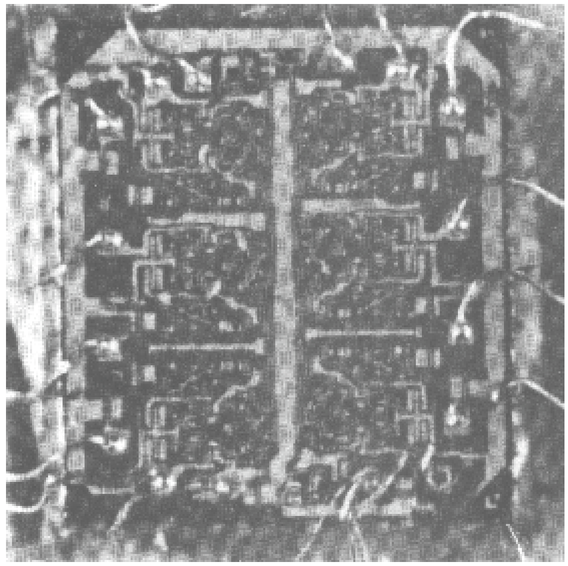
The **PRINTED CIRCUIT BOARD (pcb)** reduces weight and eliminates point-to-point wiring.



The **INTEGRATED CIRCUITS (IC)** consist of elements inseparably associated and formed on or within a single SUBSTRATE.

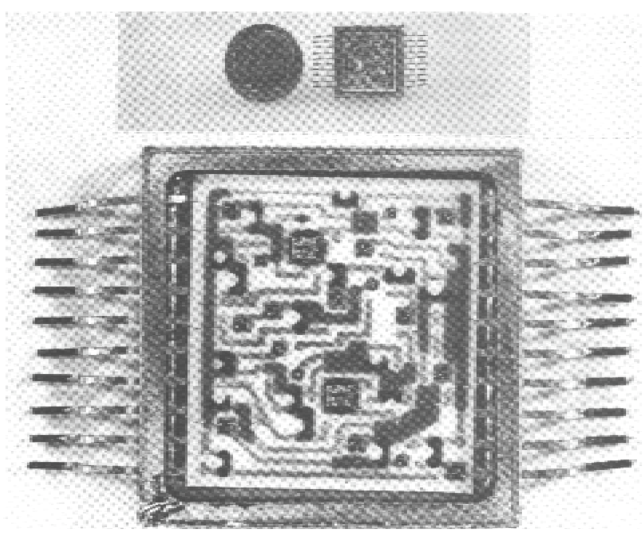
ICs are classified as three types: **MONOLITHIC, FILM**, and **HYBRID**.

The **MONOLITHIC IC**, called a chip or die, contains both active and passive elements.



FILM COMPONENTS are passive elements, either resistors or capacitors.

HYBRID ICs are combinations of monolithic and film or of film and discrete components, or any combination thereof. They allow flexibility in circuits.

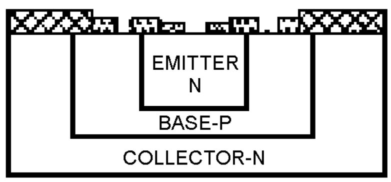


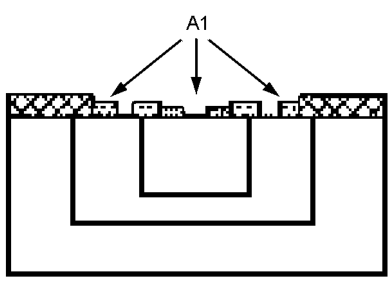
Rapid development has resulted in increased reliability and availability, reduced cost, and higher element density.

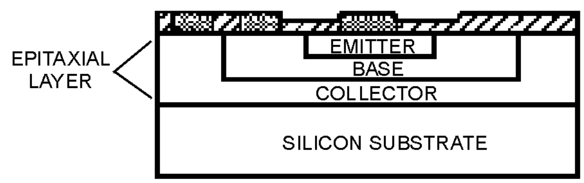
LARGE-SCALE (lsi) and **VERY LARGE-SCALE INTEGRATION (vlsi)** allow thousands of elements in a single chip.

MONOLITHIC ICs are produced by the diffusion or epitaxial methods.

DIFFUSED elements penetrate the substrate, **EPITAXIAL** do not.





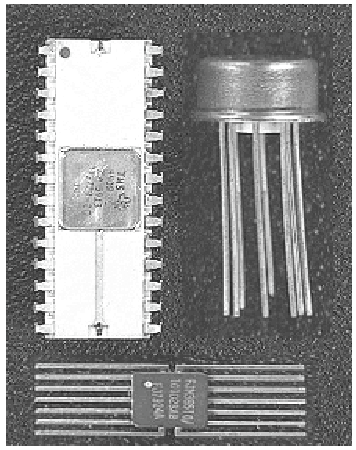


ISOLATION is a production method to prevent unwanted interaction between elements within a chip.

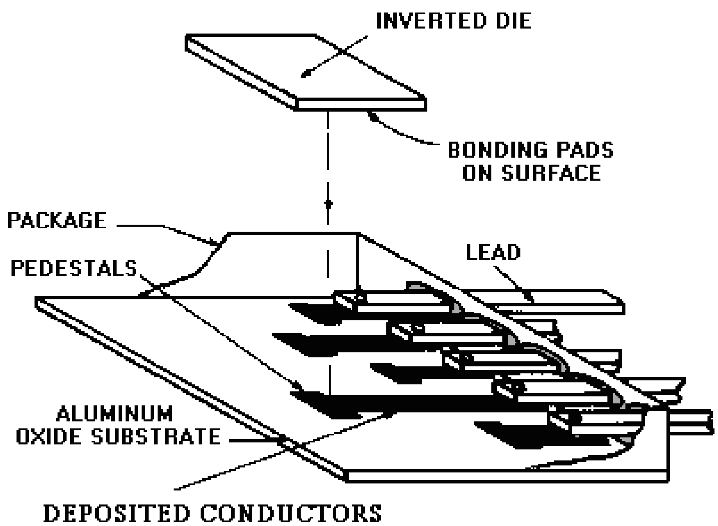
THIN-FILM ELEMENTS are produced through **EVAPORATION** or **CATHODE SPUTTERING** techniques.

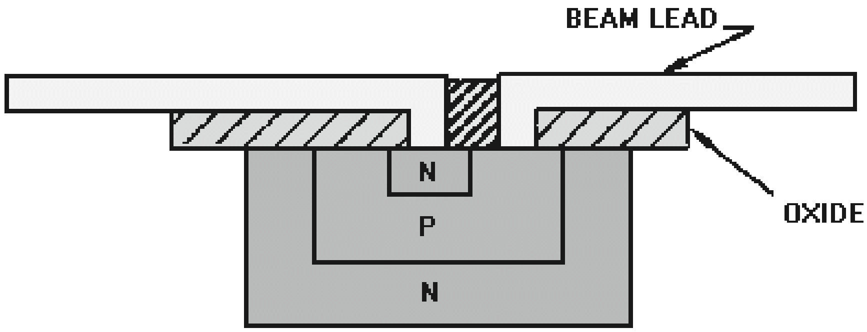
THICK-FILM ELEMENTS are screened onto the substrate.

The most common types of packages for **ICs** are **TO**, **FLAT PACK**, and **DUAL INLINE**.



FLIP CHIPS and **BEAM-LEAD CHIPS** are techniques being developed to eliminate bonding wires and to improve packaging.

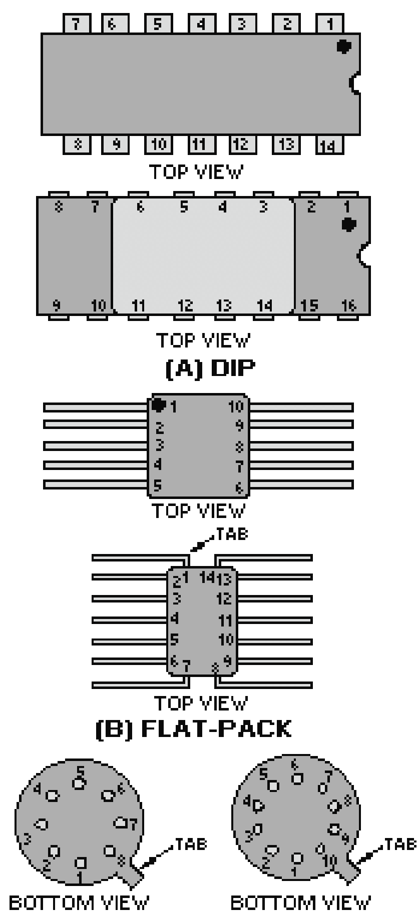




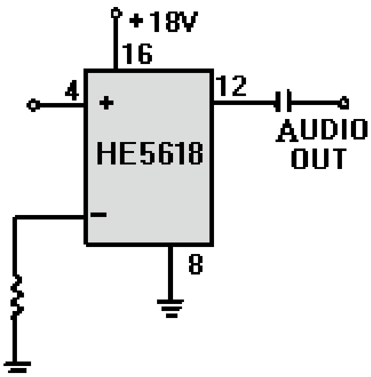
Large **DIPs** are being used to package lsi and vlsi. They can be produced with up to 64 pins and are designed to fulfill a specific need.

Viewed from the tops, **DIPS** and **FLAT-PACK LEADS** are numbered counterclockwise from the reference mark.

Viewed from the bottom, **TO-5 LEADS** are numbered clockwise from the tab.



Numbers and letters on schematics and **ICs** identify the **TYPE OF IC**.

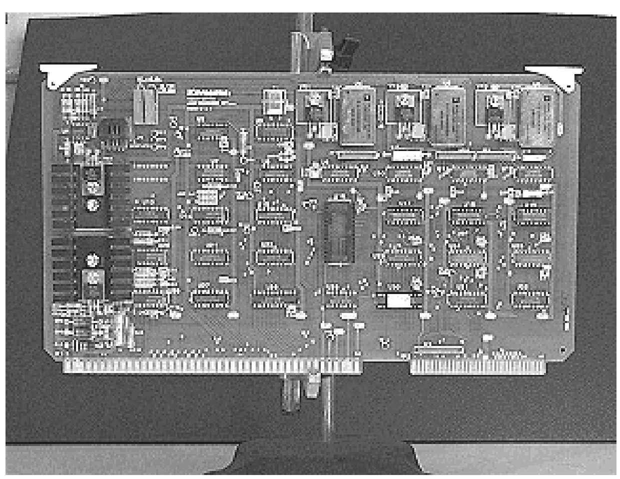


Knowledge of **TERMINOLOGY** used in microelectronics and of packaging concepts will aid you in becoming an effective technician.

STANDARD TERMINOLOGY has been adopted by the Navy to ease communication.

MICROELECTRONICS is that area of technology associated with electronic systems designed with extremely small parts or elements.

A **MICROCIRCUIT** is a small circuit which is considered as a single part composed of elements on or within a single substrate.



A **MICROCIRCUIT MODULE** is an assembly of microcircuits or a combination of microcircuits and discrete components packaged as a replaceable unit.

MINIATURE ELECTRONICS are card assemblies and modules composed exclusively of discrete electronic components.

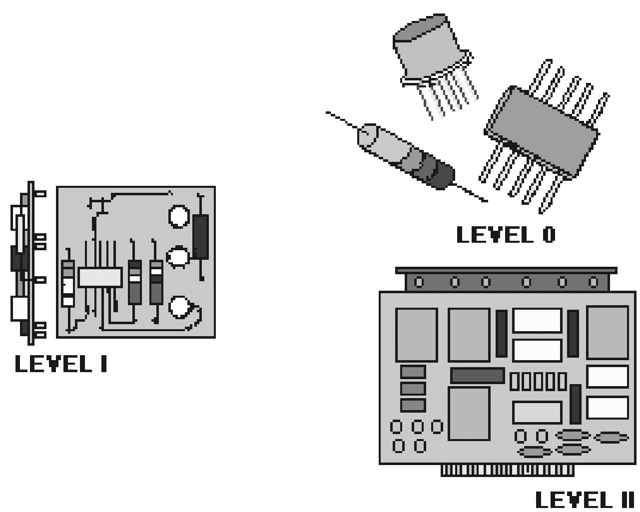
SYSTEM PACKAGING refers to the design of a system, taking into account environmental and electronic characteristics, access, and maintainability.

PACKAGING LEVELS 0 to IV are used to identify assemblies within a system. Packaging levels are as follows:

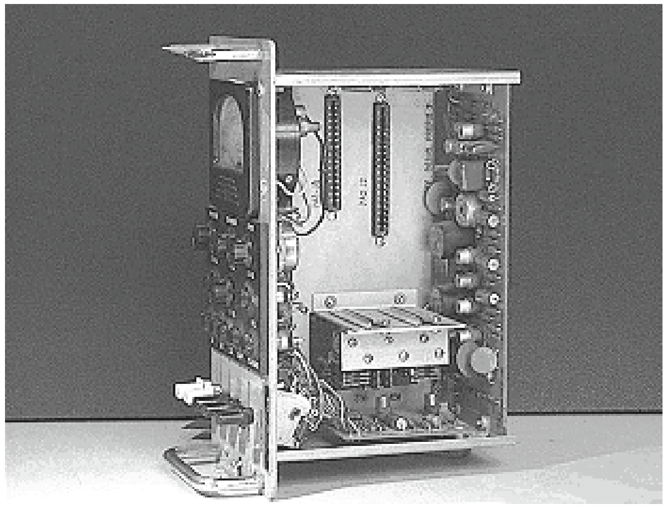
LEVEL 0-Nonrepairable parts (resistors, diodes, etc.)

LEVEL I -Submodules attached to circuit cards.

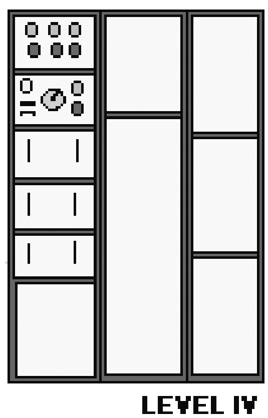
LEVEL II -Circuit cards and **MOTHER BOARDS**.



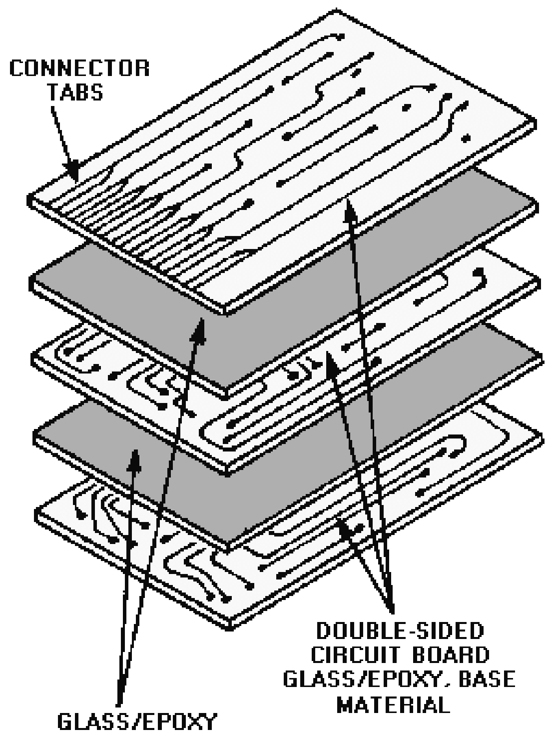
LEVEL III - Drawers.



LEVEL IV - Cabinets.



The most common **METHODS OF INTERCONNECTION** are the conventional pcb, the multilayer pcb, and modular assemblies.

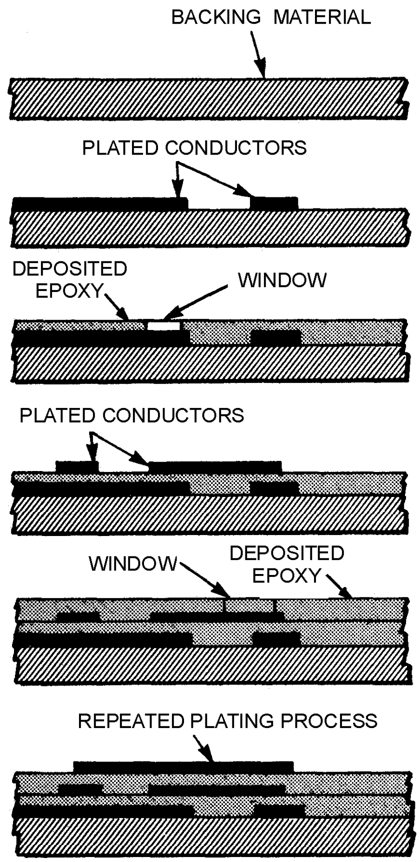


Three methods of interconnecting circuitry in multilayer printed circuit boards are the **CLEARANCE-HOLE**, the **PLATED-THROUGH-HOLE**, and **LAYER BUILD-UP**.

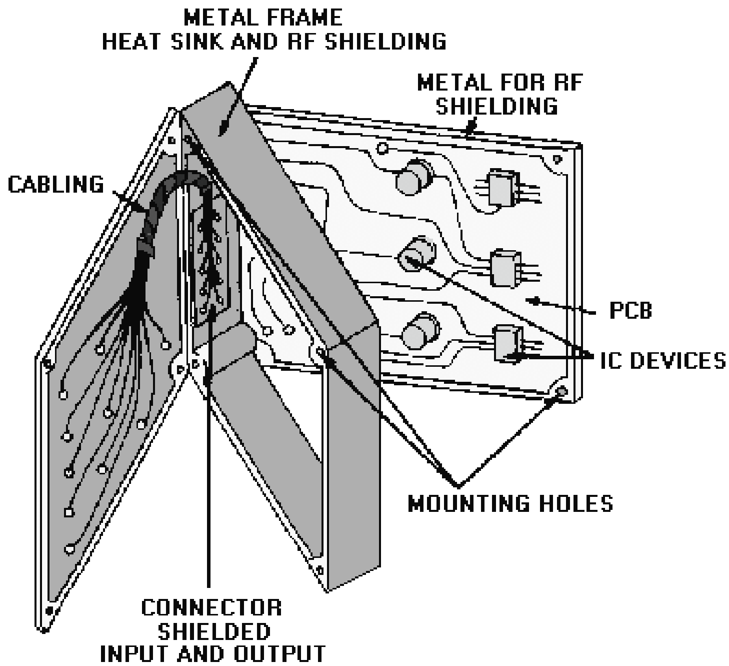
MODULAR ASSEMBLIES were devised to achieve high circuit density.

Modular assemblies have progressed from **CORDWOOD MODULES** through **MICROMODULES**. Micromodules consist of film components and discrete components to integrated and hybrid circuitry.

ENVIRONMENTAL FACTORS to be considered are temperature, humidity, shock, vibration, and rf interference.



ELECTRICAL FACTORS are overcome by using shielding and ground planes and by careful placement of components.



ANSWERS TO QUESTIONS Q1. THROUGH Q50.

A1. *Size, weight, and power consumption.*

A2. *The transistor and solid-state diode.*

A3. *Technology of electronic systems made of extremely small electronic parts or elements.*

A4. *The Edison Effect.*

A5. *Transformers, capacitors, and resistors.*

A6. *"Rat's nest" appearance and unwanted interaction, such as capacitive and inductive effects.*

A7. *Rapid repair of systems and improved efficiency.*

A8. *Differences in performance of tubes of the same type.*

A9. *Eliminate heavy chassis and point-to-point wiring.*

A10. *Components soldered in place.*

A11. *Cordwood module.*

A12. *Elements inseparably associated and formed in or on a single substrate.*

A13. *Monolithic, film, and hybrid.*

A14. *Monolithic ICs contain active and passive elements. Film ICs contain only passive elements.*

A15. Combination of monolithic ICs and film components.

A16. 1,000 to 2,000.

A17. Circuit design, component placement, suitable substrate, and depositing proper materials on substrate.

A18. Complex.

A19. Control patterns of materials on substrates.

A20. Glass or ceramic.

A21. Crystal is sliced into wafers. Then ground and polished to remove any surface defect.

A22. Diffusion; epitaxial growth.

A23. Diffusion penetrates substrate; epitaxial does not.

A24. Electrical separation of elements.

A25. Evaporation and cathode sputtering.

A26. Screening.

A27. Combination of monolithic and film elements.

A28. Circuit flexibility.

A29. Protect the IC from damage; make handling easier.

A30. TO, flat pack, DIP.

A31. Flip-chip, beam lead.

A32. Left.

A33. Counterclockwise.

A34. Clockwise.

A35. Identify the type of IC.

A36. Communications.

A37. Integrated circuits.

A38. Miniature.

A39. Level 0.

A40. Level I.

A41. Level II.

A42. Conventional printed circuit boards, multilayer printed circuit boards and modular assemblies.

A43. Clearance hole, plated-through hole, and layer build-up.

A44. Difficulty of repair of internal connections.

A45. Cordwood modules.

A46. Procurement specifications.

A47. Military Standards.

A48. Equipment designers (planners).

A49. Ground planes, shielding, component placement.

CHAPTER 2

MINIATURE/MICROMINIATURE (2M) REPAIR PROGRAM AND HIGH-RELIABILITY SOLDERING

LEARNING OBJECTIVES

Upon completion of this topic, the student will be able to:

1. State the purpose and need for training and certification of 2M repair technicians.
2. Explain the maintenance levels at which maintenance is performed.
3. Identify the specialized and general test equipment used in fault isolation.
4. Recognize the specialized types of tools used and the importance of repair facilities.
5. Explain the principles of high-reliability soldering.

INTRODUCTION

As mentioned in topic 1, advances in the field of microelectronics are impressive. With every step forward in production development, a corresponding step forward must be made in maintenance and repair techniques.

This topic will teach you how the Navy is coping with the new technology and how personnel are trained to carry out the maintenance and repair of complex equipment. The program discussed in this topic is up to date at this time, but as industry advances, so must the capabilities of the technician.

MINIATURE AND MICROMINIATURE (2M) ELECTRONIC REPAIR PROGRAM

Training requirements for miniature and microminiature repair personnel were developed under guidelines established by the Chief of Naval Operations. The Naval Sea Systems Command (NAVSEA) developed a program which provides for the proper training in miniature and microminiature repair. This program, NAVSEA Miniature/Microminiature (2M) Electronic Repair, authorizes and provides proper tools and equipment and establishes a personnel certification program to maintain quality repair.

The Naval Air Systems Command has developed a similar program specifically for the aviation community. The two programs are patterned after the National Aeronautics and Space Administration (NASA) high-reliability soldering studies and have few differences other than the administrative chain of command. For purposes of this topic, we will use the NAVSEA manual for reference.

The 2M program covers all phases of miniature and microminiature repair. It establishes the training curriculum for repair personnel, outlines standards of workmanship, and provides guidelines for specific repairs to equipment, including the types of tools to use. This part of the program ensures high-reliability repairs by qualified technicians.

Upon satisfactory completion of a 2M training course, a technician will be CERTIFIED to perform repairs. The CERTIFICATION is issued at the level at which the technician qualifies and specifies what type of repairs the technician is permitted to perform. The two levels of qualification for technicians are MINIATURE COMPONENT REPAIR and MICROMINIATURE COMPONENT REPAIR. Miniature component repair is limited to discrete components and single- and double-sided printed circuit boards, including removal and installation of most integrated circuit devices. Microminiature component repair consists of repairs to highly complex, densely packaged, multilayer printed circuit boards. Sophisticated repair equipment is used that may include a binocular microscope.

To ensure that a technician is maintaining the required qualification level, periodic evaluations are conducted. By inspecting and evaluating the technician's work, certification teams ensure that the minimum standards for the technician's level of qualification are met. If the standards are met, the technician is recertified; if not, the certification is withheld pending retraining and requalification. This portion of the program ensures the high-quality, high-reliability repairs needed to meet operational requirements.

Q1. Training requirements for (2M) repair personnel were developed under guidelines established by what organization?

Q2. What agencies provide training, tools, equipment, and certification of the 2M system?

Q3. To perform microminiature component repair, a 2M technician must be currently certified in what area?

Q4. Multilayer printed circuit board repair is the responsibility of what 2M repair technician?

LEVELS OF MAINTENANCE

Effective maintenance and repair of microelectronic devices require one of three levels of maintenance. Level-of-repair designations called SOURCE, MAINTENANCE, and RECOVERABILITY CODES (SM&R) have been developed and are assigned by the Chief of Naval Material. These codes are D for DEPOT LEVEL, I for INTERMEDIATE LEVEL, and O for ORGANIZATIONAL LEVEL.

DEPOT-LEVEL MAINTENANCE.

SM&R Code D maintenance is the responsibility of maintenance activities designated by the systems command (NAVSEA, NAVAIR, NAVELEX). This code augments stocks of serviceable material. It also supports codes I and O activities by providing more extensive shop facilities and equipment and more highly skilled technicians. Code D maintenance includes repair, modification, alteration, modernization, and overhaul as well as reclamation or reconstruction of parts, assemblies, subassemblies, and components. Finally, it includes emergency manufacture of nonavailable parts. Code D maintenance also provides technical assistance to user activities and to code I maintenance organizations. Code D maintenance is performed in shops, located in shipyards and shore-based facilities, including contractor maintenance organizations.

INTERMEDIATE-LEVEL MAINTENANCE

SM&R Code I maintenance, performed at mobile shops, tenders on shore-based repair facilities (SIMAS) provides *direct* support to user organizations. Code I maintenance includes calibration, repair, or replacement of damaged or unserviceable parts, components, or assemblies, and emergency manufacture of nonavailable parts. It also provides technical assistance to ships and stations.

ORGANIZATIONAL-LEVEL MAINTENANCE

SM&R Code O maintenance is the responsibility of the activity who owns the equipment. Code O maintenance consists of inspecting, servicing, lubricating, adjusting, and replacing parts, minor assemblies, and subassemblies.

An INTEGRATED LOGISTICS SUPPORT PLAN (ILSP) determines the maintenance level for electronic assemblies, modules, and boards for each equipment assigned to an activity. The ILSP codes the items according to the normal maintenance capabilities of that activity. This results in two additional repair-level categories - NORMAL and EMERGENCY.

Normal Repairs

Generally, 2M repairs are performed at the level set forth in the maintenance plan and specified by the appropriate SM&R coding for each board or module. Therefore, normal repairs include all repairs except organizational-level repair of D- and I-coded items and intermediate-level repair of D-coded items.

Emergent/Emergency Repairs

In the NAVSEA 2M Electronic Repair Program, emergent/emergency repairs are those arising unexpectedly. They may require prompt repair action to restore a system or piece of equipment to operating condition where normal repairs are not authorized. These Code O repairs on boards or modules are normally SM&R-coded for Code D repairs. Emergent/emergency 2M repairs may be performed only to meet an urgent operational commitment as directed by the operational commander.

SOURCE, MAINTENANCE, AND RECOVERABILITY (SM&R) CODES

The Allowance Parts List (APL) is a technical document prepared by the Navy for specific equipment/system support. This document lists the repair parts requirements for a ship having the exact equipment/component. To determine the availability of repair parts, the 2M technician must be familiar with these documents. SM&R codes, found in APLs, determine where repair parts can be obtained, who is authorized to make the repair, and at what maintenance level the item may be recovered or condemned.

Q5. What are the three levels of maintenance?

Q6. Maintenance performed by the user activity is what maintenance level?

TEST EQUIPMENT

Microelectronic developments have had a great impact on the test equipment, tools, and facilities necessary to maintain systems using this technology. This section discusses, in general terms, the importance of these developments.

Early electronic systems could be completely checked-out with general-purpose electronic test equipment (GPETE), such as multimeters, oscilloscopes, and signal generators. Using this equipment to test the microelectronics components individually in one of today's very complex electronic systems would be extremely difficult if not impossible. Therefore, improvements in system testing procedures have been necessary.

One such improvement in system testing is the design of a method that can test systems at various *functional levels*. This allows groups of components to be tested as a whole and reduces the time required to test components individually. One advantage of this method is that complete test plans can be written to provide the best sequencing of tests for wave shape or voltage outputs for each functional level. This method of testing has led to the development of special test sets, called AUTOMATED TEST EQUIPMENT (ATE). These test sets are capable of simulating actual operating conditions of the system being tested. Appropriate signal voltages are applied by the test set to the various functional levels of the system, and the output of each level is monitored. Testing sequences are prewritten and steps may be switched-in manually or automatically. The limits for each functional level are preprogrammed to give either a "go/no-go" indication or diagnose a fault to a component. A go/no-go indication means that a functional level either meets the test specifications (go) or fails to meet the specifications (no-go).

If a no-go indication is observed for a given function, the area of the system in which it occurs is then further tested. You can test the trouble area by using general-purpose electronic test equipment and the troubleshooting manual for the system. General purpose electronic test equipment (GPETE) will be discussed later in this topic. (Effective fault isolation at this point depends on the experience of the technician and the quality of the troubleshooting manual.) After the fault is located, the defective part is then replaced or repaired, depending on the nature of the defect. At this stage, the defective part is usually a circuit card, a module, or a discrete part, such as a switch, relay, transistor, or resistor.

BUILT-IN TEST EQUIPMENT

One type of fault isolation that can be either on-line or off-line is BUILT-IN TEST EQUIPMENT (BITE). BITE is any device that is permanently mounted in the prime equipment (system); it is used only for testing the equipment or system in which it is installed either independently or in association with external test equipment. The specific types of BITE are too varied to discuss here, but may be as simple as a set of meters and switches or as complex as a computer-controlled diagnostic system.

ON-LINE TEST EQUIPMENT

Functional-level testing and modular design have been successfully applied to most electronic systems in use today; however, the trend toward increasing the number of subassemblies within a module by incorporating microelectronics will make this method of testing less and less effective.

The increased circuit density and packaging possible with microelectronic components makes troubleshooting and fault location difficult or, in some cases, impossible. The technician's efforts must be aided if timely repairs to microelectronic systems are to be achieved. These repairs are particularly significant when considered in the light of the very stringent availability requirements for today's systems. This dilemma has led to the present trend of developing both ON-LINE and OFF-LINE automatic test systems. The on-line systems are designed to monitor performance continuously and to isolate faults automatically to removable assemblies. Off-line systems automatically check removable assemblies and isolate faults to the component level.

OFF-LINE TEST EQUIPMENT

Over the years several off-line testing programs have been developed. The ones mentioned below are being used throughout the fleet today.

Consolidated Automated Support System

Consolidated Automated Support System (CASS) has become the United States Navy’s Aviation community standard in automated test equipment (ATE), it has been established to replace 24 different types of computer based ATE systems with one modern, cost effective family of testers. The system increases repair capabilities and material readiness, decreases overall support and life cycle costs, and reduces the physical space currently required by other types of electronic testing equipment. Currently the system repairs over 600 different assets. As new test programs are written, it will support over 2,300 different Weapons Replaceable Assemblies (WRAs) and Shop Replaceable Assemblies (SRAs) from current and planned Naval Aviation inventory of aircraft to include F/A-18 (all models), EA-6B, F-14 (all models), H-60, E-2C, and V-22. The test set has automated capabilities for testing avionics systems whose technology encompasses frequency stimulus and measurement and digital functions. The capabilities include power conditioning, interface control, calibration, self-maintenance, instrumentation, and software functions necessary to perform end-to-end tests, fault isolation tests, and alignment or adjustment of units under test (UUT). UUTs are interfaced to the test set by test program sets (TPSs), interface devices (IDs), and accessories. Since one tester capable of meeting all these needs would be too large, CASS was designed to be modular. Based upon functional needs, configuration of CASS benches are grouped and defined as distinct configurations.

Module Test and Repair and Gold Disk Development

Module Test and Repair (MTR) Program develops and provides electrical/electronic module test and repair capabilities to organizational (O-level) and intermediate (I-level) maintenance facilities for both ashore and afloat commands. It also serves as the In-Service Engineering Agent (ISEA) for various MTR equipment, development and provides technical manuals on CD ROM, develops gold disks for circuit card assemblies, certifies gold disk developers and developer sites, provides for the maintenance and operation of MTR facilities, provides technical assistance and guidance for all aspects of the MTR Program; and is the gold disk verification agent for all DoD activities.

Gold Disk Development

A gold disk is a diagnostic troubleshooting routine used to isolate faulty components on circuit card assemblies (CCAs) or electronic modules (EMs). The same technology may also be used to determine whether an assembly is no fault evident (NFE). The primary purpose of a gold disk is to provide the technician all required logistical information to identify component level failures and complete necessary repair. The distribution of gold disk routines DoD-wide permits an organic repair capability to users possessing the AN/USM-646 (V) and AN/USM-658 (V) test stations. The diagnostic enhancement and organic repair capability supports reduction of funding expenditures and improvement of operational availability. Repair accomplished at the lowest practical maintenance level (progressive repair) assists activities in offsetting the cost of shipping suspect faulty CCAs/EMs back to a depot for repair/replacement.

GENERAL-PURPOSE ELECTRONIC TEST EQUIPMENT (GPETE)

When no automatic means of accomplishing fault isolation is available, general-purpose electronic test equipment and good troubleshooting procedures is used; however, such fault diagnosis should be attempted only by experienced technicians. Misuse of electrical probes and test equipment may permanently damage boards or microelectronic devices attached to them. The proximity of leads to one another and the effects of interconnecting the wiring make the testing of boards extremely difficult; these factors also make drift or current leakage measurements practically impossible.

Boards that have been conformally coated are difficult to probe because the coating is often too thick to penetrate for a good electrical contact. These boards must be removed for electrical probe testing. Many boards, however, are designed with test points that can be monitored either with special test sets or general-purpose test equipment. Another method of obtaining access to a greater number of test points is to use extender cards or cables. The use of extender cards or cables makes these test points easier to check.

Special care should be exercised when probing integrated circuits; they are easily damaged by excessive voltages or currents, and component leads may be physically damaged. Precautions, concerning the use of test equipment for troubleshooting equipments containing integrated circuits, are similar to those that should be observed when troubleshooting equipment containing semiconductor or other voltage and current-sensitive devices.

Voltage and resistance tests of resistors, transistors, inductors, and so forth, are usually effective in locating complete failures or defects that exhibit large changes from normal circuit characteristics; however, these methods are time-consuming and sometimes unsuccessful. The suspect device often must be desoldered, removed from the circuit, and then retested to verify the fault. If the defect is not verified, the device must be resoldered to the board again. If this procedure has to be repeated several times, or if the board is conformally coated, the defect may never be located. In fact, the circuit may be further damaged by the attempt to locate the fault. *For these reasons, the device should never be desoldered until all possible in-circuit tests are performed and the defect verified.*

Q7. List the three groups of test equipment used for fault isolation in 2M repair.

Q8. What test equipment continuously monitors electronic systems?

Q9. NELAT and VAST are examples of what type of test equipment?

REPAIR STATIONS

In addition to the requirements for special skills, the repair of 2M electronic circuits also requires special tools. Because these tools are delicate and expensive, they are distributed only to trained and certified 2M repair technicians.

2M repair stations are equipped with electrical and mechanical units, tools, and general repair materials. Such equipments are needed to make reliable repairs to miniature and microminiature component circuit boards.

Although most of the tools and equipments are common to both miniature and microminiature repair stations, several pieces of equipment are used solely with microminiature repair. Precision drill presses and stereoscopic-zoom microscopes are examples of microminiature repair equipment normally not found in a miniature repair station. A brief description of some of the tools and equipments and their uses will broaden your knowledge and understanding of 2M repair.

The 2M repair set consists of special electrical units, tools, and materials necessary to make high-reliability repairs to component circuitry. The basic repair set is made up of a repair station power unit, magnifier/light system, card holder, a high-intensity light, a Pana Vise, and a tool chest with specialized tools and materials. As mentioned previously, stations that have microminiature repair capabilities will include a stereoscopic-zoom microscope and precision drill press.

REPAIR STATION POWER UNIT

The repair station power unit is a standardized system that provides controlled soldering and desoldering of all types of solder joint configurations. The unit is shown in figure 2-1. Included in the control unit's capabilities are:

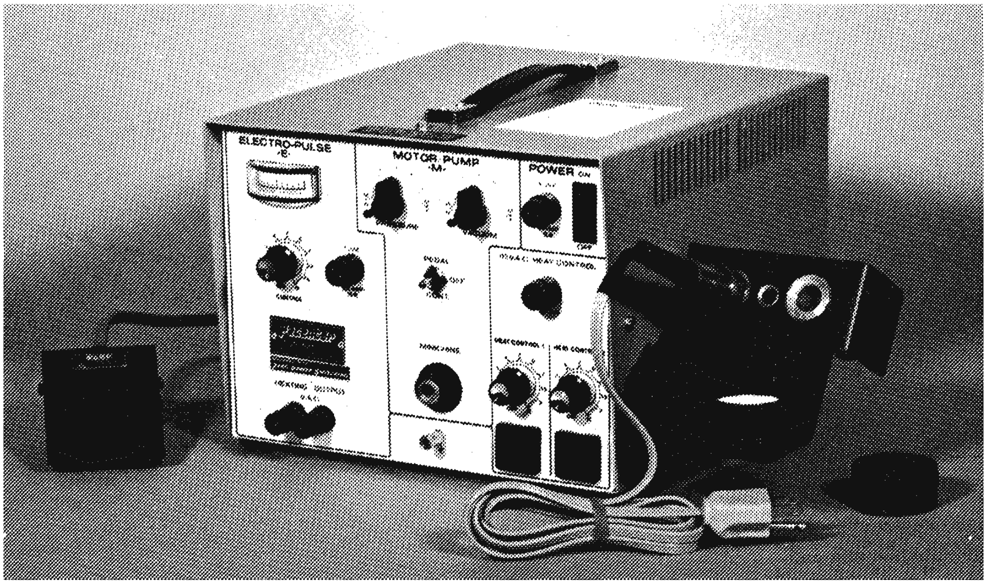


Figure 2-1.—Repair station power unit.

- "Spike free" power switching for attached electrical hand tools to eliminate damage to electrostatic discharge components.
- Abrading, milling, drilling, grinding, and cutting using a flexible shaft, rotary-drive machine. This allows the technician to remove conformal coatings, oxides, eyelets, rivets, damaged board material, and damaged platings from assemblies.
- Lap flow solder connections and thermal removal of conformal coatings.
- Resistive and conductive tweezer heating for connector soldering applications.
- Thermal wire stripping for removing polyvinyl chloride (PVC) and other synethetic wire coverings.

Power Source

The basic unit houses the power supply, power level indicator, motor control switch, hand tool temperature controls, air pressure and vacuum controls with quick connect fittings, positive ground terminal, the mechanical power-drive for the rotary-drive machine, and a vacuum/pressure pump. A two-position foot pedal, to the left of the power unit in the illustration, allows hand-free operation for all ancillary (additional) handpieces. The first detent on the pedal provides power to the voltage heating outputs. The second detent activates the motor drive or vacuum/pressure pump.

Handpieces

The handpieces used with the power unit are shown in figures 2-2 and 2-3. The lap flow handpiece, view (A) of figure 2-2, is used with the variable low-voltage power source. This handpiece allows removal of conformal coatings, release of sweat joints, and lap flow soldering capability. (Lap flow soldering will be discussed in topic 3.) The thermal wire stripper in view (B) is used to remove insulation from various sizes of wire easily and cleanly.

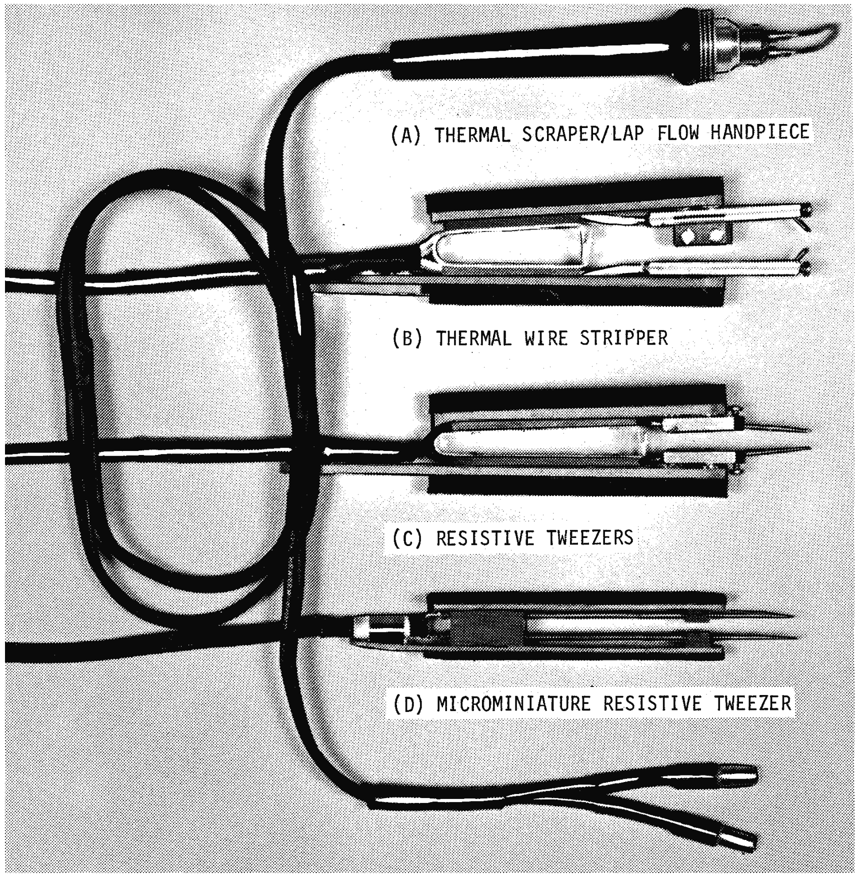


Figure 2-2.—Low voltage handpiece.

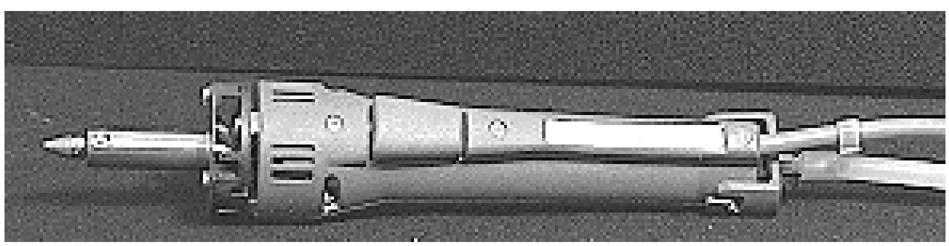


Figure 2-3.—Motorized solder extrator.

The resistive tweezers, shown in view (C), are used for soldering components. Two sizes [views (C) and (D)] are provided to meet the needs of the technician. Both the thermal stripper and the resistive tweezers are used with the low-voltage power supply.

The solder extractor, shown in figure 2-3, is connected to the variable high-voltage outlet. This handpiece allows airflow application (at controlled temperatures) of a vacuum or pressure to the selected area. Five sizes of extractor tips are provided. You can determine the one to be used by matching the tip with the circuit pad and the component being desoldered.

Soldering Irons

A soldering iron is shown in figure 2-4. This is connected to the 115-volt ac variable outlet of the power unit. You control the temperature by adjusting the voltage. The iron has replaceable tips. Chosen for their long life and good heat conductivity, soldering iron tips are high quality with iron-clad over copper construction. The tip shape and size and the heat range used are determined by the area and mass to be soldered.

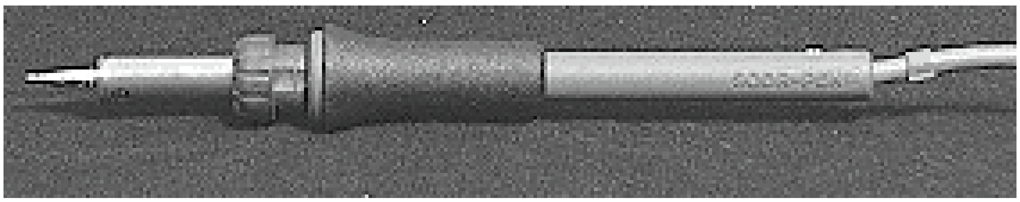


Figure 2-4.—Soldering iron.

ROTARY-DRIVE MACHINE

This variable-speed, rotary power drive adapts to standard diameter shank drill bits, ball mills, wheels, disks, brushes, and mandrels for most drilling and abrasive removal techniques (figure 2-5).

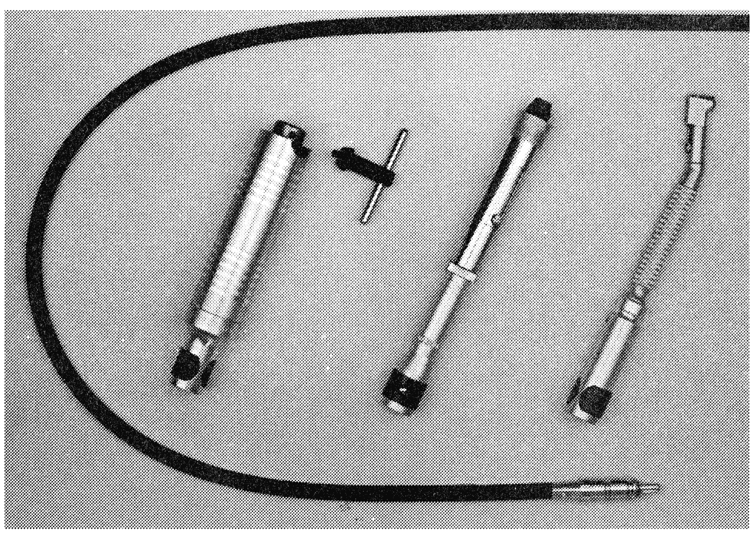


Figure 2-5.—Rotary-drive machine handpieces.

The accessories, used with the rotary-drive tool, are shown in views (A) through (F) of figure 2-6. Abrasive ball mills, wheels, discs, and brushes are either premounted on mandrels or can be mounted by the technician on the mandrels provided. These attachments are used for sanding and smoothing repaired areas, drilling holes, removing conformal coatings, and repairing burned or damaged areas. A chuck-equipped handpiece allows it to accept rotary tools with varying shank sizes.

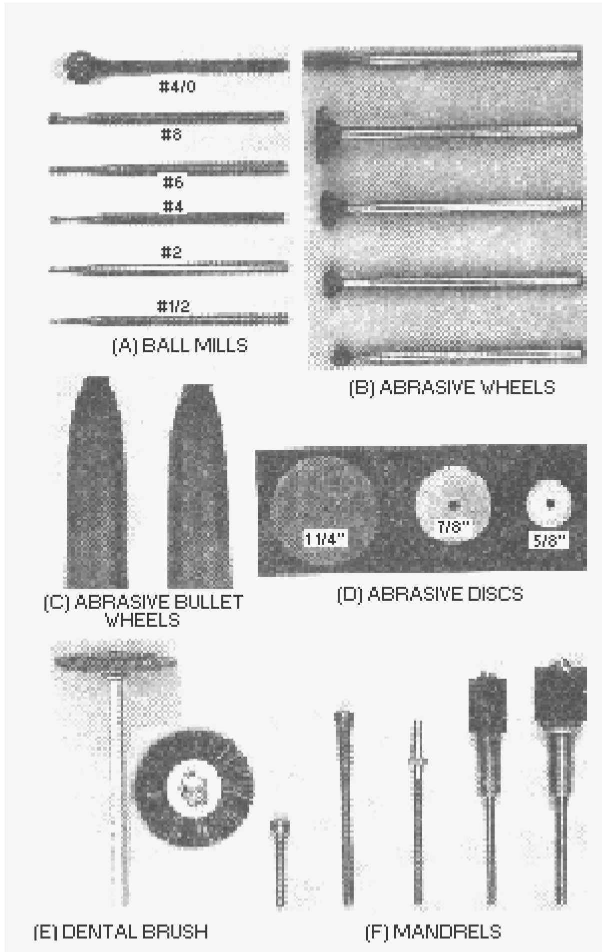


Figure 2-6.—Rotary-drive machine accessories. BALL MILLS

CIRCUIT CARD HOLDER AND MAGNIFIER

The circuit card holder is an adjustable, rotatable holder for virtually any size circuit card. Figure 2-7 shows the circuit card holder [view (A)] and the magnifier unit [view (B)]. The magnifier unit provides magnification when detail provided by a microscope is not required. The special lens allows the technician to view a rectangular area of over 14 square inches with low distortion, fine resolution, and excellent depth of field. The magnifier unit, which includes high intensity lamps, adapts to the vertical shaft of the circuit card holder.

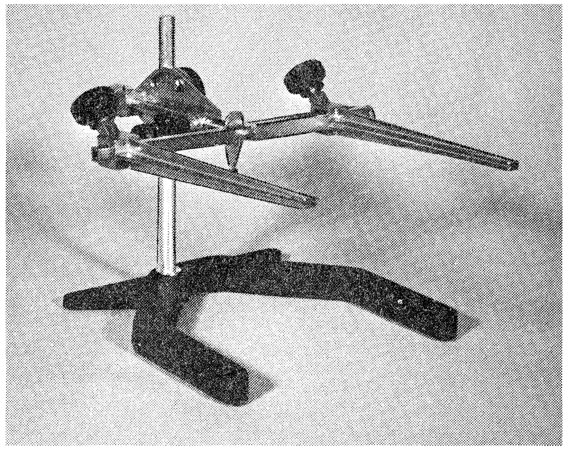


Figure 2-7.—Card holder and magnifier.

HIGH-INTENSITY LIGHT

The high-intensity light provides a variable, high-intensity, portable light source over the work area. The two flexible arms permit both front and back lighting of the workpiece and provide a balanced light that eliminates shadows (figure 2-8).

Figure 2-8.—High intensity lamp.

The high-intensity light uses 115-volt, 60-hertz input power. One brightness knob controls a flood-type bulb, and the other knob controls a spot-type bulb.

PANA VISE

This nylon-jawed, multiposition vise can rotate and tilt. With this flexibility the technician can achieve any compound angle for holding a workpiece during assembly, modification, or repair (figure 2-9).

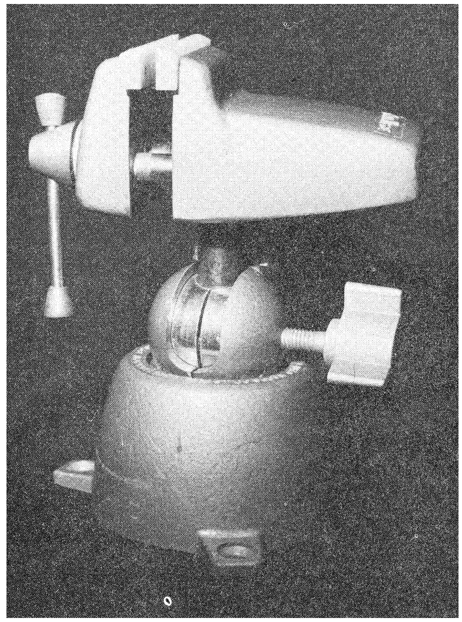


Figure 2-9.—Pana Vise.

HAND TOOLS

Figure 2-10, views (A) through (C), shows some representative types of hand tools used in 2M repair procedures.

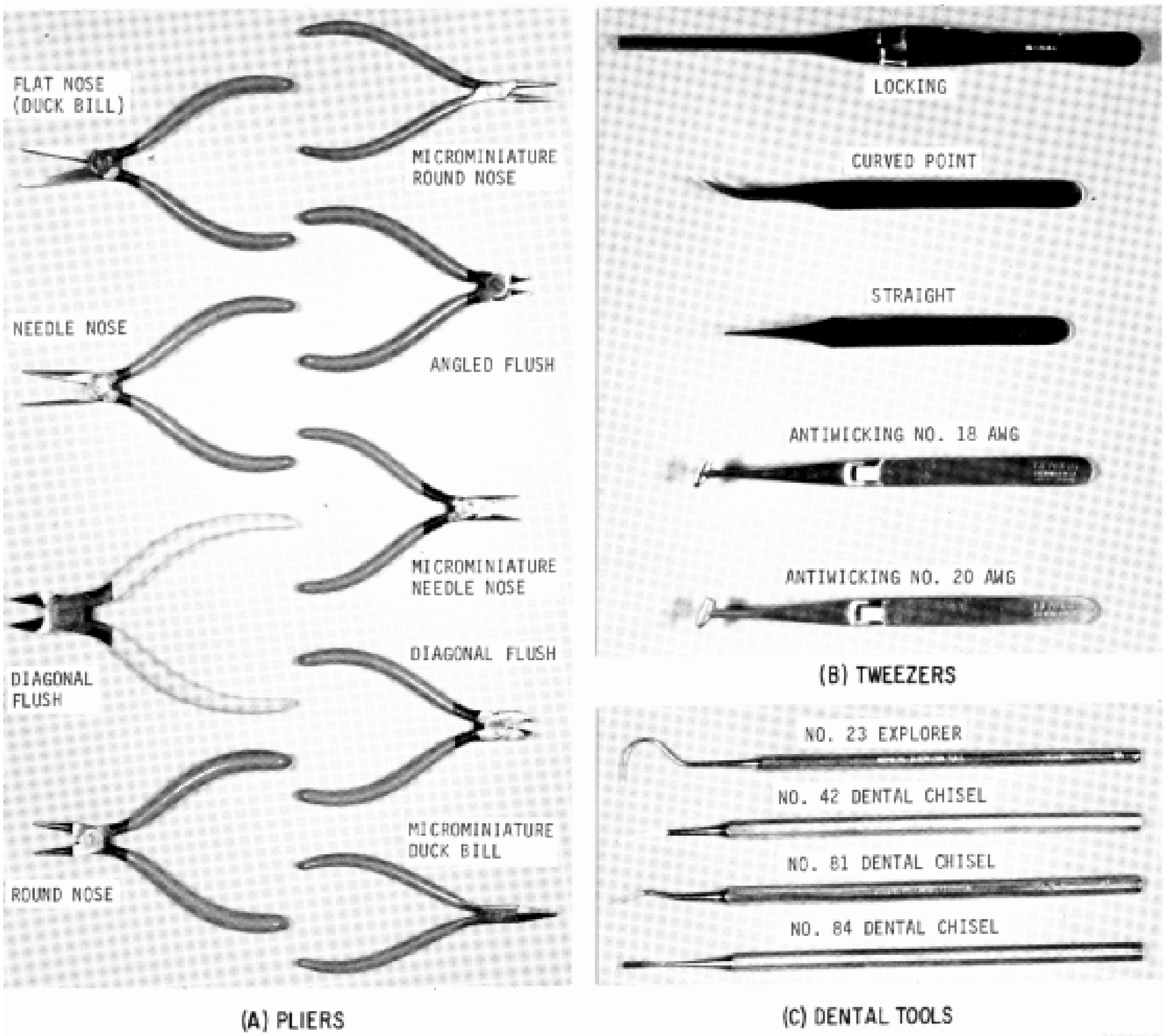


Figure 2-10.—Pliers, tweezers, and dental tools.

Pliers

In view (A), the figure shows the pliers preferred for 2M repair procedures. These precision pliers have a long and useful life if handled and cared for properly. The flush-cutting pliers are used to cut various sizes of wire and component leads. The needlenose, roundnose, and flatnose pliers are used for forming, looping, and bending wires and component leads. They are also used for gripping components and leads during removal or installation.

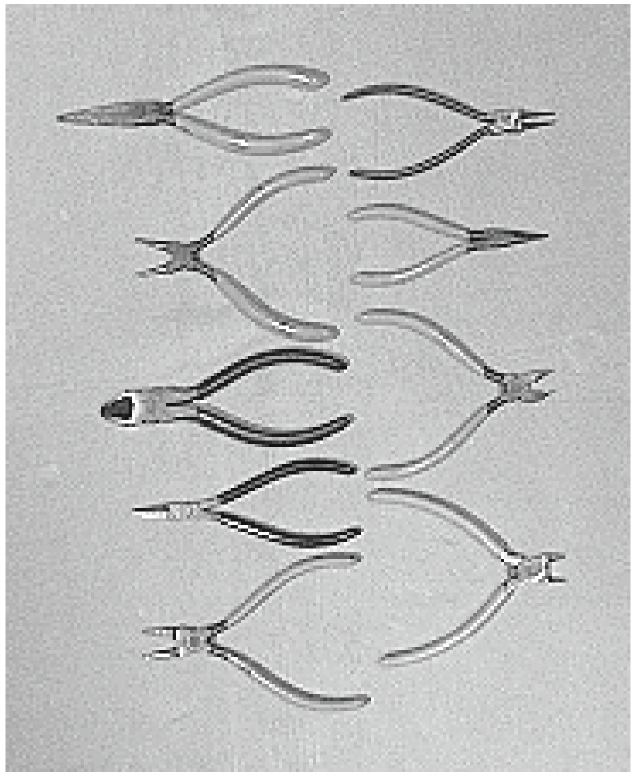


Figure 2-10a.—Pliers.

Tweezers

View (B) shows tweezers contained in the 2M repair set. The top two pairs of tweezers are used to hold small components during installation and repair procedures. The other pairs are anti-wicking tweezers used to tin and solder stranded wire leads.

Dental Tools

View (C) shows some of the dental tools contained in the 2M repair set. They are used for picking, chipping, abrading, mixing, and smoothing various conformal coatings used on printed circuit boards and other general pcb repair techniques.

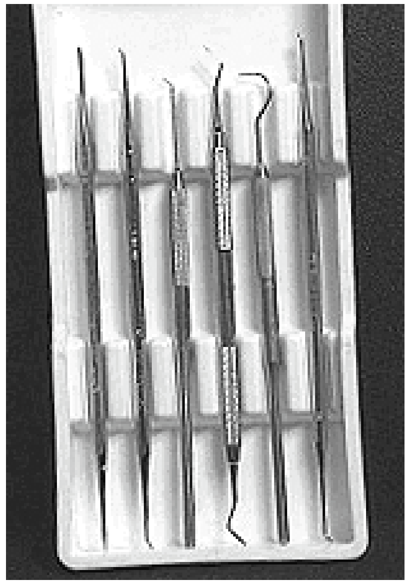


Figure 2-10c.—Dental tools.

Eyelet-Setting Tools

Among the repair procedures required of the 2M repair technician is the replacement of eyelets. Eyelets must sometimes be replaced because of the damage caused by incorrect repair procedures or complete failure of a printed circuit board. Figure 2-11 illustrates the tools used to replace these eyelets. Eyelets will be discussed in topic 3.

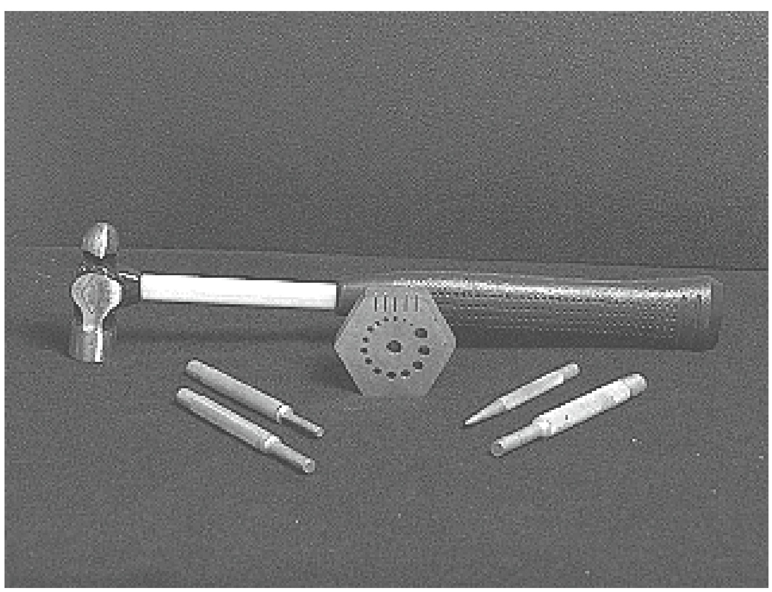


Figure 2-11.—Eyelet-setting tools.

MISCELLANEOUS TOOLS AND SUPPLIES

An assortment of some of the miscellaneous items used in 2M repair are shown in figure 2-12. A variety of brushes, files, scissors, thermal shunts, and consumables, such as solder wick, are included.

Even though all the items are not used in every repair procedure, it is extremely important that they be available for use should the need arise.

Figure 2-12.—Miscellaneous tools and supplies.

SAFETY EQUIPMENT

The nature of 2M repair requires items to be included in the tool kit for the personal safety of the technicians. The goggles and respirator illustrated in figure 2-13 have been approved for use by the technician. These should be worn at all times where dust, chips, fumes, and other hazardous substances are generated as a result of drilling, grinding, or other repair procedures.

Figure 2-13.—Safety equipment.

STEREOSCOPIC-ZOOM MICROSCOPE

The stereoscopic-zoom microscope provides a versatile optical viewing system. This viewing system is used in the fault detection, fault isolation, and repair of complex microminiature circuit boards and components. Figure 2-14 shows the microscope mounted on an adjustable stand. The microscope has a minimum of 3.5X and a maximum of 30X magnification to detect hairline cracks in conductor runs and stress cracks in solder joints.

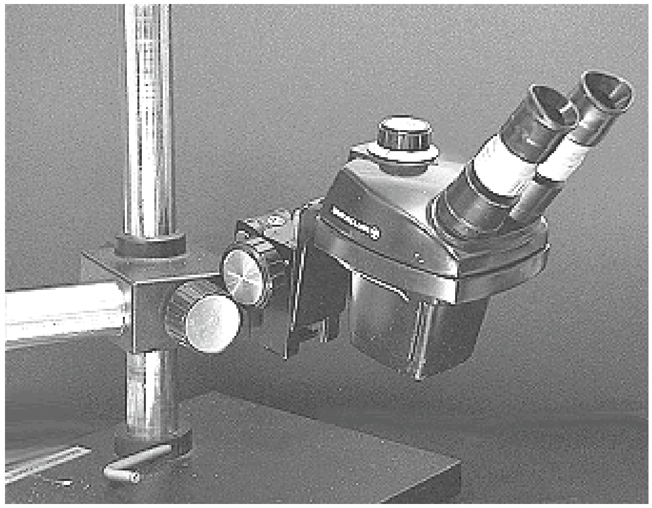


Figure 2-14.—Stereoscopic zoom microscope.

TOOL CHEST

The tool chest (not shown), provides storage space for the electronic repair hand tools, dental tools, abrasive wheels, solder and solder wicks, eyelets, abrasive disks, ball mills, various burrs, and other consumables used with the repair procedures. The chest is portable, lockable, and has variously sized drawers for convenience.

REPLACEMENT PARTS

Replacement parts are provided with the 2M repair set to ensure the technician has the capability to maintain the equipment properly. Actual preventive and corrective maintenance procedures, as well as data on additional spare parts and ordering information, are found in the technical manual for the 2M repair set equipment.

REPAIR STATION FACILITIES

To be effective, 2M electronic component repair must be performed under proper environmental conditions. Repair facility requirements, whether afloat or ashore, include adequate lighting, ventilation, noise considerations, work surface area, ESD (electrostatic discharge) protection, and adequate power availability. The recommended environmental conditions are discussed below. With the exception of requirements imposed by the Naval Environmental Health Center and other authorities for ship and shore work conditions, each activity tailors the requirements to meet local needs.

LIGHTING

The recommended lighting for a work surface is 100 footcandles from a direct lighting source. Light-colored overheads and bulkheads and off-white or pastel workbench tops are used to complement the lighting provided.

VENTILATION

Fumes from burning flux, coating materials, grinding dust, and cleaning solvents require adequate ventilation. The use of toxic, flammable substances, solvents, and coating compounds requires a duct system that vents gasses and vapors. This type of system must be used to prevent contamination often found in closed ventilation systems. This need is particularly important aboard ship. Vented hoods, ducts, or installations that are vented outside generally meet the minimum standards set by the Naval Environmental Health Center.

NOISE CONSIDERATIONS

Noise in the work area during normal work periods must be no greater than the acceptable level approved for each activity involved. Because the work is tedious and tiring, noise levels should be as low as possible. Ear protectors are required to be worn when a noise level exceeds 85 dB. Ear protectors should also be worn anytime the technician feels distracted by, or uncomfortable with, the noise level.

WORK SURFACE AREA

Work stations should have a minimum work surface of at least 60 inches wide and 30 inches deep. Standard Navy desks are excellent for this purpose. Standard shipboard workbenches are acceptable; however, off-white or pastel-colored heat-resistant tops should be installed on the workbenches. Chairs should be the type with backs and without arms. They should be comfortably padded and of the proper height to match the work surface height. Drawers or other suitable tool storage areas are usually provided.

ELECTROSTATIC DISCHARGE SENSITIVE DEVICE (ESDS) CAPABILITY

A 2M work station should be capable of becoming a static-free work station. This is specified in the Department of Defense Standard, Electrostatic DISCHARGE Control Program for Protection of Electrical and Electronic Parts, Assemblies, and Equipment. ESD will be discussed in greater detail in topic 3.

POWER REQUIREMENTS

No special power source or equipment mounting is required. The 2M repair equipment operates on 115-volt, 60-hertz power. A 15-ampere circuit is sufficient and six individual power receptacles should be available.

HIGH-RELIABILITY SOLDERING

The most common types of miniature and microminiature repair involve the removal and replacement of circuit components. The key to these repairs is a firm knowledge of solder and high-reliability soldering techniques.

Solder is a metal alloy used to join two or more metals with a metallic bond. The bonding occurs when molten solder dissolves a small amount of the metals and then cools to form a solid connection. The solder most commonly used in electronic assemblies is an alloy of tin and lead. Tin-lead alloys are identified by their percentage in the solder; the tin content is given first. Solder marked 60/40 is an alloy of 60 percent tin and 40 percent lead. The two most common alloys used in electronics are 60/40 and 63/37.

The melting temperature of tin-lead solder varies depending on the percentage of each metal. Lead melts at a temperature of 621 degrees Fahrenheit, and tin melts at 450 degrees Fahrenheit. Combinations of the two metals melt into a liquid at different temperatures. The 63/37 combination melts into a liquid at 361 degrees Fahrenheit. At this temperature, the alloy changes from a solid directly to a liquid with no plastic or semiliquid state. An alloy with such a sharp changing point is called a EUTECTIC ALLOY.

As the percentages of tin and lead are varied, the melting temperature increases. Alloy of 60/40 melts at 370 degrees Fahrenheit, and alloy of 70/30 melts at approximately 380 degrees Fahrenheit. Alloys, other than eutectic, go through a plastic or semiliquid state in their heating and cooling stages. Solder joints that are disturbed (moved) during the plastic state will result in damaged connections. For this reason, 63/37 solder is the best alloy for electronic work. Solder with 60/40 alloy is also acceptable, but it goes into a plastic state between 361 and 370 degrees Fahrenheit. When soldering joints with 60/40 alloy, you must exercise extreme care to prevent movement of the component during cooling.

USE OF FLUX IN SOLDER BONDING

Reliable solder connections can only be accomplished with clean surfaces. Using solvents and abrasives to clean the surfaces to be soldered is essential if you are to achieve good solder connections. In almost all cases, however, this cleaning process is insufficient because oxides form rapidly on heated metal surfaces. The rapid formation of oxides creates a nonmetallic film that prevents solder from contacting the metal. Good metal-to-metal contact must be obtained before good soldering joints may take place. Flux removes these surface oxides from metals to be soldered and keeps them removed during the soldering operation. Flux chemically breaks down surface oxides and causes the oxide film to loosen and break free from the metals being soldered.

Soldering fluxes are divided into three classifications or groups: CHLORIDE FLUX (commonly called ACID), ORGANIC FLUX, and ROSIN FLUX. Each flux has characteristics specific to its own group. Chloride fluxes are the most active of the three groups. They are effective on all common metals except aluminum and magnesium. Chloride fluxes, however, are NOT suitable for electronic soldering because they are highly corrosive, electrically conductive, and are difficult to remove from the soldered joint.

Organic fluxes are nearly as active as chloride fluxes, yet are less corrosive and easier to remove than chloride fluxes. Also, these fluxes are NOT satisfactory for electronic soldering because they must be removed completely to prevent corrosion.

Rosin fluxes ARE ideally suited to electronic soldering because of their molecular structure. The most common flux used in electronic soldering is a solution of pure rosin dissolved in suitable solvent. This solution works well with the tin- or solder-dipped metals commonly used for wires, lugs, and connectors. While inert at normal temperatures, rosin fluxes break down and become highly active at soldering temperatures. In addition, rosin is nonconductive.

Most electronic solder, in wire form, is made with one or more cores of rosin flux. When the joint or connection is heated and the wire solder is applied to the joint (not the iron), the flux flows onto the surface of the joint and removes the oxide. This process aids the wetting action of the solder. With enough heat the solder flows and replaces the flux. Insufficient heat results in a poor connection because the solder does not replace the flux.

Q10. Stereoscopic-zoom microscopes and precision drill presses are normally associated with what type of repair station?

Q11. Solder used in electronic repair is normally an alloy of what two elements?

Q12. In soldering, what alloy changes directly from a solid state to a liquid state?

Q13. Flux aids in soldering by removing what from surfaces to be soldered?

Q14. What type(s) of flux should never be used on electronic equipment?

SUMMARY

This topic has presented information on the Miniature and Microminiature 2M Repair Program and high-reliability soldering. The information that follows summarizes the important points of this topic.

The **MINIATURE/MICROMINIATURE (2M) REPAIR PROGRAM** provides training, tools and equipment, and certification for 2M repair personnel.

CERTIFICATION of technicians ensures the capability of high-quality, high-reliability repairs.

The three **SM&R** codes for maintenance of electronic devices are**: DEPOT (D)**, **INTERMEDIATE (I)**, and **ORGANIZATIONAL (O)**.

SM&R CODE D MAINTENANCE is characterized by extensive facilities and highly trained personnel. Code D activities are capable of the most complex type repairs.

CODE I activities provide direct support for user activities. This includes calibration, repair, and emergency manufacture of nonavailable parts.

CODE O maintenance is the responsibility of the user activity. It includes preventive maintenance and minor repairs.

ON-LINE TEST EQUIPMENT continuously monitors system performance and isolates faults to removable assemblies.

OFF-LINE TEST EQUIPMENT evaluates removable assemblies outside of the equipment and isolates faults to the component level.

FAULT ISOLATION USING GENERAL-PURPOSE ELECTRONIC TEST EQUIPMENT (GPETE) should only be attempted by experienced technicians.

2M REPAIR STATIONS are equipped according to the level of repairs to be accomplished.

ALLOYS, such as solder, which change directly from a solid state to a liquid are called eutectic alloys.

SOLDER with a tin/lead ratio of 63/37 is preferred for electronic work. A ratio of 60/40 is also acceptable.

ROSIN or **RESIN FLUXES** are the only fluxes to be used in electronic work.

ANSWERS TO QUESTIONS Q1. THROUGH Q14.

A1. Chief of Naval Operations (CNO).

A2. Naval Sea Systems Command (NAVSEASYSCOM) and Naval Air Systems Command (NAVAIRSYSCOM).

A3. Microminiature component repair.

A4. Microminiature repair technician.

A5. Depot, Intermediate, and Organizational.

A6. Organizational.

A7. On-line, off-line, and General Purpose Electronic Test Equipment (GPETE).

A8. On-line.

A9. Off-line.

A10. Microminiature repair station.

A11. Tin and lead.

A12. Eutectic.

A13. Oxides.

A14. Chloride or (acid) and organic.

CHAPTER 3

MINIATURE AND MICROMINIATURE REPAIR PROCEDURES

LEARNING OBJECTIVES

Upon completion of this topic, the student will be able to:

1. Explain the purpose of conformal coatings and the methods used for removal and replacement of these coatings.
2. Explain the methods and practices for the removal and replacement of discrete components on printed circuit boards.
3. Identify types of damage to printed circuit boards, and describe the repair procedures for each type of repair.
4. Describe the removal and replacement of the dual-in-line integrated circuit.
5. Describe the removal and replacement of the TO-5 integrated circuit.
6. Describe the removal and replacement of the flat-pack integrated circuit.
7. Describe the types of damage to which many microelectronic components are susceptible and methods of preventing damage.
8. Explain safety precautions as they relate to 2M repair.

INTRODUCTION

As you progress in your training as a technician, you will find that the skill and knowledge levels required to maintain electronic systems become more demanding. The increased use of miniature and microminiature electronic circuits, circuit complexity, and new manufacturing techniques will make your job more challenging. To maintain and repair equipment effectively, you will have to duplicate with limited facilities what was accomplished in the factory with extensive facilities. Printed circuit boards that were manufactured completely by machine will have to be repaired by hand.

To meet the needs for repairing the full range of electronic equipment, you must be properly trained. You must be capable of performing high-quality, reliable repairs to the latest circuitry.

MINIATURE AND MICROMINIATURE ELECTRONIC REPAIR PROCEDURES

As mentioned at the beginning of topic 2, 2M repair personnel must undergo specialized training. They are trained for a particular level of repair and must be certified at that level. Also, recertification is required to ensure the continued high-quality repair ability of these technicians.

CAUTION

THIS SECTION IS NOT, IN ANY WAY, TO BE USED BY YOU AS AUTHORIZATION TO ATTEMPT THESE TYPES OF REPAIRS WITHOUT OFFICIAL 2M CERTIFICATION.

In the following sections, you will study the general procedures used in the repair, removal, and replacement of specific types of electronic components. By studying these procedures, you will become familiar with some of the more common types of repair work. Before repair work can be performed on a miniature or microminiature assembly, the technician must consider the type of specialized coating that usually covers the assembly. These coatings are referred to as CONFORMAL COATINGS.

CONFORMAL COATINGS

Conformal coatings are protective material applied to electronic assemblies to prevent damage from corrosion, moisture, and stress. These coatings include epoxy, parylene, silicone, polyurethane, varnish, and lacquer. Coatings are applied in a liquid form; when dry, they exhibit characteristics that improve reliability. These characteristics are:

- *Heat conductivity* to carry heat away from components
- *Hardness and strength* to support and protect components
- *Low moisture* absorption
- *Electrical* insulation

Conformal Coating Removal

Because of the characteristics that conformal coatings exhibit, they must be removed before any work can be done on printed circuit boards. The coating must be removed from all lead and pad/eyelet areas of the component. It should also be removed to or below the widest point of the component body. *Complete* removal of the coating from the board is not done.

Methods of coating removal are thermal, mechanical, and chemical. The method of removal depends on the type of coating used. Table 3-1 shows suggested methods of removal of some types. Note that most of the methods are variations of mechanical removal.

Table 3-1.—Conformal Coating Removal Techniques

TYPES CONFORMAL COATING (LISTED IN DESCENDING ORDER OF HARDNESS)	SOLVENT	ROTARY BRISTLE BRUSH	GRIT-IMPREGNATED RUBBER WHEELS AND DISCS	HOT AIR JET AND ORANGEWOOD STICK	THERMAL PARTING TOOI	BALL MILLS, ETC.	CUTTING AND PEELING
PARYLENE			2		1 (F)	3	
EPOXY		3 (C)	4 (C)	1 *(C)	2 *	5 (B&D)	
ACRYLIC LACQUER	1 *(E)	3	4 (C)	2			
POLYURETHANE	5 (A&C)	3 (C)	4 (C)	1 *	2 *		
VARNISH	1 (A&C)	3	4	2 *			
RTV		2 (C)					1 *

* DENOTED METHOD THAT POSITIVELY IDENTIFIES TYPE OF CONFORMAL COATING.
A SPECIFIC TYPES OF COATING COMPOSITIONS ONLY.
B FOR THICK COATINGS ONLY (.025 AND THICKER).
C FOR THIN COATINGS ONLY.
D DO NOT ATTEMPT TO GRIND TO BOARD SURFACE WITH THIS METHOD.
E ORGANIC SOLVENTS ARE BEST.
F USE MODIFIED DRILL BIT FOR TIP (SHAPED TO BEVELED EDGE). CAUTION: USE WITH CARE, APPROXIMATELY 500° NEEDED AT TIP.

NOTE: THE PREFERRED ORDER FOR APPLYING INDIVIDUAL REMOVAL TECHNIQUES TO SPECIFIC COATINGS IS NUMERICALLY INDICATED. THESE REMOVAL TECHNIQUES ARE LISTED IN ASCENDING ORDER OF THEIR DAMAGEABILITY TO THE MODULE UNDER REPAIR. ANY OF THE METHODS LISTED MAY CAUSE DAMAGE IF NOT USED WITH CARE. ALWAYS TRY THE METHOD WHICH CAUSES THE LEAST AMOUNT OF DAMAGE FIRST. CONSIDER THE POSSIBILITY OF HEAT OR VIBRATION SENSITIVITY OF COMPONENTS. (VIBRATION ESPECIALLY WILL AFFECT ALL AREAS OF THE WORKPIECE.)

The coating material can best be identified through proper documentation; for example, technical manuals and engineering drawings. If this information is not available, the experienced technician can usually determine the type of material by testing the hardness, transparency, thickness, and solvent solubility of the coating. The thermal (heat) properties may also be tested to determine the ease of removal of the coating by heat. The methods of removal discussed here describe the basic concept, but not the step-by-step "how to" procedures.

THERMAL REMOVAL.—Thermal removal consists of using controlled heat through specially shaped tips attached to a handpiece. Soldering irons should never be used for coating removal because the high temperatures will cause the coatings to char, possibly damaging the board materials. Modified tips or cutting blades heated by soldering irons also are not used; they may not have proper heat capacity or allow the hand control necessary for effective removal. Also, the thin plating of the circuit may be damaged by scraping.

The thermal parting tool, used with the variable power supply, has interchangeable tips, as shown in figure 3-1, that allow for efficient coating removal. These thin, blade-like instruments act as heat generators and will maintain the heat levels necessary to accomplish the work. Tips can be changed easily to suit the configuration of the workpiece. These tips cool quickly after removal of power because their small thermal mass and special alloy material easily give up residual heat.

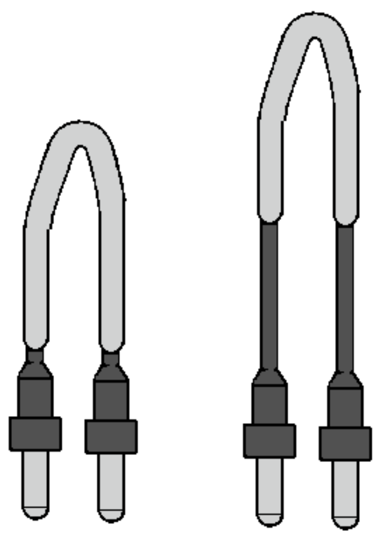


Figure 3-1.—Thermal parting tips.

The softening or breakdown point of different coatings vary, which is a concern when you are using this method. Ideally, the softening point is below the solder melting temperature. However, when the softening point is equal to or above the solder melting point, you must take care in applying heat at the solder joint or in component areas. The work must be performed rapidly to limit the heating of the area involved and to prevent damage to the board and other components.

HOT-AIR JET REMOVAL.—In principle, the hot-air jet method of coating removal uses controlled, temperature-regulated air to soften or break down the coating, as shown in figure 3-2. By controlling the temperature, flow rate, and shape of the jet, you may remove coatings from almost any workpiece configuration without causing any damage. When you use the hot-air jet, you do not allow it to contact the workpiece surface physically. Delicate work, handled in this manner, permits you to observe the removal process.

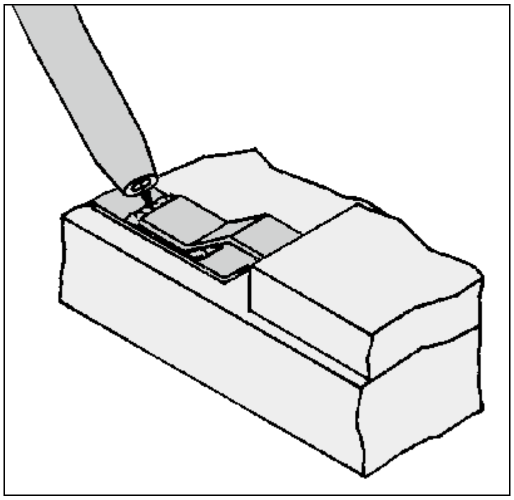


Figure 3-2.—Hot air jet conformal coating removal.

POWER-TOOL REMOVAL DESCRIPTION.—Power-tool removal is the use of abrasive grinding or cutting to remove coatings mechanically. Abrasive grinding/rubbing techniques are effective on thin coatings (less than 0.025 inch) while abrasive cutting methods are effective on coatings greater than 0.025 inch. This method permits consistent and precise removal of coatings without mechanical damage or dangerous heating to electronic components. A variable-speed mechanical drive handpiece permits fingertip-control and proper speed and torque to ease the handling of gum-type coatings. A variety of rotary abrasive materials and cutting tools is required for removal of the various coating types. These specially designed tools include BALL MILLS, BURRS, and ROTARY BRUSHES.

The ball mill design places the most efficient cutting area on the side of the ball, rather than at the end. Different mill sizes are used to enter small areas where thick coatings need to be removed (ROUTED). Rubberized abrasives of the proper grade and grit are ideally suited for removing thin, hard coatings from flat surfaces; soft coatings adhere to and coat the abrasive causing it to become ineffective. Rotary bristle brushes work better than rubberized abrasives on contoured or irregular surfaces, such as soldered connections, because the bristles conform to surface irregularities. Ball mill routing and abrasion removal are shown in figure 3-3.

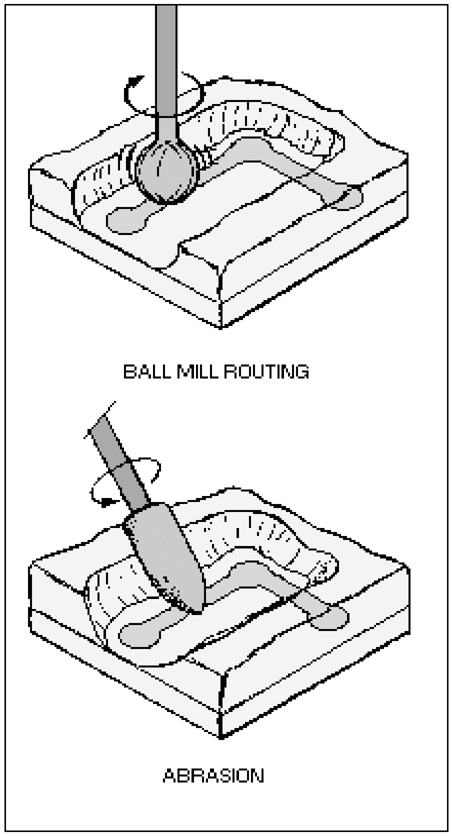


Figure 3-3.—Rotary tool conformal coating removal.

CUT AND PEEL.—Silicone coatings (also referred to as RTV) can easily be removed by cutting and peeling. As with all mechanical removal methods, care must be taken to prevent damage to either components or boards.

CHEMICAL REMOVAL.—Chemical removal uses solvents to break down the coatings. General application is not recommended as the solvent may cause damage to the boards by dissolving the adhesive materials that bond the circuits to the boards. These solvents may also dissolve the POTTING COMPOUNDS (insulating material that completely seals a component or assembly) used on other parts or assemblies. Only thin acrylic coatings (less than 0.025 inch) are readily removable by solvents. Mild solvents, such as ISOPROPYL ALCOHOL, XYLENE, or TRICHLOROETHANE, may be used to remove soluble coatings on a spot basis.

Evaluations show that many tool and technique combinations have proven to be reliable and effective in coating removal; no single method is the best in all situations. When the technician is determining the best method of coating removal to use, the first consideration is the effect that it will have on the equipment.

Conformal Coating Replacement

Once the required repairs have been completed, the conformal coating must be replaced. To ensure the same protective characteristics, you should use the same type of replacement coating as that removed.

Conformal coating application techniques vary widely. These techniques depend on material type, required thickness of application, and the effect of environmental conditions on curing. These procedures cannot be effectively discussed here.

Q1. What material is applied to electronic assemblies to prevent damage from corrosion, moisture, and stress?

Q2. What three methods are used to remove protective material?

Q3. What chemicals are used to remove protective material?

Q4. Abrasion, cutting, and peeling are examples of what type of protective material removal?

Q5. Why should the coating material be replaced once the required repair has been completed?

REMOVAL AND REPLACEMENT OF DISCRETE COMPONENTS

To perform the required repair properly, the 2M technician must be knowledgeable of the techniques used by manufacturers in the production of electronic assemblies. The techniques, materials, and types of components determine the repair procedures used.

Interconnections and Assemblies

Assemblies may range from simple, single-sided boards with standard-sized components to double-sided or multilayered boards with miniature and microminiature components. The variations in component lead termination and mounting techniques used by manufacturers present the technician with a complex task. For example, the 2M technician is concerned about the type of solder joints on the module. To determine the solder joint type, the technician must consider the board circuitry, hole reinforcement, and lead termination style.

Recall the discussion from topic 1 on printed circuit board construction and the types of interconnections used. Single-sided and some double-sided boards have UNSUPPORTED HOLES where component leads are soldered to the pad. The clearance-hole method is also an interconnection with no hole support. SUPPORTED HOLES are those that have metallic reinforcement along the hole walls.

In addition to the plated-through hole you studied earlier, EYELETS, shown in figure 3-4, view (A), view (B), and view (C), are also used in both manufacturing and repair. These hole-reinforcing devices are usually made of pure copper, but are often plated with gold, tin, or a tin-lead alloy. The copper-based eyelet is pliable; when set, it reduces the possibility of circuit board damage. Eyelets may be inserted into single-sided or double-sided boards and are of three different types - ROLL SET, FUNNEL SET, and FLAT SET. All three are types referred to as INTERFACIAL CONNECTIONS. Interfacial connections identify the procedure of connecting circuitry on one side of a board with the circuitry on the other side.

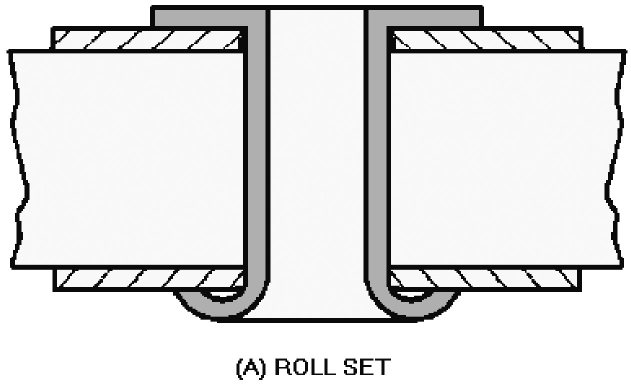


Figure 3-4A.—Eyelets (interfacial connections). ROLL SET

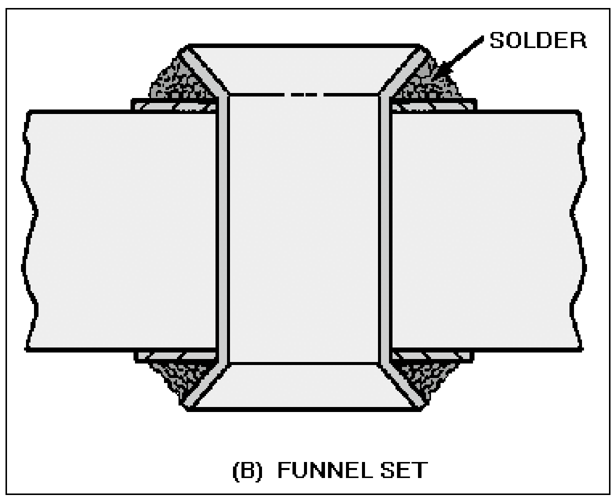


Figure 3-4B.—Eyelets (interfacial connections). FUNNEL SET

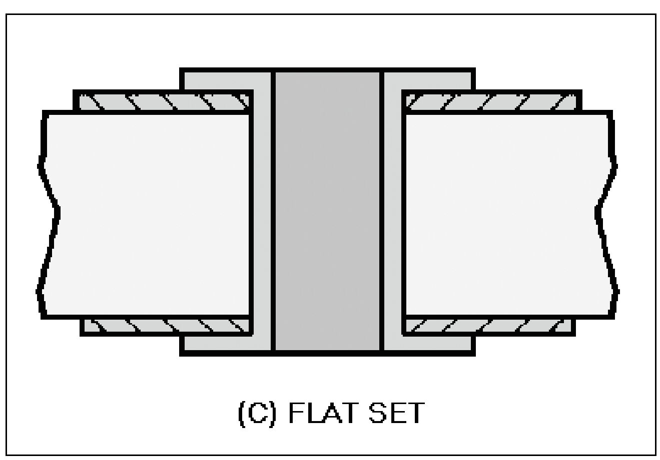


Figure 3-4C.—Eyelets (interfacial connections). FLAT SET

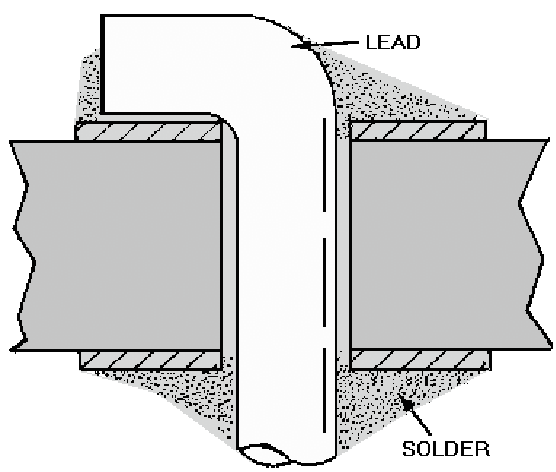
As you can see, the flat-set eyelet actually provides reinforcement for the pads on both sides of the circuit board and reinforces the hole itself. The design of the roll-set eyelet (which may trap gasses, flux, or other contaminants, and obscures view of the finished solder flow) is not acceptable as a repair technique. The funnel-set eyelet does not provide as much pad reinforcement as the other types. However, it provides better "outgassing" of flux, moisture, or solvents from the space between the eyelet and the hole wall. It also provides a better view of the finished solder connection than the roll-set eyelet.

Lead Terminations

The finished circuit board consists of conductive paths, pads, and drilled holes with components and/or wires assembled directly to it. Leads and wires may terminate in three ways: (1) through the hole in the board, (2) above the surface of the board, or (3) on the surface of the board.

THROUGH-HOLE TERMINATION.—This style provides extra support for the circuit pads, the hole, and the lead by a continuous solder connection from one side of the circuit board to the other. Three basic variations of through-hole termination are the CLINCHED LEAD (two types), STRAIGHT-THROUGH LEAD, and OFFSET PAD.

Clinched Lead.—The clinched-lead termination is usually used with unsupported holes, but is found with supported holes as well. Both clinched-lead types, FULLY CLINCHED and SEMICLINCHED (figure 3-5), provide component stability. Like the fully clinched lead, the semi-clinched lead also provides stability during assembly. However, this termination can be easily straightened to allow removal of the solder joint should rework or repair be required. Note that the fully clinched lead is bent 90 degrees while the semiclinched lead is bent 45 degrees.



(A) FULLY CLINCHED

Figure 3-5A.—Clinched leads. FULLY CLINCHED.

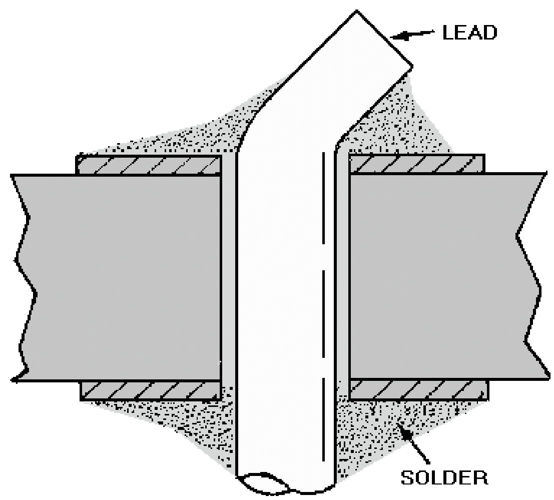


Figure 3-5B.—Clinched leads. SEMICLINCHED

Straight-Through Lead.—Straight-through terminations (figure 3-6) are used by manufacturers when the termination stability is not a prime consideration. This termination type may also be used with unsupported holes. The through-hole termination provides a better, solder-joint contact area and more solder support; the solder runs from the component side to the conductor. The straight-through termination is the easiest to remove and rework.

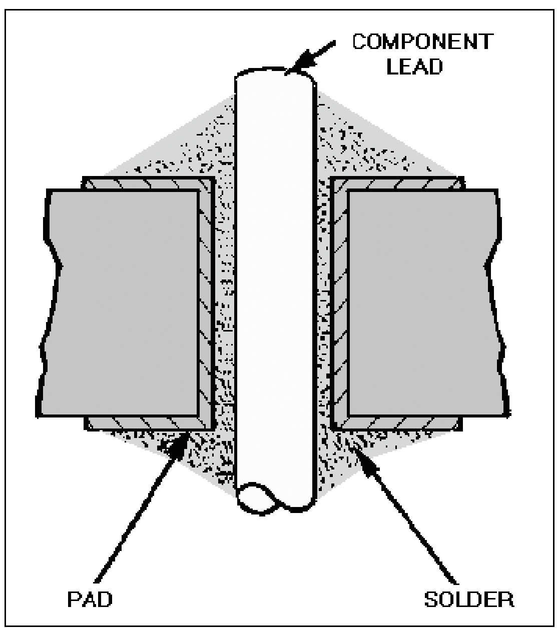


Figure 3-6.—Straight-through termination.

The Offset-Pad Termination.—This termination, shown in view (A) of figure 3-7, is a variation of clinch-lead termination. The pad is set off from the centerline of the hole. The lead clinch is also offset from the hole centerline so that it may contact the pad [view (B)].

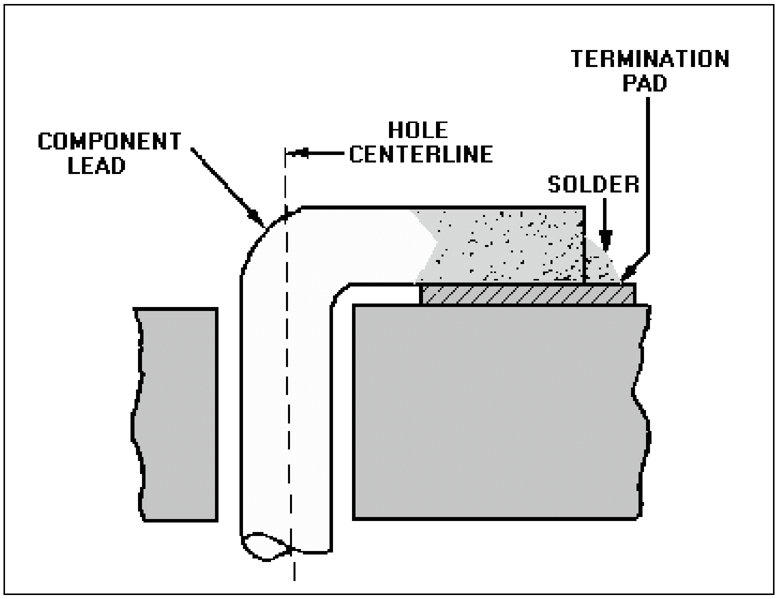


Figure 3-7A.—Offset pad termination SIDE VIEW.

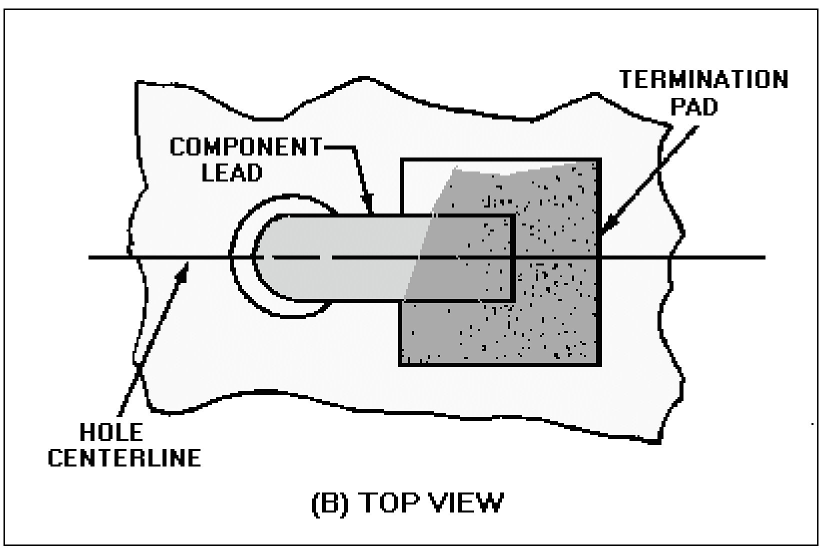


Figure 3-7B.—Offset pad termination TOP VIEW

ABOVE-THE-BOARD TERMINATION.—Above-the-board termination is accomplished through the use of terminals or posts. Terminals are used for a variety of reasons. The type of terminal depends on its use. Although many configurations are used, all terminals fall into one of the five categories covered in this section [figure 3-8, views (A) through (E)].

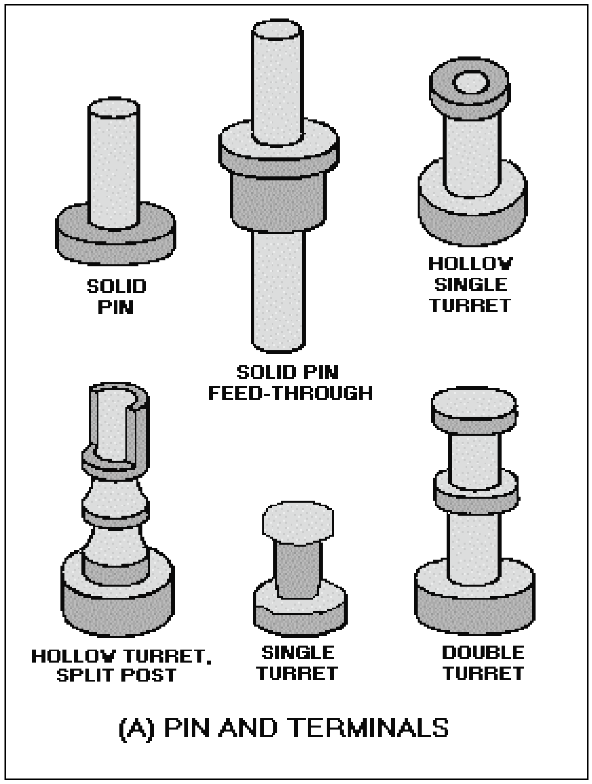


Figure 3-8A.—Terminals. PIN AND TERMINALS.

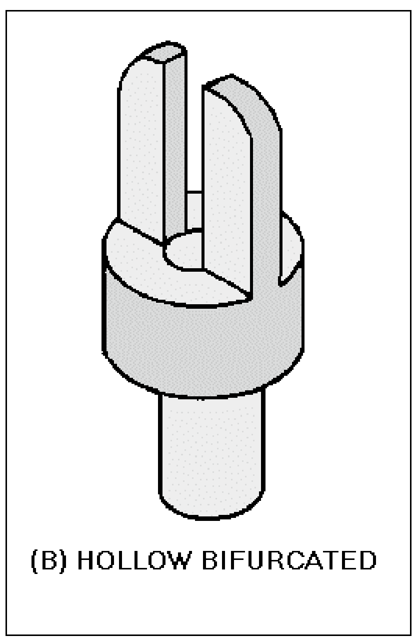


Figure 3-8B.—Terminals. HOLLOW.

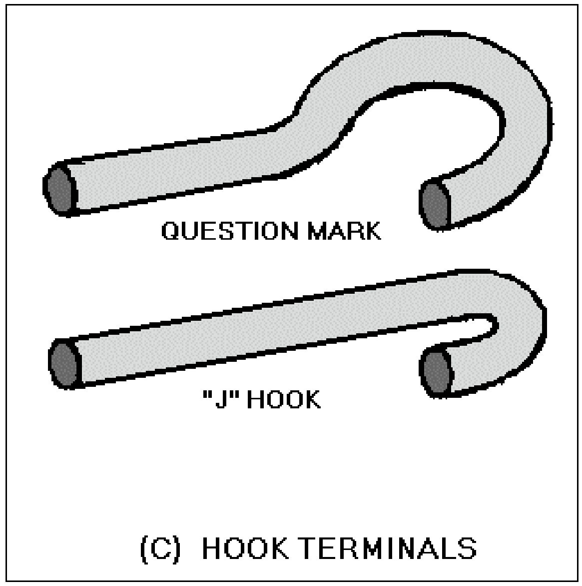


Figure 3-8C.—Terminals. HOOK TERMINALS.

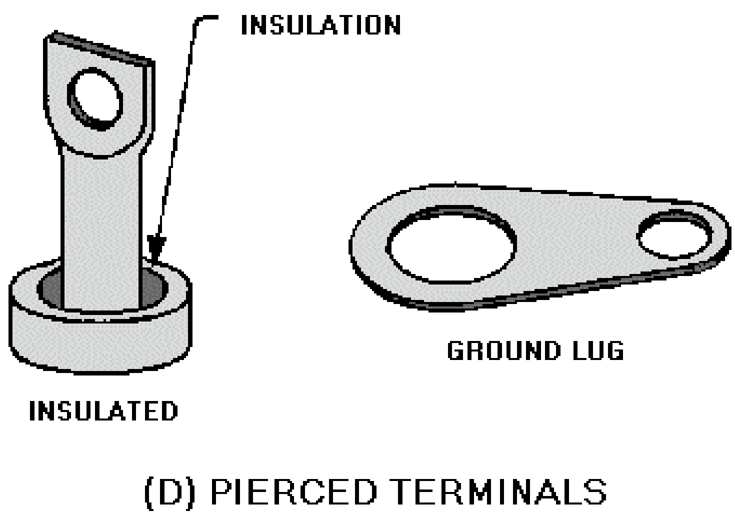


Figure 3-8D.—Terminals. PIERCED TERMINALS.

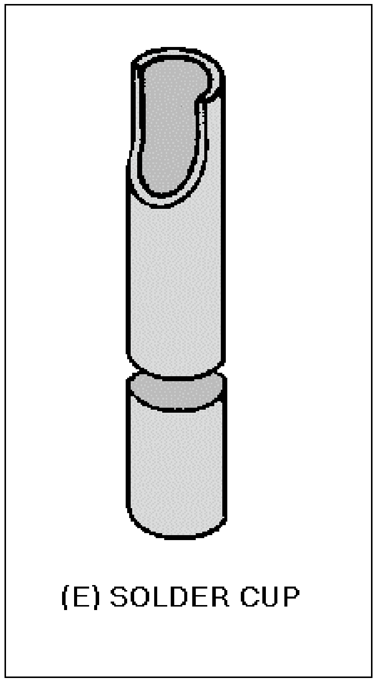


Figure 3-8E.—Terminals. SOLDER CUP

- PIN TERMINALS AND TURRET TERMINALS [view (A)] are single-post terminals, either insulated or uninsulated, solid or hollow, stud or feed-through. Stud terminals protrude from one side of a board; feed-throughs protrude from both sides.
- BIFURCATED OR FORK TERMINALS [view (B)] are solid or hollow double-post terminals.
- HOOK TERMINALS [view (C)] are made of cylindrical stock formed in the shape of a hook or question mark.
- PERFORATED OR PIERCED TERMINALS [view (D)] describe a class of terminals that uses a hole pierced in flat metal for termination (e.g., terminal lugs).
- SOLDER CUP TERMINALS [view (E)] are a common type found on connectors.

Turret and bifurcated terminals are used for interfacial connections on printed circuit boards, terminal points for point-to-point wiring, mounting components, and as tie points for interconnecting wiring. Hook terminals are used to provide connection points on sealed devices and terminal boards.

Terminals, used for wire or component lead terminations, are normally made of brass with a solderable coating. Uninsulated terminals may be installed on an insulating substrate to form a terminal board. They may also be added to a printed circuit board or installed on a metal chassis. Insulated terminals are installed on a metal chassis.

ON-THE-BOARD TERMINATION.—On-the-board termination (figure 3-9) is also called LAP FLOW termination. In a lap flow solder termination, the component lead does not pass through the circuit board. This form of planar mounting may be used with both round and flat leads.

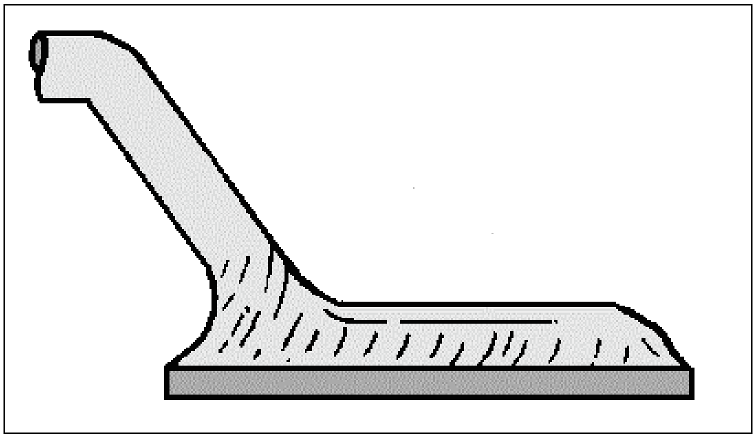


Figure 3-9.—On-the-board termination.

Q6. What term is used to identify the procedure of connecting one side of a circuit board with the other?

Q7. Name two types of through-hole termination.

Q8. Turret, bifurcated, and hook terminals are used for what type of termination?

Q9. When a lead is soldered to a pad without passing through the board, it is known as what type of termination?

Component Desoldering

Most of the damage in printed circuit board repair occurs during disassembly or component removal. More specifically, much of this damage occurs during the desoldering process. To remove components for repair or replacement, the technician must first determine the type of joint that is used to connect the component to the board. The technician may then determine the most effective method for desoldering these connections.

Three generally accepted methods of solder connection removal involve the use of SOLDER WICK, a MANUALLY CONTROLLED VACUUM PLUNGER, or a motorized solder extractor using CONTINUOUS VACUUM AND/OR PRESSURE. Of all the extraction methods currently in use, continuous vacuum is the most versatile and reliable. Desoldering becomes a routine operation and the quantity and quality of desoldering work increases with the use of this technique.

SOLDER WICKING.—In this technique, finely stranded copper wire or braiding (wick) is saturated with liquid flux. Most commercial wick is impregnated with flux; the liquid flux adds to the effectiveness of the heat transfer and should be used whenever possible. The wick is then applied to a solder joint between the solder and a heated soldering iron tip, as shown in figure 3-10. The combination of heat, molten solder, and air spaced in the wick creates a capillary action and causes the solder to be drawn into the wick.

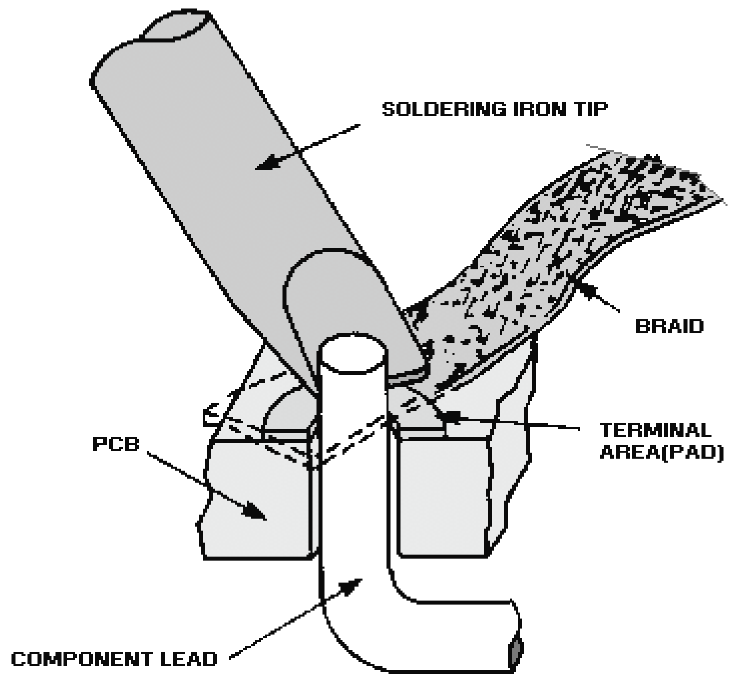


Figure 3-10.—Solder wicking.

This method should be used to remove surface joints only, such as those found on single-sided and double-sided boards without plated-through holes or eyelets. It can also remove excessive solder from flat surfaces and terminals. The reason is that the capillary action of the wicking is not strong enough to overcome the surface tension of the molten solder or the capillary action of the hole.

MANUALLY CONTROLLED VACUUM PLUNGER.—The second method of removing solder involves a manually controlled and operated, one-shot vacuum source. This vacuum source uses a plunger mechanism with a heat resistant orifice. The vacuum is applied through this orifice. Figure 3-11 shows the latest approved, manual-type desoldering tool. This technique involves melting the solder joint and inserting the solder-extractor tip into the molten solder over the soldering iron tip. The plunger is then released, creating a short pulse of vacuum to remove the molten solder. Although this method offers a positive vacuum, rather than the capillary force of the wicking method, it still has limited application. This method will not remove 100 percent of the solder and may cause circuit pad lifting because of the extremely high vacuum generated and the jarring caused by the plunger action.

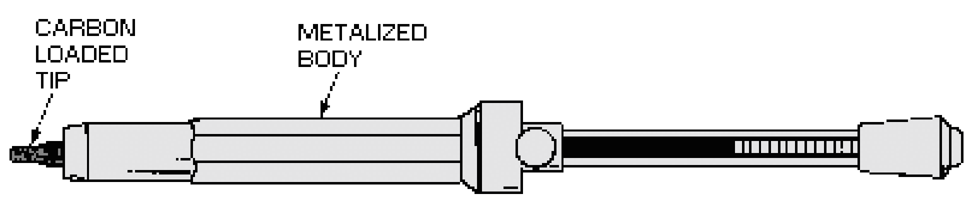


Figure 3-11.—Manual desoldering tool.

Because 100 percent of the solder cannot be removed, the extraction method is not usually successful with the plated-through solder joint. The component lead in a plated-through hole joint usually rests against the side wall of the hole. Even though most of the molten solder is removed by a vacuum, the small amount of solder left between the lead and side walls causes a SWEAT JOINT to form. A sweat joint is a paper-thin solder joint formed by a minute amount of solder remaining on the conductor lead surfaces.

MOTORIZED VACUUM/PRESSURE METHOD.—The most effective method for solder joint removal is motorized vacuum extraction. The solder extractor unit, described in topic 2, is used for this type of extraction. This method provides controlled combinations of heat and pressure or vacuum for solder removal. The motorized vacuum is controlled by a foot switch and differs from the manual vacuum in that it provides a continuous vacuum. The solder extraction device is a coaxial, in-line instrument similar to a small soldering iron. The device consists of a hollow-tipped heating element, transfer tube, and collecting chamber (in the handle) that collects and solidifies the waste solder. This unit is easily maneuvered, fully controllable, and provides three modes of operation (figure 3-12): (1) heat and vacuum (2) heat and pressure, and (3) hot-air jet. Some power source models provide variable control for pressure and vacuum levels as well as temperature control for the heated tubular tip. The extraction tip and heat source are combined in one tool. Continuous vacuum allows solder removal with a single heat application. Since the slim heating element allows access to confined areas, the technician is protected from contact with the hot, glass, solder-trap chamber. Continuous vacuum extraction is the only consistent method for overcoming the resweat problem for either dual or multilead devices terminating in through-hole solder joints.

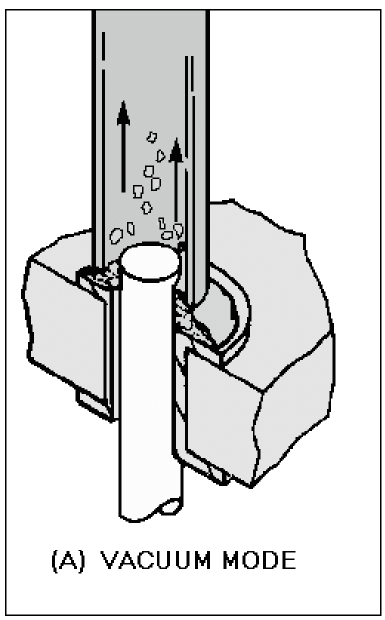


Figure 3-12A.—Motorized vacuum/pressure solder removal. VACUUM MODE.

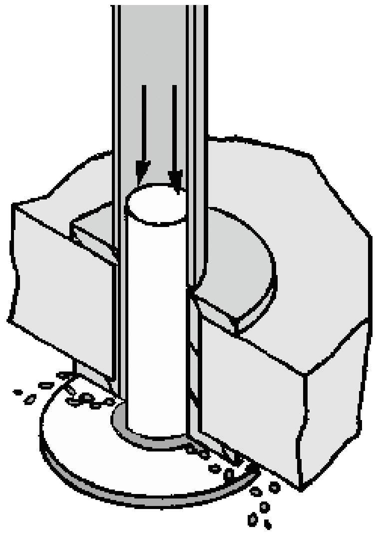


Figure 3-12B.—Motorized vacuum/pressure solder removal. PRESSURE MODE.

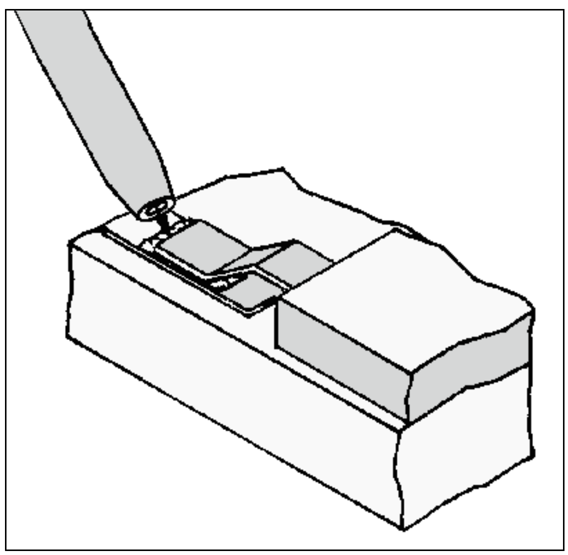


Figure 3-12C.—Motorized vacuum/pressure solder removal. HOT AIR JET MODE.

Motorized Vacuum Method.—In the motorized vacuum method, the heated tip is applied to the solder joint. When melted solder is observed, the vacuum is activated by the technician causing the solder to be withdrawn from the joint and deposited into the chamber. If the lead is preclipped, it may also be drawn into a holding chamber. To prevent SWEATING (reforming a solder joint) to the side walls of the plated-through hole joint, the lead is "stirred" with the tip while applying the vacuum. This permits cool air to flow into and around the lead and side walls causing them to cool.

Motorized Pressure Method.—In the pressure method, the tip is used to apply heat to a pin for melting a sweat joint. The air pressure is forced through the hole to melt sweat joints without contacting the delicate pad. This method is seldom used because it is not effective in preventing sweating of the lead to the hole nor for cooling the workpiece.

Hot-Air Jet Method.—The hot-air jet method uses pressure-controlled, heated air to transfer heat to the solder joint without physical contact from a solder iron. This permits the reflow of delicate joints while minimizing mechanical damage.

When the solder is removed from the lead and pad area, the technician can observe the actual condition of the lead contact to the pad area and the amount of the remaining solder joint. From these observed conditions, the technician can then determine a method of removing the component and lead.

With straight-through terminations, the component and lead may be lifted gently from uncoated boards with pliers or tweezers. Working with clinched leads on uncoated boards requires that all sweat joints be removed and that the leads be unclinched before removal.

The techniques that have been described represent the successful methods of desoldering components. As mentioned at the beginning of this section, the 2M technician must decide which method is best suited for the type of solder joint. Two commonly used but unacceptable methods of solder removal are heat-and-shake and heat-and-pull methods.

In the heat-and-shake method, the solder joint is melted and then the molten solder is shaken from the connection. In some cases, the shaking action may include striking the assembly against a surface to shake the molten solder out of the joint. This method should NEVER be used because all the solder may not be removed and the solder may splatter over other areas of the board. In addition, striking the board against a surface can lead to broken boards, damaged components, and lifted pads or conductors.

The heat-and-pull method uses a soldering iron or gang-heater blocks to melt individual or multiple solder joints. The component leads are pulled when the solder is melted. This method has many shortcomings because of potential damage and should NOT be attempted. Heating blocks are patterned to suit specific configurations; but when used on multiple-lead connections, the joints may not be uniformly heated. Uneven heating results in plated-through hole damage, pad delamination, or blistering. Damage can also result when lead terminations are pulled through the board.

When desoldering is complete, the workpiece must undergo a careful physical inspection for damage to the circuit board and the remaining components. The technician should also check the board for scorching or charring caused by component failure. Sometimes MEASLING is present. Measling is the appearance of light-colored spots. It is caused by small areas of fiberglass strands that have been damaged by epoxy overcuring, heat, abrasion, or internal moisture. No cracks or breaks should be visible in the board material. None of the remaining components should be cracked, broken, or show signs of overheating. The solder joints should be of good quality and not covered by loose or splattered solder, which may cause shorts. The technician should examine the board for nicked, cracked, lifted, or delaminated conductors and lifted or delaminated pads.

Q10. When does most printed circuit board damage occur?

Q11. What procedure involves the use of finely braided copper wire to remove solder?

Q12. What is the most effective method of solder removal?

Q13. When, if at all, should the heat-and-shake or the heat-and-pull methods of solder removal be used?

INSTALLATION AND SOLDERING OF PRINTED CIRCUIT COMPONENTS

The 2M technician should restore the electronic assembly at least to the original manufacturer's standards. Parts should always be remounted or reassembled in the same position and with termination methods used by the original manufacturer. This approach ensures a continuation of the original reliability of the system.

High reliability connections require thoroughly cleaned surfaces, proper component lead formation and termination, and appropriate placement of components on the board. The following paragraphs describe the procedures for properly installing components on a board including the soldering of these components.

Termination Area Preparation

The termination areas on the board and the component leads are thoroughly cleaned to remove oxide, old solder, and other contaminants. Old or excess solder is removed by one of the desoldering techniques explained earlier in this topic. A fine abrasive, such as an oil-free typewriter eraser, is used to remove oxides. This is not necessary if the area has just been desoldered. All areas to be soldered are cleaned with a solvent and then dried with a lint-free tissue to remove cleaning residue.

Component Lead Preparation

Component leads are formed before installation. Both machine- and hand-forming methods are used to form the leads. Improper lead formation causes many repairs to be unacceptable. Damage to the SEALS (point where lead enters the body of the component) occurs easily during the forming process and results in component failure. Consequently, lead-forming procedures have been established. To control the lead-forming operation and ensure conformity and quality of repairs, the technician should ensure the following:

1. The component is centered between the holes, and component leads are formed with proper bend-radii and body seal-to-bend distance.
2. The possibility of straining component body seals during lead forming is eliminated.
3. Stress relief loops are formed without straining component seals while at the same time providing the desired lead-to-lead distances.
4. Leads are measured and formed for both horizontal and vertical component mounting.
5. Transistor leads are formed to suit standard hole spacing.

Lead-Forming Specifications

Component leads are formed to provide proper lead spacing.

- The minimum distance between the seal (where the lead enters the body of the component) and the start of the lead bend must be no less than twice the diameter of the lead, as shown in figure 3-13.

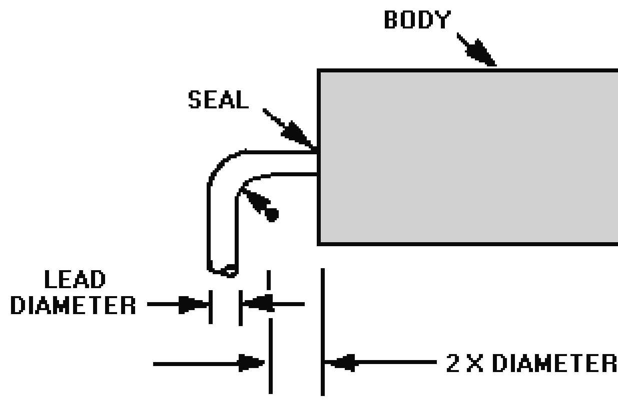


Figure 3-13.—Minimum distance lead bend to component body.

- Leads must be approximately 90 degrees from their major axis to ensure free movement in hole terminations, as shown in figure 3-14.

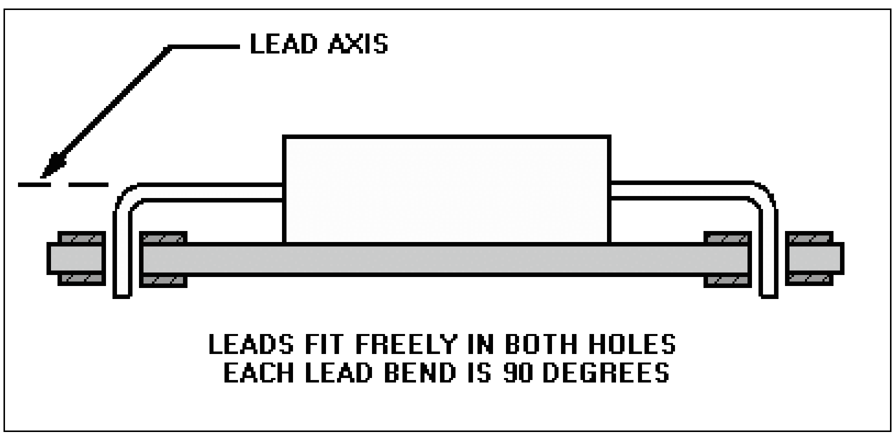


Figure 3-14.—Ideal lead formation.

- In lead-forming, the lead must not be damaged by nicking.

- Energy from the bending action must not be transmitted into the component body.

COMPONENT PLACEMENT.—Where possible, parts are remounted or reassembled as they were in the original manufacturing process. To aid recognition, manufacturers use a coding system of colored dots, bands, letters, numbers, and signs. Replacement components are mounted to make all identification markings readable without disturbing the component. When components are mounted like the original, all the identification markings are readable from a single point.

Component identification reads uniformly from left to right, top to bottom, unless polarity requirements determine otherwise, as shown in figure 3-15. To locate the top, position the board so the part number may be read like a page in a book. By definition, the top of the board is the edge above the part number.

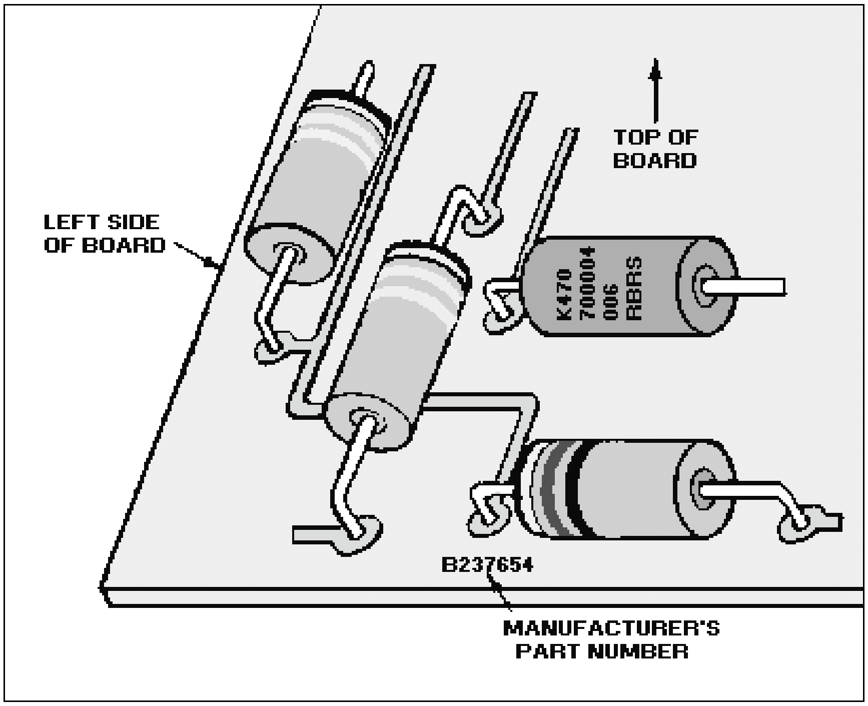


Figure 3-15.—Component arrangement.

When possible, component identification markings should be visible after installation. If you must choose between identification and electrical value markings, the priority of selection is as follows: (1) electrical value, (2) reliability level, and (3) part number.

Components are normally mounted parallel to and on the side opposite the printed circuitry and in contact with the board.

FORMATION OF PROPER LEAD TERMINATION.—After component leads are formed and inserted into the board, the proper lead length and termination are made before the lead is soldered. Generally, if the original manufacturer clinched (either full or semi) the component leads, the replacement part is reinstalled with clinched leads.

When clinching is required, leads on single- and double-sided boards are securely clinched in the direction of the printed wiring connected to the pad. Clinching is performed with tools that prevent damage to the pad or printed wiring. The lead is clinched in the direction of the conductor by bending the lead. The leads are clipped so that their minimum clinched length is equal to the radius of the pad. Under no circumstances does the clinched lead extend beyond the pad diameter. Natural springback away from the pad or printed wiring is acceptable. A gap between the lead end and the pad or printed wiring is acceptable when further clinching endangers the pad or printed wiring. These guidelines ensure uniform lead length.

Q14. To what standards should a technician restore electronic assemblies?

Q15. How is oxide removed from pads and component leads?

Q16. Leads are formed approximately how many degrees from their major axis?

Q17. When you replace components, identification marks must meet what requirements?

Q18. In what direction are component leads clinched on single- and double-sided boards?

Soldering of PCB Components

The fundamental principles of solder application must be understood and observed to ensure consistent and satisfactory results. As discussed in topic 2, the soldering process involves a metal-solvent action that joins two metals by dissolving a small amount of the metals at their point of contact.

SOLDERABILITY.—As the solder interacts with the base metals, a good metallurgical bond is obtained and metallic continuity is established. This continuity is good for electrical and heat conductivity as well as for strength. Solderability measures the ease with which molten solder wets the surfaces of the metals being joined. WETTING means the molten solder leaves a continuous permanent film on the metal surface. Wetting can only be done properly on a clean surface. All dirt and grease must be removed and no oxide layer must exist on the metal surface. Using abrasives and/or flux to remove these contaminants produces highly solderable surfaces.

HEAT SOURCE.—The soldering process requires sufficient heat to produce alloy- or metal-solvent action. Heat sources include CONDUCTIVE, RESISTIVE, CONVECTIVE, and RADIANT types. The type of heat source most commonly used is the conductive-type soldering iron. Delicate electronic assemblies require that the thermal characteristics of a soldering iron be carefully balanced and that the iron and tip be properly matched to the job. Successful soldering depends on the combination of the iron tip temperature, the capacity of the iron to sustain temperature, the time of iron contact with the joint, and the relative mass and heat transfer characteristics of the object being soldered.

SELECTION OF PROPER TIP.—The amount of heat and how it is controlled are critical factors to the soldering process. The tip of the soldering iron transfers heat from the iron to the work. The shape and size of the tip are mainly determined by the type of work to be performed. The tip size and the wattage of the element must be capable of rapidly heating the mass to the melting temperature of solder.

After the proper tip is selected and attached to the iron, the operator may control the heat by using the variable-voltage control. The most efficient soldering temperature is approximately 550 degrees Fahrenheit. Ideally, the joint should be brought to this temperature rapidly and held there for a short period of time. In most cases the soldering action should be completed within 2 or 3 seconds. When soldering a small-mass connection, control the heat by decreasing the size of the tip.

Before heat is applied to solder the joint, a thermal shunt is attached to sensitive component leads (diodes, transistors, and ICs). A thermal shunt is used to conduct heat away from the component. Because of its large heat content and high thermal conductivity, copper is usually used to make thermal shunts. Aluminum also has good conductivity but a smaller heat content; it is also used to conduct heat, especially if damage from the physical weight of the clamp is possible. Many types, shapes, and sizes of thermal shunts are available. The most commonly used is the clamp design; this is a spring clip (similar to an alligator clip) that easily fastens onto the part lead, as shown in figure 3-16.

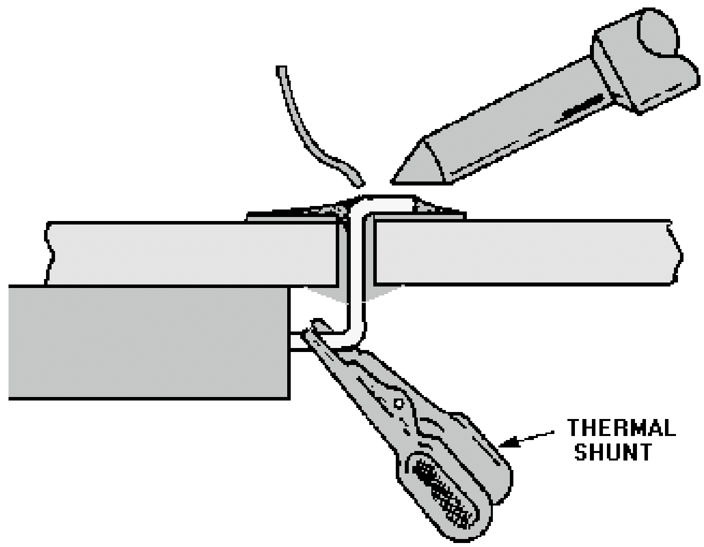


Figure 3-16.—Thermal shunt.

APPLICATION OF SOLDER AND SOLDERING IRON TIP.—Before solder is applied to the joint, the surface temperature of the parts being soldered is increased above the solder melting point. In general, the soldering iron is applied to the point of greatest mass at the connection. This increases the heat in the parts to be soldered. Solder is then applied to a clean, fluxed, and properly heated surface. When properly applied, the solder melts and flows without direct contact with the heat source and provides a smooth, even surface that feathers to a thin edge.

Molten solder forms between the tip and the joint, creating a heat bridge or thermal linkage. This heat bridge causes the tip to become part of the joint and allows rapid heat transfer. A solder (heat) bridge is formed by melting a small amount of solder at the junction of the tip and the mass being soldered as the iron is applied. After the tip makes contact with the lead and the pad and after the heat bridge is established, the solder is applied with a wiping motion to form the solder bond. The completed solder joint should be bright and shiny in appearance. It should have no cracks or pits, and the solder should cover the pad. Examples of preferred solder joints are shown in figure 3-17. They are referred to as full fillet joints.

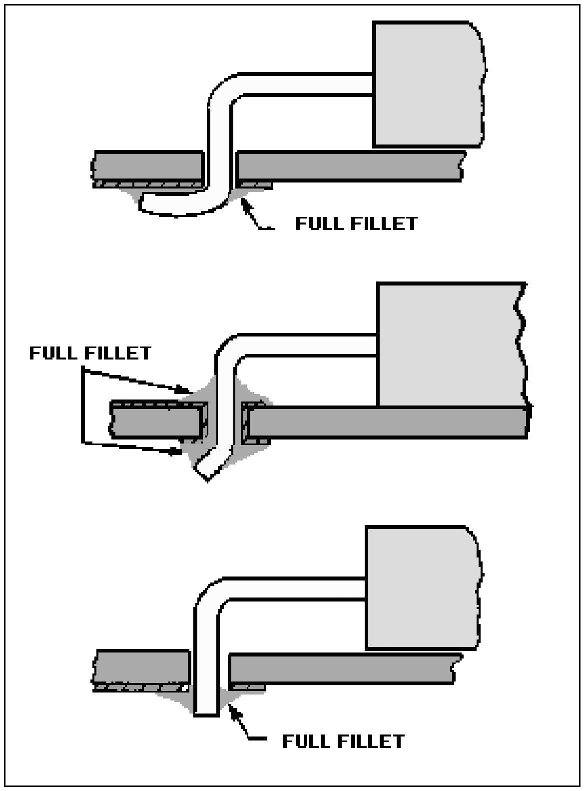


Figure 3-17.—Preferred solder joint.

When a solder joint is completed, solvent must be used to remove all flux residue. The two most highly recommended solvents, in the order of their effectiveness, are 99.5 percent pure ethyl alcohol and 99.5 percent pure isopropyl alcohol.

Q19. What is solderability?

Q20. What is the most common source of heat in electronic soldering?

Q21. What determines the shape and size of a soldering iron tip?

Q22. What term describes a device used to conduct heat away from a component?

Q23. What is the appearance of a properly soldered joint?

REMOVAL AND REPLACEMENT OF DIPS

In topic 1 you learned the advantages of DIPs. They are easily inserted by hand or machine and require no special spreaders, spacers, insulators, or lead-forming tools. Standard hand tools and soldering equipment can be used to remove and replace DIPS.

DIPs may be mounted on a board in two ways: (1) They may be mounted by plugging them into DIP mounting sockets that are soldered to the printed circuit boards or (2) they are soldered in place and may or may not be conformally coated. Although plug-ins are very easy to service, they lack the reliability of soldered-in units, do not meet MILSPECS, and are seldom used in military designed equipment. They are susceptible to loosening because of vibration and to poor electrical contact because of dust and dirt and corrosion.

Removal of Plug-In DIPs

To remove plug-in DIPs, use an approved DIP puller, such as the one shown in figure 3-18. The puller shown is a plastic device that slips over the ends of the DIP and lifts the DIP evenly out of the socket. Before the DIP is removed, the board is marked or a sketch is made of the DIP reference mark location; then the reference mark for the replacement part will be in the proper position. The DIP is grasped with the puller and gently lifted straight out of the socket. Lifting one side or one end first results in bent leads. If the removed DIP is to be placed back in the circuit, particular care is taken in straightening bent leads to prevent breaking. To straighten bent leads, the technician grasps the wide portion of the lead with one pair of smooth-jaw needle nose pliers; with another pair, the technician then bends the lead into alignment with the other leads. Tools, used for lead straightening, should be cleaned with solvent to remove contaminants.

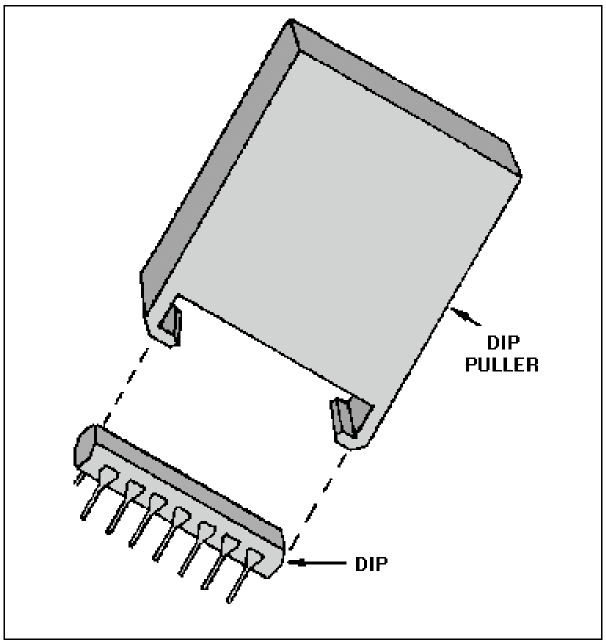


Figure 3-18.—Typical DIP puller.

To replace a plug-in DIP, the technician should clean the leads with solvent and then check the proper positioning of the reference mark. To do this, the technician holds the DIP body between the thumb and forefinger and places the part on the socket to check pin alignment. *The pins are not touched*. If all pins are properly aligned, the technician presses the part gently into the socket until the part is firmly seated. As pressure is applied, each pin is checked to ensure that all pins are going into the socket. If pins tend to bend, the part is removed and the pins are straightened. The socket is then inspected to make sure the holes are not obstructed. Then the process is repeated. After a thorough visual inspection, the card should be ready for testing.

Removal and Replacement of Soldered-In DIPs

The removal of soldered-in DIPs without conformal coatings is essentially the same as the removal of discrete components, except that a skipping pattern is always used. A skipping pattern is one that skips from pad to pad, never heating two pads next to each other. This reduces heat accumulation and reduces the chance of damage to the board. Of course, many more leads should be desoldered before the part can be removed. Special care must be exercised to make sure all leads are completely free before an attempt is made to lift the part off the board. If the part is known to be faulty or if normal removal may damage the board, then the leads should be clipped. Once this has been done, desoldering can be done from both sides of the board. After the clipped leads have been desoldered, they can be removed with tweezers or pliers.

The removal of DIPs from boards with conformal coatings should be completed in the same manner as for other components. The coating should be removed using the preferred method of removal for that particular type of material. The coating should be removed from both sides of the board after masking off the work area. Particular care should be taken when removing the material from around the delicate leads. If the part is to be reused, as much of the coating should be removed from the leads as possible. As with DIPs without conformal coatings, if the part is known to be bad or if the possibility of board damage exists, the leads are clipped; the part and leads are then removed as described earlier in this section. Once the part has been removed, the work area should be completely cleaned to remove any remaining coating or solder.

The steps for replacing a soldered-in DIP are similar to those for replacing a plug-in DIP. Once the part is in position, it is soldered using the same standard used by the manufacturer, or as close to that standard as is possible with the available equipment. The joints should be soldered as quickly as possible using only as much heat as is necessary using a skipping pattern. The repaired card should then be visually inspected for defects in workmanship, and testing of the card should take place. Once the successful repair has been accomplished, a conformal coating should be applied to the work area.

REMOVAL AND REPLACEMENT OF TO PACKAGES

You should recall from chapter 1 of this module, that TO packages are mounted in two ways—plugged-in or embedded. The term *plug-in,* when referring to TOs, should not be confused with DIP plug-ins. TOs are normally soldered in place. You will come across sockets for TOs, but not as frequently as for DIPs. Figure 3-19 shows the methods of mounting TOs. Notice that plug-ins may either be mounted flush with the board surface or above the surface with or without a spacer. The air gap or spacer may be used by the manufacturer for a particular purpose. This type of mounting could be used for heat dissipation, short circuit protection, or to limit parasitic interaction between components. The spacer also provides additional physical support for the TO. The technician is responsible for using the same procedure as the manufacturer to replace TOs or any other components.

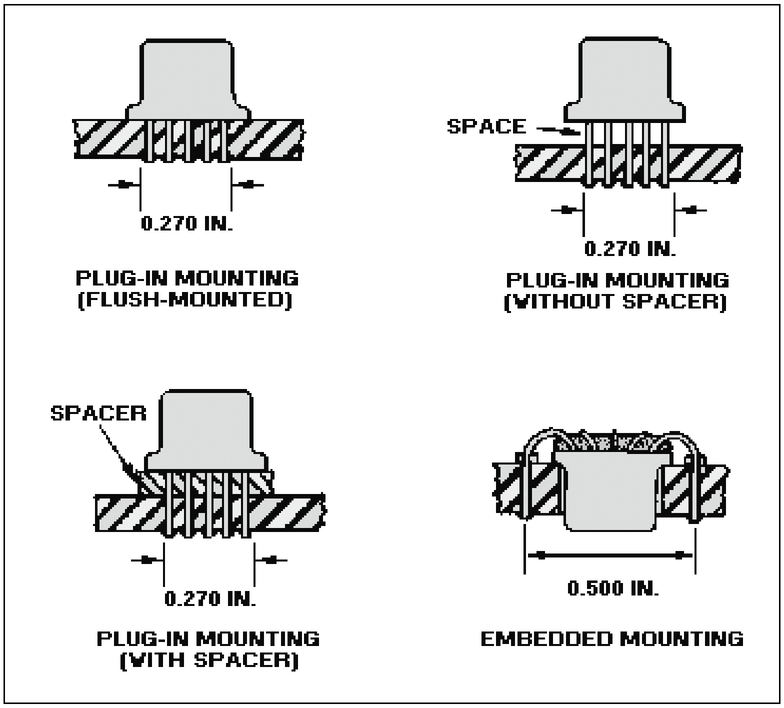


Figure 3-19.—TO mounting techniques.

The procedure for removal of plug-in TOs (with or without conformal coatings) is the same as that used for a similarly mounted DIP or discrete component. The conformal coating is removed if required. Leads are desoldered and gently lifted out of the board. Then board terminals and component leads are cleaned.

In some plug-ins, the leads must be formed before they are placed in a circuit. Care should be taken to ensure that seal damage does not occur and that formed leads do not touch the TO case. This would result in a short-circuit.

When the new part or the one that was removed is installed, the leads are slipped through the spacer if required, and the part is properly positioned (reference tab in the proper location). The leads are aligned with the terminal holes and gently pressed into position. The part is soldered into place and visually inspected. Then the card is tested and the conformal coating is replaced if required.

The removal of an imbedded TO package varies only slightly from the removal of other types of mountings. First, the work area is masked and the conformal coating is removed if required. Then the desoldering handpiece is used to remove the solder from each lead. When all leads are free, the TO is pushed out of the board. If all the leads are free, the TO should slip out of the board easily. The package should not be forced out of the board. Excessive pressure may cause additional damage. If the leads are not completely free, the leads must be clipped and removed after the package is out of the board. This process is shown in figure 3-20.

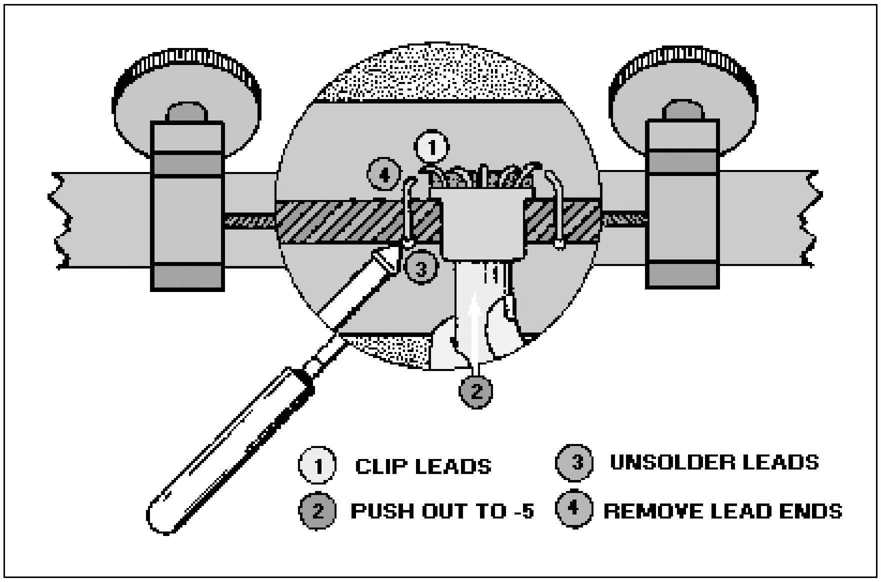


Figure 3-20.—Imbedded TO removal.

The most critical part of replacing an imbedded TO is the lead formation. The leads are formed to match the original part as closely as possible. Once the body and leads are seated, the leads can be soldered and the board inspected.

REMOVAL AND REPLACEMENT OF FLAT PACKS

Up to this point, all of the components discussed have had through-the-board leads. In addition, the removal and replacement of discrete components, DIPs, and TOs have been similar.

PLANAR-MOUNTED COMPONENTS (FLAT-PACKS)

Different techniques are used in the removal and replacement of flat packs and devices with on-the-board terminations. Lap-flow solder joints require that the technician pay particular attention to workmanship. Some of the standards of workmanship will be discussed later in this section.

Flat-Pack Removal

Prior to the removal of a flat pack, as with other ICs, a sketch should be prepared to identify the proper positioning of the part. The conformal coating should be removed as required.

To remove the flat pack, the 2M technician carefully heats the leads and lifts them free with tweezers. If the part is to be reused, special care is taken not to damage or bend the leads. The work area around the component should then be thoroughly cleaned and prepared for the new part.

Flat-Pack Replacement

Flat packs attached to boards normally have formed and trimmed leads. Manufacturers form and trim the leads in one operation with a combination die. However, most replacement flat packs are received in a protective holder (figure 3-21) and the leads must be formed and trimmed by hand. Cost prevents equipping the repair station with the variety of tools and dies to form leads because of the variety of component configurations.

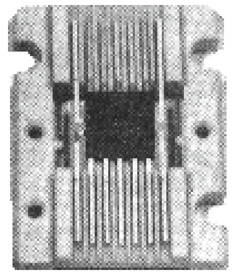


Figure 3-21.—Flat pack in protective holder.

LEAD-BENDING TECHNIQUES.—The 2M technician learns several methods of lead forming that will provide proper contact for soldering and circuit operations. The techniques used to bend leads include the use of specialized tools and such common items as flat toothpicks, bobby pins, and excess component leads. Care is taken not to stress the seal of the component during any step of the lead forming. Figure 3-22 illustrates two views, view (A) and view (B), of properly formed flat-pack leads.

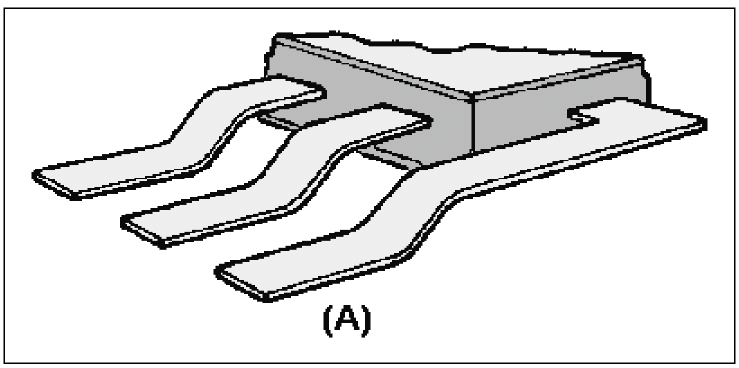


Figure 3-22A.—Properly formed flat pack leads.

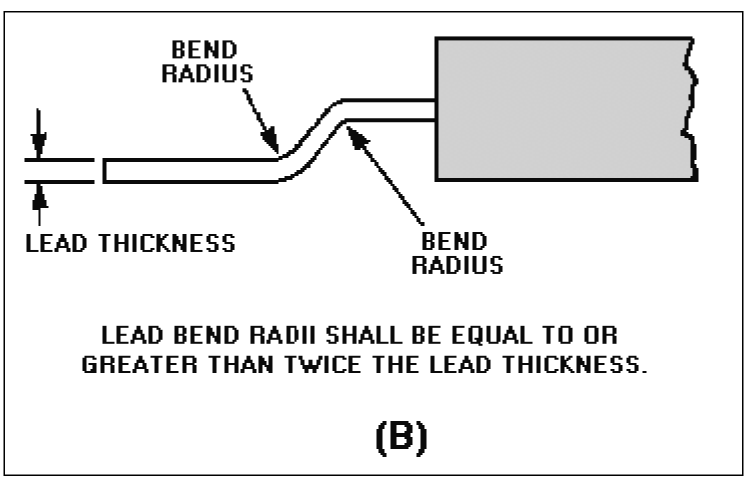


Figure 3-22B.—Properly formed flat pack leads.

Because most replacement flat packs come with leads that are longer than required, they must be trimmed before they are soldered. The removed part is used as a guide in determining lead length. Surgical scissors or scalpels are recommended for use in cutting flat-pack leads. Surgical scissors permit all leads to be cut to the required lead length in a smooth operation with no physical shock transmitted to the IC.

LAP-SOLDERING CONNECTIONS.—Before a connection is lap-soldered, the solder pads are cleaned and pretinned and the component leads are tinned. This is particularly important if they are gold plated. The IC is properly positioned on the pad areas, and the soldering process is a matter of "sweating" the two conductors together. When multilead components, such as ICs, are soldered, a skipping pattern is used to prevent excessive heat buildup in a single area of the board or component. When soldering is completed, all solder connections are thoroughly cleaned. All joints should be inspected and tested. The standards of workmanship are more specific for flat-pack installation.

Q24. When removing the component, under what circumstances may component leads be clipped?

Q25. How are imbedded TOs removed once the leads are free?

Q26. How is a flat pack removed from a pcb?

Q27. How do you prevent excessive heat buildup on an area of a board when soldering multilead components?

Q28. What are the two final steps of any repair?

REPAIR OF PRINTED CIRCUIT BOARDS AND CARDS

Removal and replacement of components on boards and circuit cards are, by far, the most common types of repair. Equally important is the repair of damaged or broken cards. Proper repair of damaged boards not only maintains reliability of the board but also maintains reliability of the system.

Cards and boards may be damaged in any of several ways and by a number of causes. *Untrained personnel making improper repairs and technicians using improper tools are two major causes of damage.* Improper shipping, packaging, storage, and use are also common sources of damage. The source of damage most familiar to technicians is operational failure. Operational failures include cracking caused by heat, warping, component overheating, and faulty wiring.

Before attempting board repairs, the technician should thoroughly inspect the damage. The decision to repair or discard the piece depends on the extent of damage, the level of maintenance authorized, operational requirements, and the availability of repair parts and materials. The following procedures will help you become familiar with the steps necessary to repair particular types of damage. *Remember, only qualified personnel are authorized to attempt these repairs.*

Repair of Conductor and Termination Pads

Conductor (run) and pad damage is very common. The technician must examine the board for nicks, tears, or scratches that have not broken the circuit, as well as for complete breaks, as shown in figure 3-23. Crack damage may exist as nicks or scratches in the conductor. These nicks or scratches must be repaired if over one-tenth of the cross-sectional area of the conductor is affected as current-carrying capability is reduced. Cracks may also penetrate the conductor.

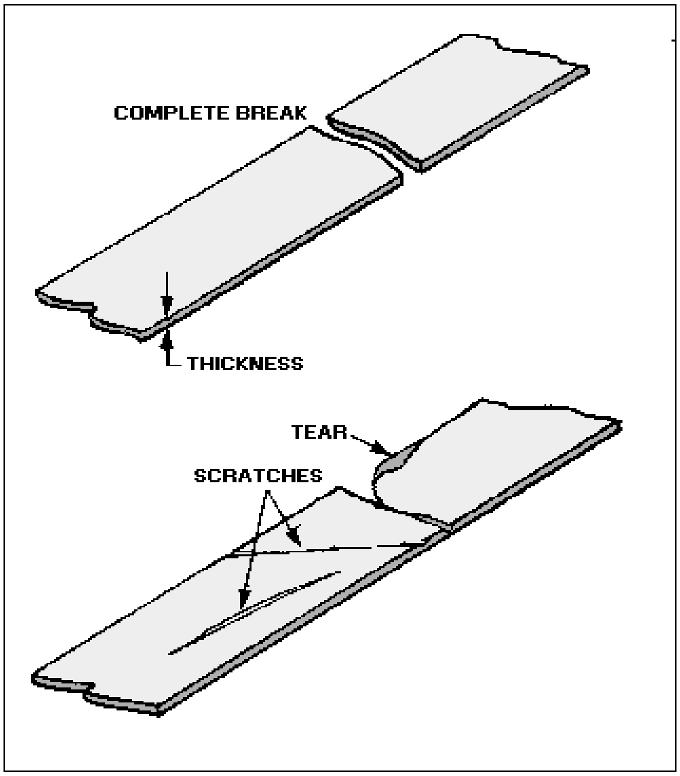


Figure 3-23.—Pcb conductor damage.

CRACK REPAIR.—Four techniques are used to repair cracks in printed circuit conductors. One method is to flow solder across the crack to form a solder bridge. This is not a high-reliability repair since the solder in the break will crack easily.

The second method is to lap-solder a piece of wire across the crack. This method produces a stronger bond than a solder bridge; but it is not highly reliable, as the solder may crack.

A third repair technique is to drill a hole through the board where the crack is located and then to install an eyelet in the hole and solder it into place.

The fourth method is to use the clinched-staple method, shown in figure 3-24. It is the most reliable method and is recommended in nearly all cases.

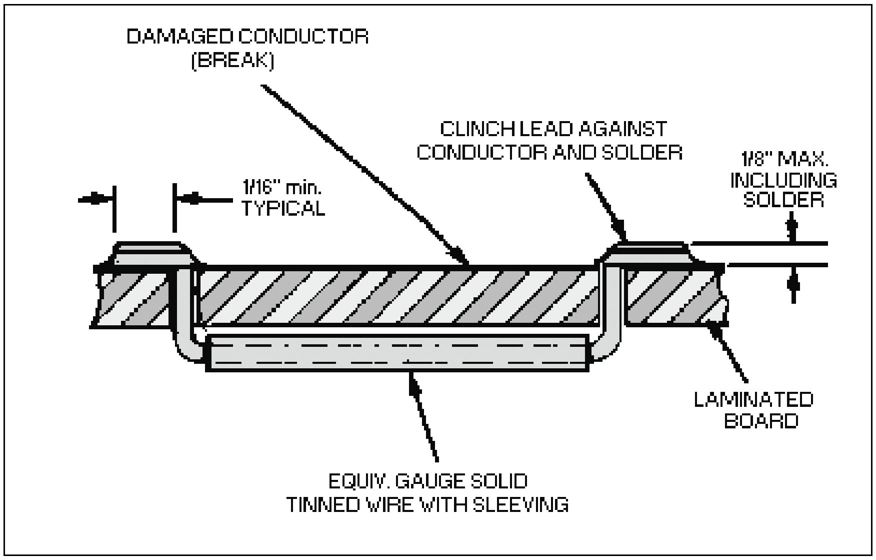


Figure 3-24.—Clinched-staple repair of broken conductor.

Pads or conductor runs may be completely missing from the board. These missing pads or runs must be replaced. Also included in this type of damage are conductors that are present but damaged beyond repair.

REPLACING DAMAGED OR MISSING CONDUCTORS.—The procedures used to replace damaged or missing conductors are essentially the same as using the clinched-staple method of conductor repair.

REPLACING THE TERMINATION PAD.—Many times the termination pad, as well as part of the conductor, is missing on the board. In these cases, a replacement pad is obtained from a scrap circuit board. Refer to figure 3-25 as you study each step.

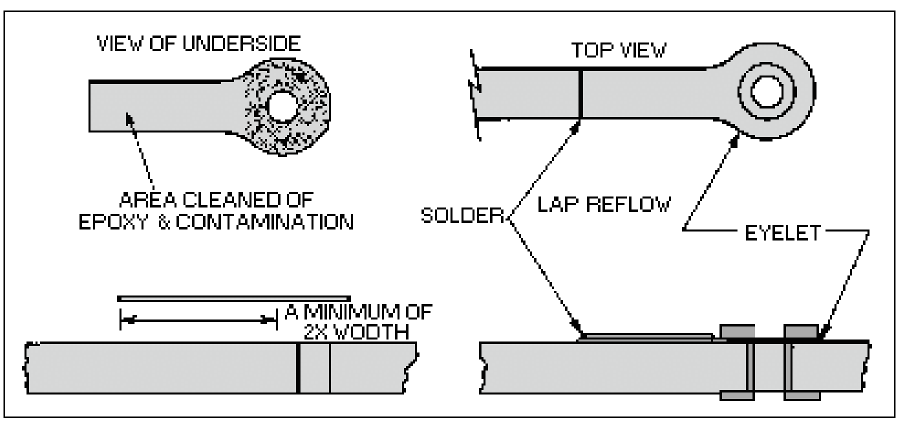


Figure 3-25.—Replacement of damaged termination pad.

The underside of the replacement pad and the area where it will be installed is cleaned. An epoxy is used to fasten the replacement pad to the board. An eyelet is installed to reinforce the pad before the epoxy sets and cures. This ensures a good mechanical bond between the board and pad and provides good electrical contact for components. After the epoxy cures, the new pad is lap-soldered to the original run.

REPAIRING DELAMINATED CONDUCTORS.—DELAMINATED CONDUCTORS (figure 3-26) are classified as conductors no longer bonded to the board surface. Separation of the laminations may occur only on a part of the conductor. Proper epoxying techniques ensure complete bonding of the conductor to the circuit board laminate. The following procedures are used to obtain a proper bond:

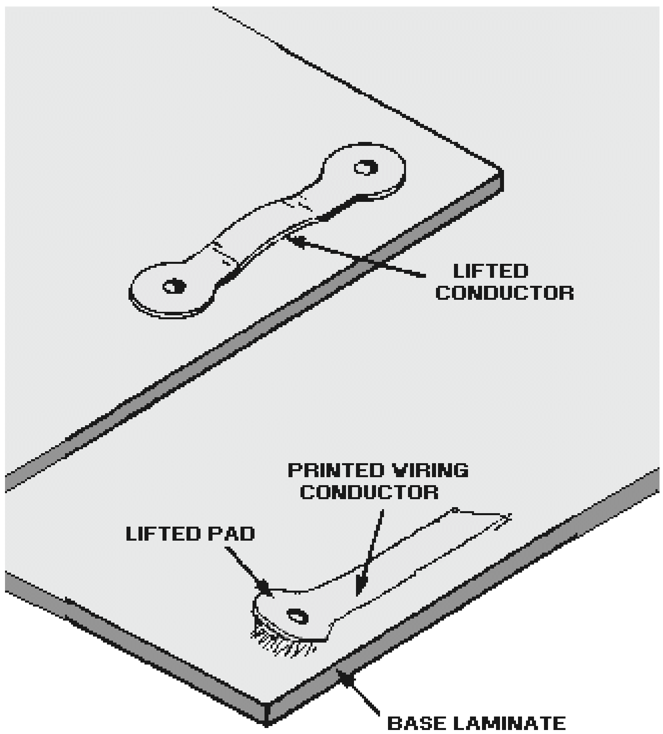


Figure 3-26.—Delaminated conductors.

1. A small amount of epoxy is mixed and applied to the conductor and the conductor path; no areas are left uncoated.

2. The conductor is clamped firmly against the board surface until the epoxy has completely cured.

REPLACING EYELETS.—Eyelets have been referred to in several places in this topic. Not only are they used for through-the-board terminations, but also to reinforce some types of board repairs. As with any kind of material, eyelets are subject to damage. Eyelets may break, they may be installed improperly, or they may be missing from the equipment. When an eyelet is missing or damaged, regardless of the kind of damage, it should be replaced. The guidelines for the selection and installation of new eyelets are far too complex to explain here. However, they do comprise a large part of the 2M technician's training.

Repair of Cracked Boards

When boards are cracked, the length and depth of the cracks must be determined. Also, the disruption to conductors and components caused by cracks must be determined by visual inspection. To avoid causing additional damage, the technician must exercise care when examining cracked boards and must not flex the board. Rebuilding techniques must be used to repair damage, such as cracks, breaks, and holes, that extend through the board. The following steps are used to repair cracks:

1. Abrasive methods are used to remove all chips and fractured material.

2. The edges of the removed area are beveled and undercut to provide bond strength.

3. A smoothly surfaced, nonporous object is fastened tightly against one side of the removed area.

4. The cutaway area is filled with a compound of epoxy and powdered fiberglass (figure 3-27). Extreme care is exercised to prevent the formation of voids or air bubbles in the mixture.

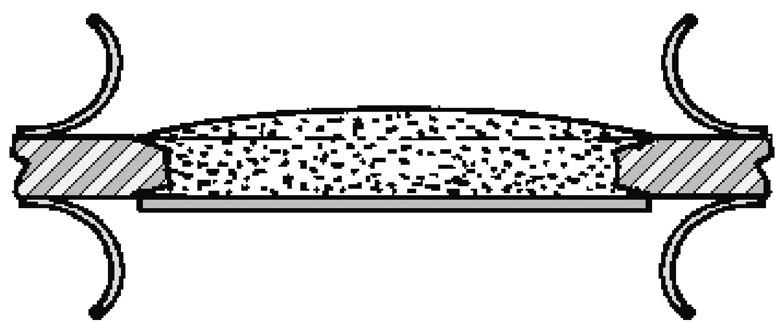


Figure 3-27.—Repair of cracked pcbs.

5. The surface of the filled area is smoothed to make it level with the surface of the original board.

6. The board is cured, smoothed, redrilled, and cleaned.

Broken Board Repair

Broken boards should be examined to determine if all parts of the board are present and if circuit conductors or components are affected by the break. They are also examined to determine if the broken pieces may be rejoined reliably or if new pieces must be manufactured.

Breaks and holes are repaired in the same manner as cracks unless broken pieces are missing or the hole exceeds 1/2 inch in diameter. In such cases, the following repair steps are used:

1. The same technique used in repairing cracks is used to prepare the damaged edge.

2. A piece as close in size to the missing area as possible is cut from a scrap board of the same type and thickness. The edges of this piece are prepared in the same manner as the edges of the hole.

3. A smooth-surfaced object is tightly fastened over one side of the repair area, and the board is firmly clamped in an immovable position with the uncovered area facing up.

4. The replacement piece is positioned as nearly as possible to the original board configuration and firmly clamped into place.

5. The repair is completed using the same epoxy-fiberglass mixture and repair techniques used in the patching repair method discussed in the following section on "Burned Board Repair."

Burned Board Repair

Scorched, charred, or deeply burned boards should be inspected to determine the size of the discolored area and to identify melted or blackened conductors and burned, melted, or blackened components. The depth of the damage, which may range from a slight surface discoloration to a hole burned through the circuit board, should also be determined. Damage not extending through the board may be repaired by patching (figure 3-28). The following procedure is used in the repair of these boards.

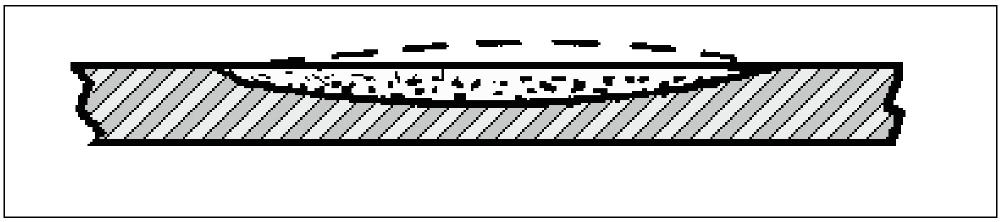


Figure 3-28.—Repair of surface damage.

1. If the board is scorched, charred, or burned, all discolored board material is removed by abrasive methods, as shown in figure 3-29. Several components in the affected area may have to be desoldered and removed before the repair is continued.

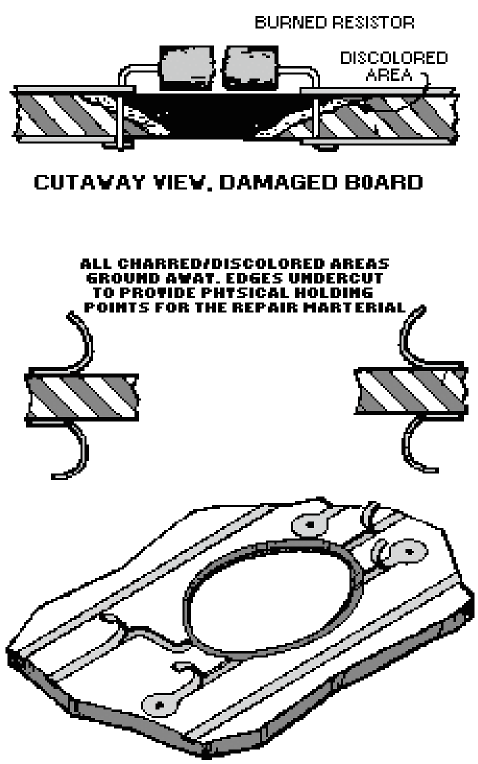


Figure 3-29.—Repair of burned boards.

2. Repairable delaminations not extending to the edge of the circuit board should be cut away by abrasive methods until no delaminated material remains.

3. Delaminated material is not removed if it is repairable.

4. After all damaged board material is removed, the edge of the removed area is beveled and undercut to provide holding points for the repair material.

5. Solvent is used to clean thoroughly and to remove all loose particles.

6. A compound of epoxy and powdered fiberglass is mixed and used to fill the cutaway area.

7. The epoxy repair mixture is cured according to the manufacturer's instructions.

8. The surface of the filled area is leveled after the compound is cured.

9. If delaminations extend to the edge of the board, the delaminated layers are filled completely with the repair mixture and clamped firmly together between two flat surfaces.

10. After the cure is completed, abrasive methods are used to smooth the repaired surface to the same level as the original board.

11. If necessary, needed holes are redrilled in the damaged area, runs are replaced, eyelets and components are installed, and the area is cleaned. Figure 3-30 shows the repaired area ready for components.

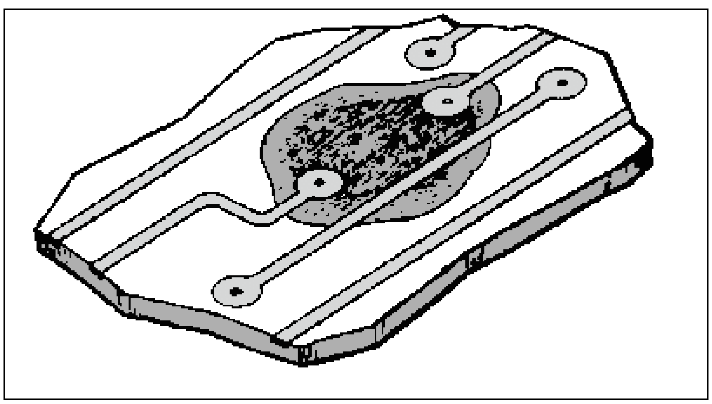


Figure 3-30.—Repaired board ready for components.

Q29. List three causes of damage to printed circuit boards.

Q30. What is the preferred method of repairing cracked runs on boards?

Q31. Damaged or missing termination pads are replaced using what procedure?

Q32. How is board damage caused by technicians?

Q33. What combination of materials is used to patch or build up damaged areas of boards?

SAFETY

Safety is a subject of utmost importance to all technical personnel. Potentially hazardous situations exist in almost any work area. The disregard of safety precautions can result in personal injury or in the loss of equipment or equipment capabilities.

In this section we will discuss two types of safety factors. First, we will cover damage that can occur to electronic components because of electrostatic discharge (ESD) and improper handling and stowage of parts and equipment. Second, we will cover personal safety precautions that specifically concern the technician.

ELECTROSTATIC DISCHARGE

Electrostatic discharge (ESD) can destroy or damage many electronic components including integrated circuits and discrete semiconductor devices. Certain devices are more susceptible to ESD damage than others. Because of this, warning symbols are now used to identify ESD-sensitive (ESDS) items (figure 3-31).

Figure 3-31.—Warning symbols for ESDS devices.

Static electricity is created whenever two substances (solid or fluid) are rubbed together or separated. This rubbing or separation causes the transfer of electrons from one substance to the other; one substance then becomes positively charged and the other becomes negatively charged. When either of these charged substances comes in contact with a conductor, an electrical current flows until that substance is at the same electrical potential as ground.

You commonly experience static build-up during the winter months when you walk across a vinyl or carpeted floor. (Synthetics, especially plastics, are excellent generators of static electricity.) If you then touch a doorknob or other conductor, an electrical arc to ground may result and you may receive a slight shock. For a person to experience such a shock, the electrostatic potential created must be 3,500 to 4,000 volts. Lesser voltages, although present and similarly discharged, normally are not apparent to a person's nervous system. Some typical measured static charges caused by various actions are shown in table 3-2.

Table 3-2.—Typical Measured Statics Charges (in volts)

ITEM	RELATIVE HUMIDITY	
	LOW (10-20%)	HIGH (65-90%)
WALKING ACROSS CARPET	35,000	1,500
WALKING OVER VINYL FLOOR	12,000	250
WORKER AT BENCH	6,000	100
VINYL ENVELOPES FOR WORK INSTRUCT.	7,000	600
POLY BAG PICKED UP FROM BENCH	20,000	1,200
WORK CHAIR PADDED WITH URETHANE FOAM	18,000	1,500

Metal oxide semiconductor (MOS) devices are the most susceptible to damage from ESD. For example, an MOS field-effect transistor (MOSFET) can be damaged by a static voltage potential of as little as 35 volts. Commonly used discrete bipolar transistors and diodes (often used in ESD-protective circuits), although less susceptible to ESD, can be damaged by voltage potentials of less than 3,000 electrostatic volts. Damage does not always result in sudden device failure but sometimes results in device degradation and early failure. Table 3-2 clearly shows that electrostatic voltages well in excess of 3,000 volts can be easily generated, especially under low-humidity conditions. ESD damage of ESDS parts or circuit assemblies is possible wherever two or more pins of any of these devices are electrically exposed or have low impedance paths. Similarly, an ESDS device in a printed circuit board, or even in another pcb that is electrically connected in a series can be damaged if it provides a path to ground. Electrostatic discharge damage can occur during the manufacture of equipment or during the servicing of the equipment. Damage can occur anytime devices or assemblies are handled, replaced, tested, or inserted into a connector.

Technicians should be aware of the many sources of static charge. Table 3-3 lists many common sources of electrostatic charge. Although they are of little consequence during most daily activity, they become extremely important when you work with ESD material.

Table 3-3.—Common Sources of Electrostatic Charge

OBJECT OR PROCESS	MATERIAL OR ACTIVITY
WORK SURFACES	• WAXED, PAINTED, OR VARNISHED SURFACES • COMMON VINYL OR PLASTICS
FLOORS	• SEALED CONCRETE • WAXED, FINISHED WOOD • COMMON VINYL TILE OR SHEETING
CLOTHES	• COMMON CLEAN ROOM SMOCKS • COMMON SYNTHETIC PERSONNEL GARMENTS • NONCONDUCTIVE SHOES • VIRGIN COTTON*
CHAIRS	• FINISHED WOOD • VINYL • FIBERGLASS
PACKAGING AND HANDLING	• COMMON PLASTIC—BAGS, WRAPS, ENVELOPES • COMMON BUBBLE PACK, FOAM • COMMON PLASTIC TRAYS, PLASTIC TOTE BOXES, VIALS, PARTS BINS
ASSEMBLY, CLEANING, TEST AND REPAIR AREAS	• SPRAY CLEANERS • COMMON PLASTIC SOLDER SUCKERS • SOLDER IRONS WITH UNGROUNDED TIPS • SOLVENT BRUSHES (SYNTHETIC BRISTLES) • CLEANING OR DRYING BY FLUID OR EVAPORATION • TEMPERATURE CHAMBERS • CRYOGENIC SPRAYS • HEAT GUNS AND BLOWERS • SAND BLASTING • ELECTROSTATIC COPIERS
PERSONNEL ITEMS	• STYROFOAM COFFEE OR PLASTIC DRINK CUPS • PLASTIC OR RUBBER HAIR COMBS OR BRUSHES • CELLOPHANE OR PLASTIC CANDY, GUM OR CIGARETTE WRAPPERS • VINYL PURSES

*VIRGIN COTTON CAN BE A STATIC SOURCE AT LOW RELATIVE HUMIDITIES (BELOW 30 PERCENT)

Prevention of ESD Damage

Certified 2M technicians are trained in procedures for reducing the causes of ESD damage. The procedures are similar for all levels of maintenance. The following procedure is an example of some of the protective measures used to prevent ESD damage.

1. Before starting to service equipment, the technician should be grounded to discharge any static electric charge built up on the body. This can be accomplished with the use of a test lead (a single-wire conductor with a series resistance of 1 megohm equipped with alligator clips on each end). One clip end is connected to the grounded equipment frame, and the other clip end is

touched with a bare hand. Figure 3-32 shows a more refined ground strap which frees both hands for work.

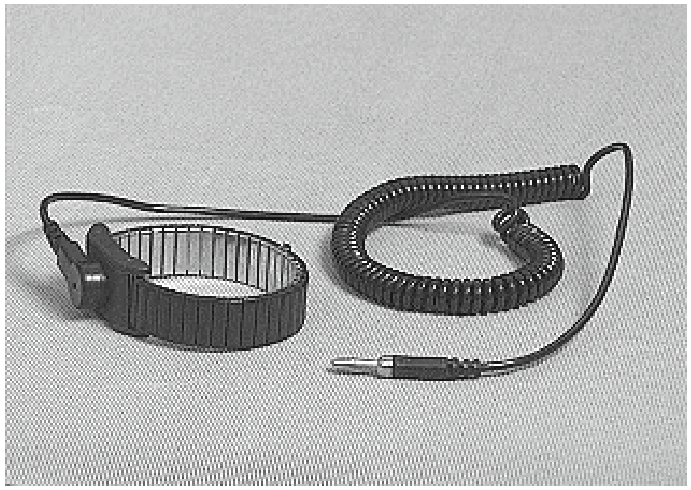


Figure 3-32.—ESD wrist strap.

2. Equipment technical manuals and packaging material should be checked for ESD warnings and instructions.

3. Prior to opening an electrostatic unit package of an electrostatic sensitive device or assembly, clip the free end of the test lead to the package. This will cause any static electricity which may have built up on the package to discharge. The other end remains connected to the equipment frame or other ESD ground. Keep the unit package grounded until the replacement device or assembly is placed in the unit package.

4. Minimize handling of ESDS devices and assemblies. Keep replacement devices or assemblies, with their connector shorting bars, clips, and so forth, intact in their electrostatic-free packages until needed. Place removed repairable ESD devices or assemblies with their connector shorting bars/clips installed in electrostatic-free packages as soon as they are removed from the equipment. ESDS devices or assemblies are to be transported and stored only in protective packaging.

5. Always avoid unnecessary physical movement, such as scuffing the feet, when handling ESDS devices or assemblies. Such movement will generate additional charges of static electricity.

6. When removing or replacing an ESDS device or assembly in the equipment, hold the device or assembly through the electrostatic-free wrap if possible. Otherwise pick up the device or assembly by its body only. Do not touch component leads, connector pins, or any other electrical connections or paths on boards, even though they are covered by conformal coating.

7. Do not permit ESDS devices or assemblies to come in contact with clothing or other ungrounded materials that could have an electrostatic charge. The charges on a nonconducting material are not equal. A plastic storage bag may have a −10,000 volt potential 1/2 inch from a +15,000 volt potential, with many such charges all over the bag. Placing a circuit card inside the bag allows the charges to equalize through the pcb conductive paths and components, thereby causing failures. Do not hand an ESD device or assembly to another person until the device or assembly is protectively packaged.

8. When moving an ESDS device or assembly, always touch (with bare skin) the surface on which it rests for at least one second before picking it up. Before placing it on any surface, touch the surface with your free hand for at least one second. The bare skin contact provides a safe discharge path for charges accumulated while you are moving around.

9. While servicing equipment containing ESD devices, do not handle or touch materials such as plastic, vinyl, synthetic textiles, polished wood, fiberglass, or similar items which create static charges; or, be sure to repeat the grounding action with the bare hands after contacting these materials. These materials are prime electrostatic generators.

10. If possible, avoid repairs that require soldering at the equipment level. Soldering irons must have heater/tips assemblies that are grounded to ac electrical ground. Do not use ordinary plastic solder suckers (special antistatic solder suckers are commercially available).

11. Ground the leads of test equipment momentarily before you energize the test equipment and before you probe ESD items.

Grounded Workbenches

Workbenches on which ESDS items will be placed and that will be contacted by personnel should have ESD protective work surfaces. These protective surfaces should cover the areas where ESD items will be placed. Personnel ground straps are also necessary for ESD protective workbench surfaces. These straps prevent people from discharging a static charge through an ESDS item to the work bench surface. The workbench surface should be connected to ground through a ground cable. The resistance in the bench top ground cable should be located at or near the point of contact with the workbench top. The resistance should be high enough to limit any leakage current to 5 milliamperes or less; this is taking into consideration the highest voltage source within reach of grounded people and all parallel resistances to ground, such as wrist ground straps, table tops, and conductive floors. See figure 3-33 for a typical ESD ground workbench.

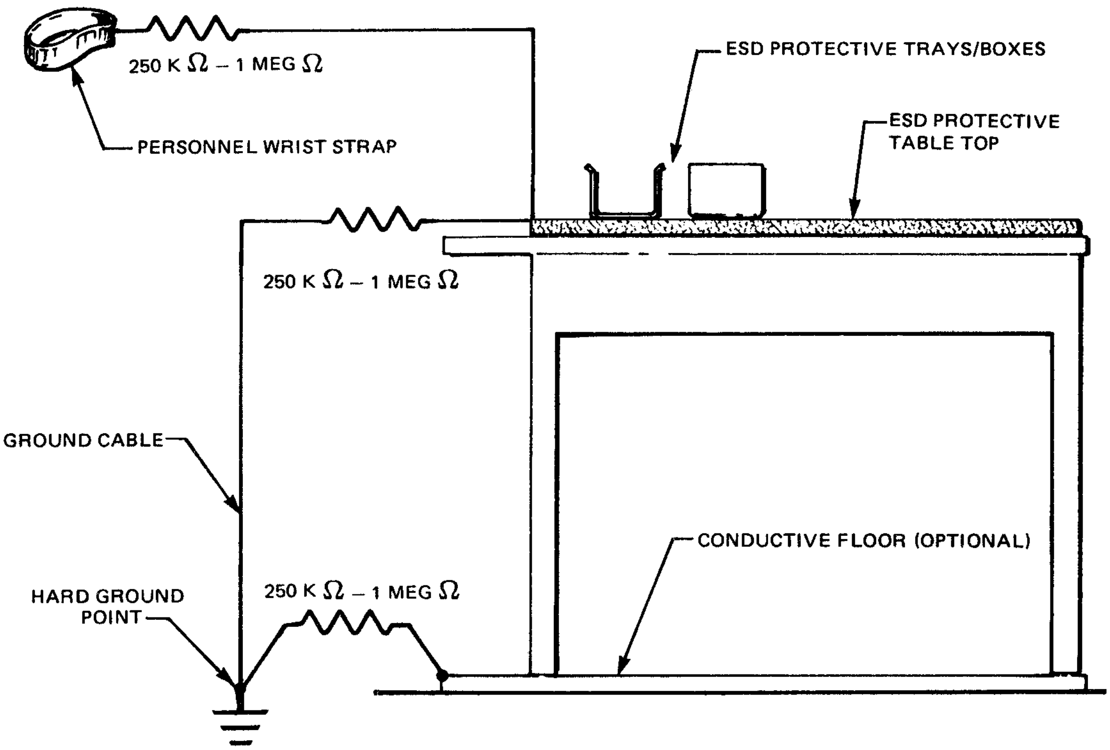


Figure 3-33.—Typical ESD ground workbench.

Energized equipment provides protection from ESD damage through operating circuitry. Circuit cards with ESD sensitive devices are generally considered safe when installed in an equipment rack; but they may be susceptible to damage if a "drawer" or "module" is removed and if connector pins are touched (even putting on plastic covers can transfer charges that do damage). *There must not be any energized equipment placed on the conductive ESD work surface.* An ESD work area is for "dead" equipment ONLY.

ESD protection is critical. If you should be assigned to 2M repair school, your education in ESD prevention will be quite extensive.

PERSONAL SAFETY

Throughout your career you will be aware of emphasis placed on safety. Safety rules remind you of potential dangers in work. Most accidents are preventable. Accidents don't happen without a cause. Most accidents are the result of not following prescribed safe operating procedures.

This would be a good time to review the safety section in topic 5 of NEETS, Module 2, *Introduction to Alternating Current and Transformers*. That section covers the basics of electrical shock and how to prevent it.

The 2M technician should be aware of other potential dangers in addition to the dangers of electrical shock. These dangers are discussed in the following paragraphs.

Power Tools

Hazards associated with the use of power tools include electrical shock, cuts, and particles in the eye. Safe tool use practices reduce or eliminate such accidents. Listed below are some of the general safety precautions that you should observe when your work requires the use of power tools.

- Ensure that all metal-cased power tools are properly grounded.
- Do not use spliced cables unless an emergency warrants the risks involved.
- Inspect the cord and plug for proper connection. Do not use any power tool that has a frayed cord or broken or damaged plug.
- Make sure that the on/off switch is in the OFF position before inserting or removing the plug from the receptacle.
- Always unplug the extension cord from the receptacle before the portable power tool is unplugged from the extension cord.
- Ensure all cables are positioned so they will not constitute a tripping hazard.
- Wear eye protection (goggles) in work areas where particles may strike the eye.
- After completing a task requiring a portable power tool, disconnect the power cord as described above and store the tool in its assigned location.

Soldering Iron

When using a soldering iron, remember the following:

- To avoid burns, always assume that a plugged-in soldering iron is HOT.
- Never rest a heated iron anywhere but in a holder provided for that purpose. Faulty action on your part could result in fire, extensive equipment damage, and/or serious injuries.
- Never use an excessive amount of solder. Drippings can cause serious skin or eye burns and can cause short circuits.
- Do not swing an iron to remove excess solder. Bits of hot solder can cause serious skin or eye burns or may ignite combustible material in the work area.
- When cleaning an iron, use a natural fiber cleaning cloth; never use synthetics, which melt. Do not hold the cleaning cloth in your hand. Always place the cloth on a suitable surface; then wipe the iron across it to avoid burning your hand.
- Hold small soldering jobs with pliers or a suitable clamping device to avoid burns. Never hold the work in your hand.
- Do not use an iron that has a frayed cord or damaged plug.

- Do not solder electronic equipment unless the equipment is electrically disconnected from the power supply circuit.

- After completing a task requiring a soldering iron other than the iron that is part of a work station, disconnect the power cord from the receptacle. When the iron has cooled, store it in its assigned stowage area.

Cleaning Solvents

The technician who smokes while using a cleaning solvent is inviting disaster. Unfortunately, many such disasters have occurred. For this reason, the Navy does not permit the use of gasoline, benzine, ether, or like solvents for cleaning since they present potential fire or explosion hazards. Only nonvolatile solvents should be used to clean electrical or electronic apparatus.

In addition to the potential hazard of accidental fire or explosion, most cleaning solvents can damage the human respiratory system where the fumes are breathed for a period of time.

The following positive safety precautions should be followed when performing cleaning operations.

- Use a blower or canvas wind chute to blow air into a compartment in which a cleaning solvent is being used.

- Open all usable port holes and place wind scoops in them.

- Place a fire extinguisher nearby.

- If it can be done, use water compounds instead of other solvents.

- Wear rubber gloves to prevent direct contact with solvents.

- Use goggles when a solvent is being sprayed on surfaces.

- Hold the nozzle close to the object being sprayed.

Where water compounds cannot be used, inhibited methyl chloroform (1.1.1 trichloroethane) should be used. Carbon tetrachloride is not used. Cleaning solvents that end with ETHYLENE are NOT safe to use. Methyl chloroform is an effective cleaner and is as safe as can be expected when reasonable care is exercised, such as adequate ventilation and the observance of fire precautions. When using inhibited methyl chloroform, avoid direct inhalation of the vapor. It is not safe for use, even with a gas mask, because its vapor displaces oxygen in the air.

Aerosol Dispensers

A 2M technician will encounter several uses for aerosol dispensers. The most common type is in applying conformal coatings.

Specific instructions concerning the precautions and procedures that must be observed to prevent physical injury cannot be given in this section because of the many available industrial sprays. However, all personnel concerned with handling aerosol dispensers containing volatile substances must clearly understand the hazards involved. They must also understand the importance of exercising protective measures to prevent personal injury. Strict compliance with the instructions printed on the aerosol dispensers will prevent many accidents that result from misapplication, mishandling, or improper storage of industrial sprays.

The rules for safe use of aerosol dispensers are listed below:

- Carefully read and comply with the instructions printed on the container.
- Do not use any dispenser that is capable of producing dangerous gases or other toxic effects in an enclosed area unless the area is adequately ventilated.
- If a protective coating must be sprayed in an inadequately ventilated space, either an air respirator or a self-contained breathing apparatus should be provided. However, fresh air supplied from outside the enclosure by exhaust fans or portable blowers is preferred. Such equipment prevents inhalation of toxic vapors.
- Do not spray protective coating on warm or energized equipment because this creates a fire hazard.
- Avoid skin contact with the liquid. Contact with some liquids may cause burns, while milder exposure may cause rashes. Some toxic materials are actually absorbed through the skin.
- Do not puncture the dispenser. Because it is pressurized, injury can result.
- Keep dispensers away from direct sunlight, heaters, and other heat sources.
- Do not store dispensers in an environment where the temperature exceeds the limits printed on the can. High temperatures may cause the container to burst.

Q34. List two causes of damage to ESD-sensitive electronic components.

Q35. What is the purpose of the wrist ground strap?

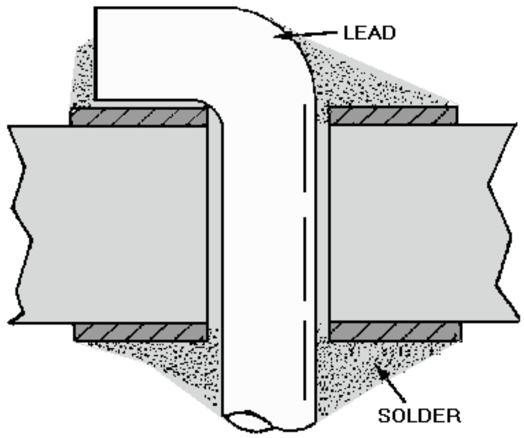
Q36. What is the cause of most accidents?

SUMMARY

This topic has presented information on miniature and microminiature (2M) repair procedures and 2M safety precautions. The information that follows summarizes the important points of this topic.

CONFORMAL COATINGS are protective materials applied to electronic assemblies to prevent damage caused by corrosion, moisture, and stress.

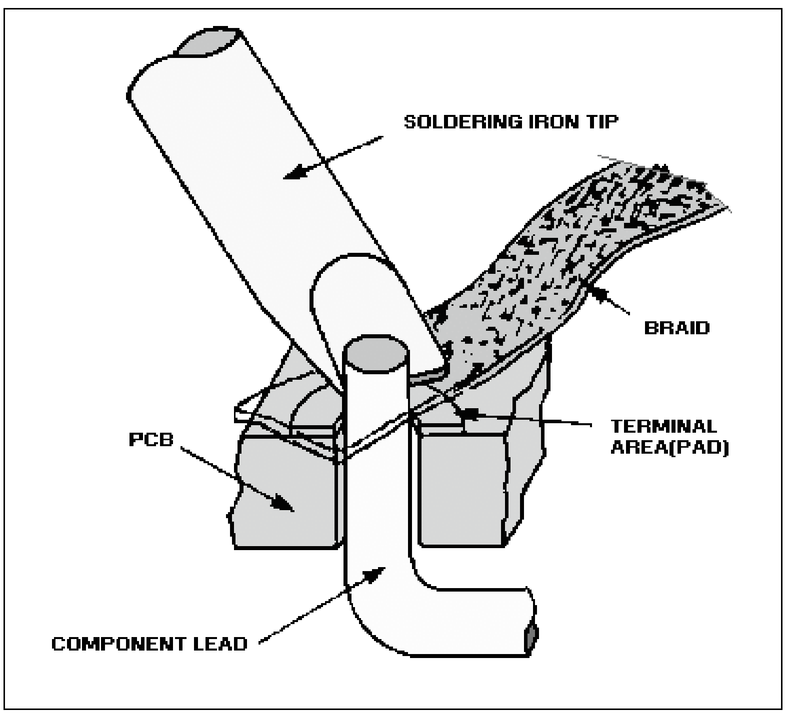
CONFORMAL COATINGS REMOVAL is accomplished mechanically, chemically, or thermally, depending on the material used.



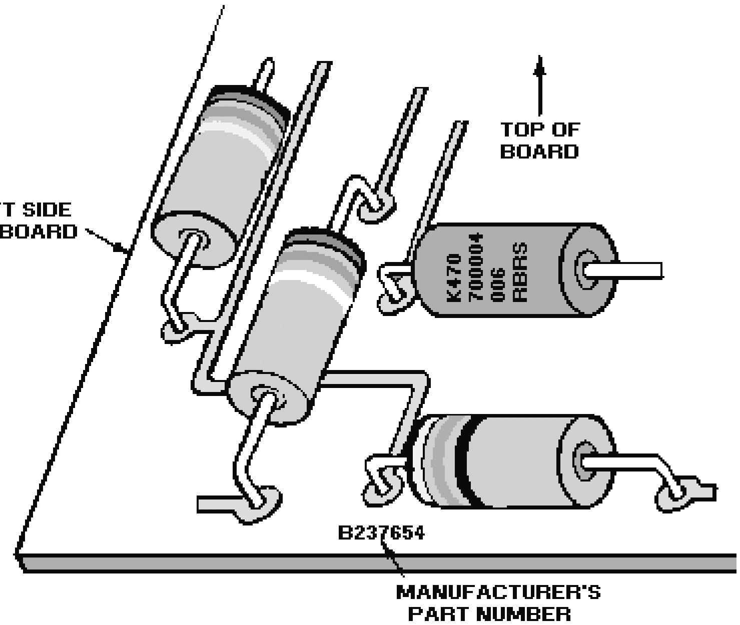
(A) FULLY CLINCHED

Component **LEADS** are terminated either through the board, above the board, or on the board.

SOLDER may be removed by wicking, by a manual vacuum plunger, or by a continuous vacuum solder extractor.



ELECTRONIC ASSEMBLIES should be restored to the original manufacturer's standards using the same orientation and termination method.

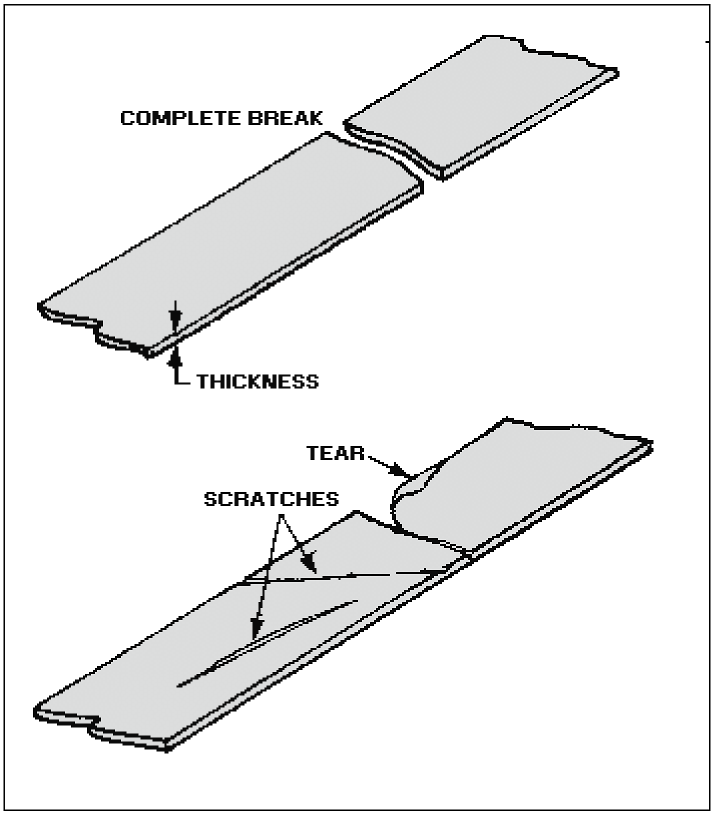


A **GOOD SOLDER JOINT** is bright and shiny with no cracks or pits.

When **REPLACING DIPs**, **TOs**, AND **FLAT PACKS**, make certain that pins are placed in the proper position.

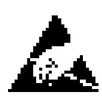
COMPONENT LEADS may be clipped prior to removal only if the part is known to be bad or if normal removal will result in board damage.

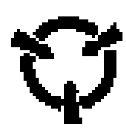
The technician must determine through **INSPECTION** what method of repair is necessary for the board.

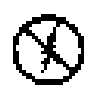


ELECTROSTATIC DISCHARGE (ESD) can damage or destroy many types of electronic components including integrated circuits and discrete components.

Special handling is required for **ELECTROSTATIC-DISCHARGE-SENSITIVE (ESDS)** devices or components.







USE PRESCRIBED SAFETY PRECAUTIONS when you use power tools, soldering irons, cleaning solvents, and aerosol dispensers.

ANSWERS TO QUESTIONS Q1. THROUGH Q36.

A1. Conformal coating.

A2. Chemical, mechanical, and thermal.

A3. Solvents or xylene and trichloroethane.

A4. Mechanical.

A5. To ensure protective characteristics are maintained.

A6. Interfacial connections.

A7. Clinched lead, straight-through, and offset pad.

A8. Above-the-board termination.

A9. On-the-board termination.

A10. During disassembly or repair.

A11. Wicking.

A12. Continuous vacuum.

A13. These methods should not be used.

A14. Manufacturer's standards.

A15. A fine abrasive.

A16. 90 degrees.

A17. They should be readable from a single point.

A18. In the direction of the run.

A19. The ease with which molten solder wets the surfaces of the metals to be joined.

A20. Conductive-type soldering iron.

A21. The type of work to be done.

A22. A thermal shunt.

A23. Bright and shiny with no cracks or pits.

A24. If the component is known to be defective or if the board may be damaged by normal desoldering.

A25. By pushing it gently out of the board.

A26. Heat each lead and lift with tweezers.

A27. Use a skipping pattern.

A28. Inspect and test.

A29. Operational failures, repairs by untrained personnel, repair using improper tools, mishandling, improper shipping, packaging, and storage.

A30. Clinched staple.

A31. Epoxy a replacement pad to the board, set an eyelet, and solder it.

A32. Repairs by untrained personnel and technicians using improper tools.

A33. Epoxy and fiberglass powder.

A34. Esd, improper stowage, and improper handling.

A35. To discharge any static charge built up in the body.

A36. Deviation from prescribed safe operating procedures.

APPENDIX I

GLOSSARY

ALLOWANCE PARTS LIST (APL)—Repair parts required for unit having the equipment/ component listed.

ALLOWANCE EQUIPAGE LIST (AEL)—Equipment requirements for a unit having the exact equipment/component listed.

BEAM-LEAD CHIP—Semiconductor chip with electrodes (leads) extended beyond the wafer.

BONDING WIRES—Fine wires connecting the bonding pads of the chip to the external leads of the package.

BUILT-IN TEST EQUIPMENT (BITE)—Permanently mounted to the equipment for the purpose of testing the equipment.

CABLE HARNESS—A group of wires or ribbons of wiring used to interconnect electronic systems and subsystems.

CATHODE SPUTTERING—Process of producing thin film components.

CERMET—A combination of powdered precious-metal alloys and an inorganic material such as alumina. Used in manufacturing resistors, capacitors, and other components for high-temperature applications.

CORDWOOD MODULE.—A method of increasing the number of discrete components in a given space. Resembles wood stacked for a fireplace.

CRYSTAL FURNACE.—Device for artificially growing cylindrical crystals for producing semi-conductor substrates.

DEPOT-LEVEL MAINTENANCE (SM&R Code D)—Supports S&R Code I and SM&R Code O activities through extensive shop facilities and equipment and more highly skilled personnel.

DICE—Uncased chips.

DIE BONDING—Process of mounting a chip to a package.

DIFFUSION—Controlled application of impurity atoms to a semiconductor substrate.

DISCRETE COMPONENTS—Individual transistors, diodes, resistors, capacitors, and inductors.

DOPING—*See* Diffusion.

DUAL IN-LINE PACKAGE (DIP)—IC package having two parallel rows of preformed leads.

ENCAPSULATED—Imbedded in solid material or enclosed in glass or metal.

EPITAXIAL PROCESS—The depositing of a thin uniformly doped crystalline region (layer) on a substrate.

EUTECTIC ALLOY—An alloy that changes directly from a solid to a liquid with no plastic or semiliquid state.

EUTECTIC SOLDER—An alloy of 63 percent tin and 37 percent lead. Melts at 361º F.

FILM ICs—Conductive or nonconductive material deposited on a glass or ceramic substrate. Used for passive circuit components, resistors, and capacitors.

FLAT PACK—IC package.

FLIP CHIP—Monolithic IC packaging technique that eliminates need for bonding wires.

FLUX—Removes surface oxides from metals being soldered.

GENERAL PURPOSE ELECTRONIC TEST EQUIPMENT (GPETE)—Multimeters, oscilloscopes, voltmeters, signal generators, etc.

GROUND PLANES—Copper planes-used to minimize interference between circuits and from external sources.

HYBRID ICs—Two or more integrated circuit types, or one or more integrated circuit types and discrete components on a single substrate.

INTEGRATED CIRCUIT (IC)—Elements inseparably associated and formed on or within a single substrate.

INTERMEDIATE-LEVEL MAINTENANCE (SR&R Code I)—Direct support and technical assistance to user organizations. Tenders and shore-based repair facilities.

ISOLATION—The prevention of unwanted interaction or leakage between components.

LANDS—Conductors or runs on pcbs.

LARGE SCALE INTEGRATION (lsi)—An integrated circuit containing 1,000 to 2,000 logic gates or up to 64,000 bits of memory.

MASK—A device used to deposit materials on a substrate in the desired pattern.

MICROCIRCUIT—A small circuit having high equivalent-circuit-element density, which is considered as a single part composed of interconnected elements on or within a single substrate to perform an electronic-circuit function.

MICROELECTRONICS—That area of electronics technology associated with electronic systems built of extremely small electronic parts or elements.

MICROCIRCUIT MODULE—An assembly of microcircuits or a combination of microcircuits and discrete components that perform one or more distinct functions.

MODIFIED TRANSISTOR OUTLINE (TO)—IC package resembling a transistor.

MODULAR PACKAGING—Circuit assemblies or subassemblies packaged to be easily removed for maintenance or repair.

MODULE—A circuit or portion of a circuit packaged as a removable unit. A separable unit in a packaging scheme displaying regularity of dimensions.

MILITARY STANDARDS (MILSTD)—Standards of performance for components or equipment that must be met to be acceptable for military systems.

MINIATURE ELECTRONICS—Modules, packages, pcbs, and so forth, composed exclusively of discrete components.

2M—Miniature/Microminiature repair program.

MONOLITHIC IC—ICs that are formed completely within a semiconductor substrate. Silicon chips.

OFF-LINE TEST EQUIPMENT—Tests andisolates faults in modules or assemblies removed from systems.

OHMS PER SQUARE—The resistance of any square area of thin film resistive material as measured between two parallel sides.

ON-LINE TEST EQUIPMENT—Continuously monitors the performance of electronic systems.

ORGANIZATIONAL-LEVEL MAINTENANCE (SM&R Code O)—Responsibility of the user organization.

PACKAGING LEVELS—System developed to assist maintenance personnel in isolating faults.

PHOTO ETCHING—Chemical process of removing unwanted material in producing printed circuit boards.

POINT-TO-POINT WIRING—Individual wires run from terminal to terminal to complete a circuit.

PRINTED CIRCUIT BOARD (peb)—The general term for completely processed printed circuit or printed wiring configurations. It includes single-layered, double-layered, and multi-layered boards.

SCREENING—Process of applying nonconductive or semiconductive materials to a substrate to form thick film components.

SHIELDING—Technique designed to minimize internal and external interference.

SOURCE, MAINTENANCE, AND RECOVER-ABILITY CODES (SM&R CODES)—Specify maintenance level for repair of components or assemblies.

SUBSTRATE—Mounting surface for integrated circuits. May be semiconductor or insulator material depending on type of IC.

THICK FILM COMPONENTS—Passive circuit components (resistors and capacitors) having a thickness of 0.001 centimeter.

THIN FILM COMPONENTS—Passive circuit elements (resistors and capacitors) deposited on a substrate to a thickness of 0.0001 centimeter.

VACUUM EVAPORATION—Process of producing thin film components.

VERY LARGE SCALE INTEGRATION (vlsi)—An integrated circuit containing over 2,000 logic gates or 64,000 bits of memory.

WAFER—A slice of semiconductor material upon which monolithic ICs are produced.

APPENDIX II

REFERENCES USED TO DEVELOP THIS NRTC

NOTE: Although the following references were current when this NRTC was published, their continued currency cannot be assured. When consulting these references, keep in mind that they may have been revised to reflect new technology or revised methods, practices, or procedures; therefore, you need to be sure that you are studying the latest references.

CHAPTER 1

Linear Integrated Circuits, Basic Electricity and Electronics Course, Module 34, CANTRAC A-100-0010, Naval Education and Training Program Development Center Detachment, Great Lakes, III., 1981.

Technical Manual, Miniature/Microminiature (2M) Electronic Repair Program, Vols. I, II, and III, NAVSEA TE000-AA-HBK-010/020/030/2M, Naval Sea Systems Command, Keyport, Wash., 1982.

CHAPTER 2

Technical Manual, Miniature/Microminiature (2M) Electronic Repair Program, Vols. I, II, and III, NAVSEA TE000-AA-HBK-010/020/030/2M, Naval Sea Systems Command, Keyport, Wash., 1982.

CHAPTER 3

General Maintenance Handbook, Electronics Installation and Maintenance Books, NAVSEA SE000-00-EIM-160, Naval Sea Systems Command, Washington, D.C., 1981.

Technical Manual, Miniature/Microminiature (2M) Electronic Repair Program, Vols. I, II, and III, NAVSEA TE000-AA-HBK-010/020/030/2M, Naval Sea Systems Command, Keyport, Wash., 1982.

MODULE 14 INDEX

ASSIGNMENT 1

Textbook assignment: Chapter 1, “Microelectronics,” pages 1-1 through 1-56. Chapter 2, “Miniature/Microminiature (2M) Repair Program and High-Reliability Soldering,” pages 2-1 through 2-22.

1-1. What term is used to describe electronic systems that are made up of extremely small parts or elements?

1. Microelectronics
2. Modular packages
3. Integrated circuits
4. Solid-state technology

1-2. During World War II, which of the following limitations were considered unacceptable for military electronics systems?

1. Large size, heavy weight, and wide bandwidth
2. Excessive power requirements, large size, and complex manning requirements
3. Large size, heavy weight, and excessive power requirements
4. Heavy weight, complex circuits and limited communications range

1-3. The development of which of the following types of components had the greatest impact on the technology of microelectronics?

1. Vacuum tubes and resistors
2. Transformers and capacitors
3. Vacuum tubes and transistors
4. Transistors and solid-state diodes

1-4. For a vacuum tube to operate properly in a variety of different circuit applications, additional components are often required to "adjust" circuit values. This is because of which of the following variations within the vacuum tube?

1. Element size
2. Warm-up times
3. Plug-in mountings
4. Output characteristics

1-5. Point-to-point wiring in a vacuum tube circuit often caused which of the following unwanted conditions?

1. Heat interactions
2. Inductive interactions
3. Capacitive interactions
4. Both 2 and 3 above

1-6. Functional blocks of a system that can easily be removed for troubleshooting and repair are called

1. sets
2. chassis
3. modules
4. vacuum tubes

1-7. Which of the following characteristics of a printed circuit board (pcb) is NOT an advantage over a point-to-point wired tube circuit?

1. The pcb weighs less
2. The pcb eliminates the need for point-to-point wiring
3. The pcb eliminates the need for a heavy metal chassis
4. The pcb contains a limited number of components

1-8. A module in which the components are supported by end plates is referred to as

1. a pcb
2. cordwood
3. a substrate
4. encapsulated

1-9. A module which is difficult to repair because it is completely imbedded in solid material is one which has been

1. balanced
2. enveloped
3. integrated
4. encapsulated

1-10. All components and interconnections are formed on or within a single substrate in which of the following units?

1. Cordwood
2. Integrated circuit
3. Equivalent circuit
4. Printed circuit board

1-11. Monolithic integrated circuits are usually referred to as

1. hybrids
2. substrates
3. silicon chips
4. selenium rectifiers

1-12. In integrated circuits, a conductive or nonconductive film is used for which of the following types of components?

1. Capacitors and diodes
2. Transistors and diodes
3. Resistors and capacitors
4. Resistors and transistors

1-13. Which of the following types of electronic circuits is NOT a hybrid integrated circuit?

1. Thick film and transistors
2. Thin film and silicon chips
3. Transistors and vacuum tubes
4. Silicon chips and transistors

1-14. What maximum number of logic gates should be expected in a large-scale integration circuit?

1. 20
2. 200
3. 2,000
4. 20,000

1-15. Integrated circuits containing more than 64,000 bits of memory are referred to as

1. hybrid integration
2. large-scale integration
3. small-scale integration
4. very large-scale integration

1-16. Which of the following pieces of equipment is used to prepare component layout in complex ICs?

1. A mask
2. A camera
3. A computer
4. A microscope

1-17. A device that allows the depositing of material in selected areas of a semiconductor substrate, but not in others, is known as a

1. blind
2. screen
3. filter
4. wafer mask

1-18. Which of the following types of material is preferred for film circuit substrates?

1. Silicon
2. Ceramic
3. Germanium
4. Fiberglass

1-19. A typical silicon wafer has approximately (a) what diameter and (b) what thickness?

1. (a) 2 inches (b) 0.01 to 0.02 inches
2. (a) 2 inches (b) 0.21 to 0.40 inches
3. (a) 3 inches (b) 0.21 to 0.40 inches
4. (a) 3 inches (b) 0.01 to 0.20 inches

1-20. Artificially grown silicon or germanium crystals are used to produce substrates for which of the following types of integrated circuits?

1. Hybrid
2. Thin-film
3. Thick-film
4. Monolithic

1-21. Elements penetrate the semiconductor substrate in (a) what type of IC but (b) do NOT penetrate the substrate in what type of IC?

1. (a) Diffused (b) thin-film
2. (a) Diffused (b) epitaxial
3. (a) Thick-film (b) epitaxial
4. (a) Thick-film (b) thin-film

1-22. Pn junctions are protected from contamination during the fabrication process by which of the following materials?

1. Oxide
2. Silicon
3. Germanium
4. Photoetch

1-23. The prevention of unwanted interaction or leakage between components is accomplished by which of the following techniques?

1. Isolation
2. Insulation
3. Integration
4. Differentiation

1-24. Vacuum evaporation and cathode sputtering are two methods used to produce which of the following types of components?

1. Diodes
2. Thin-film
3. Thick-film
4. Transistors

1-25. To deposit highly reactive materials on a substrate, you should use which of the following methods..

1. Photoetching
2. Photolithography
3. Cathode sputtering
4. Vacuum evaporation

1-26. To produce thin film resistors, you should use which of the following materials?

1. Nichrome
2. Tantalum
3. Titanium
4. Each of the above

1-27. Which of the following is a major advantage of hybrid ICs?

1. Ease of manufacture
2. Ease of replacement
3. Design flexibility
4. Easy availability

1-28. IC packaging is required for which of the following reasons?

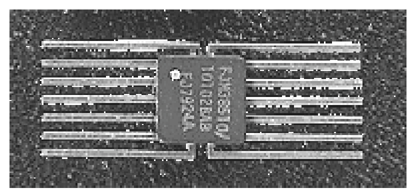
1. To dissipate heat
2. For ease of handling
3. To increase shelf life
4. To meet stowage requirements

IN ANSWERING QUESTIONS 1-29 AND 1-30, MATCH THE IC PACKING EXAMPLES IN THE QUESTIONS TO THE PACKAGING DESCRIPTIONS IN FIGURE 1A.

A.	**TO**
B.	**DIP**
C.	**Flatpack**
D.	**Hybrid**

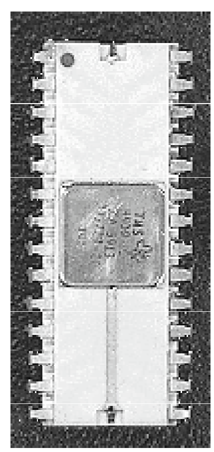
Figure 1A. —Packaging descriptions.

1-29.



1. A
2. B
3. C
4. D

1-30.



1. A
2. B
3. C
4. D

1-31. Which of the following types of DIPs are most commonly used in the Navy's microelectronics systems?

1. Glass
2. Metal
3. Ceramic
4. Plastic

1-32. In IC production, gold or aluminum bonding wires are used for which of the following purposes?

1. To bond the chip to the package
2. To provide component isolation
3. To connect the package to the circuit board
4. To connect the chip to the package leads

1-33. IC packages that may be easily installed by hand or machine on mounting boards fall into which of the following categories?

1. TO
2. DIP
3. Flatpack
4. Each of the above

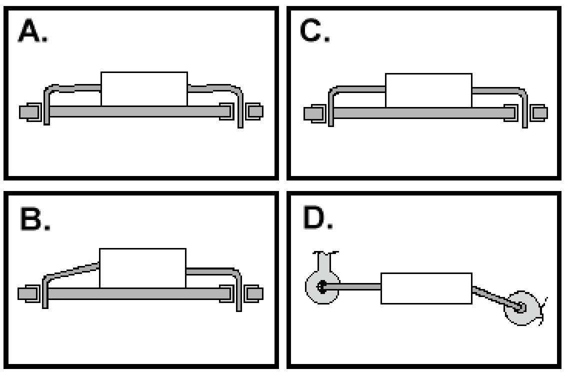
1-34. The need for bonding wires has been eliminated by which of the following production techniques?

1. LSI
2. Beam lead
3. Flip chip
4. Both 2 and 3 above

THIS SPACE LEFT BLANK INTENTIONALLY.

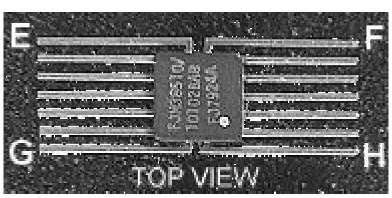
IN ANSWERING QUESTIONS 1-35 THROUGH 1-37, MATCH THE LETTER IN EACH OF THE FIGURES THAT IDENTIFIES PIN 1.

1-35.



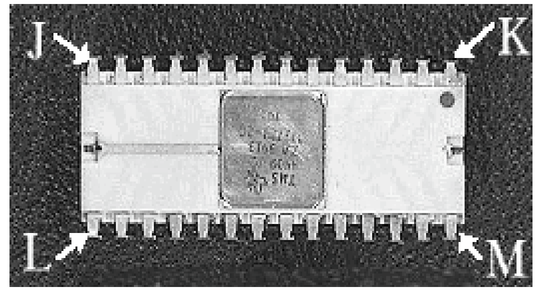
1. A
2. B
3. C
4. D

1-36.



1. E
2. F
3. G
4. H

1-37.



1. J
2. K
3. L
4. M

1-38. Letters and numbers stamped on the body of an IC serve to provide which of the following types of information?

1. Use
2. Serial number
3. Date of manufacture
4. Applicable equipment

1-39. Descriptive information about a particular type of IC may be found in which of the following documents?

1. The manufacturer's data sheet
2. The equipment Allowance Part List (APL)
3. The National Stock Number (NSN)
4. The IC identification number list

1-40. Assemblies made up EXCLUSIVELY of discrete electronic parts are classified as

1. vacuum-tube circuits
2. microcircuit modules
3. hybrid microcircuits
4. miniature electronics circuits

1-41. An assembly of microcircuits or a combination of microcircuits and discrete components is referred to as a

1. mother board
2. microprocessor
3. miniature module
4. microcircuit module

1-42. A technician has isolated a problem to a plug-in module on a printed circuit board. What is this level of system packaging?

1. Level O
2. Level I
3. Level II
4. Level III

1-43. A faulty transistor would be identified as what level of packaging?

1. Level O
2. Level I
3. Level II
4. Level III

1-44. A chassis located in a radar antenna pedestal would be identified as what level of system packaging?

1. Level I
2. Level II
3. Level III
4. Level IV

1-45. Which of the following characteristics is NOT an advantage of multilayer printed circuit boards?

1. Allows greater wiring density on boards
2. Provides shielding for a large number of conductors
3. Eliminates complicated wiring harnesses
4. Reduces the number of components per board

1-46. Which of the following circuit connection methods is NOT used in making interconnections on a multilayer printed circuit board interconnection?

1. Terminal lug
2. Clearance hole
3. Layer build-up
4. Plated-through hole

1-47. The most complex to produce and difficult to repair printed circuit boards are those made using which of the following methods?

1. Layer-buildup
2. Clearance-hole
3. Step-down-hole
4. Plated-through-hole

1-48. Environmental performance requirements for ICs are set forth in which of the following publications?

1. 2M repair manual
2. Military Standards
3. System maintenance manuals
4. Manufacturer's data sheet

1-49. Ground planes and shielding are used to prevent which of the following electrical interactions?

1. Cross talk
2. External interference
3. The generation of rf within the system
4. All of the above

1-50. Training requirements for miniature and microminiature (2M) repair personnel was established by which of the following authorities?

1. Chief of Naval Education and Training
2. Chief of Naval Technical Training
3. Chief of Naval Operations
4. Commander, Naval Sea Systems Command

1-51. The standards of workmanship and guidelines for specific repairs to equipment are contained in which of the following Navy publications?

1. Introduction to Microelectronics
2. NAVSHIPS' Technical Manual
3. Electronics Installation and Maintenance Books (EIMB)
4. Miniature/Microminiature (2M) Electronic Repair Program

1-52. A technician is authorized to perform 2M repairs upon satisfactory completion of which of the following types of training?

1. A 2M training class
2. On-the-job training
3. NEETS, Module 14
4. Any electronics class "A" school

1-53. Repairs that are limited to discrete components and single- and double-sided boards are classified as what level of repairs?

1. Intermediate
2. Organizational
3. Miniature component
4. Microminiature component

1-54. To ensure that a 2M technician maintains the minimum standards of workmanship, the Navy requires that the technician meet which of the following requirements?

1. Be licensed
2. Be certified
3. Be experienced
4. Be retrained

1-55. If a technician should fail to maintain the required standards of workmanship, the technician's certification is subject to what action?

1. Cancellation
2. Recertification
3. Reduction to next lower level
4. Withholding pending requalification

1-56. The most extensive shop facilities and highly skilled technicians are located at what SM & R level of maintenance?

1. Depot
2. Operational
3. Intermediate
4. Organizational

1-57. SM & R code D maintenance facilities are usually located at which of the following activities?

1. Shipyards
2. Contractor maintenance organizations
3. Shore-based facilities
4. All of the above

1-58. Direct support to user organizations is provided by which of the following SM & R code maintenance levels?

1. Depot
2. Operational
3. Intermediate
4. Organizational

1-59. Inspecting, servicing, and adjusting equipment is the function of which of the following SM & R code maintenance levels?

1. Depot
2. Operational
3. Intermediate
4. Organizational

1-60. The maintenance level at which normal 2M repairs are performed is set forth in the maintenance plan and specified by the

1. NAVSEA 2M Repair Program
2. Source, Maintenance, and Recoverability (SM & R) code
3. Chief of Naval Operations
4. Equipment manufacturers' documentation

1-61. Boards or modules that are SM & R code D may be repaired at the organizational level under which of the following conditions?

1. On a routine basis
2. When parts are available
3. To meet an urgent operational commitment
4. When code D repair will take six weeks or longer

1-62. Source, Maintenance, and Recovery (SM & R) codes that list where repair parts may be obtained, who is authorized to make the repair, and the maintenance level for the item are found in which of the following documents?

1. Allowance Equipage Lists (AEL)
2. Allowance Parts Lists (APL)
3. Manufacturer's Parts List
4. Navy Stock System

1-63. Test equipment that continuously monitors performance and automatically isolates faults to removable assemblies is what category of equipment?

1. On-line
2. Off-line
3. General purpose
4. Fault isolating

1-64. A dc voltmeter that is permanently attached to a power supply for the purpose of monitoring the output is an example of what type test equipment?

1. General Purpose Electronic Test Equipment (GPETE)
2. Built in Test Equipment (BITE)
3. Off-line test equipment
4. Specialized test equipment

1-65. Which of the following types of test equipment is classified as off-line automatic test equipment?

1. Centralized Automatic Test System (CATS)
2. Versatile Avionic Shop Test System (VAST)
3. General Purpose Electronic Test Equipment (GPETE)
4. Test Evaluation and Monitoring System (TEAMS)

1-66. Fault diagnosis using GPETE should only be attempted by which of the following personnel?

1. Officers
2. Technician strikers
3. Experienced technicians
4. Basic Electricity and Electronics school graduates

1-67. During fault isolation procedures, a device or component should be desoldered and removed from the circuit only at which of the following times?

1. After defect verification
2. For out-of-circuit testing
3. During static resistance checks
4. At any time the technician desires

1-68. 2M repair stations are equipped according to the types of repairs to be accomplished. The use of microscopes and precisions drill presses would be required in which of the following types of repair?

1. Miniature
2. Microminiature
3. Both 1 and 2 above
4. Emergency

1-69. In the selection of a soldering iron tip, which of the following factors should be considered?

1. The complexity of the pcb
2. The composition of the pcb
3. The area and mass being soldered
4. The type of component being soldered

1-70. The handpiece that can be used for the greatest variety of operations is the

1. solder extractor
2. rotary-drive tool
3. resistive tweezers
4. lap flow and thermal scraper hand tool

1-71. Regardless of location, 2M repair stations require adequate work surface area, lighting, power, and what other minimum requirement?

1. Heat source
2. Ventilation
3. Illumination
4. Dust-free space

1-72. Solder used in electronics is an alloy composed of which of the following metals?

1. Tin and zinc
2. Tin and lead
3. Lead and zinc
4. Lead and copper

1-73. A roll of solder is marked 60/40. What do these numbers indicate?

1. 60% tin, 40% lead
2. 60% tin, 40% copper
3. 60% lead, 40% tin
4. 60% lead, 40% copper

1-74. Which of the following alloys will melt directly into a liquid and have no plastic or semiliquid state?

1. Metallic alloy
2. Eutectic alloy
3. Zinc-lead alloy
4. Copper-zinc alloy

1-75. The PREFERRED solder alloy ratio for electronic repair is 63/37. Which of the following alloy ratios is also ACCEPTABLE for this type of repair?

1. 30/70
2. 50/50
3. 60/40
4. 70/30

ASSIGNMENT 2

Textbook assignment: Chapter 3, "Miniature and Microminiature Repair Procedures," pages 3-1 through 3-51.

2-1. Which of the following requirements must be met by a 2M technician to be authorized to perform repairs at a particular level?

1. Have knowledge of that level
2. Be certified at that level
3. Be experienced at that level
4. Be licensed at the next higher level

2-2. Protective materials applied to electronic assemblies to prevent damage caused by corrosion, moisture, and stress are called

1. conformal coatings
2. isolation materials
3. electrical insulation
4. encapsulation coatings

2-3. Before working on a pcb, the conformal coating should be removed from what part of the board?

1. The entire board
2. The component side
3. The area of the repair
4. The side opposite the component side

2-4. What are the approved methods of conformal coating removal?

1. Peeling and abrading
2. Stripping and heating
3. Mechanical and thermal only
4. Mechanical, thermal, and chemical

2-5. Most methods of conformal coating removal are variations of which of the following types of removal?

1. Thermal
2. Chemical
3. Mechanical
4. Electrical

2-6. What is the preferred method of removing epoxy conformal coatings?

1. Solvent
2. Ball mill
3. Hot-air jet
4. Thermal parting tool

2-7. A conformal coating is considered to be thin if it is less than what thickness?

1. 0.025 inches
2. 0.040 inches
3. 0.050 inches
4. 0.250 inches

2-8. Of the various mechanical methods of conformal coating removal, physical contact with the work piece is NOT required using which of the following methods?

1. Cutting
2. Grinding
3. Hot-air jet
4. Thermal parting

2-9. Cutting and peeling is an easy method of removing which of the following types of coatings?

1. Epox
2. Varnish
3. Parylene
4. Silicone

2-10. Thin acrylic coatings are readily removed in which of the following ways?

1. Routing
2. Hot-air jet
3. Cut and peel
4. Mild solvents

2-11. When you are applying an application of conformal coating, which of the following conditions is true?

1. The board should be clean and moist
2. The coating should be applied only to the component replaced
3. The coating should be the same type as that used by the manufacturer
4. The coating should be applied only to the solder joints

2-12. The procedure of connecting circuitry on one side of a board with the circuitry on the other side is known as

1. mounting
2. termination
3. hole reinforcement
4. interfacial connections

2-13. Reinforcement for circuit pads on both sides of the board is provided by which of the following types of eyelets?

1. Flat-set
2. Roll-set
3. Funnel-set

2-14. The manner in which wires and leads are attached to an assembly is described by which of the following terms?

1. Termination
2. Solder joints
3. Lead formation
4. Interfacial connections

2-15. Clinched leads, straight-through leads, and offset pads are variations of what type of termination?

1. Solder cup
2. On-the-board
3. Above-the-board
4. Through-hole

2-16. What total number of degrees of bend are (a) fully clinched leads and (b) semiclinched leads?

1. (a) 45 (b) 90
2. (a) 45 (b) 45
3. (a) 90 (b) 45
4. (a) 90 (b) 90

2-17. Which of the following types of terminals are used as tie points for interconnecting wiring?

1. Pins
2. Hooks
3. Solder cups
4. Turrets

2-18. During assembly, component stability is provided by which of the following types of lead termination?

1. Hook terminal
2. Clinched lead
3. Turret terminal
4. On-the-board (lap-flow)

2-19. Which of the following types of lead terminations is the easiest to remove and rework?

1. Offset-pad
2. Semiclinched
3. Fully clinched
4. Straight-through

2-20. Turret, fork, and hook terminals are examples of what type of termination?

1. Off-set
2. On-the-board
3. Above-the-board
4. Through-the-board

2-21. When a lead is soldered to a pad without passing it through the board, what type of termination has been made?

1. Lap-flow
2. Off-set pad
3. Clinched lead
4. Straight-through lead

2-22. Most damage to printed circuit boards occurs at which of the following times?

1. During troubleshooting
2. During component removal
3. During component replacement
4. During normal system operations

2-23. Of the solder removal methods listed below, which one is most versatile and reliable?

1. Wicking
2. Heat and shake
3. Motorized solder extractor
4. Manually controlled vacuum plunge

2-24. When you are removing solder with a solder wick, where should the wick be placed in relation to the solder joint and the iron?

1. Below both the joint and the iron
2. Between the joint and the iron
3. Above the joint and the iron

2-25. The motorized solder extractor may be operated in three different modes. Which of the following is NOT one of those modes?

1. Hot-air jet
2. One-shot vacuum
3. Heat and vacuum
4. Heat and pressure

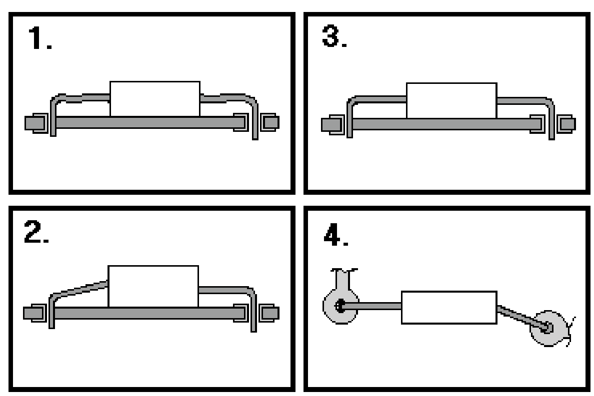
2-26. Stirring the lead during desoldering prevents which of the following unwanted results?

1. Sweat joints
2. Overheating
3. Cold solder joints
4. Pad delamination

2-27. Of the following solder removal methods, which one is acceptable for removing solder?

1. Hot-air jet
2. Heat and pull
3. Heat and shake
4. Heat and squeeze

2-28. Which of the following examples represents properly formed leads?



2-29. A 2M technician is repairing a board that is manufactured with semiclinched leads. What type of termination should the technician use in replacing components?

1. Lap flow
2. Clinched lead
3. Semiclinched lead
4. Straight-through-lead

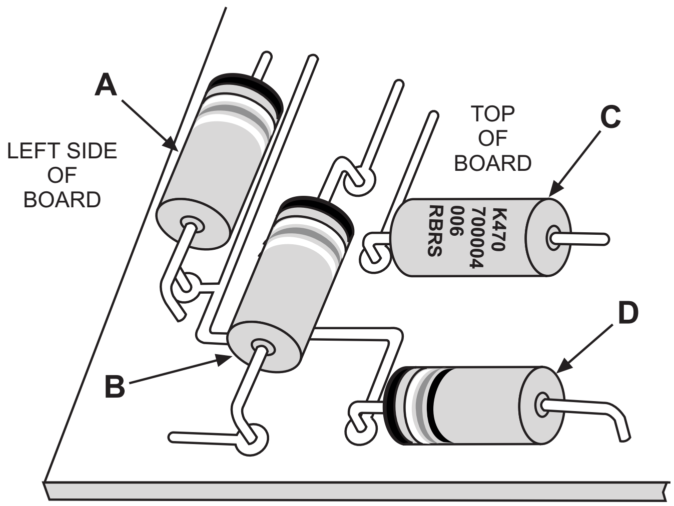


Figure 2A.—Component mounting.

IN ANSWERING QUESTION 2-30, REFER TO FIGURE 2A.

2-30. What component on the board is NOT properly mounted?

1. A
2. B
3. C
4. D

2-31. What type of heat source is a soldering iron?

1. Radiant
2. Resistive
3. Conductive
4. Convective

2-32. The shape and size of the soldering iron tip to be used is determined by which of the following factors?

1. The type of flux to be used
2. The type of work to be done
3. The type of solder to be used
4. The voltage source for the iron

2-33. A thermal shunt is attached to the leads of a transistor prior to applying solder to the joint. This shunt serves what purpose?

1. Prevents short circuits
2. Retains heat at the joint
3. Conducts heat away from the component
4. Physically supports the lead during soldering

2-34. What is the appearance of a good solder joint?

1. Dull gray
2. Crystalline
3. Bright and shiny
4. Shiny with small pits

2-35. What is the most efficient soldering temperature?

1. 360°F
2. 440°F
3. 550°F
4. 800°F

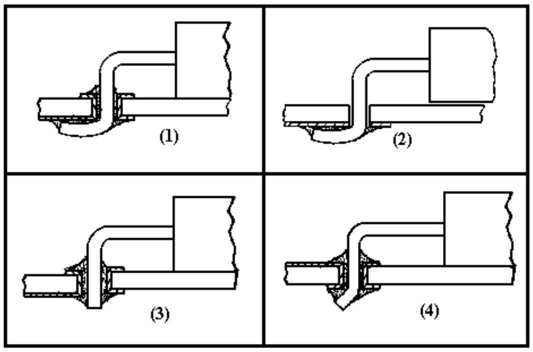


Figure 2B. —Solder joints.

IN ANSWERING QUESTION 2-36, REFER TO FIGURE 2B.

2-36. All the solder joints in the figure have two things in common. The first is that all are through-the-board terminations. The other is that they are all what type of joint?

1. Sweat
2. Full-fillet
3. Unacceptable
4. Clinched-lead

2-37. Plug-in DIPs are mounted by using which of the following aids/parts?

1. Insulators
2. Special tools
3. Clinched leads
4. Mounting sockets

2-38. Plug-in DIPs are susceptible to loosening because of which of the following causes?

1. Heat
2. Stress
3. Warpage
4. Vibration

2-39. Component leads may be clipped to aid in their removal under which of the following conditions?

1. When the component is conformally coated
2. When the component is known to be defective
3. When board damage may result from normal removal methods
4. Both 2 and 3 above

2-40. Visual inspection of a completed repair is conducted to evaluate which of the following aspects of the repair?

1. Workmanship quality
2. Component placement
3. In-circuit quality test
4. Conformal coating integrity

2-41. When speaking of TO mounting techniques, the term "plug-in" refers to the same technique as is used with DIPs.

1. True
2. False

2-42. In addition to heat dissipation and physical support, which of the following needs might justify the use of a spacer with a TO mount?

1. Vibration elimination
2. Proper lead formation
3. Proper lead termination
4. Short-circuit protection

2-43. For the removal of imbedded TOs in which all the leads are free, which of the following methods is recommended?

1. Push out gently
2. Pull out with pliers
3. Pull out with fingers
4. Tap out with soft mallet

2-44. What is the most critical step in replacing an imbedded TO?

1. Seating the leads
2. Forming the leads
3. Soldering the leads
4. Seating the body of the TO

2-45. When a TO or a DIP is replaced on a printed circuit board, what type of termination is normally used?

1. Lap flow
2. On-the-board
3. Above-the-board
4. Through-the-board

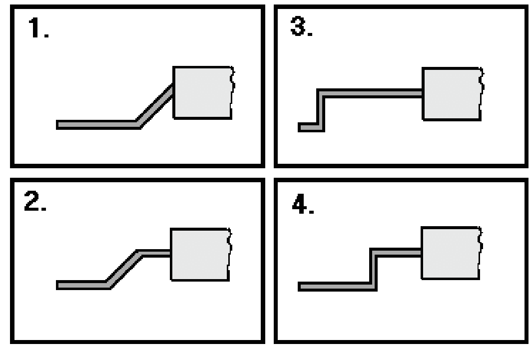
2-46. What type of termination is used in the replacement of a flat pack?

1. Lap flow
2. Solder cup
3. Solder plug
4. Full fillet

2-47. Heating the leads and lifting them free with tweezers is the preferred method of removing which of the following components?

1. TOs
2. DIPs
3. Flat packs
4. Transistors

2-48. Which of the examples shows the correct lead formation for a flat pack?



2-49. Use of a skipping pattern when soldering multilead components prevents

1. cold solder joints
2. excessive heat buildup
3. the component from moving
4. the need to visually inspect the piece

2-50. Cards and boards may be damaged under which of the following conditions?

1. When unauthorized repairs are attempted by untrained personnel
2. When technicians use improper tools
3. When improperly stored
4. Each of the above

2-51. DS3 Spark is preparing to repair a cracked conductor on a card. For this type of damage, which of the following repair methods is preferred?

1. Solder bridge
2. Clinched staple
3. Install an eyelet at the crack and solder in place
4. Lap-flow soldered wire across the break

2-52. To ensure a good mechanical bond between the board and replacement pad and to provide good electrical contact for components, you should use which of the following procedures?

1. Epoxying the pad to the run
2. Electroplating the repair area
3. Installing an eyelet in the pad
4. Lap-flow soldering the repair area

2-53. Breaks, holes, and cracks in pcbs are repaired by using a mixture of

1. fiberglass and rosin
2. epoxy and powdered carbon
3. conformal coating and RTV
4. epoxy and powdered fiberglass

2-54. What is the first step in the repair of burned or scorched boards?

1. Filling the burned area with epoxy and fiberglass
2. Removing all discolored material
3. Removing all delaminated conductors
4. Cleaning all surfaces with solvent

2-55. Which of the following statements is correct concerning the repair of repairable delaminated conductors?

1. All delaminations are removed
2. Repairable delaminations are not removed
3. All delaminations are epoxied to the board
4. All delaminations are replaced with insulated wire

2-56. Damage to some electronic components can occur at what minimum electrostatic potential?

1. 14 volts
2. 35 volts
3. 350 volts
4. 3,500 volts

2-57. Electrostatic discharge (ESD) has the greatest effect on which of the following devices?

1. Silicon diodes
2. Selenium rectifiers
3. Germanium transistors
4. Metal-oxide transistor

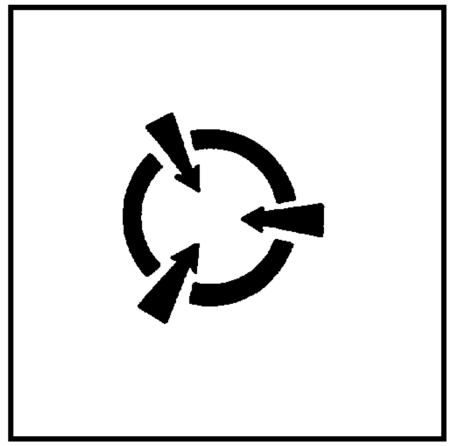


Figure 2C. —Symbol.

IN ANSWERING QUESTION 2-58, REFER TO FIGURE 2C.

2-58. The symbol shown in the figure is found on a component package. What does it indicate about the component?

1. It is a high cost item
2. It is a bipolar device
3. It is an integrated circuit
4. It is electrostatic discharge sensitive

2-59. Electrostatic charges may develop as high as which of the following voltages?

1. 3,000 volts
2. 15,000 volts
3. 20,000 volts
4. 35,000 volts

2-60. To prevent an electrostatic charge built up on the body of the technician from damaging ESDS devices, the technician should take which of the following precautions?

1. Be grounded
2. Wear gloves
3. Ground the device
4. Handle the device with insulated tools

2-61. The best source of information concerning the application, handling, and storage of any aerosol dispensers may be found in/on which of the following sources?

1. NAVSEA 2M manual
2. NEETS, Module 14, Topic 3
3. On the aerosol dispenser
4. Applicable military standards

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 15—Principles of Synchros, Servos, and Gyros

NAVEDTRA 14187

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 15 of a series.

History of the course:

Sep 1998: Original course released.
Jul 2003: Administrative update release. No significant content changes were made.

NAVSUP Logistics Tracking Number
0504-LP-026-8400

TABLE OF CONTENTS

CHAPTER 1

SYNCHROS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions placed throughout the chapters are based on the objectives. By successfully completing the Nonresident Training Course (NRTC), you indicate that you have met the objectives and have learned the information. The learning objectives for this chapter are listed below.

Upon completing this chapter, you will be able to:

1. Define the term "synchro."
2. State the primary purpose of a synchro.
3. Explain the importance of synchros in naval equipment.
4. Name the two general classifications of synchros.
5. Explain the differences between torque and control synchros.
6. Name the seven functional classes of synchros and list all inputs and outputs.
7. Name the two types of synchro identification codes.
8. Interpret all synchro markings and identify the particular codes used.
9. Draw the five standard schematic symbols for synchros and identify all connections.
10. Describe the general construction and physical appearance of synchro rotors and stators.
11. Name the two common types of synchro rotors, giving an application of each.
12. List the different synchro characteristics and give a brief explanation of each.
13. State the advantage of using 400-Hz synchros over 60-Hz synchros.
14. Explain the operation of a basic synchro transmitter and receiver.
15. State the difference between a synchro transmitter and a synchro receiver.
16. List the basic components that compose a torque synchro system.
17. Explain the operation of a simple synchro transmission system.
18. Define the term "correspondence" and explain how it is used in a simple synchro system.
19. Explain the principle behind reversing the S1 and S3 leads on a synchro receiver and how this action affects receiver operation.

20. Explain what happens when the rotor leads on a synchro transmitter or receiver are reversed.
21. State the purposes of differential synchros.
22. Name the two types of differential synchros and give a brief explanation of each.
23. Explain the difference between the torque differential transmitter and the torque differential receiver.
24. Name the components that make up the TDX and the TDR synchro systems.
25. Explain how the two differential synchro systems add and subtract.
26. State the wiring changes required to convert the differential synchro systems from subtraction to addition.
27. State the purposes and functions of control synchros.
28. Name the different types of control synchros.
29. Explain how the CX and CDX differ from the TX and TDX.
30. Explain the theory and operation of a control transformer.
31. List the basic components that compose a control synchro system.
32. Explain the operation of a control synchro system and how it is used to control a servo system.
33. State the purpose and function of the synchro capacitor.
34. Explain how synchro capacitors improve the accuracy of synchro systems.
35. Explain the method used to connect synchro capacitors in a circuit.
36. Define single and multispeed synchro systems.
37. State the purposes and functions of multispeed synchro systems.
38. Stale the purposes for zeroing synchros.
39. Name three common synchro zeroing methods and give a brief explanation of each.
40. Explain the different troubleshooting techniques used in isolating synchro malfunctions and breakdowns.

SYNCHRO FUNDAMENTALS

Synchros play a very important role in the operation of Navy equipment. Synchros are found in just about every weapon system, communication system, underwater detection system, and navigation system used in the Navy. The importance of synchros is sometimes taken lightly because of their low failure rate. However, the technician who understands the theory of operation and the alignment procedures for synchros is well ahead of the problem when a malfunction does occur. The term "synchro" is an abbreviation of the word "synchronous." It is the name given to a variety of rotary, electromechanical, position-sensing devices. Figure 1-1 shows a phantom view of typical synchro. A synchro resembles a small electrical motor in size and appearance and operates like a variable transformer. The synchro, like the transformer, uses the principle of electromagnetic induction.

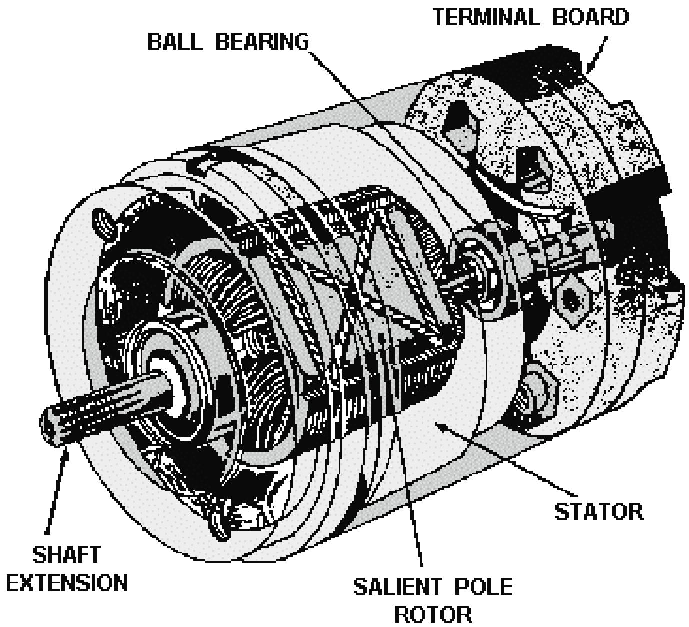


Figure 1-1.—Phantom view of a synchro.

Synchros are used primarily for the rapid and accurate transmission of information between equipment and stations. Examples of such information are changes in course, speed, and range of targets or missiles; angular displacement (position) of the ship's rudder; and changes in the speed and depth of torpedoes. This information must be transmitted quickly and accurately. Synchros can provide this speed and accuracy. They are reliable, adaptable, and compact. Figure 1-2 shows a simple synchro system that can be used to transmit different as of data or information In this system, a single synchro transmitter furnishes information to two synchro receivers located in distant spaces. Operators put information into the system by turning the handwheel. As the handwheel turns, its attached gear rotates the transmitter shaft (which has a dial attached to indicate the value of the transmitted information). As the synchro transmitter shaft turns, it converts the mechanical input into an electrical signal, which is sent through interconnecting wiring to the two synchro receivers. The receiver shafts rotate in response to the electrical signal from the transmitter. When these shafts turn, the dials attached to the shafts indicate the transmitted information.

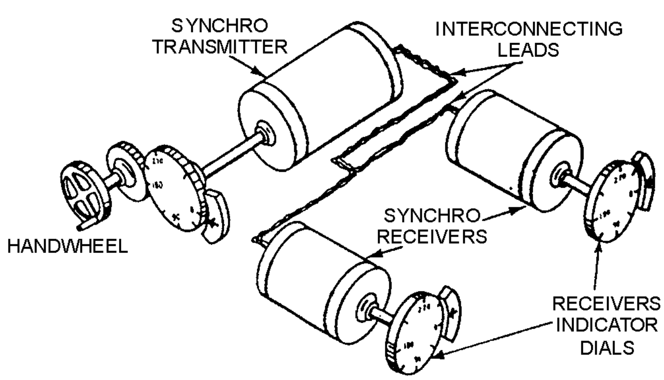


Figure 1-2.—Data transfer with synchros.

By studying the simple synchro system, you can see that information can be transmitted over long distances, from space to space, and from equipment to equipment.

In addition to supplying data by positioning dials and pointers, synchros are also used as control devices in servo systems. When the synchro and the servo are combined, they work as a team to move and position heavy loads. The methods used to accomplish this are covered in detail in the next chapter.

Q-1. What is the name given to a variety of rotary electromechanical, position sensing devices?

Q-2. What is the primary purpose of a synchro system?

SYNCHRO CLASSIFICATION

Synchros work in teams. Two or more synchros interconnected electrically form a synchro system. There are two general classifications of synchro systems—TORQUE SYSTEMS AND CONTROL SYSTEMS. Torque-synchro systems use torque synchros and control-synchro systems use control synchros. The load dictates the type of synchro system, and thus the type of synchro.

Torque-synchro systems are classified "torque" because they are mainly concerned with the torque or turning force required to move light loads such as dials, pointers, or similar indicators. The positioning of these devices requires a relatively low amount of torque. Control synchros are used in systems that are designed to move heavy loads such as gun directors, radar antennas, and missile launchers.

In addition to the two general classifications, synchros are grouped into seven basic functional classes as shown in table 1-1. Four of these are the torque type and three are the control type. Each synchro is described in the table by name, abbreviation, input, output, and the other synchro units that may be connected to it. Generally, torque and control synchros may not be interchanged. The functional operation of each of these seven synchros is covered later in this text.

Table 1-1.—Synchro Information

FUNCTIONAL CLASSIFICATION	ABBREVIATION	INPUT	OUTPUT
Torque transmitter	TX	Mechanical input to rotor (rotor energized from AC source)	Electrical output from stator representing angular position of rotor to TDX, TDR, or TR.
Control transmitter	CX	Same as TX	Same as TX except it is supplied to CDX or CT
Torque differential transmitter	TDX	Mechanical input to rotor, electrical input to stator from TX or another TDX.	Electric output from rotor representing algebraic sum or difference between rotor angle and angle represented by electrical input to TR, TDR, or another TDX.
Control differential transmitter	CDX	Same as TDX except electrical input is from CX or another CDX.	Same as TDX except output to CT or another CDX.
Torque receiver	TR	Electrical input to stator from TX or TDX. (Rotor energized from AC source)	Mechanical output from rotor. Note: Rotor has mechanical inertia damper.
Torque differential receiver	TDR	Electrical input to stator from TX or TDX, another electrical input to rotor from TX or TDX.	Mechanical output from rotor representing algebraic sum or difference between angles represented by electrical inputs. Has inertia damper.
Control transformer	CT	Electric input to stator from CX or CDX, mechanical input to rotor.	Electrical output from rotor proportional to the sine of the angle between rotor position and angle represented by electrical input to stator. Called error signal.
Torque receiver	TRX	Depending on application, same as TX.	Depending on application, same as TX or TR.

Synchros are also classified according to their operating frequency. This classification was brought about by the development of the 400-Hz synchro. Prior to this time, the 60-Hz synchro was the only one in use. Synchro operating frequencies are covered in detail in the section on synchro characteristics.

Q-3. Name the two general classifications of synchro systems.

Q-4. What is the difference between a torque synchro and a control synchro?

Q-5. Using table 1-1, name two synchros that provide a mechanical output.

STANDARD MARKINGS AND SYMBOLS

Synchros used in the Navy can be grouped into two broad categories: MILITARY STANDARD SYNCHROS and PRESTANDARD NAVY SYNCHROS. Military standard synchros conform to specifications that are uniform throughout the armed services. New varieties of equipment use synchros of this type. Prestandard synchros were designed to meet Navy, rather than servicewide, specifications. Each category has its own designation code for identification.

Military Standard Synchro Code

The military standard designation code identifies standard synchros by their physical size, functional purpose, and supply voltage characteristics. The code is alphanumerical and is broken down in the following manner. The first two digits indicate the diameter of the synchro in tenths of an inch, to the next higher tenth. For example, a synchro with a diameter of 1.75 inches has the numeral 18 as its first two digits. The first letter indicates the general function of the synchro and of the synchro system-C for control or T for torque. The next letter indicates the specific function of the synchro, as follows:

LETTER	DEFINITION
D	Differential
R	Receiver
T	Transformer
X	Transmitter

If the letter B follows the specific function designation, the synchro has a rotatable stator. The last number in the designation indicates the operating frequency-6 for 60 Hz and 4 for 400 Hz. The upper-case letter following the frequency indicator is the modification designation. The letter "A" indicates that the synchro design is original. The first modification is indicated by the letter "B." Succeeding modifications are indicated by the letters "C," "D," and so on, except for the unused letters "I," "L," "O," and "Q."

For example, an 18TR6A synchro is an original design, 60-Hz torque receiver with a diameter of between 1.71 and 1.80 inches.

A synchro designated 16CTB4B is the first modification of a 400-Hz control transformer with a rotatable stator and a diameter of between 1.51 and 1.60 inches.

All standard synchros are labeled with such a code. Synchros used in circuits supplied by 26 volts are classified in the same way, except that the symbol 26V is prefixed to the designator (for example, 26V-16CTB4A). Otherwise, a 115 volts source is assumed for the synchro system.

Navy Prestandard Synchro Code

The Navy prestandard designation code identifies prestandard synchros by size and function, using a number and letter combination. Unlike the standard code, the number does not indicate directly the diameter of the synchro. The number merely represents the approximate size of the synchro, increasing as the size increases. The approximate size and weight of the five most common sizes are shown in the following table.

SIZE	APPROX. DIAMETER	APPROX. LENGTH	APPROX WEIGHT
1	2 1/4 in	4 in	2 lb
3	3 1/10 in	5 3/8 in	3 lb
5	3 3/8 to 3 5/8 in	6 1/2 in	5 lb
6	4 1/2 in	7 in	8 lb
7	5 3/4 in	9 in	18 lb

Note that prestandard size 1 is approximately the same size as standard size 23 (2.21 to 2.30 inches in diameter). Prestandard size 3 is approximately the same size as standard size 31. Prestandard size 5 is approximately the same size as standard size 37.

The letters used in the prestandard coding system indicate the function, mounting, or special characteristics of the synchro as shown in the following chart.

LETTER	DEFINITION
G	Transmitter
F	Flange Mounted Receiver (this letter is normally omitted if letters other than H or S occur in type designation)
D	Differential Receiver
DG	Differential Transmitter
CT	Control Transformer
H	High-Speed Unit
B	Bearing Mounted Unit
N	Nozzle Mounted Unit
S	Special Unit

Navy prestandard synchros are rarely used today. They have been replaced by the standard synchro. However, by being familiar with the prestandard coding system, you will be able to identify the older synchros and make correct replacements if necessary.

Q-6. What does the code 26V-11TX4D mean on a synchro nameplate?

Q-7. Which of the two synchro designation codes is indicated by 5DG on a synchro nameplate?

Schematic Symbols

Schematic symbols for synchros are drawn by various manufacturers in many different ways. Only five symbols (as shown in figure 1-3), however, meet the standard military specifications for schematic diagrams of synchros and synchro connections. When a symbol is used on a schematic, it will be accompanied by the military abbreviation of one of the eight synchro functional classifications (TR, TX, TDX, etc.).

The symbols shown in views A and B of figure 1-3 are used when it is necessary to show only the external connections to a synchro, while those shown in views C, D, and E are used when it is important to see the positional relationship between the rotor and stator. The letters R and S, in conjunction with an Arabic number, are used to identify the rotor and stator connections; for example, R1, R2, S1, S2, and S3. The small arrow on the rotor symbol indicates the angular displacement of the rotor; in figure 1-3 the displacement is zero degrees.

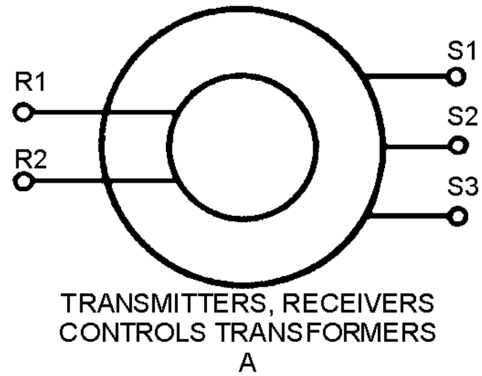


Figure 1-3A.—Schematic symbols for synchros.

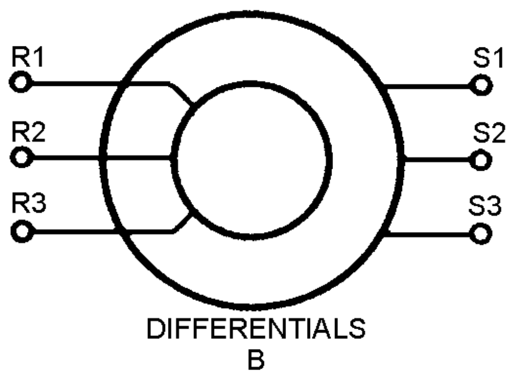


Figure 1-3B.—Schematic symbols for synchros.

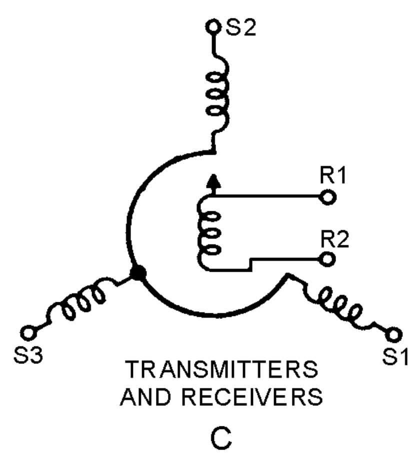


Figure 1-3C.—Schematic symbols for synchros.

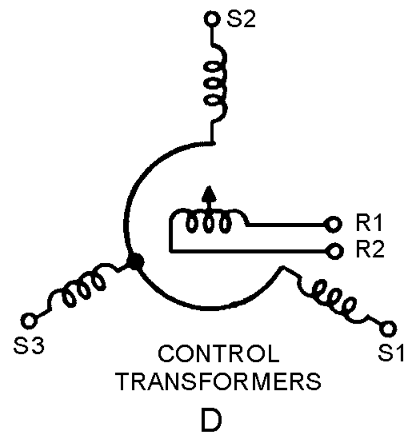


Figure 1-3D.—Schematic symbols for synchros.

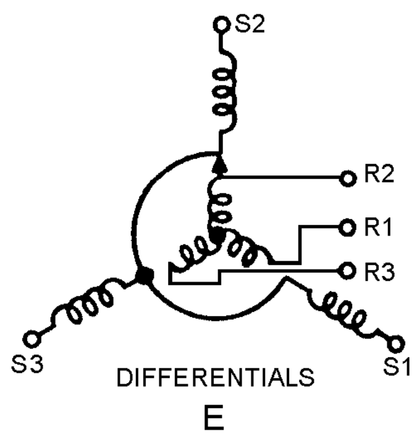


Figure 1-3E.—Schematic symbols for synchros.

Q8. On the synchro schematic symbol, what indicates the angular displacement of the rotor?

SYNCHRO CONSTRUCTION

Figure 1-4 shows a cutaway view of a typical synchro. Having the knowledge of how a synchro is constructed should enable you to better understand how synchros operate.

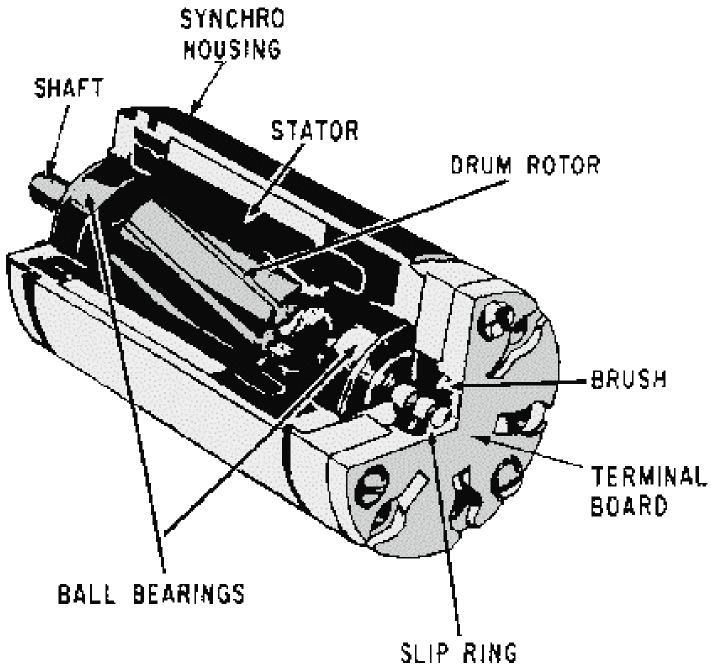


Figure 1-4.—Typical synchro assembly.

In this section we will discuss how rotors and stators are constructed and how the synchro is assembled. Each synchro contains a rotor, similar in appearance to the armature in a motor, and a stator, which corresponds to the field in a motor. The synchro stator is composed of three Y-connected windings (S1, S2, and S3). The rotor is composed of one single winding (R1 and R2). As you can see in the figure, the rotor winding is free to turn inside the stator. The rotor is usually the primary winding and receives its voltage (excitation) from an external voltage source. The stator receives its voltage from the rotor by magnetic coupling.

ROTOR CONSTRUCTION

There are two common types of synchro rotors in use-the SALIENT-POLE ROTOR and the DRUM or WOUND ROTOR. The salient-pole rotor shown in figure 1-5 has a single coil wound on a laminated core. The core is shaped like a "dumb-bell" or the letter "H."

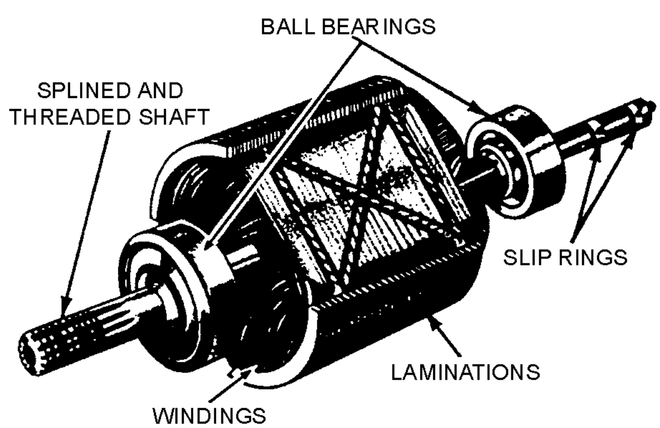


Figure 1-5.—Salient-pole rotor.

This type of winding is frequently used in both transmitters and receivers.

The drum or wound rotor has coils wound in slots in a laminated core as shown in figure 1-6. This type of rotor is used in most synchro control transformers and differential units, and occasionally in torque transmitters. It may be wound continuously with a single length of wire or may have a group of coils connected in series. The single continuous winding provides a distributed winding effect for use in transmitters. When the rotor is wound with a group of coils connected in series, a concentrated winding effect is provided for use in control transformers. When used in differential units, the rotor is wound with three coils so their magnetic axes are 120º apart.

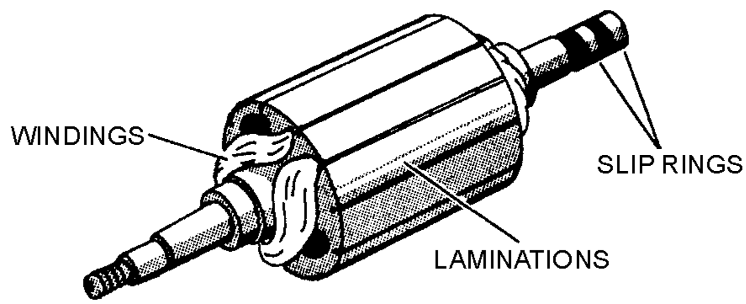


Figure 1-6.—Drum or wound rotor.

Both types of synchro rotors have their coils wound on laminated cores that are rigidly mounted on a shaft. To enable the excitation voltage to be applied to the rotor winding, two slip rings are mounted on one end of the shaft and insulated from the shaft to prevent shorting. An insulated terminal board, mounted on the end of the cylindrical frame, houses the brushes, which ride on the slip rings. These brushes provide continuous electrical contact to the rotor during its rotation. Also mounted on the rotor shaft are low-friction ball bearings, which permit the rotor to turn easily.

STATOR CONSTRUCTION

The stator of a synchro is a cylindrical structure of slotted laminations on which three Y-connected coils are wound with their axes 120º apart. In figure 1-7, view A shows a typical stator assembly consisting of the laminated stator, stator windings, and cylindrical frame; view B shows the stator lamination and the slots in which the windings are placed. Some synchros are constructed so both the stator and the rotor may be turned. Electrical connections to this type of stator are made through slip rings and brushes.

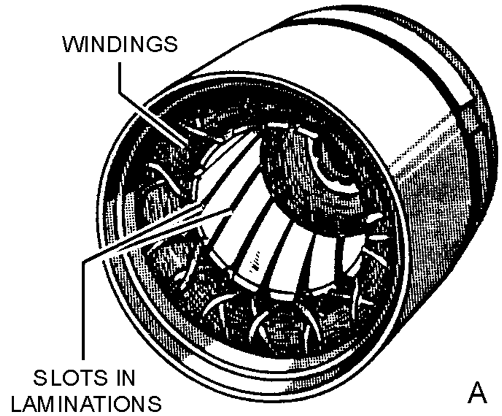


Figure 1-7A.—Typical stator.

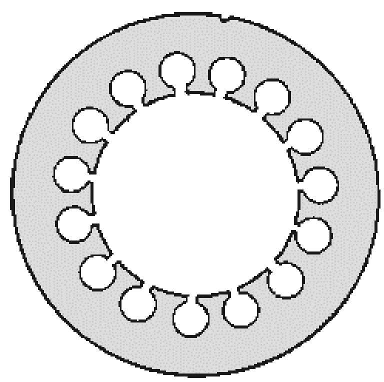


Figure 1-7B.—Stator lamination.

Now, refer to figure 1-4 for a view of a completed synchro assembly. The rotor has been placed in the stator assembly, and a terminal board has been added to provide a point at which internal and external connections can be made.

Q-9. What are the two major components of a synchro?

Q-10. Which of the two main types of rotors can have either a single winding or three Y-connected windings?

Q-11. How does the stator receive its voltage?

Q-12. Where are the external connections made on standard synchros?

SYNCHRO CHARACTERISTICS

Synchro characteristics play a very important part in synchro troubleshooting and maintenance. By closely observing these characteristics, you can generally tell if a synchro or synchro system is working properly. Low torque, overheating, and improper operating voltages are just a few of the abnormal characteristics found in synchro systems. In general, the load capacity of a synchro system is limited by the number and types of receiver units loading the transmitter, the loads on these receiver units, and the operating temperature.

TORQUE

Torque is simply a measure of how much load a machine can turn. In torque synchros, only small loads are turned; therefore, only a small amount of torque is required. The measure of torque is the product of the applied force and the distance between the point of application and the center of rotation. For instance, if a 3 ounce weight is suspended from a synchro pulley having a radius of 2 inches, the torque required to move the weight is 6 ounce-inches. In heavy machinery, torque may be expressed in pound-feet, but torque synchro measurements are in ounce-inches.

NOTE: The unit of torque is the pound-foot or ounce-inch. Do not confuse this with foot-pounds, which is the measurement of work. Many times in referring to torque, tools are marked in foot-pounds. While this use of foot-pounds is technically incorrect, common usage has made it acceptable.

The torque developed in a synchro receiver results from the tendency of two electromagnets to align themselves. Since the rotor can be turned and the stator usually cannot, the stator must exert a force (torque) tending to pull the rotor into a position where the primary and secondary magnetic fields are in line. The strength of the magnetic field produced by the stator determines the torque. The field strength depends on the current through the stator coils. As the current through the stator is increased, the field strength increases and more torque is developed.

Q-13. What major factors determine the load capacity of a torque-synchro transmitter?

Q-14. Define the term "torque."

Q-15. What unit of measurement refers to the torque of a synchro transmitter?

OPERATING VOLTAGES AND FREQUENCIES

Military standard and Navy prestandard synchros are designed to operate on either 115 volts or 26 volts. Synchros used in shipboard equipment are designed predominately for 115 volts, while most aircraft synchros operate on 26 volts.

Synchros are also designed to operate on a 60- or 400-Hz frequency. But like transformers, they are more efficient at the higher frequency. Operating a synchro at a higher frequency also permits it to be made physically smaller. This is because the lines of flux produced by the 400-Hz excitation voltage are much more concentrated than those produced by the 60-Hz excitation voltage. Hence, the core of the 400-Hz synchro can be made smaller than the core of the 60-Hz synchro. However, some 400-Hz synchro units are identical in size to their 60-Hz counterparts. This is done so that 60- and 400-Hz units can be physically interchanged without special mounting provisions. The operating voltage and frequency of each synchro is marked on its nameplate.

The use of the smaller size synchro permits the construction of smaller and more compact equipment. The most widely used frequency for airborne equipment is 400 Hz. It is being used increasingly in shipboard equipment as well. The newer gun and missile fire-control systems use 400-Hz synchros almost exclusively.

A synchro designed for 60-Hz operation may occasionally be used with a 400-Hz supply. There may be considerable loss of accuracy, but the synchro will not be damaged. This should be done only in the case of an emergency when the specified replacement is not available, and system accuracy is not critical.

CAUTION

NEVER connect a 400-Hz synchro to 60-Hz voltage. The reduced impedance results in excessive current flow and the windings quickly burn out.

Q-16. What type of equipment normally uses 26-volt 400-hertz synchros?

OPERATING TEMPERATURES AND SPEEDS

Standard synchros are designed to withstand surrounding temperatures ranging from −67º F to +257º F (-55º C to +125º C) at the terminal board. Prestandard synchros operate in a range of −13º F to +185º F (-25º C to +85º C). When a synchro is energized and not loaded, its temperature should stay within prescribed limits. Loading an energized synchro causes it to generate more heat. Similarly, overloading causes a synchro to generate much more heat than it would under normal loading conditions and could possibly result in permanent synchro damage. To meet military specifications, all standard synchros must be capable of continuous operation for 1,000 hours at 1,200 revolutions per minute (rpm) without a load.

A prestandard synchro has one of two specifications, depending upon its use in a data transmission system. Low-speed prestandard synchros must be capable of continuous operation for 500 hours at 300 rpm without a load. Low-speed prestandard synchros must be capable of continuous operation for 1,500 hours at 1200 rpm without a load.

Q-17. When will a synchro generate more heat than it is designed to handle?

THEORY OF OPERATION

Synchros, as stated earlier, are simply variable transformers. They differ from conventional transformers by having one primary winding (the rotor), which may be rotated through 360º and three stationary secondary windings (the stator) spaced 120º apart. It follows that the magnetic field within the synchro may also be rotated through 360º. If an iron bar or an electromagnet were placed in this field and allowed to turn freely, it would always tend to line up in the direction of the magnetic field. This is the basic principle underlying all synchro operations.

We will begin the discussion of synchro operation with a few basic points on electromagnets. Look at figure 1-8. In this figure, a simple electromagnet is shown with a bar magnet pivoted in the electromagnet's field. In view A, the bar is forced to assume the position shown, since the basic law of magnetism states that like poles of magnets repel and unlike poles attract. Also notice that when the bar is aligned with the field, the magnetic lines of force are shortest. If the bar magnet is turned from this position and held as shown in view B, the flux is distorted and the magnetic lines of force are lengthened. In this condition, a force (torque) is exerted on the bar magnet. When the bar magnet is released, it snaps back to its original position. When the polarity of the electromagnet is reversed, as shown in view C, the field reverses and the bar magnet is rotated 180º from its original position.

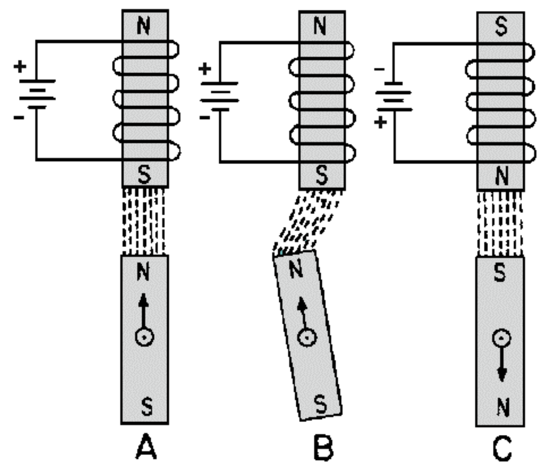


Figure 1-8.—Operation of an electromagnet with a bar-magnet rotor.

Keeping in mind these basic points, consider how the bar magnet reacts to three electromagnets spaced 120º apart as illustrated in figure 1-9. In this figure, stator coils S1 and S3, connected in parallel, together have the same field strength as stator coil S2. The magnetic field is determined by current flow through the coils. The strongest magnetic field is set up by stator coil S2, since it has twice the current and field strength as either S1 or S3 alone. A resultant magnetic field is developed by the combined effects of the three stator fields. Coil S2 has the strongest field, and thus, the greatest effect on the resultant field, causing the field to align in the direction shown by the vector in view B of the figure. The iron-bar rotor aligns itself within the resultant field at the point of greatest flux density. By convention, this position is known as the zero-degree position. The rotor can be turned from this position to any number of positions by applying the proper combination of voltages to the three coils, as illustrated in figure 1-10, view (A), view (B), view (C), view (D), view (E), view (F).

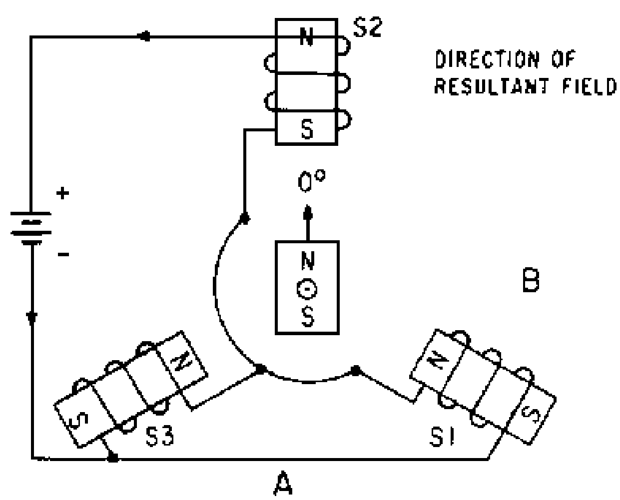


Figure 1-9.—Operation of three electromagnets spaced 120º apart.

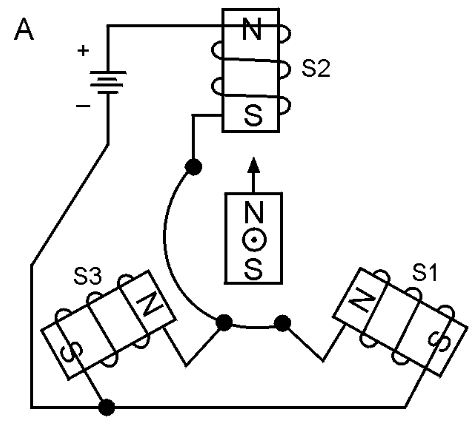


Figure 1-10A.—Positioning of a bar magnet with three electromagnets.

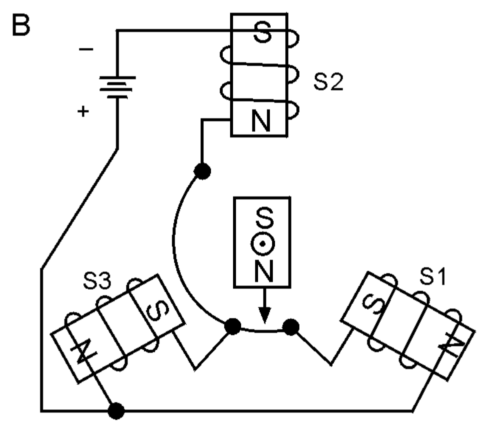


Figure 1-10B.—Positioning of a bar magnet with three electromagnets.

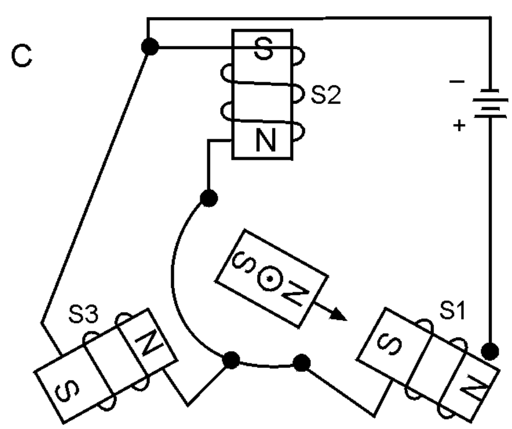


Figure 1-10C.—Positioning of a bar magnet with three electromagnets.

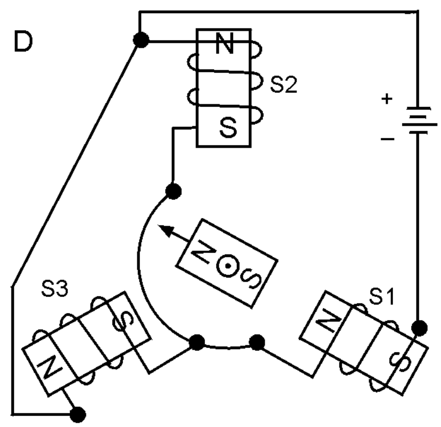


Figure 1-10D.—Positioning of a bar magnet with three electromagnets.

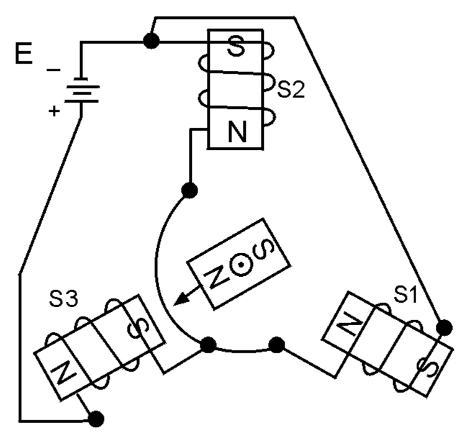


Figure 1-10E.—Positioning of a bar magnet with three electromagnets.

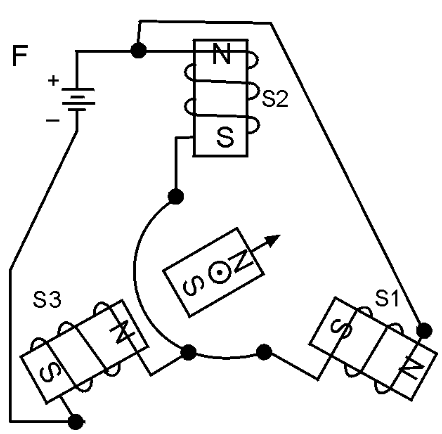


Figure 1-10F.—Positioning of a bar magnet with three electromagnets.

Notice in figure 1-10, in views A C, and E, that the rotor positions are achieved by shifting the total current through different stator windings (S1, S2, and S3). This causes the rotor to move toward the coil with the strongest magnetic field. To obtain the rotor positions in views B, D, and F, it was necessary only to reverse the battery connections. This causes the direction of current flow to reverse and in turn reverses the direction of the magnetic field. Since the rotor follows the magnetic field the rotor also changes direction. By looking closely at these last three rotor positions, you will notice that they are exactly opposite the first three positions we discussed. This is caused by the change in the direction of current flow. You can now see that by varying the voltages to the three stator coils, we can change the current in these coils and cause the rotor to assume any position we desire.

In the previous examples, dc voltages were applied to the coils. Since synchros operate on ac rather than dc, consider what happens when ac is applied to the electromagnet in figure 1-11. During one complete cycle of the alternating current, the polarity reverses twice.

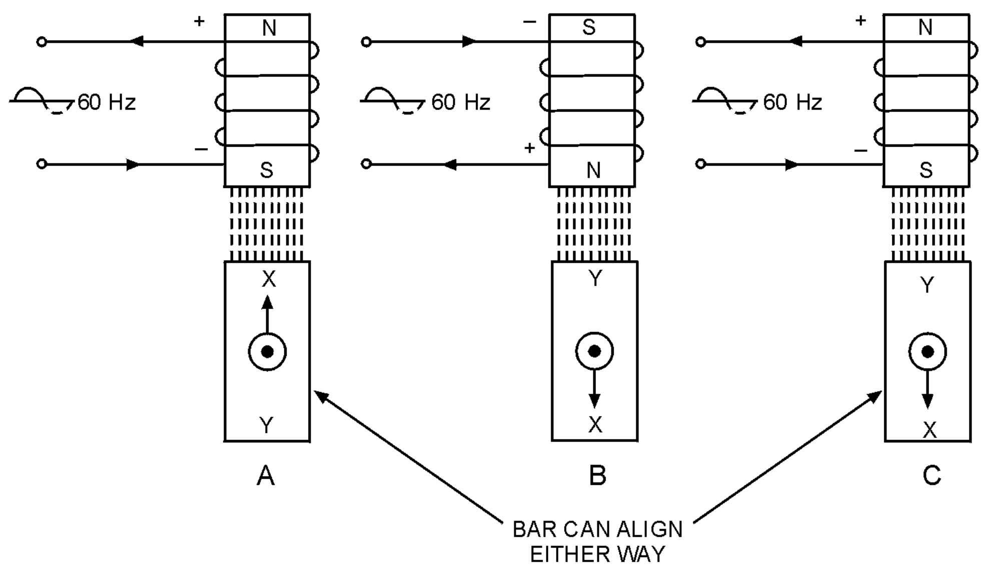


Figure 1-11.—Operation of an electromagnet with ac voltage.

Therefore, the number of times the polarity reverses each second is twice the excitation frequency, or 120 times a second when a 60-Hz frequency is applied. Since the magnetic field of the electromagnet follows this alternating current, the bar magnet is attracted in one direction during one-half cycle (view A) and in the other direction during the next half cycle (view B). Because of its inertia, the bar magnet cannot turn rapidly enough to follow the changing magnetic field and may line up with either end toward the coil (view C). This condition also causes weak rotor torque. For these reasons, the iron-bar rotor is not practical for ac applications. Therefore, it must be replaced by an electromagnetic rotor as illustrated in figure 1-12.

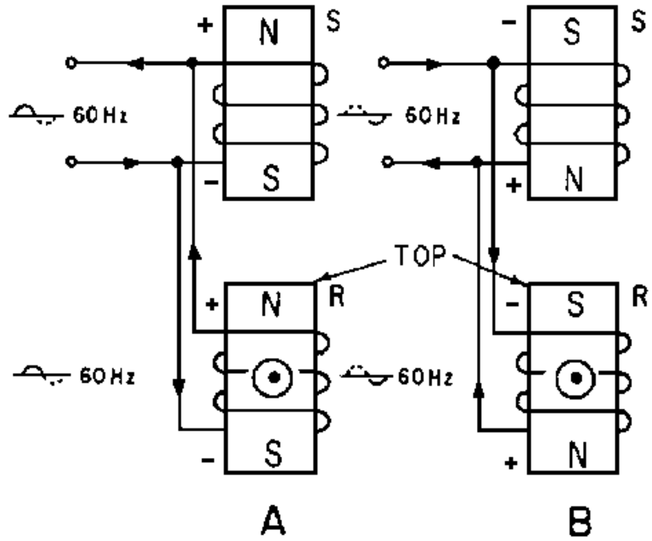


Figure 1-12.—Operation of fixed and moveable electromagnets with ac voltage.

In this figure, both stationary and rotating coils are connected to the same 60-Hz source. During the positive alternation (view A), the polarities are as shown and the top of the rotor is attracted to the bottom of the stationary coil. During the negative alternation (view B), the polarities of both coils reverse, thus keeping the rotor aligned in the same position. In summary, since both magnetic fields change direction at the same time when following the 60-Hz ac supply voltage, the electromagnetic rotor does not change position because it is always aligned with the stationary magnetic field.

Q-18. How do synchros differ from conventional transformers?

Q-19. Describe the zero-position of a synchro transmitter.

SYNCHRO TORQUE TRANSMITTER

The synchro transmitter converts the angular position of its rotor (mechanical input) into an electrical output signal.

When a 115-volt ac excitation voltage is applied to the rotor of a synchro transmitter, such as the one shown in figure 1-13, the resultant current produces an ac magnetic field around the rotor winding. The lines of force cut through the turns of the three stator windings and, by transformer action, induce voltage into the stator coils. The effective voltage induced in any stator coil depends upon the angular position of that coil's axis with respect to the rotor axis. When the maximum effective coil voltage is known, the effective voltage induced into a stator coil at any angular displacement can be determined.

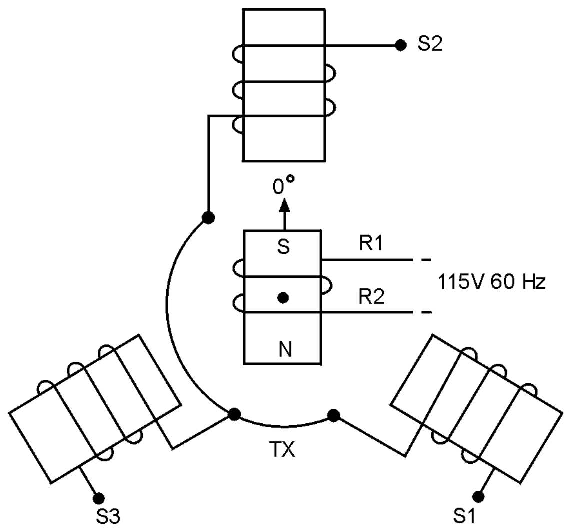


Figure 1-13.—Synchro transmitter.

Figure 1-14 illustrates a cross section of a synchro transmitter and shows the effective voltage induced in one stator coil as the rotor is turned to different positions. The turns ratios in synchros may vary widely, depending upon design and application, but there is commonly a 2.2:1 stepdown between the rotor and a single coil. Thus, when 115 volts is applied to the rotor, the highest value of effective voltage induced in any one stator coil is 52 volts. The maximum induced voltage occurs each time there is maximum magnetic coupling between the rotor and the stator coil (views A, C, and E). The effective voltage induced in the secondary winding is approximately equal to the product of the effective voltage on the primary, the secondary-to-primary turns ratio, and the magnetic coupling between primary and secondary. Therefore, because the primary voltage and the turns ratio are constant, it is commonly said that the secondary voltage varies with the angle between the rotor and the stator.

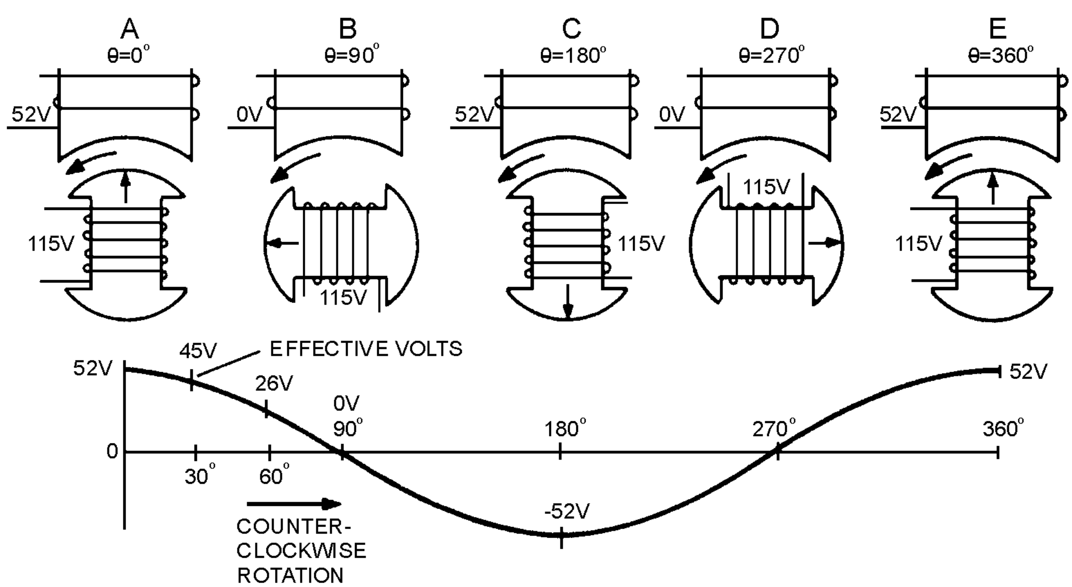


Figure 1-14.—Stator voltage vs rotor position.

When stator voltages are measured, reference is always made to terminal-to-terminal voltages (voltage induced between two stator terminals) instead of to a single coil's voltage. This is because the voltage induced in one stator winding cannot be measured because the common connection between the stator coils is not physically accessible.

In summary, the synchro transmitter converts the angular position of its rotor into electrical stator signals, which are sent through interconnecting wires to other synchro devices.

Q-20. When is the maximum voltage induced into a stator coil?

Q-21. What three factors determine the amplitude of the voltage induced into a stator winding?

SYNCHRO TORQUE RECEIVER

Synchro torque receivers, commonly called synchro receivers, are electrically identical to torque transmitters of the same size except for the addition of some form of damping. In some sizes of 400-Hz synchros, units are designated as torque receivers but may be used as either transmitters or receivers.

Unlike the transmitter, the receiver has an electrical input to its stator and a mechanical output from its rotor. The synchro receiver's function is to convert the electrical data supplied to its stator from the transmitter, back to a mechanical angular position through the movement of its rotor. This function is accomplished when the rotor is connected to the same ac source as the transmitter and assumes a position determined by the interaction of its magnetic field with the magnetic field of the stator. If you recall, this is the same concept discussed earlier under the operation of electromagnets.

Normally, the receiver rotor is unrestrained in movement except for brush and bearing friction. When power is first applied to a system, the transmitter position changes quickly; or if the receiver is switched into the system, the receiver rotor turns to correspond to the position of the transmitter rotor. This sudden motion can cause the rotor to oscillate (swing back and forth) around the synchronous position If the movement of the rotor is great enough, it may even spin. Some method of preventing oscillations or spinning must be used. Any method that accomplishes this task is termed DAMPING.

There are two types of damping methods ELECTRICAL and MECHANICAL. In small synchros the electrical method is used more frequently than the mechanical method. This method uses an additional winding placed in the synchro to retard oscillations. In larger units, a mechanical device, known as an inertia damper, is more effective. Several variations of the inertia damper are in use. One of the more common types consists of a heavy brass flywheel (inertia damper), which is free to rotate around a bushing that is attached to the rotor shaft (fig. 1-15). A tension spring on the bushing rubs against the flywheel so that the bushing and flywheel turn together during normal operation. If the rotor shaft turns or tends to change its speed or direction of rotation suddenly, the inertia of the damper opposes the changing condition.

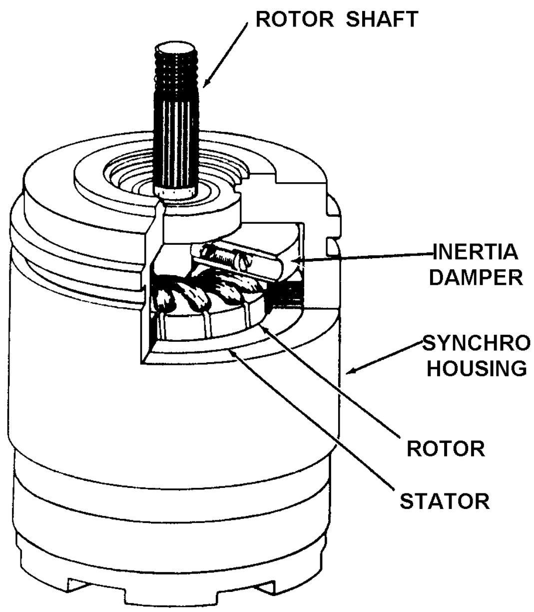


Figure 1-15.—Cutaway view of torque receiver with inertia damper.

Q-22. What is the physical difference between a synchro transmitter and a synchro receiver?

Q-23. What method is used to prevent oscillations in large synchro units?

TORQUE SYNCHRO SYSTEM

A torque transmitter (TX) and a torque receiver (TR) make up a simple torque-synchro system. Basically, the electrical construction of synchro transmitters and receivers is similar, but their intended functions are different. The rotor of a synchro transmitter is usually geared to a manual or mechanical input. This gearing may drive a visual indicator showing the value or quantity being transmitted. The rotor of the receiver synchronizes itself electrically with the position of the rotor of the transmitter and thus responds to the quantity being transmitted.

BASIC SYNCHRO SYSTEM OPERATION

A simple synchro transmission system consisting of a torque transmitter connected to a torque receiver (TX-TR) is illustrated in figure 1-16. As you can see, in this system the rotors are connected in parallel across the ac line. The stators of both synchros have their leads connected S1 to S1, S2 to S2, and S3 to S3, so the voltage in each of the transmitter stator coils opposes the voltage in the corresponding coils of the receiver. The voltage directions are indicated by arrows for the instant of time shown by the dot on the ac line voltage.

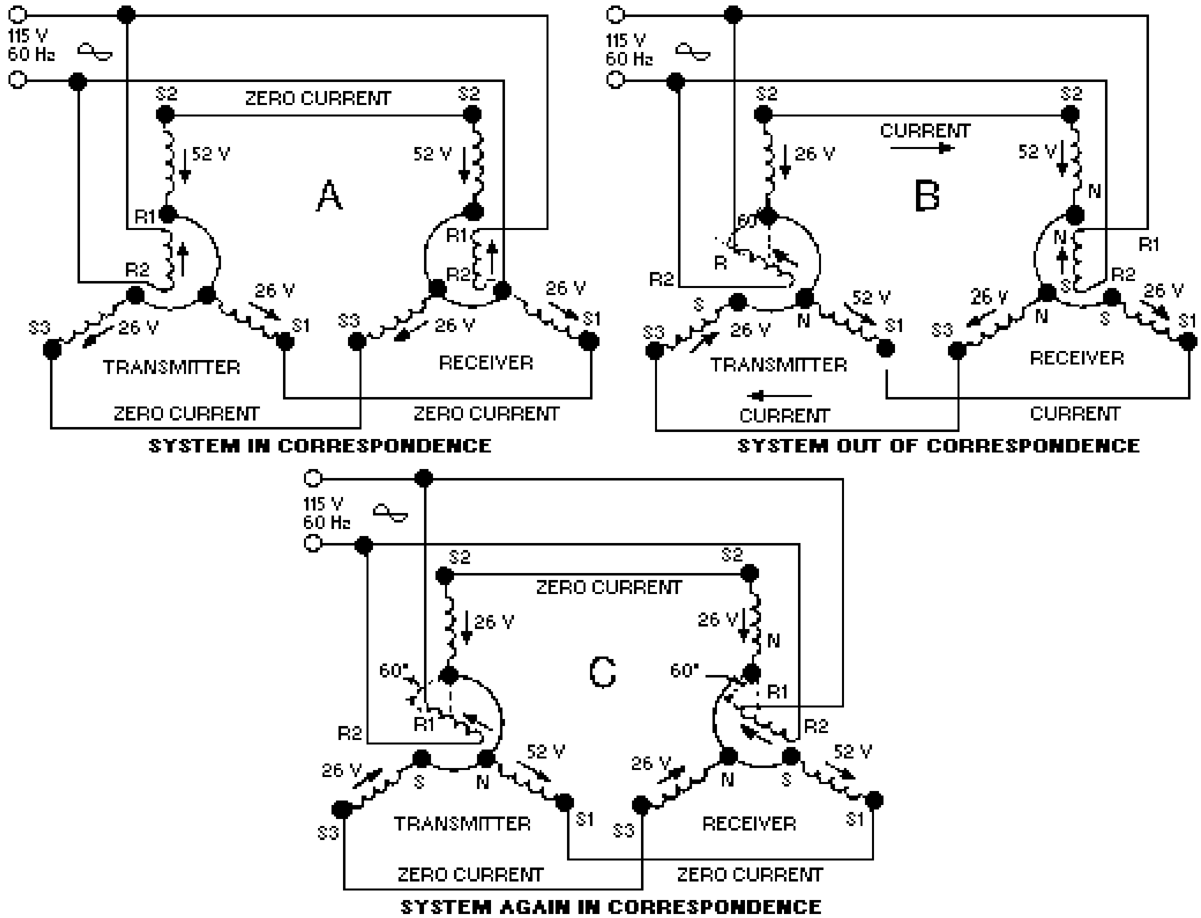


Figure 1-16.—A simple synchro transmission system.

When both transmitter and receiver rotors in a synchro system are on zero or displaced from zero by the same angle, a condition known as CORRESPONDENCE exists. In view A of figure 1-16, the transmitter and receiver are shown in correspondence. In this condition, the rotor of the TR induces voltages in its stator coils (S2 = 52V; S1 and S3 = 26V) that are equal to and opposite the voltages induced into the TX stator coils (S2 = 52V; S1 and S3 = 26V). This causes the voltages to cancel and reduces the stator currents to zero. With zero current through the coils, the receiver torque is zero and the system remains in correspondence.

The angle through which a transmitter rotor is mechanically rotated is called a SIGNAL. In view B of figure 1-16, the signal is 60º. Now, consider what happens to the two synchros in correspondence when this signal is generated

When the transmitter rotor is turned, the rotor field follows and the magnetic coupling between the rotor and stator windings changes. This results in the transmitter S2 coil voltage decreasing to 26 volts, the S3 coil voltage reversing direction, and the S1 coil voltage increasing to 52 volts. This imbalance in voltages, between the transmitter and receiver, causes current to flow in the stator coils in the direction of the stronger voltages. The current flow in the receiver produces a resultant magnetic field in the receiver stator in the same direction as the rotor field in the transmitter. A force (torque) is now exerted on the receiver rotor by the interaction between its resultant stator field and the magnetic field around its rotor. This force causes the rotor to turn through the same angle as the rotor of the transmitter. As the receiver approaches correspondence, the stator voltages of the transmitter and receiver approach equality. This action decreases the stator currents and produces a decreasing torque on the receiver. When the receiver and the transmitter are again in correspondence, as shown in view C, the stator voltages between the two synchros are equal and opposite (S1 = 52V; S2 and S3 = 26V), the rotor torque is zero, and the rotors are displaced from zero by the same angle (60º). This sequence of events causes the transmitter and receiver to stay in correspondence.

In the system we just explained, the receiver reproduced the signal from the transmitter. As you can see, a synchro system such as this could provide a continuous, accurate, visual reproduction of important information to remote locations.

Q-24. What two components make up a simple synchro transmission system?

Q-25. What leads in a simple synchro system are connected to the ac power line?

Q-26. What is the relationship between the transmitter and receiver stator voltages when their rotors are in correspondence?

Q-27. What is the name given to the angle through which a transmitters rotor is mechanically rotated?

Receiver Rotation

When the teeth of two mechanical gears are meshed and a turning force is applied, the gears turn in opposite directions. If a third gear is added, the original second gear turns in the same direction as the first. This is an important concept, because the output of a synchro receiver is often connected to the device it operates through a train of mechanical gears. Whether or not the direction of the force applied to the device and the direction in which the receiver rotor turns are the same depends on whether the number of gears in the train is odd or even. The important thing, of course, is to move the dial or other device in the proper direction. Even when there are no gears involved, the receiver rotor may turn in the direction opposite to the direction you desire. To correct this problem, some method must be used to reverse the receiver's direction of rotation. In the transmitter-receiver system, this is done by reversing the S1 and S3 connections so that SI of the transmitter is connected to S3 of the receiver and vice versa (fig. 1-17), view (A) and view (B).

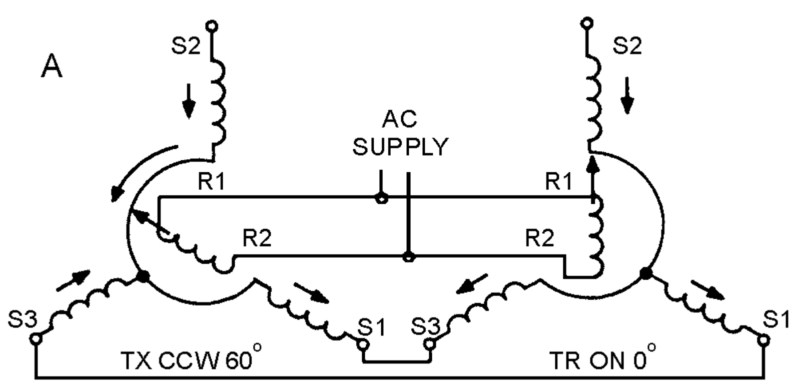


Figure 1-17A.—Effect of reversing the S1 and S3 connections between the transmitter and the receiver.

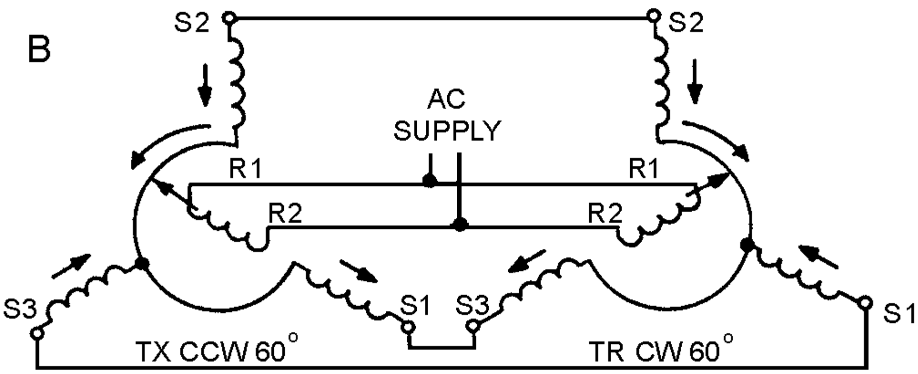


Figure 1-17B.—Effect of reversing the S1 and S3 connections between the transmitter and the receiver.

Even when the S1 and S3 connections are reversed, the system at 0º acts the same as the basic synchro system we previously described at 0º. This is because the voltages induced in the S1 and S3 stator windings are still equal and oppose each other. This causes a canceling effect, which results in zero stator current and no torque. Without the torque required to move the receiver rotor, the system remains in correspondence and the reversing of the stator connections has no noticeable effect on the system at 0º.

Suppose the transmitter rotor is turned counterclockwise 60º, as shown in view A of figure 1-17. The TX rotor is now aligned with S1. This results in maximum magnetic coupling between the TX rotor and the S1 winding. This maximum coupling induces maximum voltage in S1. Because S1 is connected to S3 of the TR, a voltage imbalance occurs between them. As a result of this voltage imbalance, maximum current flows through the S3 winding of the TR causing it to have the strongest magnetic field. Because the other two fields around S2 and S1 decrease proportionately, the S3 field has the greatest effect on the resultant TR stator field. The strong S3 stator field forces the rotor to turn 60º clockwise into alignment with itself, as shown in view B. At this point, the rotor of the TR induces canceling voltages in its own stator coils and causes the rotor to stop. The system is now in correspondence. Notice that by reversing S1 and S3, both synchro rotors turn the same amount, but in OPPOSITE DIRECTIONS.

We must emphasize that the only stator leads ever interchanged, for the purpose of reversing receiver rotation, are S1 and S3. S2 cannot be reversed with any other lead since it represents the electrical zero position of the synchro. As you know, the stator leads in a synchro are 120º apart. Therefore, any change in the S2 lead causes a 120º error in the synchro system and also reverses the direction of rotation.

In new or modified synchro systems, a common problem is the accidental reversal of the R1 and R2 leads on either the transmitter or receiver. This causes a 180º error between the two synchros, but the direction of rotation remains the same.

Q-28. What two receiver leads are reversed to reverse the rotor's direction of rotation?

Q-29. What is the most likely problem if the transmitter shaft reads 0º when the receiver shaft indicates 180º?

TORQUE DIFFERENTIAL SYNCHRO SYSTEMS

The demands on a synchro system are not always as simple as positioning an indicating device in response to information received from a single source (transmitter). For example, an error detector used in checking weapons equipment uses a synchro system to determine the error in a gun's position with respect to the positioning order. To do this, the synchro system must accept two signals, one containing the positioning order and the other corresponding to the actual position of the gun. The system must then compare the two signals and position an indicating dial to show the difference between them, which is the error.

Obviously, the simple synchro transmitter-receiver system discussed so far could not handle a job of this sort. A different type of synchro is needed, one which can accept two signals simultaneously, add or subtract the signals, and furnish an output proportional to their sum or difference. This is where the SYNCHRO DIFFERENTIAL enters the picture. A differential can perform all of these functions.

There are two types of differential units - differential transmitters and differential receivers. The differential transmitter (TDX) accepts one electrical input and one mechanical input and produces one electrical output. The differential receiver (TDR) accepts two electrical inputs and produces one mechanical output. A comparison of the TDX and TDR is shown in figure 1-18. The torque differential transmitter and the torque differential receiver can be used to form a DIFFERENTIAL SYNCHRO SYSTEM. The system can consist either of a torque transmitter (TX), a torque differential transmitter (TDX), and a torque receiver (TR), (TX-TDX-TR); or two torque transmitters (TXs) and one torque differential receiver (TDR), (TX-TDR-TX). Before beginning a discussion of the systems using differentials, we need to provide a brief explanation on the newly introduced synchros, the TDX and the TDR.

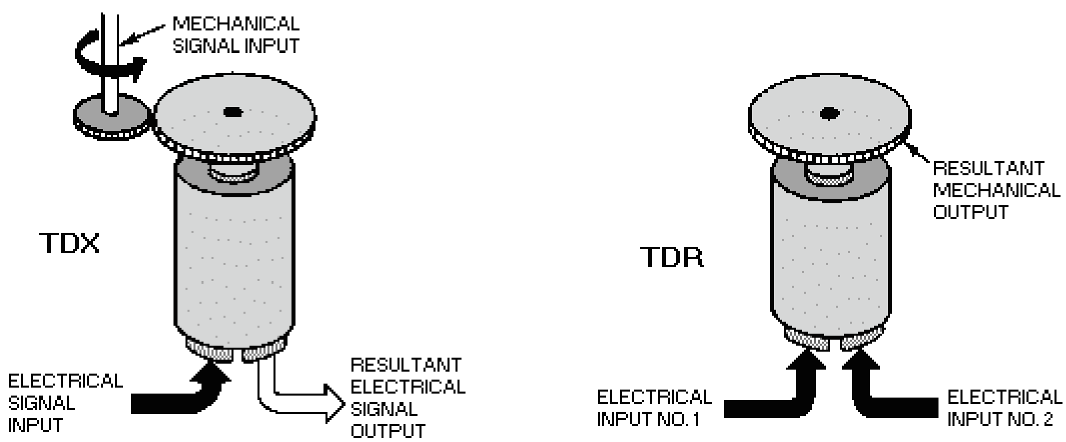


Figure 1-18.—Torque differentials.

Torque Differential Transmitter

In the torque differential transmitter, BOTH the rotor and stator windings consist of three Y-connected coils, as illustrated in view A of figure 1-19. The stator is normally the primary, and receives its input signal from a synchro transmitter. The voltages appearing across the differential's rotor terminals (R1, R2, and R3) are determined by the magnetic field produced by the stator currents, the physical positioning of the rotor, and the step-up turns ratio between the stator and the rotor. The magnetic field, created by the stator currents, assumes an angle corresponding to that of the magnetic field in the transmitter supplying the signal. The position of the rotor controls the amount of magnetic coupling that takes place between the stator magnetic field and the rotor, and therefore, the amount of voltage induced into the rotor windings. If the rotor position changes in response to a mechanical input, then the voltages induced into its windings also change. Therefore, the output voltage of the TDX varies as a result of either a change in the input stator voltage or a change in the mechanical input to the rotor. This electrical output of the TDX may be either the SUM or the DIFFERENCE of the two inputs depending upon how the three units (the TX, the TDX, and the TR) are connected.

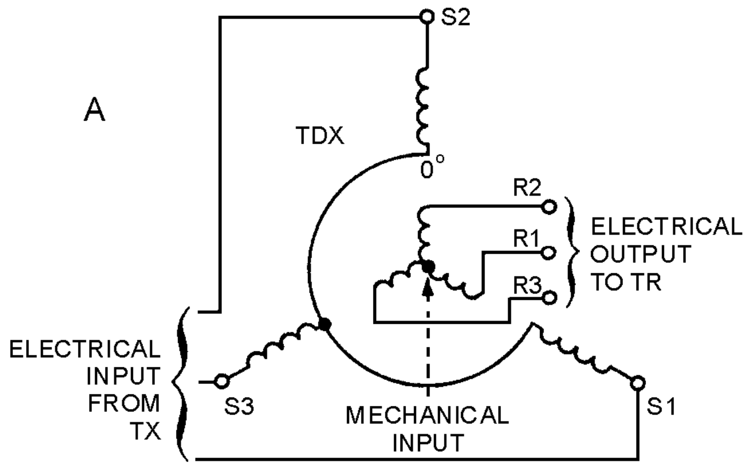


Figure 1-19A.—Torque differential transmitter.

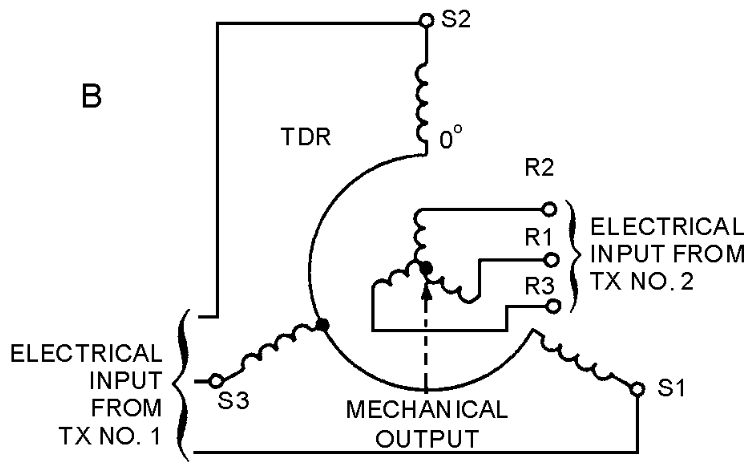


Figure 1-19B.—Torque differential receiver.

Torque Differential Receiver

The torque differential transmitter (TDX) and the torque differential receiver (TDR) are ELECTRICALLY IDENTICAL. The only difference in their construction is that the receiver (TDR) has a damper, which serves the same purposes as the damper in the TR — it prevents the rotor from oscillating. The real difference in the receiver lies in its application. It provides the mechanical output for a differential synchro system usually as the sum or difference of two electrical inputs from synchro transmitters. As in the case with the TDX, the TDR addition or subtraction function depends upon how the units in the system are connected.

Basically, the torque differential receiver operates like the electromagnets we discussed earlier in this chapter. In view B, the rotor and stator of the torque differential receiver receive energizing currents from two torque transmitters. These currents produce two resultant magnetic fields, one in the rotor and the other in the stator. Each magnetic field assumes an angle corresponding to that of the magnetic field in the transmitter supplying the signal. It is the interaction of these two resultant magnetic fields that causes the rotor in the TDR to turn.

Q-30. What is the purpose of using differential synchros instead of regular synchros?

Q-31. What are the two types of differential synchros?

Q-32. Other than their physical differences, what is the major difference between a TDX and a TDR?

Q-33. What determines whether a differential synchro adds or subtracts?

TX-TDX-TR System Operation (Subtraction)

Now that you know how the individual units work, we can continue our discussion with their application in different systems. The following sections explain how the TDX and TDR are used with other synchros to add and subtract.

To understand how a TDX subtracts one input from another, first consider the conditions in a TX-TDX-TR system when all the rotors are on 0º, as in view A of figure 1-20. In this case, the TDX is on electrical zero and merely passes along the voltages applied to its windings without any change. Therefore, the TX stator voltages are felt at the TDX rotor. With the system in perfect balance, the TDX rotor voltages equal and oppose the TR stator voltages so that no current flows in the circuit. Since there is no current to produce the torque required to move the TR rotor, the system will remain in this condition, thus solving the equation 0º – 0º = 0º.

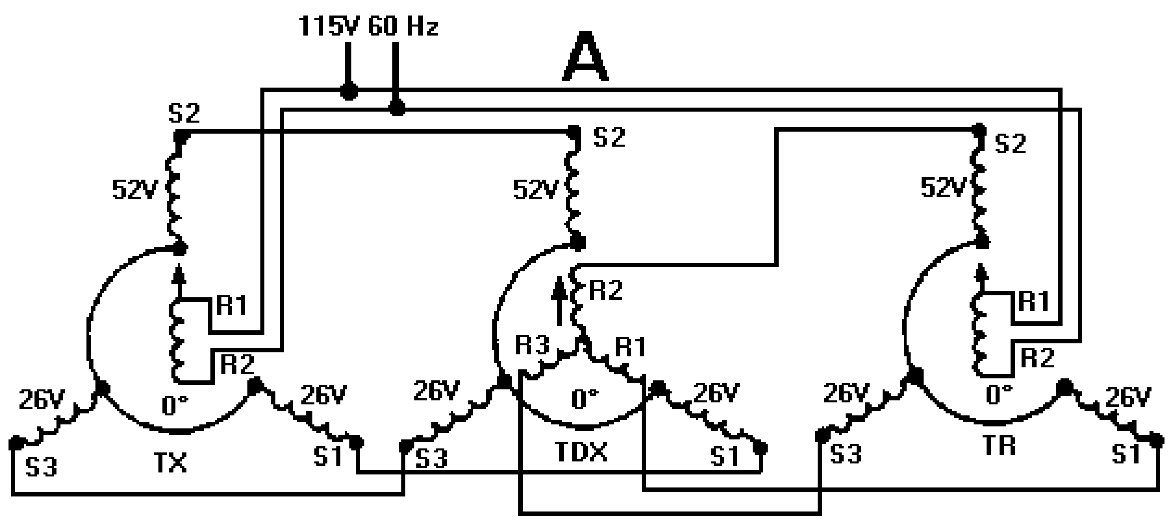


Figure 1-20A.—TX-TDX-TR system operation (subtraction).

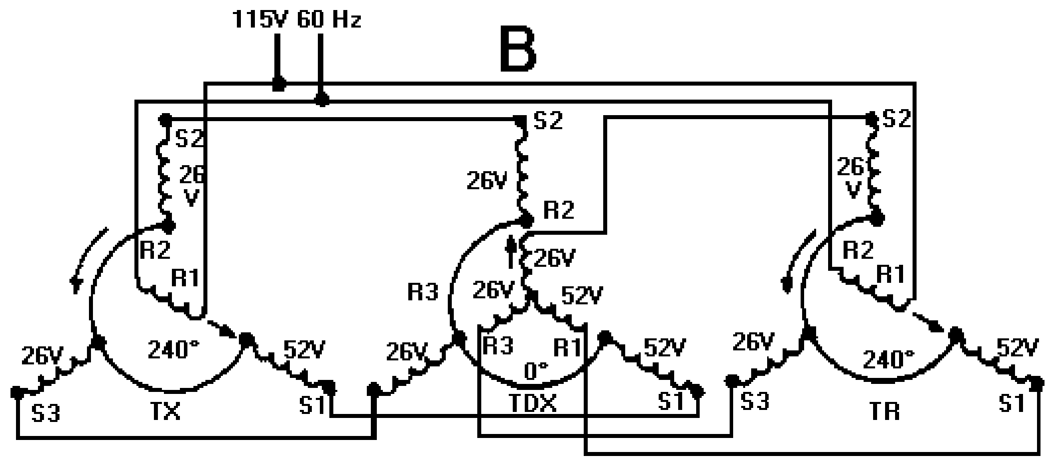


Figure 1-20B.—TX-TDX-TR system operation (subtraction).

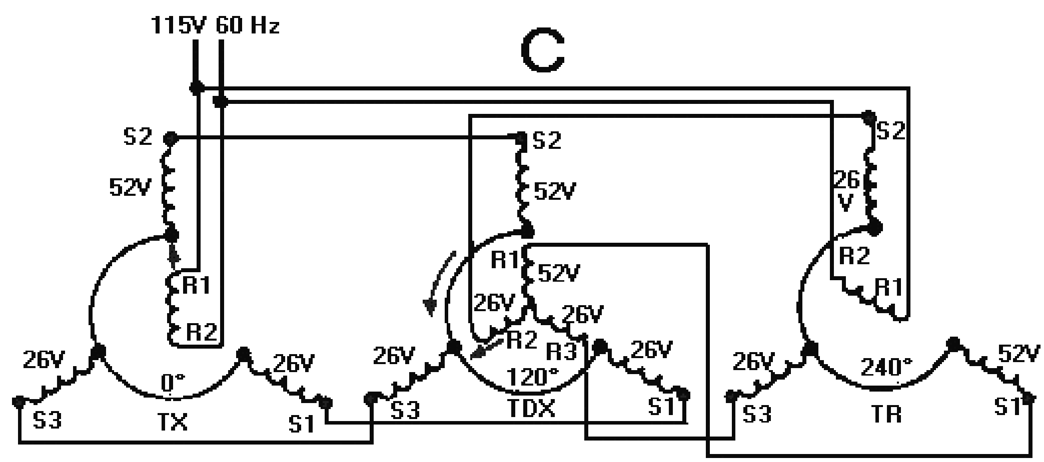


Figure 1-20C.—TX-TDX-TR system operation (subtraction).

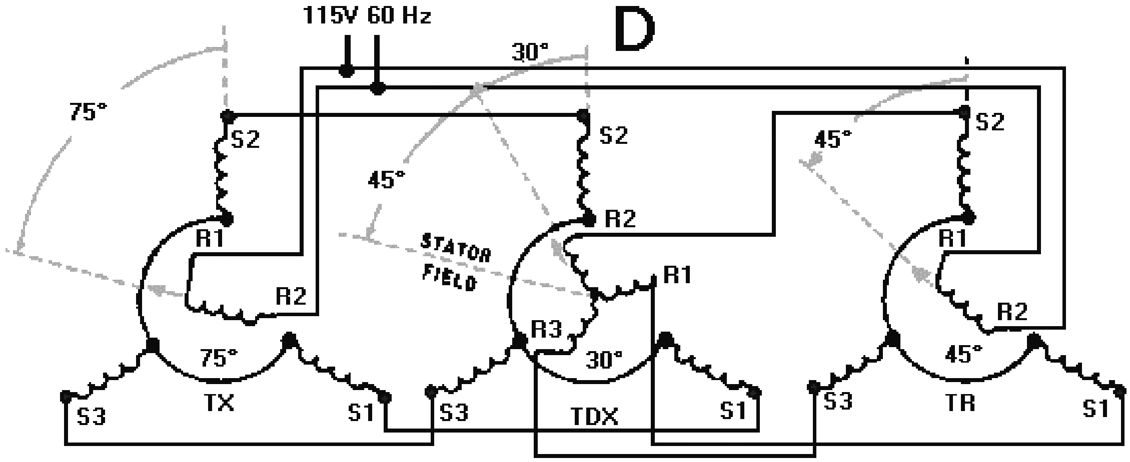


Figure 1-20D.—TX-TDX-TR system operation (subtraction).

Up to this point, we have discussed the number of degrees a rotor is turned. Now, it is important to point out the labeling of synchro positions. Labeling is necessary to determine the actual position of the synchro's rotor. Notice that synchro rotor positions are labeled from 0º, increasing in a counterclockwise direction. It is common practice to refer to a synchro transmitter as being on 120º when its rotor is pointing toward the S3 winding. Do not confuse these positions with the number of degrees a rotor is turned.

Assume that a 240º input is applied to the system, as indicated in view B, by turning the TX rotor to its 240º position. At this position maximum voltage is induced into the S1 winding of the TX and coupled to S1 of the TDX. Since the TDX rotor is on 0º, it passes this maximum voltage (via R1) along to the S1 winding of the TR. The stator magnetic field in the receiver now lines up in the direction of the S1 winding and causes the rotor to turn counterclockwise to the 240º position. This illustrates an important point:

Whenever the TDX rotor is at 0º, the TR rotor follows the TX rotor exactly. In the present case, the system has just solved the equation 240º – 0º = 240º.

Before we go to another example, you need to understand that when you subtract a higher value of degrees from a lower value of degrees, you add 360º to the lower value and subtract directly.

For example: 10º - 260º

Add 360º to lower value: 10º + 360º = 370º

Subtract: 370º – 260º = 110º

In the next example, hold the TX rotor on 0º and turn the TDX rotor to 120º, as illustrated in view C of figure 1-20. In this situation, R1 of the TDX has maximum voltage induced in its winding since it is in line with S2. With R1 of the TDX connected to S1 of the TR, the TR stator magnetic field lines up in the direction of S1 and causes the TR rotor to turn clockwise to the 240º position. Given, then, that the TX is on 360º (or the 0º position), and subtracting the 120º displacement of the TDX rotor, the difference is 240º. This is the position at which the TR rotor comes to rest. Therefore, the system has solved the equation 360º – 120º = 240º. The actual subtraction operation of the TDX is a little more apparent in the next example.

Now, consider what happens in view D when the TX rotor is turned manually to 75º and the TDX rotor is set manually on 30º. When the TX rotor is turned to 75º, magnetic coupling increases between the rotor and S1. This, in turn, increases the voltage in S1 and, therefore, the magnetic field surrounding it. At the same time, the field in S2 and S3 decreases proportionately. This causes the resultant TX stator field to line up in the direction of its rotor. The increased voltage in S1 of the TX also causes an increase in current flow through S1 in the TDX, while decreased currents flow through S2 and S3. Therefore, a strong magnetic field is established around the S1 winding in the TDX. This field has the greatest effect on the resultant TDX stator field and causes it to line up in the same relative direction as the TX stator field (75º). The TDX stator field does not move from this 75º position because it is controlled by the position of the TX rotor. However, its angular position with respect to the R2 winding decreases by 30 when the TDX rotor is turned. Therefore, the signal induced into the TDX rotor and transmitted to the TR is 45º. The TR rotor responds to the transmitted signal and turns counterclockwise to 45º. This system has just solved the equation 75º – 30º = 45º.

TX-TDX-TR System Operation (Addition)

Frequently it is necessary to set up a TX-TDX-TR system for addition. This is done by reversing the S1 and S3 leads between the TX and the TDX, and the R1 and R3 leads between the TDX and the TR. With these connections, the system behaves as illustrated in figure 1-21. Consider what happens when the TX rotor is turned to 75º, while the TDX is set at 0º view A. In the TX, with the rotor at 75º, increased coupling between the rotor and S1 increases the current in, and consequently the magnetic field around, that coil. At the same time, the field strengths of S2 and S3 decrease proportionately. This causes the resultant field of the TX stator to rotate counterclockwise and align itself with its rotor field. The system is now connected so the increased current in S1 of the TX flows through S3 of the TDX, while decreased currents flow through S1 and S2. Therefore, in the TDX, the resultant stator field is shifted 75º clockwise because of the stronger field around S3. Since the rotor of the TDX is on 0º, the voltage in the rotor is not changed but simply passed on to the TR. Remember, the R1 and R3 leads between the TDX and the TR have also been reversed. Just as in the simple TX-TR system with S1 and S3 leads interchanged, torque is developed in the TR, which turns the rotor in a direction opposite to the rotation of the TDX stator field. Therefore, the TR rotor rotates 75º counterclockwise and aligns itself with the TX rotor. Thus, the TX-TDX-TR system connected for addition behaves in the same way as the system connected for subtraction as long as the TDX rotor remains on 0º. When this condition exists, the TR rotor follows the TX rotor exactly. As you can see, the system in view. A just solved the equation 75º + 0º = 75º.

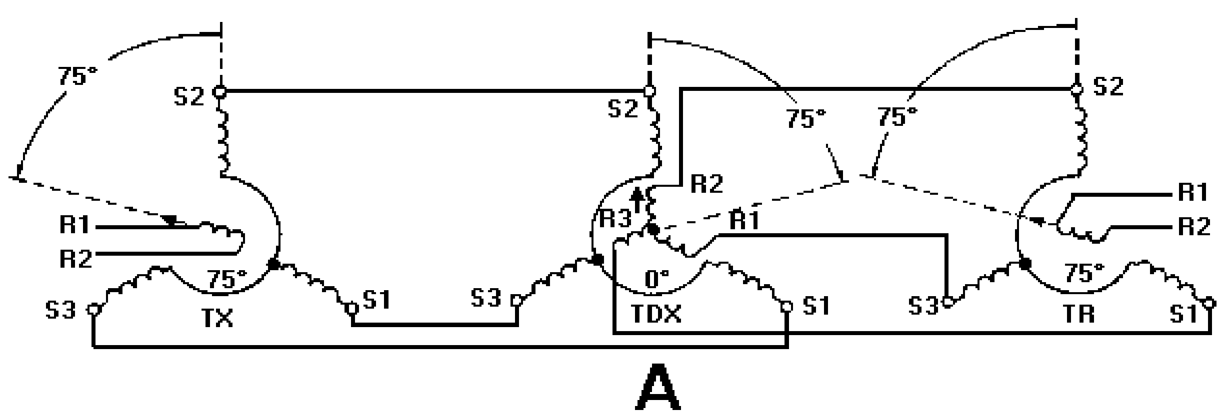


Figure 1-21A.—TX-TR system operation (addition).

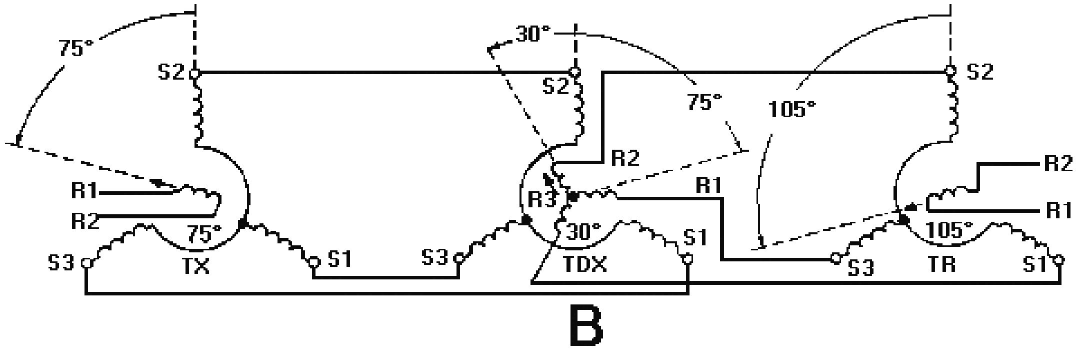


Figure 1-21B.—TX-TR system operation (addition).

Now, with the TX in the same position (75º), the TDX rotor is turned to 30º (view B). The angle between the TDX stator field and R2 is then increased by 30º. This appears to the TR as an additional rotation of the TDX stator field. In transmitting the TX signal to the TR, the TDX adds the amount its own rotor has turned. The TR rotor now turns to 105º. Thus, the equation 75º + 30º = 105º is solved.

Q-34. In a TDX system when does the TR rotor follow the TX rotor exactly?

Q-35. What is the angular position of a TX rotor when it is pointing toward the S1 winding? (Hint. Remember synchros are labeled counter clockwise from 0º.)

Q-36. In a TDX system with standard synchro connections, the TX rotor is at 120º and the TDX rotor is at 40º. What position will the TR indicate?

Q-37. What connections in a TDX system are reversed to set up the system for addition?

TX-TDR-TX System Operation (Subtraction)

As we previously explained, the differential receiver differs chiefly from the differential transmitter in its application. The TDX in each of the previous examples combined its own input with the signal from a synchro transmitter (TX) and transmitted the sum or difference to a synchro receiver (TR). The synchro receiver then provided the system's mechanical output. When the differential receiver (TDR) is used, the TDR itself provides the system's mechanical output. This output is usually the sum or difference of the electrical signals received from two synchro transmitters. Figure 1-22 shows a system consisting of two TXs (No. 1 and No. 2) and a TDR connected for subtraction.

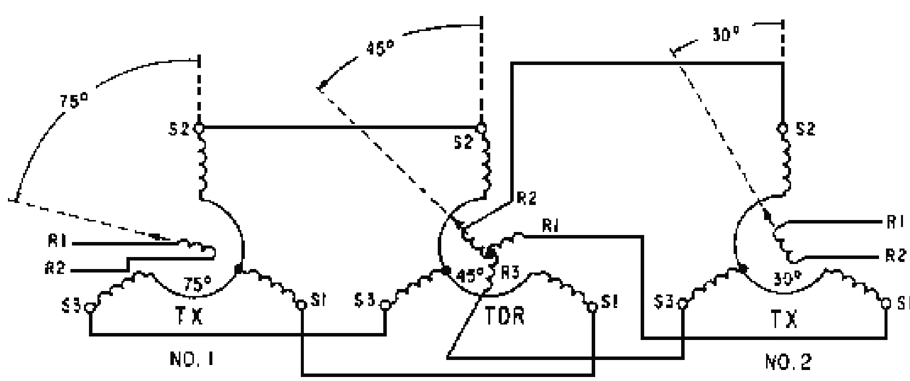


Figure 1-22.—TX-TDR-TX system operation (subtraction).

In this figure the signal from TX No. 1 rotates the resultant TDR stator field 75º counterclockwise. In a similar manner, the signal from TX No. 2 rotates the resultant TDR rotor field counterclockwise 30º. Since the two resultant fields are not rotated by equal amounts, a torque is exerted on the rotor to bring the two fields into alignment. This torque causes the rotor to turn to 45º, the point at which the two fields are aligned. To bring the two fields into alignment, the TDR rotor need turn only through an angle equal to the difference between the signals supplied by the two TXs.

TX-TDR-TX System Operation (Addition)

To set up the previous system for addition, it is necessary to reverse only the R1 and R3 leads between the TDR rotor and TX No. 2. With these connections reversed, the system operates as shown in figure 1-23.

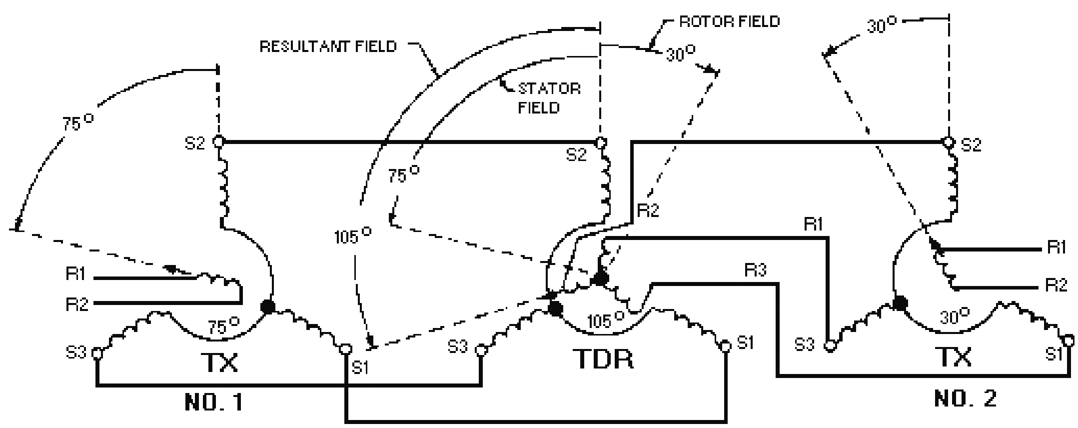


Figure 1-23.—TX-TDR-TX system operation (addition).

Assume the TDR rotor is initially at 0º. TX No. 1 is turned to 75º, and TX rotor No. 2 is turned to 30º. The TDR stator field still rotates counterclockwise 75º, but because R1 and R3 on the TDR rotor are reversed, its rotor field rotates 30º clockwise. The angular displacement of the two fields then, with respect to each other, is the sum of the signals transmitted by the two TXs. The magnetic force pulling the TDR rotor field into alignment with that of the stator turns the TDR rotor to 105º. Therefore, the system solves the equation 75º + 30º = 105º.

Q-38. What connections in a TDR system are reversed to set up the system for addition?

Q-39. In a TDR system connected for addition in what direction will the TDR rotor field turn when the TX rotor to which it is connected turns counterclockwise?

CONTROL SYNCHRO SYSTEMS

It should be clear to you from our discussion of torque synchro systems that, since they produce a relatively small mechanical output, they are suitable only for very light loads. Even when the torque system is moderately loaded, it is never entirely accurate because the receiver rotor requires a slight amount of torque to overcome its static friction.

When large amounts of power and a higher degree of accuracy are required, as in the movement of heavy radar antennas and gun turrets, torque synchro systems give way to the use of CONTROL

SYNCHROS. Control synchros by themselves cannot move heavy loads. However, they are used to "control" servo systems, which in turn do the actual movement. Servo systems are covered in depth in the next chapter in this module.

There are three types of control synchros: the CONTROL TRANSMITTER (CX), the CONTROL TRANSFORMER (CT), and the CONTROL DIFFERENTIAL TRANSMITTER (CDX). The control transmitter (CX) and the control differential transmitter (CDX) are identical to the TX and the TDX we discussed previously except for higher impedance windings in the CX and CDX. The higher impedance windings are necessary because control systems are based on having an internal voltage provide an output voltage to drive a large load. Torque systems, on the other hand, are based on having an internal current provide the driving torque needed to position an indicator. Since we discussed the theory and operation of the TX and the TDX earlier, we will not discuss their counterparts, the CX and CDX. However, we will cover the third control synchro, the CT, in depth during this discussion.

CONTROL TRANSFORMERS

A control transformer is just what its name implies-a control synchro device accurately governing some type of power amplifying device used for moving heavy equipment. Figure 1-24 shows a phantom view of a typical CT and its schematic symbols.

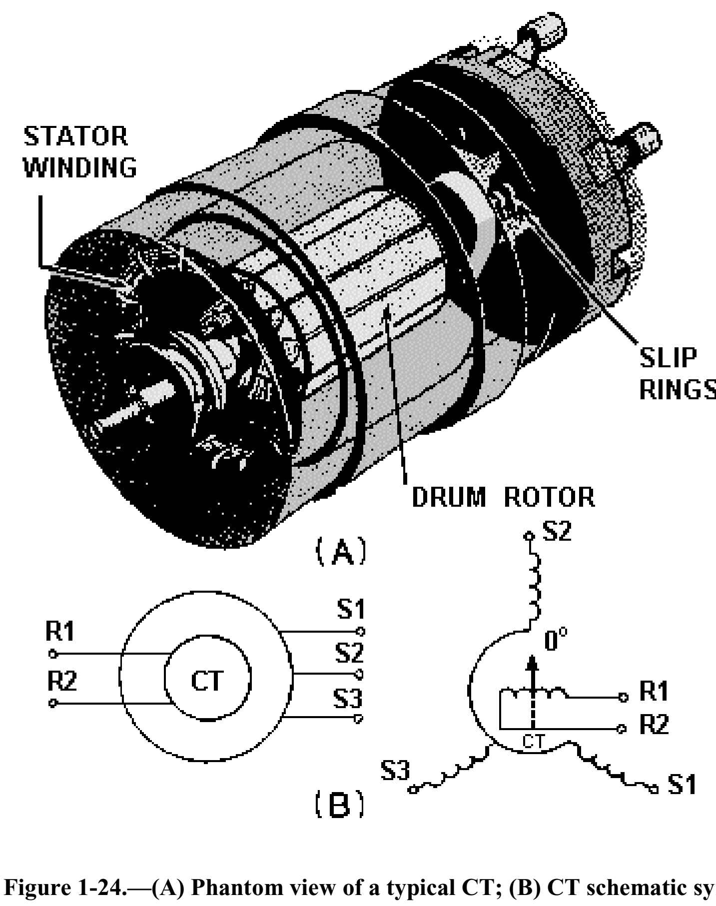


Figure 1-24.—(A) Phantom view of a typical CT; (B) CT schematic symbols.

The CT compares two signals, the electrical signal applied to its stator and the mechanical signal applied to its rotor. Its output is a difference signal that controls a power amplifying device and thus the movement of heavy equipment.

The unit construction and physical characteristics of a control transformer are similar to those of a control transmitter or torque receiver, except that there is no damper and the rotor is a drum or wound rotor rather than a salient-pole rotor.

An interesting point about the rotor is that it is never connected to an ac supply and, therefore, induces no voltages in the stator coils. As a result, the CT stator currents are determined solely by the voltages applied to the high-impedance stator windings. The rotor itself is wound so that its position has very little effect on the stator currents. Also, there is never any appreciable current flowing in the rotor because its output voltage is always applied to a high-impedance load. As a result, the CT rotor does not try to follow the magnetic field of its stator and must be turned by some external force.

The stator windings of the CT are considered to be the primary windings, and the rotor windings the secondary windings. The output, which is taken off the R1 and R2 rotor leads, is the voltage induced in the rotor windings. The phase and amplitude of the output voltage depend on the angular position of the rotor with respect to the magnetic field of the stator.

Q-40. What type of synchro is used in systems requiring large amounts of power and a high degree of accuracy?

Q-41. What are the three types of control synchros?

Q-42. How do the CX and CDX differ from the TX and TDX?

Q-43. What three things prevent a CT rotor from turning when voltages are applied to its stator windings?

CONTROL SYNCHRO SYSTEM OPERATION

A control synchro system consisting of a control transmitter and a control transformer is illustrated in figure 1-25. The stator windings of the CX are connected to the stator windings of the CT and both synchros are shown on 0º. Notice, that at 0º, the CT rotor is perpendicular to its S2 winding. This is contrary to what we have learned so far about synchros, but it is just another peculiarity of the CT. When the rotor of the CX is on 0º, the rotor's magnetic field points straight up as shown (the black arrow). The voltages induced in the CX stator windings, as a result of this field, are impressed on the CT stator windings through the three leads connecting the S1, S2, and S3 terminals. Exciting currents proportional to these voltages flow in the CT stator windings and establish a magnetic field in the CT in the same direction (white arrow) as the magnetic field (black arrow) in the CX. Observe that the rotor of the CT is perpendicular to the stator magnetic field and, therefore, the induced voltage in the rotor is zero, as indicated by the straight line on the oscilloscope presentation.

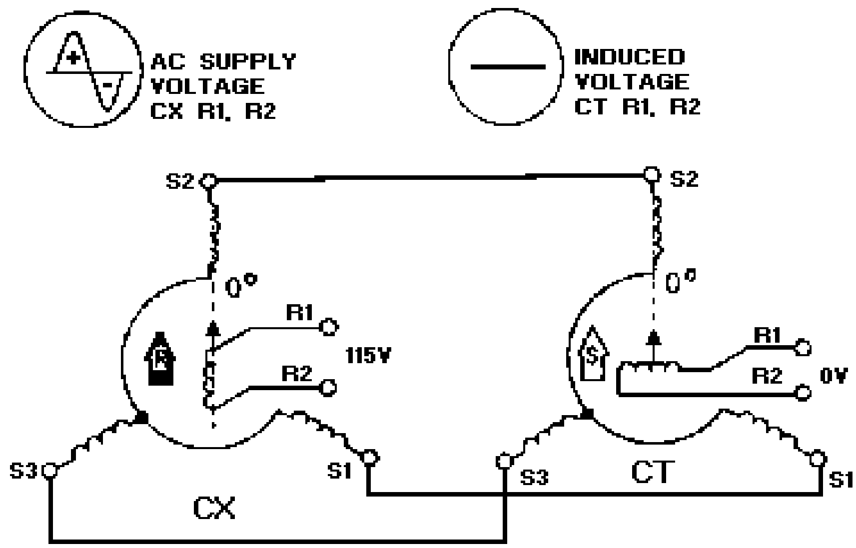


Figure 1-25.—CX-CT system operation with rotor in correspondence.

When the CT rotor is rotated 90º, as shown in figure 1-26, the rotor is parallel to the resultant stator field Maximum magnetic coupling occurs between the rotor and stator fields at this point. As a result of this coupling, the stator windings induce a maximum of 55 volts into the rotor winding. The phase of this voltage depends upon the direction in which the CT rotor is turned. The rotor of the CT is wound so that clockwise rotation of the stator magnetic field induces a voltage across the rotor which is proportional to the amount of rotation and in phase with the ac supply voltage. Counterclockwise rotation of the stator magnetic field produces a voltage that is still proportional to the amount of rotation, but 180º out of phase with the supply voltage. Keep in mind that the clockwise rotation of the CT stator magnetic field is the same as the counterclockwise rotation of the CT rotor. This phase relationship between the ac supply voltage and the CT output voltage becomes more apparent in figure 1-27.

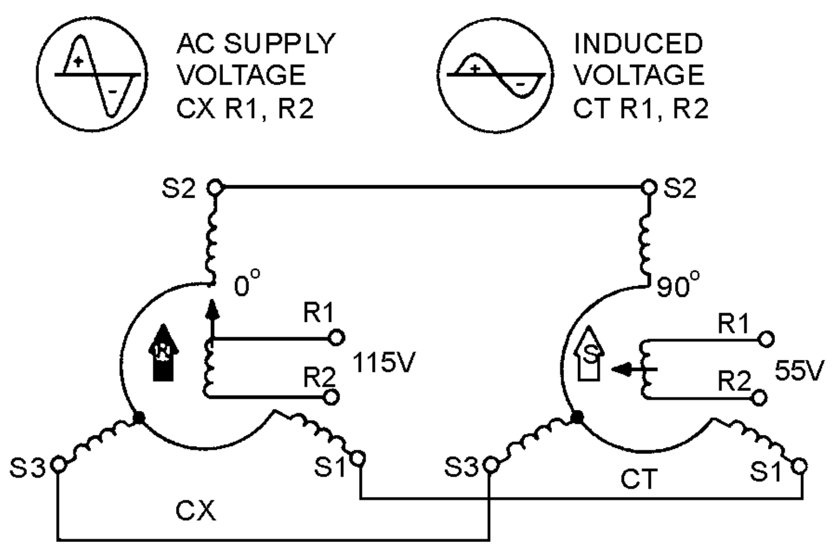


Figure 1-26.—CX-CT system operation with the CX rotor at 0º and the CT rotor at 90º.

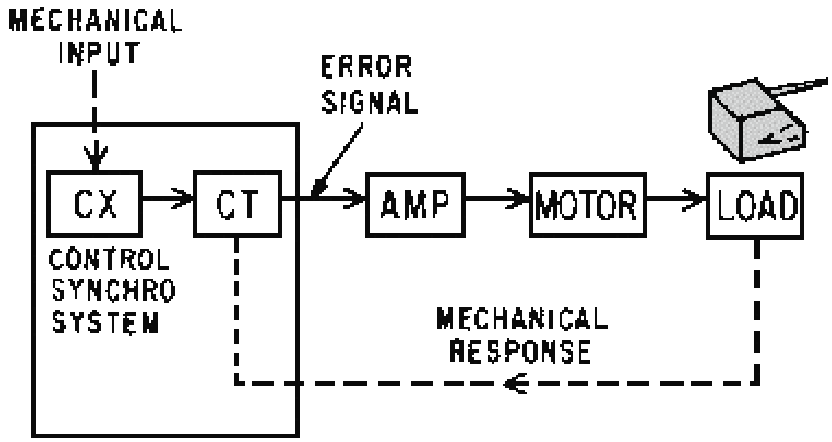


Figure 1-27A.—Control synchro system operation.

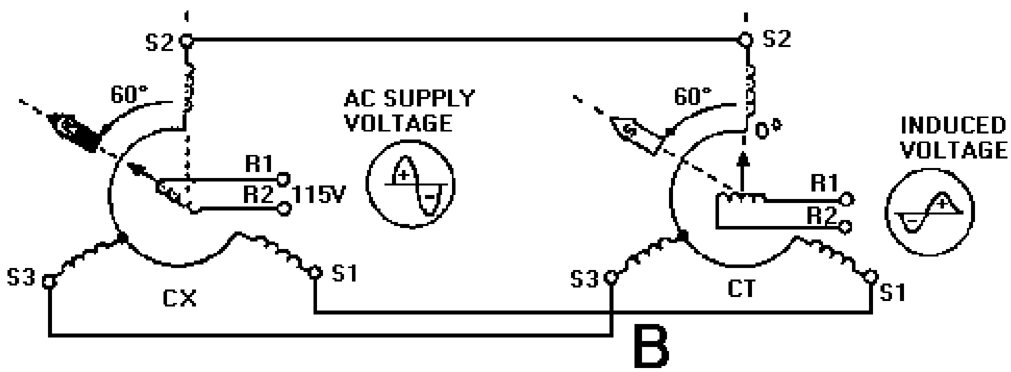


Figure 1-27B.—Control synchro system operation.

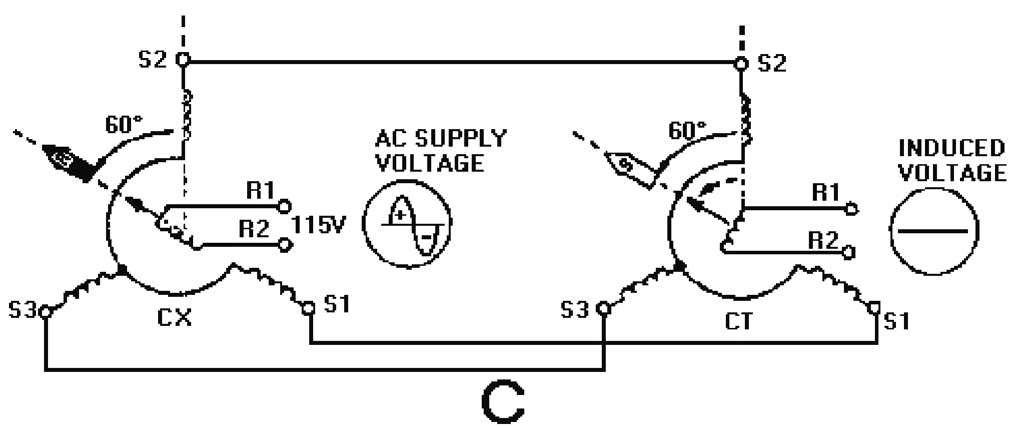


Figure 1-27C.—Control synchro system operation.

When the rotor of the CX in view A of figure 1-27 is turned 60º clockwise, the magnetic field in the CX (black arrow) and the magnetic field in the CT (white arrow) also rotate 60º clockwise. This action induces a voltage in the CT rotor that is in phase with the ac supply, as indicated by the oscilloscope presentation. If the rotor of the CX in view B is turned 60º in a counterclockwise direction from its 0º position, the magnetic field (white arrow) in the CT also rotates counterclockwise through the same number of degrees as the CX. Since the magnetic field in the stator of the CT cuts through the rotor in the opposite direction, the induced voltage in the rotor is now out of phase with the ac supply to the CX, as shown in the oscilloscope presentation.

At times it is necessary, because the CT is used to control servo systems, to have the CT output reduced to zero volts to prevent any further movement of a load. To accomplish this, it is necessary to turn the rotor of the CT through the same number of degrees and in the same direction as the rotor of the CX. This places the CT rotor perpendicular to its own stator field and reduces its output to zero volts as illustrated in view C.

The CT output voltage discussed throughout this section is commonly referred to as an ERROR SIGNAL. This is because the voltage represents the amount and direction that the CX and CT rotors are out of correspondence. It is this error signal that eventually is used in moving the load in a typical servo system.

Now that we have covered the basic operation of the control synchro system, let us see how this system works with a servo system to move heavy equipment. Figure 1-28 shows a block diagram of a typical servo system that uses a control synchro system. Assume the shaft of the CX in this system is turned by some mechanical input. This causes an error signal to be generated by the CT because the CX and the CT rotors are now out of correspondence. The error signal is amplified by the servoamplifier and applied to the servomotor. The servomotor turns the load, and through a mechanical linkage called RESPONSE, also turns the rotor of the CT. The servomotor turns the rotor of the CT so that it is once again in correspondence with the rotor of the CX, the error signal drops to zero volts, and the system comes to a stop.

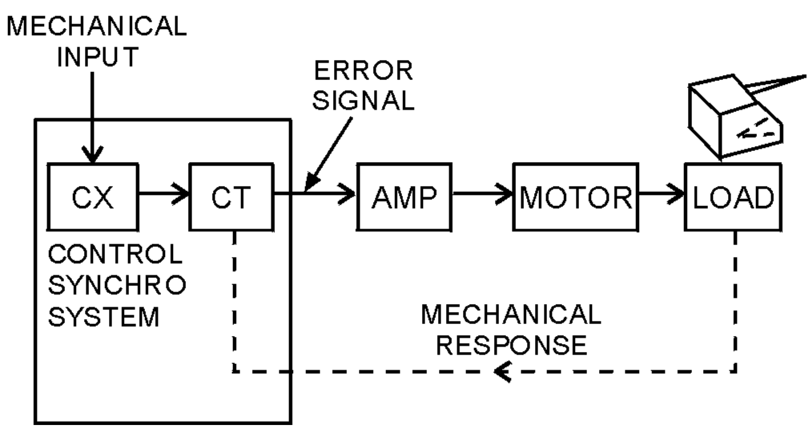


Figure 1-28.—A positioning servo system using a control synchro system.

Q-44. When a CT is on electrical zero, what is the relationship between its rotor and the S2 winding?

Q-45. What is the amplitude and voltage induced into the rotor when the CX is turned 90º while the CT remains on electrical zero?

Q-46. What is the name given to the electrical output of a CT?

Q-47. In a control synchro system, when is the output of the CT reduced to zero?

SYNCHRO CAPACITORS

As we stated earlier, the speed and accuracy of data transmission are most important. With the use of more complex synchros, like the differential and the control transformer, the accuracy of the synchro systems may be affected. The following discussion will deal with how complex synchros affect the accuracy of synchro systems and what can be done to keep this accuracy as high as possible. Synchro capacitors play a major role in maintaining a high degree of accuracy in synchro systems.

When a torque transmitter is connected to a torque receiver (TX-TR), very little, if any, current flows in the stators when the rotors are in correspondence. This is because the voltages induced in the TR windings almost exactly balance out the voltages induced in the TX windings. As a result, the TR is very sensitive to small changes in the position of the TX rotor, causing the TR to follow the TX with a high degree of accuracy.

When a synchro system contains differential synchros (TDX or CDX), the stator currents at correspondence are greater than they are in a single TX-TR system. The reason is the step-up turns ratio between the stator and rotor in the differential synchro.

In a synchro system that uses a CT, stator current at correspondence is also greater than in a TX-TR system. In this case, however, this reason is that the CT rotor is not energized and as a result no voltage is induced in the stator to oppose the voltage in the transmitter stator. The overall effect of this increase in stator current is to reduce the accuracy of the system. To maintain high accuracy in a synchro system containing either differential units or CTs, the stator currents must be kept to a minimum. This is done by connecting synchro capacitors in the circuit.

To understand the operation of a synchro capacitor and how it reduces current drain on the transmitter requires a recollection of the voltage and current relationships in inductive and capacitive circuits. As you learned in module 2 of this series, current lags voltage by 90º in a purely inductive circuit. You also know that an ideal inductor is impossible to make because there is always resistance present. Therefore, an inductor has a combination of inductive reactance and resistance. Since current and voltage are always in phase in a resistive circuit and 90º out of phase in an inductive circuit, we can say that there are two currents in an inductor-the loss current, which is the resistive (in-phase) current, and the magnetizing current, which is the inductive (out-of-phase) current. It is this magnetizing current that we would like to eliminate in the stator coils of the TDX, CDX, and CT because it makes up most of the line current.

Keeping in mind that current leads voltage by 90º in a capacitive circuit, let's see what happens to magnetizing current when a capacitor is added to the circuit.

Suppose a capacitor is hooked up across one of the stator coils of a TDX and its capacitance is adjusted so that its reactance equals the reactance of the coil. Since the two reactances are equal, the current they draw from the line must also be equal. However, these currents are going to be 180º out of phase, because the current in the coil lags the line voltage, while the capacitor's current leads it. Since the two currents are equal in magnitude but opposite in phase, they cancel. The total line current is reduced by this effect and, if a capacitor is placed across each coil in the TDX, the line current decreases even further. This, in effect, increases torque in synchro systems near the point of correspondence and, therefore, increases overall system accuracy.

Connecting capacitors across individual stator windings is impractical because it requires that the stator winding's common connection be outside the synchro. Since this is not done with synchros, another method has been devised to connect up the capacitors which works just as well. This method is shown in figure 1-29.

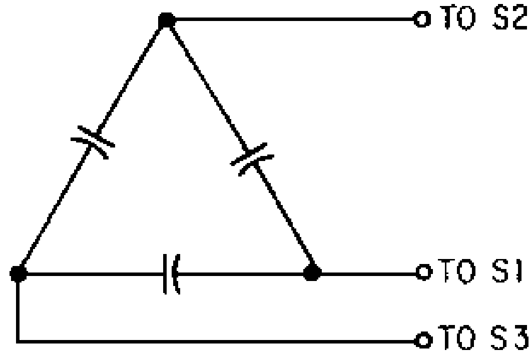


Figure 1-29.—The synchro capacitor.

The three delta-connected capacitors, shown in figure 1-29, usually come as a unit mounted in a case with three external connections. The entire unit is called a SYNCHRO CAPACITOR. The synchro capacitor is made in many sizes to meet the requirements of all sizes of standard differentials and control transformers. The synchro capacitor is rated by its total capacity, which is the sum of the individual capacities in the unit.

Figure 1-30 shows how a synchro capacitor affects the operation of a control synchro system. In this figure, the capacitor is placed between the CX and the CT. Two current meters are also placed in the circuit to show the effect the capacitor has on stator current. The meter connected between the capacitor and the CT reads normal stator current, 32 milliamperes (mA). This current would normally flow in the stator of the CX if the synchro capacitor were not connected. The other meter reads 10 mA, which is what is left of the original stator current after the magnetizing current has been canceled by the synchro capacitor. By reducing the current drain on the transmitter, the sensitivity and accuracy of the system increase.

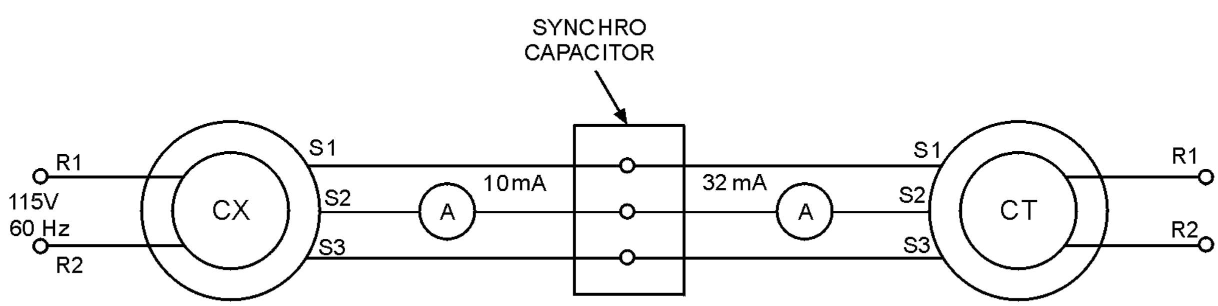


Figure 1-30.—The use of a synchro capacitor with a CT.

Figure 1-31 shows another application of a synchro capacitor; this time in a differential system in this circuit the capacitor is placed between a TX and a TDX. The meter readings show the same comparison between currents as in the previous paragraph. The only significant difference between this circuit and the one in figure 1-30 is that the differential draws more stator current than the CT.

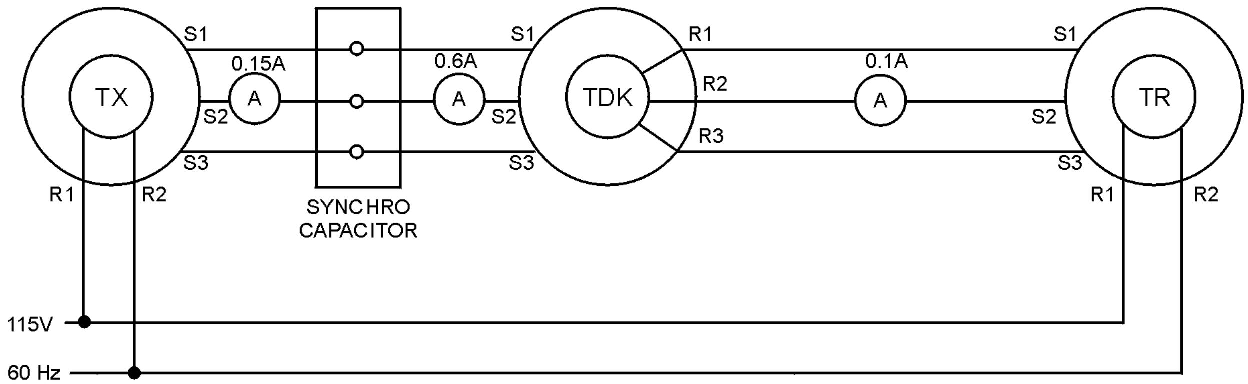


Figure 1-31.—The use of a synchro capacitor with a TDX.

Some synchro systems contain a differential and a control transformer, as illustrated in figure 1-32. In this figure, there are large stator currents flowing in the CX, since it supplies all the losses as well as the magnetizing current for both synchros. Two meters are placed in the circuit to show the value of stator current for the CDX and CT. Another meter is placed in series with the ac excitation voltage to show the amount of current being drawn from the ac line is 0.9 ampere.

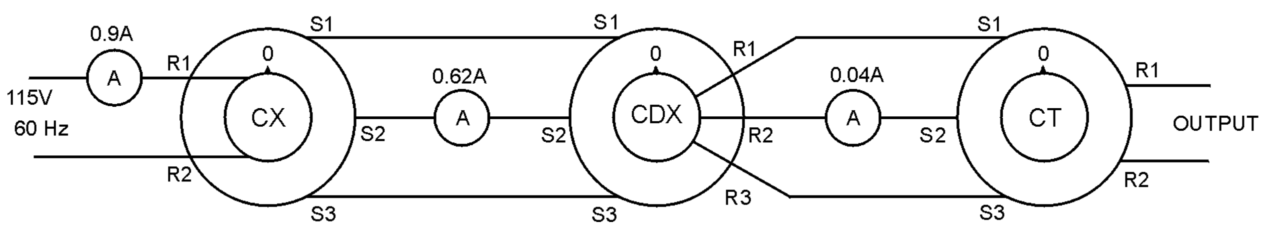


Figure 1-32.—Synchro current in a control synchro system using a CDX and a CT.

Adding synchro capacitors to this system, as shown in figure 1-33, greatly reduces the stator currents and improves the efficiency of the system. Also, notice that the line current is reduced from 0.9 ampere in figure 1-32 to 0.65 ampere in figure 1-33.

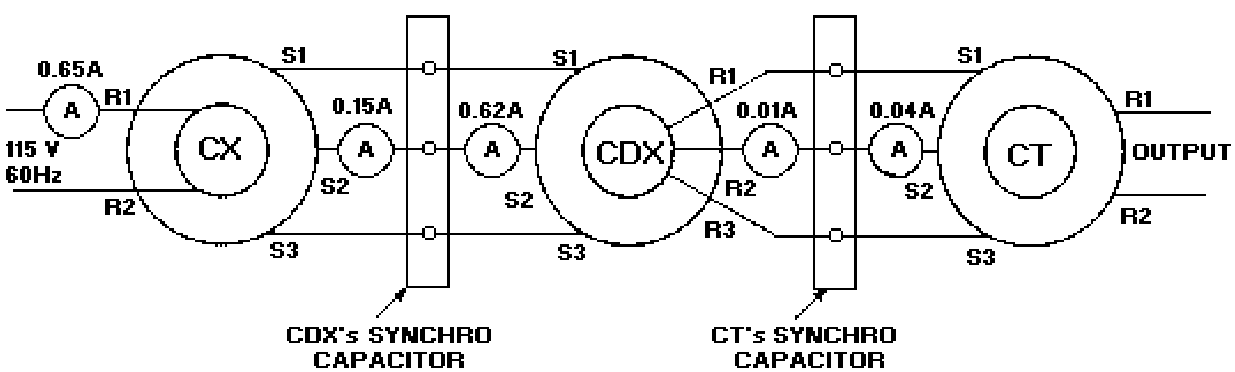


Figure 1-33.—The effects of synchro capacitors in a control synchro system using a CDX and a CT.

When a synchro capacitor is used, it is always placed physically close to the differential or control transformer whose current it corrects. This is done to keep the connections as short as possible, because high currents in long leads increase the transmitter load and reduce the accuracy of the system.

We must stress that the synchro capacitor should never be used in a simple transmitter-receiver system. This is because stator currents in this system are zero at correspondence and the addition of a synchro capacitor would only increase the stator current and throw the system out of balance.

Q-48. What is the purpose of the synchro capacitor?

Q-49. What type of synchros usually require the use of synchro capacitors?

Q-50. What type of current is eliminated by synchro capacitors?

Q-51. How are synchro capacitors connected in a circuit?

Q-52. Why are synchro capacitors placed physically close to differentials transmitters and CTs?

MULTISPEED SYNCHRO SYSTEMS

The data to be transmitted is another important factor that we must consider when we discuss the accuracy of a synchro system. If this data covers a wide range of values, the basic synchro system is unable to detect any small changes in the data. When this happens, the accuracy of the system decreases. Because of this difficulty, multispeed synchro systems were developed. They handle this type of data very effectively and still maintain a high degree of accuracy.

Multispeed synchro systems use more than one speed of data transmission. The speed of data transmission is, simply, the number of times a synchro transmitter rotor must turn to transmit a full range of values. For example, a system in which the rotors of synchro devices turn in unison with their input and output shafts is commonly called a 1-speed data transmission system. In this system, the transmitter's rotor is geared so that 1 revolution of the rotor corresponds to 1 revolution of the input. Until now, the discussion of synchro systems has dealt exclusively with this 1-speed system.

In a 36-speed data transmission system, the rotor of the synchro transmitter is geared to turn through 36 revolutions for 1 revolution of its input. Units transmitting data at one speed are frequently called 1-speed synchros; a unit transmitting data at 36-speed would be a 36-speed synchro, and so forth.

It is quite common in synchro systems to transmit the same data at two different speeds. For example, ship's course information is usually transmitted to other locations on a ship at 1-speed and 36-speed. A system in which data is transmitted at two different speeds is called a dual- or double-speed system. Sometimes a dual-speed system will be referred to by the speeds involved, for example a 1- and 36-speed system.

In summary, the speed of data transmission is referred to as 1-speed, 2-speed, 36-speed, or some other definite numerical ratio. To indicate the number of different speeds at which data is transmitted, refer to the system as being a single-speed, dual-speed, or tri-speed synchro system.

SINGLE-SPEED SYNCHRO SYSTEM

If the data to be transmitted covers only a small range of values, a single-speed system is normally accurate enough. However, in applications where the data covers a wide range of values and the accuracy of the system is most important, the 1-speed system is not adequate enough and must be replaced by a more suitable system. Increasing the speed of a single-speed system from 1-speed to 36-speed provides greater accuracy, but the self-synchronous feature of the 1-speed system is lost. If primary power is interrupted in a 36-speed system and the transmitter is turned before power is reapplied, the synchros could realign themselves in an erroneous position. The number of positions in which the transmitter and receiver rotors can correspond is the same as the transmission speed. Thus, in the 36-speed system, there are 35 incorrect positions and only 1 correct position of correspondence.

For accurate transmission of data over a wide range of values without the loss of self-synchronous operation, multispeed synchro systems must be used. Multispeed synchro systems use more than one speed of data transmission and, therefore, require more than one output shaft.

DUAL-SPEED SYNCHRO SYSTEM

A basic dual-speed synchro system consists of two transmitters and two receivers, as shown in figure 1-34. One transmitter receives the external input to the system and, through a network of gears, passes the effects of the external input to the second transmitter. The gear ratio between these two transmitters determines the two specific speeds the system will use to transmit the input data. The two speeds of this system are often referred to as fast and slow, high and low, or more often as fine and coarse.

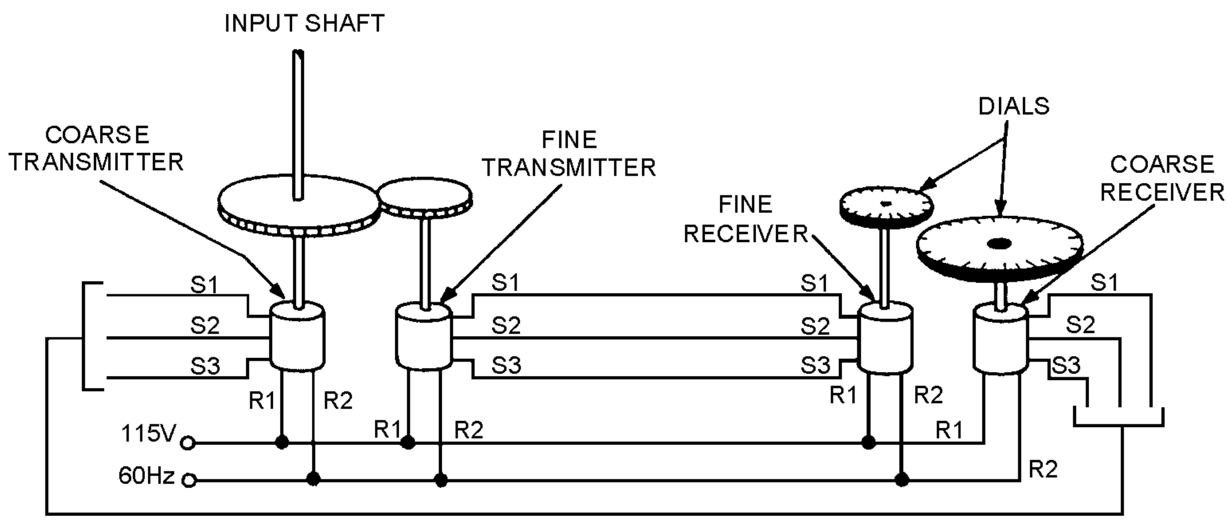


Figure 1-34.—Dual-speed synchro system.

If, for example, the gear ratio between the two transmitters is 36 to 1, 1 revolution of the rotor of the first transmitter causes 36 revolutions of the rotor of the second transmitter. The first transmitter-the one that accepts the external input-can be called the coarse transmitter, and the second one can be called the fine transmitter. Representative speeds include 1 and 36, 2 and 36, and 2 and 72.

The output of each transmitter is passed through standard synchro connections to a receiver. One receiver receives the coarse signal and the other one receives the fine signal. The two receivers may or may not be connected by a network of gears similar to the network between the two transmitters. In some dual-speed applications, a double receiver is used instead of two individual receivers.

The double receiver (fig. 1-35) consists of a coarse and a fine receiver enclosed in a common housing. It has a two-shaft output one inside the other. The coarse and fine receivers may be likened to the hour and minute hands of a clock. The coarse receiver corresponds to the hour hand, and the fine receiver corresponds to the minute hand. This double receiver has the advantage of requiring less space than two single receivers. However, it also has a disadvantage — when one receiver goes bad, both must be replaced.

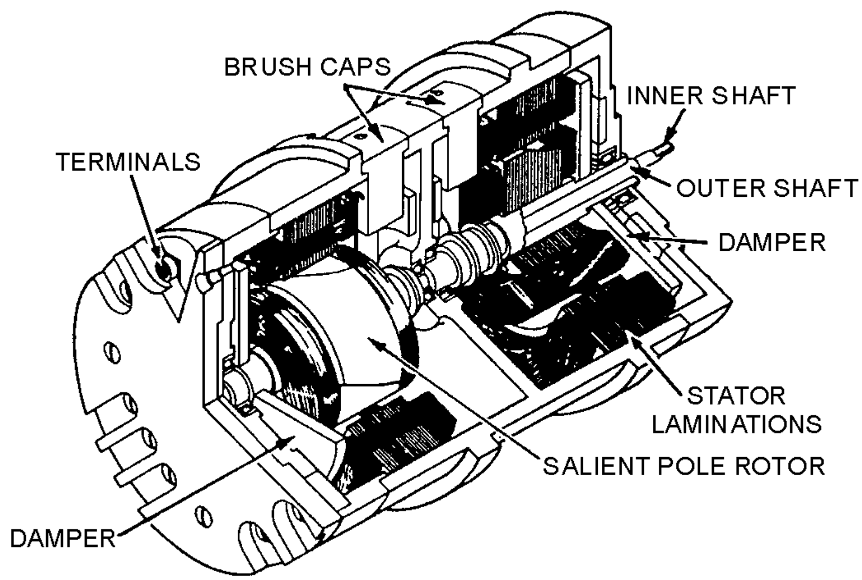


Figure 1-35.—Cutaway view of a double receiver.

In the dual-speed synchro system, data is transmitted by the coarse transmitter, while the system is far out of correspondence and then is shifted to the fine transmitter as the system approaches correspondence. This shifting from coarse to fine control is done automatically when the fine error signal exceeds the coarse error signal. The fine synchro transmitter then drives the system to the point of correspondence.

When the dual-speed synchro system includes control transformers, there is always the possibility of a 180º error being present in the system. The reason is the rotor of the CT is not energized and its error-voltage output is zero both at its proper position and also at a point 180º away from that position. To prevent the CT from locking 180º out of phase with the rest of the system, a low voltage is applied across the rotor terminals of the coarse CT as shown in figure 1-36.

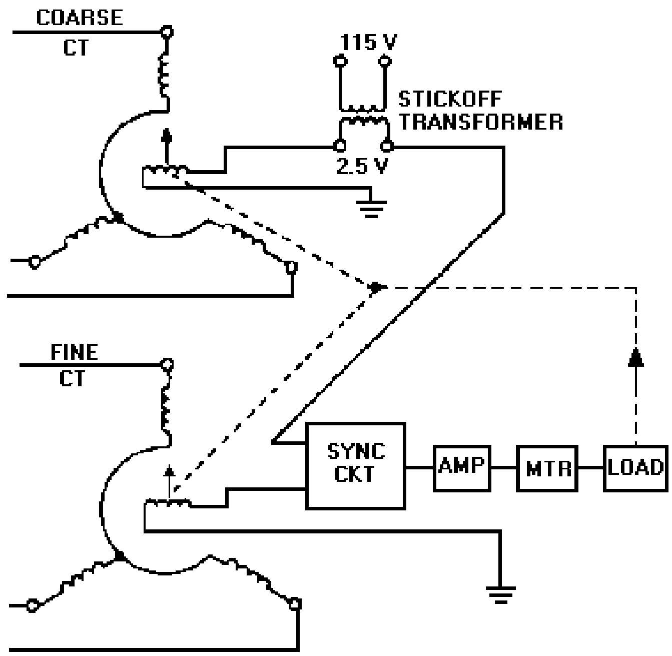


Figure 1-36.—Dual-speed synchro system using a stickoff transformer.

This voltage is normally about 2.5 volts and is commonly termed "stickoff" voltage. It is obtained from the secondary of a small transformer. The voltage induced in the secondary of the transformer shifts the 0º position of the coarse CT To reestablish a new 0º position, the stator of the coarse CT must be turned through an angle that induces an opposing 2.5 volts in the rotor to cancel the stickoff voltage. Therefore, at 0º the two voltages cancel and no input exists to drive the servo amplifier. Should the rotor of the CT stop at 180º, the same 2.5 volts would be induced in the rotor. However, it would be in phase with the stickoff voltage and no cancellation would occur. The end result is an error signal at 180º that drives the dual-speed synchro system out of any false synchronizations.

TRI-SPEED SYNCHRO SYSTEM

The advent of long-range missiles and high-speed aircraft has brought about the need for accurately transmitting very large quantities. This is best done by a tri-speed synchro system, which transmits data at three different speeds. These speeds are sometimes referred to as coarse, medium or intermediate, and fine. A typical weapons systems, for example, might transmit range in miles, thousands of yards, and hundreds of yards. By providing this range in three different scales, greater accuracy is obtained than would be possible with a dual-speed system. Representative speeds for tri-speed systems include 1, 36, and 180; 1, 36, and 360; and 1, 18, and 648.

Q-53. What is the name given to the synchro system that transmits data at two different speeds?

Q-54. What is the main reason for using a multispeed synchro system instead of a single-speed synchro system?

Q-55. In a dual-speed synchro system what determines the two specific speeds at which the data will be transmitted?

Q-56. What type of synchro system is used to transmit very large quantities?

Q-57. What is the disadvantage of using a double receiver instead of two individual receivers?

Q-58. What is the purpose of "stickoff voltage"?

ZEROING SYNCHROS

If synchros are to work properly in a system, they must be connected and aligned correctly with respect to each other and to the other devices with which they are used. The reference point for alignment of all synchro units is ELECTRICAL ZERO. The mechanical reference point for the units connected to the synchros depends upon the particular application of the synchro system. Whatever the application, the electrical and mechanical reference points must be aligned with each other. The mechanical position is usually set first, and then the synchro device is aligned to electrical zero.

There are various methods for zeroing synchros. Some of the more common zeroing methods are the voltmeter, the electrical-lock, and the synchro-tester methods. The method used depends upon the facilities and tools available and how the synchros are connected in the system. Also, the method for zeroing a unit whose rotor or stator is not free to turn may differ from the procedure for zeroing a similar unit whose rotor or stator is free to turn.

VOLTMETER METHOD

The most accurate method of zeroing a synchro is the ac voltmeter method. The procedure and the test circuit configuration for this method vary somewhat, depending upon which type of synchro is to be zeroed. Transmitters and receivers, differentials, and control transformers each require different test circuit configurations.

Regardless of the synchro to be zeroed, there are two major steps in each procedure. The first step is the coarse or approximate setting. The second step is the fine setting. The coarse setting ensures the device is zeroed on the 0º position rather than the 180º position. Many synchro units are marked in such a manner that the coarse setting may be approximated physically by aligning two marks on the synchro. On standard synchros, this setting is indicated by an arrow stamped on the frame and a line marked on the shaft, as shown in figure 1-37. The fine setting is where the synchro is precisely set on 0º.

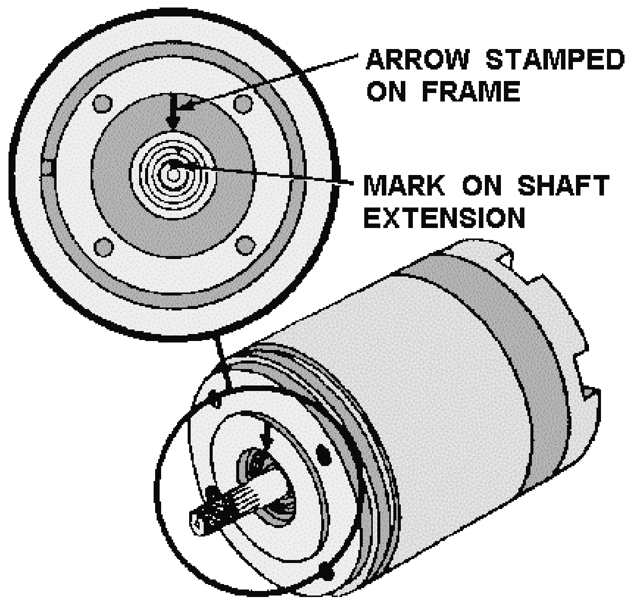


Figure 1-37.—Coarse electrical zero markings.

For the ac voltmeter method to be as accurate as possible, an electronic or precision voltmeter having 0- to 250-volt and a 0- to 5-volt ranges should be used. On the low scale this meter should also be able to measure voltages as low as 0.1 volt.

Q-59. What is the reference point for alignment of all synchro units?

Q-60. What is the most accurate method of zeroing a synchro?

Q-61. What is the purpose of the coarse setting of a synchro?

Zeroing Transmitters and Receivers (Voltmeter Method)

Since the TX, CX, and TR are functionally and physically similar, they can be zeroed in the same manner. For the TX and CX to be properly zeroed, electrical zero voltages (S2 = 52V; S1 and S3 = 26V) must exist across the stator winding when the rotor of the transmitter is set to 0º or its mechanical reference position. The synchro receiver (TR) is properly zeroed when the device it actuates assumes its zero or mechanical reference position while electrical zero voltages (S2 = 52V; S1 and S3 = 26V) exist across its stator windings. The following is a step-by-step procedure used to zero the TX, CX, and TR.

1. Carefully set the unit (antenna, gun mount, director, etc.) whose position the CX or TX transmits, accurately on 0º or on its reference position. In the case of the TR, deenergize the circuit and disconnect the stator leads before setting its rotor on zero or to its reference position. The rotor may need to be secured in this position; taping the dial to the frame is usually sufficient.

2. Deenergize the synchro circuit and disconnect the stator leads. NOTE: Many synchro systems are energized by individual switches. Therefore, be sure that the synchro power is off before working on the connections.Set the voltmeter to its 0- to 250-volt scale and connect it into the circuit as shown in view A of figure 1-38.

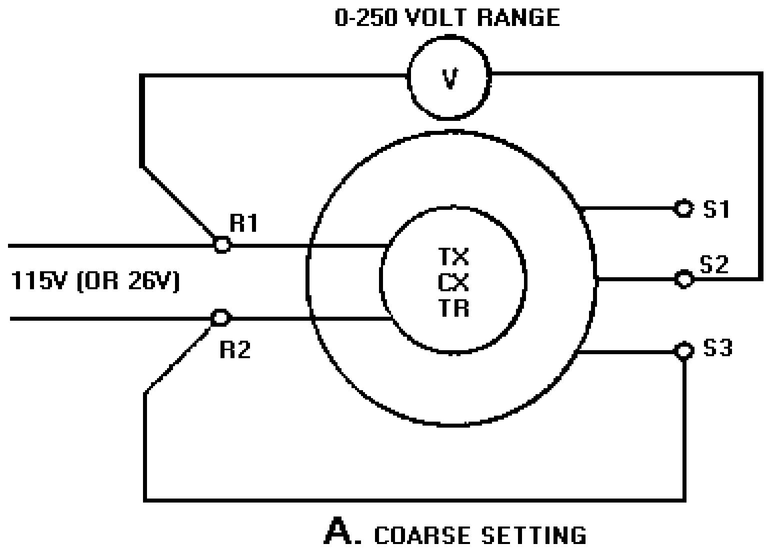


Figure 1-38A.—Zeroing a transmitter or receiver by the voltmeter method.

3. Energize the synchro circuit and turn the stator until the meter reads about 37 volts (15 volts for a 26-volt synchro). Should the voltmeter read approximately 193 volts (115 volts + 78 volts = 193 volts), the rotor is at 180º. Turn it through a half revolution to bring it back to 0º. If you cannot obtain the desired 37 (or 15) volts, use the lowest reading you can take with the meter. This is the coarse setting and places the synchro approximately on electrical zero.

4. Deenergize the synchro circuit and connect the meter as shown in view B. Start with a high scale on the meter and work down to the 0- to 5-volt scale to protect the meter movement.

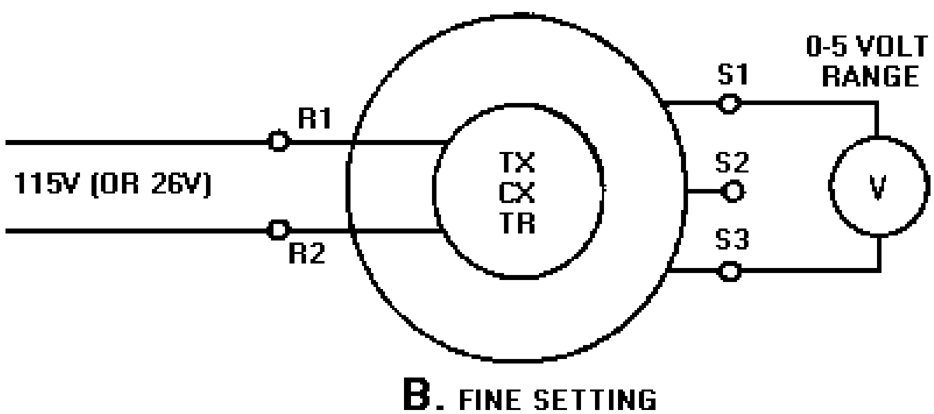


Figure 1-38B.—Zeroing a transmitter or receiver by the voltmeter method.

5. Reenergize the synchro circuit and adjust the stator for a zero or minimum voltage reading. This is the fine electrical zero position of the synchro.

When you have reconnected a TX and a TR into a system after zeroing them, you can perform a simple check on the system to see if they are accurately on electrical zero. First, place the transmitter on 0º. When the system reaches the point of correspondence, the induced voltages in the S1 and S3 stator windings of both synchros should be equal. Since the terminals of S1 and S3 are at equal potential, a jumper between these terminals at the TR should not affect the circuit. If, however, the TR rotor moves when you connect a jumper, there is a slight voltage difference between S1 and S3. This voltage difference indicates that the transmitter is not on electrical zero. If this is the case, recheck the transmitter for electrical zero.

Zeroing Differential Synchros (Voltmeter Method)

A differential synchro is zeroed when it can be inserted into a system without introducing any change. If a differential synchro requires zeroing, use the following voltmeter method:

1. Carefully and accurately set the unit whose position the CDX or TDX transmits on zero or on its reference position. In the case of the TDR, deenergize the circuit and disconnect all leads before setting its rotor to 0º or to its reference position. You may need to secure the rotor in this position; taping the dial to the frame is usually sufficient.

2. Deenergize the circuit and disconnect all leads on the differential except leads S2 and S3 when you use the 78-volt (10.2 volts for 26-volt units) supply from the transmitting unit to zero the differential. Set the voltmeter to its 0- to 250-volt scale and connect it as shown in view A of figure 1-39. If the 78-volts is not available from the transmitter or from an auto transformer, you may use a 115-volt source instead. If you use 115 volts instead of 78 volts, do not leave the synchro connected for more than 2 minutes or it will overheat and may become permanently damaged.

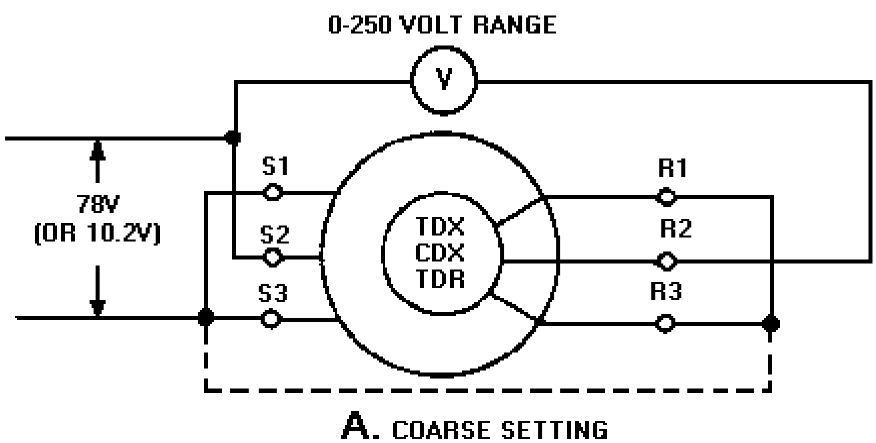


Figure 1-39A.—Zeroing differential synchros by the voltmeter method.

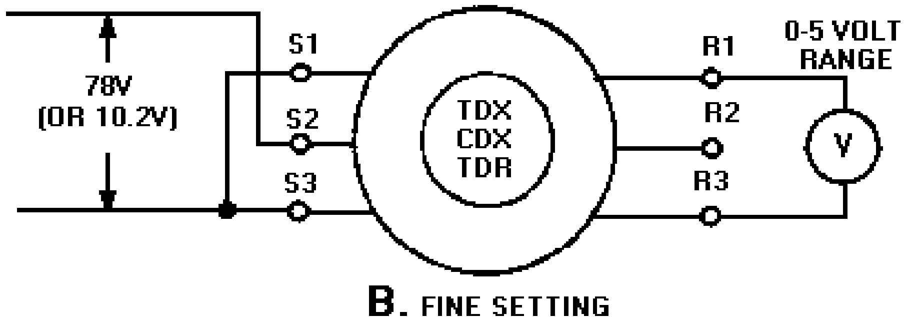


Figure 1-39B.—Zeroing differential synchros by the voltmeter method.

3. Energize the circuit, unclamp the differential's stator, and turn it until the meter reads minimum. The differential is now approximately on electrical zero. Deenergize the circuit and reconnect it as shown in view B.

4. Reenergize the circuit. Start with a high scale on the meter and work down to the 0- to 5-volt scale to protect the meter movement. At the same time, turn the differential's stator until you obtain a zero or minimum voltage reading. Clamp the differential stator in position, ensuring the voltage reading does not change. This is the fine electrical zero position of the differential.

Zeroing a Control Transformer (Voltmeter Method)

Two conditions must exist for a control transformer (CT) to be on electrical zero. First, its rotor voltage must be minimum when electrical zero voltages (S2 = 52 volts; S1 and S3 = 26 volts) are applied to its stator. Second, turning the shaft of the CT slightly counterclockwise should produce a voltage across its rotor in phase with the rotor voltage of the CX or TX supplying excitation to its stator. To zero a CT (using 78 volts from its transmitter) by the voltmeter method, use the following procedure:

1. Set the mechanism that drives the CT rotor to zero or to its reference position. Also, set the transmitter (CX or TX) that is connected to the CT to zero or its reference position.

2. Check to ensure there is zero volts between S1 and S3 and 78 volts between S2 and S3. If you cannot obtain these voltages, you will need to rezero the transmitter. NOTE: If you cannot use the 78 volts from the transmitter circuit and, an auto transformer is not available, you may use a 115-volt source. <u>If you use 115 volts instead of 78 volts, do not energize the CT for more than 2 minutes because it will overheat and may become permanently damaged.</u>

3. Deenergize the circuit and connect the circuit as shown in view A of figure 1-40. To obtain the 78 volts required to zero the CT, leave the S1 lead on, disconnect the S3 lead on the CT, and put the S2 lead (from the CX) on S3. This is necessary since 78 volts exists only between S1 and S2 or S2 and S3 on a properly zeroed CX. Now energize the circuit and turn the stator of the CT to obtain a minimum reading on the 250-volt scale. This is the coarse or approximate zero setting of the CT.

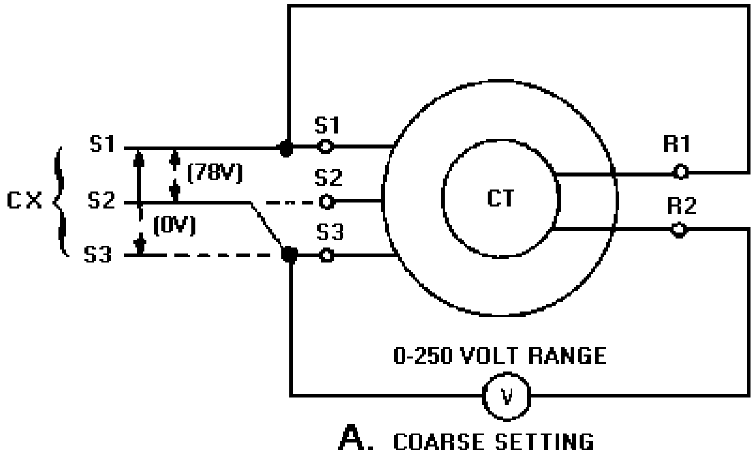


Figure 1-40A.—Zeroing a control transformer by the voltmeter method.

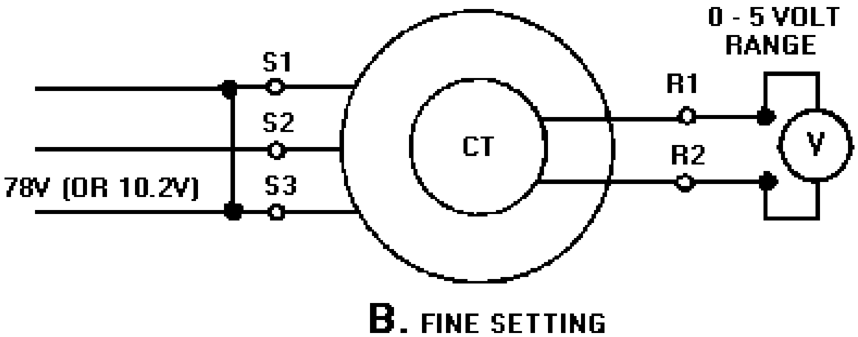


Figure 1-40B.—Zeroing a control transformer by the voltmeter method.

4. Deenergize the circuit, reconnect the S1, S2, and S3 leads back to their original positions, and then connect the circuit as shown in view B.

5. Reenergize the circuit. Start with a high scale on the meter and work down to the 0- to 5-volts scale to protect the meter movement. At the same time, turn the stator of the CT to obtain a zero or minimum reading on the meter. Clamp down the CT stator, ensuring the reading does not change. This is the fine electrical zero position of the CT.

Zeroing Multispeed Synchro Systems.

If multispeed synchro systems are used to accurately transmit data, the synchros within the systems must be zeroed together. This is necessary because these synchros require a common electrical zero to function properly in the system.

First, establish the zero or reference position for the unit whose position the system transmits. Then, zero the most significant synchro in the system first, working down to the least significant. For example, zero the coarse synchro, then the medium synchro, and finally the fine synchro. When you zero those synchros, consider each synchro as an individual unit and zero it accordingly.

Q-62. When is a synchro receiver (TR) properly zeroed?

Q-63. What should a voltmeter read when a TX is set on coarse zero?

Q-64. What precaution should you take when you use 115 volts to zero a differential?

Q-65. Why should a synchro be rechecked for zero after it is clamped down?

Q-66. What is the output voltage of a CT when it is set on electrical zero?

Q-67. When you zero a multispeed synchro system which synchro should you zero first?

ELECTRICAL LOCK METHOD

The electrical lock method, although not as accurate as the voltmeter method, is perhaps tile fastest method of zeroing synchros. However, this method can be used only if the rotors of the units to be zeroed are free to turn and the lead connections are accessible. For this reason, this method is usually used on the TR because, unlike transmitters, the TR shaft is free to turn.

To zero a synchro by the electrical lock method, deenergize the unit, connect the leads as shown in figure 1-41, and apply power. The synchro rotor will then quickly snap to the electrical zero position and lock. If the indicating device connected to the synchro shaft does not point to zero, loosen the synchro in its mounting and rotate it until the zero position of the indicator corresponds with the electrical zero of the synchro. As we stated previously, <u>you may use 115 volts as the power source instead of 78 volts, provided you do not leave the unit connected for more than 2 minutes.</u>

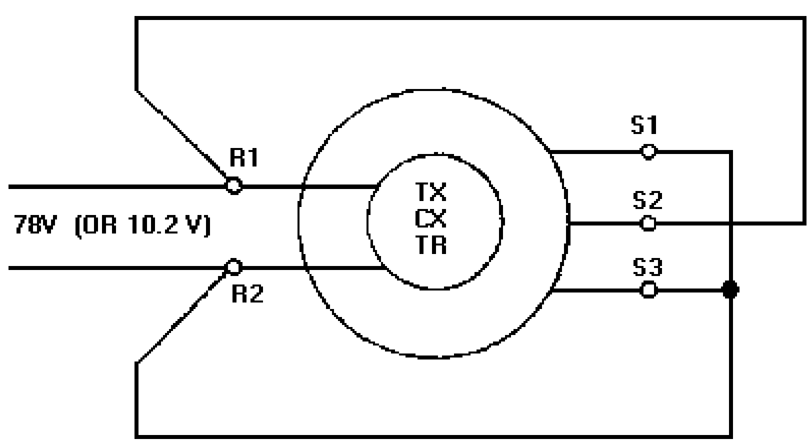


Figure 1-41.—Zeroing a synchro by the electrical lock method.

SYNCHRO TESTERS

Two types of synchro testers are shown in figure 1-42, view (A) and view (B). Each is nothing more than a synchro receiver on which a calibrated dial is mounted.

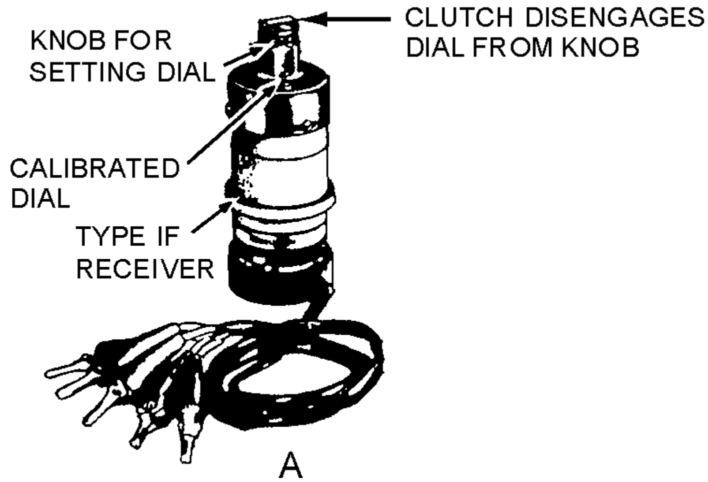


Figure 1-42A.—Synchro Testers.

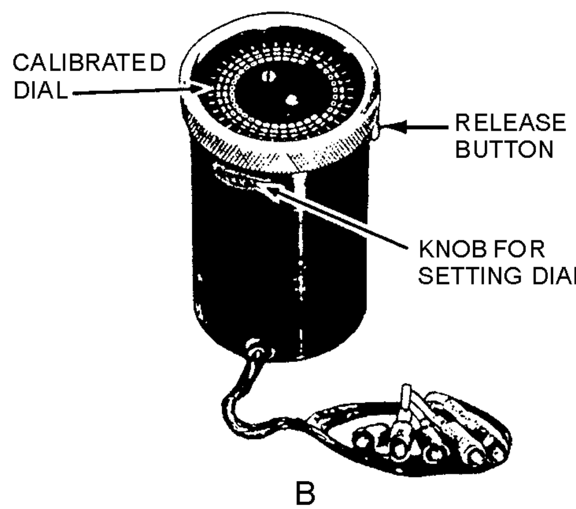


Figure 1-42B.—Synchro Testers.

These testers are used primarily for locating defective synchros. Although they do provide a method for zeroing synchros, they should not be relied on without question. It is possible for the calibrated dial to slip from its proper position, and since the dial is graduated only every 10º, it is difficult to read small angles with accuracy. Therefore, the synchro tester method of zeroing synchros is potentially less accurate than those previously described. To zero a TX, CX or TR using a synchro tester, use the following procedure:

1. Connect the synchro tester as shown in figure 1-43.

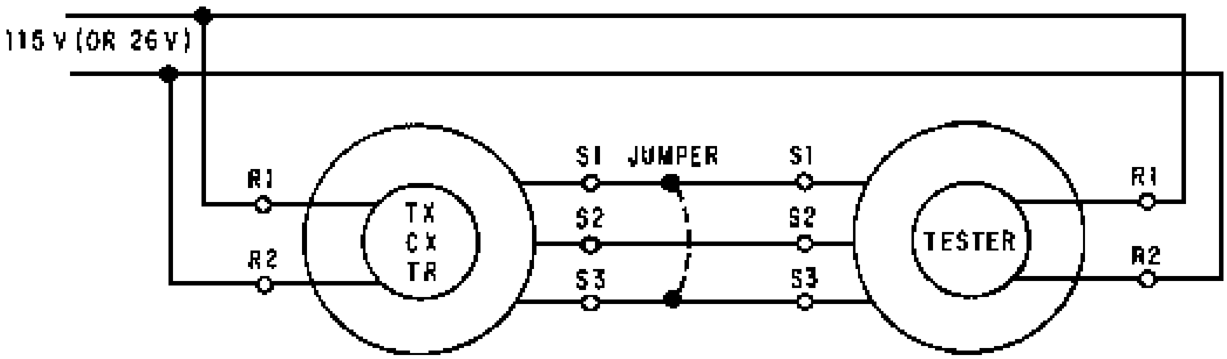


Figure 1-43.—Zeroing a synchro using a synchro tester.

2. Set the unit whose position the TX or CX transmits accurately on zero or on its reference position. In the case of the TR, set its rotor to zero or to its reference position.

3. Turn the stator of the synchro being zeroed until the synchro tester dial reads 0º. The synchro is now approximately on electrically zero.

4. Momentarily short S1 to S3 as shown. If the synchro tester dial moves when S1 is shorted to S3, the synchro is not zeroed. Check the tester dial to ensure it has not slipped. If the tester dial has not slipped, move the synchro stator until there is no movement when S1 and S3 are shorted. This is the electrical zero position of the synchro being aligned.

Q-68. What method of zeroing a synchro is perhaps the fastest but NOT necessarily the most accurate?

Q-69. What restrictions are placed on the use of the electrical lock method?

Q-70. When you zero a synchro with a synchro tester, what is indicated by a jump in the synchro tester's dial when the S1 and S3 leads are momentarily shorted?

TROUBLESHOOTING SYNCHRO SYSTEMS

One of your duties in the Navy is to keep the synchro systems in your equipment in good working order. Therefore, it is essential that you become familiar with the details of synchro maintenance and repair.

First, let's consider some of the more common problem areas you should avoid when working with synchros. As with any piece of electrical or electronic equipment, if it works—leave it alone. Do not attempt to zero a synchro system that is already zeroed just because you want to practice. More often than not, the system will end up more out of alignment than it was before you attempted to rezero it. Do not attempt to take a synchro apart, even if it is defective. A synchro is a piece of precision equipment that requires special equipment and techniques for disassembly. Disassembly should be done only by qualified technicians in authorized repair shops. A synchro, unlike an electric motor, does not require periodic lubrication. Therefore, never attempt to lubricate a synchro. Synchros also require careful handling. Never force a synchro into place, never use pliers on the threaded shaft, and never force a gear or dial onto the shaft. Finally, never connect equipment that is not related to the synchro system to the primary excitation bus. This will cause the system to show all the symptoms of a shorted rotor when the equipment is turned on; but, the system will check out good when the equipment is off.

Trouble in a synchro system that has been in operation for some time is usually one of two types. First, the interconnecting synchro wiring often passes through a number of switches; at these points opens, shorts, or grounds may occur. You will be expected to trace down these troubles with an ohmmeter. You can find an open easily by checking for continuity between two points. Similarly, you can find a ground by checking the resistance between the suspected point and ground. A reading of zero ohms means that the point in question is grounded. Secondly, the synchro itself may become defective, due to opens and shorts in the windings, bad bearings, worn slip rings, or dirty brushes. You can do nothing about these defects except replace the synchro.

Troubles in new and modified synchro systems are most often because of (1) improper wiring and (2) misalignment caused by synchros not being zeroed. You are responsible for finding and correcting these troubles. You can check for improper wiring with an ohmmeter by making a point-to-point continuity and resistance check. You can correct misalignment of a synchro system by rezeroing the entire system.

TROUBLE INDICATORS

When trouble occurs in an electronic installation that contains a large number of synchro systems, it may be very difficult to isolate the trouble to one particular system. Since it is vital that maintenance personnel locate the point of trouble and fix it in as short a time as possible, indicators, which aid in locating the trouble quickly, are included in the equipment. These indicators are usually signal lights, mounted on a central control board and connected to the different synchro systems. When trouble occurs in a synchro system, the signal light connected to it may either light or flash. Maintenance personnel identify the defective system by reading the name or number adjacent to the light.

Signal lights indicate either overload conditions or blown fuses. Overload indicators are usually placed in the stator circuit of a torque synchro system because the stator circuit gives a better indication of mechanical loading than does the current in the rotor circuit. One version of this type of indicator, as shown in figure 1-44, consists of a neon lamp connected across the stator leads of a synchro system by two transformers. The primaries, consisting of a few turns of heavy wire, are in series with two of the stator leads; the secondaries, consisting of many turns of fine wire, are in series with the lamp. The turns ratios are designed so that when excess current flows through the stator windings, the neon lamp lights. For example, when the difference in rotor positions exceeds about 18º, the lamp lights, indicating that the load on the motor shaft is excessive.

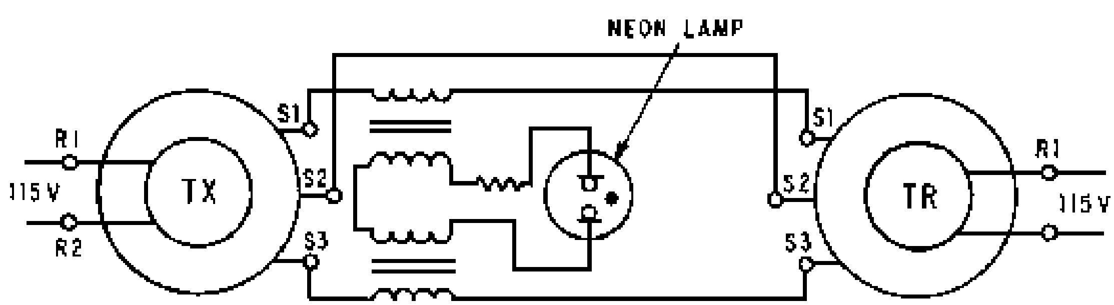


Figure 1-44.—Overload stator current indicator.

Blown fuse indicators are front panel lights which light when a protective fuse in series with the rotor blows. Figure 1-45 shows a typical blown fuse indicator. If excessive current flows in the rotor windings of this circuit because of a short or severe mechanical overload, one of the fuses will blow and the neon lamp across the fuse will light.

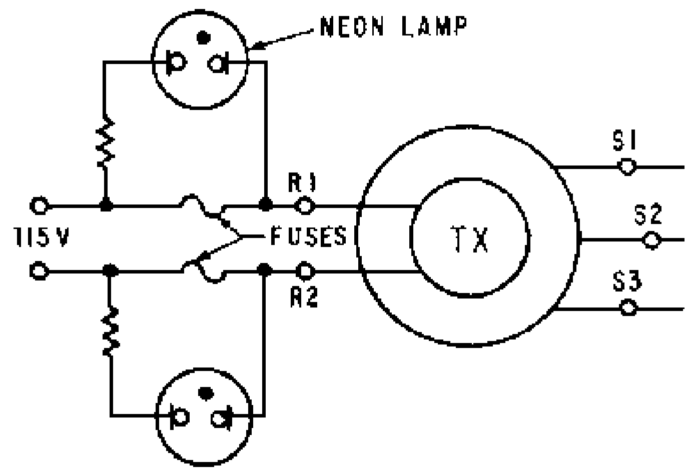


Figure 1-45.—Simple blown fuse indicator.

Another type of blown fuse indicator uses a small transformer having two identical primaries and a secondary connected, as shown in figure 1-46. With both fuses closed, equal currents flow through the primaries. This induces mutually canceling voltages in the secondary. If a fuse blows, the induced voltage from just one primary is present in the secondary, and the lamp lights.

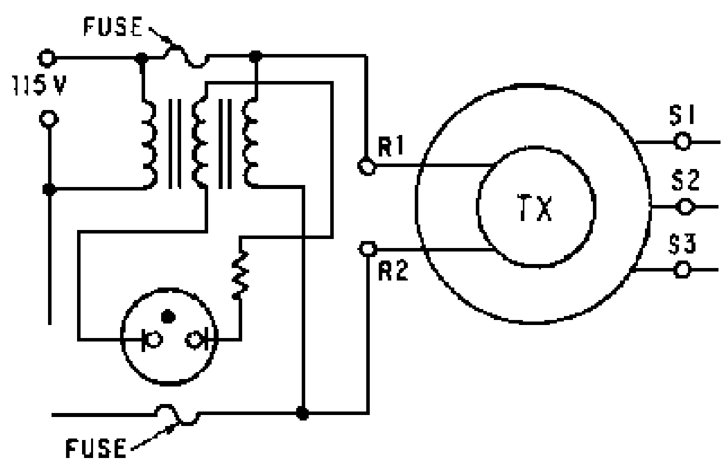


Figure 1-46.—Blown fuse indicator requiring only one lamp.

SYMPTOMS AND CAUSES

To help the technician further isolate synchro problems, many manufacturers furnish tables of trouble symptoms and probable causes with their equipment. These tables are a valuable aid in isolating trouble areas quickly. Tables 1-2 through 1-7 summarize, for a simple TX-TR system, some typical trouble symptoms and their probable causes. Keep in mind, if two or more receivers are connected to one transmitter, similar symptoms occur. However, if all the receivers act up, the trouble is usually in the transmitter or main bus. If the trouble appears in one receiver only, check the unit and its connections.

The angles shown in these tables do not apply to systems using differentials, or to systems whose units are incorrectly zeroed.

Table 1-2.—General Symptoms

<table>
<tr><td colspan="2">Preliminary Actions: Be sure TR is not jammed physically. Turn TX slowly in one direction and observe TR.</td></tr>
<tr><th>SYMPTOMS</th><th>TROUBLE</th></tr>
<tr><td>Overload Indicator lights Units hum at all TX settings One unit overheats TR follows smoothly but reads wrong</td><td>Rotor circuit open or shorted. See table 1-3.</td></tr>
<tr><td>Overload Indicator lights Units hum at all except two opposite TX settings Both units overheat TR stays on one reading half the time, then swings abruptly to the opposite one. TR may oscillate or spin.</td><td>Stator circuit shorted. See table 1-4.</td></tr>
<tr><td>Overload Indicator lights Units hum on two opposite TX settings Both units get warm TR turns smoothly on one direction, then reverses</td><td>Stator circuit open. See table 1-5.</td></tr>
<tr><td>TR reads wrong or turns backward, follows TX smoothly</td><td>Unit interconnections wrong. Unit not zeroed. See tables 1-6 and 1-7.</td></tr>
</table>

Table 1-3.—Open or Shorted Rotor

<table>
<tr><td colspan="2">Preliminary Action: Set TX to 0° and turn rotor smoothly counterclockwise.</td></tr>
<tr><th>SYMPTOMS</th><th>TROUBLE</th></tr>
<tr><td>TR turns counterclockwise from 0° in a jerky or erratic manner, and gets hot.</td><td>TX rotor open</td></tr>
<tr><td>TR turns counterclockwise from 0° or 180° in a jerky or erratic manner. TX gets hot.</td><td>TR rotor open</td></tr>
<tr><td>TR turns counterclockwise from 90° or 270°, torque is about normal, motor gets hot, and TX fuses blow.</td><td>TX rotor shorted</td></tr>
<tr><td>TR turns counterclockwise from 90° or 270°, torque is about normal, TX gets hot, and TR fuses blow.</td><td>TR rotor shorted</td></tr>
</table>

Table 1-4.—Shorted Stator

SYMPTOMS		TROUBLE
SETTING OR CONDITIONS	**INDICATION**	
When TX is on 120º or 300º but When TX is between 340º and 80º , or between 160º and 260º	Overload Indicator goes out and TR reads correctly Overload Indicator lights, units get hot and hum, and TR stays on 120º or 300º, or may swing suddenly from one point to the other.	Stator circuit shorted from S1 to S2
When TX is on 60 º or 240º but When TX is between 280º and 20º, or between 100º and 200º	Overload Indicator goes out and TR reads correctly Overload Indicator lights, units get hot and hum, and TR stays on 60º or 240º or may swing suddenly from one point to the other	Stator circuit shorted from S2 to S3Stator circuit shorted from S2 to S3
When TX is on 0º or 180º but When TX is between 40º and 140º , or between 220º and 320º	Overload Indicator goes out and TR reads correctly Overload Indicator lights, units get hot and hum, and TR stays on 0º or 180º, or may swing suddenly from one point to the other	Stator circuit shorted from S1 to S3 Stator circuit shorted from S1 to S3

Table 1-5.—Open Stator

SETTINGS OR CONDITIONS	INDICATION	TROUBLE
When TX is on 150º or 330º and When TX is held on 0º	TR reverses or stalls and Overload Indicator lights TR moves between 300º and 0º in a jerky or erratic manner	S1 stator circuit open
When TX is on 90º or 270º and When TX is held on 0º	TR reverses or stalls and Overload Indicator lights TR moves to 0º or 180º , with fairly normal torque	S2 Stator circuit open
When TX is on 30º or 210º and When TX is held on 0º	TR reverses or stalls and Overload Indicator lights TR moves between 0º and 60º in a jerky or erratic manner	S3 stator circuit open
When TX is set at 0˚, and then moved smoothly counterclockwise		

Table 1-6.—Wrong Stator Connections, Rotor Wiring Correct

SETTING OR CONDITIONS	INDICATION	TROUBLE
TX set to 0º and rotor turned smoothly counterclockwise	TR indication is wrong, turns clockwise from 240º	S1 and S2 stator connections are reversed
	TR indication is wrong, turns clockwise from 120º	S2 and S3 stator connections are reversed
	TR indication is wrong, turns clockwise from 0º	S1 and S3 stator connections are reversed
	TR indication is wrong, turns counterclockwise from 120º	S1 is connected to S2, S2 is connected to S3, and S3 is connected to S1
	TR indication is wrong, turns counterclockwise from 240º	S1 is connected to S3, S2 is connected to S1, and S3 is connected to S2

Table 1-7.—Wrong Stator and/or Reversed Rotor Connections

SETTINGS OR CONDITIONS	INDICATION	TROUBLE
TX set to 0º and rotor turned smoothly counterclockwise	TR indication is wrong, turns counterclockwise from 180º	Stator connects are correct, but rotor connections are reversed
	TR indication is wrong, turns clockwise from 60º	Stator connections S1 and S2 are reversed, and rotor connections are reversed
	TR indication is wrong, turns clockwise from 300º	Stator connections S2 and S3 are reversed, and rotor connections are reversed
	TR indication is wrong, turns clockwise from 180º	Stator connections S1 and S3 are reversed, and rotor connections are reversed
	TR indication is wrong, turns counterclockwise from 300º	S1 is connected to S2, S2 is connected to S3, S3 is connected to S1, and rotor connections are reversed
	TR indication is wrong, turns counterclockwise from 60º	S1 is connected to S3, S2 is connected to S1, S3 is connected to S2, and rotor connections are reversed

In a control system, the trouble may be slightly more difficult to isolate. However, the existence of trouble is readily indicated when the system does not properly respond to an input order. For control systems, it is easier to locate the trouble by using a synchro tester or by checking the operating voltages.

VOLTAGE TESTING

Another good way to isolate the trouble in an operating synchro system is to use known operating voltages as references for faulty operation. Since the proper operation of a system is indicated by specific rotor and stator voltages, an ac voltmeter can be used to locate the trouble. When an ac voltmeter is connected between any two stator leads, the voltage should vary from 0 to 90 volts (0 to 11.8 volts for 26-volt systems) as the transmitter rotates. The zero and maximum voltage values should occur at the following headings:

Meter Connected Between	Zero Voltage Headings	Maximum Voltage Headings
S1 and S2	120º, 300º	30º, 210º
S2 and S3	60º, 240º	150º, 330º
S1 and S3	0˚, 180˚	90˚, 270˚

The rotor voltage should remain constant at all times, either 115 volts or 26 volts. In a system where the units are close enough to permit checking, the voltage between the R1 and R2 terminal of any unit energized by the primary ac source and the corresponding R1 or R2 terminal of any other unit energized by the primary ac source should be zero. When the excitation voltage (115 volts or 26 volts) is above or below the nominal value, the maximum stator voltages will also be above or below normal.

SYNCHRO TESTERS

Synchro testers, as stated earlier, are used primarily for quickly locating a defective synchro. These testers are capable of functioning as either transmitter or receiver.

When a transmitter is suspected of being defective, a synchro tester is usually substituted in its place to simulate its actions. When the tester is used in this manner, a braking arrangement on the tester applies the necessary friction to hold its shaft in different positions so you can determine whether the transmitter is good or bad. When using the tester as a transmitter, it is usually a good idea to use only one receiver so as not to overload the tester. If the tester is connected in place of a TR or used to check the output of a transmitter, the brake is released, allowing the rotor to turn and indicate the transmitter's position. By observing the tester's response to the transmitted signal, you can determine if the TR is defective or if the transmitter's output is incorrect.

Q-71. What should you do with a synchro that has a bad set of bearings?

Q-72. Name two types of trouble you would expect to find in a newly installed synchro system.

Q-73. What type of indicator is usually placed in the stator circuit of a torque synchro system?

Q-74. What is the most probable cause of trouble in a synchro system that has all of its receivers reading incorrectly?

Q-75. If an ac voltmeter is connected between the S2 and S3 windings on a TX, at what two rotor positions should the voltmeter read maximum voltage?

Q-76. What precaution should you take when substituting a synchro tester in a circuit for a transmitter?

You should now have a good working knowledge of synchro systems. For further study and assistance in applying this knowledge to synchro troubleshooting and alignment, consult the following references:

Pertinent PMS sources:

1. Applicable maintenance instruction manuals.

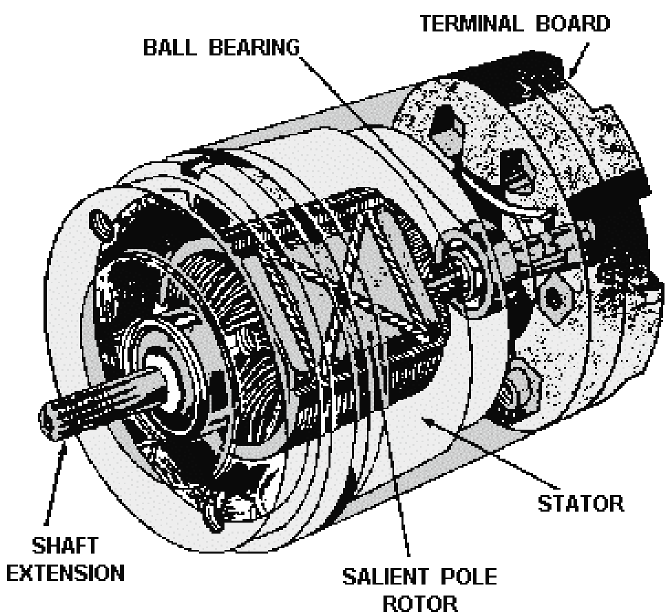
2. Military Handbook, *Synchros, Description and Operation*, MIL-HDBK-225A (March 91).

SUMMARY

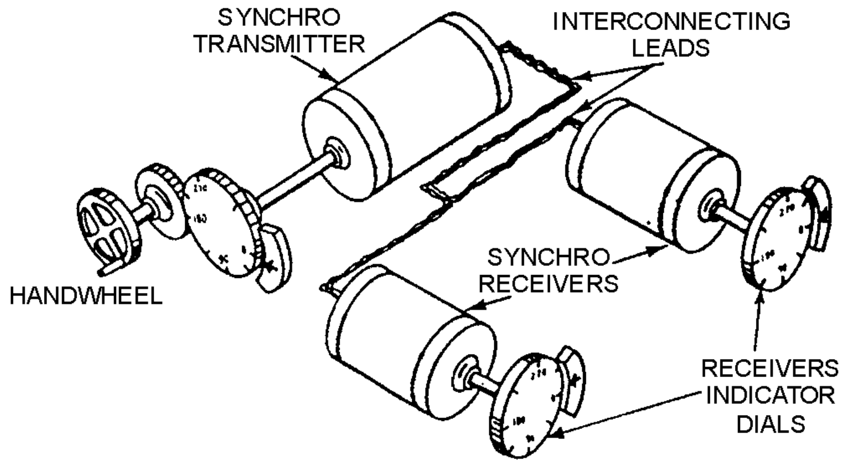
Now that you have completed this chapter, a short review is in order. The following review will refresh your memory of synchros, their principles of operation, and how they are tied together to form synchro systems.

A **SYNCHRO** resembles a small electric motor in size and appearance and operates like a variable transformer.

Synchros are used primarily for the rapid and accurate transmission of data. They are also used as control devices in servo systems.



A **SYNCHRO SYSTEM** consists of two or more synchros interconnected electrically.



TORQUE SYNCHRO SYSTEMS are systems that use torque synchros to move light loads, such as dials and pointers.

CONTROL SYNCHRO SYSTEMS are systems that use control synchros to control servo systems. The servo system, in conjunction with the control synchro system, is used to move heavy loads such as gun directors, radar antennas, and missile launchers.

MILITARY STANDARD SYNCHROS are synchros that conform to specifications that are uniform throughout the Armed Services. They replace the prestandard Navy synchros. A typical example of a military standard synchro designation code is 18TR6A. This code has the following interpretation:

18—Synchro diameter of 1.71 to 1.80 inches

T—Torque

R—Receiver

6—60-Hz frequency

A—Original design

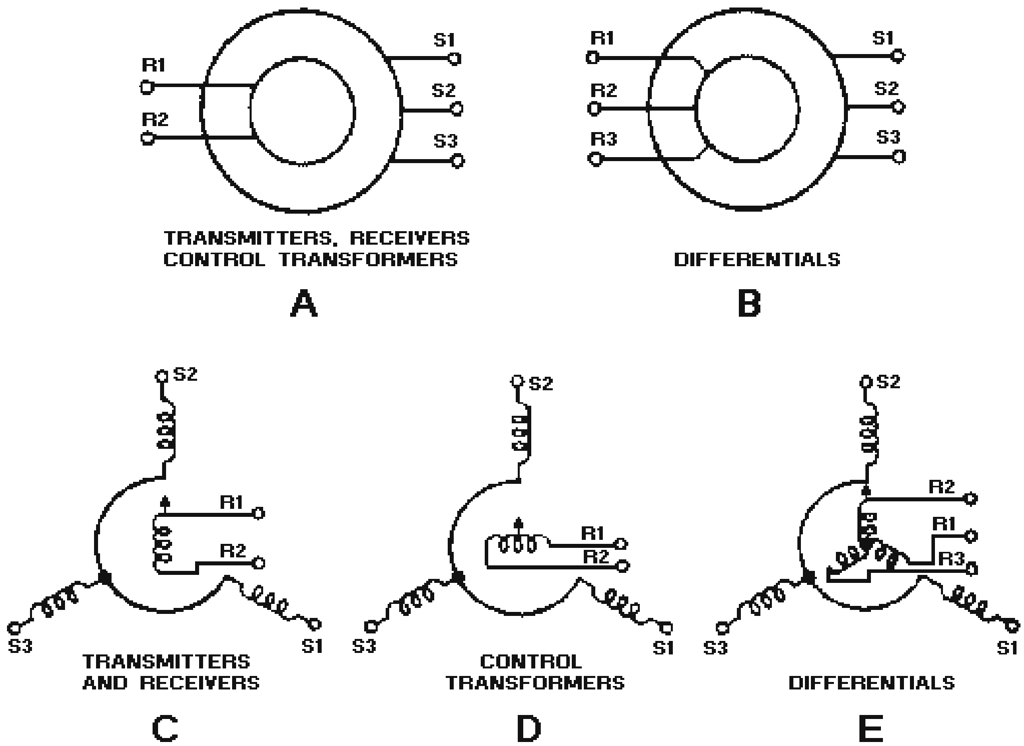
PRESTANDARD NAVY SYNCHROS are synchros designed to meet Navy, rather than service-wide, specifications. A typical example of a prestandard Navy synchro designation code is 5DG. This code has the following interpretation:

5—Synchro diameter of 3 3/8 to 3 5/8 inches, length

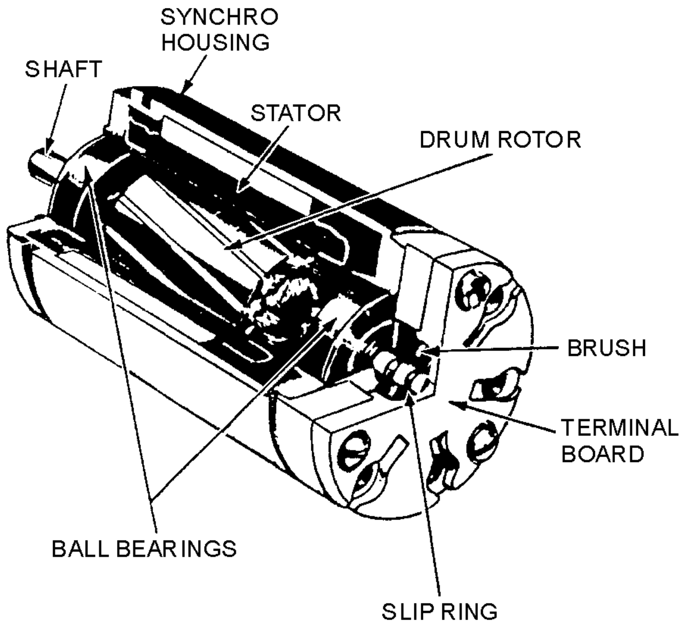
1/2 inches, weight 5 lbs.

DG—Differential transmitter

SCHEMATIC SYMBOLS for synchros are drawn in two different forms. Two of the five standard military symbols are drawn to show only the external connections to the synchro. The other three symbols are drawn to show both the external connections and the internal relationship between the rotor and the stator.

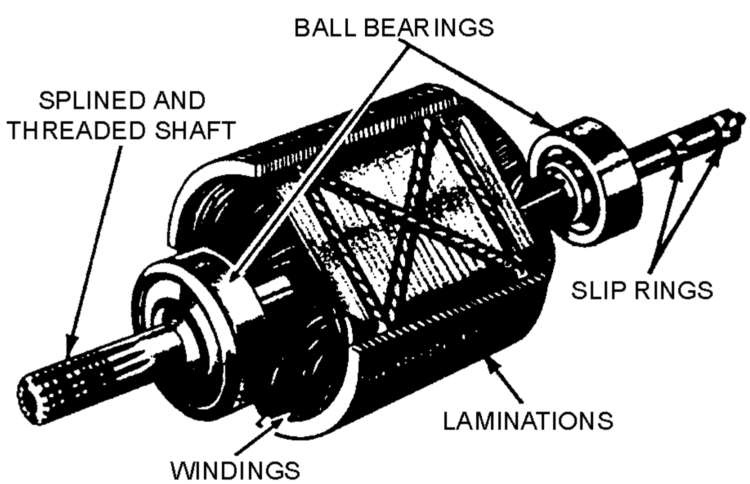


SYNCHROS ARE CONSTRUCTED like motors. Each contains a rotor, similar in appearance to the armature in a motor, and a stator which corresponds to the field in a motor.

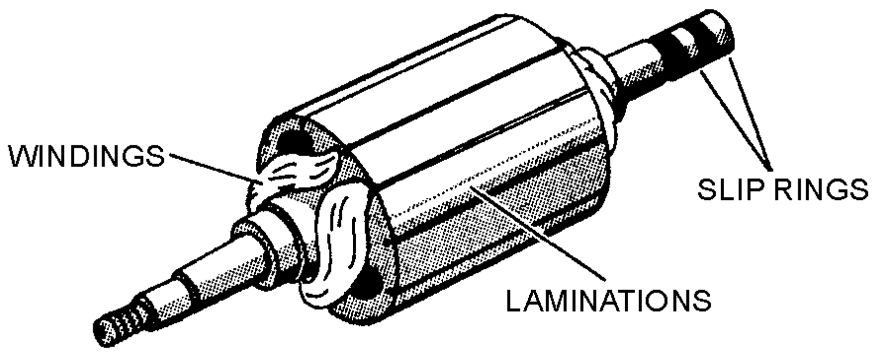


A **SYNCHRO ROTOR** is composed of either a single coil of wire wound on a laminated core or a group of coils wound in slots in a laminated core. The laminated core is rigidly mounted on a shaft that is free to turn inside the stator. Two slip rings are mounted on one end of the shaft to supply excitation voltage to the rotor. There are two common types of synchro rotors - the salient-pole rotor and the drum or wound rotor.

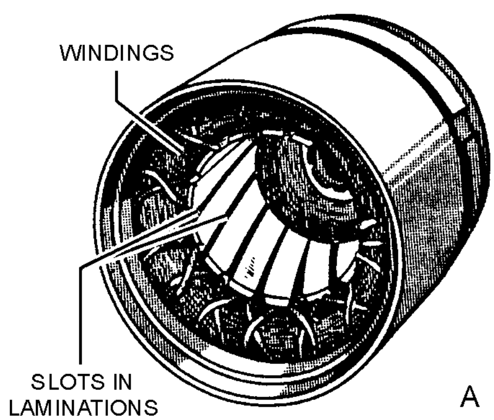
The **SALIENT-POLE ROTOR** has a single coil of wire wound on a laminated core, shaped like a dumbbell or the letter "H." This type of rotor is frequently used in transmitters and receivers.



The **DRUM OR WOUND ROTOR** may be wound continuously with a single length of wire or may have a group of coils connected in series. This type of rotor is used in most synchro control transformers and differential units, and occasionally in torque transmitters. When used in differential units, the rotor is wound with three coils so their magnetic axes are 120 degrees apart.



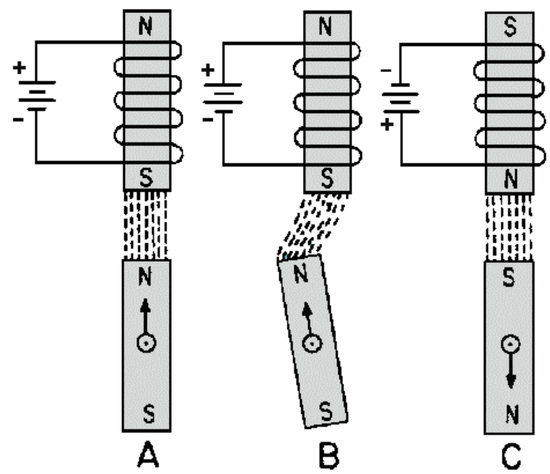
A **SYNCHRO STATOR** is a cylindrical structure of slotted lamination on which three Y-connected coils are wound with their axes 120º apart.



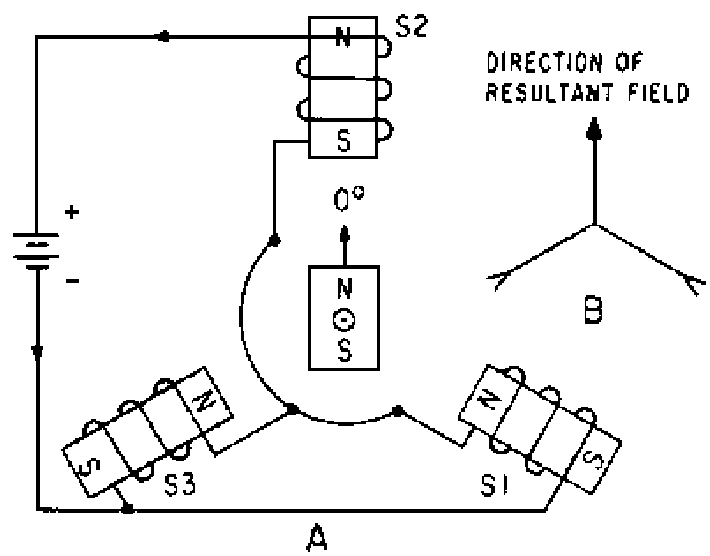
TORQUE is simply a measure of how much load a machine can turn. In heavy machinery, it is expressed in pound-feet and in torque synchro systems, it is expressed in ounce-inches.

SYNCHRO OPERATING VOLTAGES AND FREQUENCIES vary with different equipment. Synchros are designed for use on either a 115-volt or a 26-volt power source. They also operate on either a 60- or 400-Hz frequency.

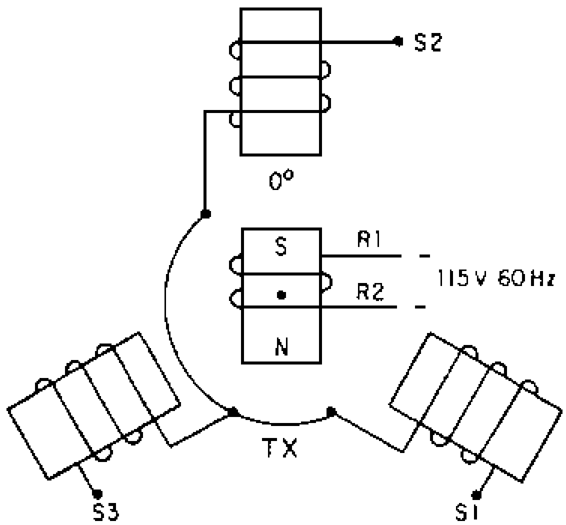
ELECTROMAGNETIC THEORY forms the basis for understanding all synchro operations.



The **RESULTANT MAGNETIC FIELD** in a synchro is the result of the combined effects of three stator fields spaced 120º apart. The stator coil with the strongest field has the greatest effect on the position of the resultant field.

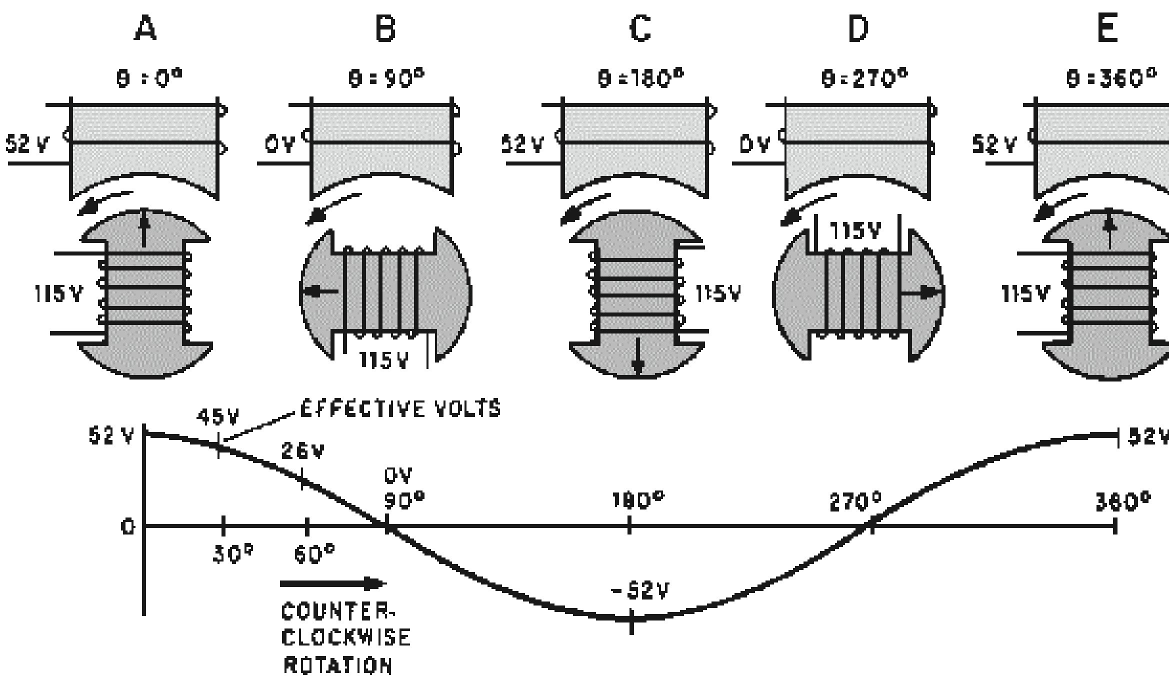


The **ZERO-DEGREE POSITION** of a synchro, transmitter is the position where the rotor and the S2 stator winding are parallel.



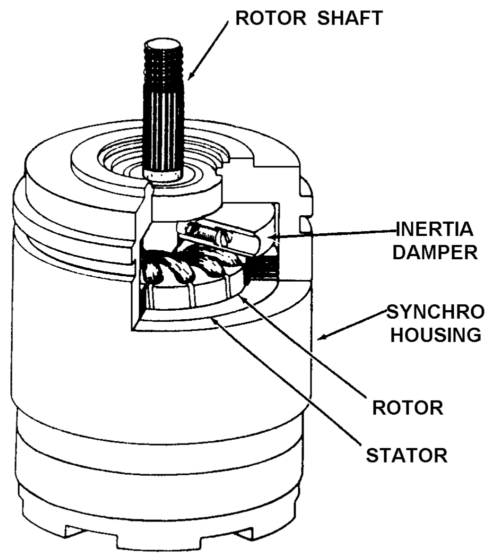
The **SYNCHRO TRANSMITTER (TX)** converts a mechanical input, which is the angular position of the rotor, into an electrical output signal. The output is taken from the stator windings and is used by a TDX, a TDR, or a TR to move light loads, such as dials and pointers.

MAXIMUM INDUCED STATOR VOLTAGE occurs in a single synchro stator coil each time there is maximum magnetic coupling between the rotor and the stator coil. This voltage is approximately equal to the product of the effective voltage on the primary, the secondary-to-primary turns ratio, and the magnetic coupling between the rotor and the stator coil.



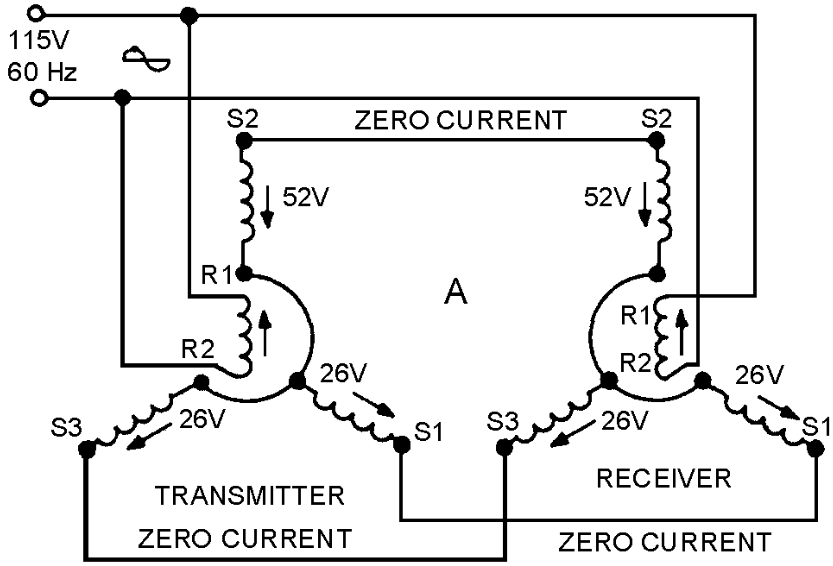
The **SYNCHRO RECEIVER (TR)** is electrically identical to the synchro transmitter. The receiver, however, uses some form of rotor damping that is not present in the transmitter. This real difference between a synchro transmitter and a synchro receiver lies in their applications. The receiver converts the electrical data, supplied to its stator from the transmitter, back to a mechanical angular output through the movement of its rotor.

DAMPING is a method used in synchro receivers to prevent the rotor from oscillating or spinning. There are two types of damping methods - ELECTRICAL and MECHANICAL. The electrical method is commonly used in small synchros, while the mechanical method is more effective in larger synchros.



A **TORQUE SYNCHRO SYSTEM** consists of a torque transmitter (TX) electrically connected to a torque receiver (TR). In this system, the mechanical input to the TX is transmitted electrically to the TR. The TR reproduces the signal from the TX and positions either a dial or a pointer to indicate the transmitted information.

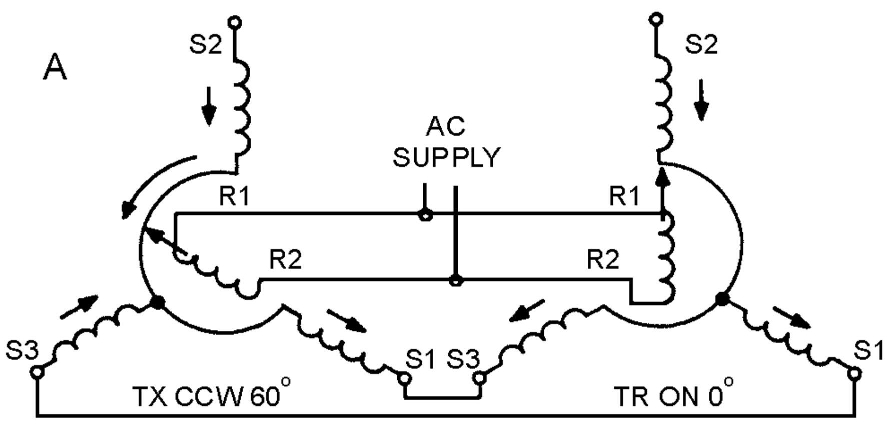
CORRESPONDENCE is the term given to the positions of the rotors of a synchro transmitter and a synchro receiver when both rotors are on 0º or displaced from 0º by the same angle.



SYSTEM IN CORRESPONDENCE

SIGNAL is defined as the angle through which a transmitter rotor is mechanically turned. This term has the same meaning as "transmitter's mechanical input."

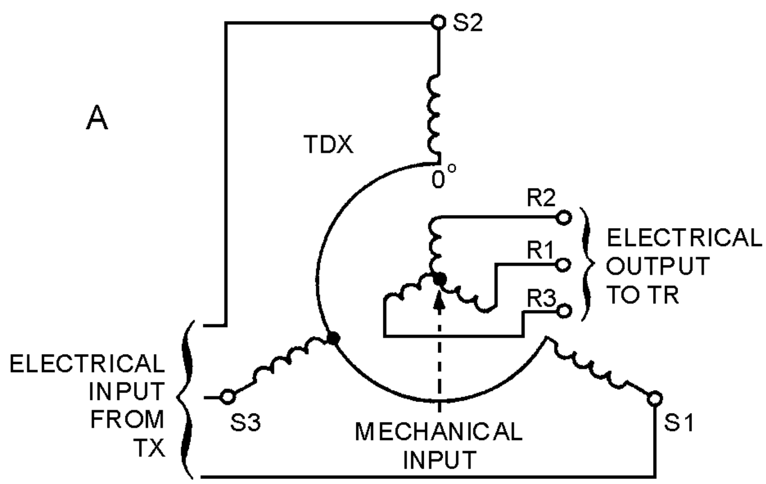
RECEIVER ROTATION may be in a direction opposite to that desired. When it is necessary to reverse receiver rotation, reverse the S1 and S3 connections on either the synchro transmitter or the synchro receiver. The causes both synchro rotors to turn through the same angle but in opposite directions.



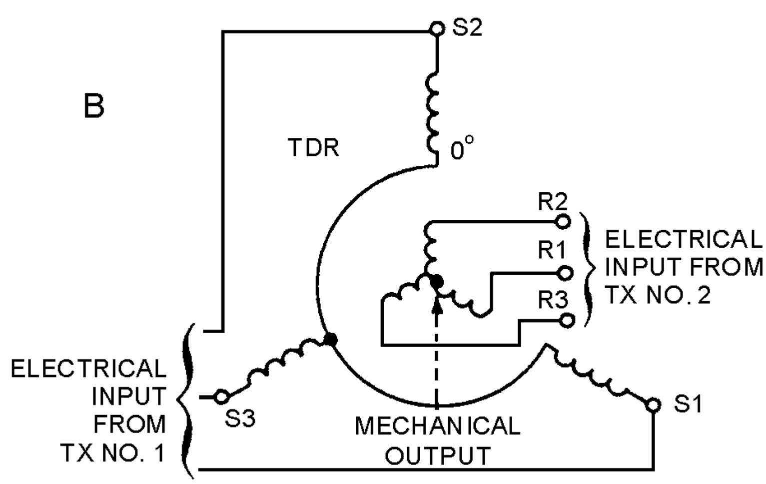
REVERSED ROTOR CONNECTIONS are common problems in new or modified synchro systems and should not be confused with the deliberate reversal of the stator connections. The reversal of the R1 and R2 connections on a synchro rotor causes a 180º error between the synchro transmitter and the synchro receiver, but the direction of rotor rotation still remains the same.

A **TORQUE DIFFERENTIAL SYNCHRO SYSTEM** consists either of a TX, a TDX, (torque differential transmitter), and a TR; or two TXs and one TDR (torque differential receiver). The system is used in applications where it is necessary to compare two signals, add or subtract the signals, and finish an output proportional to the sum of or difference between the two signals.

The **TORQUE DIFFERENTIAL TRANSMITTER (TDX)** has one electrical input to the stator and one mechanical input to the rotor. The TDX either adds or subtracts these inputs, depending upon how it is connected in the system, and provides an electrical output from its rotor proportional to the sum of or difference between the two signals.



The **TORQUE DIFFERENTIAL RECEIVER (TDR)** is electrically identical to the TDX. The only difference in their construction is that the TDR has some form of damping. The real difference between the two differentials lies in their applications. The TDR has two electrical inputs, one to the rotor and the other to the stator. The output is the mechanical position of the rotor. As is the case with the TDX, the addition or subtraction function of the TDR depends upon how it is connected in the system.



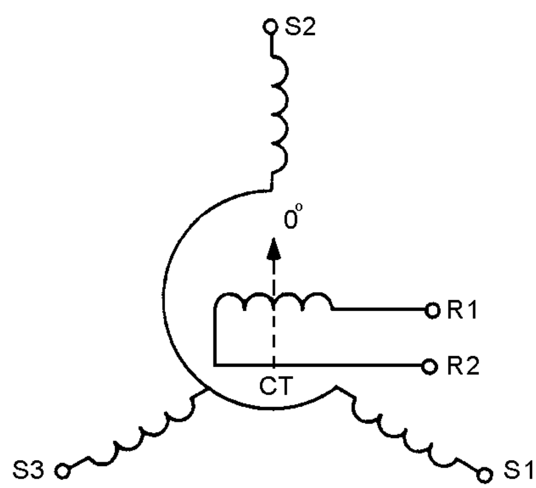
The **TX-TDX-TR SYSTEM** performs subtraction when the system contains standard synchro connections. Addition can also be performed with this system by reversing the S1 and S3 leads between the TX and the TDX, and the R1 and R3 leads between the TDX and the TR. Remember, this system works like a basic synchro system when the rotor of the TDX is on 0º; in this condition the TR rotor follows the TX rotor exactly.

The **TX-TDR-TX SYSTEM** performs subtraction when the system contains standard synchro connections. Addition can also be performed with this system when the R1 and R3 leads between the TDR rotor and TX No. 2 are reversed.

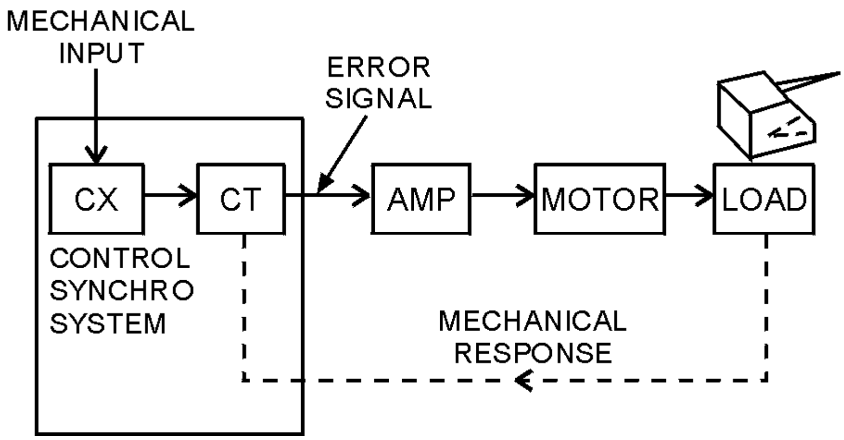
CONTROL SYNCHRO SYSTEMS contain control synchros and are used to control large amounts of power with a high degree of accuracy. These synchro systems control servos that generate the required power to move heavy loads.

CONTROL SYNCHROS are of three different types: the control transmitter (CX), the control transformer (CT), and the control differential transmitter (CDX). The CX and the CDX are identical to the TX and the TDX except for higher impedance windings. In theory, the CX and CDX are the same as the TX and TDX, respectively.

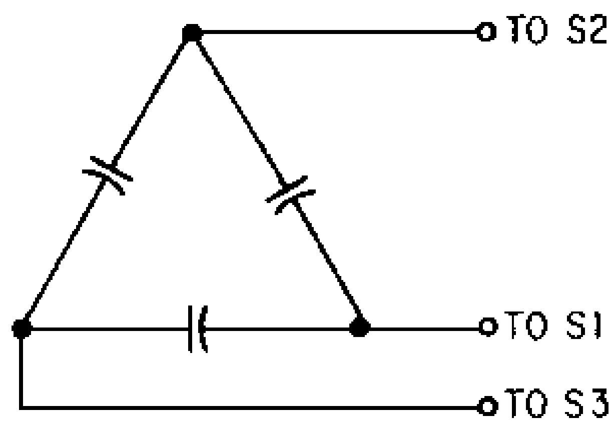
The **CONTROL TRANSFORMER (CT)** is a synchro device that compares two signals, the electrical signal applied to its stator, and the mechanical signal applied to its rotor. The output is an electrical voltage, which is taken from the rotor winding and used to control some form of power amplifying device. The phase and amplitude of the output voltage depend on the angular position of the rotor with respect to the magnetic field of the stator.



ERROR SIGNAL is the name given to the electrical output of a CT. The reason is that the electrical output voltage represents the amount and direction that the CX and CT rotors are out of correspondence. It is this error signal that eventually is used in moving the load in a typical servo system.



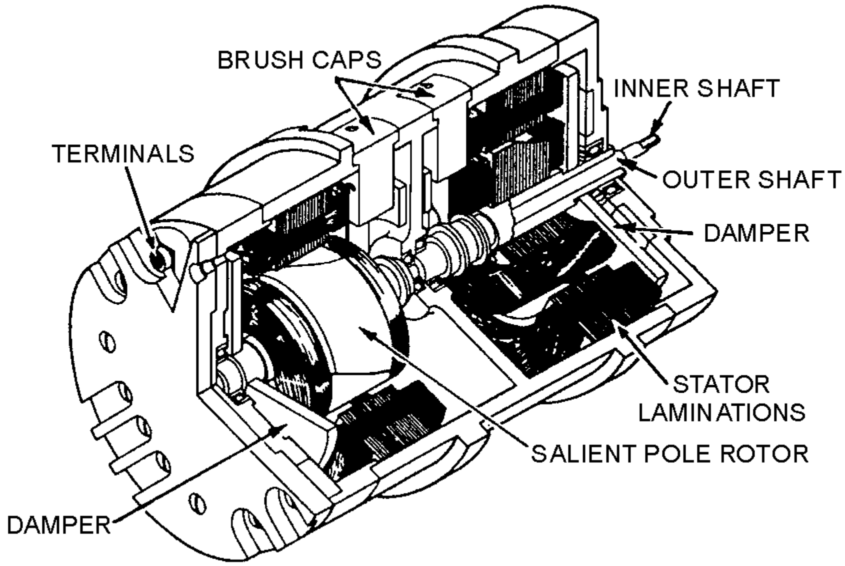
The **SYNCHRO CAPACITOR** is a unit containing three delta-connected capacitors. It is used in synchro systems containing either differential transmitters or CTs. The addition of the synchro capacitor to these systems greatly reduces the stator current and therefore increases the accuracy of the systems.



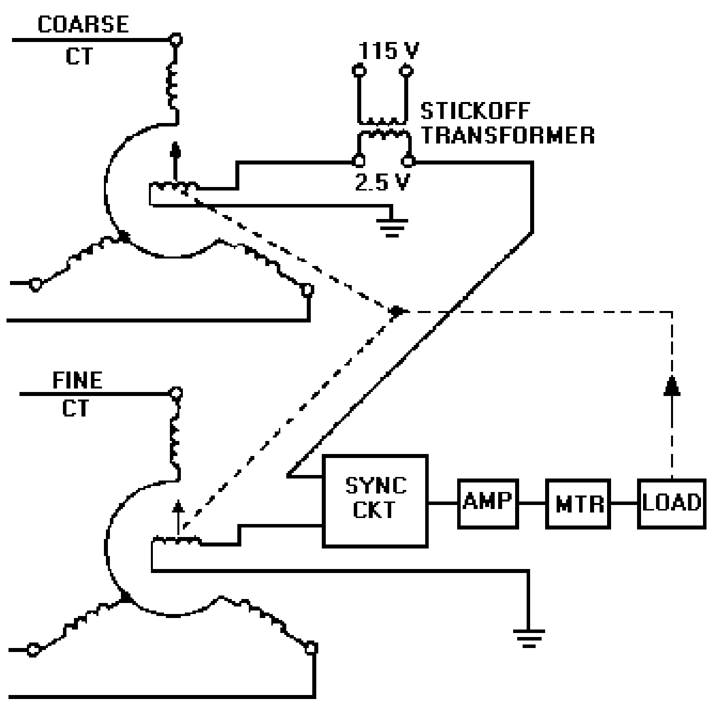
The **SPEED OF DATA TRANSMISSION** is simply the number of times a synchro transmitter rotor must turn to transmit a full range of values. You refer to the speed of data transmission as being 1-speed, 2-speed, 36-speed, or some other definite numerical ratio.

MULTISPEED SYNCHRO SYSTEMS transmit a wide range of data at different speeds and still maintain a high degree of accuracy. To indicate the number of different speeds at which data is transmitted, refer to the system as being a single-speed, dual-speed, or tri-speed synchro system.

A **DOUBLE RECEIVER** consists of a fine and a coarse receiver enclosed in a common housing. It has a two-shaft output (one inside the other), and its operation may be likened to the hour and minute hands of a clock.



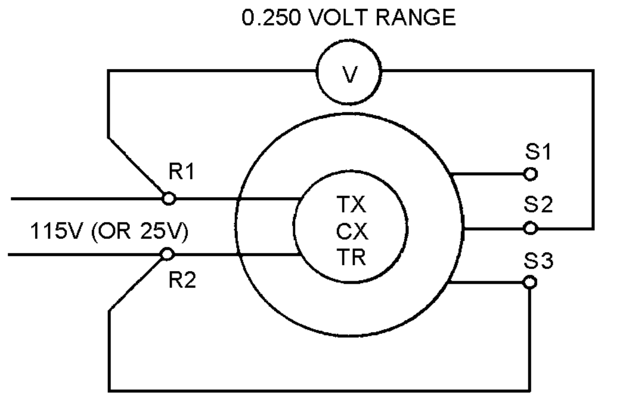
STICKOFF VOLTAGE is a low voltage used in multispeed synchro systems that contain CTs to prevent false synchronizations. The voltage is usually obtained from a small transformer and applied across the rotor terminals of the coarse CT.



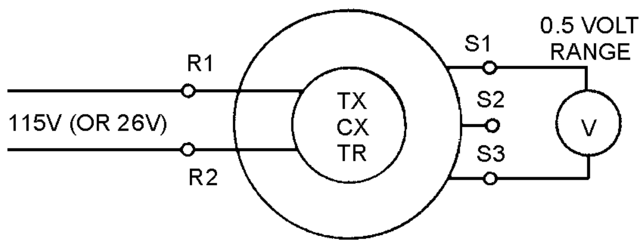
ELECTRICAL ZERO is the reference point for alignment of all synchro units.

SYNCHRO ZEROING METHODS are various and depend upon the facilities and tools available, and how the synchros are connected in the system. Some of the more common zeroing methods are the voltmeter, the electric-lock, and the synchro-tester methods.

The **VOLTMETER ZEROING METHOD** is the most accurate and requires a precision voltmeter. This method has two major steps-the coarse or approximate setting and the fine setting. The coarse setting ensures the synchro is not zeroed 180º away from its reference. This setting may be approximated physically by aligning two marks on the synchro. The fine setting is where the synchro is precisely set on 0º .

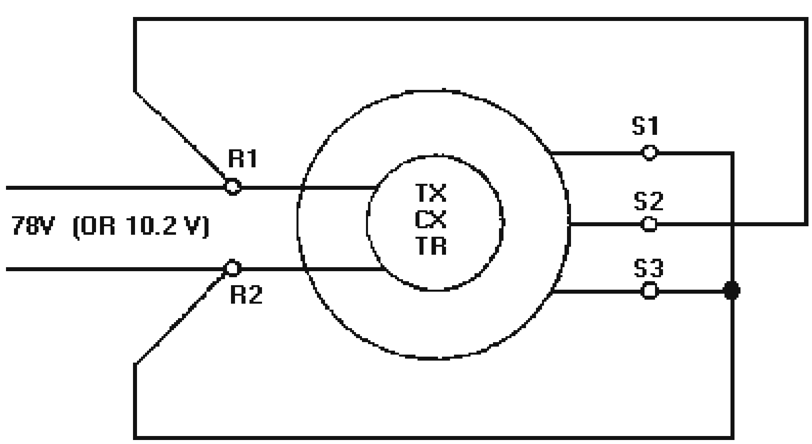


COARSE SETTING



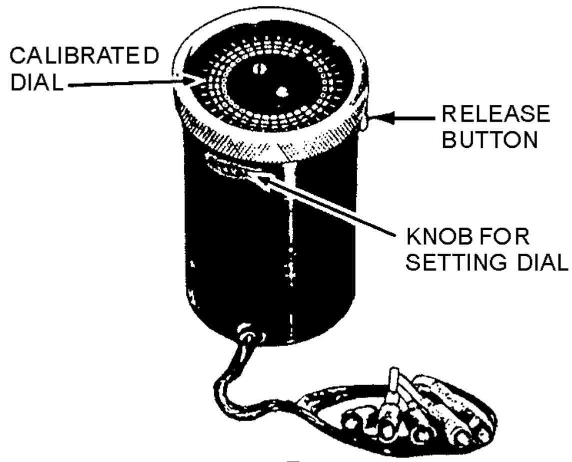
FINE SETTING

ELECTRICAL-LOCK ZEROING METHOD is perhaps the fastest. However, this method can be used only if the rotors of the synchros are free to turn and the leads are accessible. For this reason, this method is usually used on TRs.



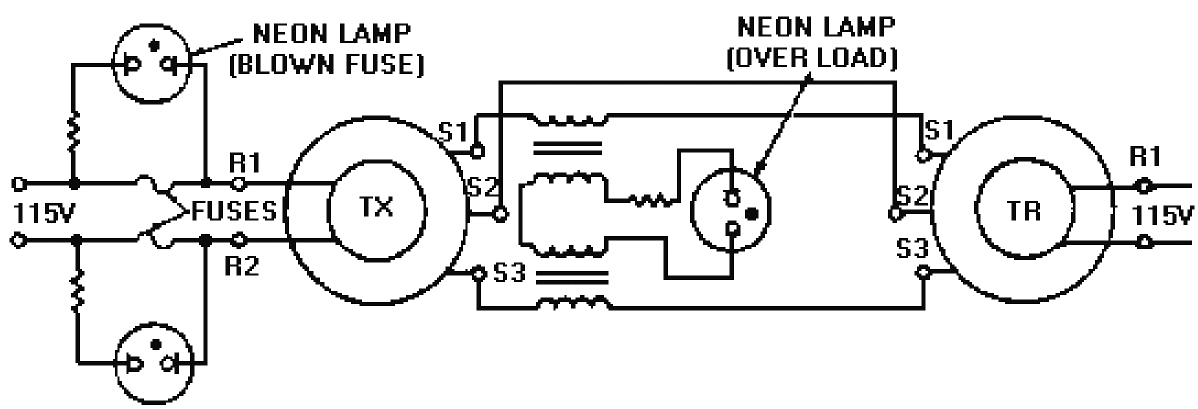
The **SYNCHRO-TESTER ZEROING METHOD** is potentially less accurate than the voltmeter or electric lock methods. This is because the dial on the tester is difficult to read and may slip from its locked position.

The synchro tester is nothing more than a synchro receiver on which a calibrated dial is mounted. The tester is used primarily for locating defective synchros but does provide a method for zeroing synchros.



B

TROUBLE INDICATORS are signal lights used to aid maintenance personnel in locating synchro trouble quickly. These lights are usually mounted on a central control board and connected to different synchro systems. The lights indicate either overload conditions or blown fuses.



SYNCHRO TROUBLESHOOTING is the locating or diagnosing of synchro malfunctions or breakdown by means of systematic checking or analysis. This is done by using trouble indicators, trouble tables furnished by manufacturers, known operating voltages as references, and synchro testers.

ANSWERS TO QUESTIONS Q1. THROUGH Q76.

A-1. The synchro.

A-2. Precise and rapid transmission of data between equipment and stations.

A-3. Torque and control.

A-4. A torque synchro is used for light loads and a control synchro is used in systems desired to move heavy loads.

A-5. The torque receiver (TR) and the torque differential receiver (TDR).

A-6. It is the third modification of a 26-volt 400-hertz (torque) synchro transmitter whose body diameter is between 1.01 and 1.10 inches.

A-7. The Navy prestandard designation code.

A-8. The position of the arrow.

A-9. The rotor and the stator.

A-10. The drum or wound rotor.

A-11. By the magnetic coupling from the rotor.

A-12. At the terminal board.

A-13. The number and type of synchro receivers, the mechanical loads on these receivers and the operating temperatures of both the transmitter and receivers.

A-14. A measure of how much load a machine can turn.

A-15. Ounce-inches.

A-16. Aircraft.

A-17. When it is overloaded.

A-18. Synchros have one primary winding that can be turned through 360º and three secondary windings spaced 120º apart.

A-19. The transmitter is in its zero-position when the rotor is aligned with the S2 stator winding.

A-20. When the rotor coil is aligned with the stator coil.

A-21. The amplitude of the primary voltage, the turns ratio, and the angular displacement between the rotor and the stator winding.

A-22. A synchro receiver uses some form of damping to retard excessive oscillations or spinning.

A-23. Mechanical damping.

A-24. A synchro transmitter and a synchro receiver.

A-25. The rotor leads.

A-26. The voltages are equal and oppose each other.

A-27. Signal.

A-28. 1 and S3.

A-29. The rotor leads on either the transmitter or the receiver are reversed.

A-30. Differential synchros can handle more signals than regular synchros and also perform addition and subtraction functions.

A-31. The TDX and the TDR.

A-32. Their application: a TDX has one electrical and one mechanical input with an electrical output.

A-33. The way the differential synchro is connected in a system is the deciding factor on whether the unit adds or subtracts its inputs.

A-34. When the TDX rotor is on 0º.

A-35. 240º.

A-36. 80º.

A-37. The S1 and S3 leads are reversed between the TX and the TDX, and the R1 and R3 leads are reversed between the TDX rotor and the TR.

A-38. The R1 and R3 leads between the TDR rotor and the TX to which it is connected.

A-39. Clockwise.

A-40. A control synchro.

A-41. CX, CT, and CDX.

A-42. The CX and CDX have higher impedance windings.

A-43. The rotor is specially wound, it is never connected to an ac supply, and its output is always applied to a high-impedance load.

A-44. They are perpendicular to each other.

A-45. The voltage is maximum and in phase with the ac excitation voltage to the CX.

A-46. Error signal.

A-47. When the CX and CT rotors are in correspondence.

A-48. To improve overall synchro system accuracy by reducing stator currents.

A-49. TDXs, CDXs, and Cts.

A-50. Magnetizing current.

A-51. They are delta-connected across the stator windings.

A-52. To keep the connections as short as possible in order to maintain system.

A-53. A dual or double-speed synchro system.

A-54. Greater accuracy without the loss of self-synchronous operation.

A-55. The gear ratio between the two transmitters.

A-56. A tri-speed synchro system.

A-57. If one of the receivers goes bad the entire unit must be replaced.

A-58. It is used in synchro systems to prevent false synchronizations.

A-59. Electrical zero.

A-60. The voltmeter method.

A-61. A61.It ensures the synchro is on 0º, not 180º.

A-62. A TR is zeroed when electrical zero voltages exist across its stator windings at the same time its rotor is on zero or on its mechanical reference position.

A-63. Approximately 37 volts.

A-64. Never leave the circuit energized for more than 2 minutes.

A-65. To ensure that it did not move off zero while it was being clamped.

A-66. Zero or minimum voltage.

A-67. The coarse synchro.

A-68. The electrical lock method.

A-69. It can be used only if the leads of the synchro are accessible and the rotor is free to turn.

A-70. The synchro under test is not on electrical zero.

A-71. Replace it.

A-72. Improper wiring and misalignment.

A-73. An overload indicator.

A-74. The transmitter or main bus.

A-75. 150º and 330º

A-76. Use only one receiver so as not to overload the tester.

CHAPTER 2

SERVOS

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Define the term "servo system" and the terms associated with servo systems, including open-loop and closed-loop control systems.
2. Identify from schematics and block diagrams the various servo circuits; give short explanations of the components and their characteristics; and of each circuit and its characteristics.
3. Trace the flow of data through the components of typical servo systems from input(s) to outputs(s) (cause to effect).

SERVOS

Servo mechanisms, also called SERVO SYSTEMS or SERVOS for short, have countless applications in the operation of electrical and electronic equipment. In working with radar and antennas, directors, computing devices, ship's communications, aircraft control, and many other equipments, it is often necessary to operate a mechanical load that is remote from its source of control. To obtain smooth, continuous, and accurate operation, these loads are normally best controlled by synchros. As you already know, the big problem here is that synchros are not powerful enough to do any great amount of work. This is where servos come into use.

A servo system uses a weak control signal to move large loads to a desired position with great accuracy. The key words in this definition are <u>move</u> and <u>great</u> <u>accuracy</u>. Servos may be found in such varied applications as moving the rudder and elevators of a model airplane in radio-controlled flight, to controlling the diving planes and rudders of nuclear submarines.

Servos are powerful. They can move heavy loads and be remotely controlled with great precision by synchro devices.

They take many forms. Servo systems are either electromechanical, electrohydraulic, hydraulic, or pneumatic.

Whatever the form, a relatively weak signal that represents a desired movement of the load is generated, controlled, amplified, and fed to a servo motor that does the work of moving the heavy load.

Q-1. What is a servo?

CATEGORIES OF CONTROL SYSTEMS

A control system is a group of components that are linked together to perform a specific purpose. Generally, a control system has a large power gain between input and output. The components used in the system and the complexity of the system are directly related to the requirements of the system's application.

Control systems are broadly classified as either CLOSED-LOOP or OPEN-LOOP.

Closed-loop control systems are the type most commonly used in the Navy because they respond and move the loads they are controlling quicker and with greater accuracy than open-loop systems.

The reason for quicker response and greater accuracy is that an automatic feedback system informs the input that the desired movement has taken place. Upon receipt of this feedback information, the system stops the motor, and motion of the load ceases until another movement is ordered by the input. This is similar to the system that controls heat in many homes. The thermostat (input) calls for heat. The furnace (output) produces heat and distributes it. Some of the heat is "fed back" to the thermostat. When this "feedback" raises the temperature of the room to that of the thermostat setting, the thermostat responds by shutting the system down until heat is again required. In such a system, the feedback path, input to output and back to input, forms what is called a "closed loop." This is a term you will hear and use often in discussions of control systems. Because closed-loop control systems are automatic in nature, they are further classified by the function they serve (e.g., controlling the position, the velocity, or the acceleration of the load being driven).

An open-loop control system is controlled directly, and only, by an input signal, without the benefit of feedback. The basic units of this system are an amplifier and a motor. The amplifier receives a low-level input signal and amplifies it enough to drive the motor to perform the desired job. Open-loop control systems are not as commonly used as closed-loop control systems because they are less accurate.

OPEN-LOOP CONTROL SYSTEM

As we stated previously, an open-loop control system is controlled directly, and only, by an input signal. The basic units of this type consist only of an amplifier and a motor. The amplifier receives a low-level input signal and amplifies it enough to drive the motor to perform the desired job.

The open-loop control system is shown in basic block diagram form in figure 2-1. With this system, the input is a signal that is fed to the amplifier. The output of the amplifier is proportional to the amplitude of the input signal. The phase (ac system) and polarity (dc system) of the input signal determines the direction that the motor shaft will turn. After amplification, the input signal is fed to the motor, which moves the output shaft (load) in the direction that corresponds with the input signal. The motor will not stop driving the output shaft until the input signal is reduced to zero or removed. This system usually requires an operator who controls speed and direction of movement of the output by varying the input. The operator could be controlling the input by either a mechanical or an electrical linkage.

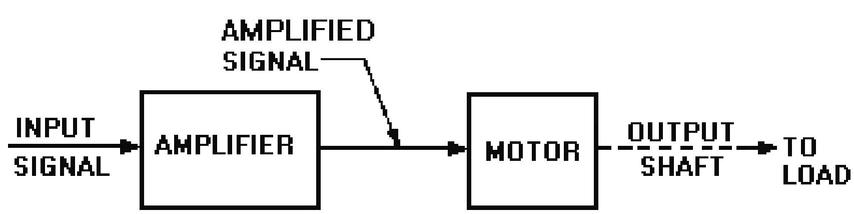


Figure 2-1.—Open-loop control system basic block diagram.

THE CLOSED-LOOP CONTROL SYSTEM

A closed-loop control system is another name for a servo system. To be classified as a servo, a control system must be capable of the following:

1. Accepting an order that defines the desired result
2. Determining the present conditions by some method of feedback
3. Comparing the desired result with the present conditions and obtaining a difference or an error signal
4. Issuing a correcting order (the error signal) that will properly change the existing conditions to the desired result
5. Obeying the correcting order

We have discussed the open- and closed-loop control systems and defined a servo system as a closed-loop control system. Although not technically accurate by definition, open-loop control systems are also often referred to in the Navy and many publications as servo systems even though they lack one of the five basic requirements, that of feedback.

Q-2. In an open-loop control system, what action reduces the input to zero so the load is stopped at the desired position?

Q-3. What basic requirement of a closed-loop system (not present in open-loops) enables present load position to be sensed?

OPERATION OF A BASIC SERVO SYSTEM

For the following discussion of a servo system, refer to figure 2-2, view (A), view (B), view (C) and view (D).This closed-loop servo system is the most common type in the Navy today. It is normally made up of electromechanical parts and consists basically of a synchro-control system, servo amplifier, servo motor, and some form of feedback (response).

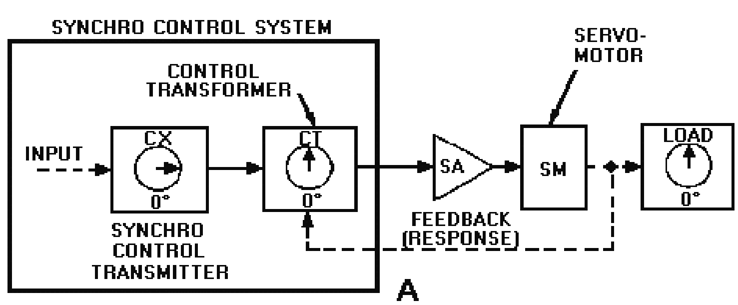


Figure 2-2A.—A basic servo system (closed-loop).

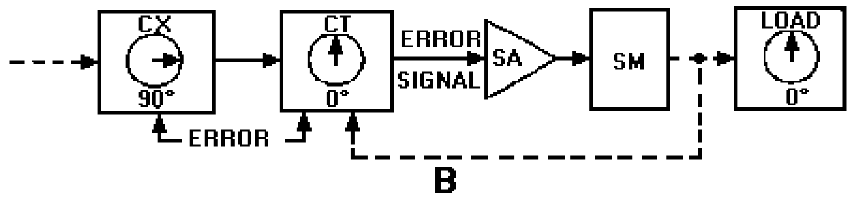


Figure 2-2B.—A basic servo system (closed-loop).

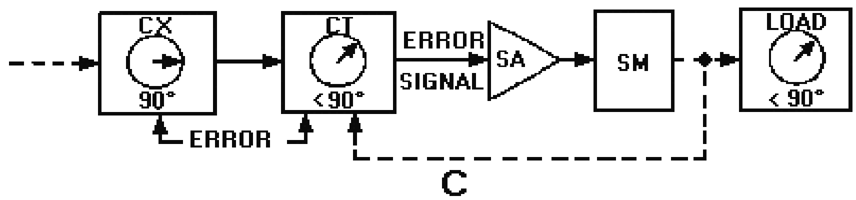


Figure 2-2C.—A basic servo system (closed-loop).

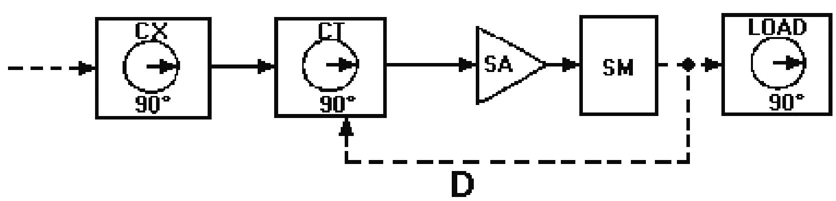


Figure 2-2D.—A basic servo system (closed-loop).

The synchro-control system provides a means of controlling the movement of the load, which may be located in a remote space. The servo amplifier and servo motor are the parts of the system in which power is actually developed (to move the load).

As you remember, the controlling signal from a CT is relatively weak, too weak to drive an electric motor directly. In views A through D of figure 2-2, assume that the control signal will be initiated by a handcrank input connected to the synchro transmitter (CX). The dials located on the CX and the CT indicate the positions of the synchro's rotors, while the dial on the load indicates the position of the load.

In view A, the dials of both the CX and the load indicate that the load is in the desired position. Because the load is where it should be, there will be no error signal present at the servo amplifier and no power to the servo motor.

In view B, the rotor of the CX has been moved by the handcrank to 90º. (This indicates that it is ordered to move the load by 90º.) Notice that the rotor of the CT is still at 0º. The CT now develops a signal, called the ERROR SIGNAL, which is proportional in amplitude to the amount the CT rotor is out of correspondence with the CX rotor. The phase of the error signal indicates the direction the CT rotor must move to reduce the error signal to zero or to "null out." The error signal is sent to the servo amplifier. In view C, the error signal has been amplified by the servo amplifier and sent on to the servo motor. The motor starts to drive in the direction that will reduce the error signal and bring the CT rotor back to the point of correspondence. In this case the motor is turning clockwise.

The mechanical linkage attached to the servo motor also moves the rotor of the CT. This feedback causes the amplitude of the error signal to decrease, slowing the speed at which the load is moving.

In view D, the servo motor has driven both the load and the rotor of the CT, so that the CT is now in correspondence with the CX rotor. As a result, the error signal is reduced to zero (nulled). The load stops at its new position. Note that in this servo system, we moved a heavy load to a predetermined position through the simple turning of a handcrank. In responding to the handcrank, the servo system performed a basic positioning function.

Two key points for you to remember, thus far, about the operation of the closed-loop servo system are:

1. The original error (movement of the CX rotor) was "detected" by the CT. For this reason the CT is called an ERROR DETECTOR.

2. The servo motor, in addition to moving the load, also provides mechanical feedback to the CT to reduce the error signal. For this reason the servo motor is also called an ERROR REDUCER.

Q-4. An error signal is the difference between what two quantities fed to the CT (error detector)?

Q-5. What are the two functions of the servo motor in the system shown in figure 2-2?

FUNCTIONAL SERVO LOOPS

Servo systems are also classified according to their functions: POSITION, VELOCITY, and ACCELERATION. We will cover the most common, POSITION and VELOCITY, in detail.

The Position Servo Loop

The primary purpose of the POSITION SERVO is to control the position of the load it is driving. It can be used to position a great number of devices (for example, valves, control surfaces, weapons, etc.). The basic servo loop we just explained using the block diagram in figure 2-2 is that of an ac position servo system. In the ac position servo system, the amplitude and phase of the ac error signal determine the amount and direction the load will be driven.

In a dc position servo system, the amplitude and polarity of a dc error signal respectively are used to determine the amount and direction the load will be driven.

Figure 2-3, view A, is a block diagram of a closed-loop de position servo. Note the Greek letter Sigma (Σ), meaning summation, surrounded by a circle.

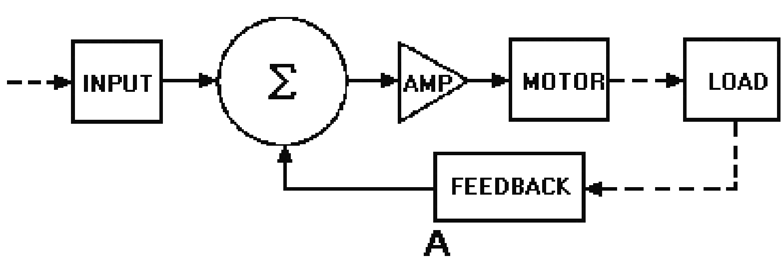


Figure 2-3A.—Block diagram of a position servo.

This is the summation, or "sum point" where the input signal, and the response signal (feedback) are summed to produce the error signal.

View B shows a more in-depth illustration of view A. With the wiper arms of R_1 and R_4 at the midpoint of travel, the voltage from the wiper arm to ground is zero volts. Therefore, zero volts would also be measured at the connection point between R_2 and R_3 (the summation point). This means that the error signal is zero. With no input signal, the amplifier output is zero; therefore, the motor shaft remains stationary.

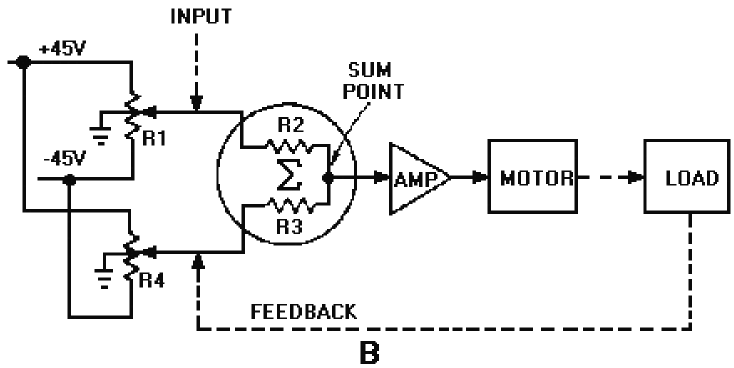


Figure 2-3B.—Block diagram of a position servo.

For the purpose of explanation, imagine that the R_1 wiper arm is mechanically moved upward to a new position where a voltage between the wiper arm and ground measures +10 volts. Further measurement shows zero volts between the wiper arm of $R_4$and ground. Since R_2 and R_3 are of equal values, +5 volts is measured between the sum point and ground because 5 volts is dropped across each resistor. The +5 volts at the sum point is the "error" signal.

As shown in figure 2-4, (view A, view B, and view C),when no error is present, the voltage at the sum point is zero. This is because the network composed of R_1, R_2, R_3, and R_4 is balanced. When the wiper of R_1 is moved toward +45 volts, the network becomes unbalanced as shown in view B. The left-hand side of R_2 becomes positive. This causes current to flow from +45 volts through R_3 and R_2 to the +10 volts at the left side of R_2. Because R_2 and R_3 are of equal value, the voltage drops then will be equal; therefore, the voltage at the sum point will equal +5 volts.

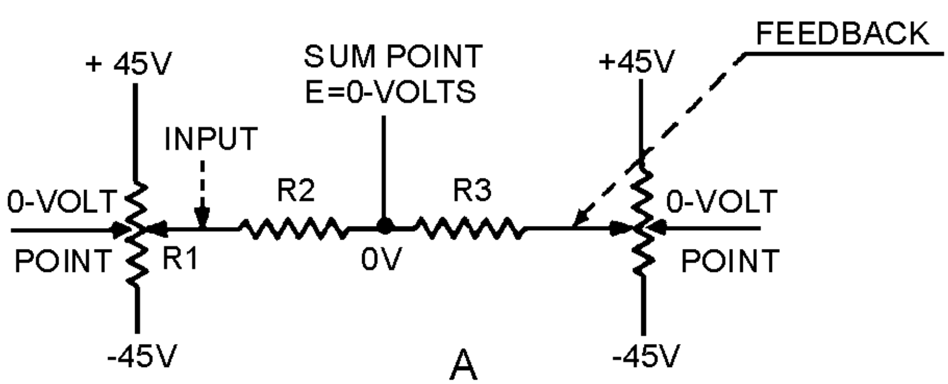


Figure 2-4A.—Development of the error signal.

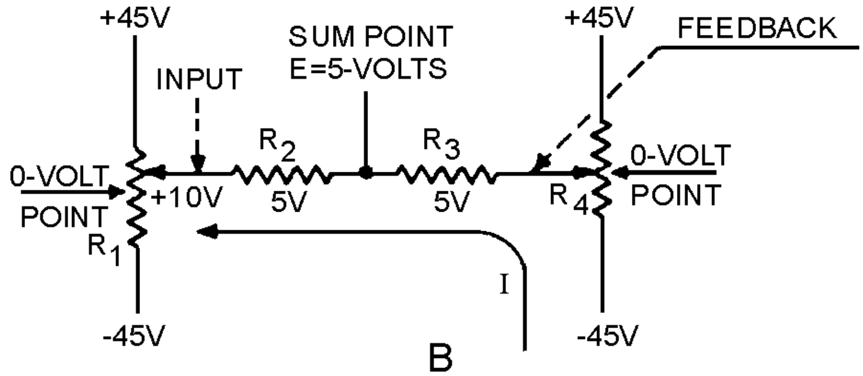


Figure 2-4B.—Development of the error signal.

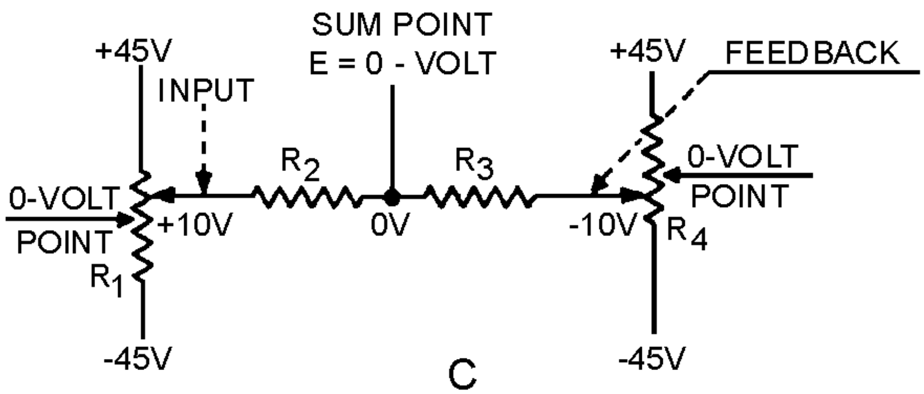


Figure 2-4C.—Development of the error signal.

The +5 volt error signal is fed into the amplifier. The amplified output starts driving the motor. The mechanical feedback from the motor drives the R_4 wiper arm down when the R_1 wiper is moved up, as shown in view C. This causes the right-hand side of R_3 to go negative. When the R_4 wiper travels far enough toward negative, it causes the right-hand side of R_3 to equal the voltage, but of opposite polarity, of the left-hand side of R_2. Simply stated, the voltage at the sum point will be zero again, and the motor will stop. This is true because R_2 and R_3 are of equal ohmic value, and when the left-hand side equals +10 volts, the right-hand side equals -10 volts. The point between the two resistors becomes zero volts at this time. At the instant that this occurs, the output shaft will have positioned the load to the new position.

Figure 2-5 shows the basic operation of a typical position servo having wide application in Navy equipment. Remember that in a position servo, an input order indicates a position in which a load is to be placed. The load in figure 2-5 is a gun turret. The purpose of the system is to position the gun by means of an order from a remote handcrank. The load is mechanically coupled through a gear train to the rotor of a CT so that the turret's position is always accurately represented by the position of the CT's rotor. An order signaling the desired position of the gun turret is fed into the servo by positioning the rotor of the CX with the handcrank. A corresponding signal immediately appears across the CT stator. This signal differs from the actual position of the gun turret, causing an error voltage to be developed across the CT rotor. The error voltage is fed from the CT rotor to the servo amplifier. At this point it is converted into power with a polarity or phase relationship that drives the motor in the direction necessary to bring the load into the desired position. As the turret moves, mechanical feedback turns the CT rotor toward agreement with the CX rotor. As the load approaches the proper position, less and less power is supplied to the motor because of the decreasing error voltage developed in the CT. When the electrical position of the CT rotor agrees with the position of the CX rotor, the error voltage reaches zero and power is removed from the motor. The turret is now in the desired position.

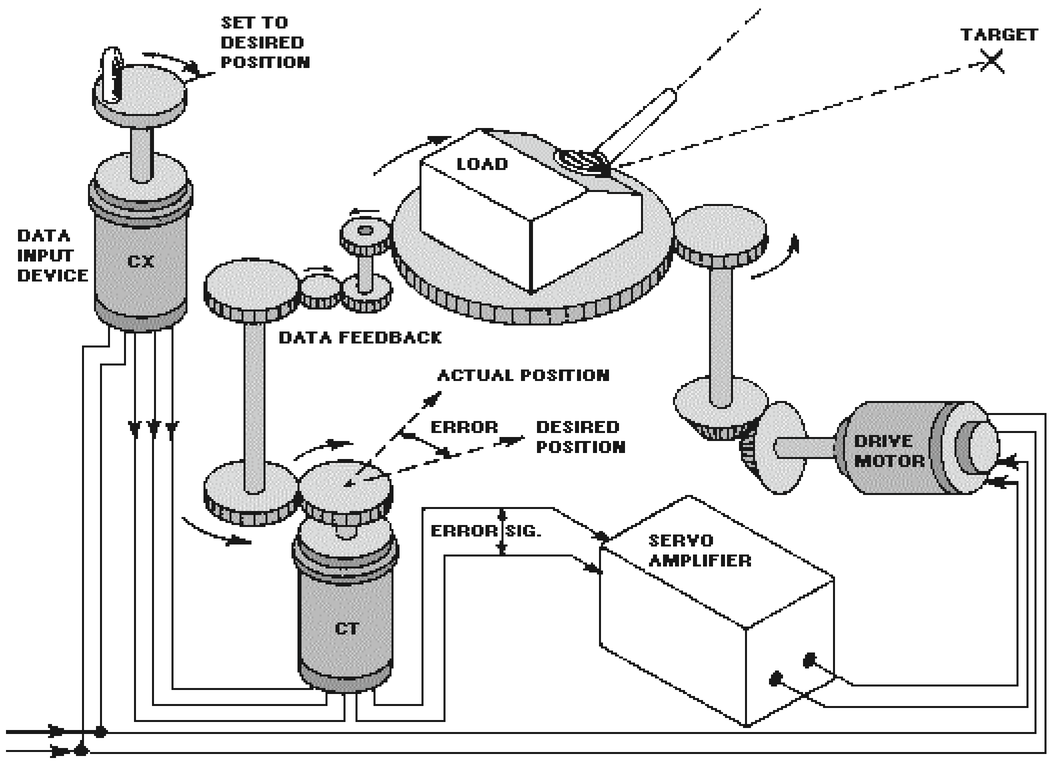


Figure 2-5.—Typical position servo.

In the actual system, the heavy gun turret's momentum tends to carry it past the desired position. This overshoot causes the rotor of the CT to move out of correspondence with the CX rotor. This, in turn, develops a new error signal that is opposite in polarity to the original input signal. The new error signal causes the turret to drive back toward the desired position — but the turret's momentum once again causes an overshoot, making the system drive in the opposite direction again. If this oscillation of the load around the desired position is allowed to go unchecked, a condition known as HUNTING results. Figure 2-6 shows graphically the result of a series of overtravels of the correspondence point (hunting). In most servos an electronic network known as an ANTIHUNT or DAMPING system is used to minimize this undesirable effect. We will cover antihunt and damping systems in depth later in this chapter.

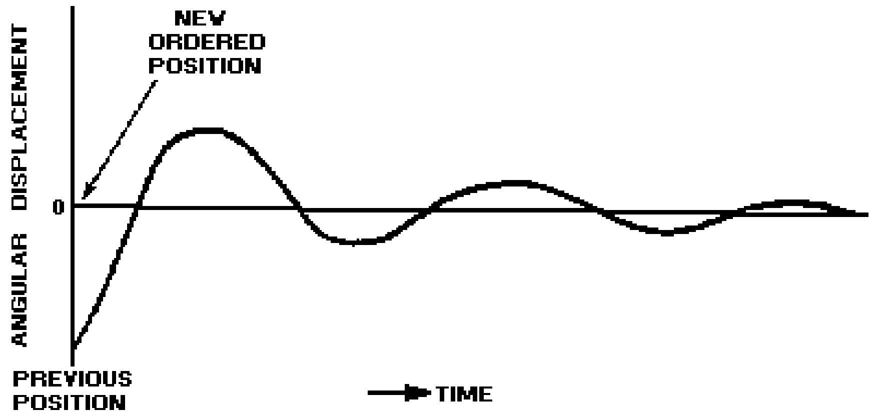


Figure 2-6.—Overtravels of the correspondence point (hunting).

Q-6. What are the three relatively common classifications of servo systems by function?

Q-7. The output of the sum point must contain information that controls what two factors of load movement in a position servo?

Q-8. What term is used for a series of overshoots in a servo system?

Velocity Servo Loop

The VELOCITY SERVO is based on the same principle of error-signal generation as the position servo, but there are some operational differences. Two major differences are as follows:

1. In this loop the VELOCITY of the output is sensed rather than the position of the load.
2. When the velocity loop is at correspondence or null position, an error signal is still present and the load is moving.

This type of servo is used in applications where the load is required to be driven at a constant speed. This speed is governed by the level of the error signal present. Radar antennas, star-tracking telescopes, machine cutting tools, and other devices requiring variable speed regulation are all examples of the types of load this servo may be used to drive.

Figure 2-7 is a block diagram of a velocity servo. It is similar to the block diagram of the position servo loop except that the velocity servo loop contains a TACHOMETER in the feedback line. The tachometer (tach) is a small generator that generates a voltage proportional to its shaft speed.

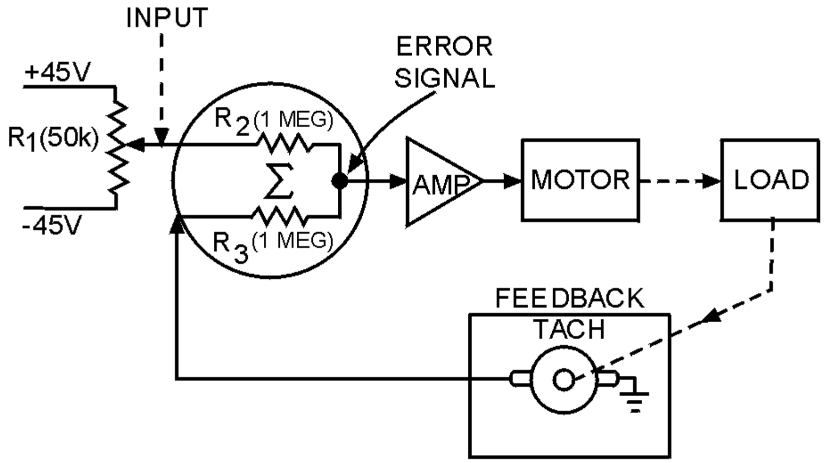


Figure 2-7.—Block diagram of a velocity servo.

In this application, the tach is used as a feedback device and is designed to produce 1 volt of feedback for each 10 rpm.

Let's assume that the motor is designed to turn 10 rpm for each volt of error signal. Figure 2-7 shows the tach mechanically connected to the load. With this arrangement, the shaft of the tach rotates as the load rotates, and the tach can be said to "sense" the speed of rotation of the load. For purposes of explanation, we win assume that the load is an antenna that we want to rotate at 30 rpm.

Initially, the wiper arm of R_1 is set at the 0-volt point (mid-position). This applies 0 volts to the left side of R_2. Since the motor is not turning, the load is not being driven, and the tach output is 0 volts. This applies 0 volts to the left side of R_3. Under these conditions, 0 volts is felt at the sum point and the motor is not driven. The voltage at the sum point is the error signal. When the wiper arm of R_1 is moved to the −9 volt point, an error signal appears at the sum point. At the first instant, the error signal (at the sum point) is −4.5 volts. This is because, at the first instant, the load and tach have not started to move. With the tach output at 0 volts, and the wiper of R_1 at −9 volts, −4.5 volts is present at the sum point. This voltage will cause the motor to start to rotate the load.

After a period of time, the load (and tach) are rotating at 10 rpm. This causes the tach to have an output of +1 volt. With +1 volt from the tach applied to the bottom of R_3, and −9 volts (from R_1 wiper) applied to the top of R_2, the voltage at the sum point (error signal) is –4 volts. Since the motor will turn 10 rpm for each volt of error signal, the motor continues to speed up. When the load reaches 30 rpm, the tach output is +3 volts. With this +3 volts at the bottom of R_3 and the −9 volts at the top of R_2, the error signal at the sum point is −3 volts. This −3 volts is the voltage required to drive the motor at 30 rpm, and places the system in balance. This satisfies the two conditions of the velocity servo. (1) The velocity of the output is sensed (by the tach), and (2) an error signal (−3 volts) is still present and the load continues to move when the velocity loop is at correspondence (30 rpm).

You may ask why the velocity loop and feedback are necessary. If this motor turns 10 rpm for each 1 volt error signal, why not simply feed −3 volts into this amplifier from the wiper of R_1 and not have a tach or summing network?

The answer is that the velocity loop will regulate the speed of the load for changing conditions. If the load in figure 2-7 were a rotating antenna on a ship, the antenna would tend to slow down as the wind opposed its movement and speed up as the wind aided its movement. Whenever the antenna slowed down, the output of the tach would decrease (since the tach is connected to the load). If the tach output decreased, the error signal would increase in amplitude and cause the motor to speed up. In the same way, if the antenna were to speed up, the tach output would increase, decreasing the error signal and the motor would slow down. Without the velocity loop to compensate for changing conditions, the load could not respond in the desired manner.

The system shown in figure 2-7 is a simplified version of a velocity loop. In practice, the reaction of the motor to error voltage and the output of the tach would not be equal (10 rpm per volt and 1 volt per 10 rpm). This would be compensated for by gearing between the motor and load and between the load and tach, or by using a summation network in which the resistors (R_2 and R_3) are riot equal. This use of unequal resistors is called a SCALING FACTOR and compensates for tach outputs and required motor inputs. This is just another way of saying that the individual components of the velocity loop must be made to work together so that each can respond in a manner that produces the desired system result.

Q-9. What are two major differences between velocity servos and position servos?

Q-10. In a typical velocity servo block diagram what device is placed in the feedback loop that is not present in the position servo?

Q-11. What is the advantage of using a closed-servo loop to control load velocity?

The Acceleration Servo

The acceleration servo is similar to the two loops we just discussed except that the acceleration of the load is sensed, rather than the position or velocity. In this loop, the tachometer of the velocity loop is replaced by an accelerometer (a device that generates a signal in response to an acceleration) as the feedback device.

We have not provided an illustration of the acceleration servo because of the complexity of its applications as well as its components. This type of servo is widely used in the rocket and missile fields, and is used whenever acceleration control is required.

SERVO CHARACTERISTICS

Servo characteristics vary primarily with the job the servo is designed to do. There are almost as many types of servos as there are jobs for servos. All servos usually have the common purpose of controlling output in a way ordered by the input. Ideally, motion and output shaft position should duplicate the track of the input shaft. However, this ideal performance is never achieved. We will discuss the major reasons for this, and show some methods used in the attempt to approach the ideal.

Because a servo compares an input signal with a feedback response, there will always be a TIME LAG between the input signal and the actual movement of the load. Also, the weight of the load may introduce an additional time lag. The time lag of the servo can be decreased by increasing the gain of the servo amplifier. If the gain is set too high, however, the servo output will tend to oscillate and be unstable. From this you can see that the gain of a servo is limited by stability considerations. Servo sensitivity must be considered along with stability to reach a "happy medium."

TIME LAG

To reduce time lag, the gain of the servo amplifier could be increased. Increasing the gain of the servo amplifier will decrease the lag time and cause the load to move faster. However, there is a serious drawback because the load is moving faster, its inertia will likely cause it to go past the desired position (overshoot). When the load attempts to drive back to the desired position, the high gain of the amplifier may cause it to overshoot in the opposite direction. Therefore, the system must be stabilized to minimize or eliminate the problem of overshoot. This is done through DAMPING. Damping can be done by either introducing a voltage in opposition to the signal voltage or placing a physical restraint on the servo output. The actual function of this antihunting is to reduce the amplitude and duration of the oscillations that may exist in the system. Every system has one or more natural oscillating frequencies that depend on the weight of the load, designed speed, and other characteristics.

The degree of damping is determined by the purpose and the use of the system. If the system is OVERDAMPED, it will not be bothered by oscillations. However, the large amount of restraint placed on the servo presents an additional problem. This is an excessive time requirement for the system to reach synchronization. Figure 2-8 is a graphic representation showing the time relationship with regard to degree of damping.

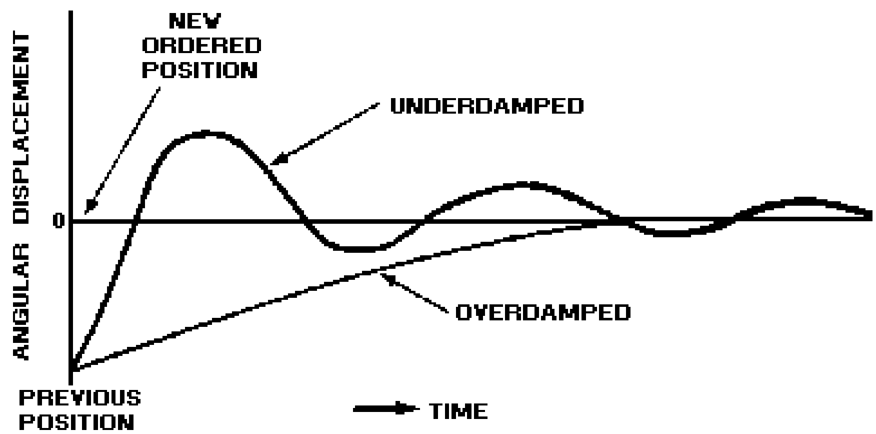


Figure 2-8.—Degree of damping.

An UNDERDAMPED servo system has other traits. The favorable one is its instantaneous response to an error signal. The unfavorable trait is an erratic operation around the point of synchronization because of the low amount of restraining force placed on the servo. Somewhere between overdamped and underdamped, there is a combination of desirable accuracy, smoothness, and moderately short synchronizing time.

The simplest form of damping is FRICTION damping. Friction damping is the application of friction to the output shaft or load that is proportional to the output velocity. The amount of friction applied to the system is critical, and will materially affect the results of the system. Friction absorbs power from the motor and converts that power to heat.

A pure friction damper would absorb an excessive amount of power from the system. However, two available systems have some of the characteristics of a friction damper, but with somewhat less power loss. These are the *friction clutch* and the *magnetic clutch*.

Q-12. If a position servo system tends to oscillate whenever a new position is selected, is the system overdamped or underdamped?

Q-13. If a position servo system does not respond to small changes of the input, is the system overdamped or underdamped?

Friction Clutch Damping

The friction clutch damper uses a friction clutch to couple a weighted flywheel to the output drive shaft of the servo motor. As the servo motor rotates, the clutch couples some of this motion to the flywheel. As the flywheel overcomes inertia and gains speed, it approaches the motor speed. The flywheel, in turning, absorbs energy (power) from the servo motor. The amount of energy stored in the flywheel is determined by its speed (velocity). Because of inertia, the flywheel resists any attempt to change its velocity.

As the correspondence point of the system is approached, the error signal is reduced and the motor begins to slow down. In an attempt to keep the output shaft turning at the same speed, the flywheel releases some of its energy into the shaft. This causes the first overshoot to be large. When the servo system drives past the point of correspondence, a new error signal is developed. The new error signal is of opposite polarity and causes the servo system motor to drive in the opposite direction. Once again the flywheel resists the motor movement and absorbs energy from the system. This causes a large reduction in the second overshoot and all subsequent overshoots of the system. The overall effect is to dampen the oscillations about the point of correspondence and reduce the synchronizing time.

The motor rotation is transmitted to the flywheel through the friction clutch. The inertia of the flywheel acts as an additional load on the motor. The friction clutch is designed to slip with a rapid change of direction or speed. This slipping effectively disconnects the flywheel instantaneously, and thus governs the amount of power the flywheel draws from the motor.

Magnetic Clutch

Another type of damper is the MAGNETIC CLUTCH. This type is similar in function to the friction-clutch damper. The main difference between the two is the method used to couple the flywheel to the shaft of the servo motor. There are two distinct types of magnetic clutch dampers. The first uses a magnetic field to draw two friction clutch plates together to produce damping. The action is similar to the friction clutch we just described.

The second version of the magnetic clutch uses the action of a magnetic field generated by two sets of coils, or one set of coils and the induced eddy currents, which result from rotation of the single set of coils near a conducting surface (the flywheel).

Coupling in this type of clutch is made by the interaction of two magnetic fields without a physical contact between the two. The two-coil or eddy-current type of magnetic clutch offers smoother operation than a pure friction clutch and has no problem of wear because of friction.

In summary, a smooth, efficient operating servo system can only be achieved by a system of compromises. As you recall, earlier we increased the gain of the amplifier to reduce time lag. This had the drawback of increasing hunting or oscillations about the point of correspondence. We overcame this difficulty through friction damping. This solved the problem of hunting and smoothed out servo operation but acted as part of the servo load. It caused a large first overshoot and increased the time lag. Some form of damping that can be used with high amplification to obtain smooth servo operation and minimum time lag is needed. The answer lies with the use of ERROR-RATE damping.

Q-14. Why is damping needed in a practical servo system?

Error-Rate Damping

Error-rate damping is a method of damping that "anticipates" the amount of overshoot. This form of damping corrects the overshoot by introducing a voltage in the error detector that is proportional to the rate of change of the error signal.

This "correction" voltage is combined with the error signal in the proper ratio to obtain the desired servo operation with reduced overshooting and minimum time lag.

The advantages of error-rate damping are as follows:

1. Maximum damping occurs when a maximum rate of change of error signal is present. This normally would occur as the servo load reverses direction.

2. Since a CHANGE in the signal causes damping, there is a minimum amount of damping when no signal, or a signal of constant strength, is present.

Error-rate voltages are generated by either electromechanical devices or electrical networks in the equipment. One electromechanical device widely used to generate an error-rate voltage is the tachometer generator. As you learned previously, its output voltage is proportional to the output velocity of the servo. Hence, the output voltage of the tachometer can be used to anticipate sudden movement changes of the load.

The compensating electrical network used for error-rate damping consists of a combination of resistors and capacitors forming an RC, differentiating or integrating network. You should recall that a differentiating circuit produces an output voltage that is proportional to the rate of change of the input voltage and that an integrating circuit produces an output proportional to the integral of the input signal.

Figure 2-9 shows a basic RC INTEGRATOR. It can be recognized by the output voltage being taken across the capacitor. R_1 is added in this circuit to develop the transient error signal (small variation in the signal from the error detector). The RC integrator is sometimes referred to as an INTEGRAL CONTROL CIRCUIT and will be used to explain electrical error-rate damping.

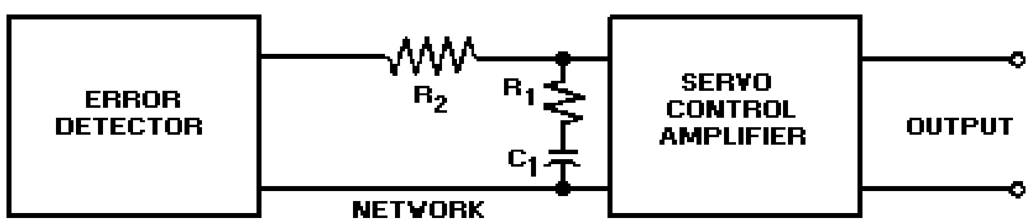


Figure 2-9.—Error rate stabilization network using an RC integrator.

The network consists of a capacitor and two resistors connected in series with the servo amplifier. The components of this circuit are designed to work with a constant or very slowly changing error signal.

Initially, all of the error voltage is divided between R_1 and R_2. But the longer the error voltage is applied, the more C_1 charges, and the greater the voltage at the input of the amplifier. Because of the RC time of the circuit, it takes time for the capacitor to charge to the value of the error input signal. Because of the long charge time of C_1, the circuit can not respond instantaneously to a rapid change in error signal.

What this means is that all error signals will be integrated (or smoothed out). The load will not respond as quickly. The inertia of the load will be reduced, and the system will be damped.

The capacitor, by not responding instantaneously to the error signal, causes the damping action. This action is used to stabilize the servo system at the new velocity. By tailoring the stabilization network (through the proper selection of the RC components) to the system's performance requirements and the type of load to be driven, undesirable load or performance characteristics can be minimized.

The various compensating networks that you will encounter will depend on the design of the individual servo system and will be covered in the associated system's technical manual.

In summary, the key to understanding compensating networks is to realize that components are chosen so the capacitor does not have time to charge and discharge in response to large, rapid fluctuations.

Q-15. Error-rate damping is effective because the circuitry has the capability of ______________the amount of overshoot before it happens.

FREQUENCY RESPONSE

The frequency response of a servo is the range of frequencies to which the system is able to respond in moving the load. It is a characteristic of the system, chosen by the designers so the system will be able to respond to whatever frequencies are expected to be present in the input signal for the particular application.

Oscillating Input Signal

At first, we considered the input order to a servo as being suddenly put at a fixed desired value. Later, we studied the case where the order slowly increased to the desired value. Actually, the input order to a servo in a given application may accelerate, start, stop, or oscillate about a fixed point. We will now consider the actions of a servo while the order oscillates. When the order is constant, oscillations of the load are undesirable. When the order oscillates, the load must oscillate in a similar manner.

Let's assume that an oscillating input signal (order) is applied to a servo. The load may behave in several ways. Ideally, it would respond in perfect sync with the order. Actually, the amplitude and phase of the load are different from those of the order, figure 2-10. As we noted above, the frequency response of the system is normally designed so the load is able to respond to the order.

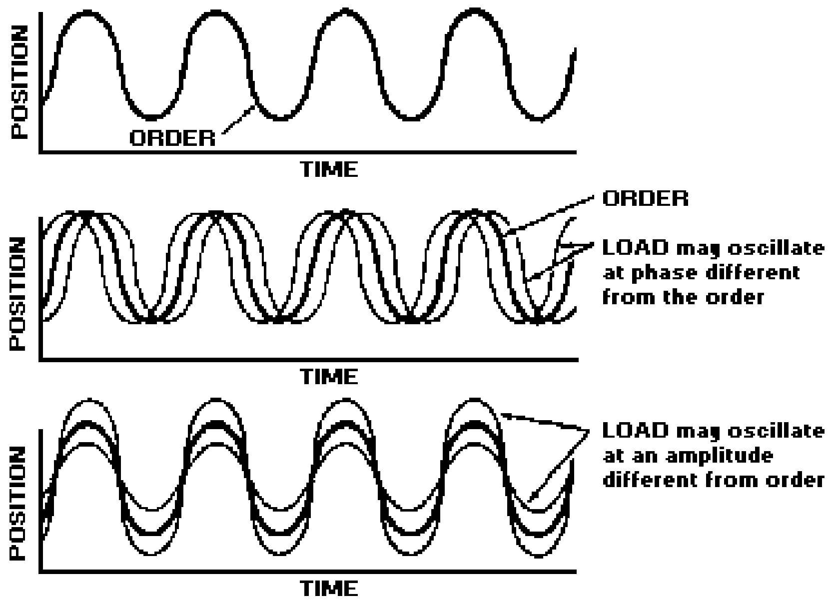


Figure 2-10.—Frequency response.

A servo may follow the order in amplitude and differ in phase; it may follow the order in phase, and differ in amplitude; or it may differ in both phase and amplitude.

Bandpass Frequencies in a Servo system

Servos are plagued by noise signals that ride through the system on desired electrical signals. These noise signals cause roughness in the servo system and must be eliminated to obtain smooth servo operation.

By examining the different signals in a servo system, we can determine which frequencies are related to the movement of the load and which ones are from noise sources, such as static, motors, harmonics, and mechanical resonances.

Filters in the signal circuit can be used to shunt some of the unwanted frequencies away from the amplifier, and allow only those frequencies that represent load movement to enter the amplifier. This can also be accomplished by designing the BANDWIDTH of the servo amplifier to accept only the range of frequencies that represents valid servo signals and to reject all others. This smooths servo response, but has the drawback of reducing amplifier gain. Reduced amplifier bandwidth is another compromise in achieving optimum servo operation.

Q-16. In a properly designed servo system that has an oscillating input (order), what should be the response of the load?

Q-17. What is the advantage of designing a limited bandwidth into a servo amplifier?

SERVO COMPONENTS AND CIRCUITS

In this section we will discuss the circuits and components that make up a typical servo system. We cannot cover all possible servo applications here because of the vast number of servo system configurations. The circuits and components discussed in the following pages are the most commonly used and represent a broad view of the systems used in the Navy today. We have not attempted to put the units into any rigid classification system. We will mention some of the more common terms used by manufacturers and the Navy to classify the devices to familiarize you with the wide variety of nomenclatures.

We will be covering much of the electronic application without discussing the theory of the units. You may want to review some of the applicable NEETS modules or other sources before or during this discussion. You will find that much of the material necessary to understand these subjects is contained in the basic theory of electricity and electronics.

POSITION SENSORS

A position sensor is a device that changes a mechanical position into a voltage that represents that position. The output of a position sensor can be either ac or dc voltage. There are many different kinds of position sensors. In the last chapter you learned about the CX, a synchro device that represents the position of its rotor by a voltage on its stators. You saw a CX used as a position sensor in a servo system earlier in this chapter. Other devices can be used as position sensors. The potentiometer is one of these devices.

Potentiometers

Potentiometer position sensors are generally used only where the input and output of the servo mechanism have limited motion They are characterized by high accuracy and small size, and may have either a dc or an ac output voltage. Their disadvantages include limited motion and a life problem resulting from the wear of the brush on the potentiometer wire. Also the voltage output of the potentiometer changes in discrete steps as the brush moves from wire to wire. A further disadvantage of some potentiometers is the high drive torque required to rotate the wiper contact.

A potentiometer is one of the simplest means of converting mechanical positional information to a proportional voltage. A schematic representation of a potentiometer is shown in figure 2-11.

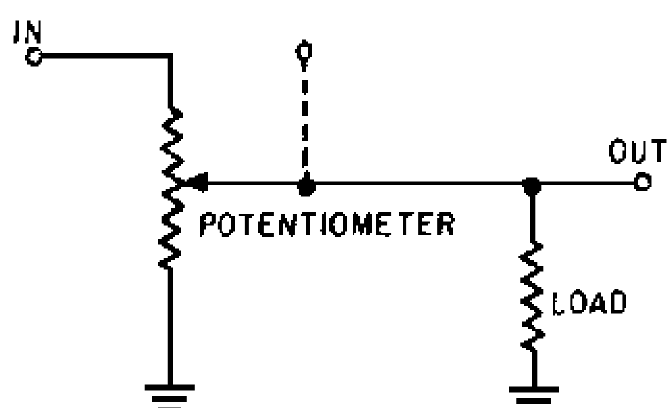


Figure 2-11.—A potentiometer.

A potentiometer is a variable voltage divider, with an output voltage that is a percentage of the input voltage. The amount of output voltage is proportional to the position of the wiper relative to the grounded end. For example, if the resistance from ground to the wiper is 50% of the total, the output voltage sensed by the load will be 50% of the total voltage across the potentiometer.

A basic, closed-loop servo system using a balanced potentiometer as a position sensor is shown in figure 2-12.

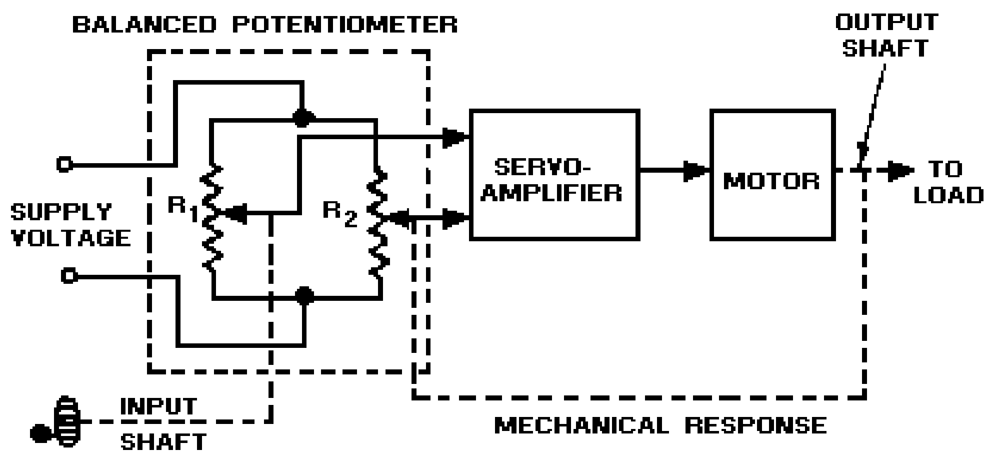


Figure 2-12.—Balanced potentiometer used In position sensing.

The command input shaft is mechanically linked to R_1, and the load is mechanically linked to R_2. A supply voltage is applied across both potentiometers.

The system is designed so that when the input and output shafts are in the same angular position, the voltages from the two potentiometers are equal and no error voltage is felt at the amplifier input. If the input shaft is rotated, moving the wiper contact of R_1, an error voltage is applied to the servo amplifier. This error voltage is the difference between the voltages at the wiper contents of R_1 and R_2. The output of the amplifier causes the motor to rotate the load and the wiper contact of R_2. This continues until both voltages are again equal. When the voltages are equal, the motor stops. In effect, the position of the output shaft has been sensed by the balanced potentiometer.

Q-18. When the input and output wipers of a balanced potentiometer are in the same angular position, what is the value of the error voltage?

ERROR DETECTORS

Electrical error detectors may be either ac or dc devices, depending upon the requirements of the servo system. An ac device used as an error detector must compare the two signals and produce an error signal in which the phase and amplitude will indicate the direction and amount of control, respectively, that are necessary for correspondence. A dc device differs in that the polarity of the output error signal determines the direction of the necessary correction. We will discuss in the following paragraphs various devices that are commonly used in servo systems.

Summing Networks

Summing networks, as we mentioned earlier, are used as error detectors in servo applications where the servo output must be proportional to the algebraic sum of two or more inputs. A typical circuit is shown in figure 2-13.

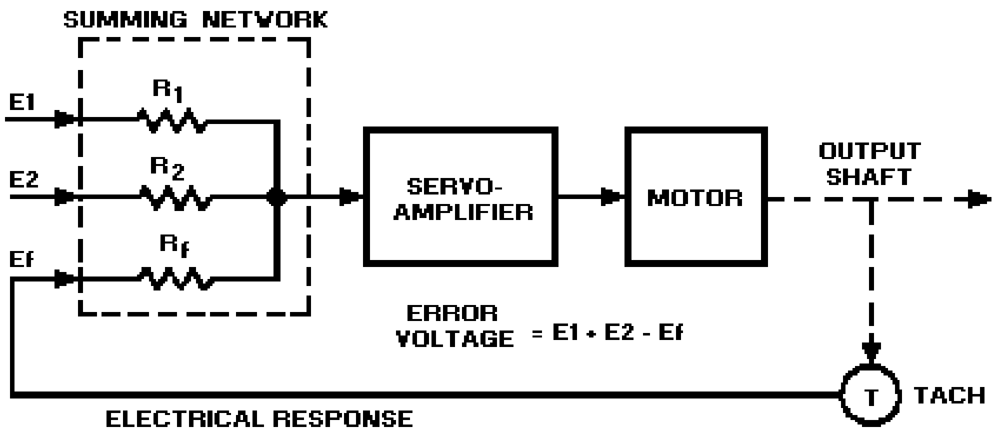


Figure 2-13.—Summing network as an error detector.

As in the case of potentiometers, the networks may use either ac or dc voltage, with the phase or polarity of the input voltage determining whether the signals are additive or subtractive. Refer to figure 2-13. If two input signals E_1 and E_2 are applied to the network, the network will provide an error voltage output that is proportional to the algebraic sum of the two signals. The servo motor drives the load and also a tachometer that supplies feedback voltage to resistor R_f. Resistor Rf nulls the error signal.

In some installations, the servo motor may position the wiper arm of a potentiometer instead of driving a tachometer to supply the feedback voltage.

E-Transformers

The E-transformer is a type of magnetic unit that is used as an error detector in systems in which the load is not required to move through large angles.

In the basic E-transformer shown in figure 2-14, an ac voltage is applied to the primary coil (2) located on the center leg of the laminated, E-shaped core. Two secondary coils (1 and 3) are wound series-opposing on the outer poles of the core. The magnetic coupling between the primary (coil 2) and the two secondaries varies with the position of the armature. The armature can be physically moved left or right in the magnetic circuit by mechanical linkage to the load. This changes the reluctance between either pole and the armature.

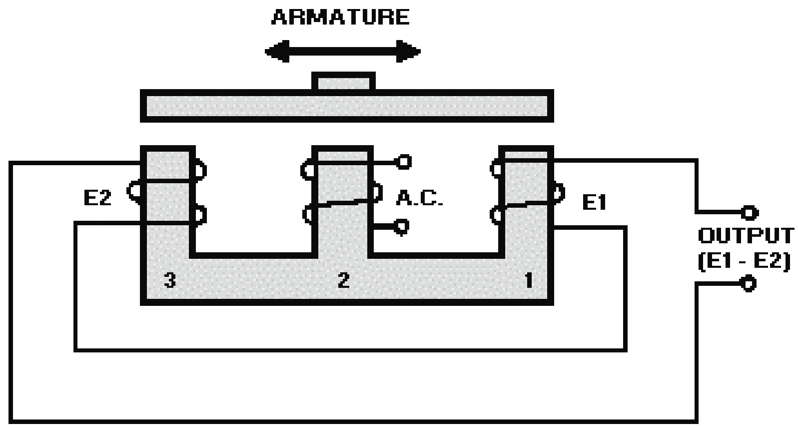


Figure 2-14.—Basic-E transformer.

When the armature is located in the center of the E-shaped core, as shown in the figure, equal and opposite voltages are induced in the secondary coils. The difference between them is zero. Thus, the voltage at the output terminals is also zero.

But, if the armature is moved, say to the tight, the voltage induced in coil 1 increases, while the voltage induced in coil 3 decreases. The two voltages are then unequal, so that the difference is no longer zero. A net voltage results at the output terminals. The amplitude of this voltage is directly proportional to the distance the armature has been moved from its center position. The phase of this output voltage, relative to the ac on the primary, controls the direction the load moves in correcting the error.

The basic E-transformer will detect movement of the armature in one axis only (either the horizontal or vertical depending upon the way the unit is mounted). To detect movement in both the horizontal and vertical axes, a CROSSED-E-TRANSFORMER is used.

If you place two E-transformers at right angles to each other and replace the bar armature with a dome-shaped one (fig. 2-15), you have the basic configuration of what is known as the crossed-E transformer, or pickoff. In most applications the dome-shaped armature is attached to a gyro, and the core assembly is fixed to a gimbal, which is the servo load.

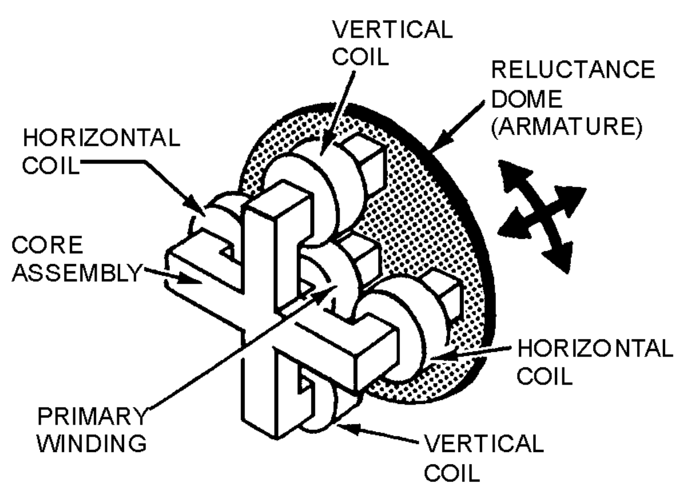


Figure 2-15.—Crossed-E transformer.

The crossed-E transformer assembly consists of five legs (poles). Each leg is encased by a coil. The coil around the center leg is the primary, which is excited by an alternating voltage. The remaining four coils are the secondaries. From this view, you can see how it gets the name, crossed-E.

When the reluctance dome (armature) is moved to the left of center, more flux links the left leg with the primary coil, and the voltage induced in the left secondary increases. The right leg has fewer flux linkages with the center coil; therefore, the voltage induced in the right coil will be less than that in the left coil. Thus there will now be a net voltage out of the pickoff. The phase of the output will be that of the larger voltage. If the dome were moved to the right, the opposite condition would exist. From this brief description, you can see that the crossed-E transformer works on the same fundamental principle as the basic type described earlier. The major difference between the two is in shape and the number of secondaries, and in the fact that the armature has universal movement.

Control Transformers

A commonly used magnetic error detector is the synchro-control transformer, which is used as a control device in servo systems. Recall that we covered the CTs operation in depth in chapter 1 of this module, and discussed its application to the servo system earlier in this chapter.

As an error detector, the CT compares the input signal impressed upon its stator with the angular position of its rotor, which is the actual position of the load. The output is an electrical (error) signal taken from the rotor, which is the difference between the ordered position and the actual position of the load.

A primary advantage of the CT over other types of error detectors is its unlimited rotation angle; that is, both the input and the output to the synchro control transformer may rotate through unlimited angles. A disadvantage is that the output supplied to the servo amplifier is always an ac error signal, and must be demodulated if it is to be used in a dc servo system.

Q-19. In the output of an ac error detector, what indicates the (a) direction and (b) amount of control necessary for correspondence?

Q-20. What two basic types of magnetic devices are used as error detectors?

RATE GENERATOR (TACHOMETER)

As we mentioned earlier, the tachometer in the velocity servo system is the heart of the feedback loop. It is used to sense the speed (velocity) of the load. The tachometer is sometimes referred to as a RATE GENERATOR. Whatever the name, it is a small ac or dc generator that develops an output voltage (proportional to its rpm) whose phase or polarity depends on the rotor's direction of rotation. The dc rate generator usually has permanent magnetic field excitation. The ac rate generator field is excited by a constant ac supply. In either case, the rotor of the tachometer is mechanically connected, directly or indirectly, to the load.

The AC Rate Generator

One type of ac rate generator used widely in the past is the drag-cup type.

The tachometer generator shown in figure 2-16 has two stator windings 90º apart, and an aluminum or copper cup rotor. The rotor rotates around a stationary, soft-iron, magnetic core. One stator winding is energized by a reference ac source. The other stator winding is the generator output, or secondary winding the voltage applied to the primary winding produces a magnetic field at right angles to the secondary winding when the rotor is stationary, as shown in view A. When the rotor is turned by mechanical linkage from the load, it distorts the magnetic field so that it is no longer 90 electrical degrees from the secondary winding. Flux lines cut the secondary winding, and a voltage is induced in the output winding as shown in views B and C. The amount of magnetic field that will be distorted is determined by the speed of the rotor. Therefore, the magnitude of the voltage induced in the secondary winding is proportional to the rotor's velocity (speed).

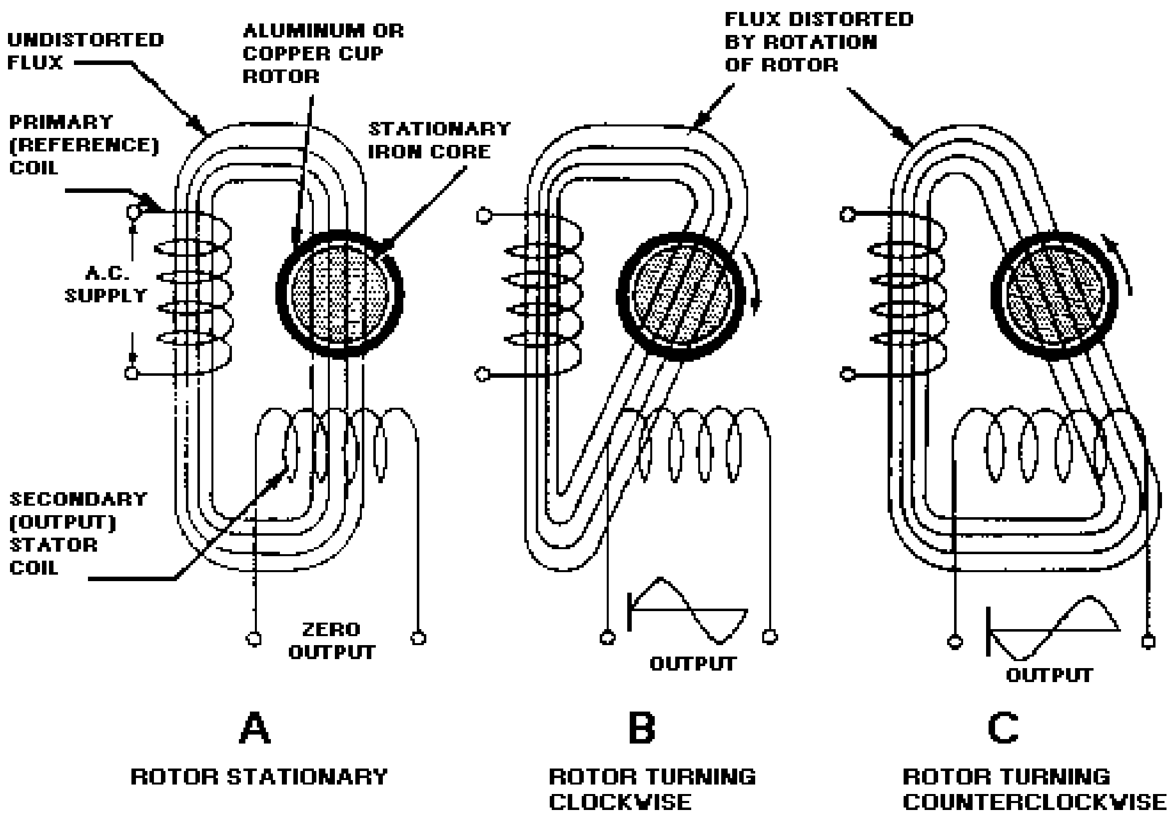


Figure 2-16.—Ac drag-cup rate generator.

The direction of the magnetic field's distortion is determined by the direction of the rotor's motion. If the rotor is turned in one direction, the lines of flux will cut the secondary winding in one direction. If the motion of the rotor is reversed, the lines of flux will cut the secondary winding in the opposite direction. Therefore, the phase of the voltage induced in the secondary winding, measured with respect to the phase of the supply voltage, is determined by the direction of the rotor's motion. The phase relationship is shown in views B and C at the output winding.

The frequency of the tachometer generator output voltage is the same as the frequency of the reference voltage. The output voltage is generated by the primary alternating flux field cutting the secondary winding; therefore, the output voltage must have the same frequency as the supply voltage.

Other types of ac tachometer generators have a squirrel-cage rotor. Otherwise their construction and principles of operation are identical to the drag-cup type.

The DC Rate Generator

The dc rate generator uses the same principles of magnetic coupling as the ac rate generator. The dc rate generator, however, has a steady (nonfluctuating) primary magnetic field. This magnetic field is usually supplied by permanent magnets. The amount of voltage induced in the rotor winding is proportional to the number of magnetic flux lines cut. The polarity of the output voltage is determined by the direction in which the rotor cuts the lines of magnetic flux.

The physical makeup and theory of operation of the dc rate generator (tach) is very similar to the dc generator (NEETS, Module 5, *Introduction to Generators and Motors*). The only major differences are size and the prime mover. The tach is much smaller and is linked mechanically to the servo motor or load instead of to a prime mover.

Tachometer generators are used in servo systems to supply velocity or damping signals and are sometimes mounted on or in the same housing as the servo motor.

Q-21. What is the basic difference between the primaries of ac and dc rate generators?

MODULATORS IN THE SERVO SYSTEM

Because of problems associated with dc amplifiers, such as drift (where the output varies with no variation of the input signal), the ac amplifier is more widely used in servo applications. This creates a need for a device to convert a dc error signal into an ac input for the servo amplifier. Such a device is referred to as a MODULATOR.

Modulator and modulating techniques vary with different types of electronic equipment. The modulator in the servo system performs a completely different function than its counterparts in radar or communications systems.

The servo modulator converts a dc error signal into an ac error signal. The modulator uses two inputs to produce the ac error signal. One input is the dc error signal (for example from an input potentiometer); the other input is an ac reference voltage from some other source, such as the swp's ac supply system. The ac output error signal must contain the same control information that is contained in the original dc error signal. This is done in the following manner:

1. The phase between the ac output and the ac reference signal is determined by the polarity of the dc input signal. The phase of the ac output indicates the direction of error (direction of the load movement).
2. The amplitude of the ac output is proportional to the amplitude of the dc input signal and indicates the amount of error signal (speed or angular displacement of the load).

These relationships of phase and amplitude must be maintained to ensure that the load will move the desired amount, or the proper speed, and in the right direction.

A typical modulator that you will see in a servo system is the CRYSTAL DIODE MODULATOR. The following paragraphs provide a brief explanation of how this modulator works.

Crystal Diode Modulators

The crystal diode modulator (fig. 2-17) consists of a diode bridge and a transformer network. When the ac reference voltage is applied to transformer T_1, diodes CR_2 and CR_3 conduct during the negative half-cycle. Conversely, diodes CR_1 and CR_4 conduct on the positive half-cycle. The diodes will conduct under these conditions because of the 180º phase reversal across T_1. Current flow during the positive and negative half-cycles is represented by dotted arrows and solid arrows, respectively. Suppose a positive, dc error signal is applied during the negative-going ac input half-cycle at the primary of T_1. Current will flow from ground, through the upper half of the primary winding of transformer T_2, through diode CR_2, and through the upper half of the secondary winding of transformer T_1 to the dc source. This produces a positive-going voltage (error signal) across the secondary of T_2 (the first half-cycle of the output signal).

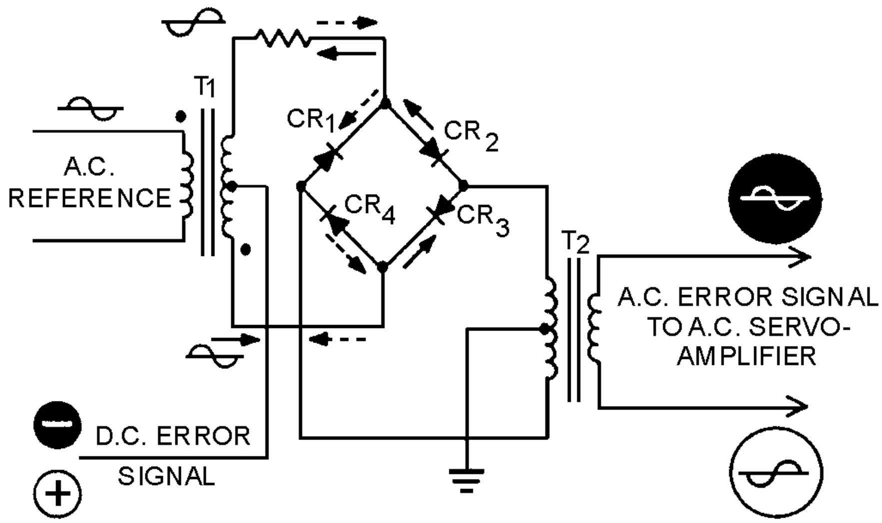


Figure 2-17.—Crystal diode modulator.

On the positive-going ac input reference voltage half-cycle, current will flow from ground, through the lower half of the primary of transformer T_2, through diode CR_4, and through transformer T_1 to the dc error signal source. This produces a negative-going voltage (error signal) across the secondary of T_2 (completing the cycle of the ac input reference). Notice that the error signal is 180º out of phase with the reference signal.

If a negative dc error signal is applied to the modulator, under the same conditions of ac reference signal, current flow through the circuit will be reversed. Keep in mind that this occurs, for example, when the load approaches the desired position from an opposite direction. This circuit will work with either a positive or a negative dc input signal, but only one condition will exist at any given time.

With a negative dc error applied, current will flow from the dc error signal source through diodes CR_3 and CR_1 (on different half-cycles of the ac reference) to ground. This causes an ac voltage to be produced across the secondary of T_2 in the same manner as previously described with the positive dc error signal input.

The only difference is that current will flow through the upper and lower halves of T_2 in a different direction (toward ground) and cause the output to be in phase with the ac reference signal.

In summary, the modulator produced an ac output, either in phase or 180º out of phase with the ac reference signal, depending upon the polarity of the dc input signal. The amplitude of the output will be proportional to the dc input signal amplitude and at the frequency of the ac reference voltage.

Q-22. What is the purpose of a modulator in a servo system?

DEMODULATORS IN THE SERVO SYSTEM

As you know, servo systems use both ac and dc servo motors depending upon the requirements of the system. Systems that are required to move light loads at constant speed use ac motors. Systems that are required to move heavy loads with a wide speed range use dc motors. When the requirements of the system call for a dc motor or other dc devices, the ac error signal within the servo system must be converted to a dc error signal before being fed to the dc servo amplifier. The conversion is made by the circuit known as a DEMODULATOR.

As with the modulator, the demodulator maintains the same relationships between its input and output signals. Just like the modulator, the demodulator's output amplitude is proportional to its input signal and its output polarity is determined by the phase of the input signal. These relationships, as in the modulator you just studied, are necessary so the "new" error signal will control the servo motor in the desired manner.

Diode Demodulator

One example of a servo demodulator is the DIODE DEMODULATOR, sometimes called a phase detector, shown in figure 2-18. This circuit is used in servo systems because it not only converts ac to dc, but it is also able to distinguish the phase of the ac signal by comparing it to a reference voltage. Do not confuse this circuit with other phase detector circuits, such as those used in radar or communications systems. This demodulator (phase detector) distinguishes signals that are either in phase or 180º out of phase. For this reason this circuit is useful in servo systems where the ac output from the error detector (CT) is either in phase with the reference signal or 180º out of phase. Whatever type of error detector is used in the servo system, the reference voltage to the error detector and to the demodulator must be IN PHASE with each other for the demodulator to do its job.

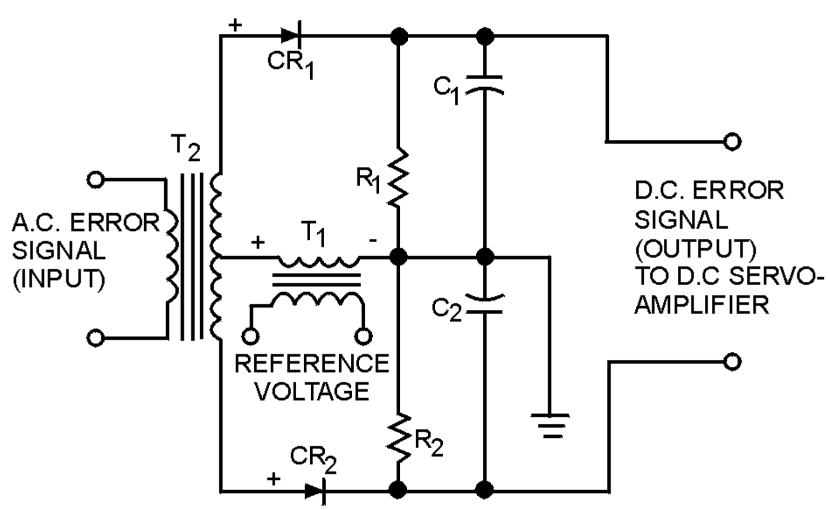


Figure 2-18.—Diode demodulator.

As shown in figure 2-18, the anodes of the two diodes are supplied with the same reference voltage.

With no ac error input signal applied to T_2 (quiescent state), both diodes will conduct equally on the positive half-cycle of the reference voltage. The voltage drops across $R_1$1 and R_2 are equal. This results in the two output terminals being at the same potential; therefore, the output voltage is zero for the positive half-cycle. During the negative half-cycle, a negative voltage is felt on the anodes of both diodes, both diodes are cut off, and zero potential is felt across the output terminals. The circuit will remain in this condition until an ac error signal is applied. As we make this circuit work, you will notice that CR_1 will conduct when the input signal is in phase with the reference voltage and then only on the positive half-cycle. CR_2 will remain in cutoff unless the phase relationship between the ac error signal and the reference voltage changes by 180º. At this time CR_1 will cut off. This change could be brought about by the error detector in the servo system sensing a change in the direction of the load. Effectively, we have a one-diode circuit for one direction of rotation.

Assume that an ac error signal is applied to T_2, making the anode of CR_1 positive and the anode of CR_2 negative. At the same time, the reference voltage on the anodes of CR_1 and CR_2 is on its positive half-cycle. Under these conditions, CR_1 will conduct and CR_2 will be cut off. A positive voltage will be developed across Ri and felt on the output terminals. During the negative half-cycle, a negative voltage will be felt on the anodes of CR_1, and CR_2 and will cut them off. The output of the circuit for one complete cycle of the reference signal will be a filtered, pulsating, dc voltage. As long as the input and reference signals are in phase, the circuit acts as a half-wave rectifier and a filter network.

As we mentioned earlier, this circuit will also respond to a 180º phase reversal between the input and reference signals. For instance, when the error signal applied to T_2 is 180º out of phase with the reference signal, CR_2 conducts and CR_1 cuts off, causing the output voltage to change polarity. You may encounter variations of the diode phase detector; however, they all depend on the same basic principle of operation.

To quickly summarize, the demodulator converted the ac input signal to a dc error signal. The polarity of the dc error signal was determined by the phase relationship between the ac error input signal and the reference signal. The amplitude of the dc error signal was directly proportional to the magnitude of the ac input signal.

Q-23. What is the purpose of a demodulator in a servo system?

SERVO AMPLIFIERS

The servo amplifiers previously discussed were used in servo systems to amplify either the ac or dc error signal to a sufficient amplitude to drive the servo motor. These amplifiers are the same amplifiers in principle as covered in NEETS Module 8, *Introduction to Amplifiers*. The basic amplifier chosen for use in the servo system must have the following characteristics:

1. Flat gain versus frequency response over the broad band of frequencies of interest.
2. Minimum phase shift with a change in input signal (zero phase shift is desired, but a small amount of phase-shift is acceptable, if constant).
3. A low output impedance.
4. A low noise level.

Up to this point in our discussion of servos, the amplifiers have been directly connected to the motor that drove the load. Servo amplifiers are also used within the system itself to amplify the error signal. For example, the signal from the demodulator or filter network may require additional amplification to maintain signal strength. In cases where the amplifier is used to feed large drive motors, to move large loads, the basic electronic amplifier that was presented earlier in this training series is not adequate to do the job. This type of work is done by large power amplifying devices such as the amplidyne generator (NEETS, Module 5, *Introduction to Generators and Motors*) and the MAGNETIC AMPLIFIER, which we will discuss later in this chapter.

AC SERVO MOTORS

Large ac motors are too inefficient for servo use. To move large loads, the ac motor draws excessive amounts of power, and is difficult to cool. Hence, ac servo motors are used primarily to move light loads. Most of the ac servo motors are of the two-phase or split-phase induction type. Fundamentally, these motors are constant-speed devices, although their speeds can be varied within limits by varying the amplitude of the voltage to one of the motors stator windings. When the load becomes heavy, the workhorse dc servo motor is used.

DC SERVO MOTORS

The control characteristics of dc servo motors are superior to those of ac servo motors. The dc servo motor can control heavy loads at variable speeds. Most dc servo motors are either the permanent magnet type, which are used for light loads, or the shunt field type, which are used for heavy loads. The direction and speed of the dc motor's rotation is determined by the armature current. An increase in armature current will increase the motor's speed. A reversal of the motor's armature current will change the motor's direction of rotation More thorough explanations of ac and dc motors are given in NEETS Module 5, *Introduction to Generators and Motors.*

SYNCHRONIZING CIRCUITS

As we explained in chapter 1, the use of a multi-speed synchro transmission system increases the accuracy of data transmission. The accuracy of the servo system depends in part upon the accuracy of the input fed from the synchro system. For example, a dual-speed synchro system operating in conjunction with a servo system uses two CTs (one coarse and one fine) to define a quantity accurately. This is done by feeding the output of the COARSE CT to the servo amplifier when the system is far out of correspondence and then shifting to the output of the FINE CT when the system is within 2 or 3 degrees of synchronization. A circuit that will perform this job is known as a SYNCHRONIZING NETWORK.

A synchronizing network (also called a crossover or switching network) senses how far the servo load is from the ordered position and then switches either the coarse signal or the fine signal into control. The signal selected by the circuit is the input to the amplifier. The selection is based on the size of the error signals the circuit receives. The coarse signal is the predominant factor in the selection, since it is a measure of the servo's output position throughout its limit of motion. The coarse signal drives the system into approximate synchronization, and then the fine signal is shifted into control.

Semiconductor-Diode Synchronizing Network

The SEMICONDUCTOR-DIODE SYNCHRONIZING NETWORK is fairly common and typical of the type used in servo systems. Let's take a look at a circuit that uses this technique. Figure 2-19 is an illustration of the circuit. In the following explanation, we will assume that the system is far out of correspondence (more than 3º). At this time, the coarse signal is large in amplitute. With this condition, CR_3 and CR_4, or CR_5 and CR_6, will be forward-biased, depending upon the polarity of the input signal. This will cause current to flow through R_1. The voltage developed across R_1 is felt on one leg of the summing newtork. A large amplitude fine signal CANNOT be present in the summing network, because CR_1 and CR_2 are designed to limit the fine amplitude to a small value. With this condition present at the summing network, the coarse signal maintains control and drives the load toward correspondence.

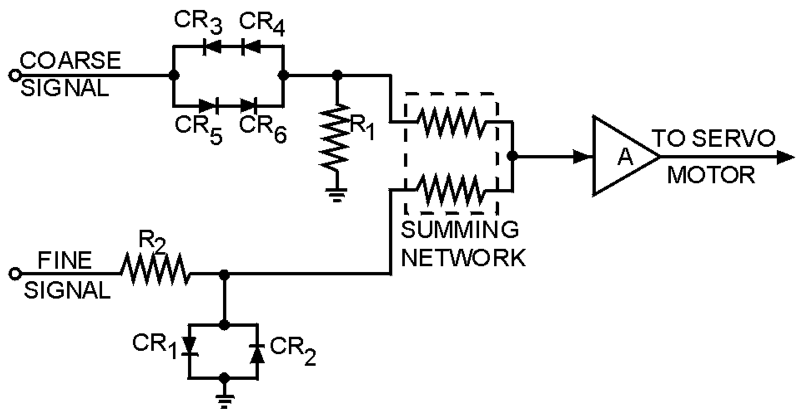


Figure 2-19.—Semiconductor diode synchronizing network.

When the load is within 3º of correspondence, the coarse signal is no longer large enough to forward bias the coarse diode network. The effect of this is to cause a large impedance across the diode network, which then drops most of the coarse signal. Practically no coarse signal voltage is felt across R_1 and one leg of the summing network. On the other hand, the fine signal is also small at this time, since the load is close to correspondence. Small fine signals are unaffected by CR_1 and CR_2. Therefore, the small fine signal is impressed across the summing network. With the fine signal being the only signal felt at the summing network, it takes control and drives the load to the exact point of correspondence. There are various types of synchronizing circuits used in servo systems. Some applications call for electron tubes, relays, and different types of semiconductor diodes. The theory of the specific type you will encounter in servo equipments will be explained in detail in the equipment's technical manual.

Q-24. What is the purpose of a synchronizing network in a servo system?

MAGNETIC AMPLIFIERS

As we stated earlier in this chapter, various types of servo amplifiers are used to drive servo motors. When the amplifier is required to produce a large amount of power, the conventional electronic amplifier becomes less efficient than some other types. The following is a brief discussion of a typical magnetic amplifier used in a servo system where large amounts of power are required to move a heavy load. If you need to refresh your memory on the theory of the magnetic amplifier, refer to Module 8 of this training series, *Introduction to Amplifiers*.

Magnetic Amplifiers in a Servo

Figure 2-20 illustrates a magnetic amplifier being used as the output stage of a servo amplifier.

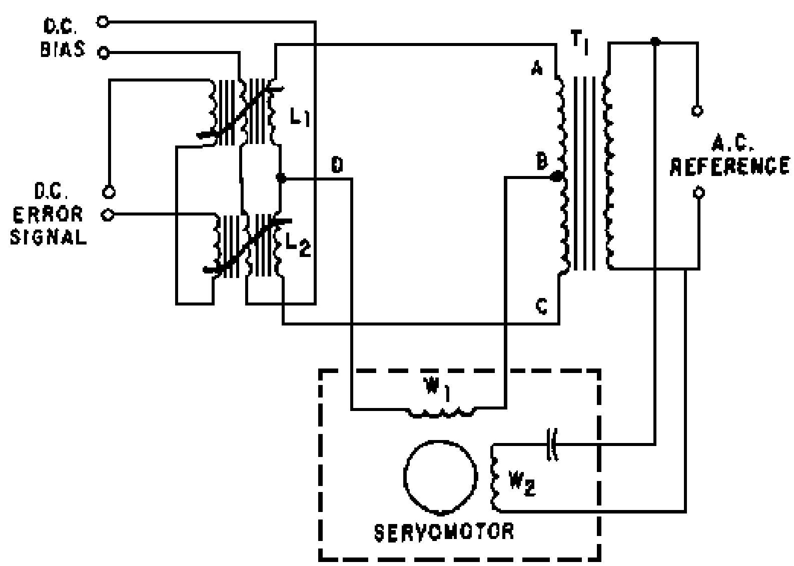


Figure 2-20.—Magnetic amplifier used to drive a servo motor.

The output of the servo amplifier is connected to one of the motor windings (controlled winding W_1). The other winding (uncontrolled winding W_2) is connected across the ac source, in series with a capacitor. The capacitor provides the required 90º phase shift necessary to cause motor rotation. The phase relationship of the current through the two windings determines the direction of rotation of the servo motor.

The magnetic amplifier consists of a transformer (T_1), and two saturable reactors (L_1 and L_2), each having three windings. The key point to the operation of this circuit lies in the fact that the error signal windings are connected in series-opposing while the bias windings are series-aiding.

With the circuit in the quiescent state (no input), the dc bias voltage causes the dc bias current to equally and partially saturate both reactors (L_1 and L_2). The reactances of L_1 and L_2 now being equal result in canceling currents through the W_1 windings of the servo motor. With only one input to the motor, it remains at rest.

Now, let's apply an error signal to the error signal windings. L_2 saturates and L_1 is driven further out of partial saturation, because the error windings are in series-opposition. This results in the error signal aiding the bias current in reactor L_2 and tending to cancel the bias current in reactor L_1. The reactance of L_2 is reduced, causing an increased current through the L_2 circuitry. In the other circuit (L_1), the reverse is true; its current decreases. This imbalance in the L_1 and L_2 circuitry results in current flow through W_1, say from left to right, and causes the motor to turn

Reversing the polarity of the error signal causes the direction of motor rotation to change. This is done by saturating reactor L_1 instead of reactor L_2 and causing current to reverse directions through W_1.

In the previous discussion, an ac motor was driven by the output of the magnetic amplifier. If a dc motor is required in the servo to move a heavy load, the ac output from the magnetic amplifier must be rectified.

NOTE: All of the components that have been described as units within a servo system are, in general, the same components used in many other electronic and electrical applications. The theory of these components has been discussed here and in other modules of the Navy Electricity and Electronics Training Series. If you have the desire or a need for an in-depth study of these components, the following are excellent references:

- Electronics Installation and Maintenance Books, NAVSEA 0967-LP-000-0130, for synchro and servo subjects.

- Electronics Installation and Maintenance Books, NAVSEA 0967-LP-000-0120, for the basic components of the servo system.

These references should be available in the technical library of your ship or station.

Q-25. What the three basic components make up the typical magnetic amplifier?

MULTI-LOOP SERVO SYSTEMS

Now that we have gone through the various servo loops and their components, let's continue our discussion with a realistic application of a servo system.

Very seldom will we find applications where one type of servo loop is used by itself. Usually several loops are combined through the use of various types of relays and switches. The many components of a complex system are caused to work together by switching them in and out as necessary.

Figure 2-21 illustrates a practical application of a multi-loop servo system. You should be able to recognize by now the different loops and components that make up this system. Nothing is really new in the system; we discussed all the loops and components earlier in this chapter.

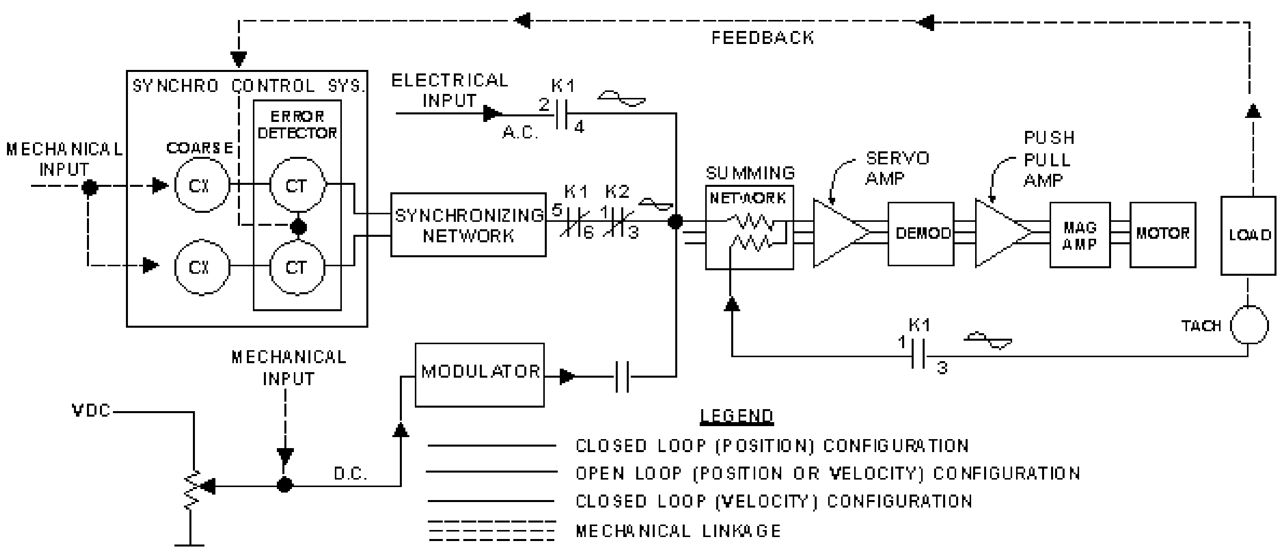


Figure 2-21.—Multi-loop servo system.

As shown by the relay conditions, the system is configured, in its normal state, as a closed-loop position servo. This is indicated by the heavy dark lines in the figure. An alternate configuration positions the load in this system by using the potentiometer. This is done by energizing relay K_2, and switching the system to an open-loop configuration. At the same time, the deenergized contacts of K_2 (1-3) open, thereby breaking the closed loop. The open loop is shown by the medium density lines in the figure. This loop is not as accurate as the closed loop, because the operator must intervene by turning the shaft of the potentiometer back to the zero voltage position to stop the load at the desired position. This type of circuit could be used by maintenance personnel to position the load for easy access to equipments, such as on an antenna or gun mount. The open loop can also function as a basic velocity loop by simply not returning the potentiometer to the zero position. This results in a constant error signal being present at the wiper arm of the potentiometer. With this condition, the load will continue to drive at some speed (rate) determined by the components in the loop.

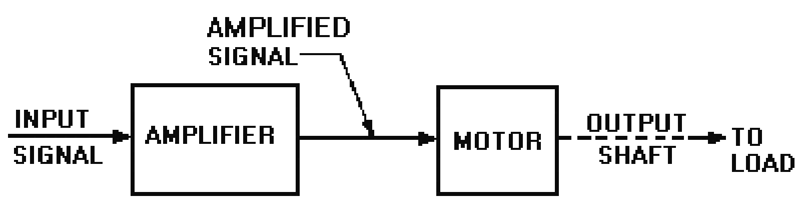
The last loop we will consider is the closed-loop velocity servo, indicated by the fine density lines.

This loop is switched into operation by energizing K_1. Notice that there are two inputs to the summing network with K_1 energized, the electrical input through contacts 2-4 and the feedback from the tach through contacts 1-3. The two signals are compared in the summing network, and their difference is used as the error signal to drive the load. When a state of equilibrium is reached in the circuit, the load will be moving at the desired velocity.

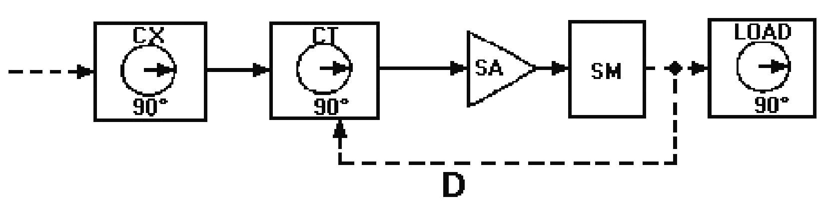
SUMMARY

This chapter has provided information basic to understanding servo systems and their components. The following is a summary of specific points in the chapter.

The **OPEN-LOOP CONTROL SYSTEM** is controlled directly, and only by an input signal. It has no feedback and is therefore less accurate than the closed-loop control system. The open-loop system usually requires an operator to control the speed and direction of movement of the output.



The **CLOSED-LOOP CONTROL SYSTEM** is the most common type used in the Navy. It can respond and move loads quickly and with greater accuracy than the open-loop system. The closed-loop system has an automatic feedback system that informs the input that the desired movement has taken place.



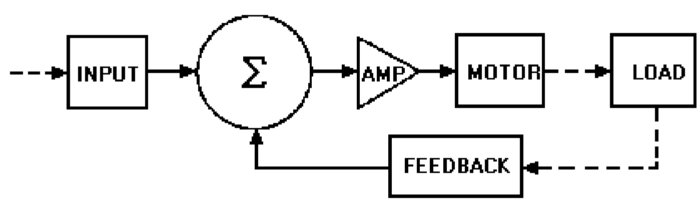
The **SERVO SYSTEM** is classified as a closed- loop system when it is capable of:

1. Accepting an order and defining the desired result,
2. Evaluating present conditions,
3. Comparing the desired result with present conditions and obtaining a difference or an error signal,
4. Issuing a correcting order, and changing the existing conditions to the desired result, and
5. Obeying the correcting order.

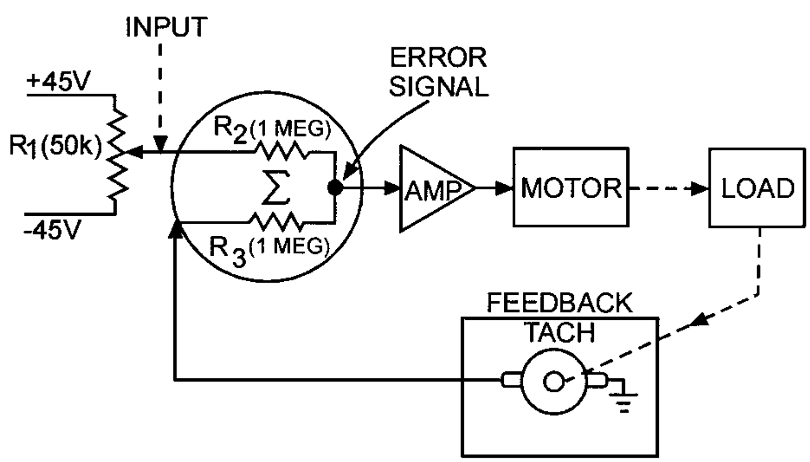
The **BASIC SERVO SYSTEM** is normally made up of electromechanical parts, and consists of a synchro-control system, servo amplifier, servo motor, and some form of feedback.

The **POSITION SERVO** has the goal of controlling the position of the load. In the ac position servo system, the amplitude and phase of the ac error signal determine the amount and direction the load will be driven.

In the dc position servo system, the amplitude and polarity of the dc error signal are used to determine the amount and direction the load will be driven.



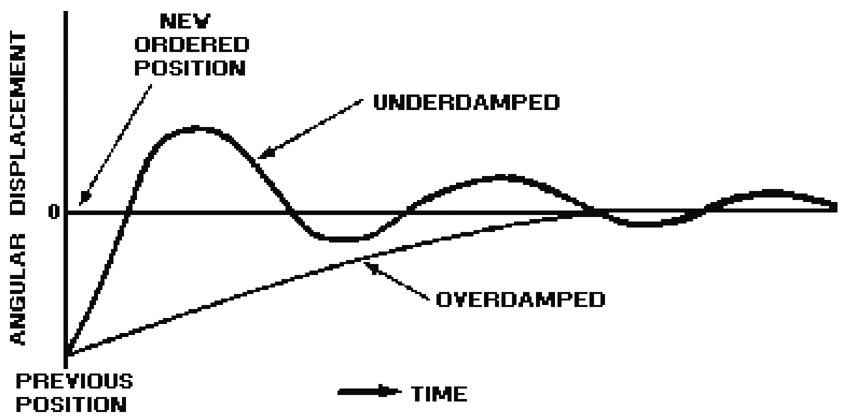
The **VELOCITY SERVO** is based on the same principle of error-signal generation as the position servo, except that the VELOCITY of the output is sensed rather than position of the load. When the velocity loop is at correspondence, an error signal is still present, and the load is moving at the desired velocity.



The **ACCELERATION SERVO** is similar to the velocity and position servos except that the acceleration of the load is being sensed rather than the position or velocity. In this loop, the tachometer of the velocity loop is replaced with an accelerometer.

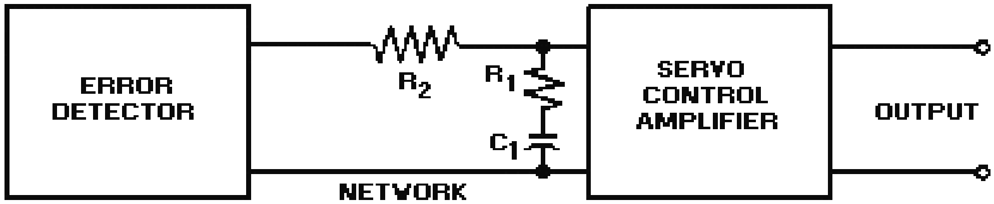
TIME LAG is a servo characteristic defined as the time between the input of the signal and the actual movement of the load. Time lag is undesirable and is reduced through the use of high-gain amplifiers. Damping systems are then added to attain smooth, efficient operation.

An **OVERDAMPED** system will not be subject to oscillations but takes an excessive amount of time to reach synchronization. An UNDERDAMPED system provides instant response to an error signal but results in the load oscillating about the point of synchronism. Somewhere between overdamped and underdamped, designers achieve adequate accuracy, smoothness, and a moderately short synchronizing time.

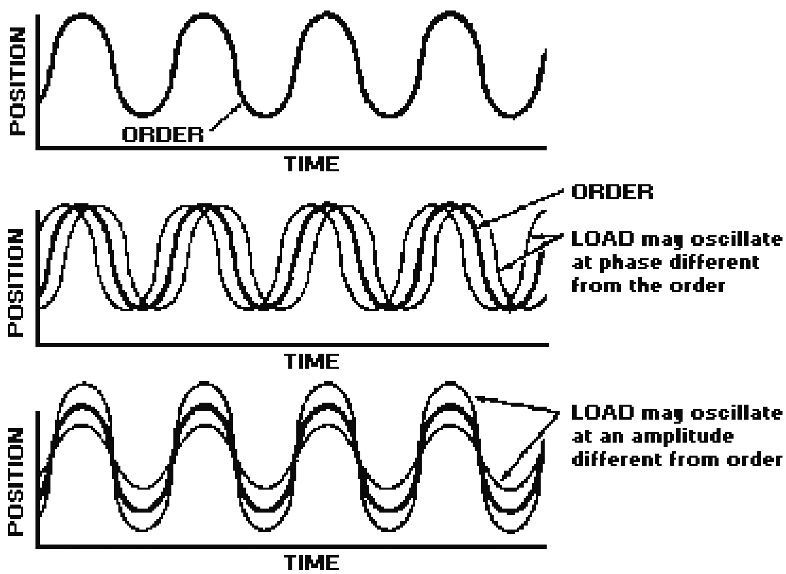


DAMPING is used to stabilize a system-to minimize or eliminate the problem of overshoot. The simplest form of damping is FRICTION CLUTCH damping. MAGNETIC CLUTCH damping is similar to friction clutch damping. The difference is in how the flywheel is coupled to the shaft of the servo motor. Magnetic coupling uses a magnetic field to draw two friction plates together to produce damping. Another method uses the magnetic field set up by a pair of coils or one coil in conjunction with a conducting surface (flywheel) to produce damping.

ERROR-RATE DAMPING is defined as a method of damping that "anticipates" the amount of overshoot. This form of damping corrects the overshoot by introducing a voltage in the error detector that is proportional to the rate of change of the error signal. The stabilization network used for error-rate damping consists of either an RC differentiating network or an integrating network. The components of the RC network are chosen to tailor the stabilization network to the requirements of the servo system.



FREQUENCY RESPONSE of a servo is the range of frequencies to which the system is able to respond in moving the load. The ideal system can respond to whatever frequencies are present in the input signal. Frequency response is a good way of judging servo performance. In a given servo system, good frequency response provides maximum stability and minimum time lag.

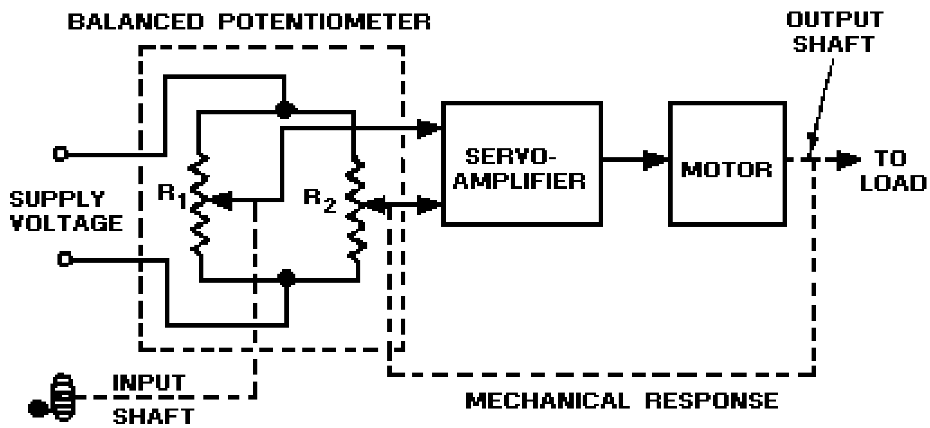


The **BANDWIDTH** of a servo amplifier, ideally, must be able to accept only the range of frequencies that represent valid servo signals.

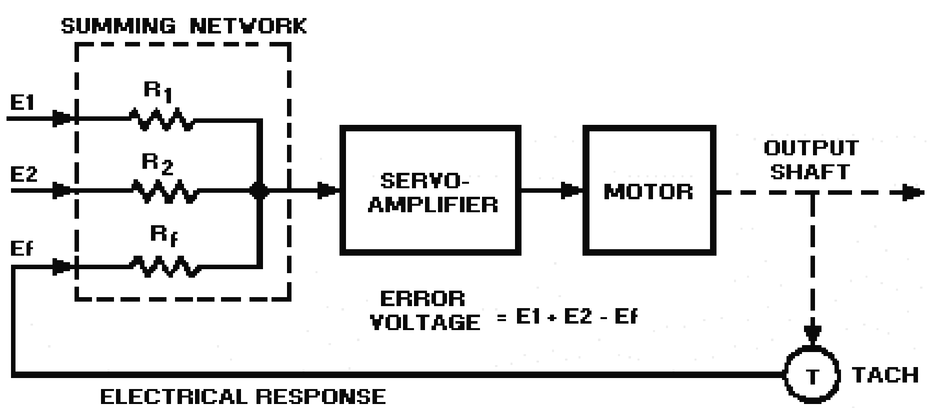
Amplifier bandwidth is another compromise in achieving optimum servo operation.

A **POTENTIOMETER** is one of the simplest position sensor devices and is generally used because of its small size, high accuracy, and output, which can be either ac or dc. Its primary disadvantages are limited motion, limited life due to wear, and high torque required to rotate the wiper contact.

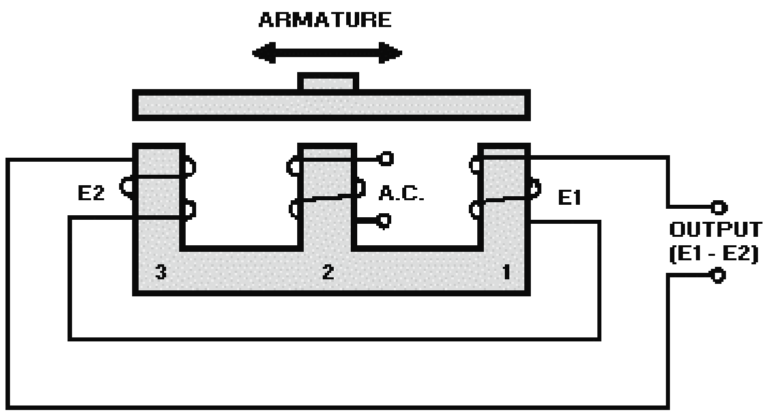
A **BALANCED POTENTIOMETER** in a closed-loop servo system is a voltage divider that functions as a position sensor and produces the error voltage that is fed to the servo amplifier.



SUMMING NETWORKS can be used as error detectors in servo systems to add algebraically two or more inputs and a feedback error voltage.



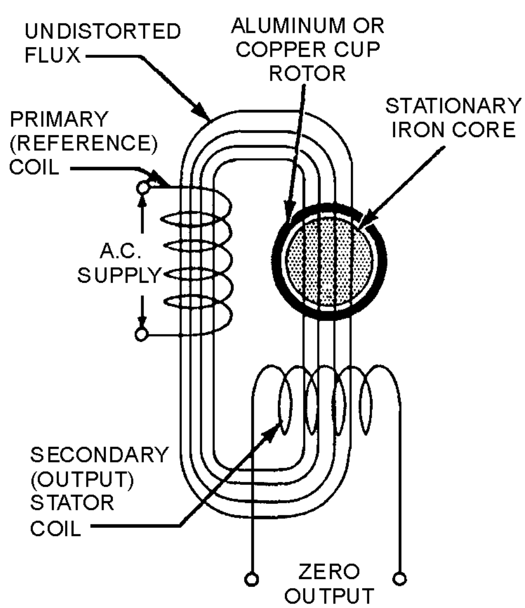
The **E-TRANSFORMER** is a magnetic error detector that can be used in systems limited by large angular movements. Output signals are either in phase, 180º out of phase, or zero, depending on the direction of the E-transformer's armature motion. The amplitude of the signal is determined by the amount of armature motion. The basic E-transformer can only detect motion in one axis.



A **CROSSED-E TRANSFORMER** (or pickoff) is two E-transformers placed at right angles to each other. This type of error detector is capable of detecting error in both horizontal and vertical directions.

A **CONTROL TRANSFORMER (CT)**, when used as a magnetic error detector, can rotate through unlimited angles. The output of this type of CT is always an ac servo error signal that must be demodulated if it is used with a dc servo motor.

A **RATE GENERATOR** (tachometer), when used in the velocity servo loop, is the heart of the feedback loop. The tach senses velocity (speed) of the load. A tachometer can be either an ac or dc rate generator. The output frequency of the ac tach is the same as the reference frequency, varying only in phase depending on the direction of rotation.



MODULATORS are used to change a dc error signal into an ac input error signal for servo amplifiers. This device is required when ac servo amplifiers are used instead of dc amplifiers.

DEMODULATORS convert ac error signals to dc error signals. The dc signal is required to drive a dc servo amplifier.

A **SERVO AMPLIFIER** used in an ac or dc servo system must have a flat gain, minimum phase shift, low output impedance, and low noise level.

AC SERVO MOTORS are used in servo systems that move light loads. Large ac motors are too inefficient for servo use when large loads are to be moved.

DC SERVO MOTORS can control heavy loads, and are widely used in servo systems. The speed and direction of the dc servo motor can be varied easily by varying the armature current.

MAGNETIC AMPLIFIERS are used when power from a conventional servo amplifier is too small to drive large servo motors (either ac or dc).

The **MULTI-LOOP SERVO SYSTEM** combines several closed and/or open servo loops together to control a common load.

ANSWERS TO QUESTIONS Q1. THROUGH Q25.

A-1. A system in which the precise movement of a large load is controlled by a relatively weak control signal.

A-2. Usually the operator senses the desired load movement and reduces the input to stop the motor.

A-3. Feedback.

A-4. Input signal and feedback.

A-5. To move the load and provide feedback data to the error detector.

A-6. Classifications in accordance with position, velocity, and acceleration functions.

A-7. Amount and direction of rotation.

A-8. Hunting.

A-9. Velocity loop senses velocity rather than position. When velocity loop is nulled, an error signal is still present and the load continues to move.

A-10. Tachometer.

A-11. The closed-servo loop can regulate load speed under changing conditions.

A-12. Underdamped.

A-13. Overdamped.

A-14. To minimize overshoot and/or oscillations.

A-15. Anticipating.

A-16. It should oscillate.

A-17. Unwanted noise-generated frequencies are rejected.

A-18. Zero.

A-19. (a) Phase. (b) Amplitude.

A-20. E-transformer and control transformers.

A-21. The method of primary excitation (ac and permanent magnet).

A-22. To convert a dc error signal into an ac error signal.

A-23. To convert an ac error signal into a dc error signal.

A-24. To switch control of the amplifier between either the coarse signal and the fine error signal.

A-25. Two saturable reactors and a transformer.

CHAPTER 3

GYROS

LEARNING OBJECTIVES

Upon completion of this chapter you should be able to:

1. Describe the characteristics of a gyroscopes.
2. List the two basic properties of gyroscopes and explain them.
3. Describe the components of a universally mounted gyro.
4. Describe the factors that determine rigidity in a gyro.
5. List the factors that determine the direction of precession in a gyro.
6. Explain the right-hand rule for gyro precession.
7. Describe the term "Degree of Freedom" as it applies to gyros.
8. Explain the effect of apparent precession (apparent rotation).
9. Explain the purposes of erection systems.
10. Describe the use of gyros with only one degree of freedom.
11. Explain the purpose of an accelerometer.
12. Explain the principle on which accelerometers operate.
13. Explain the need for a pulse-counting accelerometer.

GYROS

The word *gyroscope* was first coined by a French scientist, Leon Foucault, in 1852. It is derived from the Greek words "gyro," meaning *revolution*, and "skopien," meaning to *view*.

The gyroscope, commonly called a GYRO, has existed since the first electron was sent spinning on its axis. Electrons spin and show all the characteristics of a gyro; so does the Earth, which spins about its polar axis at over 1000 miles per hour at the Equator. The Earth's rotation about its axis provides the stabilizing effect that keeps the North Pole pointed within one degree of Polaris (the North Star).

Any rapidly spinning object—a top, a wheel, an airplane propeller, or a spinning projectile—is fundamentally a gyroscope. Strictly speaking, however, a gyroscope is a mechanical device containing a spinning mass that is universally mounted; that is, mounted so it can assume any position in space. Figure 3-1 shows a model of a gyro. As you can see, a heavy wheel (rotor) is mounted so that its spin axis is free to turn in any direction. The wheel spins about axis X; it can turn about axis Y, and it can turn about axis Z. With this mechanical arrangement, the spinning wheel can assume any position in space.

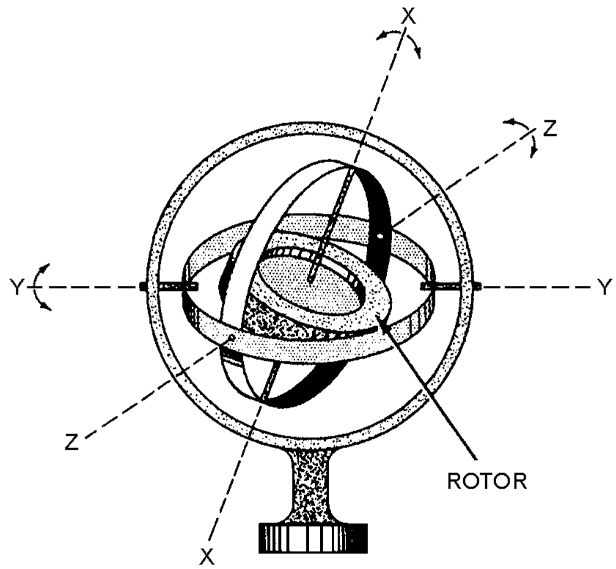


Figure 3-1.—Gyro model, universally mounted.

BASIC PROPERTIES OF GYROSCOPES

Gyroscopes have two basic properties: rigidity and precession. Those properties are defined as follows:

1. RIGIDITY — The axis of rotation (spin axis) of the gyro wheel tends to remain in a fixed direction in space if no force is applied to it.

2. PRECESSION — The axis of rotation has a tendency to turn at a right angle to the direction of an applied force.

The idea of maintaining a fixed direction in space is simple to illustrate. When any object is spinning rapidly, it tends to keep its axis pointed always in the same direction. A toy top is a good example. As long as the top is spinning fast, it stays balanced on its point. Because of this gyro action, the spinning top resists the tendency of gravity to change the direction of its axis. You can think of many more examples. A bicycle is easier to balance at high speed than when it is barely moving. At high speed, the bicycle wheels act as gyros, and tend to keep their axes (axles) parallel to the ground.

Note that it is easy to move the gyro as long as you keep the axis POINTING in the SAME DIRECTION. The gyro resists only those forces that tend to change the direction of its axis. In a bicycle, since the axis of rotation (the wheel's axles) is horizontal, the wheels win resist any force that tends to tilt or turn them to the right or left.

If you can obtain a gyroscope top, you can do some instructive experiments with it. Hold the gyro top with its axis vertical as shown in figure 3-2 and start it spinning. As long as it is spinning fast, it will stay balanced. You can balance it on a string or on the point of your finger; the axis will stay vertical as long as the top is spinning fast. As we mentioned before, this ability of a gyro to keep its axis fixed in space is called RIGIDITY.

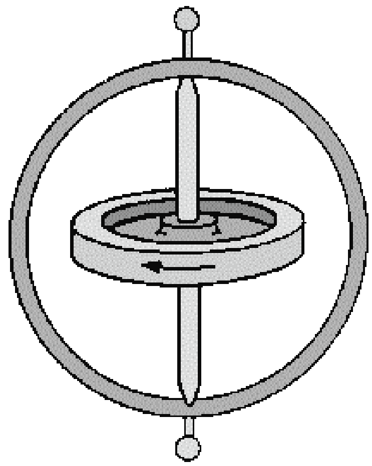


Figure 3-2.—A gyroscope top.

PRECESSION

Now, if you stop the gyro top and turn its axis horizontal, and then start it spinning again, balancing one end on a pivot, (fig. 3-3), it won't fall. The top's axis will stay horizontal, resisting the tendency of gravity to change its direction Although the gyro will RESIST the force that gravity applies to it, the gyro will still RESPOND to that force. The gyro responds by moving its axis at a RIGHT ANGLE to the APPLIED FORCE.

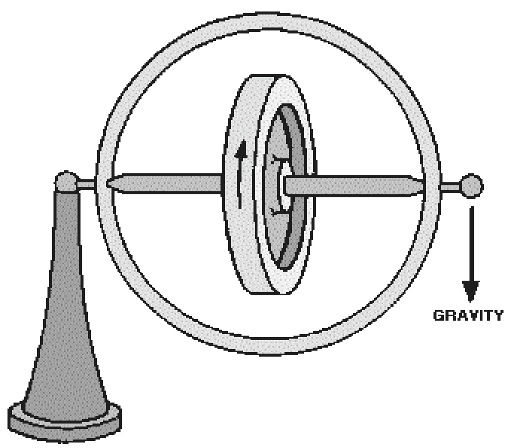


Figure 3-3.—Gyro top with axis horizontal.

The axis will tilt in a direction 90º away from the applied force. This is called PRECESSION.

Figure 3-4 is another view of the same gyroscope. Its far end is still balanced on the pivot. Gravity is pulling down on the gyro. If the gyro rotor is turning in the direction shown by the arrow, the near end of the frame (axis) will move to the left. If the rotor were turning in the opposite direction, the frame would move to the right. Note that in each of these examples the direction of movement was displaced from the applied force (gravity) by 90º. The axis stays horizontal, but the gyroscope responds to the force of gravity by rotating around the pivot.

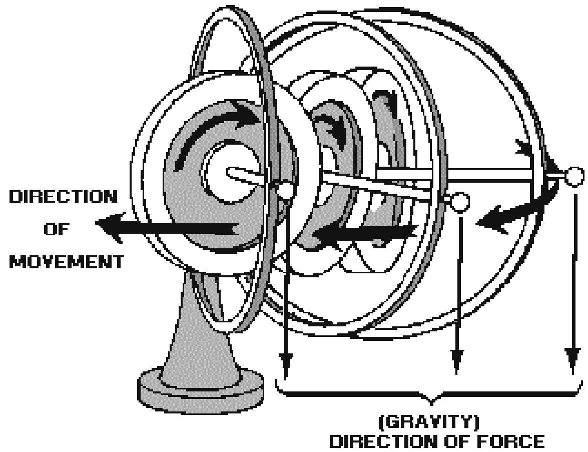


Figure 3-4.—Gyro precession.

Gyro action may be summarized as follows: A spinning gyro tends to keep its axis pointing in the same direction. This is called RIGIDITY. If you apply a force that tends to change the direction of the spin axis, the axis will move at a right angle to the direction of the applied force. The direction of precession will be 90º clockwise from the applied force if the rotor is spinning clockwise (when viewed from the "free" end of the rotor's axis); if the rotor is spinning counterclockwise, the precession will be 90º counterclockwise. If the axis is horizontal, and you try to tilt it, the axis will turn. If the axis is horizontal, and you try to turn it, the axis will tilt. This second characteristic of a gyro is called PRECESSION.

Because of precession, we can control the direction that the spin axis points. This enables us to aim the spin axis where we want it to point. Without precession, the rigidity of the gyro would be useless.

Q-1. Can any rapidly spinning object be considered a gyroscope?

Q-2. In the drawing in figure 3-1, which axis is the gyro spin axis?

Q-3. What gyro property causes the gyro to remain in a fixed position?

Q-4. What type(s) of force does a gyro resist?

Q-5. In what direction will a gyro precess in response to an outside force?

BASIC GYRO ELEMENTS

The gyro shown in figure 3-5 is a basic, universally mounted gyro, sometimes called a free gyro. Its components are rotor, inner gimbal, outer gimbal, and base or support. Gimbals are devices that permit the rotor to assume any position and retain that position when the support is tipped or repositioned. Note that in figure 3-5, the support may be moved about all axes without the rotor position being disturbed.

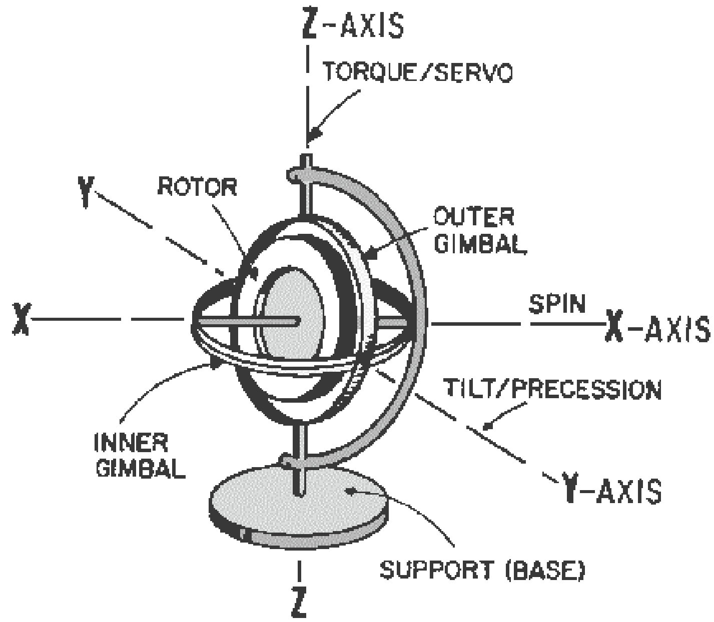


Figure 3-5.—Basic universally mounted gyro.

As you know, gravity is a force that acts along parallel lines upon each particle of matter. A plot of the resultant gravitational force on a body such as a gyro would be equivalent to the sum of these separate forces. The point at which the resultant force is applied is called the CENTER OF GRAVITY. To have a balanced gyro, the center of gravity must be located at the intersection of the three axes of the gyro.

RIGIDITY

A gyroscope is a spinning body that tends to keep its spin axis rigidly pointed in a fixed direction in space. What do we mean by "fixed direction in space"? A fixed direction on Earth is by no means fixed in space, because the Earth turns once on its axis every 24 hours, and makes a complete revolution around the sun every year. The sun itself is moving through space, taking the Earth and the other planets with it. Because of these motions, the expression "fixed direction in space" as used in this explanation is theoretical. For all practical purposes, we can say a line from the Earth to a distant star is a fixed direction in space. If the spin axis of a spinning gyro is pointed at a distant star, it will remain pointed at the star as the Earth turns.

Gyro rigidity is the strength with which a gyro resists any external force that would tilt its rotor spin axis. There are three factors that determine gyro rigidity: weight of the rotor, distribution of this weight, and rotor speed.

The gyro can be considered as an enclosed mechanical system. The energy in the system is equal to the input energy. Hence the energy necessary to spin the gyro rotor is contained in the rotor as angular momentum, which is a function of rotor weight and the speed of rotor rotation. The <u>heavier</u> the gyro rotor, the larger the torque necessary to spin it, and the <u>greater</u> the angular momentum of the rotor. If we have two rotors with identical shapes but different weights spinning at the same velocity, the heavier of the two will be more rigid in its spin axis since it has the greater angular momentum.

Now let's look at the effect of weight distribution in the rotor of a gyro. Consider three rotors of the same weight, as shown in figure 3-6, view (A), view (B), and view (C), with the diameter of one rotor half the diameter of the other two. Now, when we spin these rotors at the same speed, we find that the rotors with the greater diameter are much more rigid than the one with the smaller diameter. Next, we find that we can make the rotors equally rigid by causing the rotor with the smaller diameter to spin faster than the larger rotors. Thus rigidity depends both on speed and distribution of weight. The weight of the larger rotor being farther away from the axis of spin causes it to be more rigid. This effect is even more pronounced if we shape the rotor as shown in view C of figure 3-6. Shifting as much weight as possible to the outer rim of the rotor increases rigidity even further.

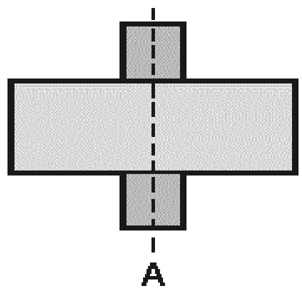


Figure 3-6A.—Gyro rotors with equal weight and unequal diameters.

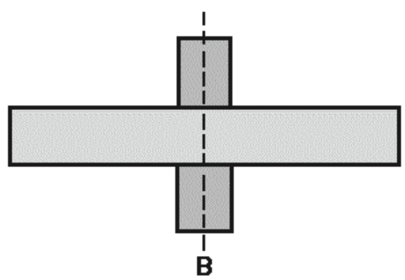


Figure 3-6B.—Gyro rotors with equal weight and unequal diameters.

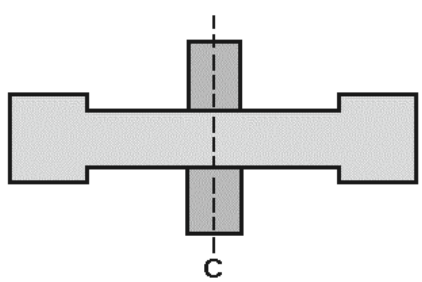


Figure 3-6C.—Gyro rotors with equal weight and unequal diameters.

Q-6. A universally mounted gyro has how many gimbals?

Q-7. What factors determine the rigidity of a gyro?

Q-8. Which gyro rotor in figure 3-6, view (A) view (B) view (C), will have the greatest rigidity if all are rotated at the same speed?

PRECESSION

To provide useful information, a gyro's spin axis must be related to some reference, usually the Earth's surface. This is done by using the second fundamental property of a gyro—PRECESSION. The gyro is precessed until its spin axis is pointed in the desired direction. So far we have covered precession in very general terms. The following paragraphs will cover this "gyro action" in more detail.

We can show precession by using the models in figure 3-7, view (A) and view (B). The gyro wheel is mounted so it is free to have its spin axis pointed in any direction. Here the wheel rotates in a flat loop called the gyro case (inner gimbal). The gyro case is pivoted in the gimbal ring (outer gimbal) and the gyro can swing about the Z axis. The gimbal ring itself turns on pivots that connect it to the fork (support). The fork permits the gyro to tilt from side to side about the Y axis.

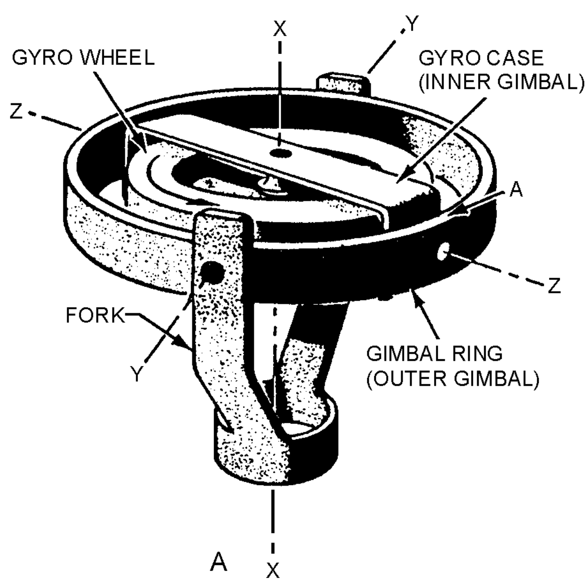


Figure 3-7A.—Gyro action.

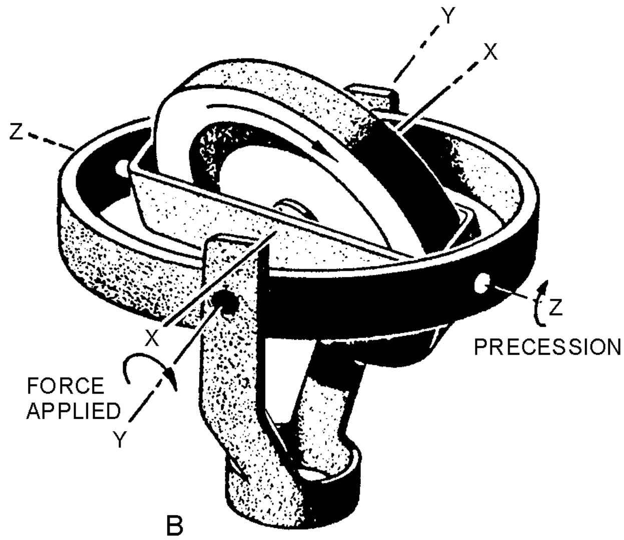


Figure 3-7B.—Gyro action.

Regardless of how the fork is placed, the spinning gyro wheel is free to lie in any given plane. That's why it is called a free gyroscope in this type of mounting.

To show the effect of precession, we can push down on the gimbal ring at point A at the nearer end of the Z-Z axis. (See view A of figure 3-7.) You might expect the ring to tilt around the Y-Y axis. Instead, the gyro case will tilt about the Z-Z axis. You can see the effect of this precession in view B.

Here's a rule that applies to all spinning gyros: THE GYRO WILL ALWAYS PRECESS AT RIGHT ANGLES TO THE DIRECTION OF THE APPLIED FORCE. Look at view A again. If we keep pushing down on the gimbal ring at point (A), the gyro case will keep turning until the spin axis of the gyro wheel is horizontal. Then there will be no further precession. At this point the gyro wheel will be spinning in the same direction in which the applied force is pushing.

Here's another rule: A GYRO ALWAYS PRECESSES IN A DIRECTION TENDING TO LINE ITSELF UP SO THAT ITS ROTOR SPINS IN THE SAME DIRECTION THAT THE APPLIED FORCE IS TRYING TO TURN IT. In other words, the direction of spin chases the applied force. When the direction of spin and the applied force are in the same direction, precession stops.

Now, compare the spin (X) axis in the two parts of figure 3-7. In view A, the spin axis is vertical. In view B, the spin axis has moved from the vertical until it is much closer to being horizontal. By applying the right amount of force in the right place, we have a method of "aiming" the spin axis so that it points to the specific fixed direction in space where we want it. The property of PRECESSION makes the property of RIGIDITY useable.

You should understand that most forces, when applied to the gyro mounting, do not cause precession. For instance, you can swing the fork around in any direction, and the motion will merely be taken up in the Y-Y and Z-Z axes. Similarly, a force applied lengthwise along one of the axes will have no effect.

Any force acting through the center of gravity of the gyroscope does not change the angle of the plane of rotation but moves the gyroscope as a unit. The position of its spin axis in space is not changed. Such forces as those stated above, operating through the center of gravity, are forces of TRANSLATION. In other words, the spinning gyroscope may be moved freely in space by means of its supporting frame, without disturbing the plane of rotation of the rotor. This condition exists because the force that is applied through the supporting frame, acting through the center of gravity produces no torque on the gyro rotor. ONLY THOSE FORCES TENDING TO TILT THE GYRO WHEEL ITSELF WILL CAUSE PRECESSION.

Let's consider further the important characteristic of gyroscopic precession. For a given amount of force, the rate of precession of a gyro is governed by the weight, shape, and speed of the rotor. These factors are the same as those that determine the rigidity of a gyro. Therefore, there is a definite relationship between the rigidity of a gyro and the rate at which a given force will cause it to precess. The greater the rigidity of a gyro, the more difficult it is to cause precession, and the less precession there will be for a given force.

A gyro will resist any force that attempts to change the direction of its spin axis. However, it will move (precess) in response to such force; NOT in the direction of the applied force, but at right angles to it.

The direction a gyro will precess also depends on the direction the gyro is spinning. Precession is actually the result of two forces: angular momentum (spinning force) and the applied force (torque). The direction of precession is always offset from the direction of the applied force. The offset is always in the direction of rotor spin.

For example, when a force is applied upward on the inner gimbal, as shown in figure 3-8, the force may be visualized as applied in an arc about axis Y-Y. This applied force is opposed by the resistance of gyroscopic inertia, preventing the gyro from rotating about axis Y-Y. With the rotor spinning clockwise, the precession will take place 90º clockwise from the point of applied force. The gyro precesses about axis Z-Z in the direction of the arrow "P".

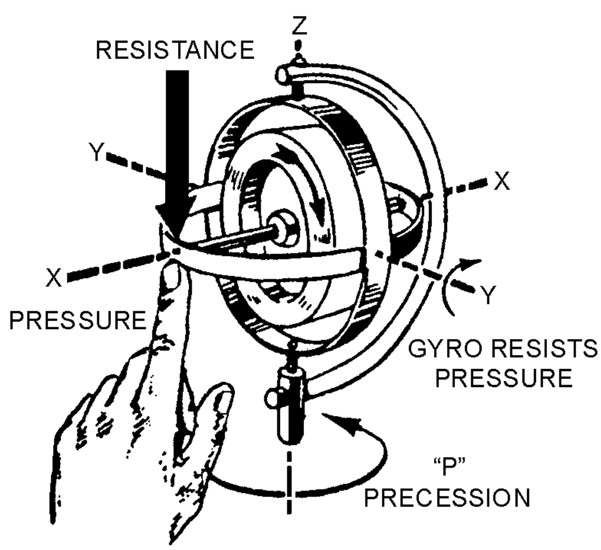


Figure 3-8.—Force applied to a gyro.

The motions of a gyroscope can be analyzed according to three basic quantities:

1. SPIN (the angular velocity of the gyro rotor).

2. TORQUE (the rotary motion applied to change the direction of the rotor axis).

3. PRECESSION (the resulting angular change of the spin axis when torque is applied).

The above quantities are related to vectors so that the relative directions may be easily compared. The SPIN VECTOR lies along the spin axis of the rotor with an arrow indicating the direction of rotation. The TORQUE VECTOR represents the axis about which the applied force is felt. The PRECESSION VECTOR represents the axis about which precession occurs. In all the above cases the direction of the vector is such that the quantity (spin, torque, or precession) is in a clockwise direction if viewed from the tail of the vector.

A simple hand rule will help you determine the direction of the SPIN VECTOR. (See fig. 3-9.) Curve the fingers of your fight hand in the direction in which the rotor is turning as if you intended to grasp the rotor. Your thumb will point in the direction of the spin vector. Similar rules will give you the direction of the TORQUE VECTOR and the PRECESSION VECTOR. With the fingers of your right hand wrapped in the direction of the applied torque (die direction the gyro would rotate if the rotor were not spinning), your thumb points in the direction of the torque vector. Placing your curved fingers in the direction of precession will place your thumb pointing in the direction of the precession vector.

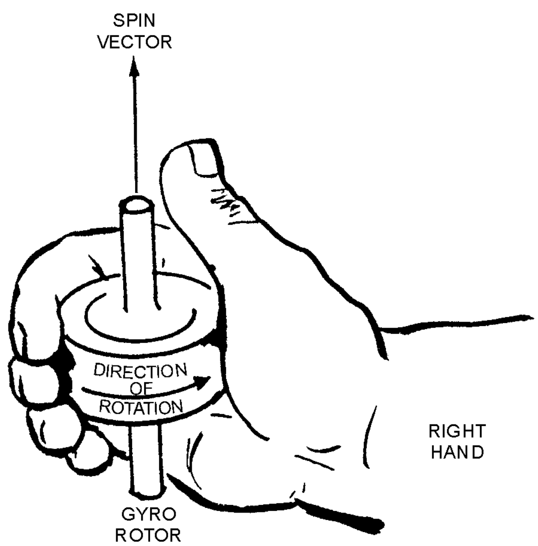


Figure 3-9.—Determining spin vector direction.

All three motions are rotary (angular) and can be represented by vectors that point in such a direction that when you look in the direction of the vector the rotary motion around the vector is clockwise.

Another hand rule convenient for determining the DIRECTION OF PRECESSION uses the fingers of the right hand. This method may not be new to you. A similar method is applied to electric motors (see NEETS, module 5).

You may represent the three vectors listed above by arranging the thumb, forefinger, and middle finger of your right hand mutually perpendicular as shown in figure 3-10. Your thumb points in the direction of the precession vector, your middle finger points in the direction of the torque vector, and your forefinger points in the direction of the spin vector. You can consider these vectors as the axes about which angular motion takes place. If you look in the directions your fingers and thumb point, you can visualize that all the rotary motions are clockwise as indicated in figure 3-10.

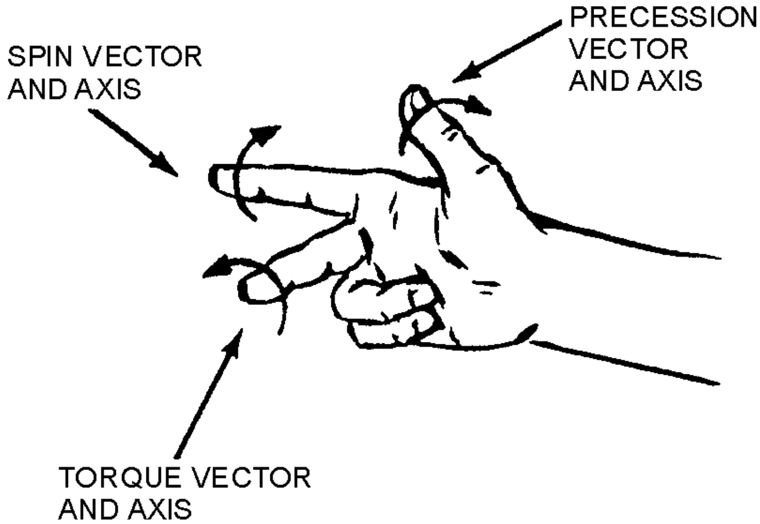


Figure 3-10.—Right-hand rule for determining direction of precession.

This three-finger rule is useful for analyzing any gyroscope motion problem because if the directions of any two of the three vectors are known, the direction of the third vector can be found and the motion around this vector may be determined.

Q-9. What type of force acts ONLY through the center of gravity of a gyro, and does NOT cause precession?

Q-10. The amount of precession that results from a given force is determined by what quantity?

Q-11. What factor determines the direction a gyro will precess in response to a particular force?

Q-12. When using the tight-hand rule to determine precession, which finger indicates the direction of the applied force?

DEGREES OF FREEDOM

A gyro can have different degrees of freedom, depending on the number of gimbals in which it is supported and the way the gimbals are arranged. Do not confuse the term "degrees of freedom" with an angular value such as degrees of a circle. The term, as it applies to gyros, is an indication of the number of axes about which the rotor is free to precess.

A gyro mounted in two gimbals has two degrees of freedom. When two gimbals are used, the gyro is said to be UNIVERSALLY MOUNTED. This arrangement provides two axes about which the gyro can precess. These two axes and the spin axis intersect at the center of gravity of the entire system (excluding the support). Because of this arrangement, the force of gravity does not exert a torque to cause precession. The rotor, inner gimbal, and outer gimbal are balanced about the three principal axes.

TWO DEGREES-OF-FREEDOM GYROS

The two-degrees-of-freedom (free) gyros can be divided into two groups. In the first group, the gyro's spin axis is perpendicular to the surface of the Earth. Thus the gyro's rotor will spin in a horizontal plane. These gyros are used to establish vertical and horizontal planes to be used where stabilized reference planes are needed.

In the second group, the gyro's spin axis is either parallel to the surface of the Earth or at some angle other than perpendicular. The spin axis of the gyro in the gyrocompass, for example, is maintained in a plane parallel to the surface of the Earth It is aligned in a plane of the north-south meridian. Once set, it will continue to point north as long as no disturbing force causes it to precess out of the plane of the meridian.

Effect of Rotation of the Earth

As you have learned, a free gyro maintains its spin axis fixed in space, and not fixed relative to the Earth's surface. To understand this, imagine yourself in a space ship somewhere out in space and looking at the South Pole of the Earth. You see a sphere rotating clockwise, with the South Pole in the center. Maneuver your ship until it is on a direct line with the South Pole and then cut in the automatic controls to keep it in this position. You will now see the Earth make a complete rotation every 24 hours.

You could keep track of that rotation by driving a big post into the Equator as shown in view A of figure 3-11. If this post were upright at 1200, the Earth's rotation would carry it around so it would be pointing to your right at 1800. Likewise, the Earth's rotation would carry the post around so that at 2400 it would be upside down. Then, at 0600 the next day, the post would be pointing to your left. Finally, at 1200 the next day the post would be back in its original position, having been carried, with the Earth, through its complete rotation. Notice that the post has many positions as you observe it—because it is attached to the Earth's surface and does not have rigidity in space.

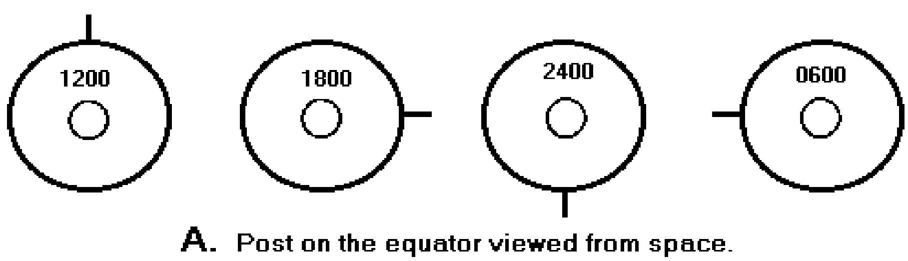


Figure 3-11A.—Fixed direction in space. Post on the equator viewed from space.

If you put a gyroscope in place of the stake, you will see a different action. Imagine a gyroscope mounted at the Equator with its spin axis aligned with the E/W axis of the Earth. The gyro is spinning and has rigidity in space. Now look at view B. At 1200 the spinning axis is horizontal with respect to the Earth's surface. At 1800 the spinning axis is vertical with respect to the Earth's surface; but the gyro is still spinning in the same plane as before, and the black end is pointing away from the Earth's surface. At 2400, the spinning axis is again horizontal. At 0600 the spinning axis is again vertical, and the black end points toward the Earth. Finally, at 1200 the next day, the gyro is in the same position as when it started. The plane of spin of the gyro wheel did not change direction in space while the gyro rotated with the Earth. This is because the gyro is rigid in space.

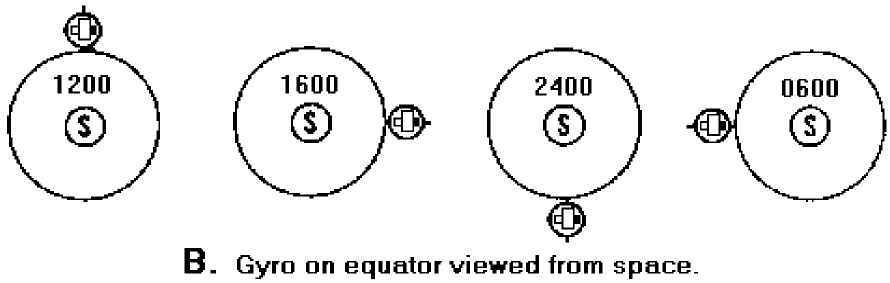


Figure 3-11B.—Fixed direction in space. Gyro on equator viewed from space.

You have just imagined observing the gyro from space. Now, let's come back to Earth and stand right next to the gyro. Look at the gyro in view C. From your viewpoint on Earth, the spinning axis <u>appears</u> to make one complete rotation in one day. As you know, the gyro is rigid, and both you and the Earth are rotating. The effect of the Earth's rotation on a gyro is sometimes called APPARENT DRIFT, APPARENT PRECESSION, or APPARENT ROTATION.

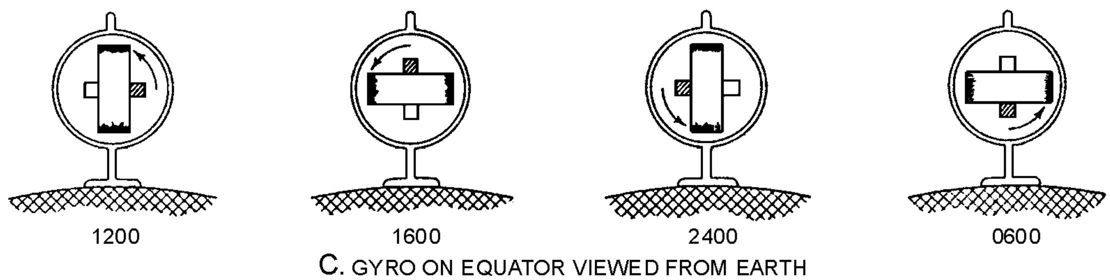


Figure 3-11C.—Fixed direction in space. Gyro on equator viewed from earth.

Effect of Mechanical Drift

A directional error in a gyro is produced by random inaccuracies caused by mechanical drift and the effect of the Earth's rotation (apparent drift).

We shall see later how it is corrected for in the equipment. First, let's consider the causes of mechanical drift.

There are three general sources of mechanical drift:

1. Unbalance. A gyro often becomes dynamically unbalanced when operated at a speed or temperature other than that for which it was designed. The static balance of the gyro is upset when its center of gravity is not at the intersection of the three major axes. Some unbalance of both types will exist in any gyro since manufacturing processes cannot produce a perfectly balanced gyro.

2. Bearing friction. Friction in the gimbal bearings results in loss of energy and incorrect gimbal positions. Friction in the rotor bearings causes mechanical drift only if the friction is not symmetrical. An even amount of friction all around in a rotor bearing results only in a change of the speed of rotation.

3. Inertia of gimbals. Energy is lost whenever a gimbal rotates because of the inertia of the gimbal. The greater the mass of the gimbal, the greater the drift from this source.

The complete elimination of mechanical drift in gyros appears to be an impossibility. However, by proper design it has been kept to a minimum. Any error that still exists can be corrected for.

Q-13. A universally mounted gyro has how many degrees of freedom?

Q-14. If a free gyro is placed at the equator at 1200 in a vertical position; in what position should it be at 1800?

Q-15. What are the three causes of mechanical drift in a gyro?

Establishing and Maintaining a Fixed Position

You now know that a free gyro maintains a fixed position in space. Because of this property, a free gyro can be used to establish a stable, unchanging reference, in any plane (horizontal, vertical, or any specific position in between). The gyro-erecting system has the function of positioning the gyro to the desired position and helping to keep it there.

Any gyro-erecting system must meet the following requirements:

1. The system must provide torques (forces) of sufficient magnitude and direction to precess the gyro so that its spin axis is brought to the desired position after the rotor is spinning at its operating speed.

2. The system must provide torques to precess the gyro back to the required position at the proper rate and direction to cancel the effects of apparent and mechanical drift.

Erection may be done mechanically or electrically, depending on the type of power available. Specific erection systems are many and varied. We will briefly discuss only two of them.

Mercury Erecting System

One of the common erection systems used for vertical gyros uses mercury as the element for sensing gyro position with respect to vertical. Mercury also provides the force to precess the gyro toward vertical when the gyro drifts.

This system consists of two tanks of mercury fastened to opposite sides of the gyro case and connected by a small mercury tube as shown in figure 3-12. A small air tube is also connected between the tanks to prevent a vacuum from forming. If the spin axis tilts away from the vertical, as shown, the mercury will flow from one tank to the other. The added weight in the left tank provides a torque which causes the gyro to precess. At this point, if you were to apply the rule for precession, you would see that the precession would be 90º away from the desired direction.

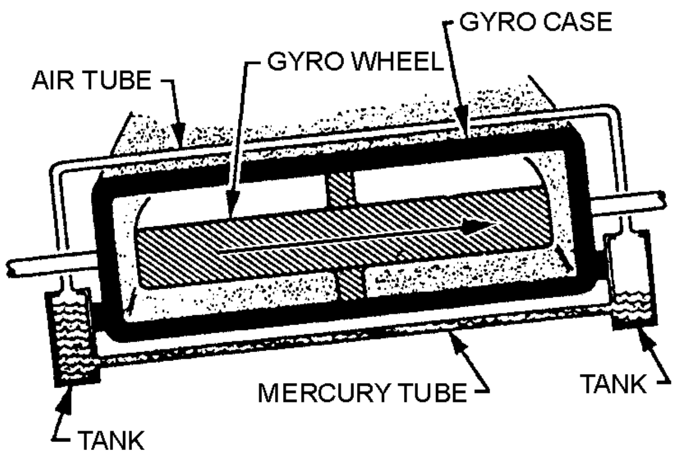


Figure 3-12.—Mercury erecting system.

To overcome this difficulty, the point where the torque is applied must be moved. The torque point is moved by causing the gimbal assembly to slowly and continuously rotate in the proper direction. This is done in the following manner. With a small mercury tube, the mercury will take nearly a second to find its level. At the same time the mercury is flowing, a small motor is rotating the gimbal supporting the gyro about 18 times a minute. Therefore, during the time that it takes the mercury to flow into the low tank, the entire gimbal assembly has rotated 90º. The torque will now be applied at a point which causes the gyro to precess in the proper direction to maintain the gyro spin axis in a vertical position.

Mercury Ballistic Erecting System

The erection system used in many horizontal gyros is very similar to the vertical gyro system just discussed. It is called the *mercury ballistic erection system*. The mercury ballistic system has the added feature of maintaining the spin axis not only in the horizontal plane, but also with the spin axis aligned North-South.

There are many different methods of causing free gyros to precess to either the vertical or the horizontal plane. All such systems use the forces of gravity to sense variation from the desired position: all systems also use the principles of precession to position the gyro property.

Q-16. What is the purpose of an erection system used with a gyro?

Q-17. What is the purpose of rotating the gimbal assembly in a gyro using a mercury erection system?

RATE GYROS

RATE GYROS are used in weapons control equipment, aircraft instrumentation, inertial navigation, and in many other applications to detect and measure angular rates of change.

A rate gyro (sometimes called a rate-of-turn gyro) consists of a spinning rotor mounted in a single gimbal, as shown in figure 3-13. A gyro mounted in this manner has one degree of freedom; that is, it is free to tilt in only one direction. The rotor in a rate gyro is restrained from precessing by some means, usually a spring arrangement. This is done to limit precession and to return the rotor to a neutral position when there is no angular change taking place. Remember, the amount of precession of a gyro is proportional to the force that causes the precession.

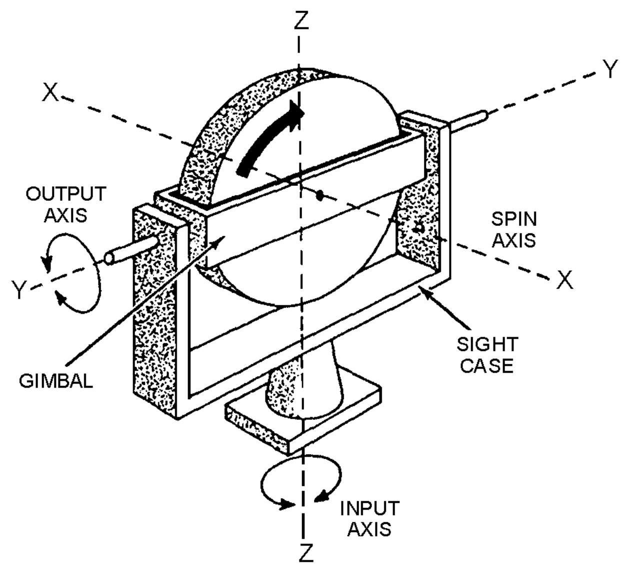


Figure 3-13.—Rate gyro (single degree of freedom).

If you attempt to change the gyro's plane of rotor spin by rotating the case about the input axis, the gyro will precess as shown in figure 3-14. From what you learned earlier in this chapter, the gyro does not appear to be obeying the rules for precession. However, turning the gyro case has the same effect as applying a torque on the spin axis. This is illustrated by arrow F in figure 3-14. You can determine the direction of precession by using the right-hand rule we discussed earlier.

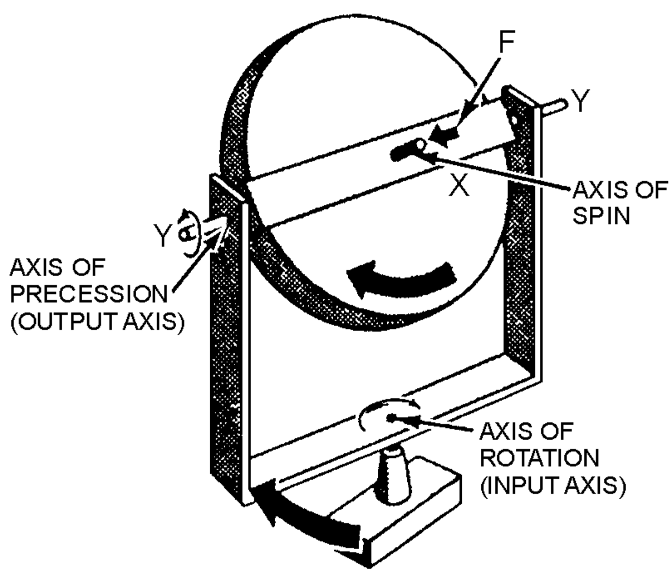


Figure 3-14.—Rate gyro precession.

The force applied at F will cause the gyro to precess at right angles to the force. Likewise, attempting to turn the gyro case will cause the same result. The gyro will precess, as shown by the arrows, around the Y-Y axis (output axis).

Since the rate of precession is proportional to the applied force, you can increase the precession by increasing the speed with which you are moving the gyro case. In other words, you have a rate gyro. The faster you turn the case, the more the gyro will precess, since the amount of precession is proportional to the rate at which you are turning the gyro case.

This characteristic of a gyro, when properly used, fits the requirements needed to sense the rate of motion about any axis.

Figure 3-15 shows a method of restraining the precession of a gyro to permit the calculation of an angle. Springs have been attached to the crossarm of the output shaft. These springs restrain the free precession of the gyro. The gyro may use other types of restraint, but no matter what type of restraint is used, the gyro is harnessed to produce some useful work.

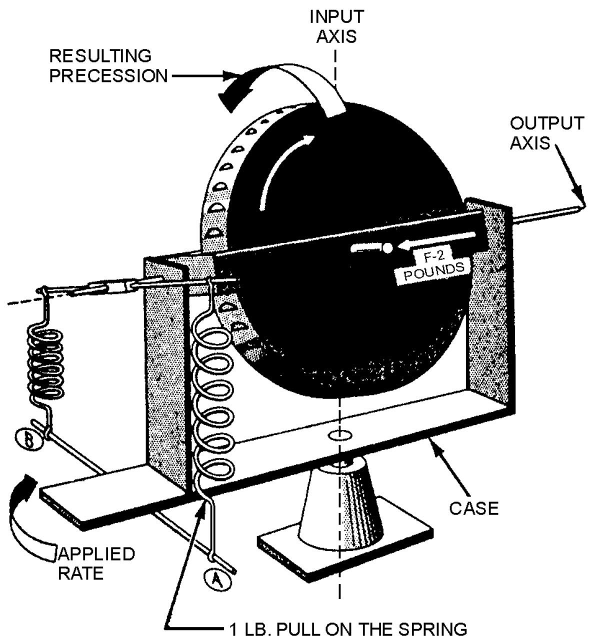


Figure 3-15.—Precession of a spring restrained rate gyro.

As the gyro precesses, it exerts a precessional force against the springs that is proportional to the momentum of the spinning wheel and the applied force. For example, suppose you rotate the gyro case (fig. 3-15) at a speed corresponding to a horizontal force of 2 pounds at F. Obviously, the gyro will precess; and as it does, it will cause the crossarm to pull up on spring A with a certain force, say 1 pound. (This amount of force would vary with the length of the crossarm.)

If you continue to turn the gyro case at this rate, the precession of the gyro will continually exert a pull on the spring. More precisely, the gyro will precess until the 1 pound pull of the crossarm is exactly counterbalanced by the tension of the spring; it will remain in a fixed position, as shown in figure 3-15. That is, it will remain in the precessed position as long as you continue to rotate the gyro case at the same, constant speed. A pointer attached to the output axis could be used with a calibrated scale to measure precise angular rates.

When you stop moving the case, you remove the force at F, and the gyro stops precessing. The spring is still exerting a pull, however, so it pulls the crossarm back to the neutral position and returns the pointer to "zero."

Suppose you now rotate the gyro case at a speed twice as fast as before, and in the same direction. This will be equal to a 4-pound force applied at F and a resulting 2-pound pull by the crossarm on spring

A. In this situation the gyro will precess twice as far before the tension on the restraining spring equals the pull on the crossarm. Precession increases when the rate of rotation increases, as shown in figure 3-16.

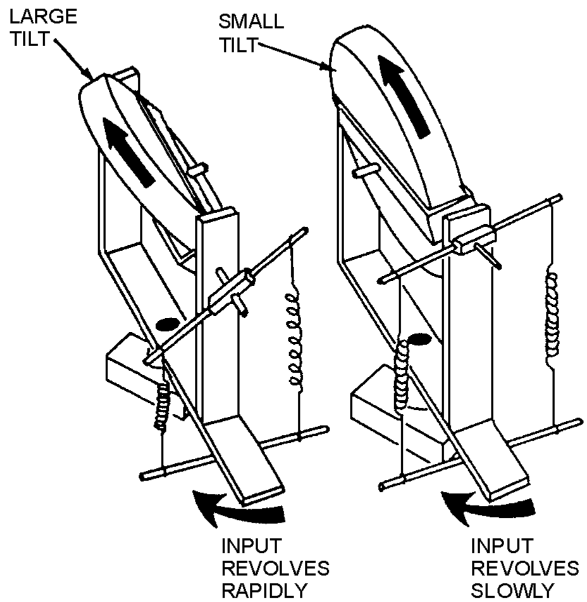


Figure 3-16.—Precession is proportional to the rate of rotation.

Another type of rate gyro (often used in inertial navigation equipment) is the floated gyro unit. This unit generally uses a restraint known as a torsion bar. The advantage of the torsion bar over the spring is that the torsion bar needs no lever arm to exert torque. The torsion bar is mounted along the output axis (fig. 3-17), and produces restraining torque in either direction by twisting instead of pulling. Also, there is no gimbal bearing friction to cause interference with gyro operation.

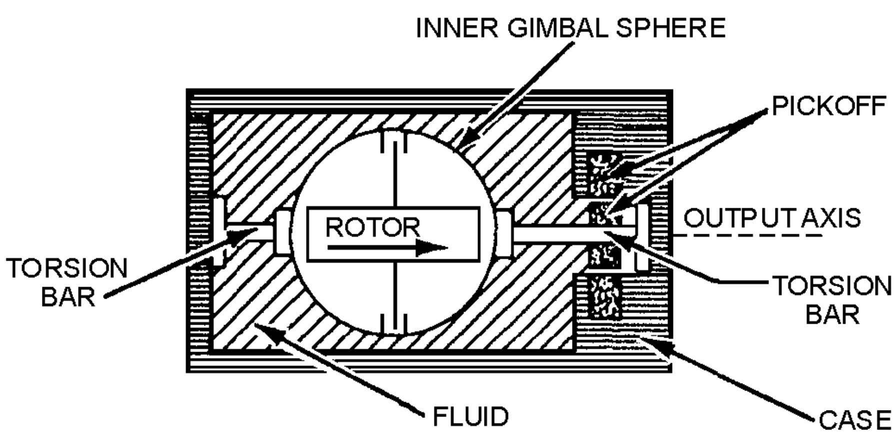


Figure 3-17.—Torsion bar-restrained floated rate gyro.

A fluid surrounds the gyro sphere and provides flotation. It also provides protection from shock, and damps the oscillations resulting from sudden changes in the angular rate input. In this gyro, the inner gimbal displacement must be measured with some type of electrical pickoff. As the gyro case is rotated about the input axis, clockwise or counterclockwise, a precession torque will be developed about the output axis that will cause the inner gimbal to exert torque against the torsion bars. The torsion bars provide a restraining torque proportional to the amount of the inner gimbal's displacement. When the exerted gimbal torque is exactly opposed by the restraining torque provided by the torsion bars, the inner gimbal displacement will be proportional to the rate of rotation of the gyro case about the input axis. The pickoff measures this displacement and provides a signal whose amplitude and polarity (or phase) represent the direction and magnitude of the input angular velocity.

The important point to remember is that every "rate" gyro measures the RATE OF ROTATION ABOUT ITS INPUT AXIS.

Up to this point, we have illustrated only basic gyros. We used these basic, or simple, gyros to explain their principles of operation In actuality, the rate gyros used in typical modern day weapon systems are considerably more complex, and in some cases, very compact. Figure 3-18 shows a cutaway view of a rate gyro used in our Navy's missile systems and aircraft.

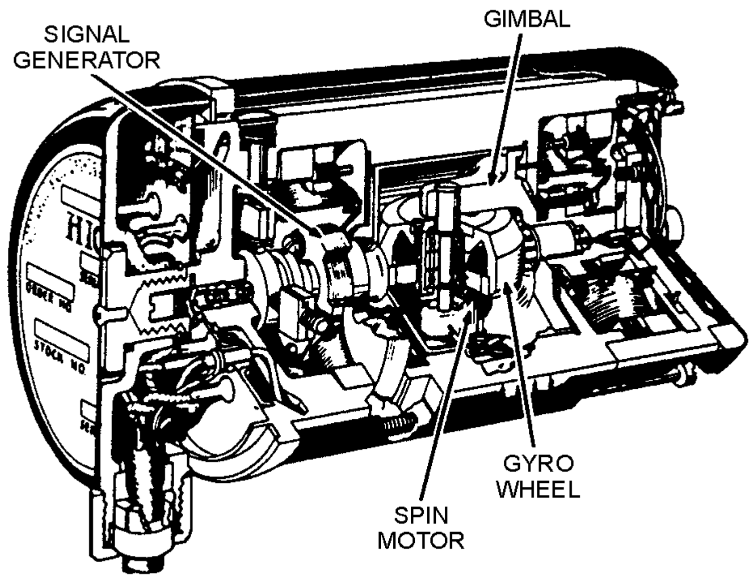


Figure 3-18.—Rate gyro, cutaway view.

Q-18. What are rate gyros primarily used for?

Q-19. How many degrees-of-freedom does a rate gyro usually have?

Q-20. What gyro characteristic provides the basis of the operation of a rate gyro?

ACCELEROMETERS

An accelerometer is a device that gives an indication, usually in the form of a voltage, that is proportional to the acceleration to which it is subjected. The operation of an accelerometer is based on the property of INERTIA (Newton's First Law of Motion). A simple demonstration of inertia happens to us almost every day. You know that if your automobile is subjected to acceleration in a forward direction, you are forced back in the seat. If your auto comes to a sudden stop, you are drawn forward. When your auto goes into a turn, you tend to be forced away from the direction of the turn-that is, if your auto turns left, you are forced to the right, and vice versa.

If we replace the human in an auto with a mass suspended in an elastic mounting system, as shown in figure 3-19, any acceleration of the auto will cause movement of the mass relative to the auto. The amount of displacement is proportional to the force causing the acceleration. The direction the mass moves is always opposite to the direction of the auto's acceleration.

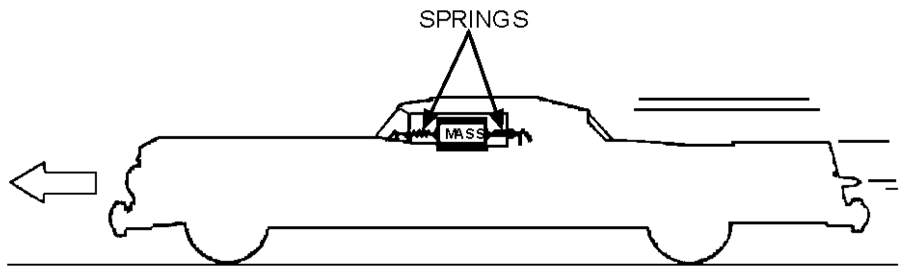


Figure 3-19.—Auto with spring-suspended now.

The mass moves according to Newton's second Law of Motion which states: when a body is acted on by force, its resulting acceleration is directly proportional to the force and inversely proportional to the mass of the body.

When no acceleration is present, the mass will be at rest. When acceleration is present, the mass will lag in proportion to the acceleration force. In other words, the car moves but the mass wants to remain at rest.

Accelerometers are used principally in inertial navigation systems. They are used in aircraft and missile navigation systems as well as aboard ship. Some common types of accelerometers are described briefly in the following paragraphs.

THE BASIC ACCELEROMETER

Figure 3-20 is a simplified drawing of a basic accelerometer. It consists of a mass that is free to slide along the sensitive axis within the case. The movement of the mass is limited by the springs. When the case is accelerated, the mass, because of its inertia, tends to remain stationary. This results in a relative movement of the mass with respect to the case. When the stretch of the springs overcomes the inertia of the mass, the springs cause the mass to stop moving with respect to the case. The displacement of the mass with respect to the case is directly proportional to the acceleration of the case. When the case stops accelerating, the springs return the mass to its zero position (the reference position). To keep the springs from causing the mass to overshoot and oscillate about the reference position, some form of damping is needed. This is usually provided by an oil-filled case with vanes for oil to bypass the mass.

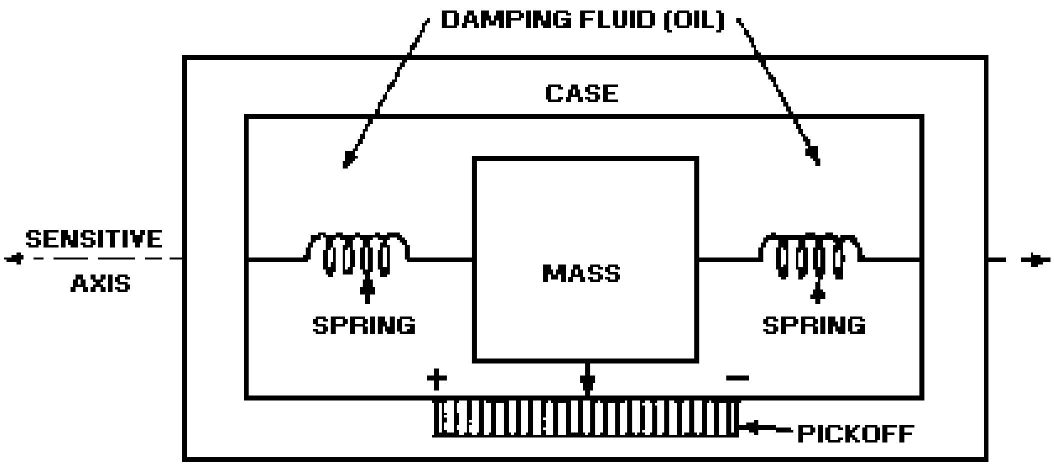


Figure 3-20.—The basic accelerometer.

An accelerometer is sensitive to gravity when its sensing axis is positioned so gravity can move or attempt to move the mass. This is useful in that we can use gravity as a reference for testing purposes, but it can be a serious problem because of the errors it may cause in acceleration measurements. If the unit is placed with the sensing axis vertical, the mass will be displaced such that the output is one "G," or one unit of gravity. This is done during testing. Then when the sensitive axis is turned so it is horizontal to the Earth, the springs center the mass, and the output of the unit is zero.

E-TRANSFORMER ACCELEROMETER

The E-transformer accelerometer (fig. 3-21) consists of a mass suspended from a calibrated leaf spring in a manner similar to a pendulum. The mass is effectively the armature of an E-transformer of the type used as an error detector in a servo system. The mass of the accelerometer is enclosed within a case that is filled with a damping liquid, which helps keep the pendulum from oscillating. The accelerometer is mounted so that acceleration in only the desired geometrical plane is detected.

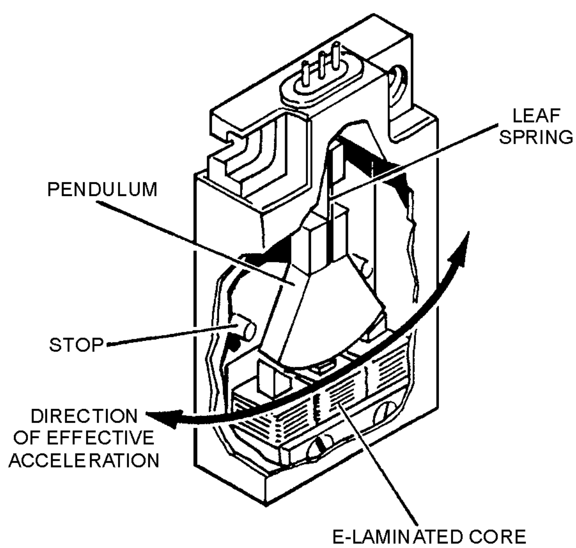


Figure 3-21.—E-transformer accelerometer.

Refer to figure 3-21. Notice that when there is no acceleration, the pendulum remains centered and the accelerometer output is zero. However, when there is acceleration the mass or pendulum swings in the direction opposite to that of the acceleration, causing an output from the E-transformer. Since the amplitude of the pendulum's swing is proportional to the amplitude of the acceleration to which it is subjected, the output of the device indicates both the direction and amplitude of the acceleration. This output is within the limits of the equipment and is limited by physical stops.

PULSE-COUNTING ACCELEROMETER

The outputs of the accelerometers discussed so far are voltages, which are proportional to acceleration. These voltages are assigned scale factors (such as units per volt). The voltage represents the quantity. In many applications there is need for accelerometer output signals to be in digital form, which means that the signal consists of a series of pulses that indicates an actual number. Pulse counting accelerometers satisfy this need. Their pulse output can be supplied directly to computer circuits and other digital logic equipment. A schematic and a pictorial diagram of a pulse counting accelerometer is shown in figure 3-22, view (A) and view (B).

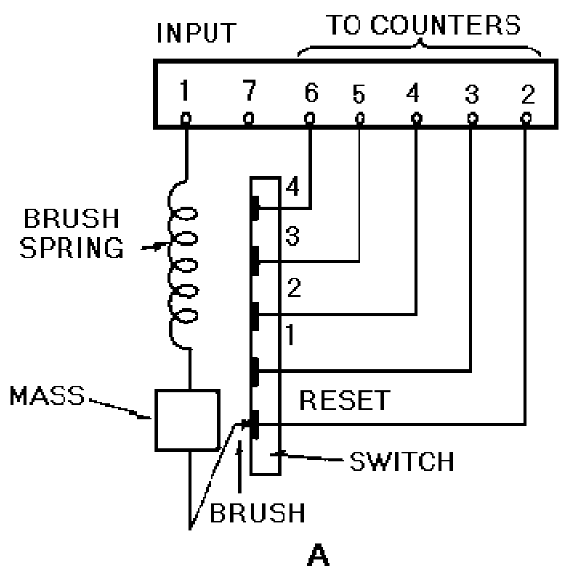


Figure 3-22A.—Pulse counting accelerometer.

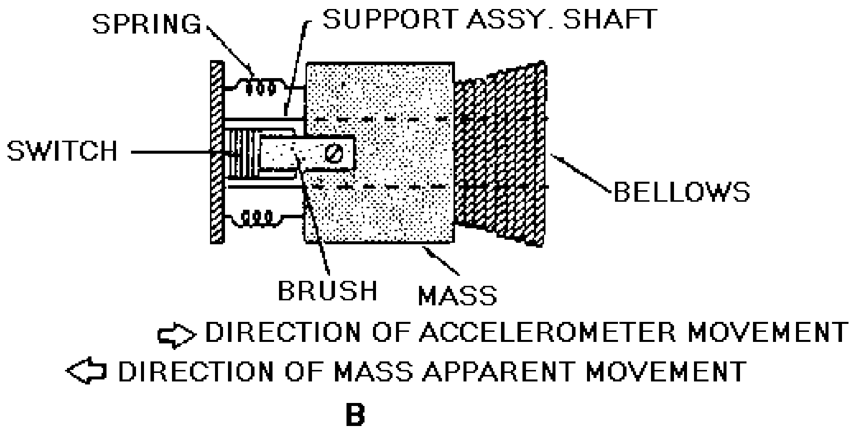


Figure 3-22B.—Pulse counting accelerometer.

When velocity remains constant, the brush spring holds the mass at null, and the brush rests on the reset contact of the switch. As acceleration occurs, the tendency of the mass to remain at a constant velocity causes the spring to compress. As the spring compresses, it allows the brush to move off the reset contact. If the acceleration is great enough, the brush will pass over the switch contacts for acceleration levels 1, 2, 3, and 4. These levels are determined by the stiffness of the spring.

As the brush passes over each contact (in a positive direction), an output pulse from each contact is coupled to one of four counters. This advances the counter one-half count. The accelerometer is designed so that as acceleration decreases, the mass tends to assume the new velocity. The counters will not advance the remaining half count until the brush once again touches the reset contact. With this type of pulse output, it is possible to record each time gravity forces have reached a predetermined level.

Q-21. Operation of an accelerometer is based on what physical property?

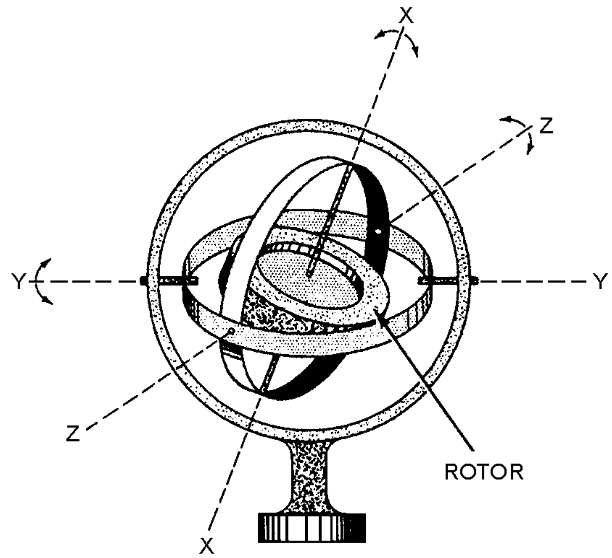
Q-22. What type of systems primarily use accelerometers?

Q-23. What special requirement is the pulse counting accelerometer designed for?

SUMMARY

In this chapter you were introduced to the subject of gyros. You studied the characteristics of gyros and several applications of their use. You were also briefly introduced to the subject of accelerometers. The following information provides a summary of the chapter.

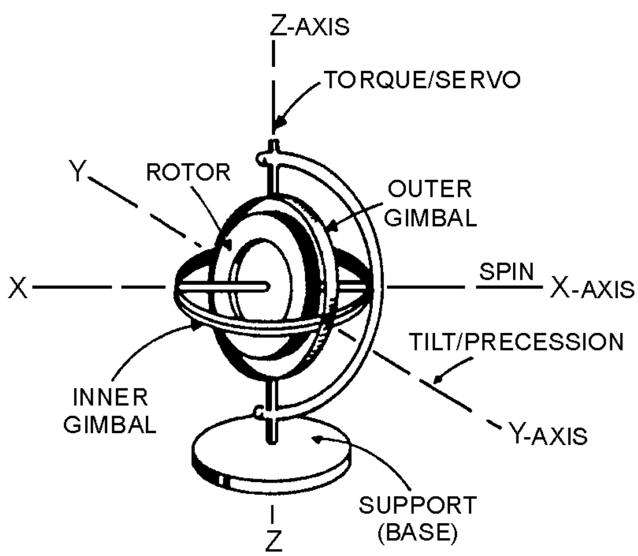
The term **GYROSCOPE** (or GYRO) may be applied to any rapidly spinning object; however, a functional gyroscope is constructed and mounted to take advantage of certain characteristics.



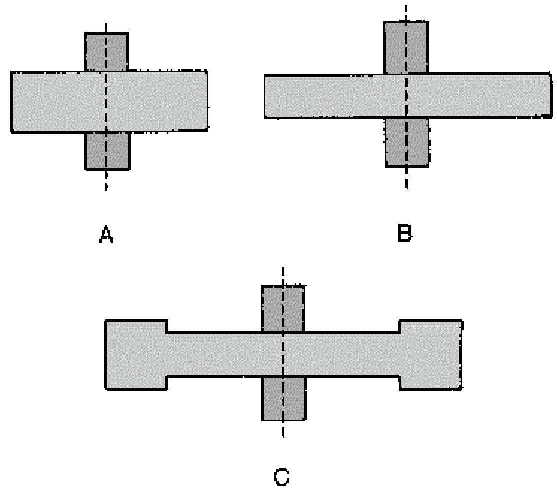
RIGIDITY in a gyro is the tendency of a spinning wheel (rotor) to remain in a fixed position in space.

PRECESSION is the property of a gyro that causes it to tilt in a direction <u>perpendicular</u> to the direction of any outside force.

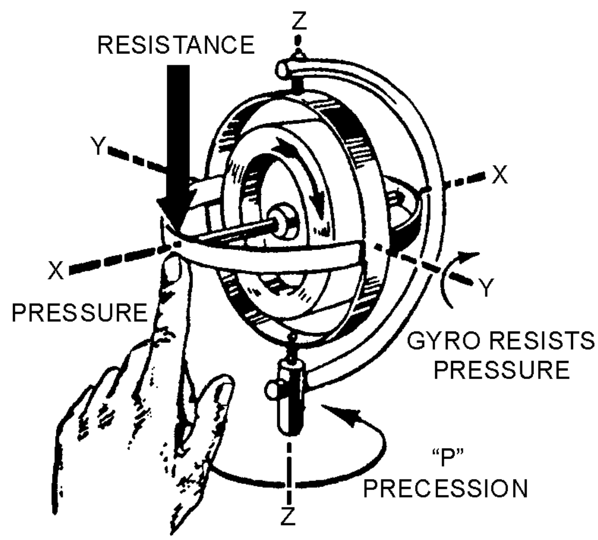
The **COMPONENTS OF A UNIVERSALLY MOUNTED GYRO** are: the Rotor, the Inner Gimbal, the Outer Gimbal, and the Base.



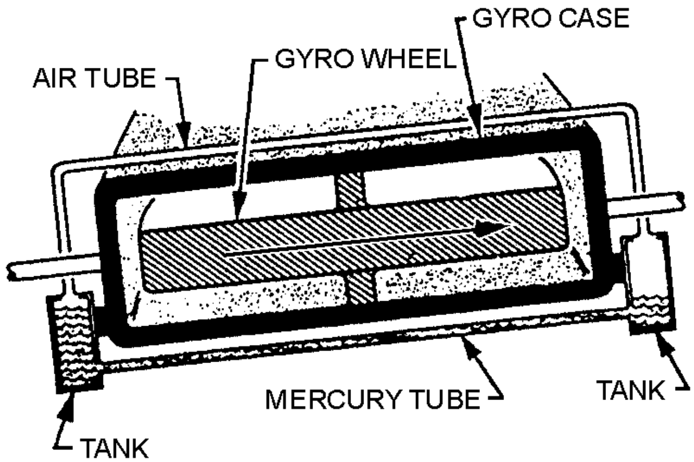
The **FACTORS THAT EFFECT RIGIDITY** are the weight, shape, and speed of rotation of the rotor.



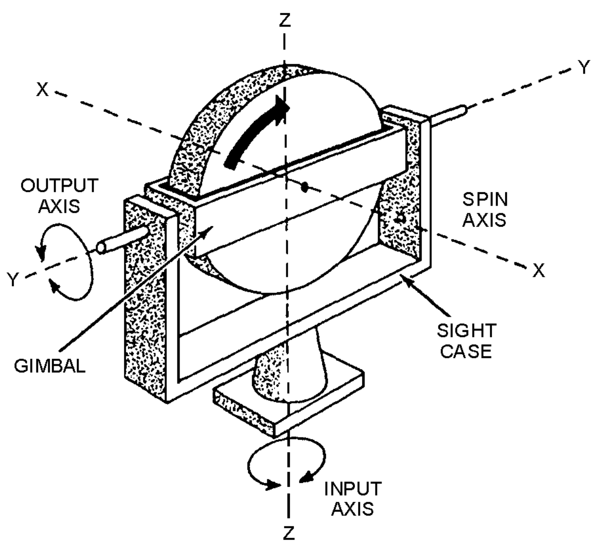
The **DIRECTION OF PRECESSION** in a gyro is always 90º from the direction of the applied force.



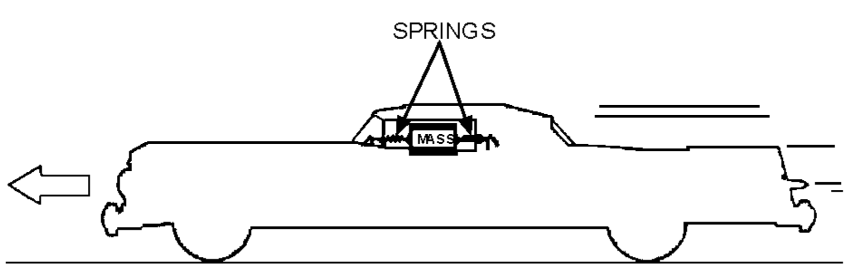
A **GYRO ERECTION SYSTEM** must be capable of sensing the difference between the present rotor position and the <u>desired</u> rotor position, and must apply forces that will cause the gyro to precess toward the desired position. The mercury erection system discussed in this chapter is only one of several possible types of gyro erection systems.



RATE GYROS are specially mounted so they are free to precess in only one direction; they are used to measure angular rates.



An **ACCELEROMETER** is a device used to detect and measure any acceleration along a particular axis.



ANSWERS TO QUESTIONS Q1. THROUGH Q23.

A-1. Yes.

A-2. X-axis.

A-3. Rigidity.

A-4. Any force that attempts to tilt the spin axis.

A-5. Perpendicular (90º) to the force.

A-6. Two.

A-7. Rotor speed, weight, shape.

A-8. C.

A-9. Force of translation.

A-10. Rigidity.

A-11. Direction of spin.

A-12. Middle finger.

A-13. Two.

A-14. Horizontal.

A-15. Unbalanced gyro, inertia of gimbals, bearing friction.

A-16. To achieve and maintain the proper operating position for the gyro (usually vertical or horizontal).

A-17. Applies torque at the proper point for correct precession.

A-18. Measuring angular rates.

A-19. One.

A-20. Precession.

A-21. Inertia.

A-22. Navigation systems.

A-23. When digital data is required from an accelerometer.

CHAPTER 4

RELATED DEVICES

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to:

1. Compare standard synchro system connections with IC synchro connections.

2. Describe the operation of a step transmitter and receiver, and list the advantages and disadvantages of a step-transmission system.

3. Compare the construction and operation of a resolver to those of a transformer, describe the solution of resolution and composition problems by a resolver.

Some other devices that logically should be included in this module are the IC synchros, step motors, and resolvers. These are all electromagnetic devices used in various shipboard and aircraft applications. They can be considered as second cousins of the synchro.

IC SYNCHROS

The engine order telegraph, steering telegraph, rudder-angle indicator, and similar position-indicating systems used on naval ships are usually simple synchro systems. Some ships, however, use IC synchros to transfer such information. These units operate on the same general principle as the synchros we discussed in chapter 1.

The interior communication synchro (IC synchro) is gradually being phased out and replaced by standard synchros when replacement is required. However, you will still find some IC synchros in use today. For that reason, you will find some background information on their purpose and theory to be beneficial. We will present these synchros in very basic form in the following paragraphs.

Because of their construction, IC synchros are sometimes called <u>reversed synchros</u>. The primary winding, consisting of two series-connected coils, is mounted physically on the stator. The secondary, consisting of three Y-connected coils, is mounted physically on the rotor. This is shown schematically in figure 4-1.

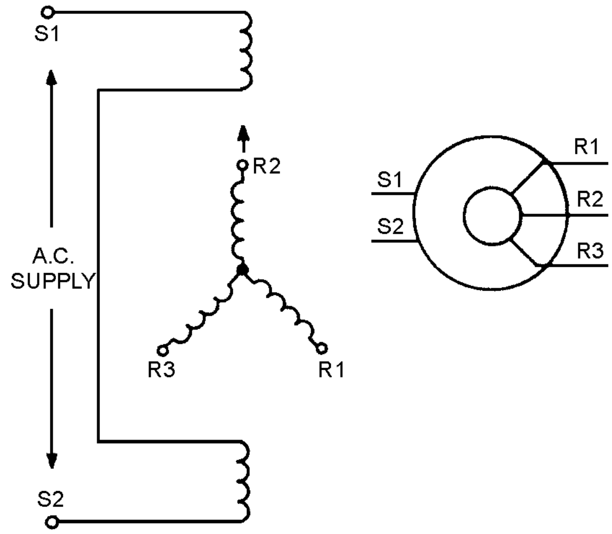


Figure 4-1.—IC synchro schematic diagrams.

IC synchros operate on the same principles of interacting magnetic fields as other synchros, but differ in direction of shaft rotation and amount of torque obtainable. When an IC transmitter and IC receiver are connected in parallel as shown in view A of figure 4-2, the shaft of the IC receiver follows the rotation of the IC transmitter shaft. In view B, the IC transmitter is replaced by a synchro transmitter; the IC receiver shaft now turns in a direction opposite to that of the synchro transmitter. Voltages that cause counterclockwise rotation of a standard synchro shaft cause clockwise rotation of an IC synchro shaft. When it is desirable to have the IC synchro receiver turn in an opposite direction from that of the transmitter, the connections are as shown in view C. For a standard synchro receiver to follow the rotation of an IC transmitter, their connections must be made as shown in view D.

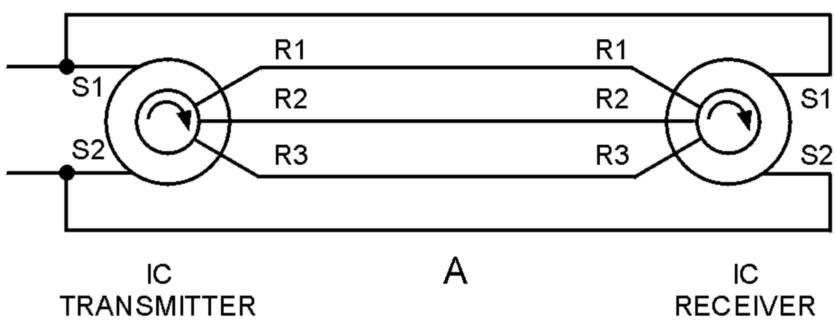


Figure 4-2A.—IC versus standard synchro connections.

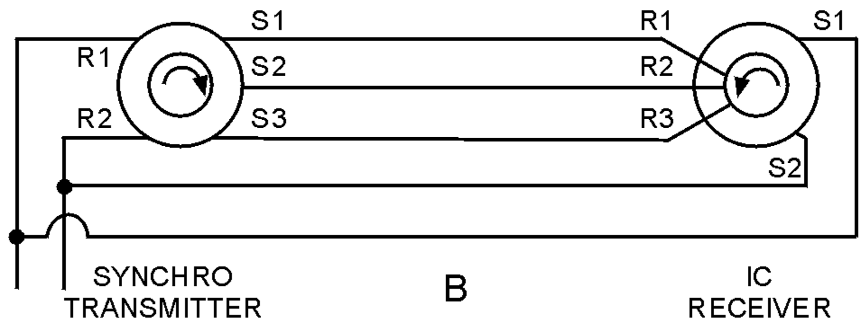


Figure 4-2B.—IC versus standard synchro connections.

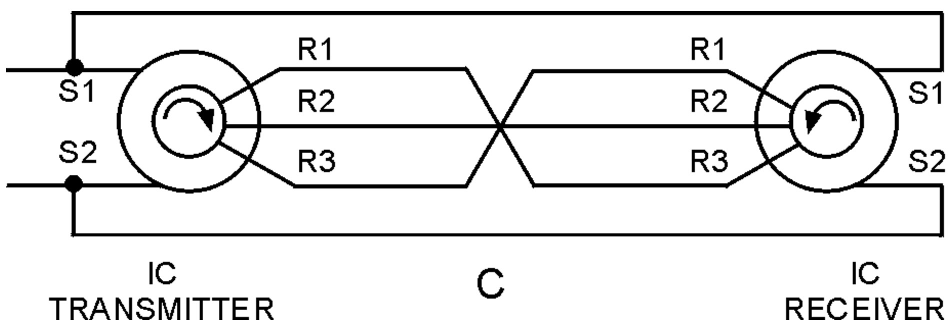


Figure 4-2C.—IC versus standard synchro connections.

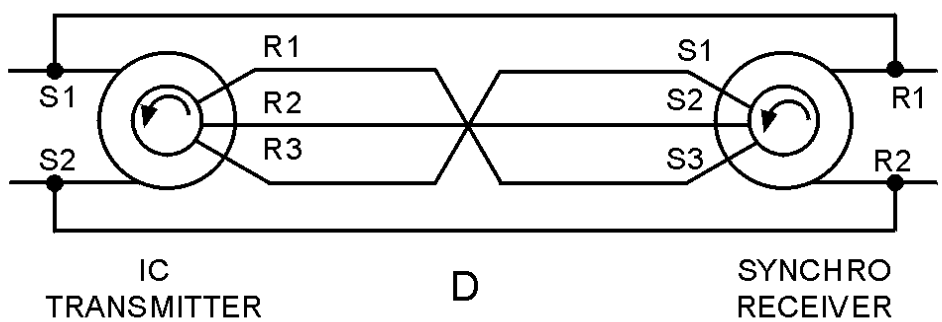


Figure 4-2D.—IC versus standard synchro connections.

The torque obtainable from either an IC synchro or a standard synchro is determined by the magnetizing power, which is limited by the allowable temperature rise. When the stator is energized, as in IC synchros, the magnetizing power can be increased with a resulting larger torque. The reason for this is that the losses are dissipated in the form of heat around the outer shell of the IC transmitter or receiver. In standard synchros, this heat loss is dissipated through the rotor, the air gap, and then the outer shell to the surrounding air.

The electrical zero position for an IC synchro is the position where rotor coil R2 is aligned with the stator as shown in figure 4-1. To zero an IC synchro, apply the same general theory as we described for other synchros.

For further information on IC synchro replacement, alignment, and theory, refer to *Military Handbook, Synchros, Description and Operation*, (MIL-HDBK-225A).

Q-1. What two characteristics of IC synchros cause them to differ from standard synchros?

STEP-TRANSMISSION SYSTEMS

All of the synchro units we have discussed operate on alternating current. In applications where alternating current is not available, there is a need for a system that can use direct current. The STEP-TRANSMISSION SYSTEM (sometimes referred to as the step-by-step system) is such a system.

A geared step-transmission system is often used to drive compass repeaters on naval vessels and merchant ships having dc power.

Although many variations are used, the simplified step-by-step transmission system shown in figure 4-3 is typical. In this system, a step transmitter and a step motor are used together to transmit angular data (information) between remote locations. When rotated, the shaft of the step transmitter periodically switches a dc excitation voltage from one pair of coils to another in the step motor. The step motor, which is the receiver in the system, responds to this varying excitation by rotating an amount that is proportional to the transmitter's shaft position.

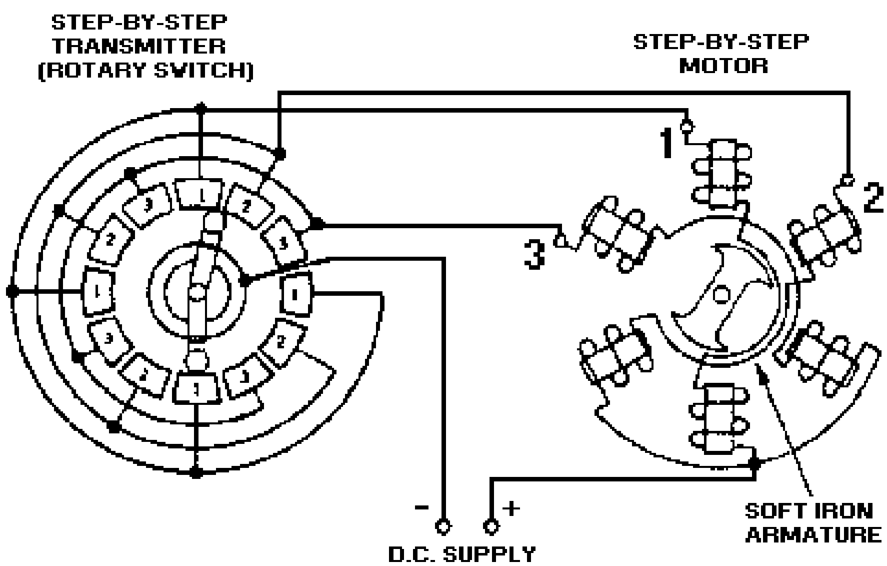


Figure 4-3.—Step-by-step transmission system.

The stator of this step motor has six field coils spaced 60 degrees apart. The coils are connected in three groups of two coils each, with opposite coils connected in series. One end of each pair of coils is connected to one of the brushes in the step transmitter. The other ends of the coils are connected to the other side of the dc supply voltage through a common lead. As the rotor of the step transmitter is turned, the corresponding coils in the step motor are energized in sequence, producing a rotating step-by-step stator field. Thus, the motor rotates in abrupt increments or steps rather than smoothly.

NOTE: The theory of the step-by-step motor is similar to that involved in positioning the bar magnet (chapter 1, fig. 1-10).

For ease of explanation, we will replace the step transmitter with a battery as we discuss the operation of the step motor. If, as shown in view A of figure 4-4, we apply the dc battery voltage across the number 1 coils only, the armature will turn to the position shown. When we apply the voltage also to the number 2 coils, the armature turns to a position midway between the number 1 and number 2 coils, view B. If we now disconnect the number 1 coils, the armature turns until it lines up with the number 2 coils, view C. View D shows the number 2 and 3 coils connected and the armature rotated one step further. As long as this process is continued, the armature can be rotated through 360 degrees.

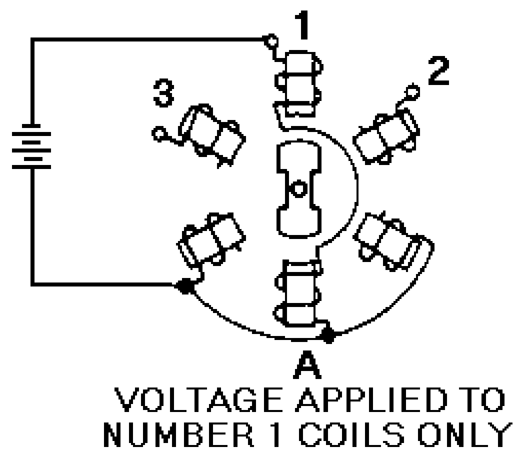


Figure 4-4A.—Step-by-step motor in various positions. VOLTAGE APPLIED TO NUMBER 1 COILS ONLY.

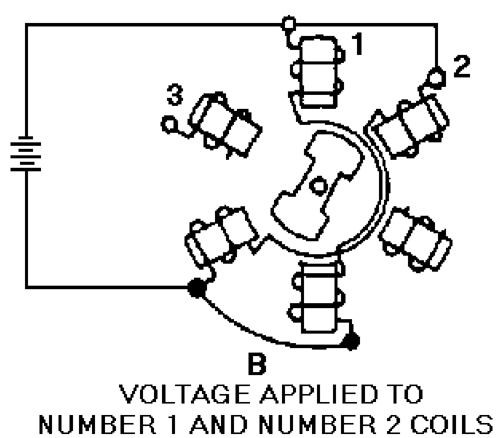


Figure 4-4B.—Step-by-step motor in various positions. VOLTAGE APPLIED TO NUMBER 1 AND NUMBER 2 COILS.

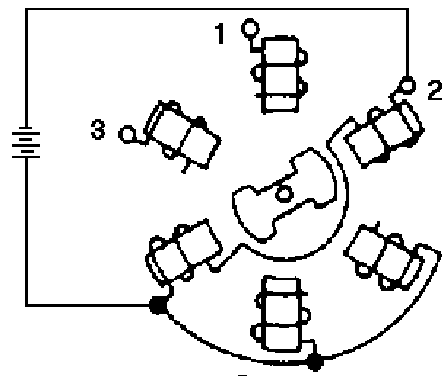


Figure 4-4C.—Step-by-step motor in various positions. VOLTAGE APPLIED TO NUMBER 2 COILS ONLY.

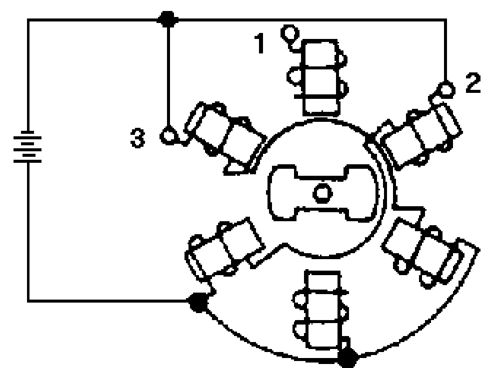


Figure 4-4D.—Step-by-step motor in various positions. VOLTAGE APPLIED TO NUMBER 2 AND NUMBER 3 COILS.

In actual operation, the step-by-step motor is driven by a step transmitter (rotary switch) as shown in figure 4-3. As the switch is rotated, it applies voltage first to coil 1, and then in sequence to 1 and 2 together, coil 2 only, coils 2 and 3 together, coil 3 only, and so on, until the complete revolution is made. As a result, the armature turns in 30-degree steps that follow the rotation of the rotary switch. The rotating arm of the switch can be turned mechanically to angles between zero and 360º in 30-degree steps. The actual angle through which the arm is rotated depends on the specific data to be transmitted by the system. Real systems may transmit data in greater or lesser steps than 30 degrees, depending on system design.

An important point in this type of system is that because the armature is soft iron, either end of the armature may turn and line up with the energized coils. For this reason, a hand reset control is provided on the step-by-step motor. This permits an operator to align the receiver with the transmitter each time the power supply is energized.

A step-transmission system is not self-synchronizing, and its limited number of steps does not permit data to be transmitted smoothly or where relatively small changes in data are required. This type of transmission system is cheap, rugged, and relatively powerful.

Q-2. Compare the power sources of synchros and step-transmission systems.

Q-3. A step transmitter is a modification of what electrical device?

RESOLVERS

The last device we will discuss is the resolver. Physically, resolvers are similar to synchros, and are used to perform mathematical computations electrically. They are rotary electromechanical devices that provide outputs that are trigonometric functions of their inputs. As you may know, the branch of mathematics that deals with the quantities and angles of a right triangle is known as trigonometry. Many "trig" problems that can be solved with paper and pencil can be solved by applying the proper electrical or mechanical quantities to a resolver. The resolver has the advantage of giving instantaneous solutions if the input quantities are changing continuously.

Resolvers are classified according to size (diameter) in the same manner as standard synchros and may be mounted with most standard synchro mounting hardware. A cutaway view of a resolver is shown in figure 4-5.

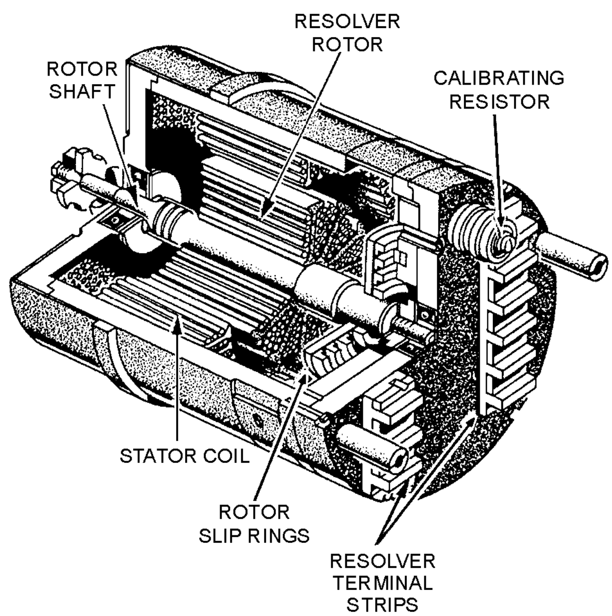


Figure 4-5.—Cutaway view of a resolver.

Notice that the stator of the resolver is a cylindrical structure of slotted laminations on which two coils are wound. The rotor is composed of a shaft, laminations, two windings, and four slip rings. Compensator components, which improve the angular accuracy of resolvers, may consist of resistors or additional windings in the stator and rotor winding circuits. Compensator windings, which increase the accuracy of the resolver, are located inside the stator. Compensating (calibrating) resistors, which compensate for voltage inaccuracies and phase shifts, may be mounted either inside or outside the resolver housing.

A cylindrical frame with a standardized mounting flange houses the assembled resolver. External and internal connections can be made to an insulated terminal board on the rear of the housing. Miniature resolvers often have lead wires brought out through the rear of the resolver, eliminating the need for a terminal block. A reference line is scribed on the face of the housing for alignment with a similar line on the end of the rotor shaft. These are used in determining coarse electrical zero.

Basically, a resolver is a transformer in which the secondary windings can be rotated with respect to the primary windings. Consequently, the amount of magnetic coupling between the primary and the secondary is variable. In the most common form, a resolver consists of a stator and a rotor, each having two separate windings placed precisely at right angles to each other as shown in figure 4-6.

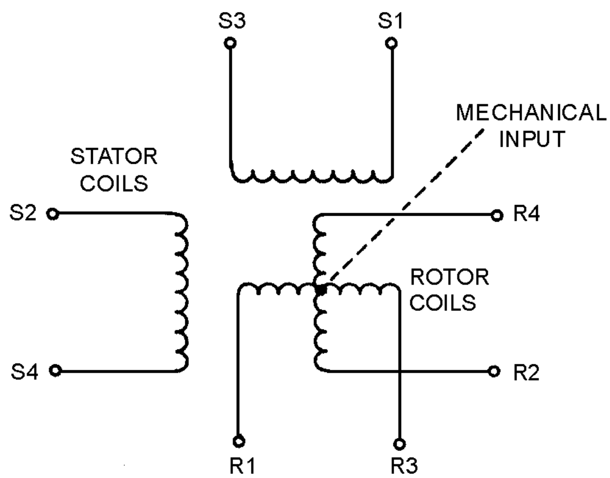


Figure 4-6.—Resolver schematic.

Since the two stator windings are physically and electrically at right angles to each other, there is no magnetic coupling between them. The stator windings are mounted on the resolver housing and are stationary with respect to it.

Similarly, the rotor windings of the resolver are wound at right angles to each other. Hence, there is no magnetic coupling between the two windings. The rotor windings are mounted on the rotor shaft and turn with it. The rotor is capable of unlimited rotation. Thus, the rotor windings can be set at any angle with respect to the stator windings.

Because of the 90º physical and electrical relationships, the resolver has the ability to separate a quantity into its two right-angle components. This is called RESOLUTION.

Figure 4-7 illustrates the use of a resolver in solving a resolution problem. Assume that a voltage, E, and an angle, R, represent the magnitude and direction of a known quantity. To determine the two right-angle components of the quantity, feed the magnitude of the quantity to one stator coil and physically turn the rotor through angle R. The input voltage (E) induces voltage E1 and E2 in the two rotor coils. The values of these rotor voltages represent the vertical and horizontal components of the known quantity and depend on both the value of E and the angle (R) through which the rotor was turned.

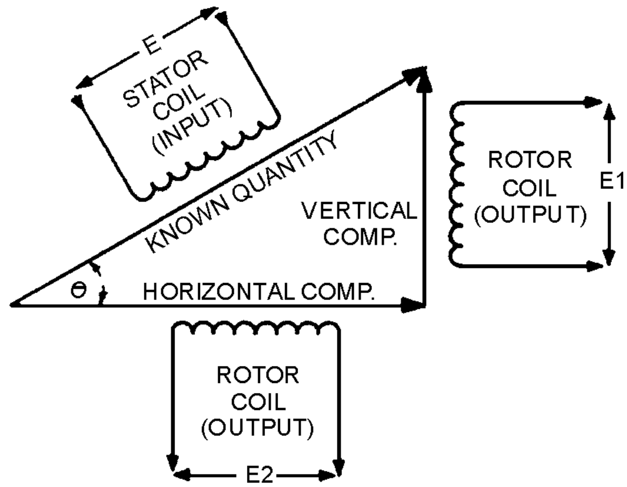


Figure 4-7.—One example of resolution.

Similarly, the resolver has the ability to add two vectors that are at right angles to each other and produce the resultant vector (hypotenuse) at the resultant angle. This is called COMPOSITION.

Figure 4-8 illustrates one use of a resolver in solving a composition problem. Assume that we have two known quantities, vertical and horizontal components, that are represented by E1 and E2, respectively. Each of these is fed to a stator coil. These two voltages induce a voltage, ET, in one of the rotor coils. ET represents a voltage that is proportional to the hypotenuse. The voltage induced in the other rotor coils is fed to a closed-loop servo, which positions the rotor shaft to the angle (direction) of the hypotenuse.

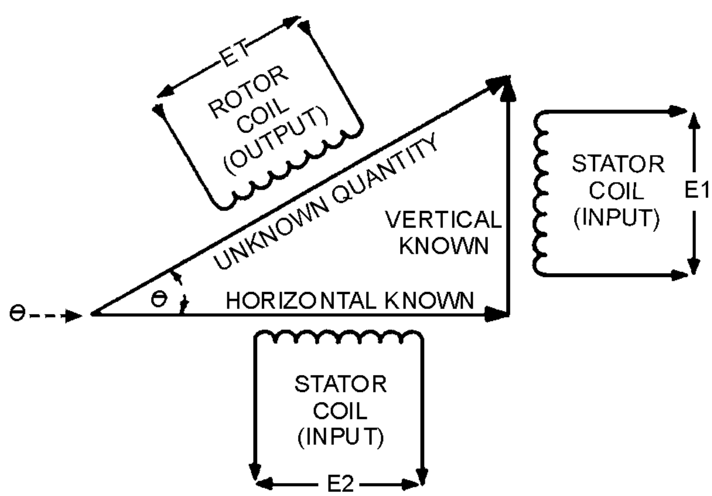


Figure 4-8.—One example of composition.

Other mathematical solutions are possible to designers who apply resolvers in equipment. Typical naval problems solved by resolvers involve distances, speeds, angular quantities, etc.

In most cases, as in figures 4-7 and 4-8, only resolver coils actively used in solving a particular problem are shown in schematics.

Resolvers can also perform a third function COMBINATION. This is the process of resolution and composition taking place simultaneously.

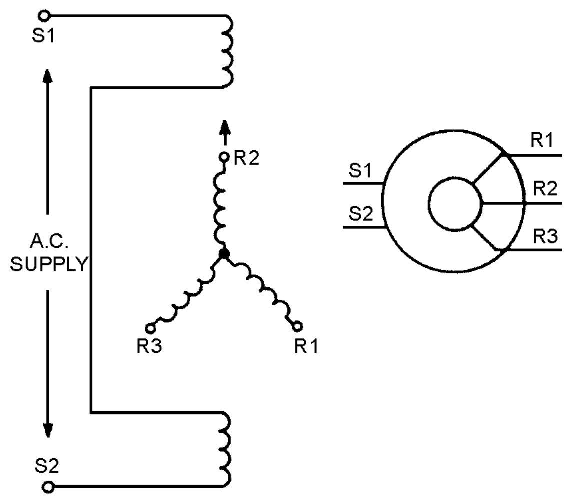
The resolver is a precision component, whose electrical characteristics are critical, and any deviation may result in excessive errors in the system. Before working on or replacing resolvers, you should check the associated equipment technical manual.

Q-4. What type of mathematical problem is solved by resolvers?

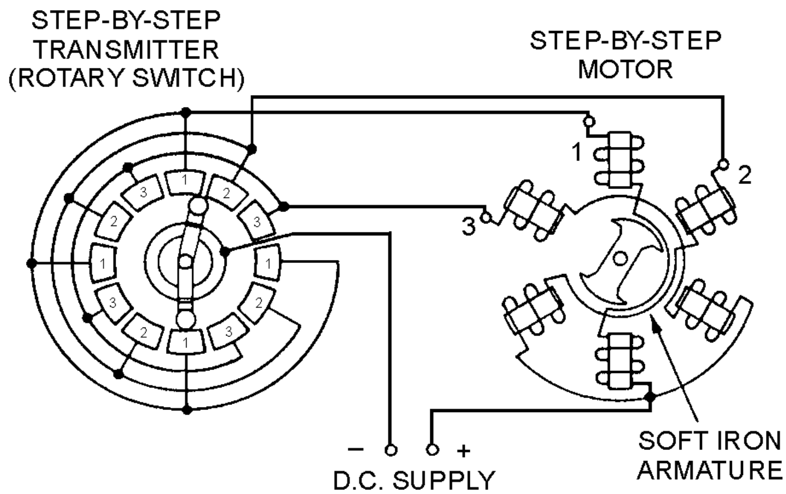
SUMMARY

The following are brief summaries of the **IC SYNCHROS, STEP-TRANSMISSION SYSTEMS**, and RESOLVERS we covered in this chapter.

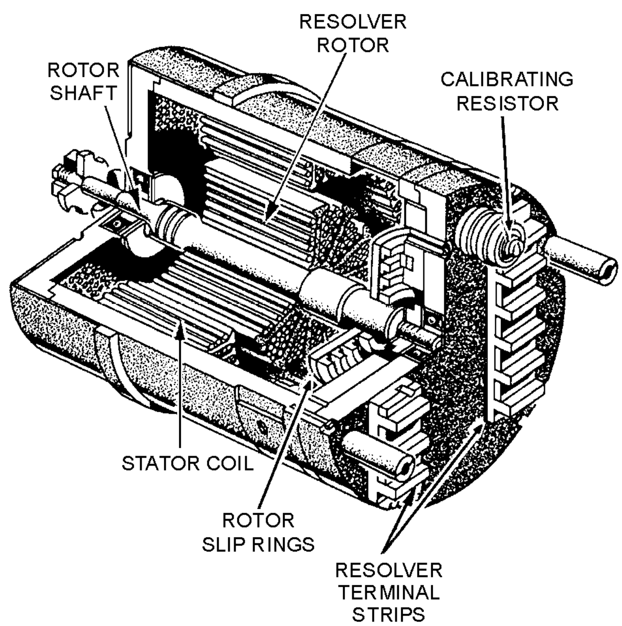
IC SYNCHROS, sometimes referred to as reversed synchros, are used in systems where transfer of information is desired. They operate on the same principles as synchros but differ in the direction of shaft rotation and amount of torque obtained.



A **STEP-TRANSMISSION SYSTEM (STEP-BY-STEP SYSTEM)** is similar to a synchro system, except that it is energized by direct current. This system uses a rotary switch to transfer angular data (information) between remote locations. The system is cheap, rugged, relatively powerful, but is not used where small changes in data are required.



RESOLVERS are physically similar to synchros and are used to perform mathematical computations. They are used extensively in analog computers, radar sets, direction and target designation equipment.



ANSWERS TO QUESTIONS Q1. THROUGH Q4.

A-1. Direction of rotation and amount of torque.

A-2. Synchros use ac; step transmission uses dc.

A-3. Rotary switch.

A-4. Right-triangle, or trigonometric, problems.

APPENDIX I

GLOSSARY

ACCELERATION SERVO SYSTEM—A servo system that controls the acceleration (rate of change in velocity) of a load.

ACCELEROMETER—A device that measures acceleration to which it is subjected and develops a signal proportional to it.

ANGULAR POSITION—The counterclockwise angular displacement of a synchro rotor, measured in degrees from its electrical zero position, as viewed from the shaft extension end of the synchro.

APPARENT DRIFT—The effect of the Earth's rotation on a gyro, which causes the spinning axis to appear to make one complete rotation in one day. Also called APPARENT PRECESSION or APPARENT ROTATION.

APPARENT PRECESSION—See apparent drift.

APPARENT ROTATION—See apparent drift.

AXIS—A straight line, either real or imaginary, passing through a body, around which the body revolves.

BANDWIDTH—The range of frequencies a servo amplifier can amplify without causing unacceptable distortion to the input signal.

CONTROL DIFFERENTIAL TRANSMITTER (CDX)—A type of synchro that transmits angular information equal to the algebraic sum or difference of the electrical input supplied to its stator, and the mechanical input supplied to its rotor. The output is an electrical voltage taken from the rotor windings.

CONTROL SYNCHRO SYSTEMS—Synchro systems that contain control synchros and are used to control large amounts of power with a high degree of accuracy. The electrical outputs of these systems control servo systems, which in turn generate the required power to move heavy loads.

CONTROL SYSTEM—A group of components, systematically organized to perform a specific control purpose. These systems are categorized as either closed- or open-loop systems. The main difference between the two is that the closed-loop system contains some form of feedback.

CONTROL TRANSFORMER (CT)—A type of synchro that compares two signals: the electrical signal applied to its stator and the mechanical signal applied to its rotor. The output is an electrical voltage, which is taken from the rotor winding and is used to control a power amplifying device. The phase and amplitude of the output voltage depend on the angular position of the rotor with respect to the magnetic field of the stator.

CONTROL TRANSMITTER (CX)—A type of synchro that converts a mechanical input, which is the angular position of its rotor, into an electrical output signal. The output is taken from the stator windings and is used to drive either a CDX or CT.

CORRESPONDENCE—The term given to the positions of the rotors of a synchro transmitter and a synchro receiver when both rotors are on 0 or displaced from 0 by the same angle.

DAMPING—A mechanical or electrical technique used in synchro receivers to prevent the rotor from oscillating or spinning. Damping is also used in servo systems to minimize overshoot of the load.

DATA TRANSMISSION—The transfer of information from one place to another or from one part of a system to another.

DEGREE-OF-FREEDOM—The number of axes about which a gyro is free to precess.

DEMODULATOR—A circuit used in servo systems to convert an ac signal to a dc signal. The magnitude of the dc output is determined by the magnitude of the ac input signal, and its polarity is determined by whether the ac input signal is in or out of phase with the ac reference voltage.

DOUBLE RECEIVER—A fine and coarse synchro receiver enclosed in a common housing with a two-shaft output (one shaft inside the other).

E-TRANSFORMER—A special form of differential transformer using an E-shaped core. The secondaries of the transformer are wound on the outer legs of the E, and the primary is on the center leg. An output voltage is developed across the secondary coils when its armature is displaced from its neutral position. This device is used as an error detector in servo systems that have limited load movements.

ELECTRICAL-LOCK—A synchro zeroing method. This method is used only when the rotors of the synchros to be zeroed are free to turn and their leads are accessible.

ELECTRICAL ZERO—A standard synchro position, with a definite set of stator voltages, that is used as the reference point for alignment of all synchro units.

ERECTING (A GYRO)—The positioning of a gyro into a desired position and the maintaining of that position.

ERROR DETECTOR—The component in a servo system that determines when the load has deviated from its ordered position, velocity, etc.

ERROR REDUCER—The name commonly given to the servo motor in a servo system. So named because it reduces the error signal by providing feedback to the error detector.

ERROR SIGNAL—In servo systems, the signal whose amplitude and polarity or phase are used to correct the alignment between the controlling and the controlled element. It is also the name given to the electrical output of a control transformer (CT).

EXCITATION VOLTAGE—The supply voltage required to excite a circuit.

FREQUENCY RESPONSE—The measure of a servo's ability to respond to various input frequencies.

GIMBAL—A mechanical frame, with two perpendicular, intersecting axes of rotation, used to support and furnish a gyro wheel with the necessary freedom to tilt in any direction.

GYRO—Abbreviation for gyroscope.

GYROSCOPE—A mechanical device containing a spinning mass mounted so that it can assume any position in space.

IC SYNCHROS—Electromechanical devices, used to transmit information, that operate on the same principles of interacting magnetic fields as synchros, but differ in their direction of rotation and the amount of torque obtainable. Because of their construction, they are sometimes called reversed synchros.

INERTIA—The physical tendency of a body in motion to remain in motion and a body at rest to remain at rest unless acted upon by an outside force (Newton's First Law of Motion).

MAGNETIC AMPLIFIER—An electromagnetic device that uses one or more saturable reactors to obtain a large power gain. This device is used in servo systems requiring large amounts of power to move heavy loads.

MILITARY STANDARD SYNCHROS—Synchros that conform to specifications that are uniform throughout the Armed Services.

MODULATOR—A circuit used in servo systems to convert a dc signal to an ac signal. The output ac signal is a sine wave at the frequency of the ac reference voltage. The amplitude of the output is directly related to the amplitude of the dc input. The circuit's function is opposite to that of a DEMODULATOR.

MULTI-LOOP SERVO SYSTEM—A servo system that contains more than one servo loop, each loop designed to perform its own function.

MULTISPEED SYNCHRO SYSTEMS—Systems that transmit data at different transmission speeds; for example, dual-speed and tri-speed synchro systems.

NEWTON'S SECOND LAW OF MOTION—If an unbalanced outside force acts on a body, the resulting acceleration is directly proportional to the magnitude of the force, is in the direction of the force, and is inversely proportional to the mass of the body.

POSITION SENSOR—A component in a servo system that measures position and converts the measurement into a form convenient for transmission as a feedback signal.

POSITION SERVO SYSTEM—A servo system whose end function is to control the position of the load it is driving.

POTENTIOMETER—An electromechanical device, used as a position sensor in servo systems, having a terminal connected to each end of a restrictive element, and a third connected to a wiper contact. The output is a voltage that is variable depending upon the position of the wiper contact. The potentiometer is commonly referred to as a variable voltage divider. It, in effect, converts mechanical information into an electrical signal.

PRECESSION—The rotation of the spin axis of a gyro in response to an applied force. The direction of precession is always perpendicular to the direction of applied force.

PRECESSION VECTOR—In a gyro, a vector representing the angular change of the spin axis when torque is applied. The precession vector represents the axis about which precession occurs.

PRESTANDARD NAVY SYNCHROS—Synchros that are designed to meet Navy, rather than servicewide, specifications.

RATE GYRO—A gyro used to detect and measure angular rates of change.

RESOLVER—A rotary, electromechanical device used to perform trigonometric computations by varying the magnetic couplings between its primary and secondary windings. It is generally used in circuits that solve vector problems, such as analog computers and conversion equipment. The resolver solves three different type problems: (1) resolution—separating a vector into two mutually perpendicular components; (2) composition—combining two components of a vector to produce a vector sum; (3) combination—the process of resolution and composition taking place simultaneously.

RESULTANT MAGNETIC FIELD—The magnetic field produced in a synchro by the combined effects of the three stator magnetic fields.

RIGIDITY—The tendency of the spin axis of a gyro wheel to remain in a fixed direction in space if no force is applied to it.

ROTOR—The rotating member of a synchro that consists of one or more coils of wire wound on a laminated core. Depending on the type of synchro, the rotor functions similar to the primary or secondary winding of a transformer. In a gyro, the rotating member is sometimes called a gyro wheel.

SCALING FACTOR—The term used to describe the use of unequal resistors in a servo's summing network to compensate for differences between input and output signal levels.

SERVO AMPLIFIER—An ac or dc amplifier used in servo systems to build up signal strength. These amplifiers usually have relatively flat gain versus frequency response, minimum phase shift, low output impedance, and low noise level. The dc amplifier is subject to excessive drift and is relatively unstable. However, the ac amplifier is considered drift free and has a constant dynamic gain over a variety of operating points.

SERVO MOTOR—An ac or dc motor used in servo systems to move a load to a desired position or at a desired speed. The ac motor is usually used to drive light loads at a constant speed, while the dc motor is used to drive heavy loads at varying speeds.

SERVO SYSTEM—An automatic feedback control system that compares a required condition (desired value, position, etc.) with an actual condition and uses the difference to adjust a control device to achieve the required condition.

SIGNAL—The angle through which a synchro transmitter rotor is mechanically turned.

SPIN VECTOR—In a gyro, a vector representing the angular velocity of the gyro rotor. The spin vector lies along the spin axis of the rotor.

STATOR—The stationary member of a synchro that consists of a cylindrical structure of slotted laminations on which three Y-connected coils are wound with their axes 120 apart. Depending on the type of synchro, the stator's functions are similar to the primary or secondary windings of a transformer.

STEP-TRANSMISSION SYSTEM—A data transmission system that operates on direct current. It consists of a step transmitter (rotary switch) and a step motor interconnected to transmit data (information) between remote locations.

STICKOFF VOLTAGE—A low voltage used in multispeed synchro systems to prevent false synchronizations.

SUMMING NETWORK—A combination of two or more parallel resistors used in servo systems as error detectors. The output of the network is the algebraic sum of the inputs.

SYNCHRO—A small motorlike device that operates like a variable transformer and is used primarily for the rapid and accurate transmission of data among equipments and stations.

SYNCHRO CAPACITOR—A unit containing three delta-connected capacitors. The synchro capacitor is used in synchro systems to increase the system's accuracy by cancelling or reducing phase shift introduced by synchro inductance.

SYNCHRO SYSTEM—Two or more synchros interconnected electrically. The system is used to transmit data among equipments and stations.

SYNCHRO TESTER—A synchro receiver with a calibrated dial. This receiver is used primarily for locating defective synchros. It can also be used for zeroing synchros.

SYNCHRO TROUBLESHOOTING—The locating or diagnosing of synchro malfunctions or breakdowns by means of systematic checking or analysis.

SYNCHRONIZING NETWORK—A circuit used in servo systems, also called a crossover or switching network, to sense how far the load is from the point of correspondence and then functions to switch the appropriate signal into control.

TACHOMETER—A small ac or dc generator, sometimes referred to as a rate generator, which converts its shaft speed into an electrical output. The tachometer is frequently used in servo systems to sense the velocity of a load.

TIME LAG—The delay in a servo system between the application of the input signal and the actual movement of the load.

TORQUE—A measure of how much load a machine can turn. This measurement is expressed either in ounce-inches for torque synchro systems or in pound-feet for heavy machinery.

TORQUE DIFFERENTIAL RECEIVER (TDR)—A type of differential synchro that takes two electrical inputs, one to the rotor and one to the stator, and produces a mechanical output. The output is the angular position of the rotor, which represents the algebraic sum or difference of the two electrical inputs.

TORQUE DIFFERENTIAL SYNCHRO SYSTEM—A synchro system containing either a TDX or a TDR. This system is used in application where it is necessary to compare two signals, add or subtract the signals, and furnish an output proportional to the sum or difference between the two signals.

TORQUE DIFFERENTIAL TRANSMITTER (TDX)—A synchro that is functionally the same as the CDX except that it is used in torque systems rather than control systems.

TORQUE RECEIVER (TR)—A type of synchro that converts the electrical input supplied to its stator back to a mechanical angular output through the movement of its rotor.

TORQUE SYNCHRO SYSTEM—A synchro system that uses torque synchros to move light loads such as dials, pointers, and other similar devices.

TORQUE TRANSMITTER (TX)—A synchro that is functionally the same as the CX except that it is used in torque synchro systems.

TORQUE VECTOR—In a gyro, a vector representing the rotary motion applied to change the direction of the rotor axis. The torque vector represents the axis about which the applied force is felt.

TRANSLATION—In a gyro, a force acting through the center of gravity of the gyro that causes no torque on the gyro rotor. Translation forces do not change the angle of the plane of rotation but move the gyroscope as a unit.

TROUBLE INDICATORS—Signal lights used to aid maintenance personnel in locating synchro troubles quickly.

TROUBLE TABLES—Tables of trouble symptoms and probable causes furnished by many manufacturers, with their equipment, to help technicians isolate synchro problems.

VELOCITY SERVO SYSTEM—A servo system that controls the speed of the load it is driving.

X-AXIS—In a gyro, the spin axis of the gyro.

Y-AXIS—In a gyro, an axis through the center of gravity and perpendicular to the spin axis.

Z-AXIS—In a gyro, an axis through the center of gravity and mutually perpendicular to both the spin (X) and Y axes.

ZEROING—The process of adjusting a synchro to its electrical zero position.

MODULE 15 INDEX

ASSIGNMENT 1

Textbook assignment: Chapter 1, “Synchros,” pages 1-1 through 1-78.

1-1. Which of the following terms accurately describes a synchro?

1. Position-sensing
2. Electromechanical
3. Rotary
4. Each of the above

1-2. What are the two general classifications of synchro systems?

1. Torque and load
2. Torque and control
3. Load and control
4. Load and lock

1-3. What is the difference in application between the two classifications of synchros?

1. Light versus heavy load
2. Mechanical versus electrical Output
3. Circular versus straight-line motion
4. High-frequency versus low-frequency operation

1-4. Which of the following types of synchro devices provides a mechanical output?

1. A control transformer
2. A torque receiver
3. A torque transmitter
4. A control transmitter

1-5. A 115-volt, 400-Hz torque transmitter with a diameter of 2.36 inches will have what military standard designation code?

1. 115 V-23CT6
2. 115 V-24TT4
3. 23TD4
4. 24TX4

1-6. A 3.5-inch diameter differential receiver will have what Navy prestandard designation code?

1. 35CR
2. 35TDR
3. 5D
4. 5DG

1-7. What does the arrow on a synchro schematic symbol indicate?

1. The direction of current flow
2. The direction of rotor movement
3. The angular position of the stator
4. The angular position of the rotor

1-8. What are the two major components of a synchro?

1. The rotor and the stator
2. The housing and the stator
3. The rotor and the shaft
4. The housing and the shaft

1-9. What type of rotor can be composed of a single winding or three Y-connected windings?

1. Salient-pole
2. Drum or wound
3. Fixed
4. "H"

1-10. How does the stator of a TX receive voltage?

1. By a physical connection with the rotor
2. By a magnetic coupling with another stator
3. By a magnetic coupling with the rotor
4. By a physical connection with a source

1-11. What part of a synchro provides a point for external connections?

1. The terminal board
2. The slip ring
3. The stator
4. The brush

1-12. Which of the following terms is defined as the amount of load a machine can turn?

1. Radian force
2. Load factor
3. Torque
4. Tension

1-13. Which of the following units should be used in measuring the amount of turning force of a synchro?

1. Ounces
2. Pounds
3. Foot-pounds
4. Ounce-inches

1-14. An overloaded synchro will probably exhibit which of the following conditions?

1. Overspeed
2. Oscillation
3. Excessive temperature
4. Noisy operation

1-15. A synchro receiver has which of the following characteristics that is NOT found in an ordinary transformer?

1. A primary that can rotate in relation to the secondary
2. A primary magnetically coupled to the secondary
3. A step-up turns ratio
4. An air core

1-16. When a synchro transmitter is in the zero-degree position, the rotor is aligned in what manner?

1. With winding S1
2. With winding S2
3. With winding S3
4. Between winding S1 and S2

1-17. Maximum voltage is induced in a stator winding of a synchro transmitter when the rotor and the stator winding have what angle between them?

1. 0 degrees
2. 30 degrees
3. 60 degrees
4. 90 degrees

1-18. Which of the following factors does NOT affect the amplitude of the voltage induced in a stator winding of a synchro transmitter?

1. The angular displacement between the rotor and stator
2. The amplitude of the primary voltage
3. The speed of data transmission
4. The turns ratio of the synchro

1-19. Damping is necessary for which of the following synchro devices?

1. Receiver
2. Transmitter
3. Control transformer
4. Differential transmitter

1-20. The primary purpose of damping is to reduce which of the following conditions in a synchro device?

1. Readings 180º out of phase
2. Overheating
3. Oscillating
4. Each of the above

1-21. What is the minimum number of synchro devices needed for a simple synchro transmission system?

1. One
2. Two
3. Three
4. Four

1-22. In a simple synchro system, what leads are connected to the source voltage?

1. R1 and R2
2. S1 and S2
3. S2 and S3
4. R1 and S1

1-23. When a synchro transmitter, and receiver are in correspondence, what is the relative value of the (a) current through the stators and (b) receiver torque?

1. (a) Maximum (b) maximum
2. (a) Maximum (b) minimum
3. (a) Minimum (b) minimum
4. (a) Minimum (b) maximum

1-24. What term applies to the angle through which a synchro transmitter rotor is rotated mechanically?

1. Lag
2. Lead
3. Gain
4. Signal

1-25. If a synchro receiver is required to rotate in a direction opposite to the rotation of the transmitter rotor, what leads should be reversed?

1. R1 and R2
2. S1 and S2
3. S2 and S3
4. S1 and S3

1-26. If a synchro receiver and transmitter are always 180 degrees out of phase with each other, what leads are reversed?

1. R1 and R2
2. S1 and S2
3. S2 and S3
4. S1 and S3

1-27. What type of synchro can accept two signals simultaneously and add or subtract?

1. Transmission
2. Differential
3. Automatic
4. Shiftless

1-28. What are the two types of synchro devices that will accept two inputs?

1. TR and TX
2. TR and TDX
3. TDR and TX
4. TDR and TDX

1-29. What types of synchro devices have (a) one electrical and one mechanical input and an electrical output; and (b) two electrical inputs and a mechanical outputs?

1. (a) TR (b) TX
2. (a) TX (b) TR
3. (a) TDX (b) TDR
4. (a) TDR (b) TDX

1-30. What determines whether a differential synchro device adds or subtracts its inputs?

1. The way it is connected in the system
2. The direction of rotor movement
3. The number of stator windings
4. The supply voltage polarity

1-31. In a TDX system, for the TR rotor to follow the TX rotor exactly, in what position must the TDX rotor be kept?

1. 0 degree position
2. 60 degree position
3. 120 degree position
4. 240 degree position

1-32. What is the angular position of a TR rotor when it is pointing to the S3 winding?

1. 0 degrees
2. 60 degrees
3. 120 degrees
4. 240 degrees

1-33. If a TDX system with standard synchro connections has the TX rotor at the 60-degree position and the TDX rotor at the 270-degree position, what is the position of the TR rotor?

1. 110 degrees
2. 150 degrees
3. 210 degrees
4. 250 degrees

1-34. For a TDX system to add its inputs rather than subtract them, what leads must be reversed between (a) the TX and TDX, and (b) the TR and TDX?

1. (a) S1 and S2 (b) R1 and R3
2. (a) S1 and S3 (b) R1 and R3
3. (a) S2 and S3 (b) R1 and R2
4. (a) S1 and S3 (b) R1 and R2

1-35. For a TDR system to add its inputs rather than subtract them, what leads must be reversed at the TDR?

1. S1 and S3
2. S1 and S2
3. R1 and R3
4. R1 and R2

1-36. If a TDR system is connected for addition and the TX rotor connected to the TDR rotor turns counterclockwise, in what direction will the TDR rotor field rotate?

1. In a direction determined by the other TX stator
2. In a direction determined by the other TX rotor
3. Counterclockwise
4. Clockwise

1-37. Which of the following types of synchros is used in a system requiring large amounts of power and high accuracy?

1. Torque
2. Control
3. Differential
4. Each of the above

1-38. What are the three types of control synchros?

1. TX, TR, CT
2. TX, CDX, CR
3. CX, CT, CR
4. CX, CT, CDX

1-39. The CX and CDX differ from the TX and TDX because the CX and CDX have which of the following characteristics?

1. Lower impedance windings
2. Higher impedance windings
3. Larger physical size
4. Smaller physical size

1-40. Which of the following is NOT a characteristic of the rotor of a control transformer (CT) rotor?

1. It is connected to a high-impedance load
2. It must be turned by an external force
3. It is connected to an ac source
4. It has a drum- or wound-type rotor

1-41. When a control transformer is at electrical zero, the rotor is perpendicular to what winding?

1. S1
2. S2
3. S3
4. R2

1-42. If a control transformer is held at electrical zero and the control transmitter is turned 90 degrees counterclockwise, what is (a) the amplitude of the induced voltage in the rotor of the control transformer, and (b) the phase relationship of this voltage and the excitation voltage to the control transmitter?

1. (a) Maximum (b) out-of-phase
2. (a) Maximum (b) in phase
3. (a) Minimum (b) out-of-phase
4. (a) Minimum (b) in phase

1-43. Which of the following terms applies to the output of a control transformer?

1. Mechanical movement
2. Deflection angle
3. Output voltage
4. Error signal

1-44. If the output of a control transformer is zero, what is the relationship of the rotors of the control transformer and the control transmitters?

1. In correlation
2. Out of correlation
3. In correspondence
4. Out of correspondence

1-45. Synchro capacitors are used to provide which of the following characteristics in a synchro system?

1. Improved accuracy
2. Reduced oscillations
3. Wider frequency response
4. Higher load-carrying capacity

1-46. Which of the following synchro devices uses a synchro capacitor?

1. TX
2. RX
3. TDR
4. CDX

1-47. What type of current is eliminated by synchro capacitors?

1. Loss
2. Rotor
3. Stator
4. Magnetizing Stator

1-48. In what configuration are synchro capacitors connected in a synchro circuit?

1. Wye, across the rotor windings
2. Delta, across the rotor windings
3. Wye, across the stator windings
4. Delta, across the stator windings

1-49. To maintain system accuracy, where are synchro capacitors physically placed in a synchro circuit?

1. Close to the TX or RX
2. Close to the TDX, CDX, or CT
3. Midway between the TX and CT
4. Far away from the TDR, CDX, or CT

1-50. Synchro systems that transmit data at two different speeds are referred to by which of the following terms?

1. Dual-speed
2. Two-speed
3. Twin-speed
4. Each of the above

1-51. Multispeed synchro systems have which of the following advantages over single-speed synchro systems?

1. Easier to troubleshoot and align
2. Fewer moving parts
3. Greater accuracy
4. All of the above

1-52. What does the gear ratio between the two transmitters in a dual-speed synchro system determine?

1. The direction of transmitter-rotation
2. The direction of receiver rotation
3. The speeds of transmission
4. The relative direction of rotation

1-53. Which of the following synchro systems, if any, should be used to transmit very large quantities?

1. Single-speed
2. Two-speed
3. Tri-speed
4. None of the above

1-54. Which of the following is a disadvantage of a double receiver as compared to two single receivers?

1. The entire unit must be replaced if one portion fails
2. It takes up much more space
3. It is much more costly
4. It is much heavier

1-55. The voltage used to prevent false synchronizations is known by what term?

1. Error voltage
2. Signal voltage
3. Source voltage
4. Stickoff voltage

1-56. What is the reference point for the alignment of all synchro units?

1. Mechanical zero
2. Electrical zero
3. Mechanical null
4. Electrical null

1-57. What is the most accurate method of aligning a synchro?

1. The dc voltmeter method
2. The ac voltmeter method
3. The synchro-tester method
4. The electric-lock method

1-58. During synchro alignment, what is the purpose of the coarse setting?

1. To ensure a setting of zero degrees rather than 180 degrees
2. To prevent the voltmeter from being overloaded
3. To keep the synchro device from overheating
4. To correct the fine setting

1-59. If a synchro receiver is properly zeroed, when do the stator windings have electrical zero voltages?

1. When the rotor is moving
2. When the rotor is stopped
3. When the rotor is at 270 degrees
4. When the rotor is at its reference position

1-60. When a 115-volt synchro transmitter is set on its coarse-zero position, approximately what voltage should be read on a voltmeter?

1. 15 volts
2. 26 volts
3. 37 volts
4. 193 volts

1-61. When a 115-volt source is used during the alignment of a differential synchro, what is the maximum time the circuit can be energized without causing damage to the synchro?

1. 1 minute
2. 2 minutes
3. 15 minutes
4. 30 minutes

1-62. After a control transformer has been zeroed and clamped down, what is the final step in the zeroing procedure?

1. Replace the fuses
2. Turn it to 270 degrees
3. Recheck the zero voltage reading
4. Disconnect all wires to the control transformer

1-63. The output voltage of a control transformer on electrical zero is which of the following relative values?

1. Equal to the supply voltage
2. Half the supply voltage
3. Maximum
4. Minimum

1-64. When a tri-speed synchro system is being zeroed, which synchro should be zeroed first?

1. Coarse
2. Medium
3. Largest
4. Fine

1-65. What method of zeroing a synchro is the fastest but NOT the most accurate?

1. The dc voltmeter method
2. The ac voltmeter method
3. The synchro-tester method
4. The electrical-lock method

1-66. The electrical-lock method of zeroing a synchro requires accessible leads and which of the following conditions?

1. A rotor free to turn
2. A stator free to turn
3. A supply voltage to the stators
4. A zero-volt potential between S1 and S2

1-67. A synchro is zeroed by the use of a synchro tester. After it is zeroed, the S1 and S3 leads are shorted together, and the synchro tester dial moves. What does this indicate?

1. The synchro is zeroed correctly
2. The synchro is not zeroed correctly
3. The supply voltage is too low
4. The supply voltage is too high

1-68. If you find that a synchro has bad bearings, which of the following actions should you take?

1. Replace the bearing
2. Lubricate the synchro
3. Replace the synchro
4. Continue to use it

1-69. Which of the following troubles is common in newly installed synchro systems?

1. Dirty brushes
2. Improper wiring
3. Worn slip rings
4. Shorted synchro windings

1-70. What type of indicating device is usually installed in the stator circuit of a torque synchro system?

1. A voltmeter indicator
2. An ohmmeter indicator
3. An overload indicator
4. A blown-fuse indicator

1-71. A synchro system with four receivers is malfunctioning. All of the receivers have incorrect readings. Which of the following is/are the most likely cause(s) of the trouble?

1. Damper failure
2. The transmitter
3. One of the receivers
4. All of the receivers

1-72. An ac voltmeter is connected between windings S1 and S3 of a synchro transmitter. Which of the following rotor positions should give a zero voltage reading?

1. 180º
2. 240º
3. 300º
4. 330º

1-73. When a synchro tester is used in place of a synchro transmitter, which of the following precautions will help to keep the tester from being overloaded?

1. Use a 26-volt supply only
2. Use a 115-volt supply only
3. Use only one syncho receiver
4. Use at least three synchro receivers

ASSIGNMENT 2

Textbook assignment: Chapter 2, “Servos,” pages 2-1 through 2-38. Chapter 3, “Gyros,” pages 3-1 through 3-27. Chapter 4, “Related Devices,” pages 4-1 through 4-12.

2-1. A servo is normally designed to move (a) what type of load to (b) what type of positions?

1. (a) Small (b) Exact
2. (a) Small (b) Approximate
3. (a) Large (b) Exact
4. (a) Large (b) Approximate

2-2. Servo systems can be found in which of the following forms?

1. Pneumatic
2. Hydraulic
3. Electromechanical
4. Each of the above

2-3. Which of the following systems are control systems?

1. Open-loop
2. Closed-loop
3. Both 1 and 2 above
4. Inductive-loop

2-4. A servo system is defined as which of the following types of control systems?

1. Open-loop
2. Closed-loop
3. Both 1 and 2 above
4. Inductive-loop

2-5. Which of the following is a basic difference between an open-loop control system and a closed-loop control system?

1. Number of loops
2. Size of the load
3. Speed of movement
4. System of feedback

IN ANSWERING QUESTIONS 2-6 THROUGH 2-8, MATCH THE FUNCTION LISTED IN COLUMN B TO THE SERVO COMPONENT LISTED IN COLUMN A THAT ACCOMPLISHES THE FUNCTION.

A. SERVO COMPONENTS	B. FUNCTIONS
2-6. Synchro control system	1. Moves the load
2-7. Servo amplifier	2. Provides power
2-8. Servo motor	3. Controls movement
	4. Converts dc to ac

2-9. In a dc position servo system, what characteristic of the error signal determines the direction in which the load is driven?

1. Amplitude
2. Frequency
3. Polarity
4. Phase

2-10. The sum point in a position servo system combines what two signals to produce an error signal?

1. Response and output
2. Feedback and Output
3. Feedback and input
4. Output and input

2-11. A position servo system exhibits a series of overtravels. This condition is known by which of the following terms?

1. Hunting
2. Overdamping
3. Undershooting
4. All of the above

2-12. A velocity servo has which of the following characteristics?

1. Senses position of the load; no error signal at correspondence
2. Senses position of the load; error signal present at correspondence
3. Senses speed of the load; no error signal at correspondence
4. Senses speed of the load; error signal present at correspondence

2-13. What device is usually used to provide feedback in a velocity servo loop?

1. Potentiometer
2. Tachometer
3. CT
4. CX

2-14. For a servo system to operate smoothly and efficiently, it must have balance between which of the following factors?

1. Acceleration and speed
2. Inertia and oscillation
3. Amplification and damping
4. Overshooting and feedback signal

2-15. When friction-clutch damping is used in a servo system, the first overshoot of the load may be characterized as

1. small
2. large
3. reversed
4. eliminated

2-16. Error-rate damping is considered to be better than friction or friction-clutch damping because of which of the following characteristics of the error-rate damping system?

1. A large error signal of short duration will not be damped
2. A small error signal of short duration will not be damped
3. A large change in the error signal causes maximum damping
4. A small change in the error signal causes maximum damping

2-17. Under what condition would a servo system that is properly designed and operating correctly have an oscillating load?

1. The input signal is large in amplitude
2. The input signal oscillates
3. Error-rate damping is used
4. Friction damping is used

2-18. A servo system is found to be "noisy." If the bandwidth of the servo amplifier were adjusted to reject the unwanted noise signals, which of the following characteristics of the servo system would be affected?

1. Amplifier gain
2. Power requirements
3. Correspondence position
4. Error-detection capability

2-19. Which of the following devices can be used as a position sensor in a servo system?

1. A summing network
2. An E-transformer
3. A potentiometer
4. A CT

2-20. Which of the following devices are magnetic error detectors?

1. CXs
2. E-transformers
3. Summing networks
4. All of the above

2-21. A dc rate generator is used in which of the following loops of a velocity servo system?

1. Prime mover
2. Feedback
3. Control
4. Error

2-22. What is the function of a modulator in a servo system?

1. To change the frequency of an ac error signal
2. To impress an ac error signal on an ac carrier
3. To convert a dc error signal to an ac error signal
4. To convert an ac error signal to a dc error signal

2-23. In a servo system that uses a modulator, what characteristic of the modulator output determines the direction of load movement?

1. Amplitude
2. Frequency
3. Polarity
4. Phase

2-24. What phase relationships between the input and reference signals are sensed by a servo demodulator?

1. 0º and 180º
2. 45º and 225º
3. 90º and 270º
4. 135º and 315º

2-25. In a properly operating servo system, what is the phase relationship between the reference voltages to the error detector and the demodulator?

1. In phase only
2. 180º out of phase only
3. Out of phase; somewhere between 0º and 180º
4. In phase or 180º out of phase, depending on the demodulator input

2-26. Which of the following should be a characteristic of a servo amplifier?

1. Narrow frequency band
2. High output impedance
3. 180º phase shift
4. Low noise level

2-27. An ac servo motor would probably be used instead of a dc servo motor in which of the following situations?

1. To move heavy loads at a constant speed
2. To move heavy loads at variable speeds
3. To move light loads at a constant speed
4. To move light loads at variable speeds

2-28. Which of the following circuits that is required in a multispeed servo system is NOT required in a single-speed servo system?

1. Position sensor
2. Error detector
3. Feedback loop
4. Synchronizer

2-29. In a two-speed servo system such as that described in the text, which of the following components controls the movement of the load at 2º but does NOT control the movement of the load at 10º?

1. Fine CT
2. Coarse CT
3. Synchronizer
4. Servoamplifier

2-30. In which of the following situations should a magnetic amplifier be used instead of a conventional amplifier?

1. When a small load is to be driven at high speeds
2. If great accuracy is required in positioning the load
3. If a dual-speed servo system is required
4. When a heavy load is to be moved

2-31. Most servo systems used in the Navy are of which of the following types?

1. Open-loop
2. Multi-loop
3. Single-loop
4. Summing-loop

2-32. Which of the following objects has gyroscopic properties?

1. A spinning top
2. A wheel on a moving bicycle
3. The moving blade assembly of an electric fan
4. Each of the above

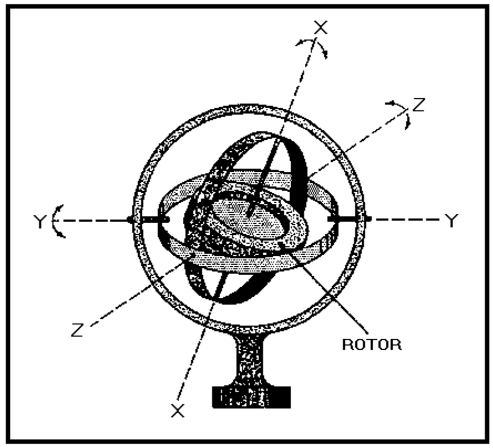


Figure 2A.—Gyro model, universally mounted.

IN ANSWERING QUESTION 2-33, REFER TO FIGURE 2A.

2-33. Which of the following axes, if any, is the gyro spin axis?

1. X-X
2. Y-Y
3. Z-Z
4. None of the above

2-34. The ability of a gyro to maintain a fixed position in space is referred to by what term?

1. Precession
2. Rigidity
3. Apparent rotation
4. Gimbal-stability

2-35. A gyro will resist all forces that attempt to change its

1. location
2. spin axis direction
3. speed of rotation
4. center of gravity

2-36. What action takes place when an outside force attempts to tilt the spin axis of a gyro?

1. The gyro precesses in the direction of the applied force
2. The gyro precesses in a direction opposite to the applied force
3. The gyro precesses in a direction at a right angle to the applied force
4. The gyro remains fixed in its original position

2-37. For a gyro to be universally mounted, it MUST have a total of how many gimbals, if any?

1. One
2. Two
3. Three
4. None

2-38. Of the following factors, which one does NOT affect rigidity?

1. Rotor speed
2. Rotor shape
3. Rotor weight
4. Rotor position

2-39. The forces that act through the center of gravity of a gyro and do NOT cause precession are referred to by what term?

1. Forces of translation
2. Forces of induction
3. Forces of isolation
4. Forces of erection

2-40. Which of the following factors determine(s) the amount of precession that will result from a given applied force?

1. Rotor speed
2. Rotor weight
3. Rotor shape
4. All of the above

2-41. Which of the following factors determine(s) the direction a gyro will precess in response to a particular force?

1. Speed of the rotor's spin
2. Shape of the rotor
3. Direction of the rotor's spin
4. All of the above

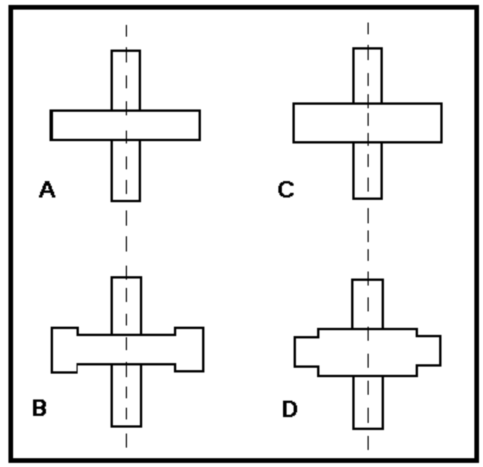


Figure 2B.—Rotors of equal weights but different shapes.

IN ANSWERING QUESTION 2-42, REFER TO FIGURE 2B.

2-42. If all of the rotors are rotated at the same speed, which one will have the greatest rigidity?

1. A
2. B
3. C
4. D

2-43. According to the right-hand rule for gyro precession, what does the thumb indicate?

1. Spin vector and axis
2. Torque vector and axis
3. Precession vector and axis
4. Axis of rotor rotation only

2-44. Which of the following is a universally mounted gyro?

1. A one-degree-of-freedom gyro
2. A two-degrees-of-freedom gyro
3. A restrained gyro
4. A rate gyro

2-45. A free gyro at the Equator appears to tilt. What is the approximate total number of degrees it will tilt in 4 hours?

1. 60º
2. 90º
3. 120º
4. 180º

2-46. Which of the following factors is NOT a cause of mechanical drift?

1. Unbalance
2. Friction
3. Apparent precession
4. Gimbal inertia

2-47. Which of the following is a purpose of a gyro-erection system?

1. To precess the gyro to its operating position
2. To prevent a gyro from precessing once the rotor is up to speed
3. To establish a vertical position to which the gyro position may be compared
4. Each of the above

2-48. Which of the following is an advantage that the mercury ballistic erection system has over the mercury erection system?

1. Greater sensitivity
2. Faster response time
3. Spin axis aligns in any desired position
4. Spin axis aligns north-south

2-49. What is the principal purpose of rate gyros?

1. To serve as gyroscopes
2. To serve as reference elements
3. To measure acceleration
4. To measure angular rates

2-50. In what maximum number of directions is a rate gyro free to precess?

1. One
2. Two
3. Three
4. Four

2-51. The amount of precession of a rate gyro is proportional to what input factor?

1. Rate of gyro case rotation
2. Amount of gyro case rotation
3. Rate of linear displacement
4. Amount of total movement

2-52. The operation of an accelerometer is based on what physical property?

1. Heat
2. Inertia
3. Gravity
4. Precession

2-53. Accelerometers find their greatest use in what type of system?

1. Navigation
2. Communication
3. Weapons control
4. Direction-indicating

2-54. Pulse-counting accelerometers are designed for use only with what type of equipment?

1. Radar sensors
2. Electronic compasses
3. Analog indicators
4. Digital computers

2-55. Which of the following is NOT a difference between IC synchros and standard synchros?

1. Amount of torque available
2. Construction of the stator
3. Construction of the rotor
4. Principle of operation

USE THE FOLLOWING INFORMATION IN ANSWERING QUESTIONS 2-56 AND 2-57. A SYNCHRO SYSTEM USING AN IC TRANSMITTER HAS THE REQUIREMENT THAT THE RECEIVER TURN IN THE OPPOSITE DIRECTION FROM THE TRANSMITTER.

2-56. If an IC receiver were used, what winding of the receiver would be connected to winding R1 of the IC transmitter?

1. R1
2. R3
3. S1
4. S1

2-57. If a standard synchro receiver were used, what winding of the receiver would be connected to winding R3 of the IC transmitter?

1. R1
2. R3
3. S1
4. S3

2-58. Angular data is to be transmitted and dc is the only power available. Which of the following systems should be used?

1. Resolver system
2. IC synchro system
3. Step-transmission system
4. Servo system using a CT and a dc servo motor

2-59. A step-transmission system is to be built in which the steps are to be smaller than the steps in the system shown on page 4-3 of the text. What must be done to the system shown in the text to accomplish this change?

1. Increase the number of coils
2. Decrease the number of coils
3. Increase the supply voltage
4. Decrease the supply voltage

2-60. Which of the following is an advantage that a step-transmission system has over a standard synchro system?

1. Smaller changes in data can be transmitted
2. Transmitted data is "smoother"
3. Synchronizing is not needed
4. Cost is considerably less

2-61. A resolver performs which of the following mathematical functions?

1. Differentiation
2. Trigonometric
3. Integration
4. Algebraic

2-62. Resolvers are used aboard a ship to keep a gun mount steady regardless of the pitch and roll of the ship. What characteristic of the resolver makes it especially useful for this application?

1. Provides instant solutions with constantly changing inputs
2. Provides higher power gain for greater accuracy
3. Uses error-rate damping for smoother solutions
4. Uses ac for greater efficiency

2-63. The (a) rotor and (b) stator of a resolver are best described by which of the following?

1. (a) A single coil
   (b) Three coils, wye-connected
2. (a) Two coils in parallel
   (b) Two coils in series
3. (a) Two coils in series
   (b) Two coils in parallel
4. (a) Two coils at right angles
   (b) Two coils at right angles

NONRESIDENT
TRAINING
COURSE
SEPTEMBER 1998

Navy Electricity and Electronics Training Series

Module 16—Introduction to Test Equipment

NAVEDTRA 14188

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Although the words “he,” “him,” and “his” are used sparingly in this course to enhance communication, they are not intended to be gender driven or to affront or discriminate against anyone.

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: To introduce the student to the subject of Test Equipment who needs such a background in accomplishing daily work and/or in preparing for further study.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and either the occupational or naval standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards,* NAVPERS 18068.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
OTMCS Richard Hall

NAVSUP Logistics Tracking Number
0504-LP-026-8410

Sailor's Creed

"I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country's Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all."

TABLE OF CONTENTS

CREDITS

The figures listed below and included in this edition of NEETS, Module 16, *Introduction to Test Equipment*, were provided by Huntron Instruments, Inc. Permission to use these illustrations is gratefully acknowledged. Permission to reproduce these illustrations must be obtained from the source.

SOURCE	FIGURE
Huntron Instruments, Inc.	5-22, 5-23, 5-24, 5-25, 5-26, 5-27, 5-28 A&B, 5-29, 5-30, 5-31, 5-32, 5-33, 5-34, 5-35, 5-36, 5-37, 5-38

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 5 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

THIS PAGE LEFT BLANK INTENTIONALLY.

Student Comments

Course Title: *NEETS Module 16*
Introduction to Test Equipment

NAVEDTRA: 14188 **Date**: ____________________

We need some information about you:

Rate/Rank and Name: ____________________ SSN: ____________ Command/Unit ____________________

Street Address: ________________________ City: ____________ State/FPO: ________ Zip ________

Your comments, suggestions, etc.:

Privacy Act Statement: Under authority of Title 5, USC 301, information regarding your military status is requested in processing your comments and in preparing a reply. This information will not be divulged without written authorization to anyone other than those within DOD for official use in determining performance.

NETPDTC 1550/41 (Rev 4-00)

CHAPTER 1

TEST EQUIPMENT ADMINISTRATION AND USE

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions included are based on the objectives. By successfully completing the NRTC, you indicate that you have met the objectives and have learned the information.

Upon completing this chapter, you should be able to:

1. Describe the Ship Configuration and Logistic Support Information System (SCLSIS).
2. State the differences between calibration and repair.
3. Explain the various calibration status labels used by the Navy.
4. List the procedures for obtaining repairs to test equipment.
5. Describe the Metrology Automated System for Uniform Recall and Reporting (MEASURE) System and the purpose of the Metrology Equipment Recall and Reporting (METER) card and recall schedule.
6. Describe major test equipment references available to you.
7. Explain the purposes and benefits of testing.
8. State the safety precautions involved in working with test equipment.
9. List three precautions you should observe to avoid damaging electric measuring instruments.
10. State the correct procedures for using a safety shorting probe.
11. Describe resistance, voltage, and current measurements in terms of purposes, methods, and instruments used.
12. Describe how capacitance and inductance are measured.
13. Explain the operation of bridges in the measurement of unknown resistances, capacitances, and inductances.

INTRODUCTION

One purpose of this chapter is to acquaint you with the practical use of test equipment. The presence of adequate test equipment in your shop is not in itself a "cure-all" for making repairs to complex electronic equipment. You must know how to best use the equipment available. First, however, you must understand the basis of electronic theory and be able to apply it to the system under repair.

Another purpose of this chapter is to introduce you to calibration and repair procedures, and basic voltage and current measurements. You will also learn how ac bridges are used for precise measurements of resistance, capacitance, and inductance.

Much of the theory of operation and practical applications of the basic types of test instruments used in electrical and electronic circuits are found in the instruction books and technical manuals that accompany various equipments. You should read and understand these books before you attempt to use any test instrument. You should also know the established safety precautions to ensure your safety and safe equipment operating procedures to protect equipment from damage.

TEST EQUIPMENT IDENTIFICATION

One of the first things you must learn as a maintenance technician is how to identify the various electronic equipment and components by their appropriate nomenclatures. You will find that several methods are used to identify test equipment used; this may be somewhat confusing to you at first. For example, a Tektronix Model 541A oscilloscope can also be identified as a CBTV-541A. The Joint Electronics Type Designation System (JETDS) is used by all branches of the military to identify equipment by a system of standardized nomenclatures.

Q-1. What system is currently used by all branches of the military to identify test equipment?

ELECTRONIC TEST EQUIPMENT CLASSIFICATION

The Electronic Test Equipment Classification Board was established in 1973 to control the increased use of undesirable electronic test equipment (ETE) in fleet and shore activities. The board classifies electronic test equipments as GENERAL PURPOSE (GPETE) or SPECIAL PURPOSE (SPETE) and assigns responsibility for their management. Items classified as general purpose are managed by the Space and Warfare Systems Command (SPAWARSYSCOM). Items classified as special purpose are managed by the individual systems command that generates the requirement.

GPETE is test equipment that has the capability, without modification, to generate, modify, or measure a range of parameters of electronic functions required to test two or more equipments or systems of basically different design.

Special-purpose electronic test equipment (SPETE) is specifically designed to generate, modify, or measure a range of parameters of electronic functions of a specific or peculiar nature required to test a single system or equipment. These special test equipments are procured by the systems command that has the responsibility for the system/equipment requiring the SPETE for maintenance.

Q-2. Name the two classes of test equipment.

Q-3. What test equipment is designed to generate, modify, or measure a range of parameters of electronic functions of a specific nature required to test a single system or equipment?

Until the ETE classification board was established, the uncontrolled increase in use of nonstandard GPETE had resulted in loss of inventory control and increased support costs. NESEA has the responsibility for evaluating requests to purchase nonstandard GPETE and for recommending its approval or disapproval to NAVSEA. NAVSEA will then forward its final decision to the originating command for such requests.

SHIP CONFIGURATION AND LOGISTIC INFORMATION SYSTEM (SCLSIS) PROGRAM

The Navy must maintain, update, and calibrate thousands of pieces of equipment. To do this, the SHIP CONFIGURATION AND LOGISTIC SUPPORT INFORMATION SYSTEM (SCLSIS) program was designed to keep track of all installed and portable equipment in the fleet. SCLSIS is used to keep up with the existence, location, and changes made to equipment. The SCLSIS program seeks to improve the quality of equipment reporting, provide information needed by other Navy management systems, and reduce record keeping. It is also designed to assist Navy supply systems that furnish spares, documentation, and training necessary to support installed and portable equipment.

Therefore, the inventory of assigned test equipment on board ship is directly related to SCLSIS records. Properly maintained SCLSIS records also show the complete inventory of test equipment on board by quantity, serial number, and location. The SCLSIS program has two basic elements: (1) VALIDATION, to establish a baseline data inventory, and (2) INVENTORY UPDATING, to correct errors or omissions and to document configuration changes.

Q-4. Name the two basic elements of the SCLSIS program.

CALIBRATION AND REPAIR PROCEDURES

The difference between the terms *calibration* and *repair* needs to be addressed before we proceed further. Calibration is little more than checking, adjusting, or systematically aligning a test instrument to a known standard. To do this, you must ensure that the equipment you send to the calibration lab is in working order.

The calibration lab is where actual repair work becomes important. Obvious problems, such as open power cords, burned components, broken meters, and missing hardware, should be repaired or replaced before sending equipment to the calibration lab. Most calibration labs with which you will deal will be part of an intermediate maintenance activity (IMA) on board a tender.

CALIBRATION STATUS

You can determine the calibration status of any test equipment by checking the calibration label or tag located on the equipment. These calibration labels or tags advise you as to whether the item is usable and within its calibration interval. Tags and labels to be used in the METROLOGY CALIBRATION (METCAL) coordination program are listed in the following paragraphs. No other calibration labels or tags are authorized to be placed on test equipment.

Calibrated Label

The CALIBRATED label, shown in view A of figure 1-1, has black lettering and a white background and comes in two sizes. It is the most commonly used label in the METCAL program. This label indicates that the instrument to which it is attached is within its applicable tolerance on all parameters. If there are any qualifying conditions for use of the instrument, one of the other labels described in the next paragraphs should be used.

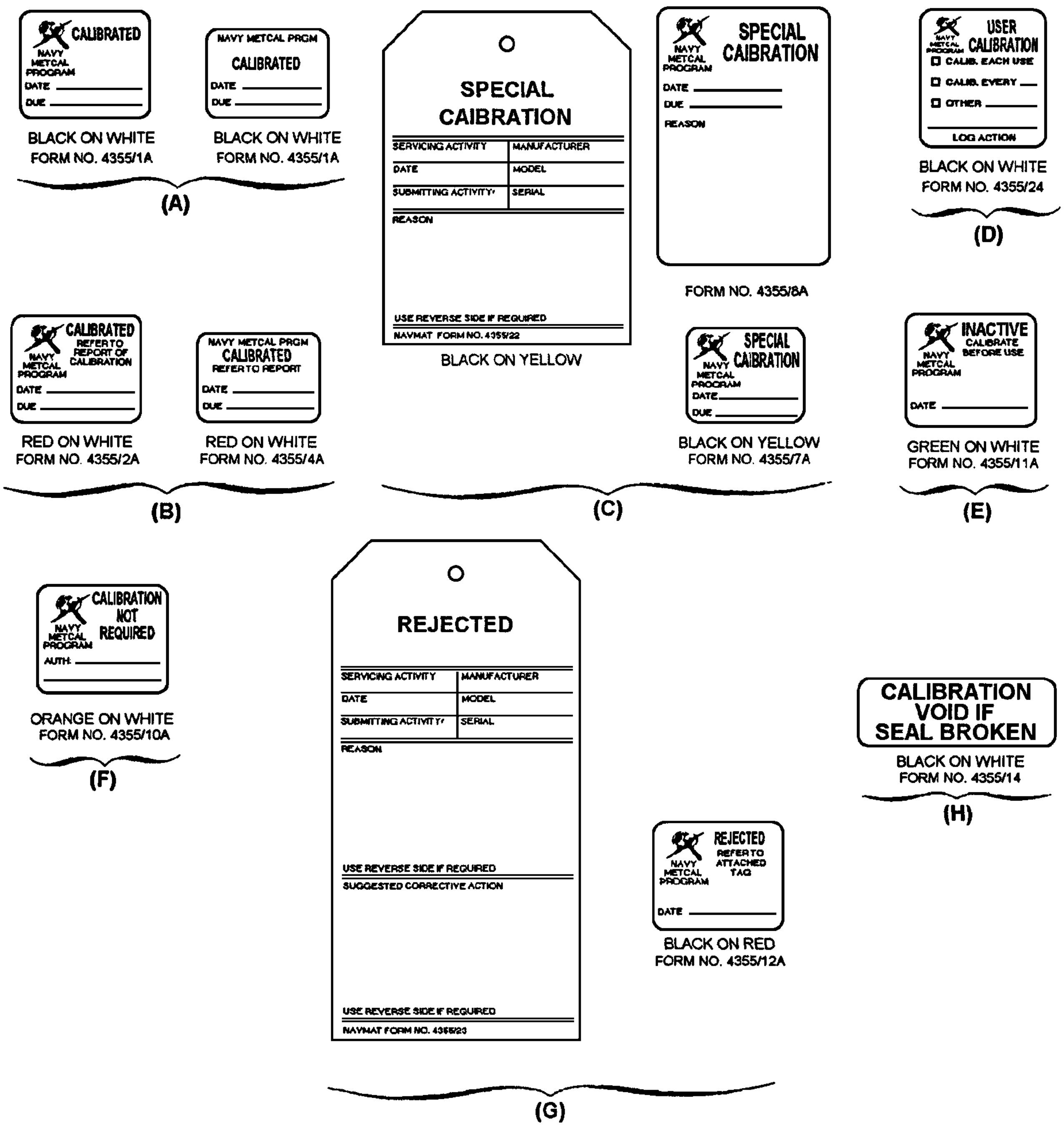


Figure 1-1.—Calibration labels and tags.

Calibrated—Refer to Report Label

The CALIBRATED—REFER TO REPORT label, shown in view B of figure 1-1, has red lettering and a white back ground. It comes in two sizes and is used when you must know the actual measurement values to use the instrument.

Q-5. What calibration label is used when actual measurement values must be known to use the test equipment?

Special Calibration Labels

Two SPECIAL CALIBRATION labels are shown in view C of figure 1-1 that have black lettering and a yellow background; the size and content of the labels are different. A SPECIAL CALIBRATION tag (figure 1-1, view C) is used with the smaller of the two labels. These labels or tag are used when some unusual or special condition in the calibration should be drawn to your attention.

Such special conditions may be deviations from usual calibration tolerances, multiple calibration intervals, or a requirement for in-place calibration. The special condition that resulted in the SPECIAL CALIBRATION label should be described on the large label when sufficient space is available on the instrument or on the tag when the small label is used. Brief descriptions of special conditions are provided in the following paragraphs.

Q-6. An instrument that must be calibrated in place requires what type of calibration label?

In cases where you do not require full instrument capability, the calibration can be performed with reduced tolerances or cover less than all ranges and parameters. This approach is often used when the instrument does not meet full calibration tolerances on certain ranges or parameters, but can still meet user requirements. On the other hand, the special calibration may be for higher accuracy than usual on a short-term basis upon your specific request.

MULTIPLE CALIBRATION INTERVALS.—Some instruments have components that require calibration less frequently than the rest of the instrument. For example, the attenuator in a signal generator may require calibration every 12 months, whereas the rest of the instrument parameters should be calibrated every 4 months. Since the attenuator calibration is time consuming and may require unavailable standards, use of the multiple-interval approach can save considerable time (man-hours) as well as permit the more frequent calibration to be performed at a lower level laboratory.

When a specific instrument has been designed for multiple calibration intervals, such information is provided in the applicable calibration procedure. The SPECIAL CALIBRATION label or tag is annotated with the words MULTIPLE INTERVAL, and the type of calibration performed is indicated; for example, partial 1 of 2, 2 of 2, complete calibration, and so forth. The calibration due date reflects the due date of the next partial or complete calibration.

CALIBRATION IN-PLACE.—Some instruments should be calibrated in-place. Annotation on the SPECIAL CALIBRATION label or tag will alert both you and the calibrator that the instrument should not be removed, but should be calibrated in-place.

User Calibration Label

Some test and measuring equipment (T&ME) should be calibrated by you instead of your referring the instrument to a calibration facility. For example, some instruments, such as hardness testers and densitometers, are provided with their own standards and should be calibrated each time used, or at least very frequently. Some instruments, such as oscillographic recorders, may require calibration before, during, and after each use.

Other automatic test equipment (ATE) have self-calibration tests that should be performed each time used or each day of use. Still other instruments are calibrated as part of checkout procedures performed daily or weekly and recorded in maintenance logs. Whenever recognized, the requirement for calibration by the user and the calibration interval (each use—daily, weekly, every 100 hours—each overhaul, and so forth) is indicated in the Metrology Requirements List (METRL).

The USER CALIBRATION label, shown in view D of figure 1-1, has black lettering and a white background and is affixed when the calibration is performed by the user; however, this label is not replaced at each calibration. When the label is first attached to the instrument, it is annotated as to the appropriate calibration interval. Records of calibrations performed, when other than each time used, should be by normal maintenance practices; that is, in the maintenance log, on maintenance action forms, and so forth.

Inactive—Calibrate Before Use Label

In the event that an individual instrument due for recalibration will not be used for some time in the future, you may indefinitely postpone the recalibration by affixing an inactive label to the instrument. As shown in view E of figure 1-1, the INACTIVE—CALIBRATE BEFORE USE label has green lettering and a white background. The INACTIVE label remains on the instrument until it is recalibrated. The instrument is not to be used while bearing this label.

Calibration Not Required Label

Test equipment standards and T&ME not requiring calibration are shown as CALIBRATION NOT REQUIRED. This label, shown in view F of figure 1-1, has orange letters and a white background. It is attached to and should remain on the instrument indefinitely unless its calibration requirements change. If the instrument is not listed in METRL, you should use the following criteria when placing instruments in the CALIBRATION NOT REQUIRED category:

- Instrument does not make quantitative measurements nor provide quantified outputs.
- The device is "fail-safe"; that is, operation beyond specified tolerances will be apparent to the user.
- All measurement/stimulus circuits are monitored during use by calibrated instruments or are dependent on external known or calibrated sources for performance within required limits. (When determining that an instrument falls into the CALIBRATION NOT REQUIRED category, you should annotate the label as to the authority for the decision, such as METRL, technical manual, letter or message from higher authority.)

Rejected—refer To Attached Tag Label

In the event that an instrument fails to meet the acceptance criteria during calibration and cannot be adequately repaired, a REJECTED—REFER TO ATTACHED TAG label is placed on the instrument and all other servicing labels removed. This label, as shown in view G of figure 1-1, has black letters and a red background. In addition to the REJECTED label, a REJECTED tag, giving the reason for rejection and other information as required, is attached to the instrument. Both the label and tag remain on the instrument until it is repaired and recalibrated. The instrument is not to be used while bearing a REJECTED label.

Calibration Void If Seal Broken Label

The CALIBRATION VOID IF SEAL BROKEN label, shown in view H of figure 1-1, has black letters and a white background. It is placed over readily accessible (usually exterior) adjustments to prevent tampering by the user when such tampering could affect the calibration. The label should not be used to cover adjustments or controls that are part of the normal use and operation of the instrument. This label may also be used to prevent removal and/or interchange of plug-ins, modules, subassemblies, and so forth, when such removal or interchange would affect the calibration.

REPAIR PROCEDURES

If you are unable to replace a known failed component with onboard spares, you can often locate the replacement component from other supply sources. The replacement component can then be delivered, along with the inoperative equipment, to the IMA. So by sending the repair part along with the equipment, you can reduce repair time considerably. This is particularly true when your unit is getting under way and no time is available for you to complete the repair before calibration. Most operational commands have a higher supply priority for purchase of repair parts than the IMA can use.

"No Reject" Policy

IMAs have a "no reject" policy on test equipment to provide operational test equipment in a more timely manner. The "no reject" policy says, in effect, that test equipment submitted to the IMA for calibration, which is later found to require repair, will be repaired by the repair department of the IMA. Before this policy, any equipment found inoperative by the calibration lab was marked REJECTED, the reasons stated, and the equipment returned uncalibrated to the ship for repairs. The "no reject" policy does not relieve you of your responsibility to ensure your equipment is in working order prior to submitting it for calibration. Its purpose is to streamline the procedure and cut down delays in returning your equipment to you calibrated and ready to use.

Responsibility for Repair and Maintenance of Test Equipment

Generally, the responsibility for repair and maintenance of test equipment is placed on maintenance personnel. In some cases, however, maintenance personnel are not authorized to make repairs. Then the test instrument must be sent to a shore repair/calibration facility.

Q-7. Responsibility for repair and maintenance of test equipment generally rests with what group of personnel?

When test equipment is sent for calibration and repair, all accessories, such as probes, adapters, and calibration sheets, should be included. Only in emergencies or special situations should partial repair or calibration be attempted on test equipment designated as nonrepairable. Such emergency repairs should be noted on a tag attached to the unit and an entry made on the MEASURE card (discussed shortly). The equipment should then be sent at the earliest opportunity to an authorized facility so that permanent repairs can be made and the unit calibrated.

STOWAGE AND HANDLING OF TEST EQUIPMENT

Most electronic test equipment is precision equipment. Such equipment must be handled with care to properly perform its designed functions. Rough handling, excessive heat, moisture, and dust all affect the useful life of the equipment. Bumping or dropping a test instrument may ruin the calibration of a meter, cause short circuits, or damage electronic elements inside the case. Sharp bends, creases, or dents in coaxial test cables can alter the expected attenuating effect and cause false meter readings or measurements. Forced air cooling, dust filters, and heaters are used in many pieces of equipment. This test equipment requires clean air filters for proper ventilation and a warm-up period that permits units in the equipment to maintain calibrated standards.

Electronic test equipment should be stowed in a dry location with the dust cover (if provided) in place. Dust covers for spare plug-in units should be constructed for such stowage. For ease in performing maintenance, the test equipment should be stowed at a location convenient to equipment spaces. If possible, related test equipment should be mounted in the equipment spaces. This reduces the problem of finding adequate stowage space elsewhere.

In stowage spaces, individual pieces of test equipment should be held in place by stretch seat-belt-type straps. If bars are used to hold equipment on shelves, meters and control knobs should be protected by blocking the equipment to prevent it from rolling and sliding on the shelf. Test equipment too large for shelf stowage should be kept in stowage cases and tie-downs provided to secure the cases. Refer to *Stowage Guide for Portable Test Equipment*, NAVSEA ST000-AB-GYD-010/GPETE, to determine adequate stowage space and proper weight support requirements.

THE METROLOGY AUTOMATED SYSTEM FOR UNIFORM RECALL AND REPORTING (MEASURE)

For the sake of simplicity, we will use the more commonly used acronym MEASURE instead of the full name to describe this system in the next discussion.

MEASURE is a data processing system designed to provide a standardized system for the recall and scheduling of test, measurement, and diagnostic equipment (TMDE) into calibration facilities. It also provides for the documentation of data pertaining to the calibration actions performed by these facilities.

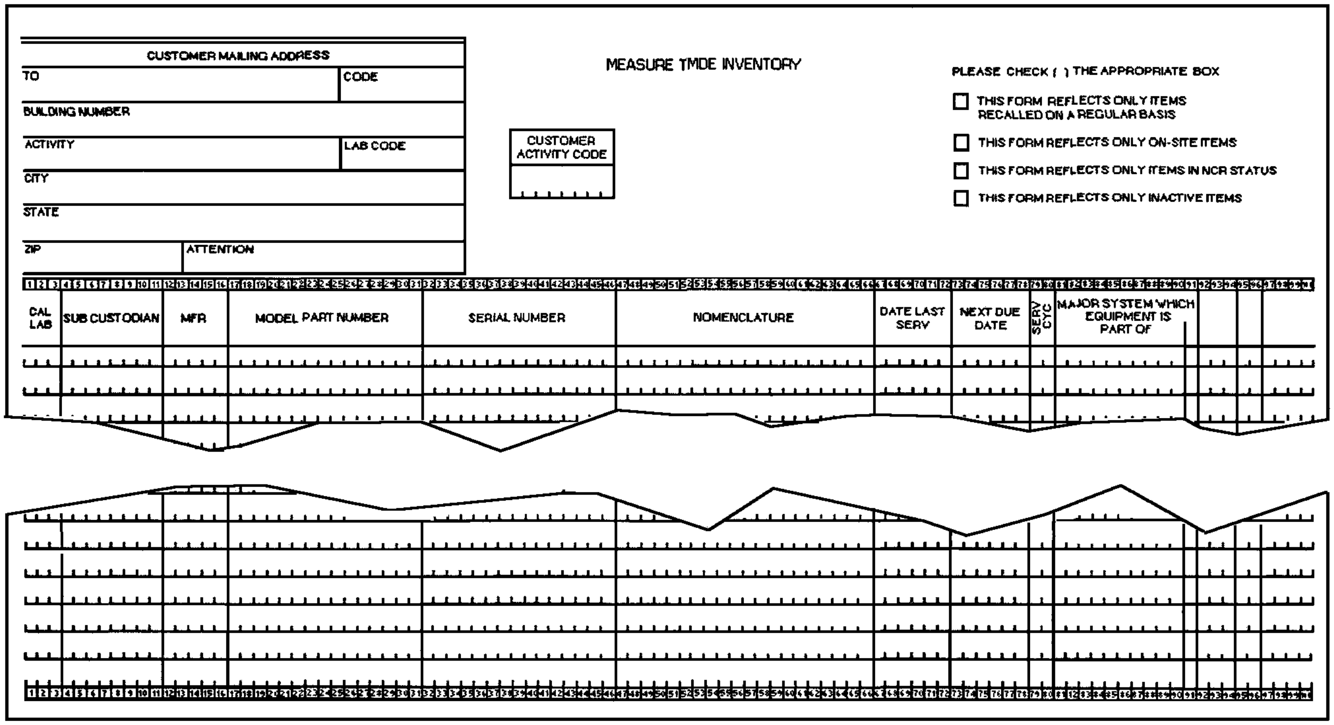
The primary reference document that describes the operation of the MEASURE system is *Metrology Automated System for Uniform Recall and Reporting (MEASURE) Users Manual*, OP 43P6A. The Chief of Naval Operations oversees this program and establishes policy and guidelines.

Q-8. What Navy office oversees the MEASURE program?

Each naval activity must ensure that the test equipment for which it has been assigned primary responsibility is submitted on a timely basis to a calibration activity for required calibration.

The MEASURE program is designed to assist these naval activities in the fulfillment of this responsibility. MEASURE does this by providing for the automatic scheduling and recall of all such test equipment for calibration.

Each activity submits an initial inventory, using the form shown in figure 1-2, to its Metrology Calibration Representative (METCALREP) for approval. The METCALREP then forwards the inventory to the Measure Operational Control Center (MOCC). The MOCC, based on the information contained on these inventory report forms, provides the necessary preprinted Metrology Equipment Recall and Reporting (METER) card. Figure 1-3 illustrates a MEASURE METER card.



CUSTOMER MAILING ADDRESS

TO

CODE

BUILDING NUMBER

ACTIVITY

LAB CODE

CITY

STATE

ZIP

ATTENTION

MEASURE TMDE INVENTORY

CUSTOMER ACTIVITY CODE

PLEASE CHECK ( ) THE APPROPRIATE BOX

☐ THIS FORM REFLECTS ONLY ITEMS RECALLED ON A REGULAR BASIS

☐ THIS FORM REFLECTS ONLY ON-SITE ITEMS

☐ THIS FORM REFLECTS ONLY ITEMS IN NCR STATUS

☐ THIS FORM REFLECTS ONLY INACTIVE ITEMS

CAL LAB	SUB CUSTODIAN	MFR	MODEL PART NUMBER	SERIAL NUMBER	NOMENCLATURE	DATE LAST SERV	NEXT DUE DATE	SERV CYC	MAJOR SYSTEM WHICH EQUIPMENT IS PART OF			

Figure 1-2.—MEASURE TMDF Inventory report form.

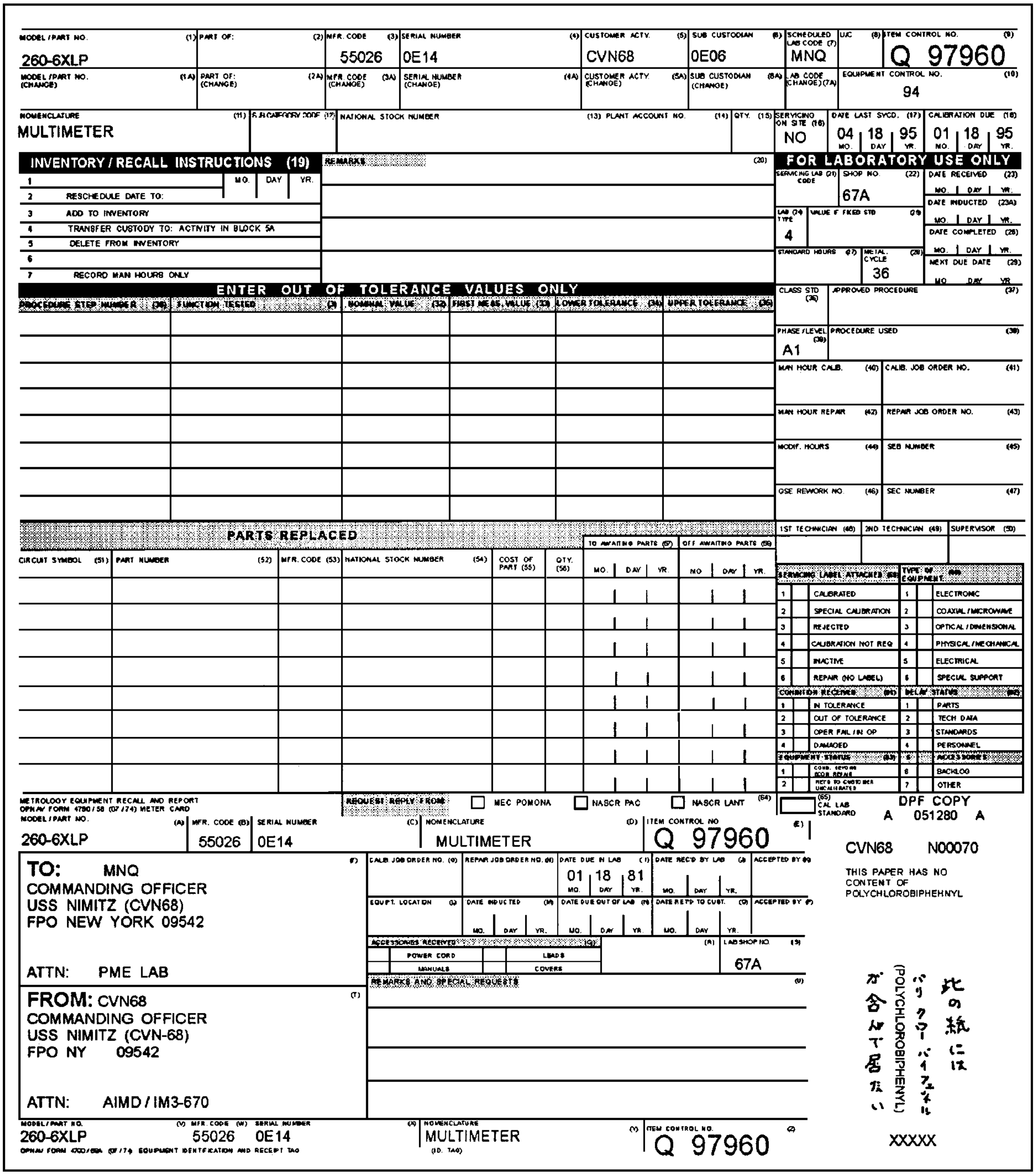


Figure 1-3.—MEASURE METER card.

In part, the METER card is preprinted with information taken from the initial inventory data submitted on the inventory report forms together with such updated data as may appear on any prior METER card. The remaining information required is entered on the card by the user of the equipment or the calibration activity, as appropriate.

The METER card is used to report changes, additions, or deletions to the user activity's inventory. It is also used to report changes in custody of the item of test equipment. The procedure for filling out the METER card is outlined in the appendixes of the *MEASURE Users Manual*. Blank METER cards can be obtained through the responsible METCALREP.

A computer printout recall schedule is also generated by the MEASURE system. The purpose of this printout is to list those items of equipment that are due for calibration. Each recall schedule is composed of a set of four identical copies. One set is provided to the calibration activity as an aid to workload planning; a second set is sent to the user's activity. The recall schedule is one of several products/formats sent automatically by the MEASURE Operation Control Center to the user activity on a regular basis. The MOCC automatically distributes the following products to user activities at the intervals shown:

DOCUMENT TITLE	TYPE DOCUMENT	INTERVAL
Format 215	Unmatched listing	As required
Format 310	Test equipment inventory	Monthly
Format 350	Test equipment inventory in sub-custodian order	Monthly
Format 804	Recall schedule for on-site equipment	Monthly/Quarterly
Replenishment cards	Preprinted METER card	As required
Blank METER cards		Initial issue

TEST EQUIPMENT REFERENCES

Several publications that contain information concerning test equipment are required to be maintained aboard ship by type commander instructions. These requirements are usually found in the inspection checkoff list. Other publications, while not required by directive, are necessary to you as reference and study material so you will be able to administer an effective test equipment program. Technicians should become familiar with the publications/directives listed in appendix II of this module.

INTRODUCTION TO TROUBLESHOOTING

Our military forces increasingly rely on electrical and electronic equipment to help perform their mission. The effectiveness of our tactical forces depends on many types of electronic systems, such as communications systems, detection systems, and fire control systems. The reliability of such equipment is determined by many factors; however, the primary factors are the quality of the equipment in use, the availability of spare parts, and the ability of maintenance personnel to perform adequate maintenance.

Maintenance is work done to correct, reduce, or counteract wear, failure, and damage to equipment. Maintenance of electrical and electronic equipment is divided into two main categories: PREVENTIVE (routine) and CORRECTIVE maintenance. Preventive maintenance consists of mechanical, electrical, and electronic checks to determine whether equipment is operating properly. It also consists of visual inspections of cabling and equipment for damage and to determine if lubrication is needed. Corrective maintenance isolates equipment failure by means of test techniques and practices; it also replaces defective parts and realigns or readjusts equipment to bring it back to proper performance.

Q-9. What are the two main categories of maintenance?

Q-10. What type of maintenance involves isolating equipment troubles and replacing defective parts?

Testing and troubleshooting are the areas of maintenance that require the greatest technical skill. Testing procedures are referred to as **measurements**, **tests**, and **checks**. The definitions of these terms often overlap, depending on their use and the results obtained. For example, a power measurement and a frequency check could constitute a test of the operation of the same radio transmitter.

Troubleshooting is a term which we in the electronics field use daily. But what does it mean? Troubleshooting is sometimes thought to be the simple repair of a piece of equipment when it fails to function properly. This, however, is only part of the picture. In addition to repair, you, as a troubleshooter, must be able to evaluate equipment performance. You evaluate performance by comparing your knowledge of how the equipment should operate with the way it is actually performing. You must evaluate equipment both before and after repairs are accomplished.

Equipment performance data, along with other general information for various electronic equipments, is available to help you in making comparisons. This information is provided in performance standards books for each piece of equipment. It illustrates what a particular waveform should look like at a given test point or what amplitude a voltage should be, and so forth. This data aids you in making intelligent comparisons of current and baseline operating characteristics for the specific equipment assigned to you for maintenance. ("Baseline" refers to the initial operating conditions of the equipment on installation or after overhaul when it is operating according to design.)

Remember, *maintenance* refers to all actions you perform on equipment to retain it in a serviceable condition or to restore it to proper operation. This involves inspecting, testing, servicing, repairing, rebuilding, and so forth. Proper maintenance can be performed only by trained personnel who are thoroughly familiar with the equipment. This familiarity requires a thorough knowledge of the theory of operation of the equipment.

A logical and systematic approach to troubleshooting is of the utmost importance in your performance of electronics maintenance. Many hours have been lost because of time-consuming "hit-or-miss" (often referred to as "easter-egging") methods of troubleshooting.

GENERAL TEST EQUIPMENT INFORMATION

In any maintenance training program, one of your most important tasks is to learn the use of test equipment in all types of maintenance work. To be effective in maintenance work, you must become familiar not only with the common types of measuring instruments, but also with the more specialized equipment. Some examples of common types of typical measuring instruments are the ammeter, voltmeter, and ohmmeter; examples of specialized test equipment are the spectrum analyzer, dual-trace oscilloscope, and power and frequency meters.

TEST EQUIPMENT SAFETY PRECAUTIONS

The electrical measuring instruments included in test equipment are delicately constructed and require certain handling precautions to prevent damage and to ensure accurate readings. In addition, to prevent injury to personnel, you must observe precautions while using test equipment. You can find a list of applicable instructions in appendix II of this module.

Instrument Precautions

To prevent damage to electrical measuring instruments, you should observe the precautions relating to three hazards: mechanical shock, exposure to magnetic fields, and excessive current flow.

MECHANICAL SHOCK.—Instruments contain permanent magnets, meters, and other components that are sensitive to shock. Heavy vibrations or severe shock can cause these instruments to lose their calibration accuracy.

EXPOSURE TO STRONG MAGNETIC FIELDS.—Strong magnetic fields may permanently impair the accuracy of a test instrument. These fields may impress permanent magnetic effects on permanent magnets, moving-coil instruments, iron parts of moving-iron instruments, or in the magnetic materials used to shield instruments.

EXCESSIVE CURRENT FLOW.—This includes various precautions, depending on the type of instrument. When in doubt, use the maximum range scale on the first measurement and shift to lower range scales only after you verify that the reading can be made on a lower range. If possible, connections should be made while the circuit is de-energized. All connections should be checked to ensure that the instrument will not be overloaded before the circuit is reenergized.

Other Instrument Precautions

Precautions to be observed to prevent instrument damage include the following:

- Keep in mind that the coils of wattmeters, frequency meters, and power meters may be carrying large quantities of current even when the meter pointer is on scale.
- Never open secondaries of current transformers when the primary is energized.
- Never short-circuit secondaries of potential transformers the primary is energized.
- Never leave an instrument connected with its pointer off-scale or deflected in the wrong direction.
- Ensure that meters in motor circuits can handle the motor starting current. This may be as high as six to eight times the normal running current.
- Never attempt to measure the internal resistance of a meter movement with an ohmmeter since the movement may be damaged by the current output from the ohmmeter.
- Never advance the intensity control of an oscilloscope to a position that causes an excessively bright spot on the screen; never permit a sharply focused spot to remain stationary for any period of time. This results in burn spots on the face of the cathode-ray tube (CRT).
- In checking electron tubes with a tube tester that has a separate "short test," always make the short test first. If the tube is shorted, no further test should be made.
- Before measuring resistance, always discharge any capacitors in the circuit to be tested. Note and record any points not having bleeder resistors or discharge paths for capacitors.
- Always disconnect voltmeters from field generating or other highly inductive circuits before you open the circuit.

Q-11. Which quantity (voltage or current) determines the intensity of an electrical shock?

Situations can arise during the use of test equipment that are extremely dangerous to personnel. For example, you may have an oscilloscope plugged into one receptacle, an electronic meter plugged into another, and a soldering iron in still another. Also, you may be using an extension cord for some equipments and not others or may be using other possible combinations. Some of the hazards presented by situations such as these include contact with live terminals or test leads. In addition, cords and test leads may be cross connected in such a manner that a potential difference exists between the metal cases of the instruments. This potential difference may cause serious or fatal shocks.

Test leads attached to test equipment should, if possible, extend from the back of the instruments away from the observer. If this is not possible, they should be clamped to the bench or table near the instruments.

At times, you may use instruments at locations where vibration is present, such as near a diesel engine. At such times, the instruments should be placed on pads of folded cloth, felt, or similar shock-absorbing material.

WORKING ON ENERGIZED CIRCUITS

Insofar as is practical, you should NOT undertake repair work on energized circuits and equipment. However, it could become necessary, such as when you make adjustments on operating equipment. In such cases, obtain permission from your supervisor, then proceed with your work, but carefully observe the following safety precautions:

- DO NOT WORK ALONE.
- Station an assistant near the main switch or circuit breaker so the equipment can be immediately de-energized in case of an emergency.
- Someone qualified in first aid for electrical shock should be standing by during the entire operation.
- Ensure that you have adequate lighting. You must be able to see clearly if you are to perform the job safely and properly.
- Be sure that you are insulated from ground by an approved rubber mat or layers of dry canvas and/or wood.
- Where practical, use only one hand, keeping the other either behind you or in your pocket.
- If you expect voltage to exceed 150 volts, wear rubber gloves.
- DO NOT work on any type of electrical apparatus when you are wearing wet clothing or if your hands are wet.
- DO NOT wear loose or flapping clothing.
- The use of thin-soled shoes and shoes with metal plates or hobnails is prohibited.
- Flammable articles, such as celluloid cap visors, should not be worn.

- Remove all rings, wristwatches, bracelets, and similar metal items before working on the equipment. Also ensure that your clothing does not contain exposed metal fasteners, such as zippers, snaps, buttons, and pins.

- Do not tamper with interlock switches; that is, do not defeat their purpose by shorting them or blocking them open.

- Ensure that equipment is properly grounded before energizing.

- De-energize equipment before attaching alligator clips to any circuit.

- Use only approved meters and other indicating devices to check for the presence of voltage.

- Observe the following procedures when measuring voltages in excess of 300 volts:

  - Turn off the equipment power.

  - Short-circuit or ground the terminals of all components capable of retaining a charge.

  - Connect the meter leads to the points to be measured.

  - Remove any terminal grounds previously connected.

  - Turn on the power and observe the voltage reading.

  - Turn off the power.

  - Short circuit or ground all components capable of retaining a charge.

  - Disconnect the meter leads.

- On all circuits where the voltage is in excess of 30 volts and where decks, bulkheads, or workbenches are made of metal, you should insulate yourself from accidental grounding by using approved insulating material. The insulating material should have the following qualities:

  - It should be dry, without holes, and should not contain conducting materials.

  - The voltage rating for which it is made should be clearly marked on the material. The proper material should be used so that adequate protection from the voltage can be supplied.

  - Dry wood may be used or, as an alternative, several layers of dry canvas, sheets of phenolic (resin or plastic) insulating material, or suitable rubber mats.

  - Care should be exercised to ensure that moisture, dust, metal chips, and so forth, which may collect on insulating material, are removed at once. Small deposits of such materials can become electrical hazards.

  - All insulating materials on machinery and in the area should be kept free of oil, grease, carbon dust, and so forth, since such deposits destroy insulation.

SAFETY SHORTING PROBE

A representative shorting probe is shown in figure 1-4. An approved shorting probe is shown in NAVSEA 0967-LP-000-0100, EIMB, *General*, Section 3.

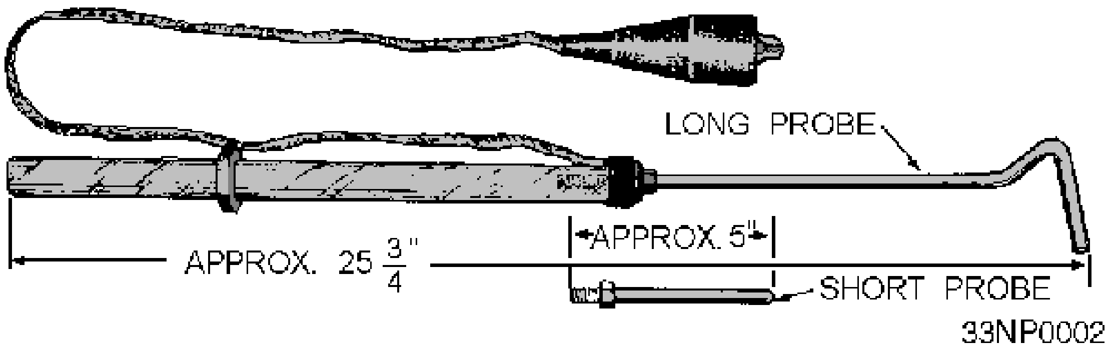


Figure 1-4.—Representative safety shorting probe.

CAUTION

Capacitors and cathode-ray tubes may retain their charge for a considerable period of time after having been disconnected from the power source.

Always assume there is a voltage present when working with circuits having high capacitance, even when the circuit has been disconnected from its power source.

An approved type of shorting probe should be used to discharge capacitors and cathode-ray tubes individually.

When using the safety shorting probe, always be sure to first connect the test clip to a good ground (if necessary, scrape the paint off the grounding metal to make a good contact). Then hold the safety shorting probe by the insulated handle and touch the probe end of the shorting rod to the point to be shorted out. The probe end is fashioned so that it can be hooked over the part or terminal to provide a constant connection by the weight of the handle alone. Always take care not to touch any of the metal parts of the safety shorting probe while touching the probe to the exposed "hot" terminal. It pays to be safe; use the safety shorting probe with care.

Some equipments are provided with walk-around shorting devices, such as fixed grounding studs or permanently attached grounding rods. When that is the case, the walk-around shorting devices should be used rather than the safety shorting probe.

Q-12. What tool is used to de-energize capacitors in a circuit that has been disconnected from its power source?

WORKING ON DE-ENERGIZED CIRCUITS

When any electronic equipment is to be repaired or overhauled, certain general safety precautions should be observed. They are as follows:

- Remember that electrical and electronic circuits often have more than one source of power. Take time to study the schematics or wiring diagrams of the entire system to ensure that all sources of power have been disconnected

- If pertinent, inform the remote station regarding the circuit on which work will be performed.

- Use one hand when turning switches on or off.

- Safety devices, such as interlocks, overload relays, and fuses, should never be altered or disconnected except for replacement. In addition, they should never be changed or modified in any way without specific authorization.

- Fuses should be removed and replaced only after the circuit has been de-energized. When a fuse "blows," the replacement should be of the same type and have the same current and voltage ratings. A fuse puller should be used to remove and replace cartridge fuses.

- All circuit breakers and switches from which power could possibly be supplied should be secured (locked if possible) in the OPEN or OFF (safe) position and danger tagged in accordance with procedures in the *Standard Organization and Regulations of the U.S. Navy*, OPNAVINST 3120.32.

- After the work has been completed, the tag (or tags) should be removed **only** by the same person who signed it (them) when the work began.

- Keep clothing, hands, and feet dry if at all possible. When you must work in wet or damp locations, place a rubber mat or other nonconductive material on top of a dry, wooden platform or stool; then use the platform or stool to sit and stand on. Use insulated tools and insulated flashlights of the molded type when you are required to work on exposed parts.

GROUNDING OF POWER TOOLS AND EQUIPMENT

The possibility of electrical shock can be reduced by ensuring that all motor and generator frames, metal bases, and other structural parts of electrical and electronic equipment are at ground potential.

Normally, on steel-hull vessels, such grounds are inherently provided because the metal cases or frames of the equipment are in contact with one another and with the metal structure of the vessel. In some instances where such inherent grounding is not provided by the mounting arrangements, such as equipment supported on shock mounts, suitable ground connections must be provided.

The grounding wire used for this purpose is generally made of flexible material (copper or aluminum) that provides sufficient current-carrying capacity to ensure an effective ground. In this manner, equipment cases and frames that are not intended to be above ground potential are effectively grounded; also, the possibility of electrical shock to personnel coming in contact with metal parts of the equipment is minimized. The secondary purpose of grounding equipment is to improve the operation and continuity of service of all equipments.

Paint, grease, or other foreign matter can interfere with the positive metal-to-metal contact at the ground connection point. Therefore, all bonding surfaces (connection points or metallic junctions) must be securely fastened and free of such matter. In all instances where equipment grounding is provided, certain general precautions and preventive maintenance measures must be taken. A few of these precautions are listed below:

- Periodically clean all strap-and-clamp connectors to ensure that all direct metal-to-metal contacts are free from foreign matter.

- Check all mounting hardware for mechanical failure or loose connections.

- Replace any faulty, rusted, or otherwise unfit grounding strap, clamp, connection, or component between the equipment and the ground to the ship's hull.

- When replacing a part of the ground connection, make certain that the metallic contact surfaces are clean and that electrical continuity is re-established.

- After completing the foregoing steps, recheck to be sure that the connection is securely fastened with the correct mounting hardware. Paint the ground strap and hardware in accordance with current procedures.

Because of the electrical shock hazards that could be encountered aboard ship, plugs and convenience outlets for use with portable equipment and power tools normally are standard three-prong type. Both plugs and outlets are keyed so that the plug must be in the correct position before it can be inserted into the receptacle. To ensure that the safety factors incorporated in these devices are in serviceable condition and are safe for use, you must perform the following precautions and inspections:

- Inspect the pins of the plug to see that they are firmly in place and are not bent or damaged.

- Check the wiring terminals and connections of the plug. Loose connections and frayed wires on the plug surface must be corrected and any foreign matter removed before the plug is inserted into the receptacle.

- Use a meter to ensure that the ground pin has a resistance of less than 1 ohm equipment ground.

- Do not attempt to insert a grounded-type plug into a grounded receptacle without first aligning the plug properly.

CAUTION

Never use a power tool or a piece of portable test equipment unless you are absolutely sure that it is equipped with a properly grounded conductor.

BASIC MEASUREMENTS

Electronic measurements involve the fundamental electrical quantities of voltage and current and the inherent characteristics of resistance, capacitance, and inductance. In circuits being tested, voltage and current are dependent upon resistance, capacitance, and inductance for their distribution; therefore, voltage and current measurements are valuable aids in determining circuit component conditions and in the evaluation of symptoms. Practically any reading obtained from the use of test equipment will depend on these basic measured quantities of resistance, capacitance, and inductance.

VOLTAGE AND CURRENT MEASUREMENTS

Voltage measurements may be made as part of either preventive or corrective maintenance. These measurements are made using a voltmeter. When compared with voltage charts, these measurements are a valuable aid in locating a trouble quickly and easily. However, if the sensitivity of the test voltmeter differs from that of the voltmeter used in preparing the chart, the voltage measurements must be evaluated before the true circuit conditions can be determined. (Sensitivity in voltmeters was discussed in NEETS, Module 3, *Introduction to Circuit Protection, Control, and Measurement*.)

Since many of the troubles you find in equipments and systems are the result of abnormal voltages, voltage measurements are a valuable aid in locating trouble. You can measure voltage with a voltmeter without interrupting circuit operation.

Point-to-point voltage measurement charts, usually found in equipment technical manuals, contain the normal operating voltages found in the various stages of the equipment. These voltages are usually measured between indicated points and ground unless otherwise stated. When you begin recording voltage measurements, it is a smart and safe practice to set the voltmeter on the highest range before measuring. This ensures that excessive voltages existing in the circuit will not cause overloading of the meter.

Q-13. On what range should you set the voltmeter prior to taking a voltage measurement?

To increase accuracy, you should then set the voltmeter to the appropriate range for the proper comparison with the expected voltage in the voltage charts. When checking voltages, remember that a voltage reading can be obtained across a resistance, even if that resistance is open. The resistance of the meter itself forms a circuit resistance when the meter probes are placed across the open resistance. Therefore, the voltage across the component may appear to be normal or near-normal as you read the meter, but may actually be abnormal when the meter is disconnected from the circuit.

If the internal resistance of the voltmeter is approximately the same value as the resistance being tested, it will indicate a considerably lower voltage than the actual voltage present when the meter is removed from the circuit. The sensitivity (in ohms per volt) of the voltmeter used to prepare the voltage charts is provided on those charts. If a meter of similar sensitivity is available, you should use it to reduce the effects of loading.

The following precautions are general safety measures that apply to the measurement of voltages. Remember that nearly all voltages are dangerous and have often proved fatal to careless technicians. When measuring voltages, be sure to observe the following precautions:

- Set test equipment to the HIGHEST range.
- Make sure safety observer knows where to secure power for the equipment under test.
- Connect the ground lead of the voltmeter first.
- Use only one hand to take measurements (when possible), and put the other hand in your pocket or behind your back.
- If the voltage to be measured is less than 300 volts, place the end of the test probe on the point to be tested; use the polarity switch to select positive or negative readings.

- If the voltage to be measured is more than 300 volts, proceed as follows:

  1. Shut off circuit power.
  2. Discharge all filter capacitors with a shorting probe.
  3. Temporarily ground the point to be measured.
  4. Connect (clip on) the proper test lead to the high-voltage point.
  5. Move away from the voltmeter.
  6. Turn on circuit power and read the voltmeter.
  7. Turn off circuit power.
  8. Discharge all capacitors before disconnecting the meter.

Q-14. When taking a voltage measurement, which lead of the voltmeter should you connect to the circuit first?

Current measurements are not often taken in the course of preventive maintenance or testing. This is because the ammeter (or other current-measuring instrument) must become an actual part of the equipment being tested. The circuit must be opened (desoldered) to connect the ammeter in series with the circuit being tested. Usually, you can take a voltage measurement and use this factor to calculate the circuit current by applying Ohm's law.

Q-15. Is an ammeter connected in series or in parallel with the circuit under test?

RESISTANCE MEASUREMENTS

Resistance measurements are a valuable aid to you in locating defective circuits and components during corrective maintenance. Maintenance handbooks for the equipment can often be used to help you take these measurements. These handbooks often contain resistance charts that are referenced to accessible test points within the equipment. Without these charts, taking resistance measurements in a complex circuit is a slow process. The process is slow because one side of the circuit component must often be desoldered to get a true resistance measurement. However, resistance tolerances vary so widely that approximate resistance readings are adequate for most jobs.

Once the most accessible test point is found, an ohmmeter is usually used to take the resistance measurement. Because of the degree of accuracy needed when an ohmmeter is used, proper calibration and understanding of the meter scales is a must. (Topic 2 of this module will discuss these requirements in detail.) When using an ohmmeter, you must observe the following precautions:

- The circuit being tested must be completely de-energized.

- Any meters or transistors which can be damaged by the ohmmeter current must be removed before any measurement is made.

Q-16. What must be done to a circuit before you can use an ohmmeter for testing?

CAPACITANCE MEASUREMENTS

Capacitance measurements are usually taken with a capacitance meter. Capacitance tolerances vary even more widely than resistance tolerances. Capacitance tolerances depend on the type of capacitor, the value of capacitance, and the voltage rating. The actual measurement of capacitance is very simple; however, you must make the important decision of whether to reject or to continue to use the capacitor after it has been tested.

The POWER FACTOR of a capacitor is important because it is an indication of the various losses of a capacitor. Power losses can be traced to the dielectric, such as current leakage and dielectric absorption. Current leakage is of considerable importance, especially in electrolytic capacitors.

Q-17. What is the term used to refer to the losses which can be traced to the dielectric of a capacitor?

INDUCTANCE MEASUREMENTS

Inductance measurements are seldom required in the course of troubleshooting. However, inductance measurements are useful in some cases; therefore, bridges (discussed in the next section) are available for making this test. You will find that many capacitance test sets can be used to measure inductance. Most capacitance test sets are furnished with inductance conversion charts if the test equipment scale is not calibrated to read the value of inductance directly.

CAPACITANCE, INDUCTANCE, AND RESISTANCE BRIDGES

You can measure capacitance, inductance, and resistance for precise accuracy by using ac bridges. These bridges are composed of capacitors, inductors, and resistors in a wide variety of combinations. These bridges are operated on the principle of a dc bridge called a WHEATSTONE BRIDGE.

Wheatstone Bridge

The Wheatstone bridge is widely used for precision measurements of resistance. The circuit diagram for a Wheatstone bridge is shown in figure 1-5. Resistors R1, R2, and R3 are precision, variable resistors. The value of R_x is an unknown value of resistance that must be determined. After the bridge has been properly balanced (galvanometer G reads zero), the unknown resistance may be determined by means of a simple formula. The galvanometer (an instrument that measures small amounts of current) is inserted across terminals **b** and **d** to indicate the condition of balance. When the bridge is properly balanced, no difference in potential exists across terminals **b** and **d**; when switch S2 is closed, the galvanometer reading is zero.

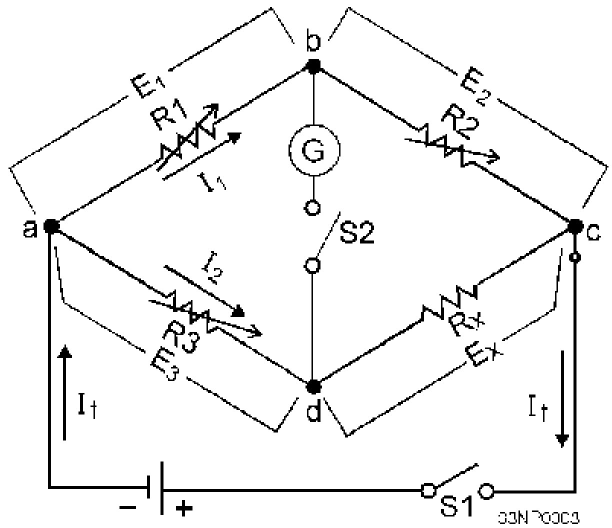


Figure 1-5.—Wheatstone bridge.

The operation of the bridge is explained in a few logical steps. When the battery switch S1 is closed, electrons flow from the negative terminal of the battery to point **a**. Here the current divides as it would in any parallel circuit. Part of it passes through R1 and R2; the remainder passes through R3 and R_x. The two currents, I_1 and I_2, unite at point **c** and return to the positive terminal of the battery. The value of I_1 depends on the sum of resistance R1 and R2, and the value of I_2 depends on the sum of resistances R3 and R_x. In each case, according to Ohm's law, the current is inversely proportional to the resistance.

R1, R2, and R3 are adjusted so that when S1 is closed, no current flows through **G**. When the galvanometer shows no deflection, there is no difference of potential between points **b** and **d**. All of I_1 follows the **a b c** path and all I_2 follows the **a b c** path. This means that a voltage drop E_1 (across R1 between points **a** and **b**) is the same as voltage drop E_3 (across R3 between points a and d). Similarly, the voltage drops across R2 and R_x (E_2 and E_x) are also equal. Expressed algebraically,

$$E_1 = E_3$$
$$I_1R1 = I_2R3$$

and

$$E_2 = E_X$$
$$I_1R2 = I_2R_X$$

With this information, we can figure the value of the unknown resistor R_x. Divide the voltage drops across R1 and R3 by their respective voltage drops across R2 and R_x as follows:

$$\frac{I_1R1}{I_1R2} = \frac{I_2R3}{I_2R_X}$$

We can simplify this equation:

$$\frac{R1}{R2} = \frac{R3}{R_X}$$

then we multiply both sides of the expression by R_x to separate it:

$$R_X = \frac{R2R3}{R_X}$$

For example, in figure 1-5, we know that R1 is 60 ohms, R2 is 100 ohms, and R3 is 200 ohms. To find the value of R_x, we can use our formula as follows:

$$R_X = \frac{R2R3}{R1}$$

$$R_X = \frac{100 \times 200}{60}$$

$$R_X = \frac{20{,}000}{60}$$

$$R_X = 333.33 \text{ ohms}$$

Use of ac Bridges

A wide variety of ac bridge circuits (such as the Wheatstone) may be used for the precision measurement of ac resistance, capacitance, and inductance. Let's look at ac bridges in terms of functions they perform.

RESISTANCE BRIDGE.—An ac signal generator, as shown in figure 1-6, is used as the source of voltage. Current from the generator passes through resistors R1 and R2, which are known as the ratio arms, and through R_s and R_x. Again, R_x is known as resistance. R_s has a standard value and replaces R3 in figure 1-6. When the voltage drops across R2 and R_s are equal, the voltage drops across R2 and R_x are also equal; no difference of potential exists across the meter and no current flows through it. As we discovered with the Wheatstone bridge, when no voltage appears across the meter, the following ratio is true:

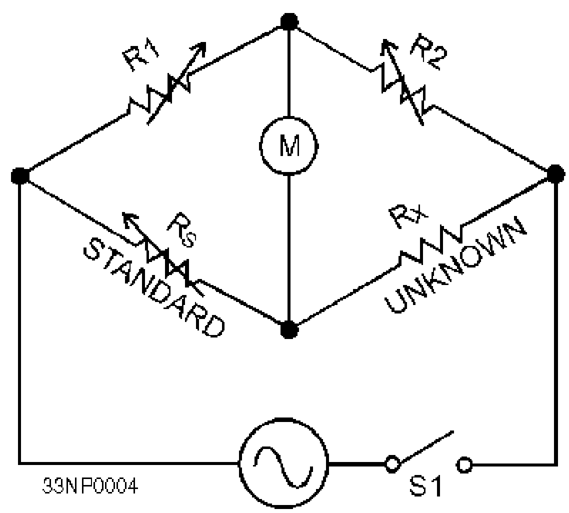


Figure 1-6.—Resistance bridge (ac).

For example, if in figure 1-6 we know that R1 is 20 ohms, R2 is 40 ohms, and R_s is 60 ohms, we can find the value of R_x using our formula as follows:

$$R_X = \frac{R2R_S}{R1}$$

$$R_X = \frac{40 \times 60}{20}$$

$$R_X = \frac{2,400}{20}$$

$$R_X = 120 \text{ ohms}$$

With the ac signal applied to the bridge, R1 and R2 are varied until a zero reading is seen on the meter. Zero deflection indicates that the bridge is balanced. (**NOTE**: In actual practice, the variables are adjusted for a minimum reading since the phase difference between the two legs will not always allow a zero reading.)

CAPACITANCE BRIDGE.—Because current varies inversely with resistance and directly with capacitance, an inverse proportion exists between the four arms of the bridge in figure 1-7; the right side of our expression is inverted from the resistance bridge expression as follows:

$$\frac{R1}{R2} = \frac{C_X}{C_S}$$

or

$$C_X = \frac{R1C_S}{R2}$$

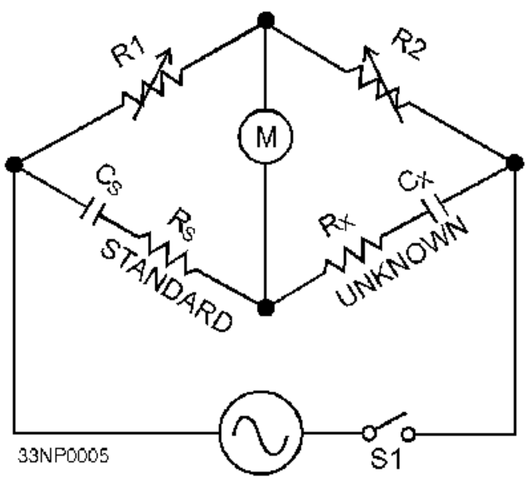


Figure 1-7.—Capacitance bridge.

Q-18. What effect does an increase in capacitance have on a capacitor's opposition to current flow?

Because R1 and R2 are expressed in the same units, the equation R1/R2 becomes a simple multiplication factor. This equation provides a numerical value for C_x and will be in the same units as C_s (farad, microfarad, and so forth).

Similarly, the following resistance ratio exists between the four arms of the bridge, just as in the resistance bridge expression discussed earlier:

$$\frac{R1}{R2} = \frac{R_S}{R_X}$$

or

$$R_X = \frac{R2R_S}{R1}$$

Thus, both the unknown resistance and capacitance, R_x and C_x, can be estimated in terms of known resistance R1, R2, R_s, and known capacitance C_s.

In figure 1-7, for example, we know that R1 is 20 ohms, R2 is 40 ohms, R_s is 60 ohms, and C_s is 10 microfarads. We can find the values of C_x and R_x by using the respective formulas as follows:

$$C_X = \frac{R1C_S}{R2}$$

$$C_X = \frac{20 \times 10}{40}$$

$$C_X = \frac{200}{40}$$

$$C_X = 5 \text{ microfarads}$$

and

$$R_X = \frac{R2R_S}{R1}$$

$$R_X = \frac{40 \times 60}{20}$$

$$R_X = \frac{2{,}400}{20}$$

$$R_X = 120 \text{ ohms}$$

Q-19. When a bridge is used to measure resistance, what is the value of R_x if R1 equals 80 ohms, R2 equals 120 ohms, and R3 equals 280 ohms?

INDUCTANCE BRIDGE.—The value of the unknown inductance L_x may be determined by means of the simple bridge circuit shown in figure 1-8. Ratio arms R1 and R2 are accurately calibrated resistors. L_s is a standard inductor with a known inductance; R_s is the known resistance, and R_x represents the resistance of the unknown inductor.

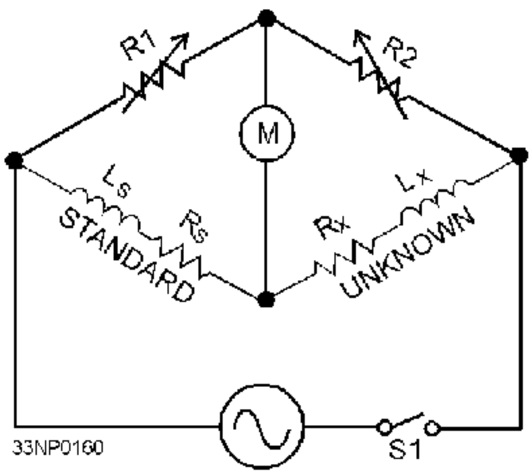


Figure 1-8.—Inductance bridge.

The ac signal is applied to the bridge, and variable resistors R1 and R2 are adjusted for a minimum or zero deflection of the meter, indicating a condition of balance. When the bridge is balanced, the following formulas may be used to find L_x.

(NOTE: The right side of this expression is NOT inverse as it was in the capacitance bridge.)

$$\frac{R1}{R2} = \frac{L_S}{L_X}$$

$$\frac{R2L_S}{R1} = L_X$$

and

$$\frac{R1}{R2} = \frac{R_S}{R_X}$$

or

$$R_X = \frac{R2R_S}{R1}$$

In figure 1-8, for example, the values of R1, R2, and R_s are 20, 40, and 60 ohms, respectively. The value of L_s is 10 millihenries. We can find the values of R_x and L_x by using their respective formulas as follows:

$$L_X = \frac{R2L_S}{R1}$$

$$L_X = \frac{40 \times 10}{20}$$

$$L_X = \frac{400}{20}$$

$$L_X = 20 \text{ millihenries}$$

and

$$R_X = \frac{R2R_S}{R1}$$

$$R_X = \frac{40 \times 60}{20}$$

$$R_X = \frac{2,400}{20}$$

$$R_X = 120 \text{ ohms}$$

Thus, both the unknown resistance and inductance can be estimated in terms of the known values for R1, R2, R_s, and L_s.

Q-20. When an unknown capacitance is tested with a bridge, what is the value of C_x if R1 equals 70 ohms, R2 equals 150 ohms, and C_s equals 550 microfarads?

SUMMARY

The important points of this chapter are summarized in the following paragraphs:

The **JETDS SYSTEM** is jointly used by all branches of the military to identify equipments by a system of standardized nomenclatures.

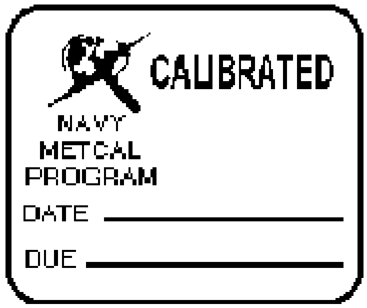
GPETE is test equipment that has the capability, without modifications, to generate, modify, or measure a range of parameters of electronic functions required to test two or more equipments or systems of basically different design. All GPETE are listed in *Standard General Purpose Electronic Test Equipment*, MIL-STD-1364 (series).

SPETE is test equipment that is specifically designed to generate, modify, or measure a range of parameters of electronic functions of a specific or peculiar nature required to test a single equipment or system.

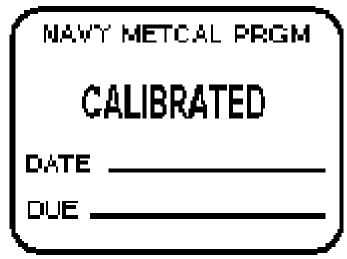
The **SHIP CONFIGURATION AND LOGISTICS INFORMATION SYSTEM (SCLSIS)** program is designed to keep track of equipment configuration changes in the fleet.

The **SCLSIS** program has two basic elements, VALIDATION and INVENTORY UPDATING.

The **CALIBRATION STATUS** of any items of test equipment can be determined by the information recorded on the calibration label or tag located on the equipment.



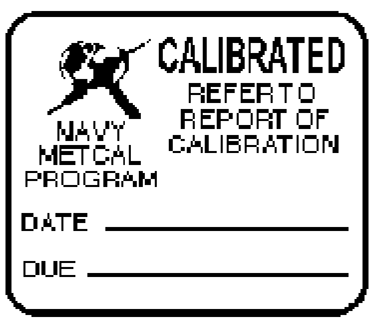
BLACK ON WHITE
FORM NO. 4355/1A



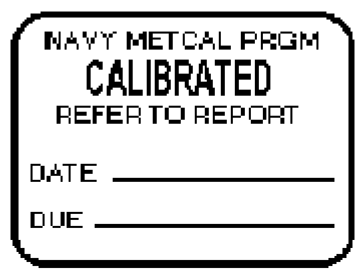
BLACK ON WHITE
FORM NO. 4355/3A

The **CALIBRATED** label, with black lettering on a white background, indicates the instrument to which it is attached is within tolerance on all scales.

The **CALIBRATED—REFER TO REPORT** label, with red lettering on a white background, is used when actual measurement values must be known to use the instrument.

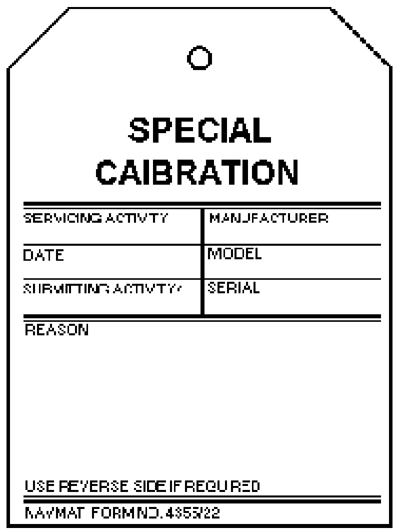


RED ON WHITE
FORM NO. 4355/2A

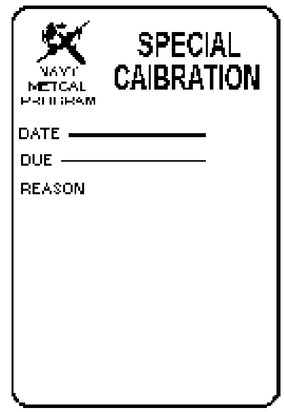


RED ON WHITE
FORM NO. 4355/4A

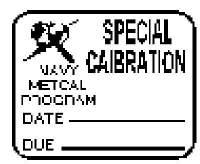
The **SPECIAL CALIBRATION** label, with black lettering on a yellow background, is used when some unusual or special condition in the calibration should be drawn to your attention.



BLACK ON YELLOW

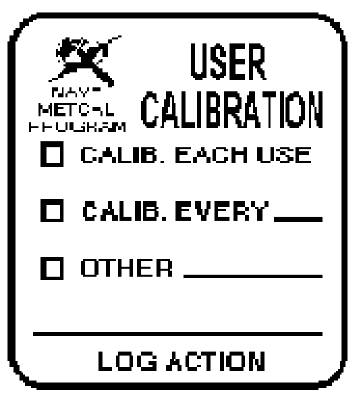


FORM NO. 4355/8A



BLACK ON YELLOW
FORM NO. 4355/7A

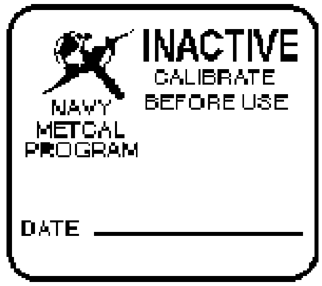
The **USER CALIBRATION** label indicates that you should calibrate the test and measuring instrument instead of sending the instrument to a calibration facility.



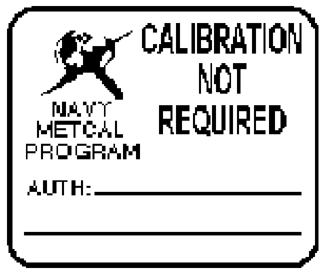
BLACK ON WHITE
FORM NO. 4355/24

The **INACTIVE—CALIBRATE BEFORE USE** label is used when a piece of test equipment due for recalibration will not be used for some time in the future.

The **CALIBRATION NOT REQUIRED** label is used on test instruments listed in the Metrology Requirements List (METRL) as not requiring calibration.

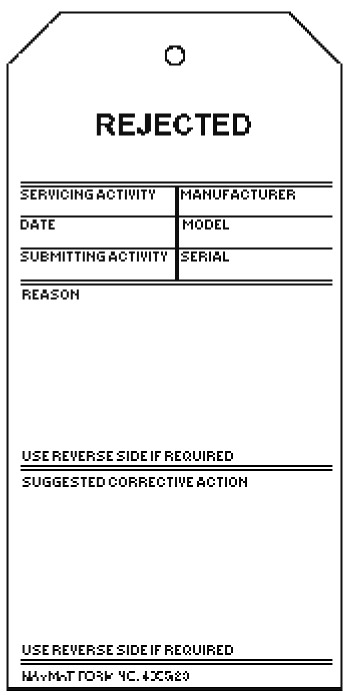


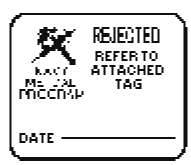
GREEN ON WHITE
FORM NO. 4355/11A



ORANGE ON WHITE
FORM NO. 4355/10A

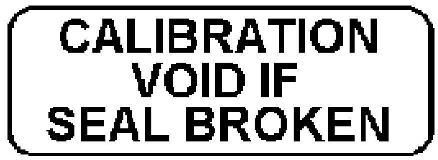
The **REJECTED** label is attached to a test instrument that fails to meet the acceptance criteria during calibration and cannot be repaired.





BLACK ON RED
FORM NO. 4355/12A

The **CALIBRATION VOID IF SEAL BROKEN** label is placed over readily accessible adjustments to prevent tampering by the user when such tampering could affect the calibration.



BLACK ON WHITE
FORM NO. 4355/14

The **MEASURE** system is designed to standardize the recall and scheduling of test, measurement, and diagnostic equipment into calibration facilities and for the documentation of actions performed by the calibration facility.

MAINTENANCE is work done to correct, reduce, or counteract wear and damage to equipment.

PREVENTIVE MAINTENANCE consists of checks to determine weather equipment is functioning properly. It also consists of visual inspections of cabling and equipment for damage and to determine if lubrication is needed.

CORRECTIVE MAINTENANCE is used to isolate troubles by means of test techniques and practices that realign or readjust equipment or otherwise bring the equipment back up to proper performance.

SENSITIVITY of the voltmeter is always given on the voltage charts for a particular piece of equipment. You should always use a voltmeter of similar sensitivity to the equipment to diminish the effects of circuit loading.

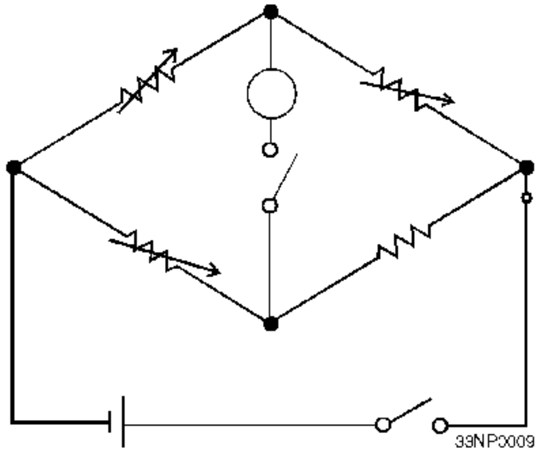
CURRENT MEASUREMENTS are not often taken in the course of testing because the ammeter (or other current measuring device) must become an actual part of the equipment being tested. The circuit must be opened for necessary connection of the meter. Usually you can use a voltage measurement to calculate the circuit current by applying Ohm's law.

You should observe the following PRECAUTIONS when using an ohmmeter:

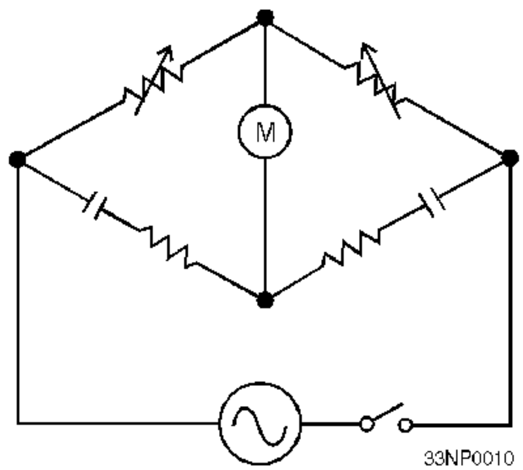
1. The circuit being tested must be completely de-energized.

2. Any circuit components which can be damaged by ohmmeter current must be removed before any measurement is made.

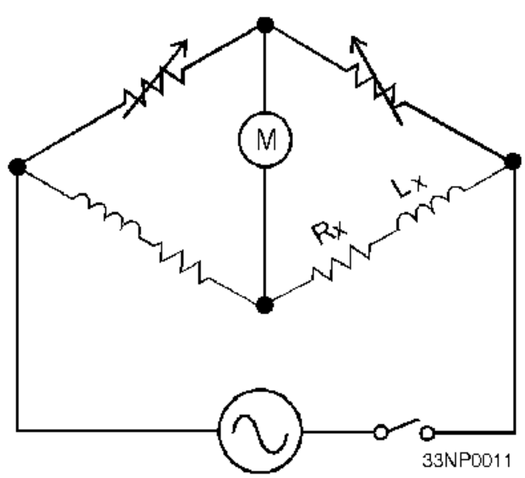
The **WHEATSTONE BRIDGE** is used for precise measurement of resistance.



The **CAPACITANCE BRIDGE** is used for measuring an unknown capacitance.



An **INDUCTANCE BRIDGE** is used to find the value of an unknown inductance.



ANSWERS TO QUESTIONS Q1. THROUGH Q20.

A-1. Joint Electronics Type Designation System (JETDS).

A-2. General-purpose electronic test equipment (GPETE) and special-purpose electronic test equipment (SPETE).

A-3. Special-purpose electronic test equipment.

A-4. Validation and updating.

A-5. CALIBRATED—REFER TO REPORT.

A-6. SPECIAL CALIBRATION label.

A-7. Maintenance personnel.

A-8. The Chief of Naval Operations.

A-9. Preventive and corrective maintenance.

A-10. Corrective maintenance.

A-11. Current.

A-12. Shorting probe.

A-13. Highest.

A-14. Ground.

A-15. In series.

A-16. It must be de-energized.

A-17. Power losses.

A-18. Opposition to current flow decreases.

A-19. 420 ohms.

A-20. 256 microfarads.

CHAPTER 2

MISCELLANEOUS MEASUREMENTS

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to:

1. Define and explain the use of the terms "dB" and "dBm" as they apply to power measurements.
2. Describe the use of resistive loads, bolometers, and thermocouples in power measurements.
3. Explain the measurement of mechanical rotation using the tachometer, stroboscope, and the strobotac.
4. Explain the measurement of frequency in various ranges using vibrating reeds, tuned circuits, heterodyne frequency meters, absorption wavemeters, cavity wavemeters, and frequency counters.
5. Describe the use of frequency-measurement devices, oscilloscopes, and spectrum analyzers in waveform analysis and maintenance.
6. Describe semiconductor testing and applicable terms in maintenance.

INTRODUCTION

In chapter 1, you studied test equipment administration and the basic measurements that all technicians are responsible for performing. Chapter 2 presents miscellaneous measurements that are fairly common; keep in mind, however, that you may not routinely perform these measurements in your particular job. This chapter introduces you to several test instruments and components found in those test instruments. It will also serve as a review of some of the basics of electronic theory related to test equipment.

POWER MEASUREMENTS

You may be required to check the power consumption and the input-signal power levels of electronic equipment. The determination of dc power is fairly simple; recall that the unit of power, the **watt**, is the product of the potential in volts and the current in amperes ($P = E \times I$).

As discussed in NEETS, Module 2, *Introduction to Alternating Current and Transformers*, the phase angle of the voltage and current must be considered for accurate ac power measurements. The measurement of ac power is further complicated by the frequency limitations of various power meters. If there is no phase angle difference, you can compute ac power in the same manner as dc power; that is, by determining the effective value of the product of the voltage and current.

For equipments that operate in the audio-frequency (af) range, power levels have to be determined in the performance of routine checks and during corrective maintenance procedures.

Power measurements for af circuits are usually indicated in terms of decibels (dB) or decibels referenced to 1 milliwatt (dBm). Because the actual calculation of decibel measurements is seldom required, the following explanation is somewhat simplified. Most test equipment is designed to measure and indicate decibels directly. This eliminates the need for you to perform complicated calculations. Nevertheless, a basic explanation of the decibel measurement system is necessary for you to understand the significance of dB readings and amplifier-gain ratings that are expressed in decibels.

THE DECIBEL SYSTEM

The basic unit of measurement in the system is not the decibel; it is the **bel**. The bel is a unit that expresses the logarithmic ratio between the input and the output of any given component, circuit, or system. It may be expressed in terms of voltage, current, or power. Most often, it is used to show the ratio between input and output power to figure gain. You can express the power gain of the amplifier (N) in bels by dividing the output (P_1) by the input (P_2) and taking the base 10 logarithm of the resulting quotient. The formula for determining this gain is:

$$\log_{10}\frac{P_1}{P_2}$$

If an amplifier doubles the input power, the quotient of P_1 to P_2 will be 2. If you consult a logarithm table, you will find that the base 10 logarithm of 2 is 0.3, making the power gain of the amplifier 0.3 bel.

Q-1. What is the logarithmic ratio between the input and output of a given circuit called?

Experience has shown that because the bel is a rather large unit, it is difficult to apply. A more practical unit, and one that can be used more easily, is the decibel (1/10 bel). You can convert any figure expressed in bels to decibels by multiplying that figure by 10 or simply by moving the decimal point one place to the right. Applying this rule, we find that the above ratio of 0.3 bel is equal to 3 decibels.

The decibel (dB) cannot be used to represent actual power; only the ratio of one power compared to another. To say that an amplifier has a 3 dB gain means that the output power is twice the input power. This gives no indication of the actual power represented. You must be able to state the input power for it to be meaningful. In many **applications**, a mathematical expression represents the actual power, not a power ratio. One standard reference is the dBm.

The dBm is an abbreviation used to represent power levels above or below 1 milliwatt. Negative dBm (−dBm) represents power levels below 1 milliwatt, and positive dBm (+dBm) represents power levels above 1 milliwatt. In other words, a dBm value is a specific amount of power; 0 dBm is equal to 1 milliwatt. Briefly stated, the amount of power in a given value of dBm is the power which results if 1 milliwatt is amplified or attenuated by that dB value. For example, 40 dBm represents an actual power level (watts or milliwatts) that is 40 dB above 1 milliwatt, whereas −10 dBm represents a power level that is 10 dB below 1 milliwatt. The formula for finding dBm is a variation of the dB power formula:

$$\text{dBm} = 10\ \log\frac{\text{actual power } (P_2)}{.001 \text{ watt } (P_1)}$$

Q-2. What term is used to represent power levels above or below a 1-milliwatt reference?

You do not need to use the formula in most applications. The following shows conversions of dBm to mW:

+20dBm	=	100mW
+10dBm	=	10mW
+7dBm	=	5mW
+6dBm	=	4mW
+4dBm	=	2.5mW
+3dBm	=	2mW
0dBm	=	1mW
-3dBm	=	.5mW
-10dBm	=	.1mW

For a +10 dBm level, start with the 1 milliwatt reference and move the decimal point one place to the right (+10 dBm = 10 mW). Another 10 dB increment brings the power level to +20 dBm, thereby moving the decimal point another place to the right (+20 dBm = 100 mW). For a –10 dBm level, again start with 1 milliwatt, but this time move the decimal point one place to the left (–10 dBm = .1 mW). An additional 10 dB decrease results in another decimal point shift to the left (–20 dBm = .01 mW).

For a 3 dB increase, you double the power. For a 3 dB decrease, you reduce the power by one-half (+3 dBm = 2 mW and –3 dBm = .5 mW). A +6 dBm level is an additional 3 dB change from +3 dBm. In this case, you just double the power level of the +3 dBm (+6 dBm = 4 mW).

Q-3. What milliwatt value is equal to +6 dBm?

The dB change can be made in either direction. For example, +7 dBm is a decrease from +10 dBm. Reducing the +10 dBm power by one-half, we have +7 dBm, or 5 mW. A +4 dBm power level is a 3 dB decrease from +7 dBm (+4 dBm – 2.5 mW). By using this simple method, you can quickly find any power level that corresponds to a given dBm.

Some test instruments you will be using are calibrated in decibels and have a 1 milliwatt zero reference level. Figure 2-1 illustrates such an instrument. Notice that this is an ac voltmeter in which the upper scale of the meter indicates ac voltage and the lower scale indicates decibels. The zero power-level indicator on the decibel scale is located at, or near, center scale. If the power in the line being measured is more than the reference value, the meter will indicate a value to the right of the zero mark (+dB). If the power is less than the reference value, the meter will indicate a value to the left of the zero mark (–dB). Such meters are useful when recording measurements where a direct indication in decibels is desired. However, you must remember that this meter is still a voltmeter and that power measurements are not meaningful unless the circuit impedance is known. If you feel the need to review how to calculate power in ac circuits, refer to NEETS, Module 2.

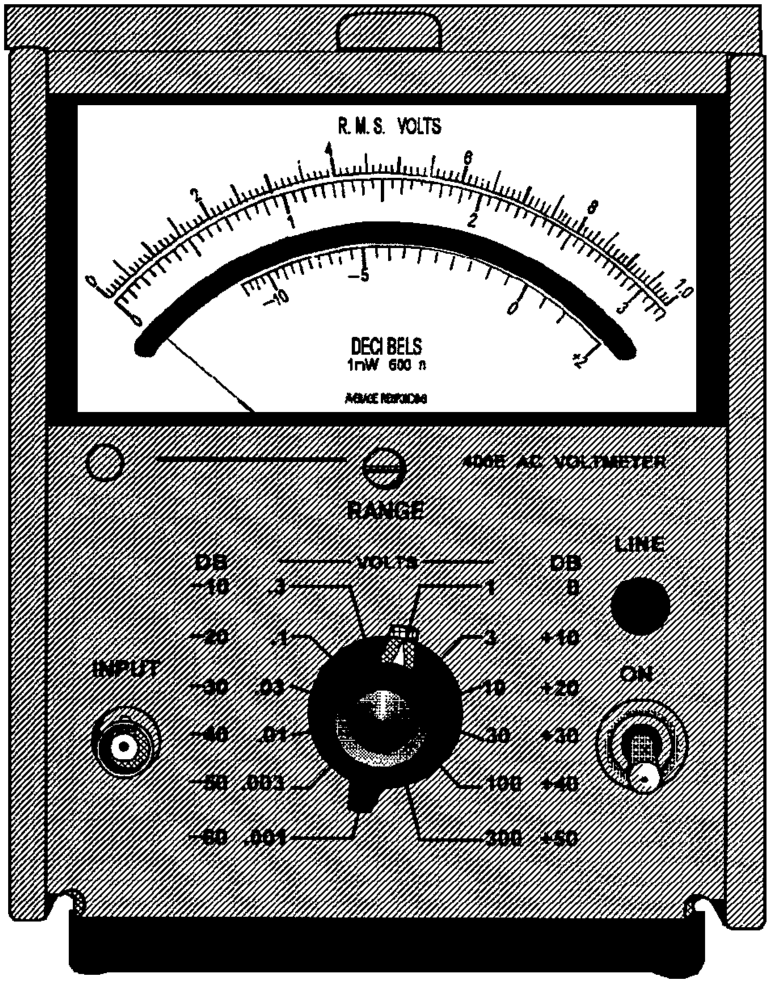


Figure 2-1.—Ac voltmeter.

MEASUREMENT METHODS

At radio frequencies below the UHF range, power is usually determined by voltage, current, and impedance measurements. One common method used to determine the output power of radio-frequency (rf) oscillators and radio transmitters consists of connecting a known resistance to the equipment output terminals. Current flowing through this resistance is then measured and the power is calculated as the product of I^2R.

Because power is proportional to the current squared, the meter scale can be calibrated to indicate power units directly. A THERMOCOUPLE AMMETER can be used in this manner for measuring rf power. The resistor used to replace the normal load is specially designed to have low reactance and the ability to dissipate the required amount of power. Such resistors are commonly called DUMMY LOADS or DUMMY ANTENNAS.

Q-4. What name is given to a resistor used to replace the normal load in a circuit?

In the UHF and SHF frequency ranges, accurately measuring the voltage, current, and resistance is difficult. These basic measurements can vary greatly, depending on where in the circuit the measurements are made. They are also affected by small changes in parts placement in the vicinity of tuned circuits.

To measure the output of microwave radio or radar transmitters, you can use test instruments that convert rf power to another form of energy, such as light or heat. These instruments can be used to indirectly measure the power. A method used to measure the effect of a resistor load on a stream of passing air can also be used to indirectly measure power. Accurate measurement of large-magnitude power also can be achieved by measuring the temperature change of a water load. The most common type of power meter for use in this frequency range employs a BOLOMETER.

Bolometer

The bolometer is a loading device that undergoes changes of resistance as changes in dissipated power occur. The two types of bolometers are the BARRETTER and the THERMISTOR. The barretter is characterized by an increase in resistance as the dissipated power rises. The thermistor decreases in resistance as the power increases. In either case, resistance is measured before and after the application of rf power. If the same change in resistance is then produced by a variable dc source of power, then the rf power is equal to the measured dc power. This relationship makes possible the direct calibration of a bridge circuit in units of power. In other words, one condition of balance exists when no rf power is applied; but in the presence of power, a second condition of balance exists because of the resistance changes of the bolometer. It is this change of resistance that is calibrated in power.

Q-5. What are the two types of bolometers?

BARRETTER.—The construction of a typical barretter is shown in figure 2-2. The fine wire (usually tungsten) is extremely small in diameter. This thin diameter allows the rf current to penetrate to the center of the wire. The wire is supported in an insulating capsule between two metallic ends, which act as connectors. Because of these physical characteristics, the barretter resembles a cartridge-type fuse. The enclosure is a quartz capsule made in two parts. One part is an insert cemented in place after the tungsten wire has been mounted. In operation, the barretter is matched to the rf line after power is applied.

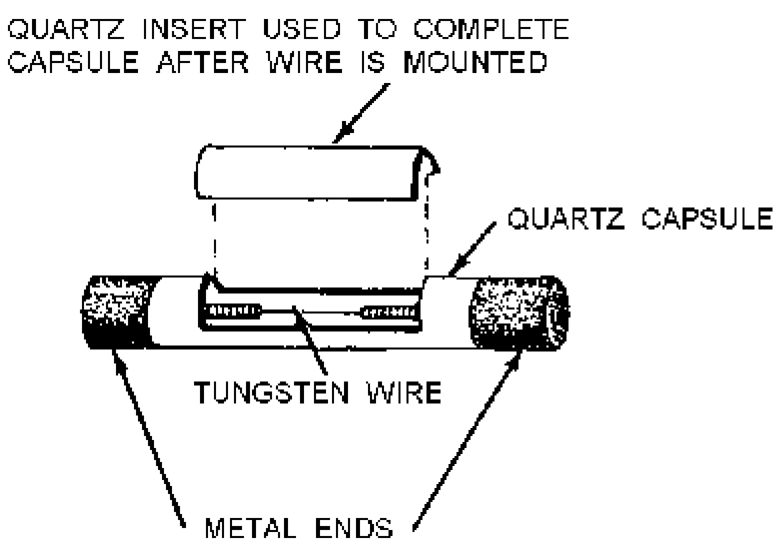


Figure 2-2.—Typical barretter.

THERMISTOR.—A high degree of precision is made possible by the thermistor; therefore, it is widely used. Figure 2-3 shows the typical construction of a bead-type thermistor. The negative-temperature coefficient comes from the use of a semiconductor as the active material. Notice that the active material is shaped in the form of a bead. It is supported between two pigtail leads by connecting wires. The pigtail ends are embedded in the ends of the surrounding glass capsule.

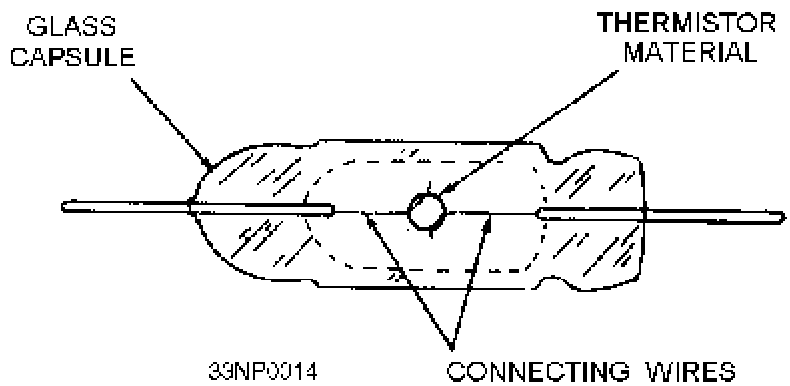


Figure 2-3.—Bead-type thermistor.

The negative-resistance temperature coefficient of thermistors is desirable. This is because excessive power has the effect of changing the resistance of the thermistor to an extent that causes a pronounced rf mismatch. The resulting decrease in power transfer reduces the likelihood of burnout.

Thermistor Bridge

Figure 2-4, views A and B, is an example of a THERMISTOR BRIDGE used for rf power measurements. A thermistor bridge circuit includes other thermistor elements, referred to as *compensating thermistors*. These thermistors respond to fluctuations in ambient temperature so that the bridge balances and calibration are maintained over a wide temperature range. Compensating thermistors are usually in disc form so that they can be mounted on a flat metal surface, such as a chassis or a waveguide. The thermistor bridge in view B is located in the terminating section of a waveguide and contains RT-2, a bead thermistor, and two compensating thermistors, RT-1 and RT-3, on the outside of the waveguide (view A). RA-1 in view B, a calibrated attenuator, controls the amount of rf energy applied to RT-2.

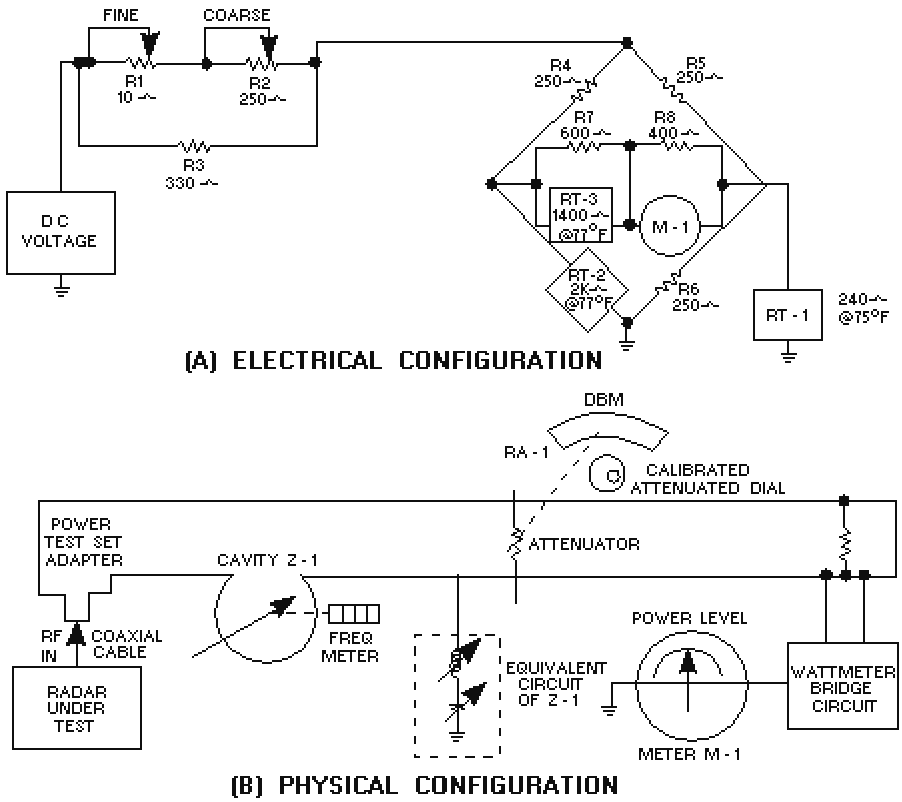


Figure 2-4.—Thermistor bridge.

Before power is applied, R1 and R2 in view A of figure 2-4 are used to adjust the current through RT-2. When the resistance of RT-2 reaches the equivalent parallel resistance of R6 and RT1 (122.4 ohms), the bridge is balanced. Meter M-1 reads 0 at this time. The rf signal being measured is connected to the test set and applied via the calibrated attenuator to RT-2. This causes the temperature of RT-2 to increase, thus reducing its resistance. The bridge becomes unbalanced, causing meter M-1 to deflect an amount proportional to the decrease in resistance of RT-2. Meter M-1, because of the operation of RT-2, reads average power.

Q-6. As the dissipated power increases, what effect does this have on the resistance of a thermistor?

If the ambient temperature rises, the resistance of RT-1 decreases. This shunts more current around the bridge network and allows RT-2 to cool. The resistance of RT-3 decreases, maintaining meter sensitivity independent of temperature changes. Cavity Z-1 in view B of figure 2-4 is an ABSORPTION-TYPE FREQUENCY METER. This type of meter will be discussed later.

FREQUENCY MEASUREMENTS

Frequency measurements are an essential part of preventive and corrective maintenance for electric and electronic equipment. Some examples of the various frequency measurements follow:

- Rotation frequencies of some electro-mechanical devices, such as electric motors, must be determined.

- The output frequency of electric power generators is checked when the engine is started and during preventive maintenance routines.

- Equipment that operates in the af range must be adjusted to operate at the correct frequencies.

- Radio transmitters must be accurately tuned to the assigned frequencies to provide reliable communications and to avoid interference with radio circuits operating on other frequencies.

- Radar sets must be properly tuned to obtain satisfactory performance.

As you can see from the above examples, frequency measurement does indeed play a valuable role in maintenance. These measurements can be divided into two broad categories: MECHANICAL-ROTATION FREQUENCY measurement and ELECTRICAL-OUTPUT FREQUENCY measurement. Depending upon your job and/or the type of command to which you are assigned, you may be tasked with performing one or both of these types of measurements.

MECHANICAL-ROTATION FREQUENCY MEASUREMENT

The rotating frequency (speed in revolutions per minute) of armatures in electric motors and engine-driven generators, as well as the blade speed in turbines, is measured with devices called TACHOMETERS, STROBOSCOPES, and STROBOTACS.

Tachometer

A tachometer is an instrument that measures the rate at which a shaft is turning. Although tachometers are installed on machinery, such as generators and engines, you may need to determine the speed of a rotating machine that is not equipped with a tachometer. In these instances, you will be required to use a PORTABLE TACHOMETER. Portable hand-held tachometers measure speed by direct contact with the shaft of the measured unit. Portable tachometers are for use only during testing and should not be used continuously. The common types of portable tachometers are the CENTRIFUGAL and the CHRONOMETRIC.

CENTRIFUGAL TACHOMETER.—A centrifugal-type tachometer is illustrated in figure 2-5, view A. View B shows the internal arrangement of the centrifugal tachometer; refer to view B in this discussion. In the centrifugal tachometer, centrifugal force acts upon fly weights that are connected by links to upper and lower collars. The upper collar is affixed to a drive shaft; the lower collar is free to move up and down the shaft. A spring, which fits over the shaft, connects the upper and lower collars.

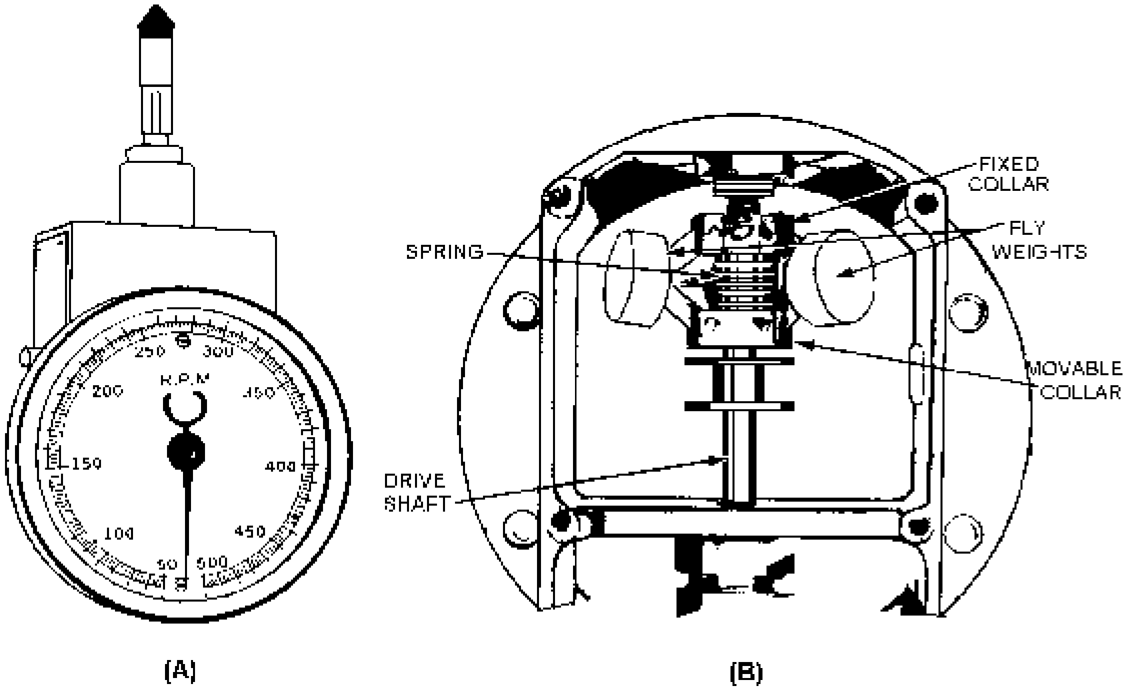


Figure 2-5.—Centrifugal tachometer.

Each portable centrifugal tachometer has a small rubber-covered wheel and a number of hard rubber tips. You fit the appropriate tip or wheel on the end of the tachometer drive shaft, and hold it against the shaft to measure speed of rotation. As the drive shaft begins to rotate, the fly weights rotate with it. Centrifugal force tends to pull the fly weights away from the center, causing the lower collar to rise and compress the spring. The lower collar is attached to a pointer, and its upward motion, restricted by the spring tension, causes an increase in the indication on the dial face.

When properly used, a centrifugal tachometer will indicate correct shaft speed as long as it is in contact with the machine shaft under test. A portable centrifugal tachometer has three ranges: low (50 to 500 rpm), medium (500 to 5,000 rpm), and high (5,000 to 50,000 rpm).

CHRONOMETRIC TACHOMETER.—The chronometric tachometer (figure 2-6) is a combination watch and revolution counter. It measures the average number of revolutions of a shaft per minute. The chronometric tachometer also comes with hard rubber tips, which must be inserted over the drive shaft.

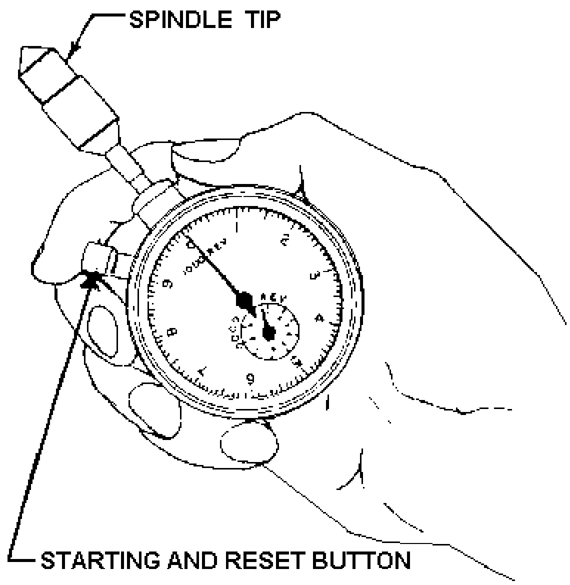


Figure 2-6.—Chronometric tachometer.

When applied to a rotating shaft, the outer drive shaft of this tachometer runs free until a starting button is depressed to start the timing element. In figure 2-6, note the starting button beneath the index finger. The chronometric tachometer retains readings on its dial after its drive shaft has been disengaged from a rotating shaft and until the pointers are returned to 0 by the reset button (usually the starting button). The range of a chronometric tachometer is usually from 0 to 10,000 rpm and from 0 to 3,000 feet per minute (fpm).

Stroboscope

The rotation frequencies of recording devices and teletypewriter motors can be measured by the use of a STROBOSCOPE. The stroboscope is an instrument that allows you to view rotating or reciprocating objects intermittently and produces the optical effect of a slowing down or stopping motion. For example, electric fan blades revolving at 1,800 rpm will appear stationary if you look at them under a light that flashes uniformly 1,800 times per minute. At 1,799 flashes per minute, the blades will appear to rotate forward at 1 rpm; at 1,801 flashes per minute, they will appear to rotate backward at 1 rpm.

When the flashing rate of the light is adjustable, you can calibrate the control in flashes (or revolutions) per minute. The stationary image you see when the rate of the lamp and the rotational rate of a shaft are equal lets you record a very precise speed measurement.

Strobotac

The STROBOTAC (figure 2-7) is an electronic flash device in which the flash duration is very short (a few millionths of a second). (Table 2-1 contains a description of the controls and indicators shown on the strobotac in figure 2-7.) Because of this short flash duration, the strobotac can measure very rapid motion. The box contains a swivel mount with a STROBOTRON LAMP in a reflector, an electronic pulse generator to control the flashing rate, and a power supply that operates from the ac power line. The flashing rate is controlled by the large knob; the corresponding speed (rpm) is indicated on an illuminated dial that is viewed through windows in the knob.

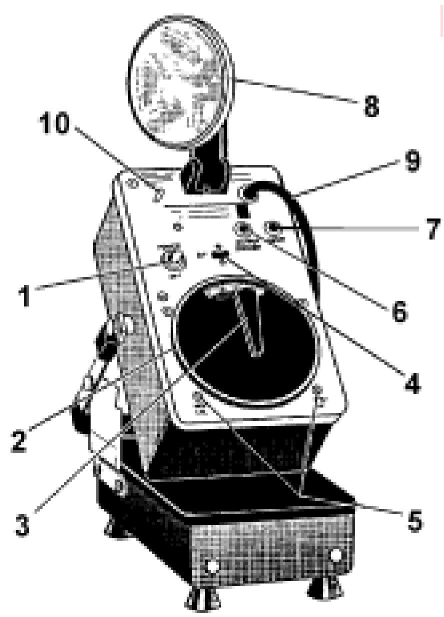


Figure 2-7.—Electronic strobotac.

Table 2-1.—Strobotac Controls and Indicators

	CONTROLS AND INDICATORS (see figure 2-7)	
NO	**NAME**	**USE**
1.	POWER switch	Turns power on and off.
2.	RPM control	Controls the flashing rate of light as the fluted rim is rotated. Dial is calibrated directly in revolutions per minute (rpm).
3.	Range switch	Selects any of three rpm (internal oscillator) ranges, plus three external-input positions: Rpm ranges Intensity External input 110-690 rpm High 700 rpm max 170-4170 rpm Med 4000 rpm max* 4000-25,000 rpm Low 25,000 rpm max
4.	CALibration indicator lamp	Indicates the correct setting of CALibration adjustments for calibrating the RPM dial to power-line frequency.
5.	HIGH CAL, LOW CAL	Calibration adjustments used to calibrate the RPM dial.
6.	OUTPUT TRIGGER jack	A trigger pulse is available at this jack for strobotac types 1531, 1538, stroboslave type 1539, and strobolume type 1532.
7.	INPUT jack	Used for connecting the stroboscope to an external synchronizing signal from the electrical device or mechanical contactor.
8.	Reflector-lamp assembly	Produces and aims the flashing light.
9.	Power cord	A permanently attached 6-foot power cord. For storage, the cord is wound clockwise around the range-switch knob and reflector. The plug is secured by sliding it onto the holding pin.
10.	Holder pin	Used to secure the plug-end of the power cord when unit is to be stored in its case.

* Flashes at 3600 rpm until external signal is plugged in.

The normal speed range is from 110 to 25,000 rpm. At speeds below 600 rpm, "flicker" becomes a problem because the human eye cannot retain successive images long enough to create the illusion of continuous motion. The life of the strobotron lamp is approximately 250 hours if used at flashing speeds of less than 5,000 rpm, or 100 hours if used at higher speeds.

ELECTRICAL OUTPUT FREQUENCY

All alternating voltage sources are generated at a set frequency or range of frequencies. A FREQUENCY METER provides a means of measuring this frequency. The electrical output frequency of ac power generators can be measured by a vibrating reed, a tuned circuit, or by a crossed-coil, iron-vane type meter. The vibrating-reed device is the simplest type of frequency meter. It has the advantage of being rugged enough to be mounted on generator control panels.

A simplified diagram of a vibrating-reed frequency meter is shown in figure 2-8, views A through D. In view A, you can see that the current to be measured flows through the coil and exerts maximum attraction on the soft-iron armature twice during each cycle. The armature is attached to the bar, which is mounted on a flexible support. Reeds of suitable dimensions to have natural vibration frequencies of 110, 112, 114, and so forth, up to 130 hertz are mounted on the bar (view B). The reed with a frequency of 110 hertz is marked 55 hertz; the one with a frequency of 112 hertz is marked 56 hertz; the one with a frequency of 120 hertz is marked 60 hertz, and so forth.

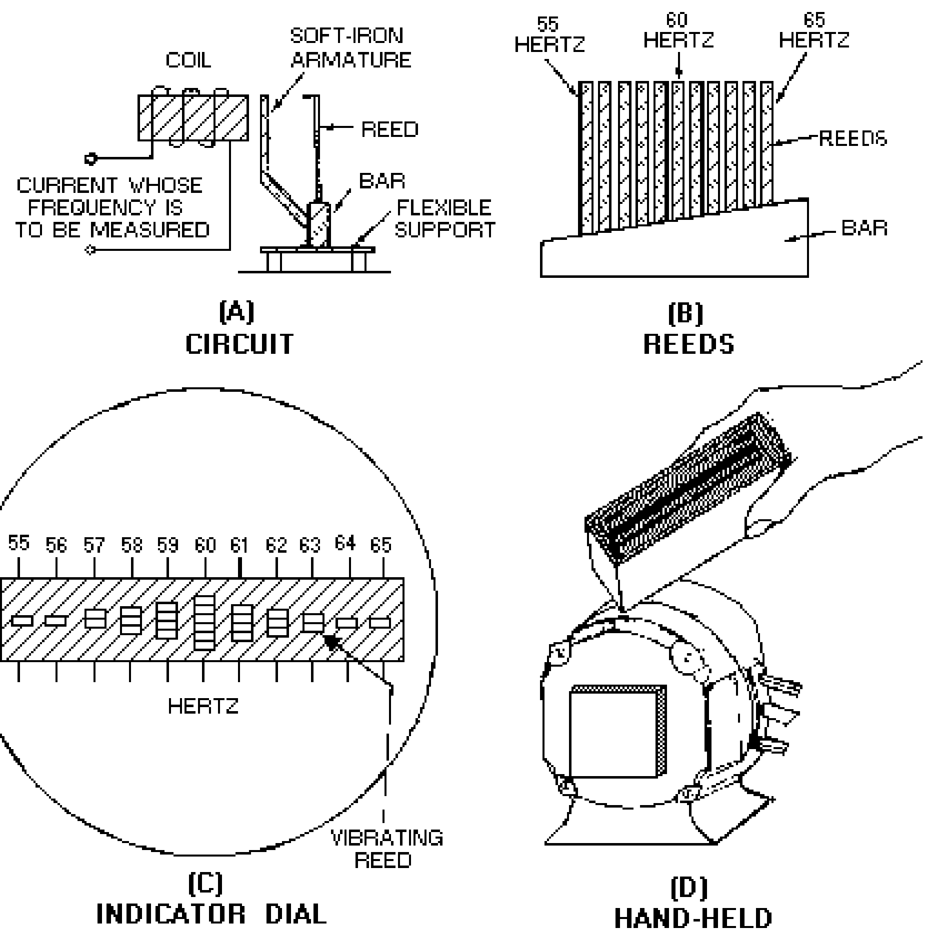


Figure 2-8.—Vibrating-reed frequency meter.

When the coil is energized by a current with a frequency between 55 and 65 hertz, all the reeds are vibrated slightly; but, the reed having a natural frequency closest to that of the energizing current vibrates through a larger amplitude. The frequency is read from the scale value opposite the reed having the greatest amplitude of vibration.

In some instruments, the reeds are the same length; but they are weighted by different amounts at the top so they will have different natural rates of vibration. An end view of the reeds in the indicator is shown in view C. If the energizing current has a frequency of 60 hertz, the reed marked 60 will vibrate the greatest amount, as shown. View D shows a hand-held vibrating-reed frequency meter mounted on the casing of a motor-generator.

Tuned Circuits

TUNED CIRCUITS are used as filters for the passage or rejection of specific frequencies. BANDPASS FILTERS and BAND-REJECT FILTERS are examples of this type. Tuned circuits have certain characteristics that make them ideal for certain types of filters, especially where a high degree of selectivity is desired. A series-tuned circuit offers a low impedance to currents of the particular frequency to which the circuit is tuned and a relatively high impedance to currents of all other frequencies. A parallel-tuned circuit, on the other hand, offers a very high impedance to currents of its natural, or resonant, frequency and a relatively low impedance to others. If you feel you need to review the subject of tuned circuits at this time, refer to NEETS, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*, for more information on these circuits and their applications.

AUDIO FREQUENCIES

Frequency measurements in the af range can be made by the comparison method or the direct-reading frequency meter. Frequency comparisons can be made by the use of a calibrated af generator in conjunction with either an oscilloscope or a modulator and a zero-beat indicating device. Direct-reading frequency measurements can be made by instruments using series, frequency-selective electrical networks, bridge test sets having null indicators, or counting-type frequency meters.

Heterodyne Frequency Meters

Heterodyne frequency meters are available in several varieties. They measure the frequency of the unknown signal by matching the unknown signal with a locally generated signal of the same frequency obtained from a calibrated, precision oscillator. This method is normally referred to as *zero beating*. When a perfect frequency match is obtained, it is indicated by the absence of a beat note (zero beat). The technician generally uses a set of headphones to detect a zero-beat condition in the equipment being tested.

The basic heterodyne meter (figure 2-9) is a calibrated variable oscillator, which heterodynes against the frequency to be measured. Coupling is accomplished between the frequency meter and the output of the equipment under test. (NOTE: This coupling should be in accordance with the step-by-step procedures listed in the technical manual for the frequency meter.) The calibrated oscillator is then tuned so that the difference between the oscillator frequency and the unknown frequency is in the af range. This difference in frequency is known as the BEAT FREQUENCY. As the two frequencies are brought closer to the same value, the tone in the headset will decrease in pitch until it is replaced by a series of rapid clicks. As the process is continued, the clicks will decrease in rapidity until they stop altogether. This is the point of zero beat; that is, the point at which the frequency generated in the oscillator of the frequency meter is equal to the frequency of the unknown signal being measured.

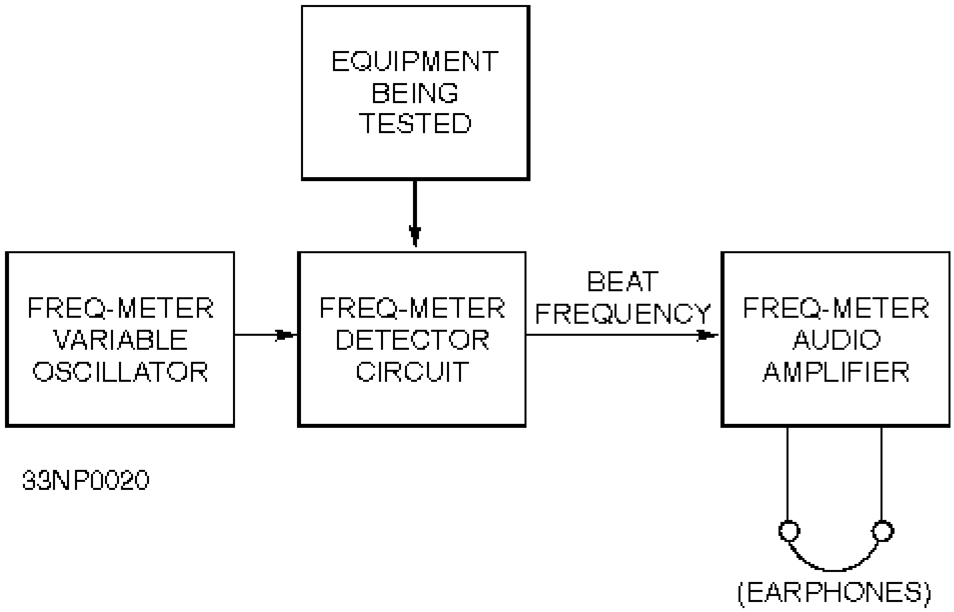


Figure 2-9.—Basic heterodyne meter (block diagram).

Q-7. In a heterodyne-type frequency meter, what is the difference between the oscillator frequency and the unknown frequency?

For all practical purposes, the point of zero beat can be assumed when the clicks are heard at infrequent intervals. Figure 2-10 illustrates the zero-beat concept. Maintaining a condition of absolute silence in the earphones is extremely difficult when you are making this measurement. When the incoming signal is strong, the clicks are sharp and distinct. When the signal is weak, the zero-beat condition is evidenced by a slowly changing "swishing" or "rushing" sound in the headset. After the zero beat is obtained, the dial reading corresponds to the frequency measured.

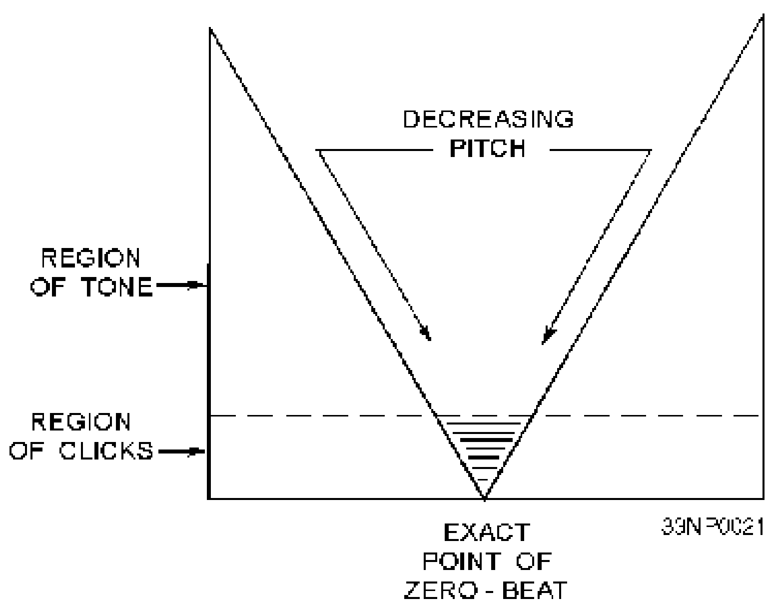


Figure 2-10.—Graph of sound heard in earphone when zero beating.

The manufacturer's calibration book is a very important part of the frequency meter package; in fact, the book is so important that it bears the same serial number as the heterodyne-type frequency meter itself. Contained in this book is a list of the dial settings and the corresponding frequencies produced by that meter at those dial settings. Operating instructions for the meter are also included.

Absorption Wavemeter

WAVEMETERS are calibrated resonant circuits used to measure frequency. The accuracy of wavemeters is not as high as that of heterodyne-type frequency meters; however, they have the advantage of being comparatively simple and can be easily carried.

Q-8. What equipment uses a calibrated resonant circuit to measure frequency?

Any type of resonant circuit can be used in wavemeter applications. The exact kind of circuit used depends on the frequency range for which the meter is intended. Resonant circuits consisting of coils and capacitors are used for VLF through VHF wavemeters.

The simplified illustration of an absorption wavemeter, shown in figure 2-11, consists of a pickup coil, a fixed capacitor, a lamp, a variable capacitor, and a calibrated dial. When the wavemeter's components are at resonance, maximum current flows in the loop, illuminating the lamp to maximum brilliance. The calibrated dial setting is converted to a frequency by means of a chart, or graph, in the instruction manual. If the lamp glows very brightly, the wavemeter should be coupled more loosely to the circuit. For greatest accuracy, the wavemeter should be coupled so that its indicator lamp provides only a faint glow when tuned to the resonant frequency.

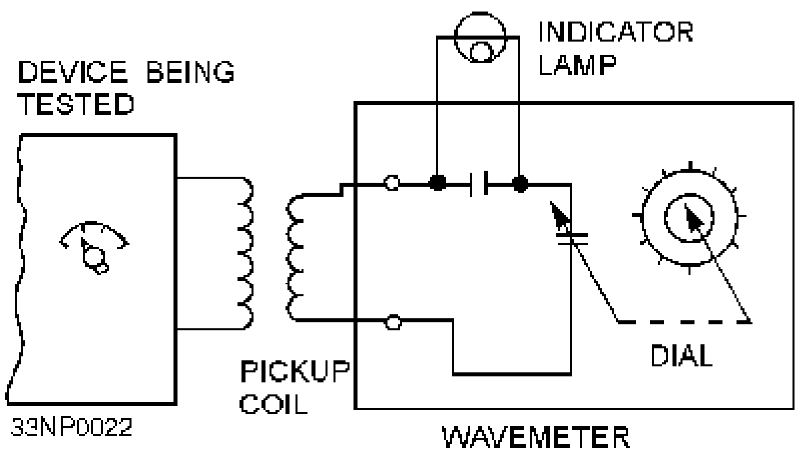


Figure 2-11.—Absorption wavemeter circuit.

FREQUENCIES ABOVE THE AUDIO RANGE

The signal frequencies of radio and radar equipments that operate in the UHF and SHF ranges can be measured by resonant, cavity-type wavemeters or resonant, coaxial-line-type wavemeters. When properly calibrated, resonant-cavity and resonant-coaxial line wavemeters are more accurate and have better stability than wavemeters used for measurements in the LF to VHF ranges. These frequency-measuring instruments are often furnished as part of the equipment. They are also available as general-purpose test sets.

Although many wavemeters are used in performing various functions, the cavity-type wavemeter is the type most commonly used. Only this type is discussed in some detail.

Cavity Wavemeter

Figure 2-12 shows a typical CAVITY WAVEMETER. The wavemeter is of the type commonly used for the measurement of microwave frequencies. The device uses a resonant cavity. The resonant frequency of the cavity is varied by means of a plunger, which is mechanically connected to a micrometer mechanism. Movement of the plunger into the cavity reduces the cavity size and increases the resonant frequency. Conversely, an increase in the size of the cavity (made by withdrawing the plunger) lowers the resonant frequency. The microwave energy from the equipment being tested is fed into the wavemeter through one of two inputs, **A** or **B**. The crystal rectifier then detects (rectifies) the signal. The rectified current is indicated on current meter **M**.

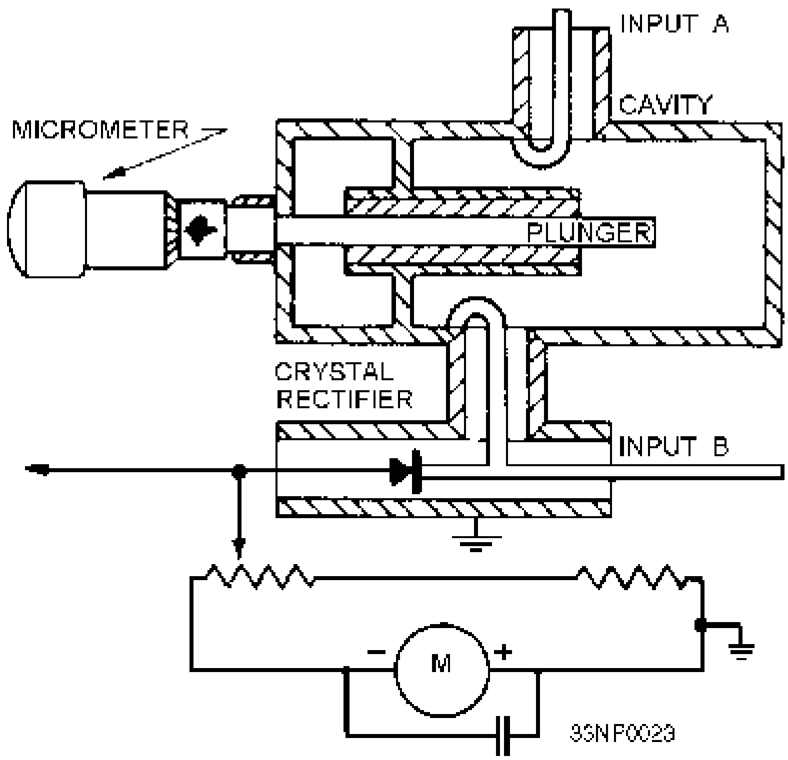


Figure 2-12.—Typical cavity wavemeter.

Electronic Frequency Counters

Another device used to measure frequencies above the audio range is the ELECTRONIC FREQUENCY COUNTER. Since this instrument will be covered in detail in a later chapter, only a brief description is provided at this time.

The electronic frequency counter is a high-speed electronic counter with an accurate, crystal-controlled time base. This combination provides a frequency counter that automatically counts and displays the number of events occurring in a precise time interval. The frequency counter itself does not generate any signal; it merely counts the recurring pulses fed to it.

WAVEFORM ANALYSIS

WAVEFORM ANALYSIS can be made by observing displays of voltage and current variations with respect to time or by harmonic analysis of complex signals. Waveform displays are particularly valuable for adjusting and testing pulse-generating, pulse-forming, and pulse-amplifying circuits. The waveform visual display is also useful for determining signal distortion, phase shift, modulation factor, frequency, and peak-to-peak voltage.

Waveform analysis is used in various electrical and electronic equipment troubleshooting. This section will briefly discuss the oscilloscope and spectrum analyzer to provide you with basic knowledge of this test equipment.

Q-9. Name two instruments used to analyze waveforms.

USE OF THE OSCILLOSCOPE

The CATHODE-RAY OSCILLOSCOPE (CRO or O-SCOPE) is commonly used for the analysis of waveforms generated by electronic equipment. Several types of cathode-ray oscilloscopes are available for making waveform analysis. The oscilloscope required for a particular test is determined by characteristics such as input-frequency response, input impedance, sensitivity, sweep rate, and the methods of sweep control. The SYNCHROSCOPE is an adaptation of the cathode-ray oscilloscope. It features a wide-band amplifier, triggered sweep, and retrace blanking circuits. These circuits are desirable for the analysis of pulse waveforms.

Oscilloscopes are also part of some harmonic analysis test equipments that display harmonic energy levels. To effectively analyze waveform displays, you must know the correct wave shape. The maintenance instructions manual for each piece of equipment illustrates what waveforms you should observe at the various test points throughout the equipment. Waveforms that will be observed at any one selected test point will differ; each waveform will depend on whether the operation of the equipment is normal or abnormal.

The display observed on a cathode-ray oscilloscope is ordinarily one similar to those shown in figure 2-13. Views A and B show the instantaneous voltage of the wave plotted against time. Elapsed time (view A) is indicated by horizontal distance, from left to right, across the etched grid (graticule) placed over the face of the tube. The amplitude (view B) of the wave is measured vertically on the graph.

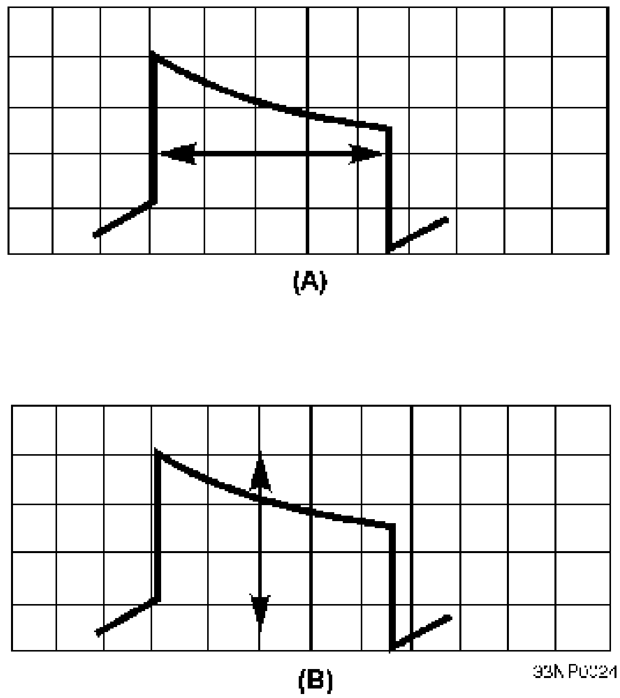


Figure 2-13.—Typical waveform displays.

The oscilloscope is also used to picture changes in quantities other than simply the voltages in electric circuits. For example, if you need to see the changes in waveform of an electric current, you must first send the current through a small resistor. You can then use the oscilloscope to view the voltage wave across the resistor. Other quantities, such as temperatures, pressures, speeds, and accelerations, can be translated into voltages by means of suitable transducers and then viewed on the oscilloscope. A detailed discussion of the oscilloscope is presented in chapter 6 of this module.

USE OF THE SPECTRUM ANALYZER

The SPECTRUM ANALYZER is a device that sweeps over a band of frequencies to determine (1) what frequencies are being produced by a specific circuit under test and (2) the amplitude of each frequency component. To accomplish this, the spectrum analyzer first presents a pattern on a display. Then the relative amplitudes of the various frequencies of the spectrum of the pattern are plotted (see figure 2-14). On the vertical, or Y axis, the **amplitudes** are plotted; on the horizontal, or X axis, the frequencies (time base) are plotted. The overall pattern of this display indicates the proportion of power present at the various frequencies within the SPECTRUM (fundamental frequency with sideband frequencies).

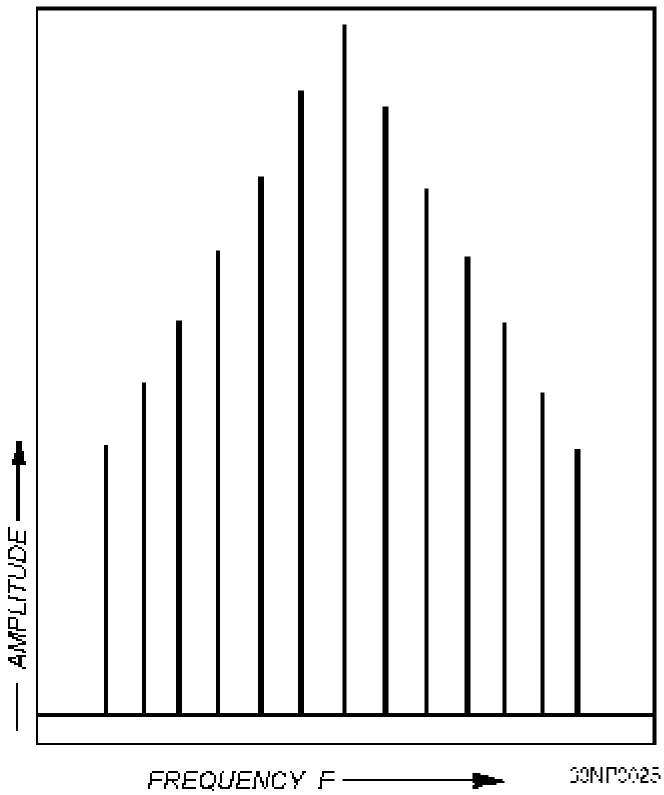


Figure 2-14.—Spectrum analyzer pattern.

Q-10. What device sweeps a band of frequencies to determine frequencies and amplitudes of each frequency component?

The spectrum analyzer is used to examine the frequency spectrum of radar transmissions, local oscillators, test sets, and other equipment operating within its frequency range. Proper interpretation of the displayed frequency spectrum enables you to determine the degree of efficiency of the equipment under test. With experience, you will be able to determine definite areas of malfunctioning components within equipment. In any event, successful spectrum analysis depends on the proper operation of a spectrum analyzer and your ability to correctly interpret the displayed frequencies. Later, in chapter 6, we will discuss the various controls, indicators, and connectors contained on the spectrum analyzer.

TESTING SEMICONDUCTOR DEVICES

Because of the reliability of semiconductor devices, servicing techniques developed for transistorized equipment differ from those normally used for electron-tube circuits. Electron tubes are usually considered to be the circuit component most susceptible to failure and are normally the first components to be tested. Transistors, however, are capable of operating in excess of 30,000 hours at maximum rating without failure. They are often soldered in the circuit in much the same manner as resistors and capacitors. Therefore, they are NOT so quickly removed for testing as tubes.

Substitution of a semiconductor diode or transistor known to be in good condition is one method of determining the quality of a questionable semiconductor device. This method should be used only after you have made voltage and resistance measurements. This ensures the circuit has no defect that might damage the substitute semiconductor device. If more than one defective semiconductor is present in the equipment section where trouble has been localized, the semiconductor replacement method becomes cumbersome. Several semiconductors may have to be replaced before the trouble is corrected. To determine which stage(s) failed and which semiconductors are not defective, you must test all the removed semiconductors. You can do this by observing whether the equipment operates correctly as you reinsert each of the removed semiconductor devices into the equipment.

TESTING DIODES

Semiconductor diodes, such as general-purpose germanium and silicon diodes, power silicon diodes, and microwave silicon diodes, can be tested effectively under actual operating conditions. However, crystal-rectifier testers are available to determine dc characteristics that provide an indication of crystal-diode quality.

A common type of crystal-diode test set is a combination ohmmeter-ammeter. Measurements of forward resistance, back resistance, and reverse current can be made with this equipment. Using the results of these measurements, you can determine the relative condition of these components by comparing their measured values with typical values obtained from test information furnished with the test set or from the manufacturer's data sheets. A check that provides a rough indication of the rectifying property of a diode is the comparison of the back-and-forward resistance of the diode at a specified voltage. A typical back-to-forward-resistance ratio is on the order of 10 to 1, and a forward-resistance value of 50 to 80 ohms is common.

Q-11. What is the typical back-to-forward resistance ratio of a good-quality diode?

Testing Diodes with an Ohmmeter

A convenient test for a semiconductor diode requires only an ohmmeter. The back-and-forward resistance can be measured at a voltage determined by the battery potential of the ohmmeter and the resistance range at which the meter is set. When the test leads of the ohmmeter are connected to the diode, a resistance will be measured that is different from the resistance indicated if the leads are reversed. The smaller value is called the FORWARD RESISTANCE, and the larger value is called the BACK RESISTANCE. If the ratio of back-to-forward resistance is greater than 10 to 1, the diode should be capable of functioning as a rectifier. However, keep in mind that this is a very limited test that does not take into account the action of the diode at voltages of different magnitudes and frequencies. (**NOTE**: This test should never be used to test crystal mixer diodes in radars. It will destroy their sensitivity.)

Testing Diodes with Oscilloscopes

An oscilloscope can be used to graphically display the back-and-forward resistance characteristics of a crystal diode. A circuit used in conjunction with an oscilloscope to make this test is shown in figure 2-15. This circuit uses the oscilloscope line-test voltage as the test signal. A series circuit (composed of resistor R1 and the internal resistance in the line-test circuit) decreases a 3-volt, open-circuit test voltage to a value of approximately 2 volts peak to peak.

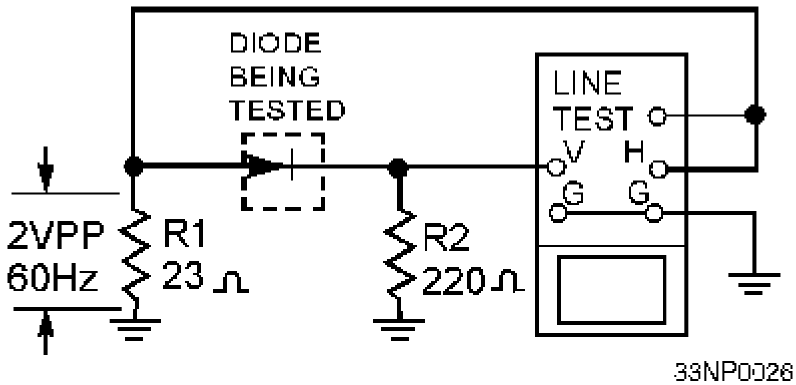


Figure 2-15.—Testing semiconductor diodes with an oscilloscope.

The test signal applied to the crystal diode is also connected to the horizontal input of the oscilloscope. The horizontal sweep represents the voltage applied to the diode under test. The voltage developed across current-measuring resistor R2 is applied to the vertical input of the oscilloscope. Because this voltage is proportional to the current through the diode being tested, the vertical deflection will indicate crystal current. The resulting oscilloscope trace for a normal diode is similar to the curve shown in figure 2-16.

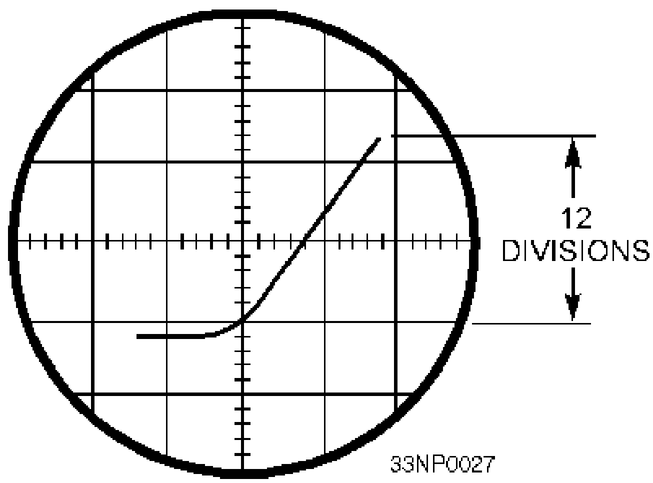


Figure 2-16.—Characteristic curve of a semiconductor diode.

To test Zener diodes, you must use a higher voltage than the oscilloscope line-test signal. This test can be made with a diode test set or with the circuit shown in figure 2-17. In this circuit, rheostat R1 is used to adjust the input voltage to a suitable value for the Zener diode being tested. Resistor R2 limits the current through the diode. The signal voltage applied to the diode is also connected to the horizontal input of the oscilloscope. The voltage developed across current-measuring resistor R3 is applied to the vertical input of the oscilloscope. The horizontal sweep represents the applied voltage, and the vertical deflection indicates the current through the diode being tested. Figure 2-18 shows the characteristic pattern of a Zener diode. Note the sharp increase in current at the Zener voltage (avalanche) point. For the Zener diode to be usable, this voltage must be within limits specified by the manufacturer.

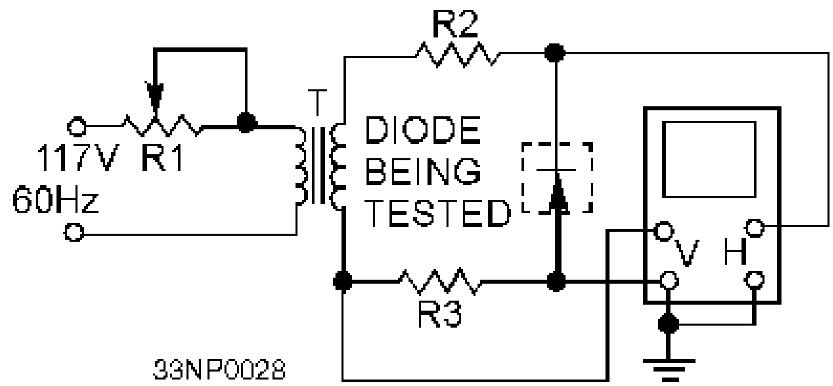


Figure 2-17.—Testing a Zener diode.

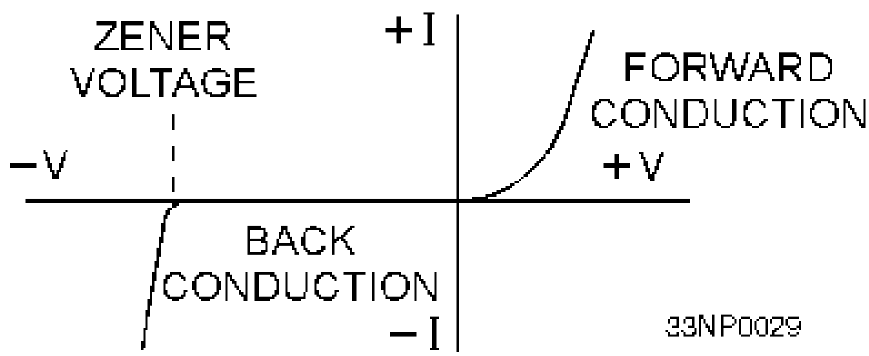


Figure 2-18.—Zener diode characteristic curve.

TESTING TRANSISTORS

Most transistorized equipments use printed circuit boards on which components are neatly arranged. This arrangement makes the transistors and other components easy to reach while you are troubleshooting and servicing the equipment. While investigating with test probes, however, you must be careful to prevent damage to the printed wiring.

One of the outstanding advantages of transistors is their reliability. Tube failures account for over 90 percent of the failures in electron-tube equipments. Transistors, however, are long lived. This factor, among others, decreases maintenance required to keep transistorized equipment operating. The techniques used in testing transistorized equipment are similar to those for maintaining electron-tube circuits. Basically, these techniques include several checks and inspections.

Power Supply Checks

When using test equipment to localize a trouble, you should check the power supply to see that its output voltages are present and of the correct values. Improper power supply voltages can cause odd effects. You will prevent many headaches by checking the power supply first.

Visual Inspection

Visual inspection is a good maintenance technique. Occasionally, you will find loose wires or faulty connections, making extensive voltage checks unnecessary.

Transistor Checks

Transistors can be checked by substitution. Transistors, however, have a characteristic known as *leakage current*, which may affect the results obtained when the substitution method is used.

The leakage current may influence the current gain or amplification factor of the transistor. Therefore, a particular transistor might operate properly in one circuit and not in another. This characteristic is more critical in certain applications than in others. As the transistor ages, the amount of leakage current tends to increase. One type of transistor checker used is the semiconductor test set. This test set can be used either for in-circuit or out-of-circuit tests or for collector leakage current or current gain. You should use extreme care when substituting transistors. More and more transistors have specific current and breakdown voltage requirements that may affect how they operate within a given circuit.

Q-12. As a transistor ages, what happens to the leakage current?

Voltage Checks

Voltage measurements provide a means of checking circuit conditions in a transistorized circuit just as they do in checking conditions in a tube circuit. The voltages, however, are much lower than in a tube circuit. The bias voltage between the base and emitter, for instance, is usually 0.05 to 0.20 volts. When making checks, observe polarity.

Resistance Checks

Transistors have little tendency to burn or change value because of low voltage in their circuits. They can, however, be permanently damaged by high-voltage conditions that occur when the collector voltage is increased. They can also be permanently damaged when the ambient temperature increases and causes excessive collector current flow. Transistors are easily damaged by high current; therefore, resistance measurements must not be taken with an ohmmeter that provides a maximum current output in excess of 1 milliampere. If you are not sure that the range of ohmmeter you want to use is below the 1 milliampere level, connect the ohmmeter to a milliammeter and check it. See figure 2-19 for a method of measuring the current from an ohmmeter.

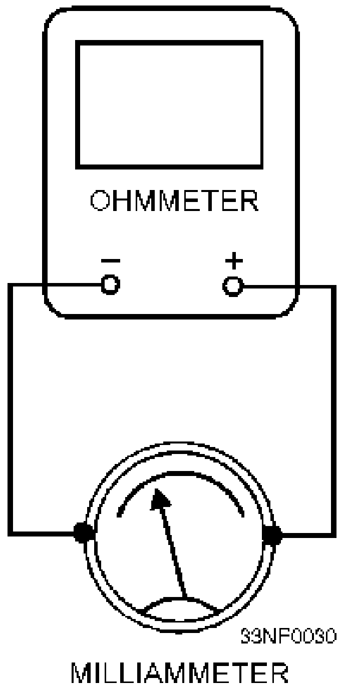


Figure 2-19.—Measuring current passed by an ohmmeter.

Resistance measurements usually are not made in transistorized circuits, except when you are checking for open windings in transformers and coils. When a resistance check is required, the transistors are usually removed from the circuit. Resistance checks cannot test all the characteristics of transistors, especially transistors designed for high frequencies or fast switching. The ohmmeter is capable of making simple transistor tests, such as open and short tests.

Refer to NEETS, Module 7, *Introduction to Solid-State Devices and Power Supplies*, for a review of transistor and semiconductor terms and theory.

SUMMARY

The important points of this chapter are summarized in the following paragraphs:

The **BEL** is a unit that expresses the logarithmic ratio between the input and output of any given component, circuit, or system and can be expressed in terms of voltage, current, or power.

Any figure expressed in bels can be converted to DECIBELS by multiplying the figure by 10. The decibel cannot be used to represent actual power, only a ratio of one power to another.

The abbreviation dBm is used to represent power levels above or below a 1 milliwatt reference level.

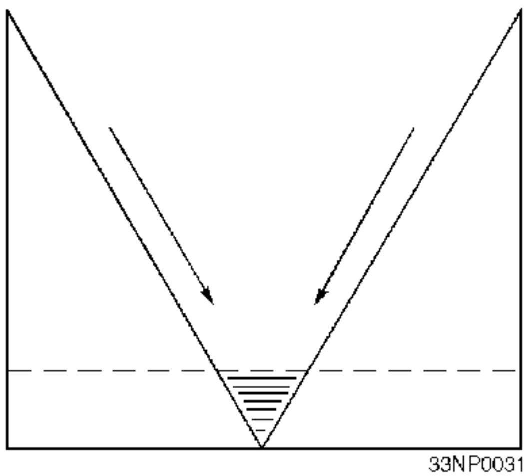
A **BOLOMETER** is a device that undergoes changes in resistance as changes in dissipated power occur. The two types of bolometers most often used are the barretter and the thermistor.

FREQUENCY MEASUREMENTS can be divided into two broad categories: mechanical-rotation frequency and electrical-output frequency measurements.

MECHANICAL ROTATION frequency is measured using a device called a TACHOMETER. Three basic tachometers are used for measuring mechanical rotation frequency—the CENTRIFUGAL tachometer, the CHRONOMETRIC tachometer, and the STROBOSCOPIC tachometer.

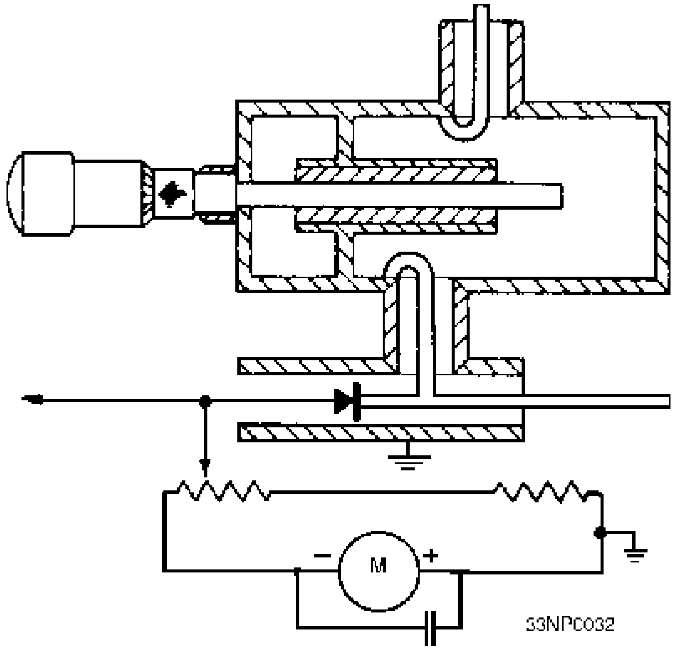
ELECTRICAL-OUTPUT frequencies of ac generators can be measured by VIBRATING-REED devices or TUNED CIRCUITS.

AUDIO FREQUENCIES can be measured by a process known as ZERO BEATING. This is done by matching an unknown signal with a locally generated signal of the same frequency obtained from a calibrated high-precision oscillator. As the two frequencies are brought closer to the same value, they reach a point of zero beat. This is when the frequency generated in the oscillator is equal to the frequency of the unknown signal being measured. Another term for zero beating is HETERODYNING.



WAVEMETERS are calibrated resonant circuits used to measure frequency. Any type of resonant circuit can be used in wavemeter applications. The type used depends on the frequency range for which the meter is intended.

For measuring frequencies in the microwave range, the CAVITY WAVEMETER is the type most commonly used.



The **CATHODE-RAY OSCILLOSCOPE** and the SPECTRUM ANALYZER are used to perform WAVEFORM ANALYSIS.

A typical **BACK-TO-FORWARD-RESISTANCE** ratio for a diode is 10 to 1.

ANSWERS TO QUESTIONS Q1. THROUGH Q12.

A-1. Bel.

A-2. dBm.

A-3. 4 mW.

A-4. Dummy load or dummy antenna.

A-5. Barretter and thermistor.

A-6. It increases.

A-7. Beat frequency.

A-8. Wavemeter.

A-9. Oscilloscope and spectrum analyzer.

A-10. Spectrum analyzer.

A-11. 10-to-1 ratio.

A-12. It tends to increase.

CHAPTER 3

BASIC METERS

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to:

1. Describe the basic theory of the galvanometer.
2. Describe the basic theory of the D'Arsonval meter movement.
3. State the proper procedure for connecting an ammeter to a circuit.
4. Define ammeter sensitivity.
5. State the proper procedure for connecting a voltmeter to a circuit.
6. Describe possible effects on a circuit caused by the connection of a voltmeter.
7. Define voltmeter sensitivity.
8. Describe the internal operation of an ohmmeter with the use of a block diagram.
9. Describe the operating procedure for using a megohmmeter.
10. Describe the use of the electrodynamometer-type meter as a voltmeter, ammeter, and wattmeter.
11. Describe the factors that limit wattmeter capability.
12. Describe an open circuit, a ground, a short, and the tests used to check for these conditions.

INTRODUCTION

When troubleshooting, testing, or repairing electronic equipment, you will use various meters and other types of test equipment to check for proper circuit voltages, currents, resistances, and to determine if the wiring is defective. You may be able to connect these test instruments to a circuit and take readings without knowing just *how* the instruments operate. However, to be a competent technician, you need to be able to do more than merely read a test instrument. You need a basic knowledge of how test instruments operate. This chapter discusses the operating principles of some of the test instruments you will use in equipment troubleshooting.

METERS

The best and most expensive measuring instrument is of no use to you unless you know what you are measuring and what each reading indicates. Remember that the purpose of a meter is to measure quantities existing within a circuit. For this reason, when the meter is connected to the circuit, it must not change the condition of the circuit.

METER POWER SOURCE

Meters are either SELF-EXCITED or EXTERNALLY EXCITED. Self-excited meters operate from their own power sources. Externally excited meters get their power from the circuit to which they are connected. Most common meters (voltmeters, ammeters, and ohmmeters) that you use in your work operate on the electromagnetic principle. All measuring instruments must have some form of indicating device, usually a meter, to be of any use to you. The most basic indicating device used in instruments that measure current and voltage operates by using the interaction between the magnetic fields associated with current flow in the circuit. Before continuing, you might want to review the properties of magnetism and electromagnetism in NEETS, Module 1, *Introduction to Matter, Energy, and Direct Current*.

Q-1. What meters operate from their own power sources?

BASIC METER MOVEMENT

A stationary, permanent-magnet, moving-coil meter is the basic meter movement used in most measuring instruments used for servicing electrical equipment. When current flows through the coil, a resulting magnetic field reacts with the magnetic field of the permanent magnet and causes the movable coil to rotate. The greater the intensity of current flow through the coil, the stronger the magnetic field produced; the stronger the magnetic field produced, the greater the rotation of the coil. The GALVANOMETER is an example of one type of stationary, permanent-magnet, moving-coil measuring instrument.

Galvanometer

The galvanometer is used to measure very low currents, such as those in bridge circuits. In modified form, the galvanometer has the highest sensitivity of any of the various types of meters in use today. A simplified diagram of a galvanometer is shown in figure 3-1. It is different from other instruments used for the same purpose because its movable coil is suspended by means of metal ribbons instead of a shaft and jewel-bearing arrangement often used in other instruments.

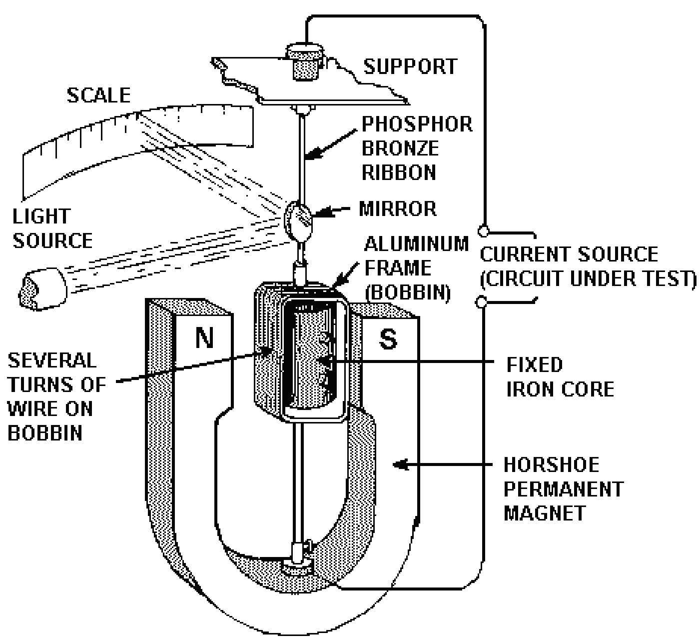


Figure 3-1.—Simplified galvanometer.

The movable coil is wrapped around the aluminum frame of the galvanometer. The coil is suspended between the poles of the magnet by means of thin, flat ribbons of phosphor bronze. These ribbons provide a conduction path for the current between the circuit being tested and the movable coil. The ribbons allow the coil to twist in response to the interaction of the applied current through the coil and the magnetic field of the permanent magnet. They also provide the restoring force for the coil. Basically, the restoring force is that force necessary to return the movable frame to its resting position after a reading. The ribbons restrain or provide a counterforce to the magnetic force acting on the coil. When the driving force of the coil current is removed, the restoring force provided by the ribbons returns the coil to its zero position.

Q-2. What physical component of a galvanometer provides the restoring force for the coil?

To determine the amount of current flow, we must have a means to indicate the amount of coil rotation. Either of two methods may be used: (1) the **POINTER** arrangement or (2) the **LIGHT AND MIRROR** arrangement.

Q-3. In a galvanometer, what two methods are used to indicate the amount of coil rotation?

In the pointer arrangement, one end of the pointer is mechanically connected to the rotating coil; as the coil moves, the pointer also moves. The other end of the pointer moves across a graduated scale and indicates the amount of current flow. The overall simplicity of this arrangement is its main advantage. However, a disadvantage of this arrangement is that it introduces a mechanical coil balancing problem, especially if the pointer is long.

Q-4. What is the primary disadvantage of the pointer arrangement for indicating coil rotation?

In the light and mirror arrangement, the use of a mirror and a beam of light simplifies the problem of coil balance. When this arrangement is used to measure the turning of the coil, a small mirror is mounted on the supporting ribbon, as shown in figure 3-1. An internal light source is directed to the mirror and then reflected to the scale of the meter. As the movable coil turns, so does the mirror. This causes the light reflection to move across the graduated scale of the meter. The movement of the reflection is proportional to the movement of the coil; therefore, the intensity of the current being measured by the meter is accurately indicated.

If the beam of light and mirror arrangement is used, the beam of light is swept to the right or left across a translucent screen (scale). The translucent screen is divided uniformly with the zero reading located at center scale. If the pointer arrangement is used, the pointer is moved in a horizontal plane to the right or left across a scale that is divided uniformly with the zero reading at the center. The direction in which the beam of light or the pointer moves depends on the direction (polarity) of current through the coil.

D'Arsonval Meter Movement

Most dc instruments use meters based on some form of the D'Arsonval meter movement. In D'Arsonval-type meters, the length of the conductor and the strength of the field between the poles of the magnet are fixed. Therefore, any change in current causes a proportional change in the force acting on the coil. Figure 3-2 is a simplified diagram showing the principle of the D'Arsonval movement.

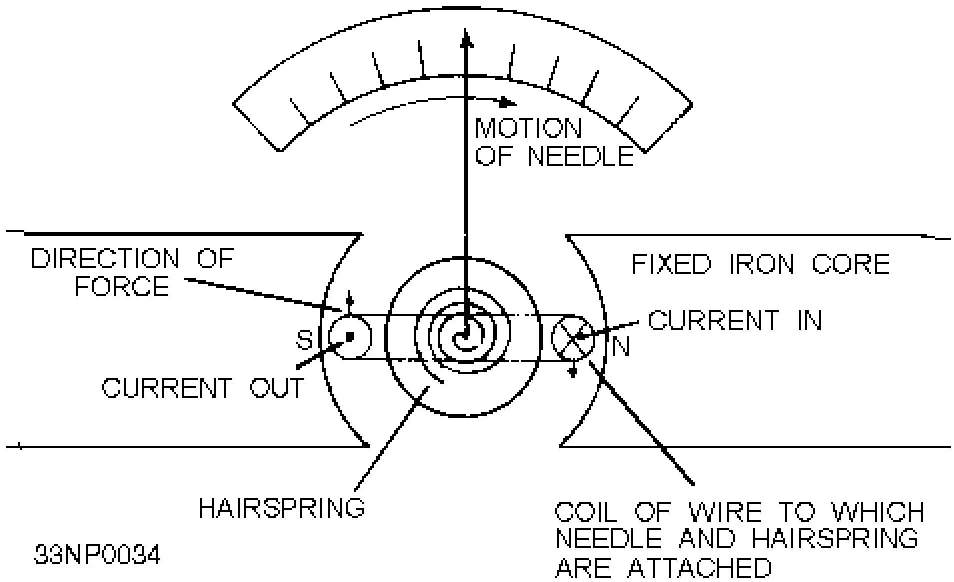


Figure 3-2.—D'Arsonval meter movement.

In the figure, only one turn of wire is shown; however, in an actual meter movement, many turns of fine wire would be used, each turn adding more effective length to the coil. The coil is wound on an aluminum frame (bobbin) to which the pointer is attached. Oppositely wound hairsprings (only one is shown in the figure) are also attached to the bobbin, one at either end. The circuit to the coil is completed through the hairsprings. In addition to serving as conductors, the hairsprings serve as the restoring force that returns the pointer to the zero position when no current flows.

Q-5. What component of the D'Arsonval meter movement completes the circuit for current flow to the coil?

COIL MOVEMENT.—As we discussed previously, the deflecting (moving) force on the coil is proportional to the current flowing through the coil. This deflecting force tends to cause the coil to rotate against the restraining force of the hairsprings. When the deflecting force and the restraining force are equal, the coil and the pointer stop moving. As we have just stated, the deflecting force is proportional to the current in the coil, the angle (amount) of rotation is proportional to the deflecting force; therefore, the angle of rotation is proportional to the current through the coil. When current stops flowing through the coil, the deflecting force stops, and the restoring force of the springs returns the pointer to the zero position.

Q-6. What component supplies restoring force to the coil of the D'Arsonval meter movement?

DIRECTION OF FORCE.—The current through the single turn of wire is in the direction indicated in the figure (away from you on the right-hand side and toward you on the left-hand side). If we apply the right-hand motor rule, the direction of force is upward on the left-hand side and downward on the right-hand side; therefore, the direction of motion of the coil and pointer is clockwise. If the current were reversed in the wire, the direction of motion of the coil and pointer would be reversed. For a review of the right-hand rule for motors, refer to NEETS, Module 5, *Introduction to Generators and Motors*.

PRINCIPLE OF OPERATION.—A more detailed view of the basic D'Arsonval movement, as it is used in ammeters and voltmeters, is shown in figure 3-3. The principle of operation is the same as that discussed in the simplified version. The iron core is rigidly supported between the pole pieces; it serves to concentrate the flux in the narrow space between the iron core and the pole piece. Current flows into one hairspring, through the coil, and out the other hairspring. The restoring forces of the spiral springs return the pointer to the normal zero position when the current through the coil is interrupted. Conductors connect the hairsprings with the outside terminals of the meter. If the instrument is not DAMPED to absorb the energy of the moving element, the pointer will oscillate (vibrate) for a period of time before coming to a stop in its final position. Damping is an energy-absorbing system that prevents this.

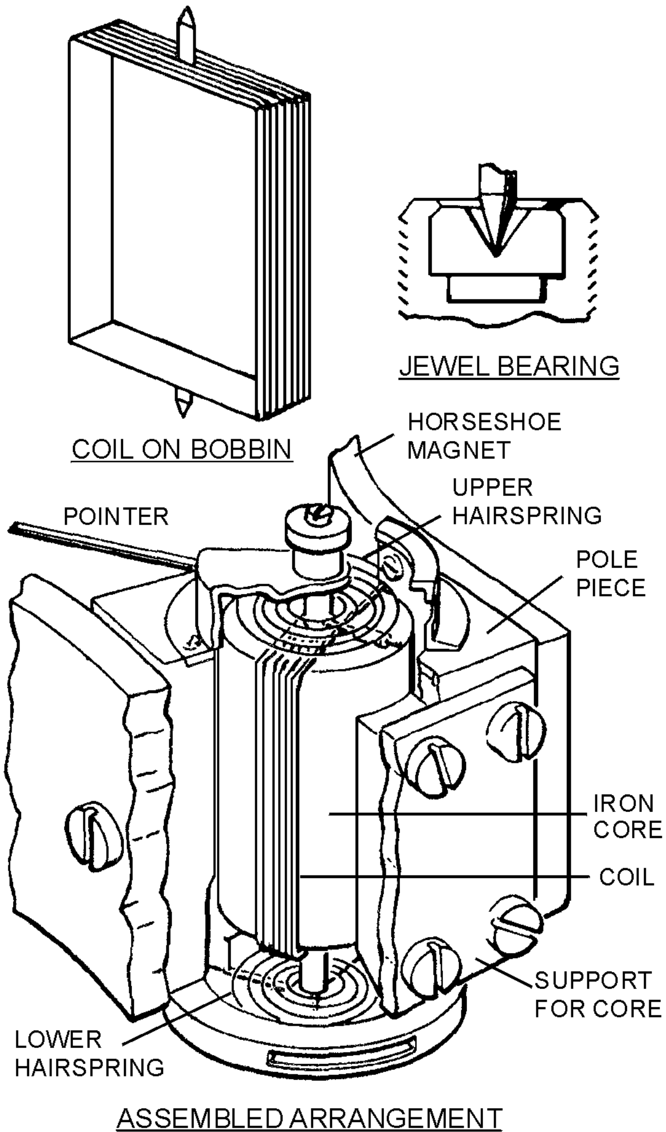


Figure 3-3.—Detailed view of the basic D'Arsonval meter movement.

DAMPING.—This is accomplished in many D'Arsonval movements by means of the motion of the aluminum bobbin on which the coil is wound. As the bobbin rotates in the magnetic field, an electromotive force is induced into it as it cuts through the lines of force. Induced currents flow in the bobbin in a direction opposite to the motion; this causes the bobbin to go beyond its final position only once before stopping. The overall sensitivity of the meter can be increased by the use of a lightweight rotating assembly (bobbin, coil, and pointer) and by the use of jewel bearings, as shown in figure 3-3.

POLE CONSTRUCTION.—Note that the pole pieces in figures 3-2 and 3-3 have curved faces. You can see the advantage of this type of construction if you remember that lines of force enter and leave a magnetic field in the air gap at right angles to the coil, regardless of the angular position of the coil. Because of this type of construction, a more linear scale is possible than if the pole faces were flat.

Q-7. What advantage is gained by using pole pieces with curved faces in the D'Arsonval meter movement?

DC AMMETER

The movable coil of the D'Arsonval meter movement we have been discussing up to now uses small-size wire in its windings. This small-size wire places limits on the amount of current that can be safely passed through the coil. Therefore, the basic D'Arsonval movement discussed can be used to indicate or measure only very small currents. Certain circuit changes must be made to the basic D'Arsonval meter movement for it to be practical in everyday use. To measure large currents, you must use a SHUNT with the meter.

Shunts

A shunt is a physically large, low-resistance conductor connected in parallel (shunt) with the meter terminals. It is used to carry the majority of the load current. Such a shunt is designed with the correct amount of resistance so that only a small portion of the total current flows through the meter coil. The meter current is proportional to the total load current. If the shunt is of such a value that the meter is calibrated in milliamperes, the instrument is called a MILLIAMMETER. If the shunt has such a value that the meter must be calibrated in terms of amperes, it is called an AMMETER.

Q-8. What structurally large, low-resistance conductor is connected in parallel with the meter movement to prevent damage?

SHUNT RESISTANCE.—A single, standardized meter movement is normally used in all ammeters, no matter what the range is for a particular meter. For example, meters with working ranges of 0 to 10 amperes, 0 to 5 amperes, or 0 to 1 ampere all use the same meter movement. The various ranges are achieved through the use of different values of shunt resistance with the same meter movement. The designer of the ammeter simply calculates the correct shunt resistance required to extend the range of the meter movement to measure any desired value of current. This shunt is then connected across the meter terminals. Shunts may be located inside the meter case (internal shunts) with the proper switching arrangements for changing them. They may also be located outside the meter case (external shunts) with the necessary leads to connect them to the meter.

EXTERNAL SHUNTS.—An external-shunt circuit is shown in figure 3-4, view A. Typical external shunts are shown in view B. View C shows a meter movement mounted within the case. The case provides protection against breakage, magnetic shielding in some cases, and portability.

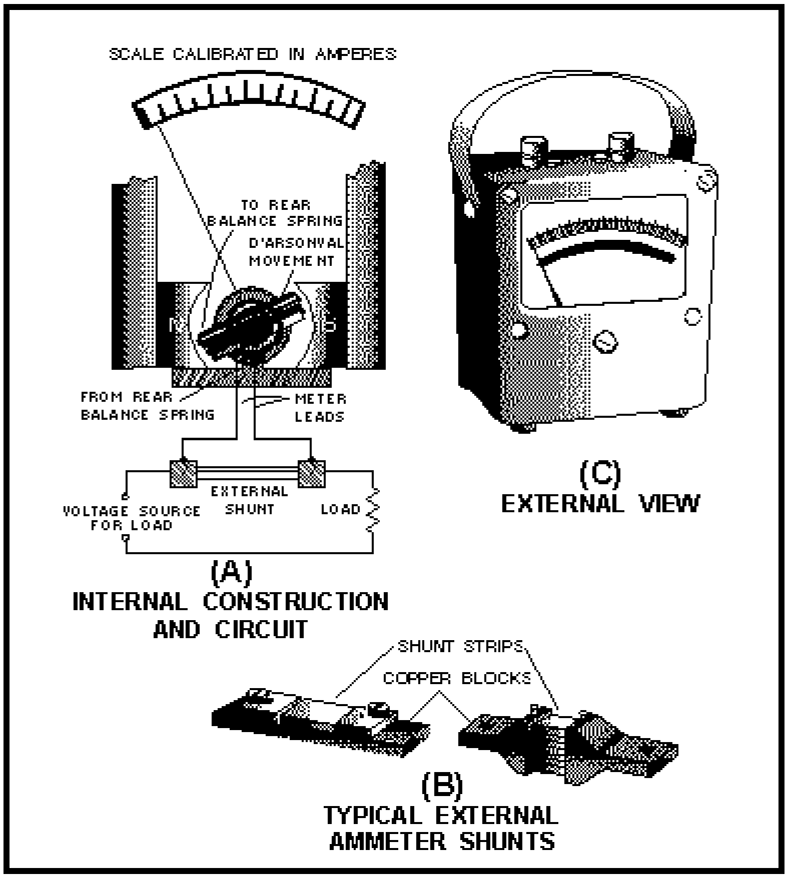


Figure 3-4.—Dc ammeter using the D'Arsonval movement with external shunts.

SHUNT CONSTRUCTION.—The shunt strips (view B of figure 3-4) are usually made of the alloy *Manganin*. Manganin has a temperature coefficient of almost zero. The zero-temperature coefficient property is desirable because of the heavy currents that often flow through shunts producing heat. A zero-temperature coefficient material is not affected by this heat; therefore, it remains stable in temperature. Most other materials increase their resistance as they are heated. If shunts were made of these materials, they would carry less current. More and more current would flow through the meter movement, and the chances of damage would increase. Using shunts constructed with zero-temperature coefficient materials eliminates this problem.

Q-9. What type of temperature coefficient material does not produce increased heat in response to increased current flow?

The ends of the shunt strips are embedded in heavy copper blocks. The blocks are attached to the meter coil leads and the line terminals. To ensure accurate readings, you should not interchangeably use the meter leads for a particular ammeter with those for a meter of a different range. Slight changes in lead length and size may vary the resistance of the meter circuit. If this happens, current will also change and cause incorrect meter readings. External shunts are generally used where currents greater than 50 amperes must be measured.

SHUNT SELECTION.—When using an external-shunt ammeter, you should select a suitable shunt so that the scale deflection can be easily read. For example, if the scale has 150 divisions and the load current you want to measure is known to be between 50 and 100 amperes, a 150-ampere shunt would be the correct choice. Under these conditions, each division of the scale represents 1 ampere. In other words, a full-scale deflection of the pointer would rest on the 150th division mark, indicating that 150 amperes of load current is flowing. At half-scale deflection, the pointer would rest on the 75th division mark, indicating that 75 amperes of load current is flowing.

A shunt having exactly the same current rating as the expected normal load current should **never** be selected. If you were to select such a shunt, higher than normal load currents could possibly drive the pointer off scale and damage the meter movement. A good choice of shunt values will place the indicating needle somewhere near the midscale indication when the load current you are reading is normal. For example, assume that the meter scale is divided into 100 equal divisions and you want to measure a current of 60 amperes. The shunt to use would be a 100-ampere shunt. This would make each division of the scale equal to 1 ampere. The meter indication would fall on the 60th division showing that 60 amperes of load current is flowing. Therefore, an allowance (40 amperes) remains for unexpected surge currents.

Q-10. A good choice of shunt resistance will place the indicating pointer near what part of the meter scale with a normal load?

INTERNAL SHUNTS FOR METERS IN THE 0- TO 50-AMPERE RANGE.—When measuring current ranges below 50 amperes, you will most often use internal shunts (R_{shunt}). In this way, you can easily change the range of the meter by means of a switching arrangement. A switch will select the correct internal shunt with the necessary current rating and resistance. Before you can calculate the required resistance of the shunt for each range, the total resistance of the meter movement must be known. For example, suppose you desire to use a 100-microampere D'Arsonval meter with an internal coil resistance of 100 ohms to measure line currents up to 1 ampere. The meter will deflect to its full-scale position when the current through the deflection coil is 100 microamperes.

Since the coil resistance is 100 ohms, you can calculate the coil's voltage (E_{coil}) by using Ohm's law, as follows:

$$\begin{aligned} E_{coil} &= I \times R_{coil} \\ &= 0.0001 \text{ amperes} \times 100 \text{ ohms} \\ &= 0.01 \text{ volt} \end{aligned}$$

When the pointer is deflected to full scale, 100 microamperes of current flows through the coil and 0.01 volt drops across it. Remember, 100 microamperes is the maximum safe current for this meter movement. Exceeding this value will damage the meter. The shunt must carry any additional load current.

The meter coil has a 0.01 volt drop across it, and, because the shunt and coil are in parallel, the shunt also has a voltage drop of 0.01 volt. The current that flows through the shunt is the difference between the full-scale meter current and the line current being fed into the shunt. In this case, meter current is 100 microamperes. Full-scale deflection is desired only when the total current is 1 ampere. Therefore, the shunt current must equal 1 ampere minus 100 microamperes, or 0.9999 ampere. Ohm's law is again used to provide the approximate value of required shunt resistance (R_{shunt}), as follows:

$$\begin{aligned} R_{shunt} &= \frac{E}{I} \\ &= \frac{0.01 \text{ volt}}{0.999 \text{ ampere}} \\ &= 0.01 \text{ ohm} \end{aligned}$$

To increase the range of the 100-microampere meter to 1 ampere (full-scale deflection), place a 0.01-ohm shunt in parallel with the meter movement.

You can convert the 100-microampere instrument to a 10-ampere meter by using a proper shunt. The voltage drop for a full-scale deflection is still 0.01 volt across the coil and the shunt. The meter current is still 100 microamperes. The shunt current must therefore be 9.9999 amperes under full-scale deflection. Again, this is an approximate figure found by the application of Ohm's law.

You can also convert the same instrument to a 50-ampere meter by using the proper shunt resistance, as follows:

$$\begin{aligned} R_{shunt} &= \frac{E}{I} \\ &= \frac{0.01 \text{ volt}}{49.999 \text{ amperes}} \\ &= 0.0002 \text{ ohm} \end{aligned}$$

INTERNAL SHUNTS FOR METERS IN THE MILLIAMPERE RANGE.—The above method of computing the shunt resistance is satisfactory in most cases; however, it can only be used when the line current is in the ampere range and the meter current is relatively small compared to the load current. In such cases, you can use an approximate value of resistance for the shunt, as was done above. However, when the line current is in the milliampere range and the coil current becomes an appreciable percentage of the line current, a more accurate calculation must be made. For example, suppose you desire to use a meter movement that has a full-scale deflection of 1 milliampere and a coil resistance of 50 ohms to measure currents up to 10 milliamperes. Using Ohm's law, you can figure the voltage (E_{coil}) across the meter coil (and the shunt) at full-scale deflection, as follows:

$$\begin{aligned} E_{coil} &= I \times R \\ &= 0.001 \text{ ampere} \times 50 \text{ ohms} \\ &= 50 \text{ millivolts} \end{aligned}$$

The current that flows through the shunt (I_{shunt}) is the difference between the line current and the meter current, as figured below:

$$\begin{aligned} I_{shunt} &= I_{total} - I_{meter} \\ &= 10 \text{ mA} - 1 \text{ mA} \\ &= 9 \text{ mA or } 0.009 \text{ ampere} \end{aligned}$$

The shunt resistance (R_{shunt}) may then be figured, as follows:

$$\begin{aligned} R_{shunt} &= \frac{E}{I} \\ &= \frac{0.05 \text{ volt}}{0.009 \text{ ampere}} \\ &= 5.55 \text{ ohms} \end{aligned}$$

Notice that, in this case, the exact value of shunt resistance has been used rather than an approximation.

The formula for determining the resistance of the shunt is given by $R_s = I_m / I_s$ times R_m, where R_s is the shunt resistance in ohms; I_m is the meter current at full-scale deflection; I_s is the shunt current at full-scale deflection; and R_m is the resistance of the meter coil. If the values given in the previous example are used in this equation, it will yield 5.55 ohms, the value previously calculated.

SWITCHING SHUNT VALUES.—Various values of shunt resistance can be used, by means of a suitable switching arrangement, to increase the number of current ranges that can be covered by the meter. Two switching arrangements are shown in figure 3-5. View A is the simpler of the two arrangements when a number of shunts are used to calculate the values of the shunt resistors. However, it has two disadvantages:

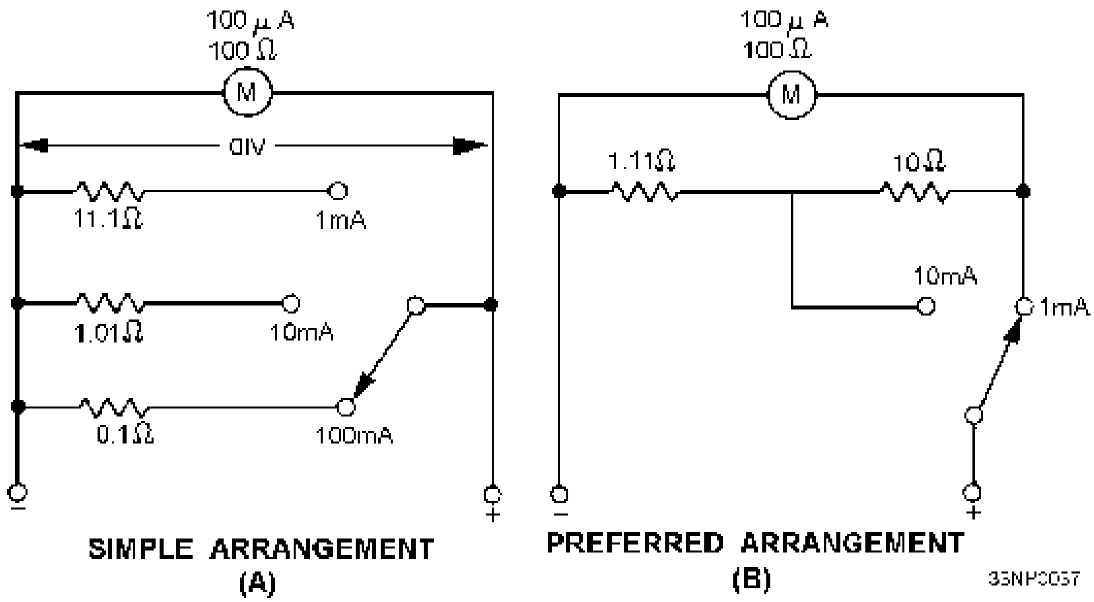


Figure 3-5.—Ways of connecting internal meter shunts.

1. When the switch is moved from one shunt resistor to another, the shunt is momentarily removed from the meter. The line current then flows through the meter coil. Even a momentary surge of current could easily damage the coil.

2. The contact resistance (resistance between the blades of the switch when they are in contact) is in series with the shunt, but not with the meter coil. In shunts that must pass high currents, this contact resistance becomes an appreciable part of the total shunt resistance. Because the contact resistance is of a variable nature, the ammeter indication may not be accurate.

The generally preferred method of range switching is shown in (figure 3-5, view B). Although only two ranges are shown, as many ranges as needed can be used. In this type of circuit, the contact resistance of the range-selector switch is external to the shunt and meter in each range position. The contact resistance in this case has no effect on the accuracy of the current measurement.

Ammeter Connections

When you are using ammeters, a primary rule of safety is that such current-measuring instruments must always be connected **in series** with a circuit, never in parallel with it. When an ammeter is connected across a constant-potential source of appreciable voltage, the low internal resistance of the meter bypasses the circuit resistance. This results in the application of the source voltage (or a good portion of it) directly to the meter terminals. The resulting excessive current burns up the meter coil and renders the meter useless until repaired.

Q-11. In what manner are current-measuring instruments connected to a circuit?

If you do not know the approximate value of current in the circuit, you should take a reading at the highest range of the ammeter; then you should switch progressively to lower ranges until a suitable reading is obtained. Most ammeter scales indicate the current being measured in increasing values from left to right. If you connect the meter without observing proper polarity, the pointer may be deflected backwards (from right to left). This action often damages the meter movement. You should ensure that the ammeter is always connected so that the current will flow into the negative terminal and out the positive terminal. Figure 3-6 shows various circuit arrangements and the proper ammeter connection methods to measure current in various portions of the circuit.

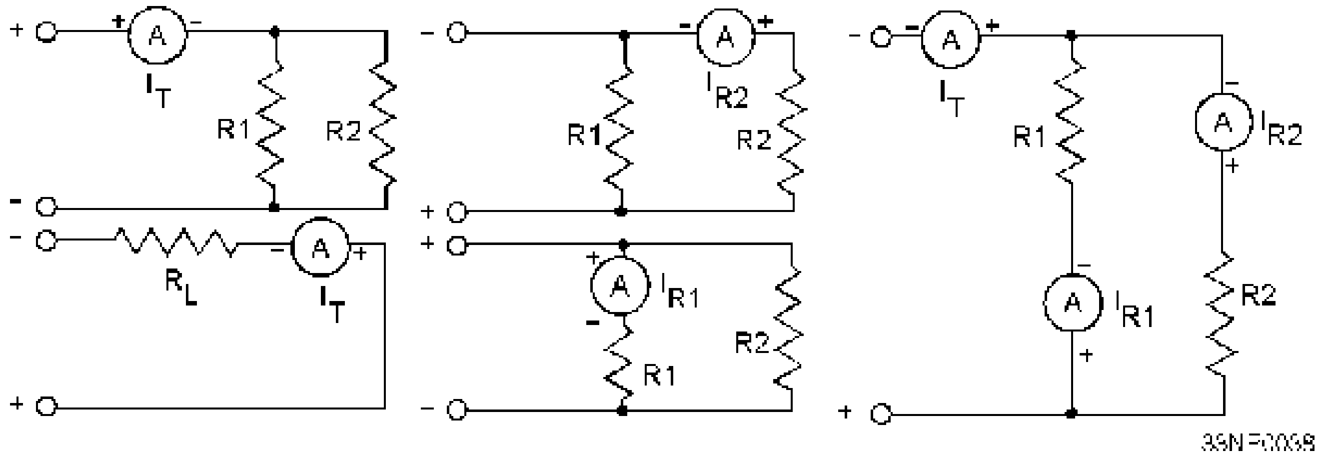


Figure 3-6.—Proper ammeter connection.

Q-12. An ammeter should always be connected so that current will flow into what terminal and out of what terminal?

Ammeter Sensitivity

Ammeter sensitivity is determined by the amount of current required by the meter coil to produce full-scale deflection of the pointer. The smaller the amount of current required to produce this deflection, the greater the sensitivity of the meter. A meter movement that requires only 100 microamperes for full-scale deflection has a greater sensitivity than a meter movement that requires 1 milliampere for the same deflection.

Q-13. (True or False) The larger the current required to produce full-scale deflection of the meter coil, the better the sensitivity of the meter.

Good sensitivity is especially important in ammeters to be used in circuits in which small currents flow. As the meter is connected in series with the load, the current flows through the meter. If the internal resistance of the meter is a large portion of the load resistance, an effect known as METER-LOADING will occur. Meter-loading is the condition that exists when the insertion of a meter into a circuit changes the operation of that circuit. This condition is not desirable. The purpose of inserting a meter into a circuit is to allow the measurement of circuit current in the normal operating condition. If the meter changes the circuit operation and changes the amount of current flow, the reading you obtain will be in error. An example of this is shown in figure 3-7.

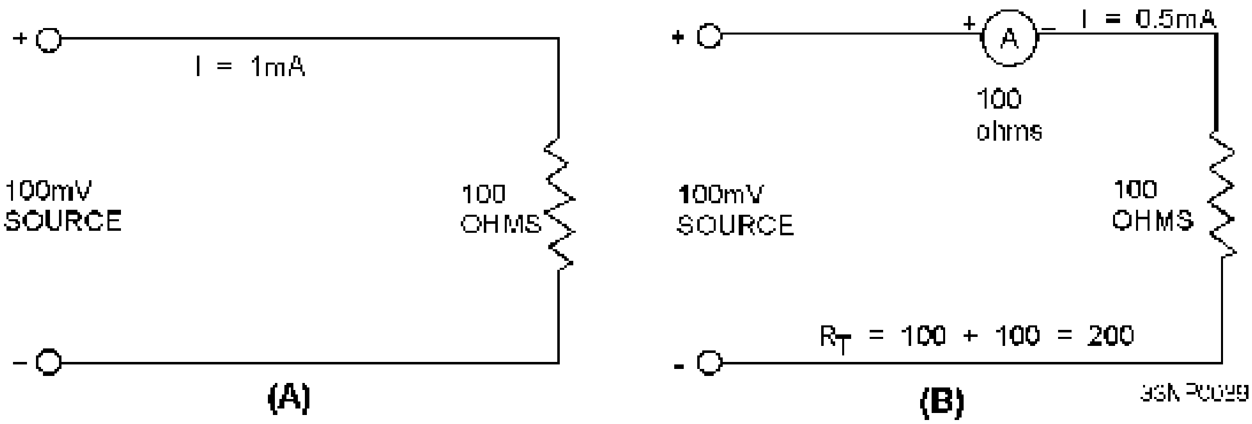


Figure 3-7.—Ammeter loading effect.

Q-14. What condition exists when the insertion of a meter into a circuit changes the operation of the circuit?

In view A of figure 3-7, the circuit to be tested has an applied voltage of 100 millivolts and a resistance of 100 ohms. The current normally flowing in this circuit is 1 milliampere. In view B, an ammeter that requires 1 milliampere for full-scale deflection and that has an internal resistance of 100 ohms has been inserted. Since 1 milliampere of current flow is shown in view A, you might naturally assume that with the meter inserted into the circuit, a full-scale deflection will occur. You might also assume that the 1 milliampere of circuit current will be measured. However, neither of these assumptions is correct. With the ammeter inserted into the circuit, as shown in view B, the total resistance of the circuit is 200 ohms. With an applied voltage of 100 millivolts, applying Ohm’s law shows the actual current ($I_{circuit}$) to be 0.5 milliampere.

$$I = \frac{E}{R}$$

$$I_{\text{circuit}} = \frac{100 \times 10^{-3}\,\text{volts}}{200\ \text{ohms}}$$

$$= 0.5 \times 10^{-3}\,\text{ampere}$$

$$= 0.0005\ \text{ampere or } 0.5\ \text{milliampere}$$

Since the meter reads 0.5 milliampere instead of the normal value of current, the meter reveals that a definite loading effect has taken place. In cases such as this, the use of ammeters, which have a lower internal resistance and a better current sensitivity, is desirable.

DC VOLTMETER

Up to this point, we have been discussing the 100-microampere D’Arsonval movement and its use as an ammeter. However, it can also be used to measure voltage if a MULTIPLIER (high resistance) is placed in series with the moving coil of the meter. For low-voltage instruments, this resistance is physically mounted inside the meter case with the D’Arsonval movement. The series resistance is constructed of a wire-wound resistance that has a low temperature coefficient wound on either a spool or card frame. For high-voltage ranges, the series resistance can be connected externally. A simplified diagram of a voltmeter is shown in figure 3-8.

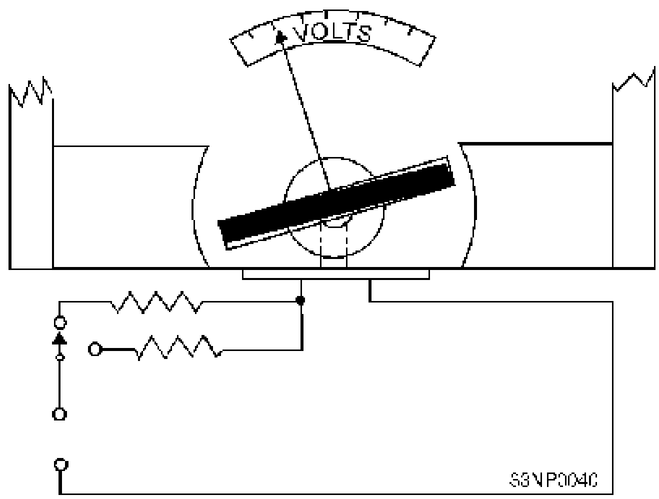


Figure 3-8.—Internal construction and circuit of a simplified voltmeter.

Q-15. What modification is made to the D’Arsonval meter movement to enable the meter to measure voltage?

Keep in mind that the D’Arsonval meter movement uses current flow to produce a magnetic field that is proportional to the current. The meter movement is, therefore, an indicator of current flow rather than voltage. The addition of the series resistance is what allows the meter to be calibrated in terms of voltage; that is, the meter movement of a voltmeter operates because of the current flow through the meter, but the scale is marked in volts. For example, the meter movement shown in figure 3-9 has an internal resistance of 100 ohms, requires 100 microamperes for full-scale deflection, and has a voltage drop of 10 millivolts when full-scale deflection is reached. If you were to place this meter directly across a 10-volt source, an excessive current (in milliamperes) would flow. The meter would be destroyed because of the excessive current flowing through the meter movement. This can be seen in the following Ohm’s law application:

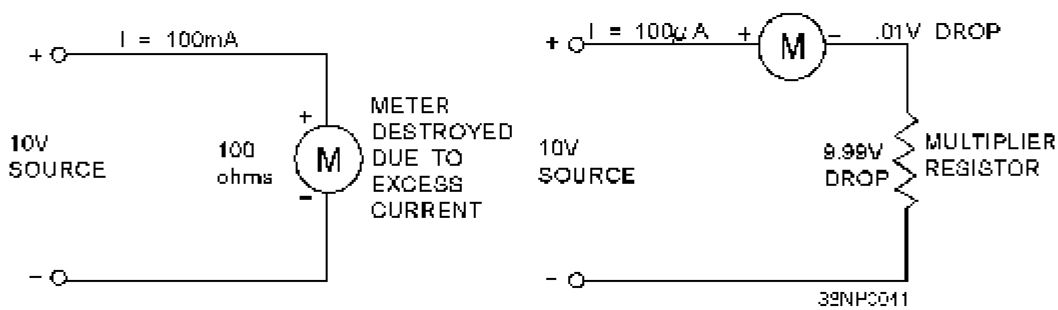


Figure 3-9.—Use of multiplier resistors with D'Arsonval meter movement.

Using this equation, you can see that a current through the meter of 100 milliamperes is excessive and will cause damage.

$$
\begin{aligned}
I &= \frac{E}{R} \\
&= \frac{10 \text{ volts}}{100 \text{ ohms}} \\
&= 100 \text{ milliamperes}
\end{aligned}
$$

Since the normal voltage drop for the meter is 10 millivolts at full-scale deflection, some means must be supplied to drop the extra 9.99 volts without applying it directly to the meter. This is done by the addition of a multiplier resistor, as shown in figure 3-9.

Extending Voltmeter Ranges

The value of series resistance is determined by the current required for full-scale deflection and by the range of the voltages to be measured. Since the current through the meter circuit is directly proportional to the applied voltage, the meter scale can be calibrated directly in volts for a fixed value of series resistance. For example, let's assume that the basic meter is to be made into a voltmeter with a full-scale deflection of 1 volt. The coil resistance of the basic meter is 100 ohms, and 100 microamperes of current causes full-scale deflection. The resistance (R_{meter}) required to limit the total current in the circuit to 100 microamperes can be found as follows:

$$
\begin{aligned}
R_{meter} &= \frac{E}{I} \\
&= \frac{1 \text{ volts}}{100 \text{ microamperes}} \\
&= 10 \text{ kilohms}
\end{aligned}
$$

Because the meter coil already measures 100 ohms, the series resistance required is equal to 10 kilohms minus 100 ohms, or 9.9 kilohms.

Q-16. What factors determine the value of the multiplier resistor?

Multirange voltmeters use one meter movement. The required resistances are connected in series with the meter by a switching arrangement. A schematic diagram of a multirange voltmeter with three ranges is shown in figure 3-10. The total meter resistance (R_{meter}) for each of the three ranges, beginning with the 1-volt range, is figured by the application of Ohm's law, as follows:

1 – volt range

$$R_{meter} = \frac{1}{.0001} = 10 \text{ kilohms}$$

100 – volt range

$$R_{meter} = \frac{100}{.0001} = 10 \text{ megohm}$$

1,000 – volt range

$$R_{meter} = \frac{1000}{.0001} = 10 \text{ megohms}$$

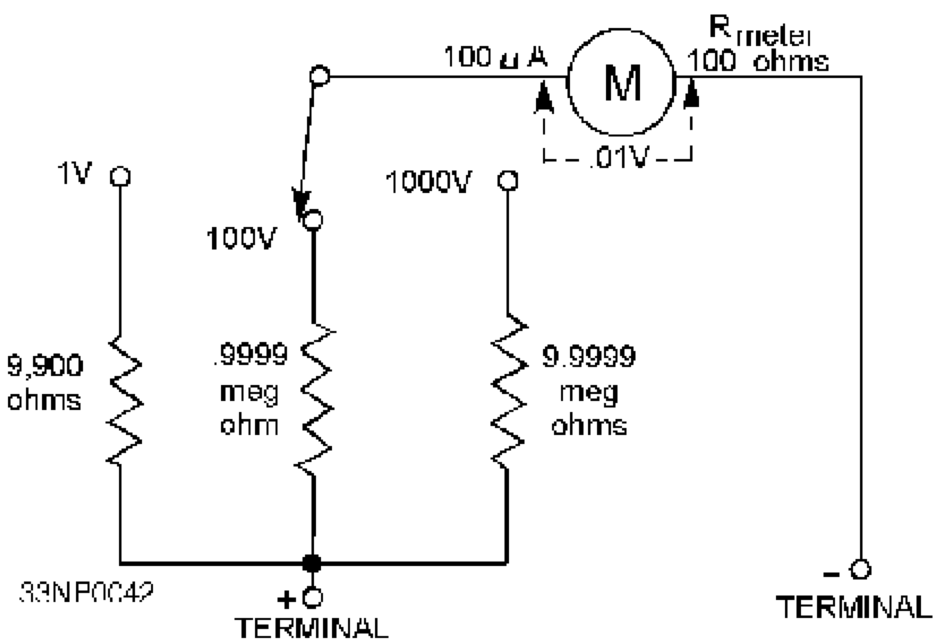


Figure 3-10.—Multirange voltmeter.

The actual value of the multiplying series resistor (R_{series}) for each of these circuits is 100 ohms less than the total resistance. This allows for the resistance of the meter coil (R_{coil}).

Voltmeter Circuit Connections

When voltmeters are used, a primary rule of safety is that such voltage-measuring instruments must always be connected in **parallel** with (across) a circuit. If you are unsure of the level of the voltage to be measured, take a reading at the highest range of the voltmeter and progressively (step by step) lower the range until a suitable reading is obtained. In many cases, the voltmeter you will be using will not be a center-zero- (0 reading is in the center) indicating instrument. Observing the correct polarity is important when connecting the instrument to the circuit. Voltmeter polarity is the same as for the dc ammeter; that is, current flows from negative to positive.

Q-17. In what manner are voltage-measuring instruments connected to the circuit to be measured?

Influence of a Voltmeter in a Circuit

The purpose of a voltmeter is to indicate the potential difference between two points in a circuit. When a voltmeter is connected across a circuit, it shunts the circuit. If the voltmeter has a low resistance, it will draw a substantial amount of current. This action lowers the effective resistance of the circuit and changes the voltage reading. When you are making voltage measurements in high-resistance circuits, use a HIGH-RESISTANCE VOLTMETER to prevent the shunting action of the voltmeter. The effect is less noticeable in low-resistance circuits because the shunting effect is less. The problem of voltmeter shunting (sometimes called circuit loading) is illustrated in figure 3-11.

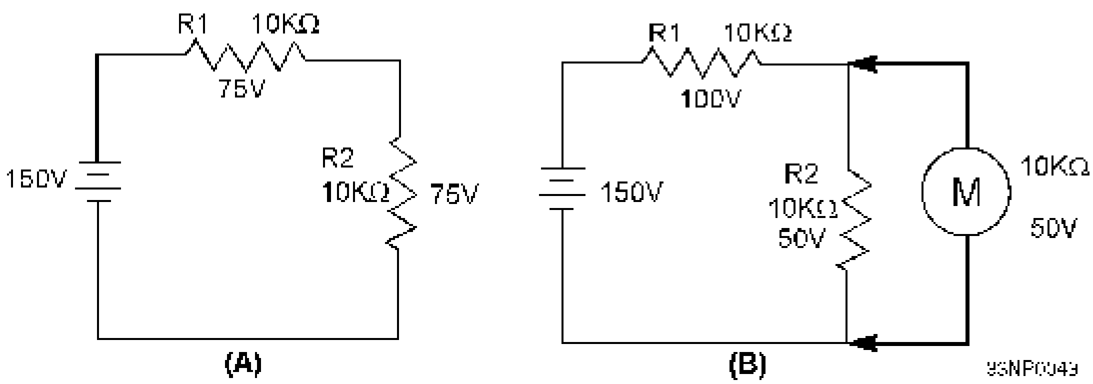


Figure 3-11.—Shunting action caused by a voltmeter.

Q-18. When making voltage measurements in a high-resistance circuit, you should always use a voltmeter with what relative value of resistance?

In view A of figure 3-11, a source of 150 volts is applied to a series circuit consisting of two 10-kilohm resistors. View A shows the voltage drop across each resistor to be 75 volts. In the 150-volt range, the voltmeter to be used has a total internal resistance of 10 kilohms. View B shows the voltmeter connected across the circuit. The parallel combination of R2 and the meter now present a total resistance of 5 kilohms. Because of the addition of the voltmeter, the voltage drops change to 100 volts across R1 and 50 volts across R2. Notice that this is not the normal voltage drop across R2. Actual circuit conditions have been altered because of the voltmeter.

Voltmeter Sensitivity

The sensitivity of a voltmeter is given in ohms per volt. It is determined by dividing the sum of the resistance of the meter (R_{meter}), plus the series resistance (R_{series}), by the full-scale reading in volts. In equation form, sensitivity is expressed as follows:

$$\text{sensitivity} = \frac{R_m + R_S}{E}$$

This is the same as saying the sensitivity is equal to the reciprocal of the full-scale deflection current. In equation form, this is expressed as follows:

$$\text{sensitivity} = \frac{\text{ohms}}{\text{volt}} = \frac{1}{\text{volt}/\text{ohms}} = \frac{1}{\text{ampere}}$$

Therefore, the sensitivity of a 100-microampere movement is the reciprocal of 0.0001 ampere, or 10,000 ohms per volt.

$$\text{sensitivity} = \frac{1}{\text{ampere}} = \frac{1}{.0001} = 10{,}000 \text{ ohms per volt}$$

Q-19. What term is used to express the sensitivity of a voltmeter?

METERS USED FOR MEASURING RESISTANCE

The two instruments you will use most often to check continuity, or to measure the resistance of a circuit or circuit component, are the OHMMETER and the MEGGER (MEGOHMMETER). The ohmmeter is widely used to measure resistance and to check the continuity of electrical circuits and devices. Its range usually extends to only a few megohms. The megger is widely used for measuring insulation resistance, such as that between a wire and the outer surface of its insulation, and the insulation resistance of cables and insulators. The range of a megger can be extended to more than 1,000 megohms.

Q-20. What instrument is used for measuring the insulation resistance of cables?

The Ohmmeter

A simple ohmmeter circuit is shown in figure 3-12. The ohmmeter consists of the dc milliammeter, discussed earlier in this chapter, and the added features shown below:

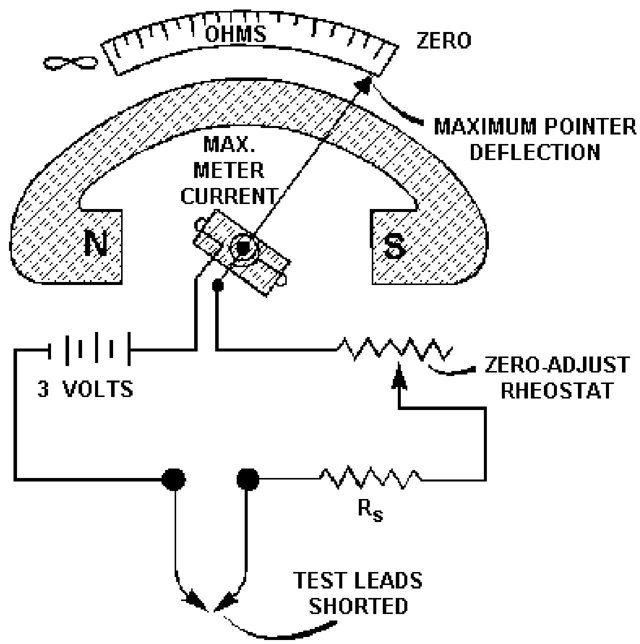


Figure 3-12.—Simple ohmmeter circuit.

- A source of dc potential; and
- One or more resistors (one of which is variable).

Q-21. What added features enable a dc milliammeter to function as an ohmmeter?

The deflection of the pointer of an ohmmeter is controlled by the amount of battery current passing through the moving coil. Before you can measure the resistance of an unknown resistor or electrical circuit, you must calibrate the ohmmeter to be used. If the value of resistance to be measured can be estimated within reasonable limits, select a range on the ohmmeter that will give approximately half-scale deflection when the resistance is inserted between the probes. If you cannot estimate the resistance to be measured, then set the range switch on the highest scale. Whatever range you select, the meter must be calibrated to read zero before the unknown resistance is measured.

To calibrate the meter, you first short the test leads together, as shown in figure 3-12. With the test leads shorted, a complete series circuit exists. The complete series circuit consists of the 3-volt source, the resistance of the meter coil (R_{meter}), the resistance of the zero-adjust rheostat, and the series multiplying resistor (R_{series}). The shorted test leads cause current to flow and the meter pointer to deflect.

Notice that the zero point on the ohmmeter scale (as opposed to the zero points for voltage and current) is located at the extreme right side of the scale. With the test leads shorted, the zero-adjust potentiometer is set so that the pointer rests on the zero mark. Therefore, a full-scale deflection indicates zero resistance between the leads.

Q-22. A full-scale deflection on an ohmmeter scale indicates what resistance between the leads?

If you change the range on the meter, you must "zero" (calibrate) the meter again to obtain an accurate reading. When you separate the test leads, the pointer of the meter will return to the left side of the scale. This action, as explained earlier, is caused by the restoring force of the spring tension acting on the movable coil assembly. The reading at the left side of the scale indicates an infinite resistance.

After you have adjusted the ohmmeter for zero reading, it is ready to be connected to a circuit to measure resistance. A typical circuit and ohmmeter arrangement is shown in figure 3-13. You must ensure that the power switch of the circuit to be measured is in the de-energized (OFF) position. This prevents the source voltage of the circuit from being applied to the meter, a condition that could cause severe damage to the meter movement.

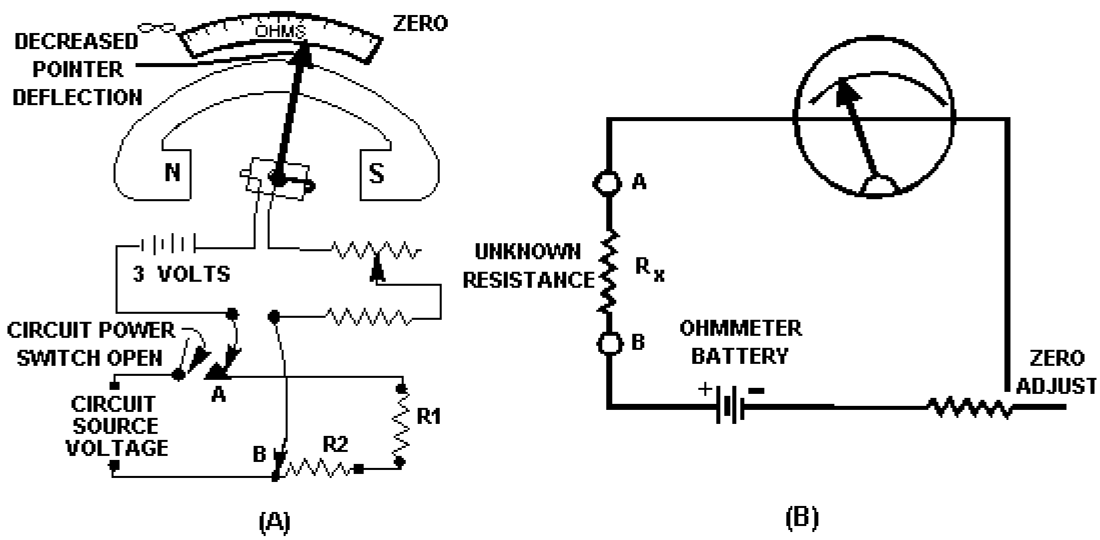


Figure 3-13.—Measuring circuit resistance with an ohmmeter.

Remember that the ohmmeter is an open circuit when the test leads are separated. To take a resistance reading with a meter, you must provide a path for current flow produced by the meter's battery. In view A of figure 3-13, the meter is connected at points **A** and **B** to produce this path. Connecting these test leads places resistors R1 and R2 in series with the resistance of the meter coil, the zero-adjust potentiometer, and the series multiplying resistor. Since you previously calibrated the meter, the amount of coil movement now depends only on the resistances of R1 and R2.

The addition of R1 and R2 into the meter circuit raises the total series resistance and decreases the current. This decreases the amount of pointer deflection. The pointer comes to rest at a scale reading that indicates the combined resistance of R1 and R2. If you were to replace either R1 or R2, or both, with a resistor having a larger ohmic value, the current flow in the moving coil of the meter would be decreased even more. This would further decrease the pointer deflection, and the scale indication would read a still higher circuit resistance. View B is a simplified version of the circuitry in view A.

From our ohmmeter discussion, two facts should be apparent: (1) Movement of the moving coil is proportional to the amount of current flow, and (2) the scale reading of the ohmmeter is inversely proportional to current flow in the moving coil.

The amount of circuit resistance to be measured may vary over a wide range. In some cases, it may only be a few ohms; in other cases, it may be as great as 1 megohm. Scale multiplication features are built into most ohmmeters so that they will indicate any ohmic value being measured and offer the least amount of error. Most ohmmeters are equipped with a selector switch for selecting the multiplication scale desired. For example, view A of figure 3-14 shows a typical meter that has a six-position switch. The positions are marked on the meter in multiples of 10, from R × 1 through R × 100K.

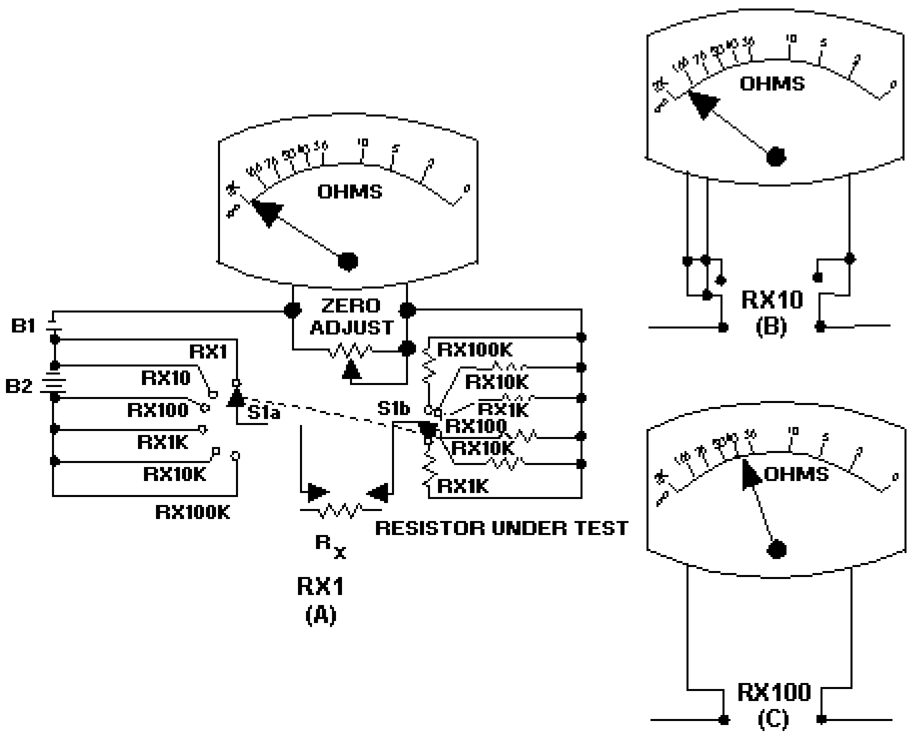


Figure 3-14.—Ohmmeter with multiplication switch.

The range used to measure any particular unknown resistance (R_x in view A of figure 3-14) depends on the approximate ohmic value of the unknown resistance. For instance, the ohmmeter scale of the figure is calibrated in divisions from 0 to infinity. Note that the divisions are easier to read on the right-hand portion of the scale than on the left. For this reason, if R_x is greater than 1,000 ohms and if you are using the R × 1 range, you will be unable to accurately read the indicated resistance. This happens because the combined series resistance of resistors R_x is too large for range R × 1 to allow enough battery current to flow to deflect the pointer away from infinity. You need to turn the range switch to the R × 10 position to obtain the 1,000-ohm reading.

Let's assume that you have changed the range switch to the R × 10 position and the pointer now deflects to a reading of 375 ohms, as shown in view B of figure 3-14. This would indicate to you that unknown resistance R_x has 3,750 (375 times 10) ohms of resistance. The change of range caused the deflection because resistor R × 10 has only 1/10 the resistance of resistor R × 1. Therefore, selecting the smaller series resistance allowed a battery current of sufficient value to cause a readable pointer deflection. If the R × 100 range were used to measure the same 3,750 ohm resistor, the pointer would deflect still further to the 37.5-ohm position, as shown in view C. This increased deflection would occur because resistor R × 100 has only 1/10 the resistance of resistor R × 10.

Q-23. The R × 100 resistance selection on an ohmmeter has what amount of resistance compared to the R × 10 selection?

The circuit arrangement in view A of figure 3-14 allows the same amount of current to flow through the moving meter coil. The same amount is allowed to flow whether the meter measures 10,000 ohms on the R × 1 scale, 100,000 ohms on the R × 10 scale, or 1,000,000 ohms on the R × 100 scale.

The same amount of current must always be used to deflect the pointer to a certain position on the scale (midscale position, for example), regardless of the multiplication factor being used. Since the multiplier resistors are of different values, you must always "zero" the meter for each multiplication scale selected. When selecting a range on the ohmmeter, select the one that will result in the pointer coming to rest as close to the midpoint of the scale as possible. This will enable you to read the resistance more accurately because scale readings are more easily interpreted at or near midpoint.

The Megohmmeter

An ordinary ohmmeter cannot be used for measuring multimillion ohm values of resistances, such as those in conductor insulation. To test for such insulation breakdown, you need to use a much higher potential than that supplied by the battery of an ohmmeter. This potential is placed between the conductor and the outside of the insulation. A megger (megohmmeter) is used for these tests. The megger, shown in figure 3-15, is a portable instrument consisting of two main elements: (1) a hand-driven dc generator, which supplies the necessary voltage for making the measurement, and (2) the instrument portion, which indicates the value of the resistance you are measuring. The instrument portion is of the opposed-coil type, as shown in view A. Coils **a** and **b** are mounted on movable member **c**. A fixed angular relationship exists between coils, and they are free to turn as a unit in a magnetic field. Coil **b** tends to move the pointer counterclockwise, and coil **a** tends to move it clockwise.

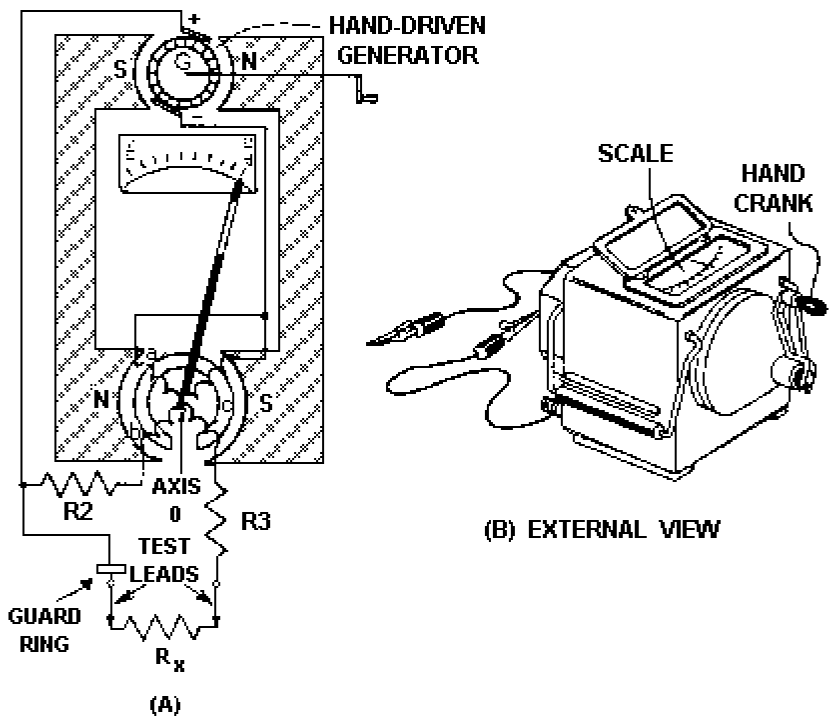


Figure 3-15.—Megger internal circuit and external view.

Coil **a** is connected in series with R3 and unknown resistance R_x. The combination of coil **a**, R3, and R_x forms a direct series path between the + and – brushes of the dc generator. Coil **b** is connected in series with R2, and this combination is also connected across the generator. Notice that the movable member (pointer) of the instrument portion of the megger has no restoring springs. Therefore, when the generator is not being operated, the pointer will float freely and may come to rest at any position on the scale.

The guard ring, shown in view A of figure 3-15, shunts any leakage currents to the negative side of the generator. This prevents such current from flowing through coil **a** and affecting the meter reading.

Q-24. What is the purpose of the guard ring in a megohmmeter?

If the test leads are open, no current will flow in coil **a**. However, current will flow internally through coil **b** and deflect the pointer to infinity. This reading indicates a resistance too large to measure. When a resistance, such as R_x, is connected between the test leads, current also flows in coil **a**; the pointer tends to move clockwise. At the same time, coil **b** still tends to move the pointer counterclockwise. Therefore, the moving element, composed of both coils and the pointer, comes to rest at a position in which the two forces are balanced. This position depends upon the value of R_x, which controls the amount of the current in coil **a**. Because changes in voltage affect both coils in the same proportion, the position of the moving element is independent of the voltage. If you short the test leads together, the pointer will come to rest at zero because the current in coil **a** is relatively large. Since R3 limits the current, the instrument will not be damaged under these circumstances. The external appearance of one type of megger is shown in view B of figure 3-15.

Most meggers you will use are rated at 500 volts; however, there are other types. Meggers are usually equipped with friction clutches, which are designed to slip if the generator is cranked faster than its rated speed. This prevents the generator speed and output voltage from exceeding rated values. A 1,000-volt generator is available for extended ranges. When an extremely high resistance, such as 10,000 megohms or more, is to be measured, a high voltage is needed to cause enough current flow to actuate the meter movement.

CAUTION

When using a megger, you can easily be injured or damage equipment if you do not observe the following MINIMUM safety precautions:

- Use meggers on high-resistance measurements only (such as insulation measurements or to check two separate conductors on a cable).
- Never touch the test leads while the handle is being cranked.
- De-energize and discharge the circuit completely before connecting a megger.
- Whenever possible, disconnect the component being checked from other circuitry before using a megger

Q-25. Most meggers you will use are rated at what voltage?

Q-26. The development of excessive test voltages is avoided by the use of meggers equipped with what device?

ELECTRODYNAMOMETER-TYPE METERS

The electrodynamometer-type meter differs from the galvanometer types we have just studied in that two fixed coils are used to produce the magnetic field instead of a permanent magnet. Two movable coils are also used in the electrodynamometer meter. The electrodynamometer meter is most commonly found in various types of power meters.

Q-27. What components in an electrodynamometer-type meter movement produce the magnetic field?

As shown in figure 3-16, the fixed coils are connected in series and positioned coaxially (in line) with a space between them. The two movable coils are also positioned coaxially and are connected in series. The two pairs of coils (fixed pair and movable pair) are also connected in series with each other. The movable coil is pivot-mounted between the fixed coils. The main shaft on which the movable coils are mounted is restrained by spiral springs that restore the pointer to zero when no current is flowing through the coil. These springs also act as conductors for delivering current to the movable coils. Since these conducting springs are very small, the meter cannot carry a high value of current.

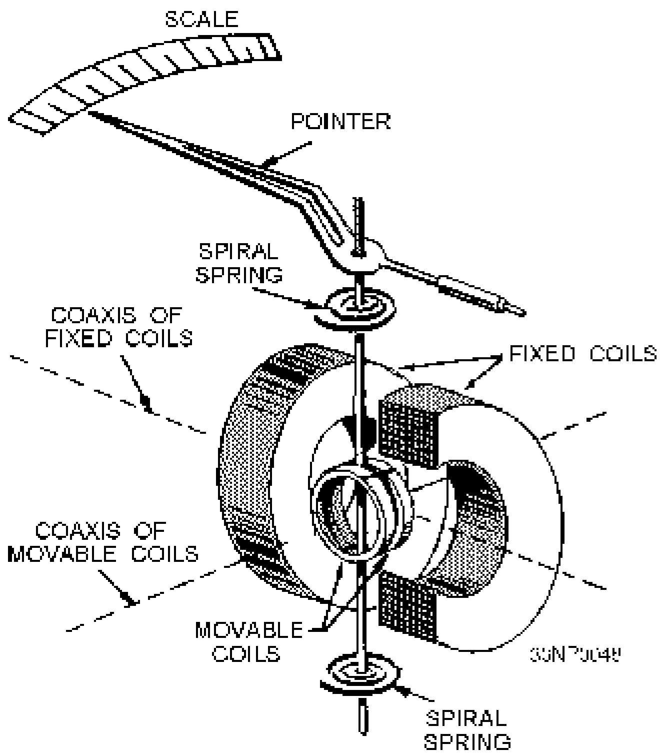


Figure 3-16.—Internal construction of an electrodynamometer.

Q-28. What is the limiting factor as to the amount of current an electrodynamometer meter movement can handle?

METER ACCURACY

The meter is mechanically damped by means of aluminum vanes that move in enclosed air chambers. Although very accurate, electrodynamometer-type meters do not have the sensitivity of the D'Arsonval-type meter movement. For this reason, you will not find them used outside of the laboratory environment to a large extent.

METER MOVEMENT

The primary advantage of the electrodynamometer-type meter movement is that it can be used to measure alternating as well as direct current. If you apply alternating current to the standard galvanometer-type meter, it will not produce a usable reading. Instead, the meter will vibrate at or near the zero reading. On one-half cycle of the ac, the meter is deflected to the left and on the other half cycle to the right. Since the frequencies you will be measuring are 60 hertz or greater, the meter is incapable of mechanically responding at this speed. The result is simply a vibration near the zero point; in addition, no useful reading of voltage or current is obtained. This problem does not exist with the electrodynamometer-type movement. Current flow through the stationary (fixed) coils sets up a magnetic field. Current flow through the moving coils sets up an opposing magnetic field. With two magnetic fields opposing, the pointer deflects to the right. If the current reverses direction, the magnetic fields of *both* sets of coils will be reversed. With both fields reversed, the coils still oppose each other, and the pointer still deflects to the right. Therefore, no rectifying devices are required to enable the electrodynamometer meter movement to read both ac and dc. Rectifying devices are required for the D'Arsonval-type meter movement to enable it to be used for measuring ac voltages and currents.

Q-29. What is the primary advantage of the electrodynamometer-type meter over the D'Arsonval-type meter?

VOLTMETER

When an electrodynamometer is used as a voltmeter, no problems in construction are encountered because the current required is not more than 0.1 ampere. This amount of current can be handled easily by the spiral springs. When the electrodynamometer is used as a voltmeter, its internal connections and construction are as shown in view A of figure 3-17. Fixed coils **a** and **b** are wound of fine wire since the current flow through them will not exceed 0.1 ampere. They are connected directly in series with movable coil **c** and the series current-limiting resistor.

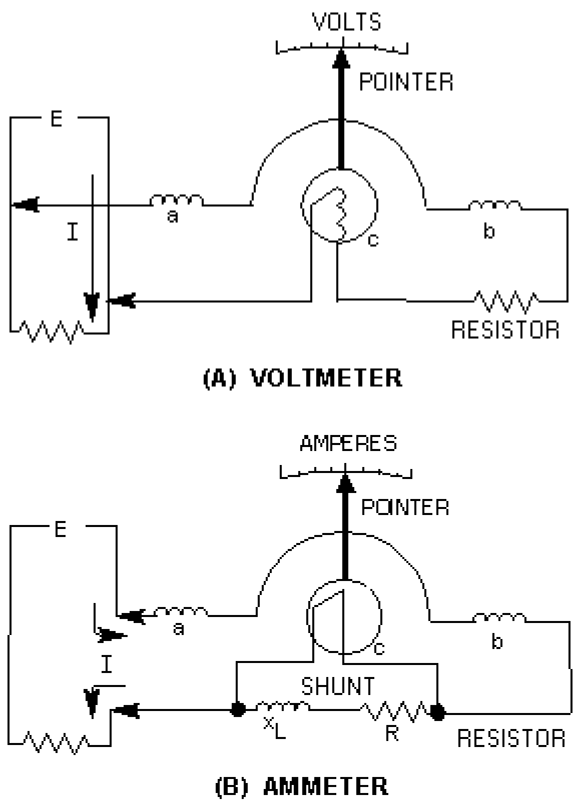


Figure 3-17.—Circuit arrangement of electrodynamometer for use as a voltmeter and an ammeter.

AMMETER

When the electrodynamometer is used as an ammeter, a special type of construction must be used. This is because the large currents that flow through the meter cannot be carried through the moving coils. In the ammeter in view B of figure 3-17, stationary coils a and b are wound of heavier wire to carry up to 5.0 amperes. An inductive shunt (X_L) is wired in parallel with the moving coils and permits only a small part of the total current to flow through the moving coil. The current flowing through the moving coil is directly proportional to the total current flowing through the instrument. The shunt has the same ratio of reactance to resistance as the moving coil does. Therefore, the instrument will be reasonably correct at frequencies at which it is used if ac currents are to be measured.

WATTMETER

Electric power is measured by means of a wattmeter. This instrument is of the electrodynamometer type. As shown in figure 3-18, it consists of a pair of fixed coils, known as current coils, and a moving coil, called the voltage (potential) coil. The fixed current coils are wound with a few turns of a relatively large conductor. The voltage coil is wound with many turns of fine wire. It is mounted on a shaft that is supported in jeweled bearings so that it can turn inside the stationary coils. The movable coil carries a needle (pointer) that moves over a suitably graduated scale. Coil springs hold the needle at the zero position in the absence of a signal.

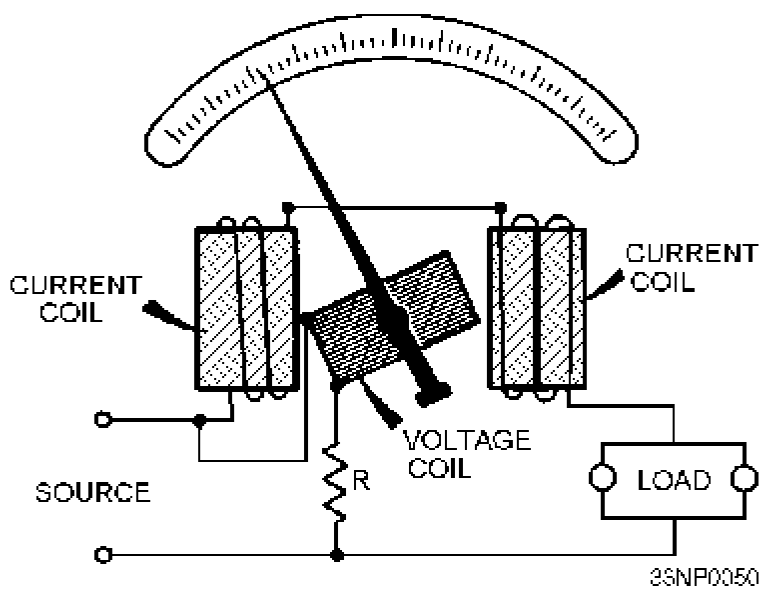


Figure 3-18.—Simplified electrodynamometer wattmeter circuit.

Wattmeter Connection

The current coil of the wattmeter is connected in series with the circuit (load), and the voltage coil is connected across the line. When line current flows through the current coil of a wattmeter, a field is set up around the coil. The strength of this field is in phase with and proportional to the line current. The voltage coil of the wattmeter generally has a high-resistance resistor connected in series with it. The purpose for this connection is to make the voltage-coil circuit of the meter as purely resistive as possible. As a result, current in the voltage circuit is practically in phase with line voltage. Therefore, when voltage is impressed on the voltage circuit, current is proportional to and in phase with the line voltage. Figure 3-19 shows the proper way to connect a wattmeter into a circuit.

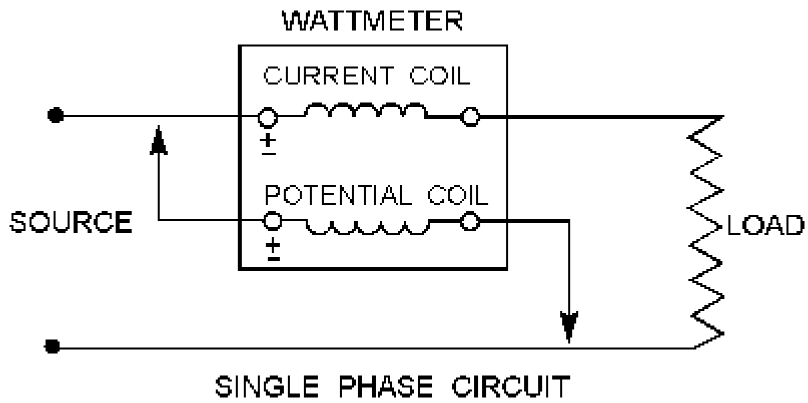


Figure 3-19.—Wattmeter connection.

Wattmeter Errors

Electrodynamic wattmeters are subject to errors arising from such factors as temperature and frequency. For example, heat through the coils eventually causes the small springs attached to the pointer to lengthen and lose tension, which produces deflection errors. Large currents through the wattmeter also produce a noticeable deflection error. These errors are caused by the heat (I^2R) loss through coils from the application of high currents. Because of this, the maximum current range of electrodynamic wattmeters is normally restricted to approximately 20 amperes. The voltage range of wattmeters is usually limited to several hundred volts because of heat dissipation within the voltage circuit. However, the voltage range can be extended by the use of voltage multipliers.

Good-quality, portable wattmeters usually have an accuracy of 0.2 to 0.25 percent. You must remember, though, that electrodynamic wattmeter errors increase with frequency. For the higher frequency and power ranges, special types of wattmeters are made specifically for those ranges. We will discuss two such wattmeters in chapter 5 of this module.

Wattmeter Overloads

The wattmeter consists of two circuits, either of which will be damaged if too much current passes through them. You should be especially aware of this fact because the reading on the instrument will not tell you whether or not the coils are being overheated. If an ammeter or voltmeter is overloaded, the pointer will indicate beyond the upper limit of its scale. In the wattmeter, both the current and potential circuit may carry such an overload that their insulations burn; yet the pointer may be only part of the way up the scale. This is because the position of the pointer depends upon the power factor of the circuit as well as upon the voltage and current. Therefore, a low power-factor circuit will provide a very low reading on the wattmeter. The reading will be low, even when the current and voltage circuits are loaded to the maximum safe limit. The safe rating for each wattmeter is always distinctly rated, not in watts, but in volts and amperes.

TECHNIQUES FOR METER USE

We have considered the more common meters; now let's consider some of the techniques employed in their use. The techniques suggested here are not all-inclusive. You will find, as you develop your technical skills, other variations and techniques in use. Consider the techniques for measuring current in a circuit. You can accomplish this by placing an ammeter in series with the circuit or by measuring the voltage across a resistor of known value and using Ohm's law to figure current. This last technique has the advantage of eliminating the necessity of opening the circuit to connect the ammeter.

CONTINUITY TESTS

Open circuits are those in which the flow of current is interrupted by a broken wire, defective switch, or any means by which the current cannot flow. The test used to detect open circuits (or to see if the circuit is complete or continuous) is continuity testing.

An ohmmeter (which contains its own batteries) is excellent for use in a continuity test. Normally, continuity tests are performed in circuits where the resistance is very low, such as the resistance of a copper conductor. An open is indicated in these circuits by a very high or infinite resistance between two continuously connected points.

Figure 3-20 shows a continuity test of a cable that connects two electronic units. Notice that both plugs are disconnected and the ohmmeter is in series with conductor **D** under test. The power should be off. When checking conductors **A**, **B**, and **C** (connection of ohmmeter to conductors not shown), the current from the ohmmeter flows through plug **2** (female) through conductor **A**, **B**, or **C** to plug **1** (female). From plug **1**, current passes through the jumper to the chassis, which is "grounded" to the ship's structure. The metal structure serves as the return path to the chassis of unit **2** and completes the circuit through the series-connected ohmmeter. The ohmmeter indicates a low resistance because no break exists in conductors **A**, **B**, or **C**. However, checking conductor **D** reveals an open. The ohmmeter is shown indicating maximum resistance because current cannot flow in an open circuit. With an open circuit, the ohmmeter needle is all the way to the left since it is a series-type ohmmeter (reads right to left).

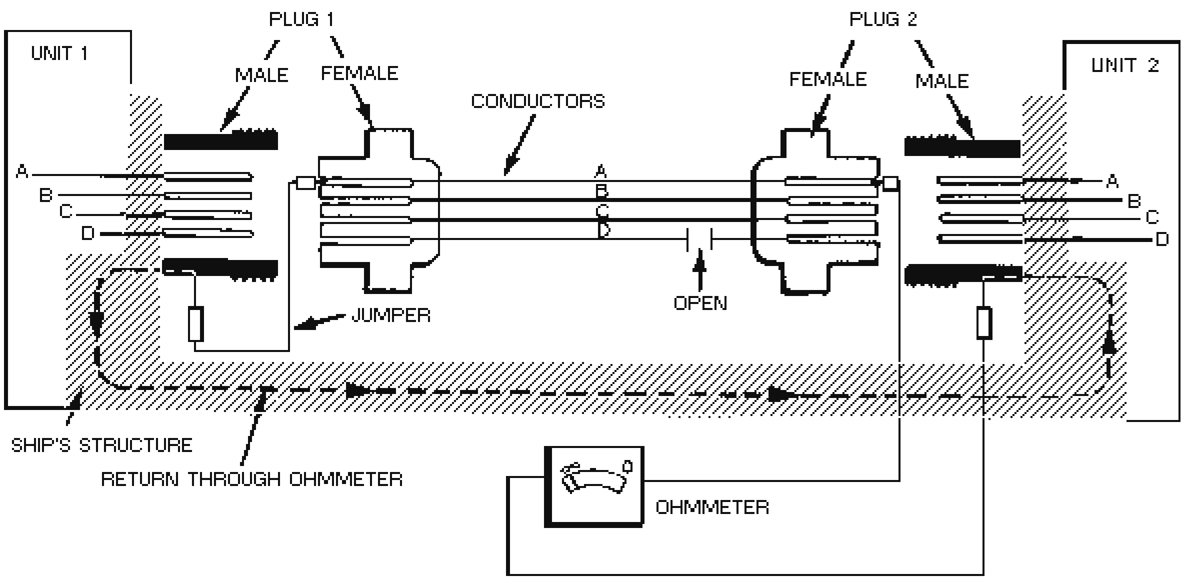


Figure 3-20.—Continuity test.

Where conditions are such that the ship's structure cannot be used as the return path, one of the other conductors (known to be good) may be used. For example, to check **D**, you can connect a jumper from pin **D** to pin **A** of plug **1** (female) and the ohmmeter leads to pins **D** and **A** of plug **2** (female). This technique will also reveal the open in the circuit.

TESTING FOR GROUNDS

Grounded circuits are caused by some conducting part of the circuit making contact either directly or indirectly with the metallic structure of the ship. Grounds can have many causes. The two most common are the fraying of insulation from a wire and moisture-soaked insulation. The fraying of insulation from a wire allows bare wire to come into contact with the metal ground. Moisture-soaked insulation causes reduced insulation resistance (also classified as a ground).

Grounds are usually indicated by blown fuses or tripped circuit breakers. Blown fuses or tripped circuit breakers, however, can also result from a short circuit other than a ground. A high-resistance ground can also occur when current is increased significantly but not enough to rupture the fuse or trip the circuit breaker.

CAUTION

Before testing any circuit, ensure the circuit under test has been de-energized and checked with a safety shorting probe.

In testing for grounds, you may use a megger or an ohmmeter. Measuring the resistance to ground from points in a circuit determines if the point is grounded. Referring again to figure 3-20, you can see one possible means of testing a cable for grounds. If the jumper is removed from pin **D** of plug **1** (female), a test for ground can be made for each conductor in the cable. You can do this by connecting one meter lead to ground and the other to each of the pins of either of the plugs. A low resistance indicates that some part of that conductor or one of the plug assemblies is grounded. Both plugs must be removed from their units; if only one plug is removed, a false indication is possible because a conductor may be grounded through the unit.

TESTING FOR SHORTS

A short circuit, other than a grounded one, is one where two conductors touch each other directly or through another conducting element. Two conductors with frayed insulation may touch and cause a short. Too much solder on the pin of a connector may short to the adjacent pin. In a short circuit, enough current may or may not flow to blow a fuse or open a circuit breaker. A short may occur between two cables carrying signals but might not be indicated by a blown fuse.

Shorts occur in many components, such as transformers, motor windings, and capacitors. The major test method used to detect shorts in such components is to measure resistance. The indicated resistance is then compared with the resistance given on schematics or in the equipment technical manuals to determine whether the measured value is within specifications.

An ohmmeter is the device used to check for shorts. You can use the ohmmeter to detect a short between two conductors by measuring the resistance between them (be sure electrical power has been disconnected). A low resistance reading indicates a short. You can test the circuit in figure 3-20 for a short by first removing the jumper and disconnecting both plugs; you then measure the resistance between the two suspended conductors.

WARNING

The following section discusses voltage measurements on live circuits. BE SURE YOU ALWAYS FOLLOW PRESCRIBED SAFETY RULES WHEN MEASURING VOLTAGES.

VOLTAGE TESTS

Voltage tests must be made with the power applied; therefore, the prescribed safety precautions must be followed to prevent injury to personnel and damage to the equipment. You will find in your maintenance work that the voltage test is of utmost importance. It is used not only in isolating casualties to major components but also in the maintenance of subassemblies, units, and circuits. Before checking a circuit voltage, you should check the voltage of the power source to be sure that the normal voltage is being applied to the circuit.

The voltmeter is used for voltage tests. In using the voltmeter, make certain that the meter used is designed for the type of current (ac or dc) to be tested and has a scale with a suitable range. Since defective parts in a circuit can cause higher than normal voltages to be present at the point of test, the highest voltmeter range available should be used first. Once you have obtained a reading, determine if a lower scale can be used that will cause no damage to the meter movement. If so, use the lower scale. This provides a more accurate reading.

Another consideration in the circuit voltage test is the resistance and current in the circuit. A low resistance in a high-current circuit could result in considerable voltage drop, whereas the same resistance in a low-current circuit may be minimal. Abnormal resistance in part of a circuit can be checked with either an ohmmeter or a voltmeter. Where practical, an ohmmeter should be used because the test is then carried out with a "dead" circuit.

The majority of the electronic circuits you will encounter in equipment will be low-current circuits, and most voltage readings will be direct current. Also, many of the schematics will indicate the voltages at various test points. Therefore, if you suspect that a certain stage is defective, you can check the voltage by connecting a voltmeter from the test point to ground. If the suspected stage is not defective, the voltmeter readings should match the voltages given on the schematic.

Some technical manuals also contain voltage charts on which all the voltage measurements are tabulated. These charts usually indicate the sensitivity of the meter (for example, 20,000 ohms/volt) used to obtain the voltage readings for the chart. To obtain comparable results, you must use a voltmeter of the same sensitivity (or greater) as that specified. Make certain that the voltmeter is not "loading" the circuit while taking a measurement. If the meter resistance is not considerably higher than the circuit resistance, the reading will be markedly lower than the true circuit voltage because of the voltmeter's loading effect. (To calculate meter resistance, multiply the rated ohms-per-volt sensitivity value of the meter by the scale in use. For example, a 1,000-ohms-per-volt meter set to the 300-volt scale will have a resistance of 300,000 ohms.)

RESISTANCE TESTS

Before checking the resistance of a circuit or of a part, make certain that the power has been turned off. Also make sure capacitors in the associated circuit are fully discharged. To check continuity, always use the lowest ohmmeter range. If the highest range is used, the meter may indicate zero, even though appreciable resistance is present in the circuit. Conversely, to check a high resistance, use the highest scale since the lower range scale may indicate infinity, even though the resistance is less than a megohm. In making resistance tests, you must remember that even though the external ohmmeter leads are connected in parallel with the circuit to be measured, the internal meter circuitry is electrically connected in series.

In making resistance tests, take into account that other circuits containing resistances and capacitances may be in parallel with the circuit to be measured. Erroneous conclusions may be drawn from readings obtained in such cases. Remember, a capacitor blocks the dc flow from the ohmmeter. To obtain an accurate reading when other parts are connected across the suspected circuit, disconnect one end of the circuit to be measured from the equipment. For example, many of the resistors in major components and subassemblies are connected across transformer windings. To obtain a valid resistance measurement, you must isolate the resistors to be measured from the shunt resistances of the coils of the transformers.

Resistance tests are also used to check a component for grounds. In these tests, the component to be tested should be disconnected from the rest of the circuit so that no normal circuit ground will exist. Dismounting the component to be checked is not necessary. The ohmmeter is set for a high-resistance range. Then the ohmmeter is connected between ground and each electrically separate circuit of the component being tested. Any resistance reading less than infinity indicates at least a partial ground. You can also check capacitors suspected of being short-circuited by measuring the resistance. To check a capacitor suspected of being open, temporarily shunt a known good capacitor then recheck the performance of the circuit.

CAUTION

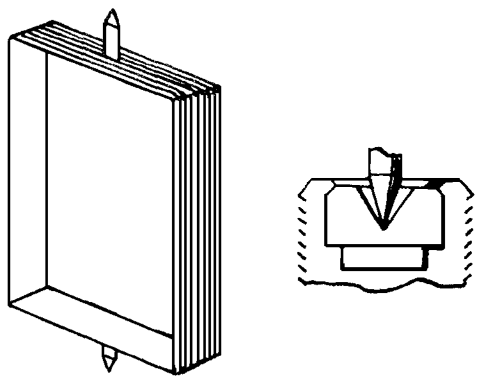
To avoid possible damage to equipment during resistance tests, observe the following precautions:

- Always connect an ammeter in series—*never* in parallel.
- Connect a voltmeter in parallel.
- *Never* connect an ohmmeter to a live circuit.
- Observe polarity when using a dc ammeter or a dc voltmeter.
- View meters directly from the front. When viewed from an angle off to the side, an incorrect reading will result because of OPTICAL PARALLAX. (Parallax was covered in NEETS, Module 3, *Introduction to Circuit Protection, Control, and Measurement.)*
- Always choose an instrument suitable for the measurement desired.
- Select the highest range first and then switch to the proper range.
- In using a meter, choose a scale that will result in an indication as near midscale as possible.
- Do not mount or use instruments in the presence of a strong magnetic field.
- Remember, a low internal resistance voltmeter (low sensitivity) may shunt the circuit being measured and result in incorrect readings.

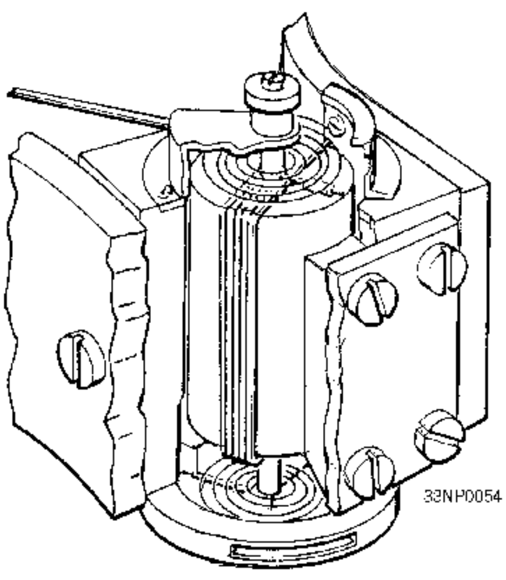
SUMMARY

The important points of this chapter are summarized in the following paragraphs. You should be familiar with these points before continuing with your study of test equipment.

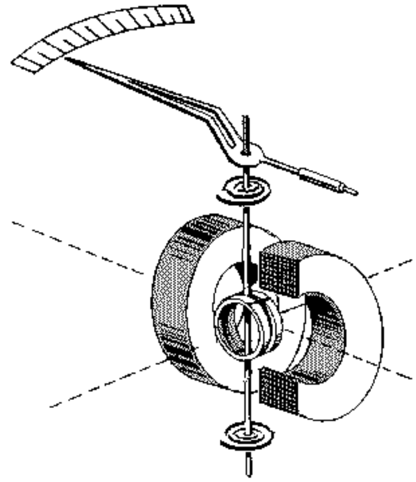
A permanent-magnet, moving-coil meter movement (**D'ARSONVAL** movement) uses the interaction of magnetic fields to produce movement.



DAMPING is used to smooth out the vibration and to help prevent overshooting of the meter pointer.



ELECTRODYNAMOMETER movements are usually used in wattmeters. They operate much like the D'Arsonval meter movement, except field coils are used instead of a permanent magnet. Electrodynamometer movements measure either ac or dc without the use of a rectifier.

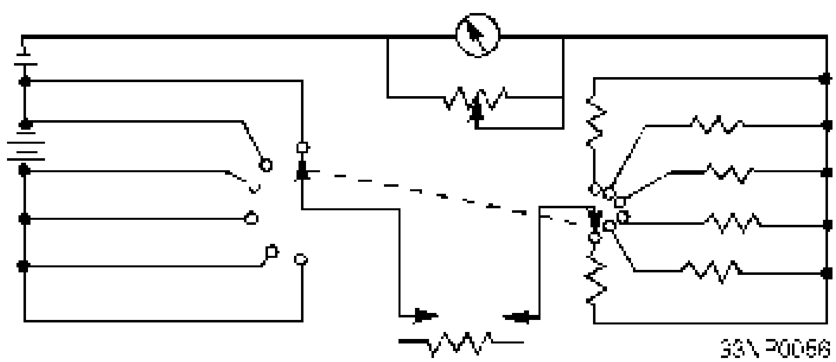


A **SHUNT** is a physically large, low-resistance conductor connected in parallel with the meter terminals. It carries the majority of the load current so that only a small portion of the total current will flow through the meter coil.

An **AMMETER** measures current and is always connected in series with the circuit being measured. An ammeter should have a low resistance so that the effect of the ammeter on the circuit will be kept to a minimum.

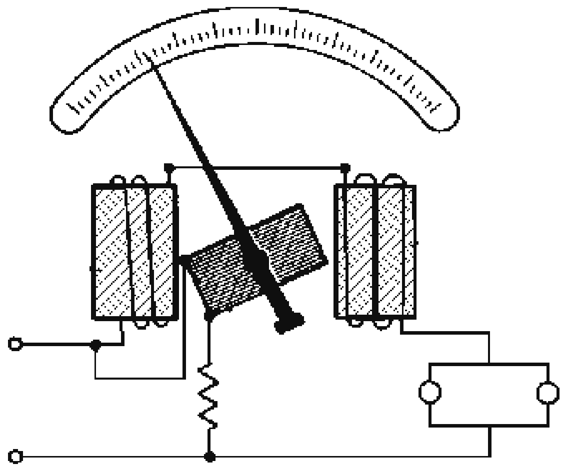
VOLTMETERS are used to measure voltage and are always connected in parallel with the circuit being measured. A voltmeter should have a high resistance compared to the circuit being measured to minimize the loading effect. Voltmeter sensitivity is expressed in ohms per volt.

OHMMETERS are used to measure resistance and to check continuity. An ohmmeter is electrically connected in series with the resistance being measured. The ohmmeter range, which allows a midscale deflection, should be used.



A **MEGOHMMETER (MEGGER)** is used to measure very high resistance, such as the insulation of wiring.

A **WATTMETER** is usually an electrodynamometer and is used to measure power.



A **CONTINUITY TEST** is accomplished with an ohmmeter. This test is used to check for opens (or to see if the circuit is complete or continuous).

GROUNDED CIRCUITS are caused by some conducting part of the circuit making contact either directly or indirectly with the metallic structure of the ship or chassis. In testing for grounds, you may use either an ohmmeter or a megger.

A **SHORT CIRCUIT**, other than a grounded one, is where two conductors touch each other directly or through another conducting element. An ohmmeter is used to test for shorts.

ANSWERS TO QUESTIONS Q1. THROUGH Q29.

A-1. Self-excited.

A-2. Phosphor bronze ribbons.

A-3. The pointer arrangement and the light and mirror arrangement.

A-4. Coil balance.

A-5. Hairspring.

A-6. Hairspring.

A-7. Makes it possible to have a more linear scale than if the poles were flat.

A-8. Shunt.

A-9. Zero-temperature coefficient.

A-10. Midscale.

A-11. In series.

A-12. Negative, positive.

A-13. False.

A-14. Meter-loading.

A-15. A multimeter (high resistance) is placed in series with the coil of the meter.

A-16. The current required for full-scale deflection, and the range of the voltage to be measured.

A-17. In parallel.

A-18. High.

A-19. Ohms per volt.

A-20. Megohmmeter (megger).

A-21. 1. A source of dc potential. 2. One or more resistors (one of which is variable).

A-22. Zero.

A-23. 1/10.

A-24. Shunts leakage current, which prevents false readings.

A-25. 500.

A-26. Friction clutches.

A-27. Fixed coils.

A-28. Size of spiral conducting.

A-29. The electrodynamometer-type meter can be used to measure both ac and dc currents.

CHAPTER 4

COMMON TEST EQUIPMENT

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to:

1. Describe the proper operating procedures for using the multimeter.
2. Describe the proper operating procedures for using the digital multimeter.
3. Describe the proper operating procedures for using the differential voltmeter.
4. Describe the proper operation of the transistor tester.
5. Describe the proper procedure for using the RCL bridge to measure resistance, capacitance, and inductance.

INTRODUCTION

In the previous chapters, you have learned how to use some basic and miscellaneous measuring instruments to perform required maintenance and upkeep of electronic systems and components. You were also introduced to the construction and operation of basic meter movements in test equipment. This chapter will introduce you to some of the testing instruments commonly used in the Navy today.

MULTIMETERS

During troubleshooting, you will often be required to measure voltage, current, and resistance. Rather than using three or more separate meters for these measurements, you can use the MULTIMETER. The multimeter contains circuitry that allows it to be used as a voltmeter, an ammeter, or an ohmmeter. A multimeter is often called a VOLT-OHM-MILLIAMMETER (VOM).

One of the greatest advantages of a VOM is that no external power source is required for its operation; therefore, no warm-up is necessary. Other advantages are its portability, versatility, and freedom from calibration errors caused by aging tubes, line voltage variations, and so forth.

Q-1. What is one of the greatest advantages of a VOM?

Two disadvantages are that (1) the VOM tends to "load" the circuit under test, and (2) the meter movement is easily damaged as a result of improper testing procedures.

CAUTION

Never press down on or place any object on the glass face of any multimeter. This can disable the meter movement from operating properly or cause damage.

MEASURING RESISTANCE, VOLTAGE, AND CURRENT WITH A VOM

In the discussion that follows, you will become familiar with the operation and use of the multimeter in measuring resistance, voltage, and current.

The meter selected for this discussion is the Simpson 260 multimeter, as shown in figure 4-1. The Simpson 260 is a typical VOM used in the Navy today.

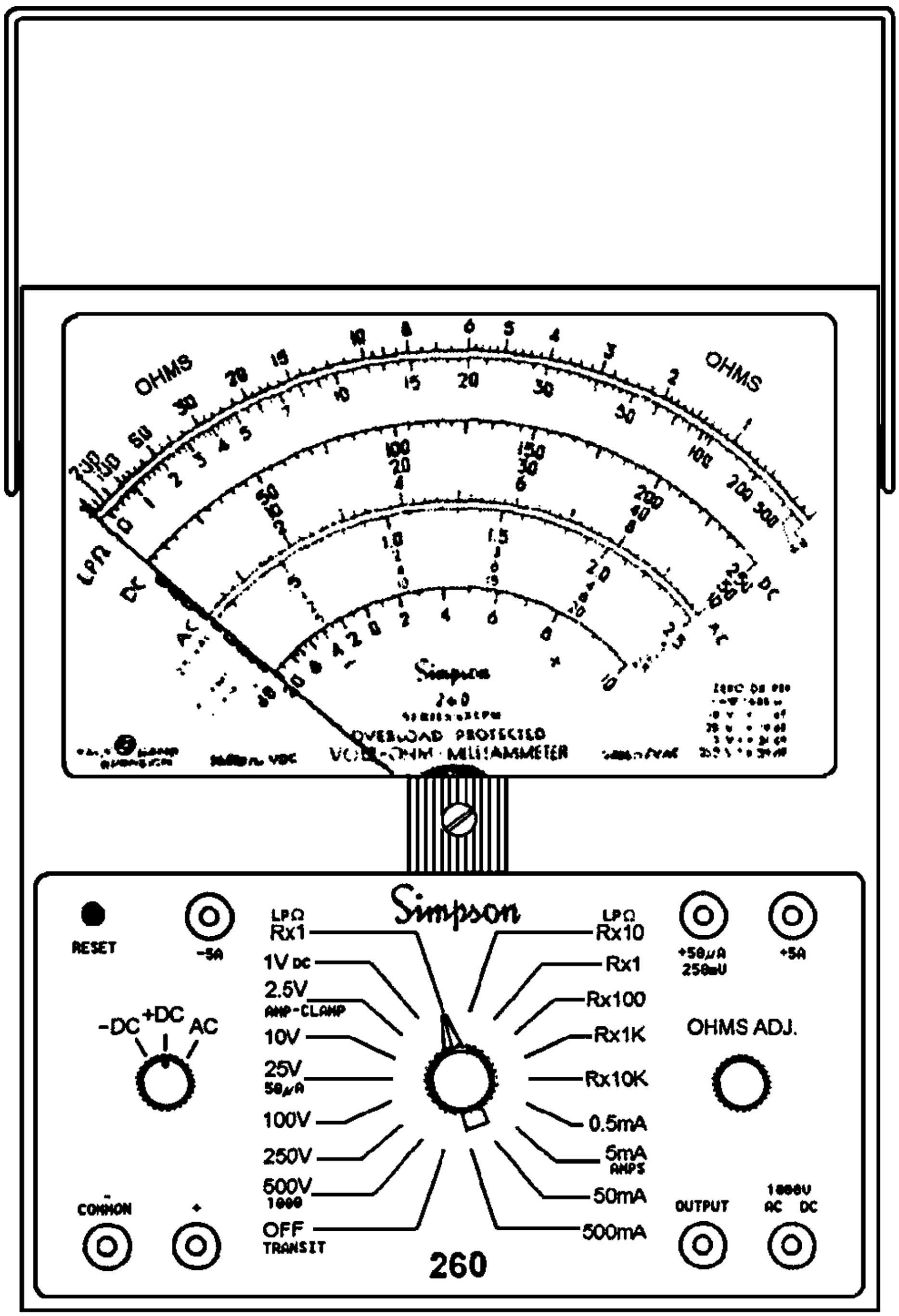


Figure 4-1.—Simpson 260 Series 6XLP Volt-Ohm-Milliammeter (VOM).

The multimeter has two selector switches. The switch on the lower left is the function switch, and the one in the lower center is the range switch. The function switch selects the type of current you will be measuring (+dc, –dc, or ac). The range switch is a 12-position switch that selects the range of ohmmeter, voltmeter, or milliammeter measurements you will make.

The multimeter is equipped with a pair of test leads; red is the positive lead and black is the negative, or common, lead. Eight jacks are located on the lower part of the front panel. To prepare the meter for use, simply insert the test leads into the proper jacks to obtain the circuit and range desired for each application. In most applications, the black lead will be inserted into the jack marked at the lower left with a negative sign (–) or with the word COMMON.

Measuring Resistance

Before proceeding, you should be aware of the following important safety precaution that must be observed when using the ohmmeter function of a VOM:

CAUTION

Never connect an ohmmeter to a "hot" (energized) circuit. Be sure that no power is applied and that all capacitors are discharged.

Q-2. Before you connect a VOM in a circuit for an ohmmeter reading, in what condition must the circuit be?

The internal components of the multimeter use very little current and are protected from damage by an overload protection circuit (fuse or circuit breaker). However, damage may still occur if you neglect the safety precaution in the CAUTION instructions above.

Because no external power is applied to the component being tested in a resistance check, a logical question you may ask is, Where does the power for deflection of the ohmmeter come from? The multimeter contains its own two-battery power supply inside the case. The resistive components inside the multimeter are of such values that when the leads are connected together (no resistance), the meter indicates a full-scale deflection. Because there is no resistance between the shorted leads, full-scale deflection represents zero resistance.

Before making a measurement, you must zero the ohmmeter to ensure accurate readings. This is accomplished by shorting the leads together and adjusting the OHMS ADJ control so that the pointer is pointing directly at the zero mark on the OHMS scale. The ZERO OHMS control is continuously variable and is used to adjust the meter circuit sensitivity to compensate for battery aging in the ohmmeter circuits.

An important point to remember when you are making an accurate resistance measurement is to "zero" the meter each time you select a new range. If this is not done, the readings you obtain will probably be incorrect.

When making a resistance measurement on a resistor, you must give the following considerations to the resistor being tested:

- The resistor must be electrically isolated. In some instances, a soldered connection will have to be disconnected to isolate the resistor. Generally, isolating one side of the resistor is satisfactory for you to make an accurate reading.

- The meter leads must make good electrical contact with the resistor leads. Points of contact should be checked for dirt, grease, varnish, paint or any other material that may affect current flow.

- Touch only the insulated portions of the test leads. Your body has a certain amount of resistance, which the ohmmeter will measure if you touch the uninsulated portions of the leads.

Figure 4-2 is a functional block diagram of the ohmmeter circuit in a VOM. The proper method of checking a resistor is to connect the red lead to one end of the resistor and the black lead to the other end of the resistor.

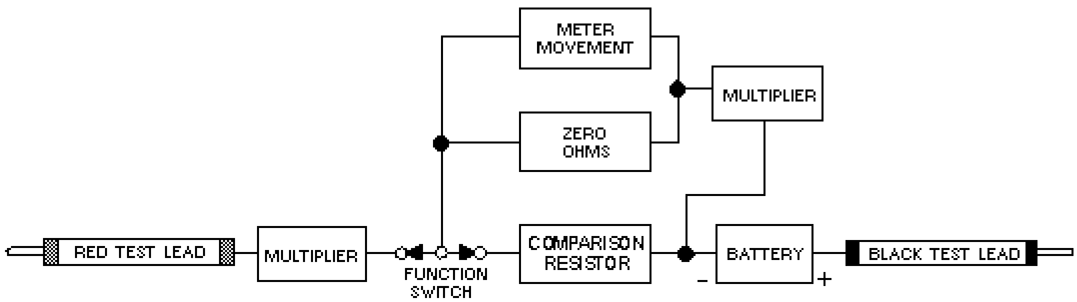


Figure 4-2.—Functional block diagram of an ohmmeter circuit.

Because zero resistance causes full-scale deflection, you should realize that the deflection of the meter is inversely proportional to the resistance being tested; that is, for a small resistance value, the deflection will be nearly full scale; and for a large resistance value, the deflection will be considerably less. This means that the left portion of the OHMS scale represents high resistance; the right side of the scale represents low resistance. Zero resistance (a short circuit) is indicated on the extreme right side of the scale; infinite resistance (an open circuit) is located on the extreme left side of the scale.

Notice that you read the OHMS scale on the multimeter from RIGHT to LEFT. For example, the pointer of the multimeter in figure 4-3 indicates 8.0 ohms. To determine the actual value of a resistor, multiply the reading on the meter scale by the range switch setting ($R \times 1$, $R \times 100$, or $R \times 10,000$).

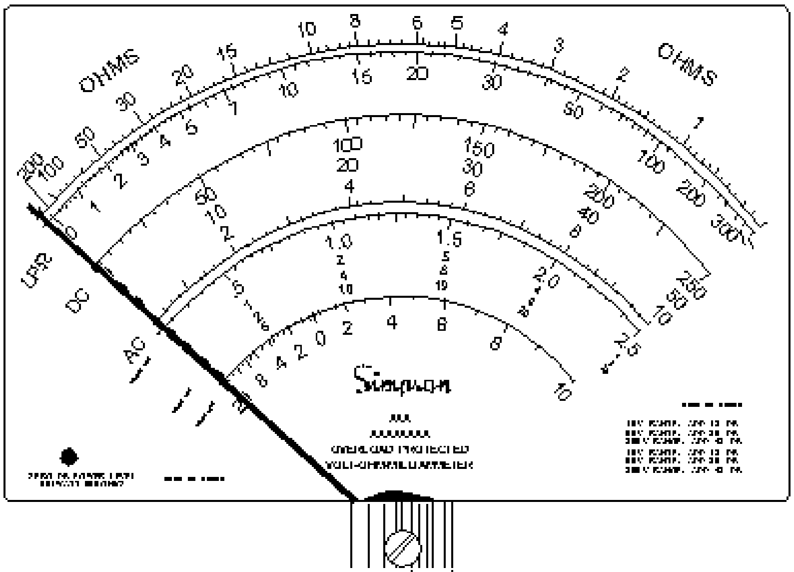


Figure 4-3.—Ohmmeter scale.

Notice that the scale marks are crowded on the left side of the OHMS scale, which makes them difficult to read. Therefore, the best range to select is one in which the pointer will fall in the space from midscale to slightly to the right side of midscale. The divisions in this area of the scale are evenly spaced and provide for easier reading and greater accuracy.

Q-3. When taking resistance readings with a VOM, you will obtain the most accurate readings at or near what part of the scale?

To explain the relationship between the meter readings and the range switch setting, let's use an example. Suppose you have a 2,400-ohm resistor, which you have identified by the resistor color code. With the range switch in the R × 1 position, you connect the meter across the resistor. The meter point then deflects between 200 and the point labeled with the infinity symbol (∞) on the extreme left side of the scale. Because the R × 1 range is selected, you multiply the reading by 1. Obviously, the scale reading is not accurate enough. Therefore, you move the range selector switch to the next higher scale position (R × 100) to obtain a more easily read value.

In the R × 100 position, you again zero the meter. This time, the pointer moves to the 24 mark on the scale. Because the R × 100 scale is selected, the reading is multiplied by 100. This gives a more accurate reading of 2,400 ohms (24 times 100).

If you position the range switch to the R × 10,000 scale, accuracy decreases. The most accurate readings are obtained at or near midscale. Other VOM instruments have ranges with other settings, such as R × 10, R × 100, or R × 1,000, to make it easier to make such readings.

Another thing to remember when you are measuring resistance is the tolerance of the resistor. If the tolerance of the resistor in the preceding example is 10 percent, we would expect a reading between approximately 2,160 and 2,640 ohms. If the reading is not within these limits, the resistor has probably changed value and should be discarded.

An open resistor will indicate no deflection on the meter. A shorted resistor causes full-scale deflection to the right on the lowest range scale, such as if the leads were shorted together.

Measuring dc Voltages

You set the multimeter to operate as a dc voltmeter by placing the function switch in either of two positions: +DC or –DC. The meter leads, as in the case of the ohmmeter function, must be connected to the proper meter jacks. When you measure dc voltages, be sure the red lead is the positive lead and the black lead is the negative, or common, lead. View A of figure 4-4 is a functional block diagram of dc voltage circuits in a multimeter. View B shows the jacks and switch positions for measuring dc voltages.

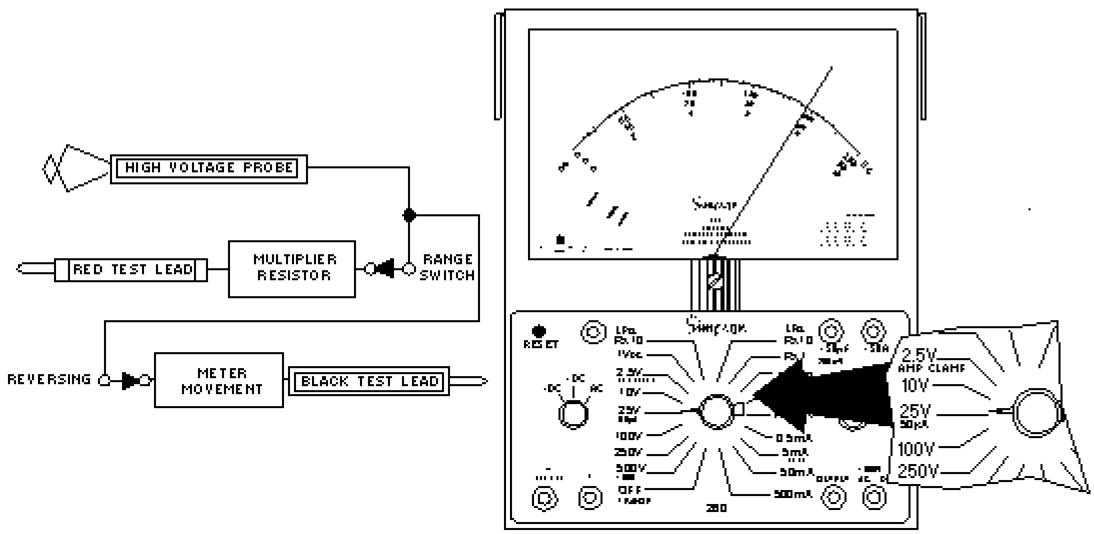


Figure 4-4.—Functional block diagram of dc voltage circuits.

When the meter is connected in a circuit, it becomes a circuit component. Because all meters have some resistance, they alter the circuit by changing the current. The resistance presented by the voltmeter depends on the amount of voltage being measured and the position of the function switch.

Some multimeters use a 20,000 ohms-per-volt meter sensitivity for measuring dc voltage and a 5,000 ohms-per-volt sensitivity for measuring ac voltage. The higher the meter resistance, the less it will load the circuit. The idea is to keep circuit loading to an absolute minimum so that the circuit under test is unaffected by the meter. In this way, you can get a clearer picture of what the circuit malfunction is, not the effect of the meter on the circuit.

Again, refer to figure 4-4. With the function switch set to either +DC or –DC, let's consider the effect of the range switch on the meter scale to be used. When measuring dc voltages, you have eight voltage ranges available: .25V, 2.5V, 10V, 50V, 250V, and 500V (1- and 1,000-volt special application plug-ins are also available). The setting of the range switch determines the maximum value represented on the meter. When measuring dc voltages, use the scale marked DC (figure 4-3). The last number at the extreme right side of the DC scale indicates the maximum value of the range being used. When the range switch is in the 2.5V position, the scale represents a maximum of 2.5 volts.

To simplify the relationship between the digits on the meter scale and the setting of the range switch, always use the multiple of the full-scale-deflection digits on the meter face that correspond to the numbers of the range switch. For example, use the 250 scale for the 250MV jack, 2.5V, and 250V ranges; the 50 on the scale for 50V and 500V ranges; and the 10 on the scale for the 10V and 1,000V ranges.

For explanation purposes, let's assume you wish to measure 30 volts dc. In this case, select the next higher range position, 50V. When you place the range switch to the 50V position (as shown in view B of figure 4-4), the meter pointer should rise from a little more than midscale to 30, which represents 30 volts dc.

When measuring a *known* dc voltage, position the range switch to a setting that will cause approximately midscale deflection. Readings taken near the center of the scale are the most accurate. When measuring an *unknown* dc voltage, always begin on the highest voltage range. Using the range switch, work down to an appropriate range. If the meter pointer moves to the left, you should reverse the polarity of the function switch.

CAUTION

Always check the polarity before connecting the meter.

Q-4. Besides setting up the meter for expected voltage ranges, what must be strictly observed when taking dc voltage readings?

Now let's discuss how you take a voltage measurement on a component within a circuit. As an example, let's measure the voltage drop across the resistor shown in figure 4-5.

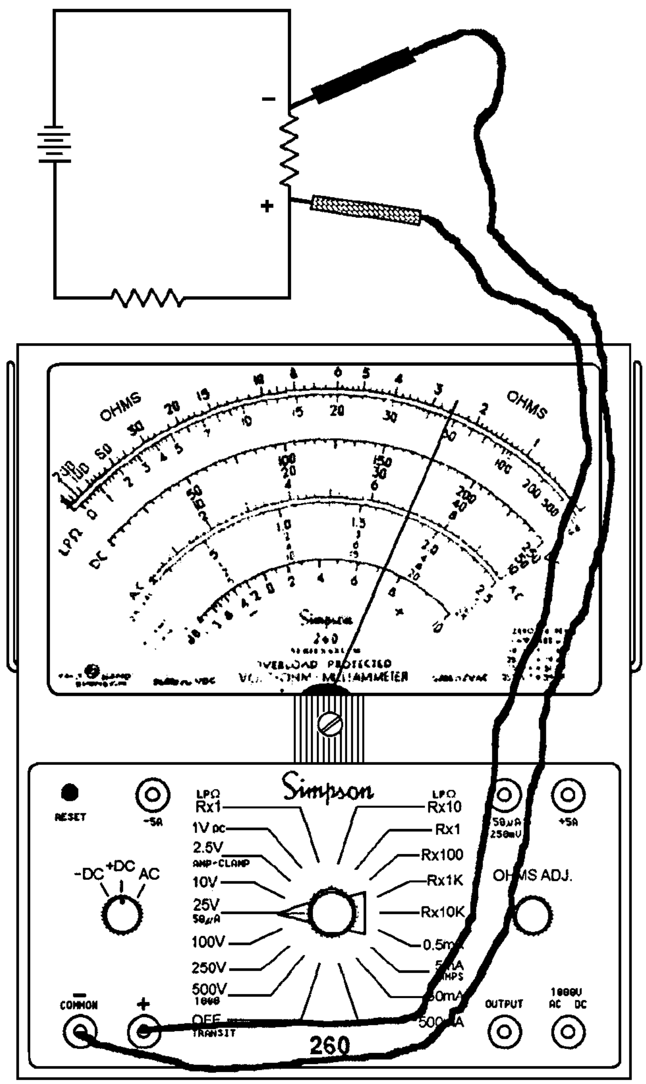


Figure 4-5.—Measuring the voltage drop of a resistor.

When measuring a dc voltage drop across a component in a circuit, you must connect the voltmeter in parallel with the component. As you can see in figure 4-5, the positive (red) lead is connected to the positive side of the resistor, and the negative (black) lead is connected to the negative side. A voltage reading is obtained on the meter when current flows through the resistor.

Some voltmeter readings will require the use of a ground as a reference point. Under these conditions, one voltmeter lead is connected to the equipment ground, and the other lead is connected to the test point where voltage is to be measured. Be sure to observe polarity.

Measuring ac Voltages

To measure ac voltages, you must set the function switch to the AC position. The same procedure used to measure dc voltages applies, except that in reading the voltage, you use the AC volts scale (the polarity of the test leads is not important). When measuring very low or high frequencies of ac voltages, you should be aware that the multimeter has a tendency to be inaccurate. View A of figure 4-6 is a functional block diagram of the ac and output voltage circuits in the multimeter. View B shows the jacks and switch positions used to measure ac voltages.

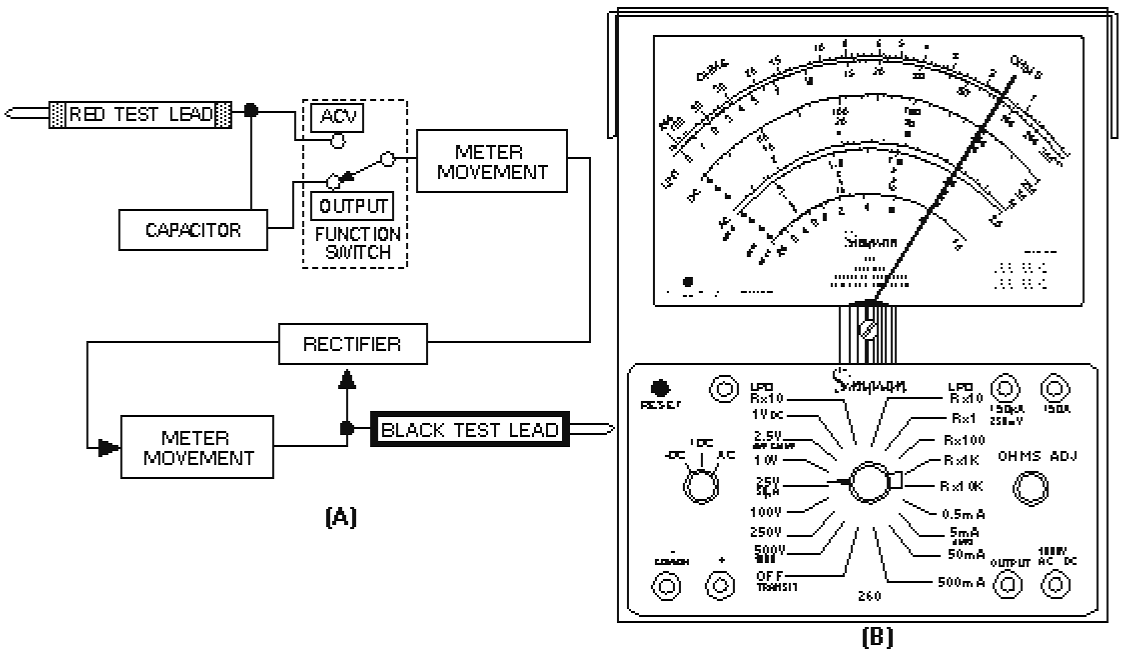


Figure 4-6.—Functional block diagram of ac and output voltage circuits.

Measuring Output Voltages

You will often measure the ac component of an output voltage where both ac and dc voltage levels exist. This occurs primarily in amplifier circuits.

The multimeter has a 0.1-microfarad, 400-volt blocking capacitor in series with the OUTPUT jack. The capacitor blocks the dc component of the current in the circuit under test, but allows the ac component to pass on to the indicating circuits.

CAUTION

When using OUTPUT, do not attempt to use the meter in a circuit in which the dc voltage component exceeds the 400-volt rating of the blocking capacitor.

To use the multimeter to measure output voltage, you must follow these steps:

1. Set the function switch to AC.
2. Plug the black test lead into the COMMON jack and the red test lead into the OUTPUT jack.
3. Set the range switch at the appropriate range position, marked as 2.5V, 10V, 50V, or 250V.
4. Connect the test leads to the component being measured with the black test lead to the negative side of the component.

5. Turn on the power in the test circuit. Read the output voltage on the appropriate ac voltage scale. For the 2.5V range, read the value directly on the scale marked 2.5. For the 10V, 50V, or 250V range, use the red scale marked AC and read the black figures immediately above the scale.

Measuring Current

The multimeter can function as an ammeter to measure current flow.

CAUTION

When using the multimeter as a current-indicating instrument, *NEVER* connect the test leads directly across a voltage. ALWAYS connect the instrument in series with the load.

To use the multimeter as an ammeter, you must take the following steps:

1. Set the function switch at +DC (assuming the current to be positive).
2. Plug the black test lead in the COMMON jack and the red test lead into the + jack.
3. Set the range switch at one of the five ampere-range positions.
4. Ensure the equipment is OFF and then physically open the circuit in which the current is being measured.
5. Connect the VOM in series with the circuit, ensuring that proper polarity is observed when making this connection.
6. Turn the equipment ON and then read the current on the DC scale. (This is the same scale used to measure dc voltages.)

The setting of the range switch determines the maximum value represented by the DC scale. Always use the range scale that corresponds to the range switch setting.

CAUTION

Never attempt to measure currents greater than the setting of the range switch. Increase the range with a shunt, if necessary, but do not exceed the marked current.

When measuring unknown currents, follow the same procedures as when measuring voltages. Always start with the highest range available and work down. Use the range that gives approximately half-scale deflection. If this procedure isn't followed, the meter could be burned out.

Figure 4-7 is a functional block diagram of the dc current circuits in a multimeter.

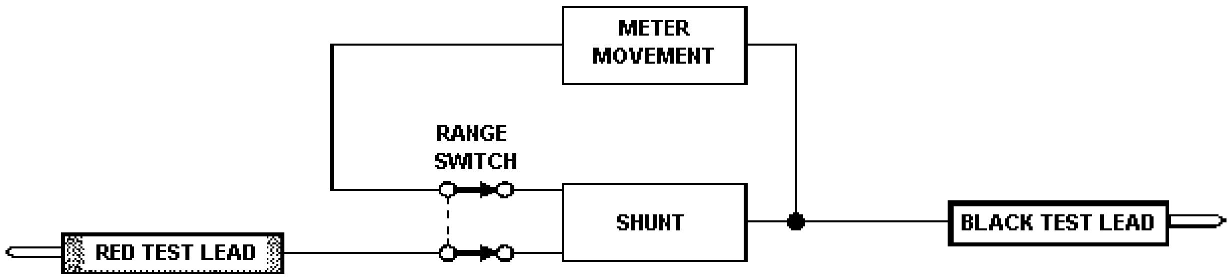


Figure 4-7.—Functional block diagram of dc current circuits.

Accessories

A dc high-voltage probe is available for use with the multimeter. The probe extends the range of the multimeter in a safe and convenient manner. It is primarily used to measure high-voltage, low-power, dc-current sources, such as the anode supplies in television receivers and other cathode-ray tube circuitry.

CAUTION

Do not use this probe on electrical equipment that can deliver high power under short-circuit conditions, such as from a large dc motor-generator set.

Also available is an ac high-voltage probe. The 10,000-volt ac probe is similar to the high-voltage dc probe with the following exceptions:

- The ac high-voltage probe is designed primarily to extend the range of a 5,000-ohms-per-volt VOM.
- The probe is used with the VOM in the 10V AC position.
- You take readings on the 0-10V AC scale and multiply by 1,000.

ELECTRONIC DIGITAL MULTIMETER

As you studied in chapter 3 (externally excited meters), placing a meter into a circuit causes energy to be taken from the circuit. The amount of energy taken depends on the sensitivity of the meter. In some cases, this energy loss cannot be tolerated. For example, in extremely sensitive circuits, such as oscillator grid circuits and automatic volume control circuits, degradation of normal circuit operation will occur. This often results in failure to obtain a usable indication of the fault. The use of electronic multimeters is practical in these sensitive electronic circuits. The higher the input impedance of a meter, the less the loading effect and the more accurate the measurements taken. Electronic multimeters have considerably greater input impedances than do nonelectronic multimeters.

One example of a typical electronic multimeter in use within the Navy is the electronic Model 8000A Digital Multimeter. Most electronic digital multimeters overcome the disadvantage of requiring a continuous external power source by combining an external ac source with an internal rechargeable battery. Another advantage of this meter is that it can be read directly and does not use a scale. Figure 4-8 shows the model 8000A multimeter.

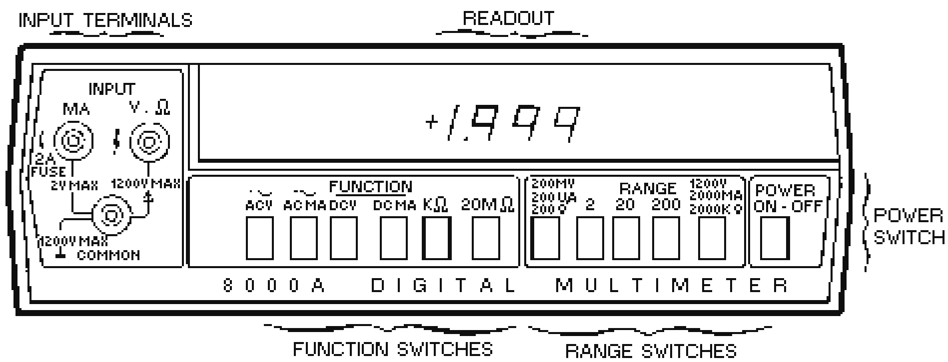


Figure 4-8.—Digital voltmeter 8000A operating features.

Operating Features

The locations of all controls, connectors, and indicators are shown in figure 4-8. The INPUT terminals, located on the left-hand side of the meter face, provide input connections for voltage or resistance (V-?) and milliampere current (MA) measurements with respect to the common terminal. The readout section, located across the upper half of the meter face, contains light-emitting diode (LED) indicators. They display the measured input and polarity signs for dc measurements. The POWER switch, located on the lower right-hand side of the meter face, is a push-button switch used to energize the instrument. The RANGE switches, located on the lower, middle, right-hand portion of the meter face, select the voltage (200 millivolts, 2, 20, 200, or 1,200 volts), current (200 microamperes, 2, 200, or 2,000 milliamperes), and resistance (200 ohms, 2, 20, 200, or 2,000 kilohms) ranges. The FUNCTION switches, located on the lower, middle, left-hand portion of the meter face, select the voltage, current, or resistance modes. The MA input terminal is also a fuse holder for the current protection fuse.

Internal Battery Models

Power is supplied by internal rechargeable batteries that allow the instrument to operate for at least 8 hours. Recharging the batteries is accomplished by switching the POWER switch to OFF and connecting the instrument to an ac power line. You can use the instrument when recharging the batteries on ac power, but the recharging time will be extended.

Q-5. Power for the electronic digital multimeter is normally supplied by what internal power source?

Overload Protection

An overload condition is indicated by the simultaneous flashing of the display readouts. The dc voltage function can withstand up to 1,200 volts dc or 1,200 volts root-mean-square (rms) on any range.

Q-6. How is an overload condition indicated by the electronic digital multimeter?

The ac voltage function can sustain up to 1,200 volts rms on the 20-, 200-, and 1,200-volt ranges and 500 volts rms on the 200-millivolt and 2-volt ranges. The current input fuse is protected above 2 amperes rms. Protection for the resistance function is to 130 volts rms in the 200-ohm and 2-kilohm ranges, and 250 volts rms in the 20-kilohm through 20-megohm ranges.

Basic Digital Multimeter Measurement

Table 4-1 lists the proper function push buttons, range push buttons, and input terminal connections for performing specific measurements with the model 8000A.

Table 4-1.—Basic Measurement Instructions

MEASUREMENT	FUNCTION	RANGE	INPUT CONNECTION	MAXIMUM OVERLOAD	REMARKS
DC Volts	DCV	200MV, 2, 20, 200, or 1200V	V—Ω and COMMON	1200V dc or 1200V rms (sinusoidal)	Auto-polarity
DC Milliamperes	DC MA	200µA, 2, 20, 200, or 2000MA	MA and COMMON	2A (fuse protected)	
AC Volts	ACV	200MV, 2, 20, 200, or 1200V	V—Ω and COMMON	1200V rms (sinusoidal), not to exceed 107 V-Hz on 20, 200, 1200V ranges. 500V rms (sinusoidal) on 200mV and 2V ranges.	
AC Milliamperes	AC MA	200µA, 2, 20, 200, or 2000MA	MA and COMMON	2A (fuse protected)	
Kilohms	KΩ	200Ω, 2, 20, 200, or 2000KΩ	V—Ω and COMMON	130V rms, 200Ωand 2KΩ ranges. 250V rms, 20kΩ thru 2000kΩ ranges.	
Megohms	20MΩ	Any	V—Ω and COMMON	250V rms	Ranges switches non-functional

Block Diagram Analysis

Figure 4-9 is a block diagram of an electronic digital multimeter. Note that the block diagram divides the instrument into three major sections: the SIGNAL CONDITIONING section, the ANALOG-TO-DIGITAL CONVERTER section, and the DISPLAY section.

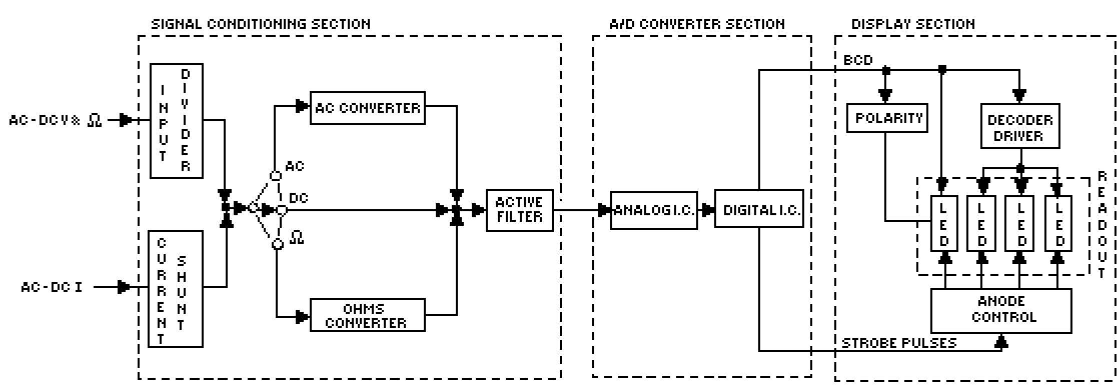


Figure 4-9.—Model 8000A block diagram.

The signal conditioning section provides a dc analog voltage, characteristic of the applied input, to the analog-to-digital converter section. This task is accomplished by the input voltage divider, current shunts, ac converter, active filter, and associated switching.

The analog-to-digital (a/d) converter section changes the dc output voltage from the signal conditioning section to digital information. The a/d converter uses a voltage-to-frequency conversion technique. A dc voltage at the input of the a/d converter is changed to a frequency by the analog integrated circuit (ic). This frequency is characteristic of the magnitude and polarity of the dc input voltage. Counting of the output frequency from the analog ic is accomplished by the digital ic. The resulting count is transferred in binary format to the display section. (Binary number systems are covered in NEETS, Module 13, *Introduction to Number Systems, Boolean Algebra, and Logic Circuits*.)

The display section takes the digital (binary) information from the a/d converter section, decodes it, and visually displays it. The decoded digital information is displayed on numerical LED readouts.

Q-7. In an electronic digital multimeter, the digital information is displayed by what type of numerical readouts?

Accessories

Several accessories are available for use with the electronic digital multimeter. One accessory is the test lead kit, shown in figure 4-10. The kit contains two color-coded test leads with threaded adapters. These adapters attach to banana plugs, pin tips, test prod tips, alligator clips, and binding post lugs.

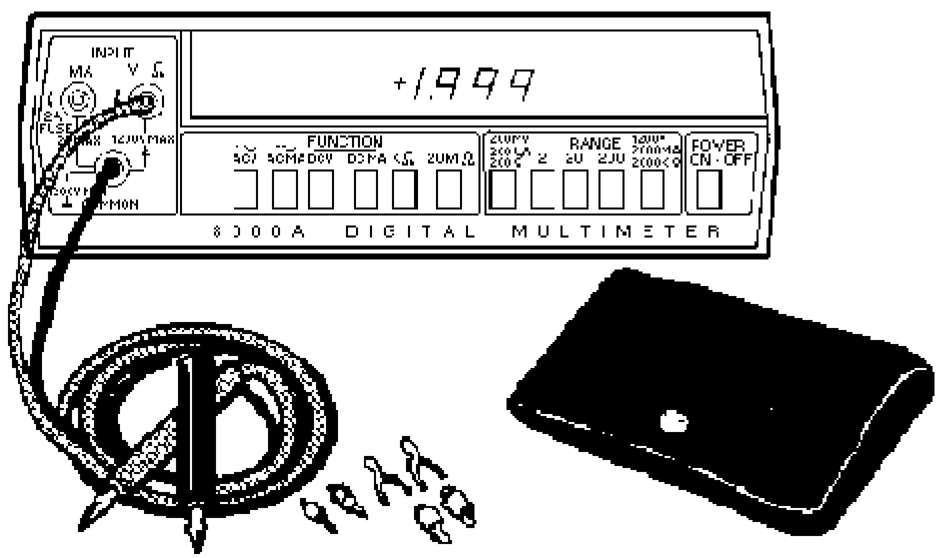


Figure 4-10.—Test lead kit.

Figure 4-11 shows a high-current probe. This probe extends the ac current measurement capability from 2 to 600 amperes at frequencies up to 400 hertz.

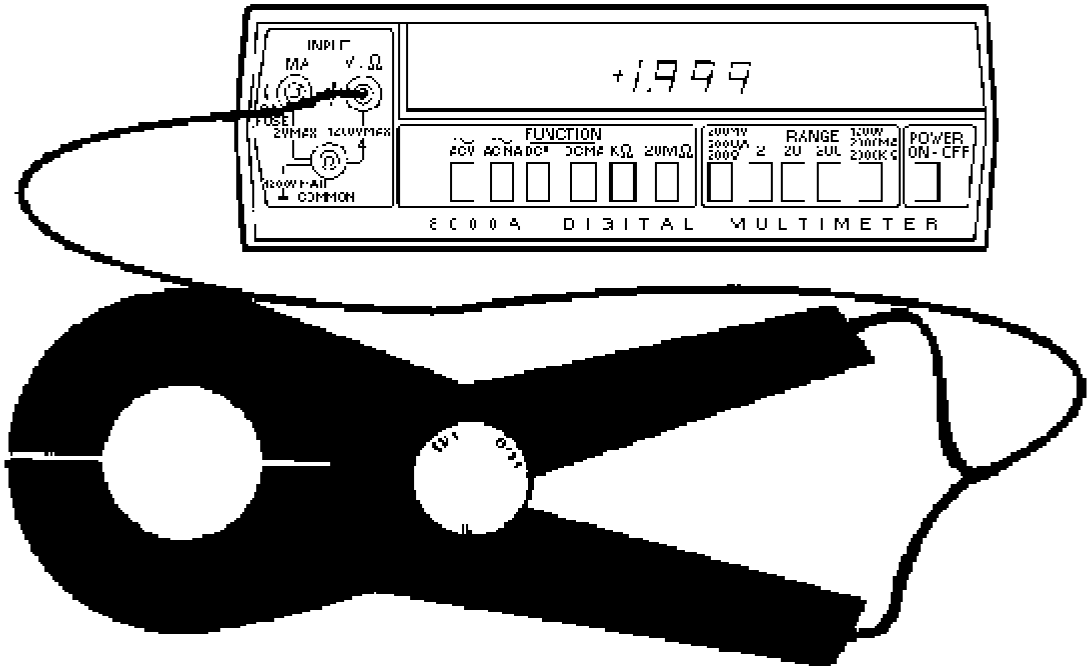


Figure 4-11.—Ac high-current probe.

Figure 4-12 shows a high-voltage probe. The probe extends the dc voltage range to 30 kilovolts.

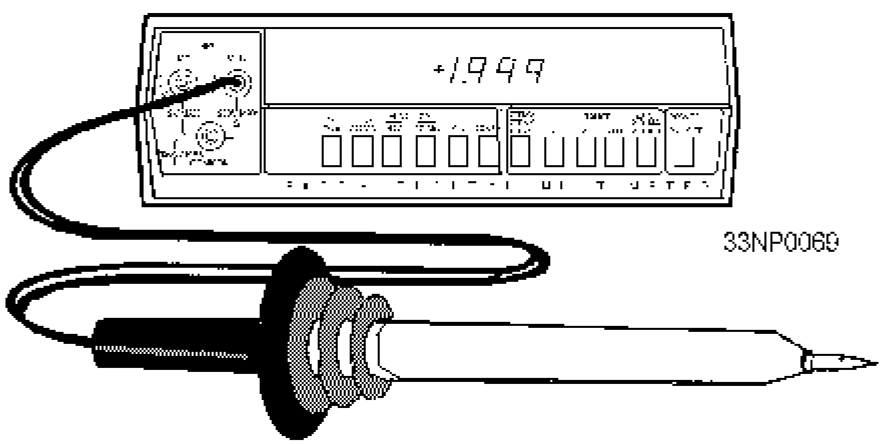


Figure 4-12.—High-voltage probe.

Figure 4-13 shows a high-frequency probe, which allows measurements over a frequency range of 10 kilohertz to 500 megahertz.

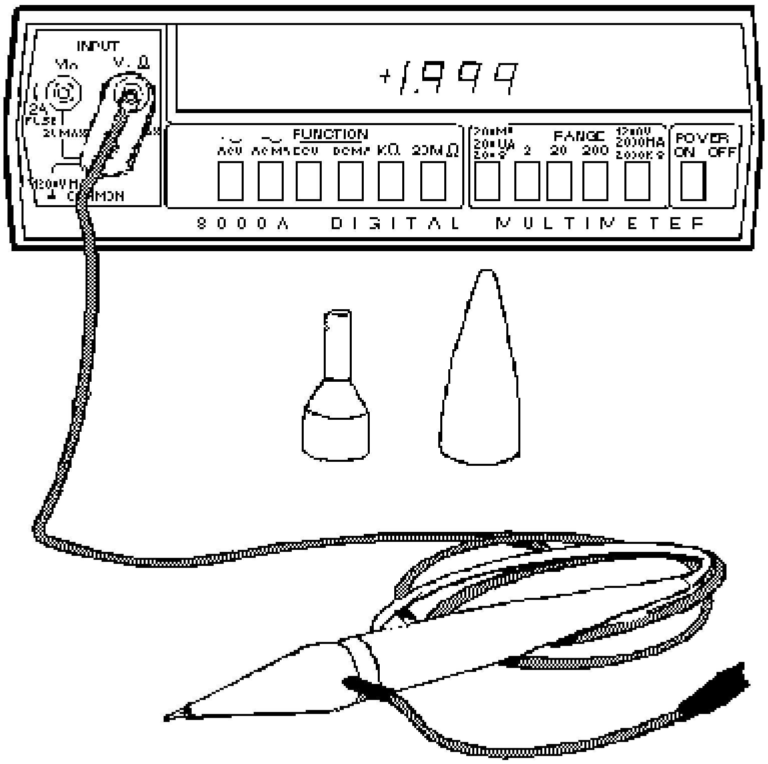


Figure 4-13.—High-frequency probe.

AC/DC DIFFERENTIAL VOLTMETER

The DIFFERENTIAL VOLTMETER provides extremely accurate voltage measurements and is a highly reliable piece of precision test equipment. Its general function is to compare an unknown voltage with a known internal reference voltage and to indicate the difference in their values. The differential voltmeter in common use in the Navy today is the model 893A (figure 4-14).

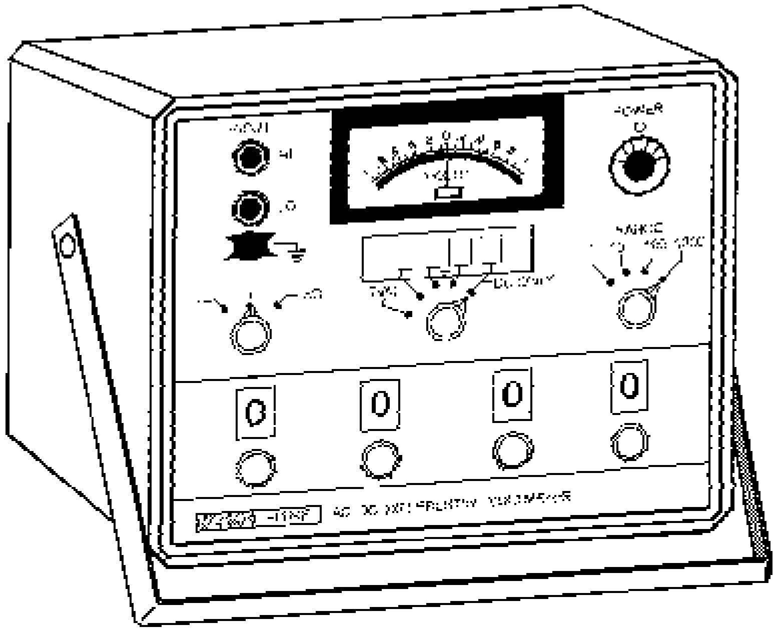


Figure 4-14.—Ac/dc differential voltmeter.

Q-8. What is the general function of the differential voltmeter?

The differential voltmeter can be used as a conventional TRANSISTORIZED ELECTRONIC VOLTMETER (TVM) and a DIFFERENTIAL NULL VOLTMETER. It can also be used to measure variations of a voltage near some known value (NULL DETECTOR), high resistance values (MEGGOMETER), and for dBm measurements.

METER DESIGN CHARACTERISTICS

The differential voltmeter is a solid-state instrument that provides the capability of making dc voltage measurements from +/− 10 microvolts to +/− 1,100 volts. Ac voltages from 0.001 to 1,100 volts can be measured over a frequency range from 5 hertz to 100 kilohertz. Both of these measurements can be made without concern for loading the circuit. The differential voltmeter has four voltage readout dials that vary the resistance of the divider assembly as described above.

The differential voltmeter uses a built-in NULL DETECTOR to measure an unknown voltage. The meter circuitry compares the unknown voltage to a known, adjustable reference voltage supplied by the meter. The reference voltage is provided by a high-voltage dc power supply and decade resistor divider assembly strings that are set by voltage readout dials. In this way, the output from the high-voltage power supply can be precisely divided into increments as small as 10 microvolts. The readout dials are used to adjust the meter pointer to 0 and the unknown voltage is then read from the voltage dials.

A primary feature of the differential voltmeter is that it does not draw current from the unknown source for dc measurements when the measurement is obtained. Therefore, the determination of the unknown dc potential is independent of its source.

FRONT PANEL CONTROLS

The front panel of a typical differential voltmeter is shown in figure 4-15. With a few differences, the controls and terminals are similar to those used on other differential voltmeters. The NULL SENSE switch selects the conventional TVM mode of operation and the various full-scale null detector sensitivity ranges when the instrument is operated in the differential mode of operation. The RANGE switch allows selection of the desired input voltage range, positions the readout dial decimal point, and selects the various ranges of the NULL SENSE switch. The readout dials provide a digital readout of the measured voltage when the instrument is in the differential mode.

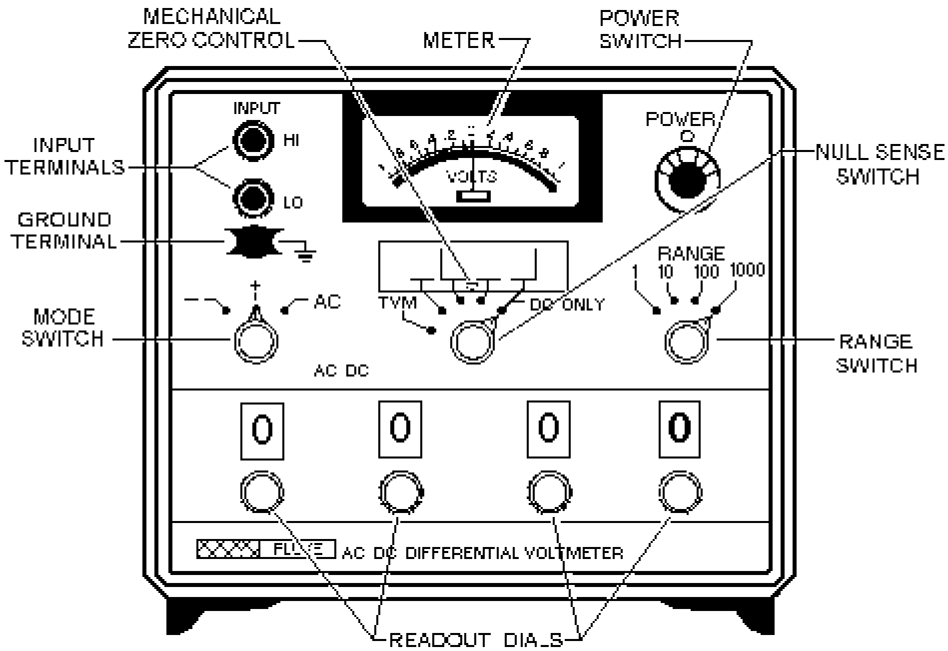


Figure 4-15.—Controls, terminals, and indicators.

MODES OF OPERATION

There are two primary modes of operation: the conventional transistorized voltmeter mode and the differential null mode. These modes are described in the next paragraphs.

Conventional Transistorized Voltmeter (TVM) Mode

When the instrument is used as a conventional transistorized dc voltmeter, the circuitry is connected as shown in figure 4-16. The null detector drives the front panel meter and provides a full-scale meter deflection for any full-scale input. Positive or negative voltage measurements are made by reversing the meter terminals through the contacts of the MODE switch.

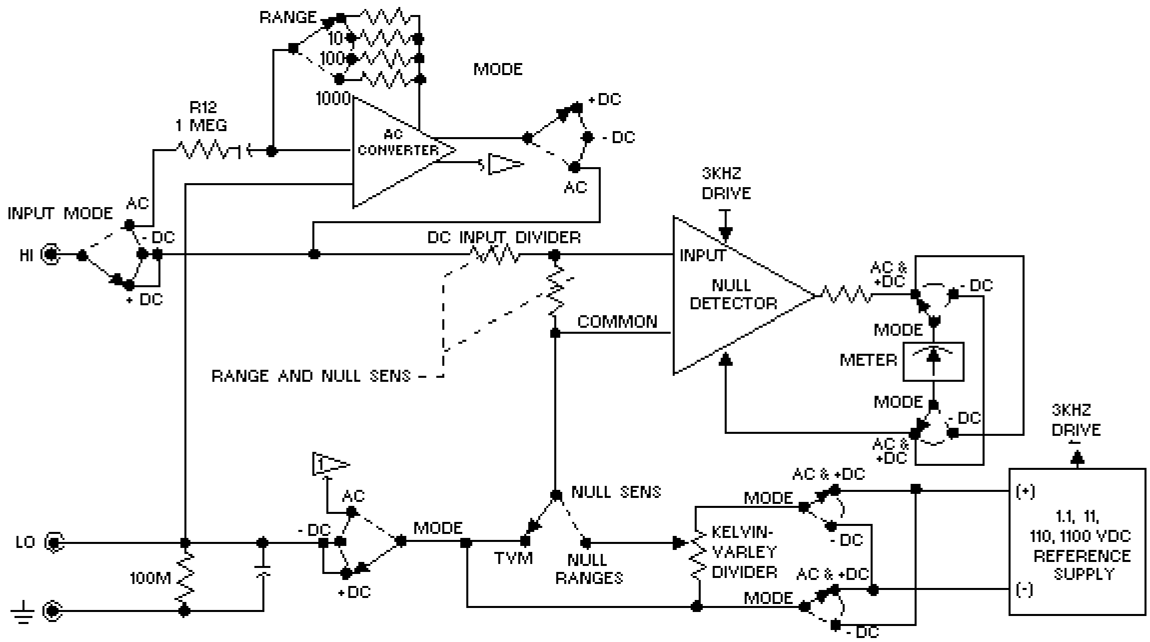


Figure 4-16.—Ac/dc differential voltmeter block diagram.

FUNCTIONAL BLOCK DIAGRAM

Figure 4-16 is a functional block diagram of a differential voltmeter. The circuitry is made up of a reference supply, a resistive divider, a dc input divider, an ac converter, a null detector, and a meter. The circuitry is interconnected by various switching arrangements when you perform the desired ac or dc conventional or differential voltage measurements.

Placing the MODE switch in figure 4-16 to the AC position connects the instrument circuitry as a conventional transistorized ac voltmeter. A full-scale input voltage at the input terminals of the instrument results in a voltage being applied to the input of the null detector. The null detector drives the front panel meter that indicates the value of the measured ac voltage.

Differential Null Mode

When the instrument is used as a dc differential voltmeter, the MODE and NULL SENS switches in figure 4-16 are placed to their respective +/– dc and desired full-scale meter sensitivity positions. In this mode of operation, the NULL SENS switch selects a suitable resistance value to determine the full-scale sensitivity of the meter. The dc input voltage applied to the instrument is then compared with the null detector, and any resulting difference is used to drive the meter. The meter terminals can be reversed through the contacts of the MODE switch for +/– dc voltage measurements.

TRANSISTOR TESTERS

Laboratory transistor test sets are used in experimental work to test characteristics of transistors. For maintenance and repair, however, checking all transistor parameters is not necessary. A check of two or three performance characteristics is usually sufficient to determine whether a transistor needs to be replaced.

Two of the most important parameters used for transistor testing are the transistor CURRENT GAIN (BETA) and the COLLECTOR LEAKAGE or REVERSE CURRENT (ICO). Two other tests that can be accomplished include the electrode resistance and diode measurements. You may want to review NEETS, Module 7, *Introduction to Solid-State Devices and Power Supplies*, for a review of transistors before continuing this section.

The Semiconductor Test Set AN/USM-206A (figure 4-17) is a rugged, field-type tester designed to test transistors and semiconductor diodes. The set will measure the beta of a transistor, the resistance appearing at the electrodes, and the reverse current of a transistor or semiconductor diode. It will also measure a shorted or open condition of a diode, the forward transconductance of a field-effect transistor, and the condition of its own batteries.

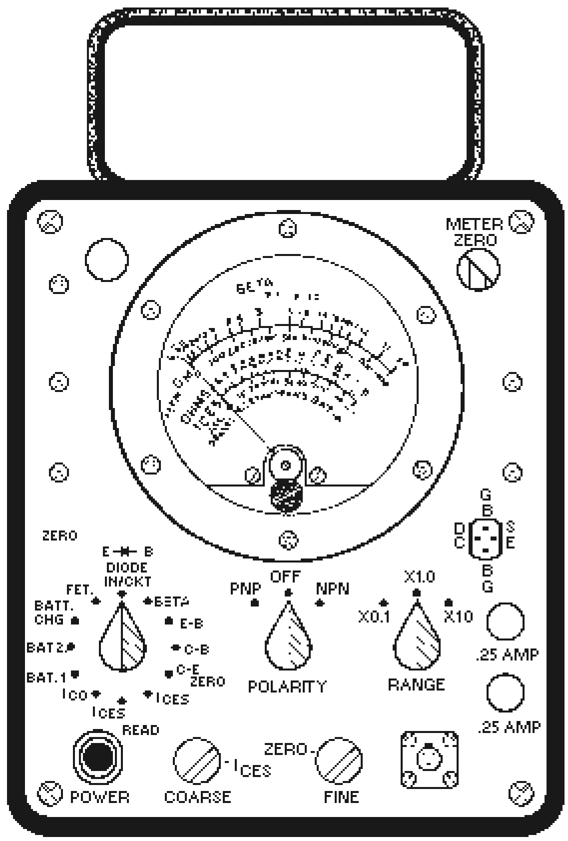


Figure 4-17.—Semiconductor test set.

To assure that accurate and useful information is gained from the transistor tester, you should make the following preliminary checks of the tester before testing any transistors:

1. With the POLARITY switch in the OFF position, the meter pointer should indicate exactly zero. (When required, rotate the meter ZERO ADJUST KNOB on the front of the meter to fulfill this requirement.) To prevent battery drain, be sure to leave the POLARITY switch in the OFF position when measurements are not actually being made.

2. Always check the condition of the test set batteries. To make this check, disconnect the test set power cord, place the polarity switch in the PNP position, and place the function switch first to BAT. 1and then to BAT. 2. In both BAT positions, the meter pointer should move so as to indicate within the red BAT box.

BETA MEASUREMENTS

If the transistor is to be tested out of the circuit, plug it into the test jack located on the right-hand side below the meter. If the transistor is to be tested in the circuit, at least 300 ohms must exist between E-B (emitter to base), C-B (collector to base), and C-E (collector to emitter) for accurate measurement. Initial setting of the test set controls is performed as follows:

1. Set the function switch to BETA.
2. Set the POLARITY switch to PNP or NPN (depending on the type of transistor under test).
3. Set the RANGE switch to X10.
4. Adjust METER ZERO for zero meter indication (transistor disconnected).
5. The POLARITY switch should remain OFF while the transistor is connected to or disconnected from the test set; it should then be set to PNP or NPN, as in step 2 above.

If the beta reading is less than 10, perform the following steps:

1. Reset the RANGE switch to X1 and reset the meter to zero.
2. After connecting the yellow test lead to the emitter, the green test lead to the base, and the blue test lead to the collector, plug the test probe (not shown) into the jack located at the lower right-hand corner of the test set.
3. When testing grounded equipment, unplug the 115-volt line cord and use battery operation. A beta reading is attained by multiplying the meter reading times the RANGE switch setting. Refer to the transistor characteristics book provided with the tester to determine if the reading is normal for the type of transistor under test.

ICO MEASUREMENTS

Adjust the METER ZERO control for a zero meter indication. Plug the transistor to be tested into the jack, or connect the test leads to the device. Set the PNP/NPN switch to correspond with the type of transistor under test. Set the function switch to ICO and the RANGE switch to X0.1, X1.0, or X10, as specified by the transistor data book for allowable leakage. Read leakage on the bottom scale and multiply by the range setting figure as required.

ELECTRODE RESISTANCE MEASUREMENTS

Connect the in-circuit probe test leads to the transistor with the yellow lead to the emitter, the green lead to the base, and the blue lead to the collector. Set the function switch to the OHMS E-B position and read the resistance between the emitter and base electrode on the center scale of the meter marked OHMS.

To read the resistance between the collector and base and the collector and emitter, set the function switch to OHMS C-B and OHMS C-E, respectively. These in-circuit electrode resistance measurements are used to correctly interpret the in-circuit beta measurements. The accuracy of beta times 1 and 10 range is ±15 percent only when the emitter-to-base load is equal to or greater than 300 ohms.

DIODE MEASUREMENTS

Diode in-circuit quality measurements are made by connecting the green test lead to the cathode and the yellow test lead to the anode. Set the function switch to DIODE IN/CKT and the RANGE switch to times 1 position. Ensure that the meter has been properly zeroed on this scale. If the meter reads down-scale, reverse the polarity switch. If the meter reads less than midscale, the diode under test is either open or shorted. The related circuit impedance of this test is less than 25 ohms.

RESISTANCE-CAPACITANCE-INDUCTANCE (RCL) BRIDGES

Resistance, capacitance, and inductance can be measured with precise accuracy by alternating-current bridges. These bridges are composed of capacitors, inductors, and resistors in a wide variety of combinations. These bridges operate on the principle of the Wheatstone bridge; that is, an unknown resistance is balanced against known resistances and, after the bridge has been balanced, the unknown resistance is calculated in terms of the known resistance.

The universal Impedance Bridge, Model 250DE (shown in figure 4-18) is used to measure resistance, capacitance, and inductance (RCL) values. It is also used to make other special tests, such as determining the turns ratio of transformers and capacitor quality tests. This instrument is self-contained, except for a source of line power, and has an approximate 500-hour battery life expectancy. It has its own source of 1,000-hertz bridge current with a sensitive bridge balance indicator and an adjustable source of direct current for electrolytic capacitor and resistance testing. The bridge also contains a meter with suitable ranges to test for current leakage on electrolytic capacitors.

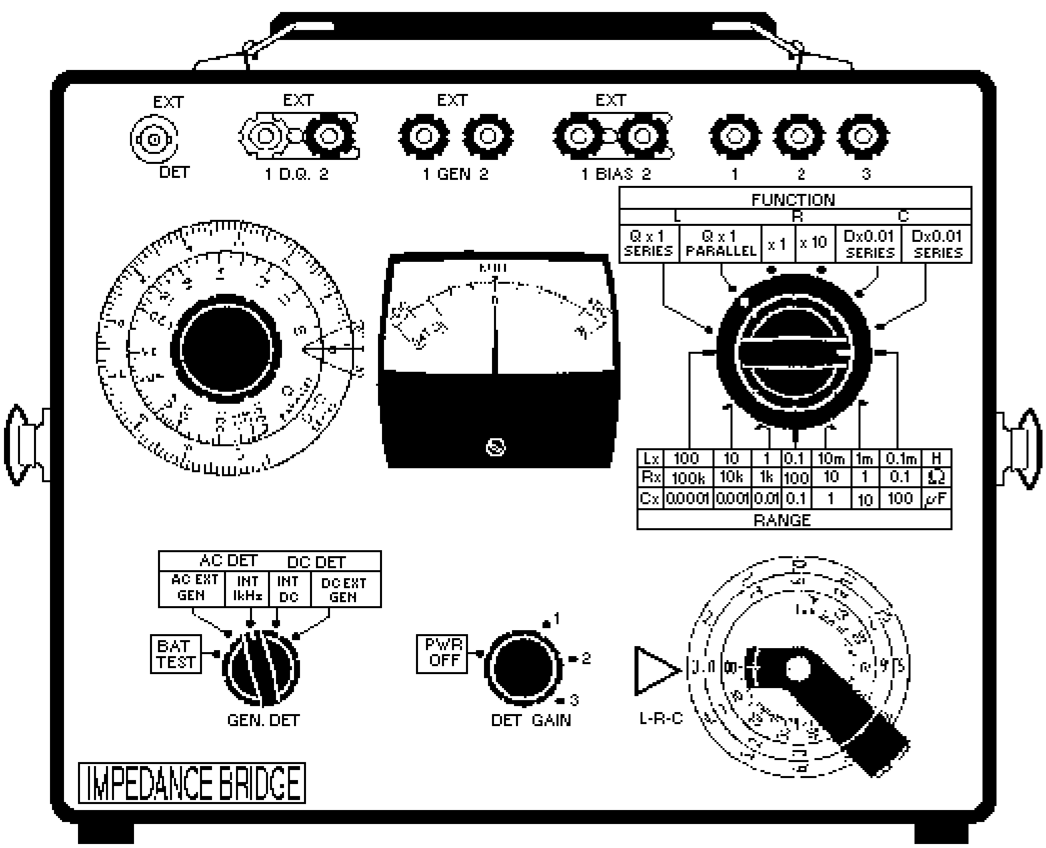


Figure 4-18.—Resistance-capacitance-inductance bridge.

CONTROLS

Figure 4-18 is a panel view of the model 250DE bridge switches, dials, controls, and connections. Refer to the figure as we briefly discuss some of the switches and dials below.

- The FUNCTION switch selects the type of bridge circuit that will measure resistance, capacitance, or inductance.
- The RANGE switch selects the multiplier for each function.
- L-R-C decade dials are a DEKASTAT decade resistor that is the main balancing element of the bridge. The setting of the dials after the bridge is balanced indicates the value of inductance, resistance, or capacitance.
- The D-Q dial is used to balance the phase of the capacitance or inductance of the bridge. The setting of the dial after the bridge is balanced indicates the value of dissipation factor (D) or storage factor (Q).
- The GEN-DET switch selects bridge generator and detector connections, ac or dc, internal or external generator. The switch also connects the internal batteries to the battery test circuit.
- The DET GAIN control adjusts the sensitivity of the ac-dc detector and turns on power to the generator.

CONNECTIONS

L, R, and C terminals 1, 2, and 3 are used to connect unknown resistors, inductors, and capacitors to the bridge. Resistors and inductors are connected between terminals 1 and 2, and capacitors are connected between terminals 2 and 3. **EXT BIAS** terminals are normally connected with a shorting lug. They allow insertion of a dc voltage or current to bias capacitors or inductors. **EXT DET** connector is a BNC coaxial socket that allows an external detector to be used with the instrument. It is connected to the bridge at **ALL TIMES**.

EXT D-Q terminals are normally connected with a shorting lug. They allow an external rheostat to extend the range of the D-Q dial. **EXT GEN** terminals provide a connection to the bridge for an external generator. When the GEN-DET switch is in the AC EXT GEN position, the terminals connect an isolation transformer so that a grounded external generator can be used. When the GEN-DET switch is in the DC EXT GEN position, the terminals are connected directly to the bridge.

BATTERY

The model 250DE bridge has a battery supply consisting of four 1.5 V dc batteries with an expected life of 500 hours. The battery power supply should be checked before each day's operation. Turn DET GAIN control to 1 and set GEN-DET switch to BATT. TEST (battery test). If the meter deflects beyond the BAT OK mark, the battery is good.

RESISTANCE MEASUREMENTS

Resistance is usually measured with direct current for maximum accuracy. The model 250DE bridge can be used to measure resistance with alternating current, but external reactance compensation is usually required. On high-resistance ranges, care should be taken to avoid leakage across a resistor under test. Insulation with a resistance of 109 ohms, which is adequate for most purposes, will cause a measurement error of 1 percent if it shunts a 10-megohm resistor. Using the following steps, you will be able to measure dc resistance ONLY:

1. Turn the DET GAIN control to 2.
2. Set the FUNCTION switch to R × 1 or R × 10.
3. Set L-R-C decade dials to 3.000.
4. Connect the unknown resistor to R-L terminals 1 and 2.
5. Set the GEN-DET switch to INT DC.
6. Adjust the RANGE switch for minimum detector deflection.
7. Adjust L-R-C decade dials for null, turning the DET GAIN control clockwise to increase sensitivity as necessary.
8. The measured resistance is the product of the L-R-C decade dial setting times the RANGE and FUNCTION switch settings.

CAPACITANCE MEASUREMENTS

Capacitance is measured in terms of a two-element equivalent circuit consisting of a capacitor in series with a resistor. The internal ac generator and detector of the model 250DE bridge are tuned to 1 kilohertz. Other frequencies can be used, but an external generator and detector are required. The D and Q ranges of the bridge can be extended by use of an external rheostat connected to the terminals provided. The measured capacitance is the product of the L-R-C dial setting times the setting of the RANGE switch. Using the following steps, you can make a standard capacitance measurement:

1. Turn the DET GAIN control to 1.
2. Set the FUNCTION switch to C, D × 0.1 or D × 0.01 SERIES.
3. Set L-R-C decade dials to 3.000 and D-Q dial to 0.
4. Connect the unknown capacitor to C terminals 2 and 3.
5. Set the GEN DET switch to INT 1 kHz.
6. Adjust the RANGE SWITCH for minimum detector deflection.
7. Adjust L-R-C decade dials and D-Q dial alternately for a minimum meter deflection, turning the DET GAIN control clockwise to increase sensitivity as necessary.
8. The measured capacitance is the product of the L-R-C decade dial settings.
9. The measured dissipation factor (D) is the product of the D-Q setting times the FUNCTION switch setting.

INDUCTANCE MEASUREMENTS

Inductance is measured in terms of a two-element equivalent circuit consisting of an inductance either in series or in parallel with a resistance. The internal ac generator and detector of the model 250DE bridge are tuned to 1 kHz. Other frequencies can be used, but like capacitance measurements, an external generator and detector are required. When inductance is being measured in ac or dc, it should be realized that iron-core inductors are sensitive to current variations. Quantitative measurements of dc effects can be made by supplying current to the unknown inductor through the EXT BIAS terminal. Use the following steps to make inductance measurements:

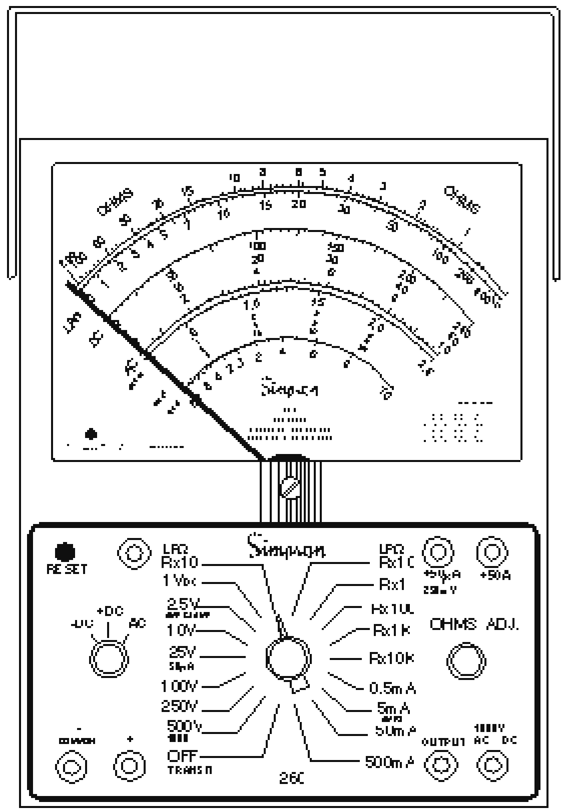
1. Turn the DET GAIN control to 1.
2. Set the FUNCTION switch to L PARALLEL if Q is greater than 10 to L SERIES if Q is less than 10.
3. Set L-R-C decade dials to 3.000 and D-Q dial to maximum.
4. Connect the unknown inductor to the R-L terminals 1 and 2
5. Set the GEN DET switch to INT 1 kHz.
6. Adjust the RANGE switch for minimum detector deflection.
7. Adjust the L-R-C decade dials and D-Q dial alternately for a minimum meter deflection, turning the DET GAIN control clockwise to increase sensitivity as necessary.
8. The measured inductance is the product of the L-R-C decade dial setting times the RANGE switch setting.
9. The measured storage factor (Q) is read directly from the D-Q dial, inner scale for parallel and outer scale for series inductance.

SUMMARY

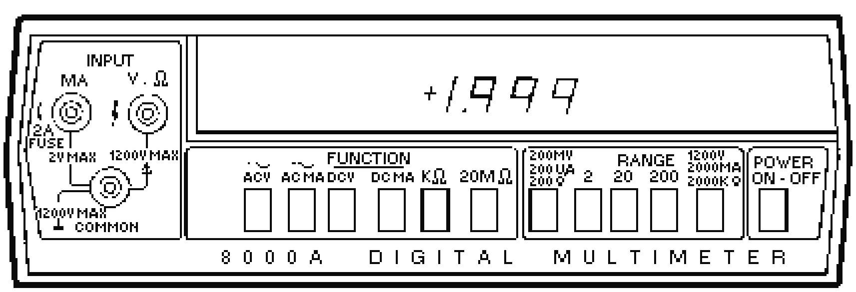
The important points of this chapter are summarized in the following paragraphs. You should be familiar with these points before continuing with your studies of test equipment.

A **MULTIMETER** is a single meter that combines the functions of a dc ammeter, a dc voltmeter, an ac ammeter, an ac voltmeter, and an ohmmeter. Observe the following safety precautions when using a multimeter:

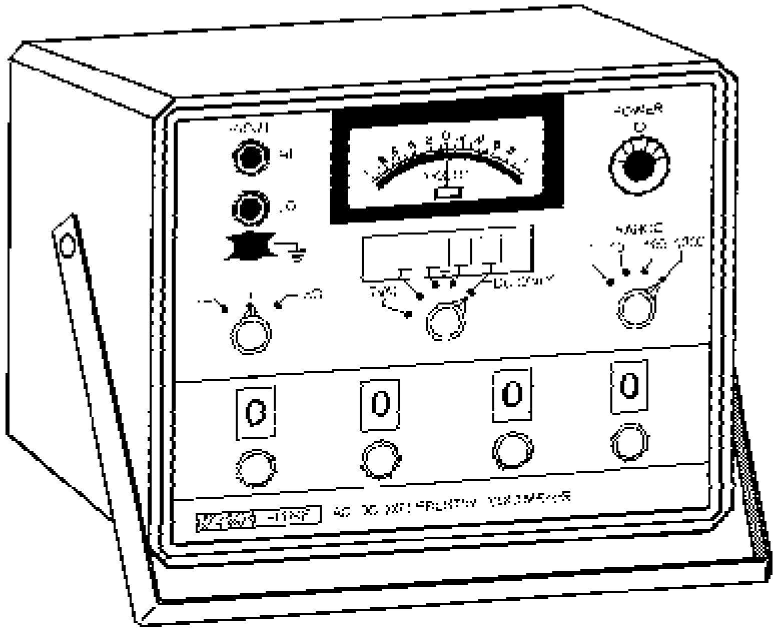
- De-energize and discharge the circuit completely before connecting a multimeter.
- Never apply power to the circuit while you are measuring resistance with an ohmmeter.
- Connect the ammeter **in series** for current measurements and **in parallel** for voltage measurements.
- Be certain the multimeter is switched to ac before attempting to measure ac circuits.
- Observe proper dc polarity when measuring dc circuits.
- Always start with the highest voltage or current range.
- Select a final range that allows a reading near the middle of the scale.
- Adjust the "0 ohms" reading after changing resistance ranges and before making a resistance measurement.



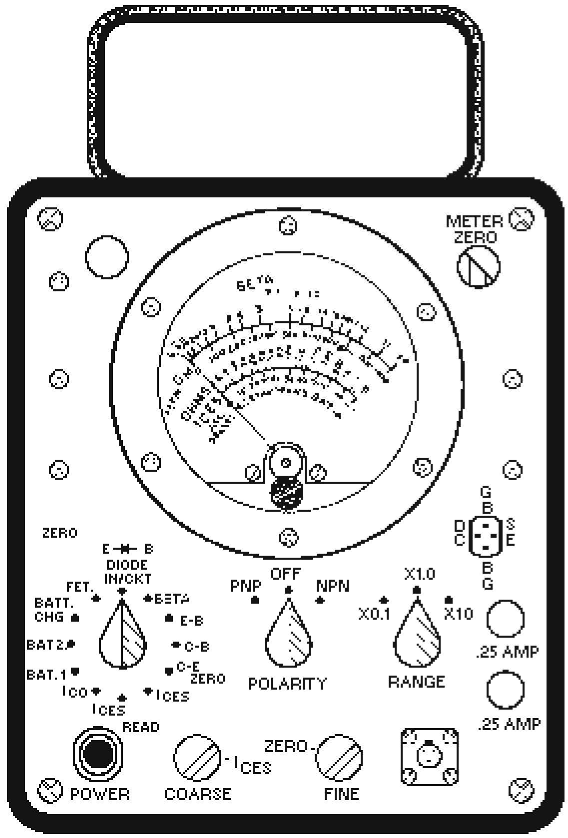
An **ELECTRONIC DIGITAL MULTIMETER** is used in sensitive electronic circuits where only extremely small amounts of energy can be extracted without disturbing the circuits under test, or causing them to be inoperative.



The **DIFFERENTIAL VOLTMETER** is a precision piece of test equipment used to compare an unknown voltage with an internal reference voltage and to indicate the difference in their values.



A **SEMICONDUCTOR TEST SET** is used to measure the beta of a transistor, the resistance appearing at the electrodes, and the reverse current of a transistor or semiconductor diode. It also measures a shorted or open condition of a diode, the forward transconductance of a field-effect transistor, and the condition of its own batteries.



Resistance, capacitance, and inductance are measured for precise accuracy by **RCL BRIDGES**. They are composed of capacitors, inductors, and resistors and operate on the principle of the Wheatstone bridge.

ANSWERS TO QUESTIONS Q1. THROUGH Q8.

A-1. No external power source is required.

A-2. De-energized.

A-3. Midscale.

A-4. Polarity.

A-5. Rechargeable batteries.

A-6. Simultaneous flashing of display readouts.

A-7. Light-emitting diodes.

A-8. To compare an unknown voltage with a known reference voltage and indicate the difference in their values.

CHAPTER 5

SPECIAL-APPLICATION TEST EQUIPMENT

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to:

1. Explain the theory of operation of two types of power meters.
2. Describe the purpose of the controls and indicators found on power meters.
3. Describe the proper procedure for taking power measurements for incident and reflected energy.
4. Describe the uses and purposes of the controls and indicators found on the signal generator.
5. Explain the theory of operation of a typical frequency counter.
6. Describe the uses and purposes of the controls and indicators found on the frequency counter.
7. Explain the uses and purposes of the controls and indicators found on the Huntron Tracker 2000.
8. Describe the proper procedures for troubleshooting with a logic probe.
9. Describe the proper procedures for troubleshooting using the Huntron Tracker 2000.

INTRODUCTION

In chapters 3 and 4, you studied the more common pieces of test equipment. As a technician, you will routinely use this test equipment to troubleshoot and perform maintenance on electronic equipment. However, the equipments you will study in this chapter may or may not be found in your shop. This is because these equipments have specific or specialized uses. Unless your rating is involved with the equipment with which they are used, you may not have reason to use them. They are presented here so that you will be familiar with their overall function should the need arise. The equipments you will study in this chapter are power meters, signal generators, frequency counters, and integrated circuit-testing devices.

POWER METERS

As a technician, you will use a POWER METER to measure power. There are various types of power meters, some of which are called WATTMETERS. Figure 5-1 shows the AN/URM-120 wattmeter, which is one type of power meter commonly used in the Navy. This particular power meter measures power directly; that is, you connect it directly between the transmitter output (rf source) and the load, most likely an antenna.

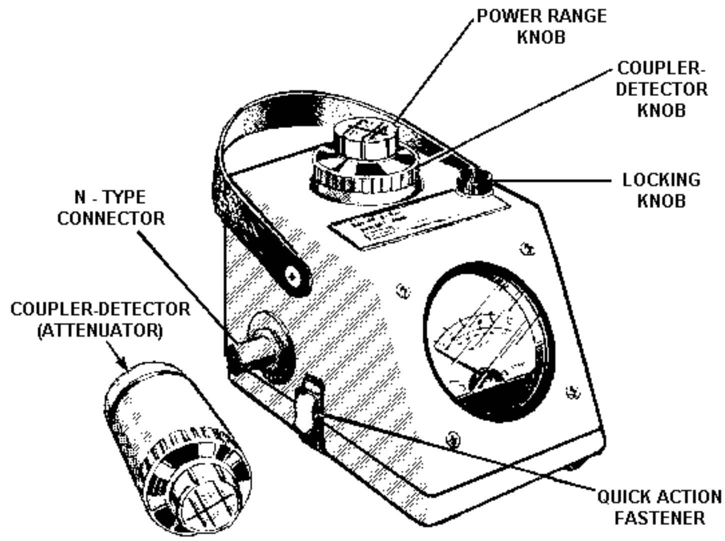


Figure 5-1.—Wattmeter (AN/URM-120).

Other types of power meters measure power indirectly; that is, they sample power in other ways—but not by being placed directly between the output of the transmitter and the load. Let's discuss the direct-measuring power meter first; then we'll talk about an indirect-measuring power meter.

DIRECT-MEASURING POWER METERS

The direct-measuring power meter is designed to measure incident (forward) and reflected (reverse) rf power from 50 to 1,000 watts at 2 MHz to 30 MHz and 10 to 500 watts at 30 MHz to 1,000 MHz. Three separate COUPLER-DETECTORS (sometimes called ATTENUATORS), each rated to cover a portion of the frequency and power ranges, are provided with the wattmeter. These devices couple the rf signal into the wattmeter and detect the signal. The coupler-detector knob projects through the top of the wattmeter case, as shown on the AN/URM-120 wattmeter in figure 5-1.

A nameplate on the top of the POWER RANGE knob indicates the power range. The POWER RANGE knob can be rotated 360º to the desired power range. The coupler-detector rotates 180º inside the metal case for either forward or reverse power measurements. Also located inside the metal case are the indicating meter and cable for interconnecting the meter to the coupler-detector. The LOCKING knob locks the coupler-detector and POWER RANGE knobs in place.

Two N-TYPE connectors (one male and one female) are located on either side of the wattmeter case to connect the instrument between the power source and the load. The upper and lower parts of the wattmeter are held together with quick-action fasteners, which permit easy access to the inside of the wattmeter.

Power measurements are made by inserting the proper coupler-detector and connecting the wattmeter in the transmission line between the load and the rf power source. To measure *incident power* with the wattmeter, rotate the arrow on the COUPLER-DETECTOR knob toward the load, and position the POWER RANGE knob for peak meter reading. To measure *reflected power*, position the arrow toward the rf power source.

In effect, rotating the coupler-detector causes the coupler to respond only to a wave traveling in a particular direction, either *to* (incident) or *from* (reflected) the load. It will be unaffected by a wave traveling in the opposite direction. A diode rectifier in the coupler rectifies the energy detected by the coupler. This detected rf energy is measured across a known impedance to obtain either incident or reflected power.

Operating the Wattmeter

Always de-energize and tag the rf power source before measuring *incident* power. Insert the proper coupler-detector for the rf power being measured into the wattmeter case. Remove the wire shunt (not shown in figure 5-1) from the meter terminals, then connect the wattmeter into the transmission line, either at the load or the rf source. Ensure that all connections are tight.

Position the POWER RANGE knob to a value higher than the rated power of the rf source.

CAUTION

If the rated power to be measured is not known, place the POWER RANGE knob in the highest power position before turning on the power source.

Rotate the coupler-detector so that the arrow indicating power flow points toward the load. Turn on the rf power source. Rotate the POWER RANGE knob to the proper range for measuring and observe the point at which the indicating meter peaks.

Q-1. To measure incident power, you must rotate the coupler-detector of the wattmeter so that the arrow indicating power flow points toward which end of the transmission line?

Reflected power is measured in the same manner as described for incident power, except that the coupler-detector is rotated so that the arrow points toward the rf source.

After completing power measurements, de-energize the rf source, disconnect the wattmeter from the transmission line, and place the wire shunt on the meter terminals.

Interpreting Power Measurements Made by the Wattmeter

The rf power measurements made by the wattmeter are used to determine the voltage standing wave ratio (VSWR) of the load and the power absorbed by the load. (VSWR is covered in NEETS, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*.) The VSWR can be determined from a chart provided in the wattmeter technical manual, or it can be calculated (as shown in the following example for a UHF transmitter) by the formula below (Pi is the incident power, and Pr is the reflected power as measured by the wattmeter):

Where:

Pi = 30 watts

Pr = 0.5 watts

$$
\begin{aligned}
\text{VSWR} &= \frac{\sqrt{P_i}+\sqrt{P_r}}{\sqrt{P_i}-\sqrt{P_r}} \\
&= \frac{\sqrt{30}+\sqrt{0.5}}{\sqrt{30}-\sqrt{0.5}} \\
&= \frac{5.47+0.71}{5.47-0.71} \\
&= \frac{6.18}{4.76} \\
&= 1.30
\end{aligned}
$$

The example above results in a standing wave ratio expressed as 1.3 to 1. In a perfectly matched transmission line where there is no reflected power ($P_r = 0$), the standing wave ratio would be 1 to 1. A standing wave ratio of 1.5 to 1 indicates a 5-percent reflection of energy (loss) and is considered to be the maximum allowable loss. So, our example is within allowable limits.

If the standing wave ratio is greater than 1.5 to 1, then the transmission line efficiency has decreased and troubleshooting is necessary. An excellent discussion of the reasons for standing wave ratio increases is presented in EIMB, *Test Methods and Practices*, NAVSEA 0967-LP-000-0130.

You can determine the rf power absorbed by the load simply by subtracting the reflected power reading from the incident power reading made by the wattmeter (30 watts - 0.5 watts = 29.5 watts).

The power meter just discussed is often described as an IN-LINE POWER METER because readings are taken while the power meter is connected in series with the transmission line. Another type of power meter used by the Navy measures power indirectly. An example of an indirect-measuring power meter is described in the next section.

INDIRECT-MEASURING POWER METERS

An example of an indirect-measuring power meter is the HP-431C, shown in figure 5-2. The controls, connectors, and indicators for the power meter are illustrated in figure 5-3. This power meter can be operated from either an ac or dc primary power source. The ac source can be either 115 or 230 volts at 50 to 400 hertz. The dc source is a 24-volt rechargeable battery. Overall circuit operation of the power meter is shown in the block diagram in figure 5-4.

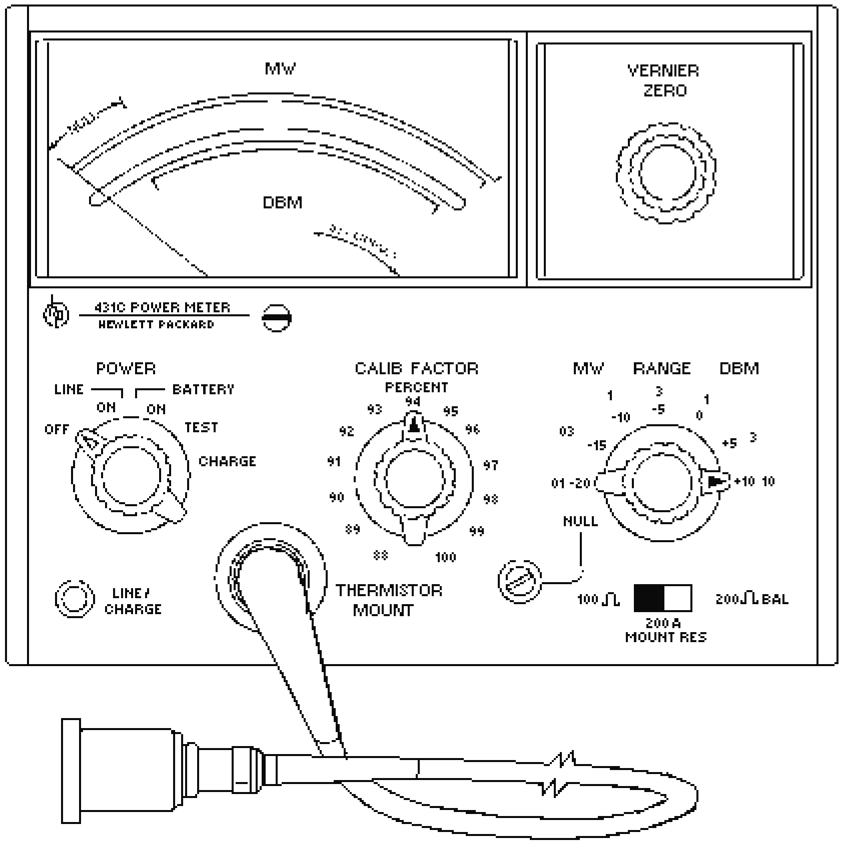


Figure 5-2.—Power meter (HP-431C).

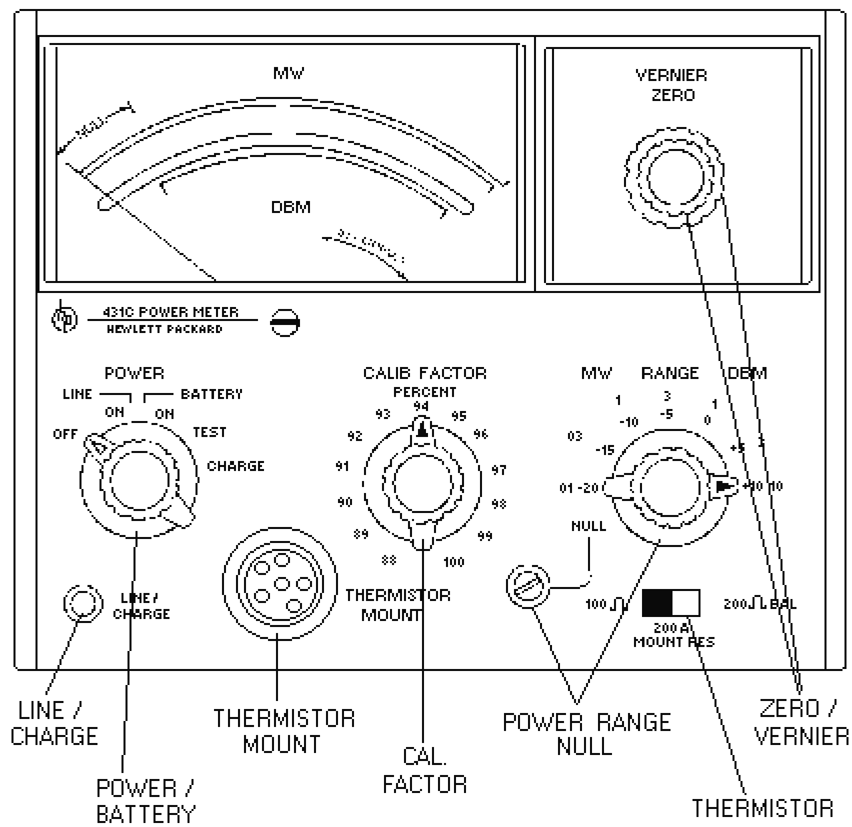


Figure 5-3.—Power meter controls, indicators, and connectors.

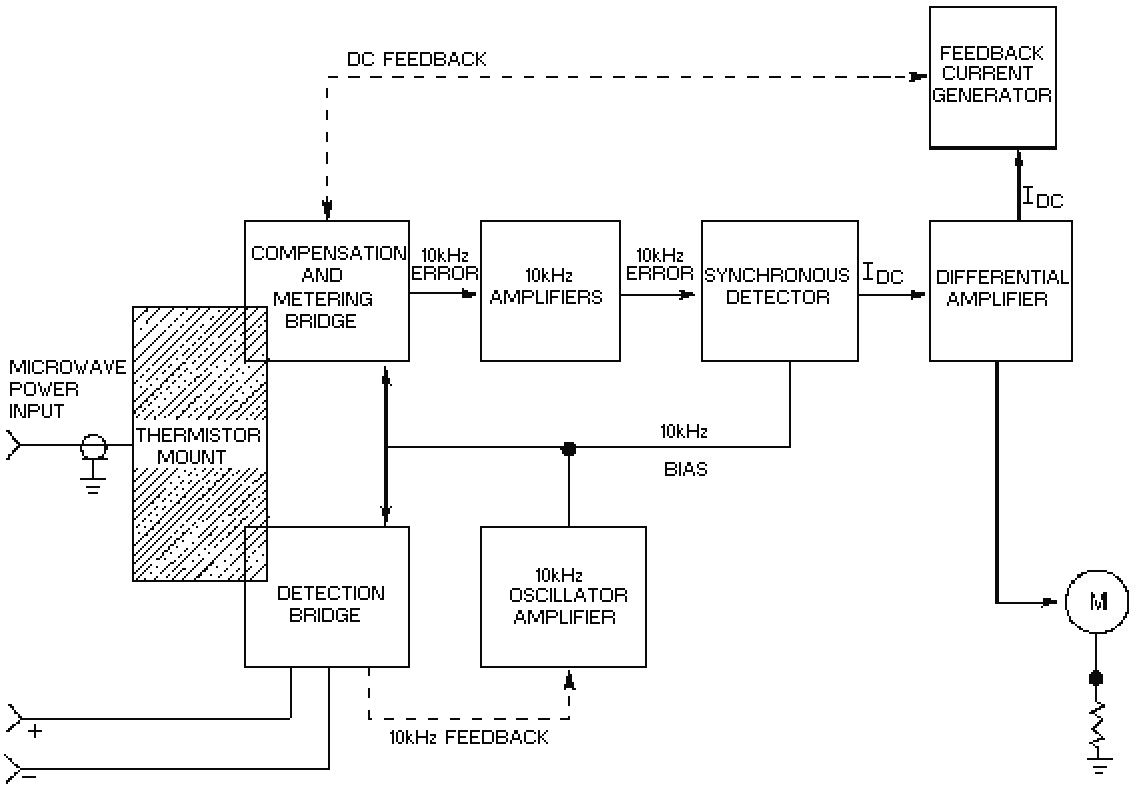


Figure 5-4.—Power meter block diagram.

The HP-431C power meter *indirectly* measures microwave frequency power by using two bridge circuits—the detection bridge and the compensation and metering bridge. The detection bridge incorporates a 10-kilohertz (kHz) oscillator in which the amplitude is determined by the amount of heating of the thermistors in that bridge caused by microwave power. (Thermistors were covered in chapter 2 of this module.) The compensation and metering bridge contains thermistors that are affected by the same microwave power heating as those of the detection bridge.

An unbalance in the metering bridge produces a 10-kHz error signal. This error signal, plus the 10-kHz bias that is taken directly from the 10-kHz OSCILLATOR-AMPLIFIER, is mixed in the SYNCHRONOUS DETECTOR. The synchronous detector produces a dc current (I_{dc}) that is proportional to the 10-kHz error signal. The I_{dc} error signal is fed back to the compensation and metering bridge, where it substitutes for the 10-kHz power in heating the thermistor and drives the bridge toward a state of balance. The dc output of the synchronous detector also operates the meter circuit for a visual indication of power.

Q-2. What condition produces the 10-kHz error signal generated by the metering bridge in the HP-431C power meter?

The HP-431C power meter measures rf power from 10 microwatts (–20 dBm) to 10 milliwatts (+10 dBm) full scale in the 10 MHz to 18 GHz for a 50-ohm coaxial system and 2.6 GHz to 40 GHz for a waveguide system.

SIGNAL GENERATORS

Standard sources of ac energy, both audio frequency (af) and radio frequency (rf), are often used in the maintenance of electronic equipment. These sources, called SIGNAL GENERATORS, are used to test and align all types of transmitters and receivers. They are also used to troubleshoot various electronic devices and to measure frequency.

The function of a signal generator is to produce alternating current (ac) of the desired frequencies and amplitudes with the necessary modulation for testing or measuring circuits. (Modulation was discussed in NEETS, Module 12, *Modulation Principles*.) It is important that the amplitude of the signal generated by the signal generator be correct. In many signal generators, output meters are included in the equipment to adjust and maintain the output at standard levels over wide ranges of frequencies.

When using the signal generator, you connect the output test signal into the circuit being tested. You can then trace the progress of the test signal through the equipment by using electronic voltmeters or oscilloscopes. In many signal generators, calibrated networks of resistors, called ATTENUATORS, are provided. You use attenuators in signal generators to regulate the voltage of the output signal. Only accurately calibrated attenuators can be used because the signal strength of the generators must be regulated to avoid overloading the circuit receiving the signal.

Q-3. In signal generators, what device is used to regulate the voltage of the output signal?

There are many types of signal generators. They are classified by use and the frequency range covered as AUDIO-FREQUENCY (AF) GENERATORS, VIDEO SIGNAL GENERATORS, RADIO-FREQUENCY (RF) GENERATORS, FREQUENCY-MODULATED RF GENERATORS, and other special types, which combine frequency ranges.

AUDIO AND VIDEO SIGNAL GENERATORS

AUDIO SIGNAL GENERATORS produce stable af signals used for testing audio equipment. VIDEO SIGNAL GENERATORS produce signals that include the audio range and extend into the rf range. These signal generators are used to test video amplifiers and other wideband circuits. In both audio and video signal generators (figure 5-5), major components include a POWER SUPPLY, an OSCILLATOR (or oscillators), one or more AMPLIFIERS, and an OUTPUT CONTROL.

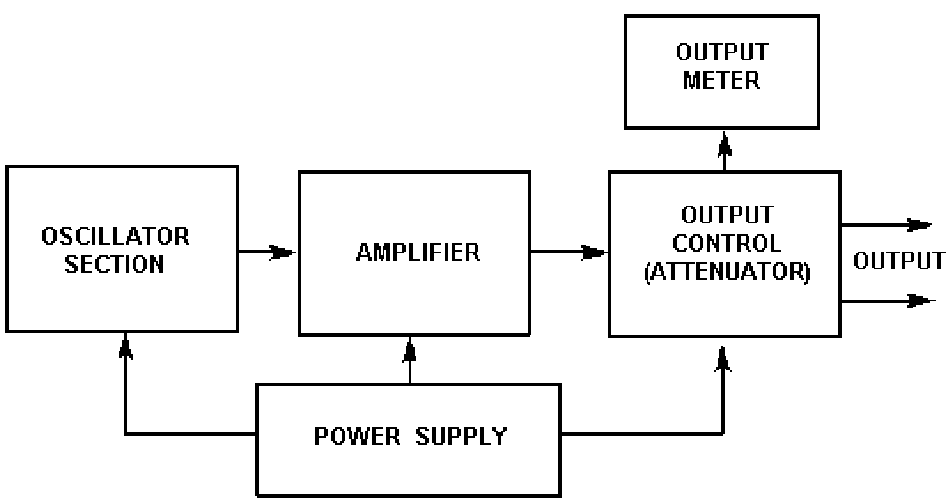


Figure 5-5.—Af and video signal generator block diagram.

In the audio and video generators that produce a beat-frequency, the output frequency is produced by mixing the signals of two separate rf oscillators. One is fixed in frequency, and the other is variable. The difference between the frequencies of the two oscillators is equal to the desired audio or video frequency.

Audio signal generators often include resistance-capacitance (rc) oscillators in which the af is directly produced. In these signal generators, a resistance-capacitance circuit is the frequency-determining part of the oscillator. The frequency varies when *either* the resistance or the capacitance is changed in value.

In other signal generators, however, the capacitance alone is often chosen as the *only* variable element. The change in frequency that can be produced by this method is limited, and it is usually necessary to cover the entire range of the generator in frequency steps. This is usually accomplished by providing several rc circuits, each corresponding to a specific portion of the entire range of frequency values. The circuits in the oscillator are switched one at a time to provide the desired portion of the af range.

The amplifier section of the block diagram (figure 5-5) usually consists of a voltage amplifier and one or two power amplifiers, which are coupled by means of rc networks. The output of the final power amplifier is often coupled to the output control (attenuator) by means of an output transformer.

The output control section regulates the amplitude of the signal. A commonly used af signal generator is the model SPN audio oscillator shown in figure 5-6. The model SPN is a programmable synthesized signal generator designed to provide a stable, low-distortion, wide-amplitude range signal over a 1-Hz to 1.3-MHz frequency range. Additionally, the equipment provides a square wave output and means for swept operation with an external signal and can be remotely controlled via an IEEE-488 bus. For the proper operation of any test equipment, you should always refer to the specific technical manual that describes its use.

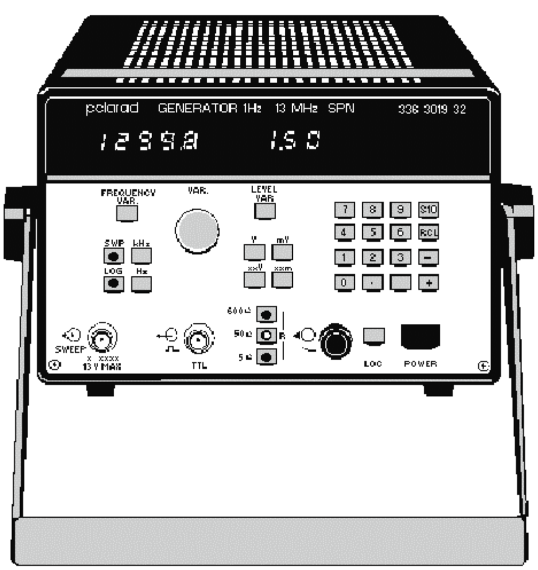


Figure 5-6.—Model SPN audio oscillator.

RADIO-FREQUENCY (RF) SIGNAL GENERATORS

In addition to the necessary power supply, a typical rf signal generator contains three other main sections: an OSCILLATOR CIRCUIT, a MODULATOR, and an OUTPUT CONTROL CIRCUIT. The modulator modulates the rf signal of the oscillator. In addition, most rf generators are provided with connections through which an external source of modulation of any desired waveform can be applied to the generated signal. Metal shielding surrounds the unit to prevent signals from the oscillator from affecting the circuit under test.

Q-4. Name the three main sections of a typical rf signal generator.

A block diagram of a representative rf signal generator is shown in figure 5-7. The function of the oscillator stage is to produce a signal that can be accurately set in frequency at any point within the range of the generator. The type of oscillator circuit used depends on the range of frequencies for which the generator is designed. In lower frequency rf signal generators, the oscillating circuit consists of one of a group of coils combined with a variable capacitor. One of the coils is selected by the position of a range selector switch that connects the coil to a capacitor to provide an inductance-capacitance circuit. The inductive-capacitance circuit then has the correct range of resonant frequencies.

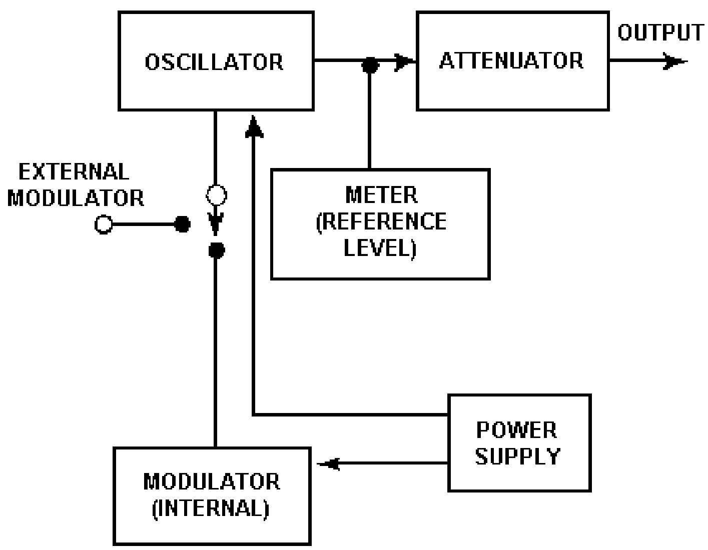


Figure 5-7.—Rf signal generator block diagram.

The function of the modulator is to produce an audio (or video) modulating signal that can be superimposed on the rf signal produced by the oscillator. The modulating signal may be provided by an audio oscillator within the generator. This is termed INTERNAL MODULATION. It may also be derived from an external source. This is termed EXTERNAL MODULATION. In some signal generators, either of these two methods of modulation can be employed. In addition, a means of disabling the modulator section is available so that the pure, unmodulated signal from the oscillator can be used when desired.

Q-5. What is the function of the modulating circuit?

The type of modulation selected depends on the application of the particular signal generator. The modulating signal may be a sine wave, a square wave, or pulses of varying duration. In some special generators, provision is made for pulses over a wide range of repetition rates and widths.

The output circuit of the rf signal generator usually contains a calibrated attenuator and an output level meter. The output level meter provides an indication and permits control of the output voltage of the generator. The attenuator allows you to select the amount of this output. The attenuator is made up of a group of resistors that form a voltage-dropping circuit.

It is controlled by a control calibrated in microvolts. When the control is adjusted so that the output meter reads unity (1.0), the reading on the attenuator control gives the exact value (no multiplication factor) of the output in microvolts. If an output voltage at a lower value is desired, the attenuator control is varied until the meter indicates some decimal value less than 1. This decimal is multiplied by the attenuator reading to give the actual output in microvolts.

An rf signal generator currently in wide use by the Navy is the HP 8640B (figure 5-8). The HP 8640B signal generator covers the frequency range of 500 kilohertz to 512 megahertz, and can be extended to 1,024 MHz by using adapters.

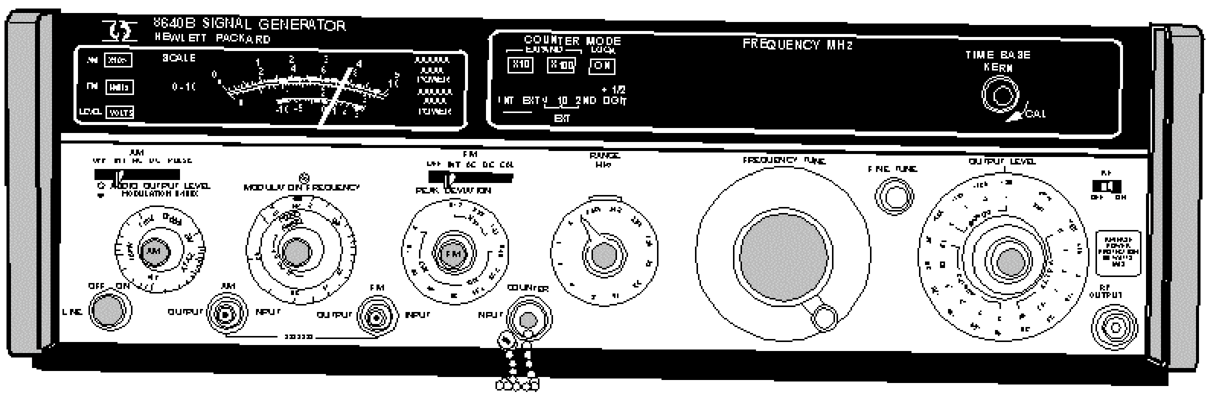


Figure 5-8.—Rf signal generator (HP-model 8640B).

This completes our discussion of signal generators. The following section deals with an instrument that *measures* frequency—the FREQUENCY COUNTER.

FREQUENCY COUNTERS

The signal generators you studied in the previous section *provide* signals for use in testing, aligning, and troubleshooting electronic equipment. Now, we will study the FREQUENCY COUNTER, an instrument that *measures* frequencies. Frequency counters are used to measure frequencies already in existence. An example of a typical frequency counter, the model 5328A, is shown in figure 5-9.

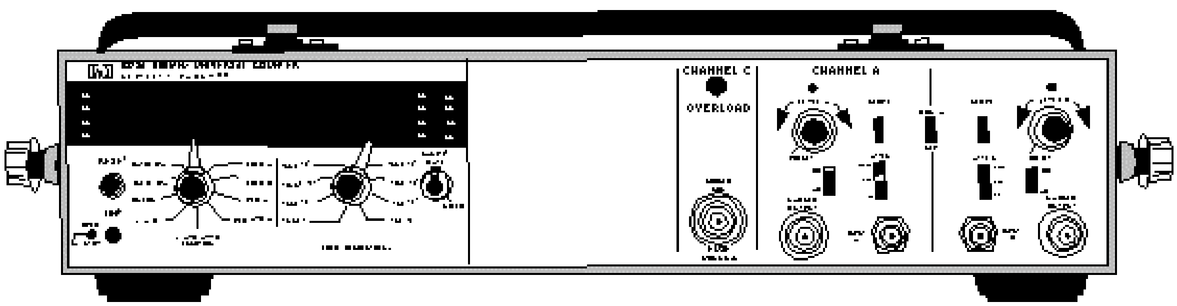


Figure 5-9.—Model 5328A 500 MHz universal frequency counter.

GENERAL DESCRIPTION OF THE FREQUENCY COUNTER

The following description is for the model 5328A counter only. Other counters use different techniques to derive the displayed frequency. The outputs available vary from counter to counter. For further information on other types of counters and their uses, refer to EIMB, *Test Equipment*, NAVSEA 0967-LP-000-0040.

The model 5328A is a portable, solid-state electronic frequency counter. It is used to precisely measure and display, using a nine-digit LED readout, frequency, period, period average, time interval, time interval average, and ratio of electronic frequency signals. This frequency counter can also provide a 1-MHz and 10-MHz output signal through the back-panel BNC.

Q-6. What frequencies are provided through the back-panel BNC?

The model 5328A frequency counter can be divided into four major internal subsections: the main counter section, input section, power supply, and interface bus section. Additionally, two separate (front panel) input channels provide for time interval measurements. Each channel has an attenuator, trigger slope selector, level control, ac or dc coupling, and an oscilloscope marker output. A third input channel is provided to allow the measurement of 30 MHz to 500 MHz with a maximum input of 5 volts rms with a fused-protected connection.

Front-panel controls are provided for you to do the function selection, frequency resolution, sample rate, and reset display. Also, a push-button control on the front panel allows the unit to be used in the operational or standby mode (power applied to the crystal oven to eliminate warm-up). Rear-panel connectors provide the gated output frequencies and an input for an external frequency standard. A detachable front cover is used to store accessory cables and connectors.

Controls and Indicators

Figures 5-10 and 5-11 show all the front- and rear-panel operating controls and indicators. Refer to tables 5-1 and 5-2 for a description of each of the numbered controls and connectors shown in figures 5-10 and 5-11, respectively.

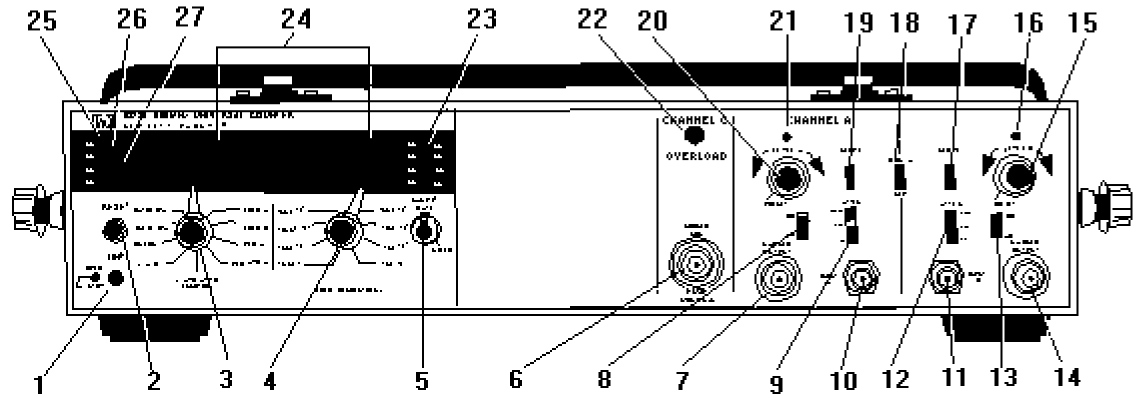


Figure 5-10.—Front-panel controls and indicators.

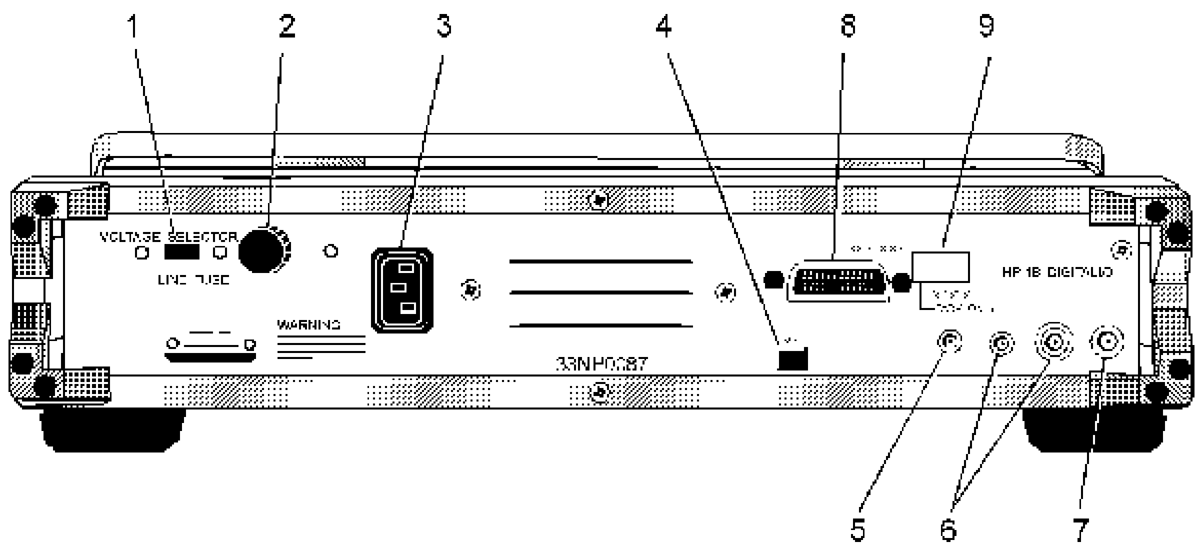


Figure 5-11.—Rear-panel controls and connectors.

Table 5-1.—Front Panel Controls and Connectors

NO.	NAME	DESCRIPTION
1	LINE switch	This is a two-position switch that can be selected for STBY or OPER. When in STBY, the power supply provides power to the oven of the high stability time base to maintain a constant temperature for the crystal. When in OPER, the power supplies normal power to the unit for operation.
2	RESET	This button, when pressed, resets the display and internal count to zero. When pressed continuously, this button lights all segments of the LED display and all annunciator LEDs for a LED test.
3	FUNCTION	This is a 10-position selector used to select the mode of operation.
	FREQ A	This position, when selected, allows the counter to measure frequency at channel A.
	FREQ C	This position, when selected, allows the counter to measure the frequency at channel C.
	PER A	This position, when selected, allows the counter to measure the period at channel A.
	PER AVG A	This position, when selected, allows the counter to make a period average measurement of the signal at channel A. The number of periods over which the average is made is determined by the RESOLUTION switch selection.
	T. I. AVG A→B	This position, when selected, allows the counter to make a time interval measurement of a signal applied to channel A. The number of time intervals over which the average will be made is determined by the RESOLUTION switch selection made.
	T. I. A→B	This position, when selected, allows the counter to make a time interval measurement. The start signal would be applied to channel A and the stop signal applied to channel B.
	CHECK	This position, when selected, applies a 10-MHz signal to the decade counting assemblies, verifying operation of the SAMPLE RATE control, RESOLUTION switch, and RESET.
3 Continued	RATIO C/A	This position, when selected, allows the counter to measure the ratio of the frequency at channel C to the frequency at channel A.
	RATIO B/A	This position, when selected, allows the counter to measure the ratio of the frequency at channel B to the frequency at channel A.
	Top	This position is blank and has no function.
4	FREQ RESOLUTION, N	This is an eight-position switch that is used to select the resolution in frequency measurements and N for totalizing and averaging measurements. This also determines how long the main gate is open for frequency measurements. N / GATE TIME (seconds) / RESOLUTION (Hz) 1 / 1×10^{-6} / 1 M 10 / 10×10^{-6} / 100 k 10^2 / 100×10^{-6} / 10 k 10^3 / 1×10^{-3} / 1 k 10^4 / .01 / 100 10^5 / .1 / 10 10^6 / 1 / 1 10^7 / 10 / .1
5	SAMPLE RATE control	This is used to vary the time between measurements continuously from approximately 2 milliseconds to an indefinite hold of the display.
6	500 MHz, 50 Ω	This is the channel C input connector.
7, 14	OUTPUT MARKER	These are the channel A and B Schmidt trigger outputs used to indicate when a channel is triggered by 0 to 300 mV levels with less than 20 nanoseconds delay.

Table 5-1.—Front Panel Controls and Connectors—Continued

8, 13	AC/DC	These control the selection of ac or dc coupling for the input signal. When the input amplifier control switch (No. 18) is in COM A, channel B coupling is determined by setting of channel A coupling switch.
9, 12	ATTEN switches	These select the attenuation of the input signal. The signal amplitude is reduced by 10 in ×10 and by 100 in ×100. When the input amplifier control switch (No. 18) is in COM A, the channel B attenuation is determined by the channel A attenuation switch.
10, 11	INPUT A and B	These are the input BNC connectors for channels A and B.
15, 20	LEVEL A/B	These controls are used in conjunction with the ATTEN (Nos. 9, 12) to select the voltage at which triggering will occur. With X1 attenuator selected, the level is variable at ±2.5 volts, and ×100 at ±250 volts.
16, 21	Channel A and B triggering lights	The light will blink for the associated channel when triggering is occurring. When the light is off, the input signal is below triggering level and on when the signal is above triggering level.
17, 19	Channel A and B SLOPE switch	These switches control the selection of triggering on either the positive or negative slope of the input signal.
18	COM A/SEP	This is the input amplifier control switch that selects independent operation of channels A and B in SEP (separate) position. When in COM A (common A) position, the signal at A is also applied to channel B; this disconnects the channel B input circuitry. Channel B coupling and attenuation are then determined by the channel A setting.
22	OVERLOAD	This indicator will flash on and off if more than 5 volts is applied to channel C input.
23	K, S, M, n, and Hz	These will light up to show the appropriate units multiplier of the measurement being taken.
24	LED display	This is a nine-digit LED display that shows the numerical measurement taken.
NO.	**NAME**	**DESCRIPTION**
25	OVFL	This indicates an overflow of one or more of the most significant digits (leftmost from the decimal point) are not displayed.
26	RMT	Lights when the unit is in remote operation.
27	GATE	Lights when the counter's main gate is open and a measurement is in progress.

Table 5-2.—Rear-Panel Controls and Connectors

NO.	NAME	FUNCTION
1	VOLTAGE SELECTOR	Used to select 115 or 230 volt operation.
2	LINE FUSE	This requires the insertion of a 2.0 amp fuse for 115 volt or 1.0 amp fuse for 230 volt operation.
3	Input ac connector	Used to connect the input ac to the unit.
4	ARM	When this switch is in the OFF position, the counter is armed by the signal that is selected for measurement. In the ON position, the measurement is armed by an input other than the input being measured.
5	EXT OSC	This input connector allows a separate outside signal to be used for the time base.
6	1 MHz OUT and 10 MHz OUT	These connectors allow an internal oscillator signal to be used externally when connected.
7	GATE/MARKER OUT	Applies a high to the output when the main gate is open.
8	HP-IB	This is an interface bus connector that allows the unit to receive programming instructions.

Frequency Measurement

As discussed previously, the model 5328A frequency counter is capable of measuring frequency, time period (inverse of frequency), ratio, and time interval. We will start with frequency. When the FUNCTION selector is in the FREQ A or FREQ C position, the counter measures the frequency, f, by accumulating the number of cycles, n, of the input signal that occurs over the time period, t. This is expressed by:

$$f = \frac{n}{t}$$

The basic counter elements necessary to perform this measurement are shown in figure 5-12. The INPUT AMPLIFIER/TRIGGER essentially conditions the input signal to a format that is compatible with the internal circuitry of the counter. As figure 5-12 indicates, the output of the amplifier/trigger corresponds directly to the input signal.

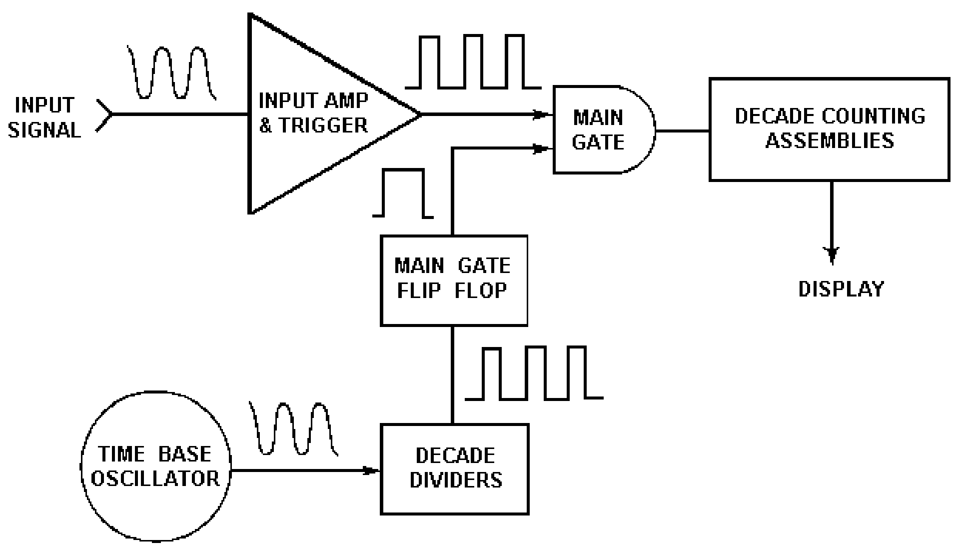


Figure 5-12.—Basic elements of the frequency counter.

The TIME BASE OSCILLATOR is a 10-MHz temperature-controlled (oven-regulated) precision, crystal oscillator used for the time base element from which time, t, is derived. DECADE DIVIDERS take the time base oscillator signal as the input and provide a pulse train, whose frequency is variable in decade steps. This frequency can be controlled by the FREQ RESOLUTION, N switch. The time, t, is determined by the period of this pulse train.

The heart of the counter is the MAIN GATE. When the gate is opened, pulses from the amplifier/trigger are allowed to pass through. The opening and closing of the main gate is controlled by the decade divider output to the main gate flip-flop. The output of the MAIN GATE is then sent to the DECADE COUNTING ASSEMBLIES (DCAs), where the pulses are combined and displayed after the gate is closed.

If the FREQUENCY RESOLUTION, N selection switch, is set for 10^6, the main gate is open for 1 second, and the decade counting assemblies display the frequency of the input signal in hertz (refer to figure 5-10, FREQUENCY RESOLUTION, N selection switch).

Period Measurement

Period, the inverse of frequency, can be measured with the counter by reversing the inputs to the main gate. With the FUNCTION selector switch in the PER A position, the input signal controls the duration over which the main gate is open and the decade divider output is counted by the DCAs (see figure 5-13). The duration of the count is one cycle or period of the input signal. When the FUNCTION selector is in the PER AVG A position, the unused decades in the decade divider chain are used to divide the amplifier/trigger output so that the gate remains open for decade steps of each input period rather than a single period. This is the basis for multiple period averaging. Period and period averaging techniques are used to increase measurement accuracy on low-frequency measurements.

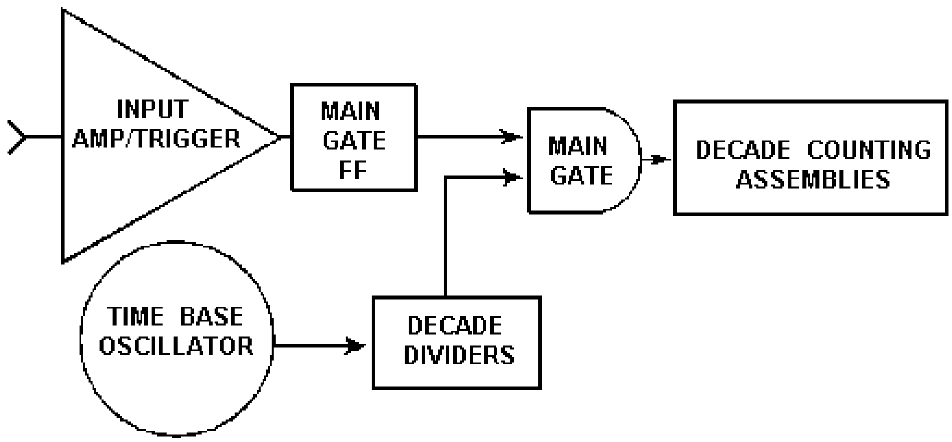


Figure 5-13.—Measuring period.

Ratio Measurement

Placing the FUNCTION selector switch to RATIO C/A OR B/A sets the counter to measure the ratio of the signal frequency at channel C or B to the signal frequency at channel A. Using the same configuration as in figure 5-13 and replacing the time base with a second input frequency, f_2, you can measure the ratio of f_2/f. The signal at frequency f can be divided into decade steps in the same manner as multiple period averaging for higher resolution.

Time Interval Measurement

Figure 5-14 illustrates the configuration for the measurement of time between two events or time interval. This is done by placing the FUNCTION selector in the T.I.AVG AB position. The START input opens the main gate, and the STOP input closes it. The START input is applied to channel A, and the STOP input is applied to channel B. The decade divider output (clock pulses) is counted, and the display shows the elapsed time between START and STOP signals, as shown in figure 5-15.

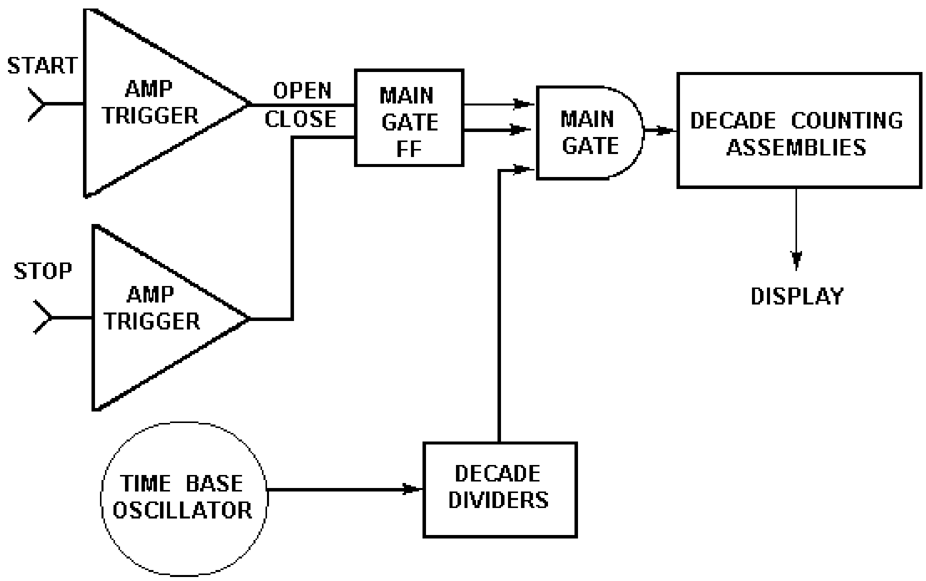


Figure 5-14.—Basic elements of a time interval counter.

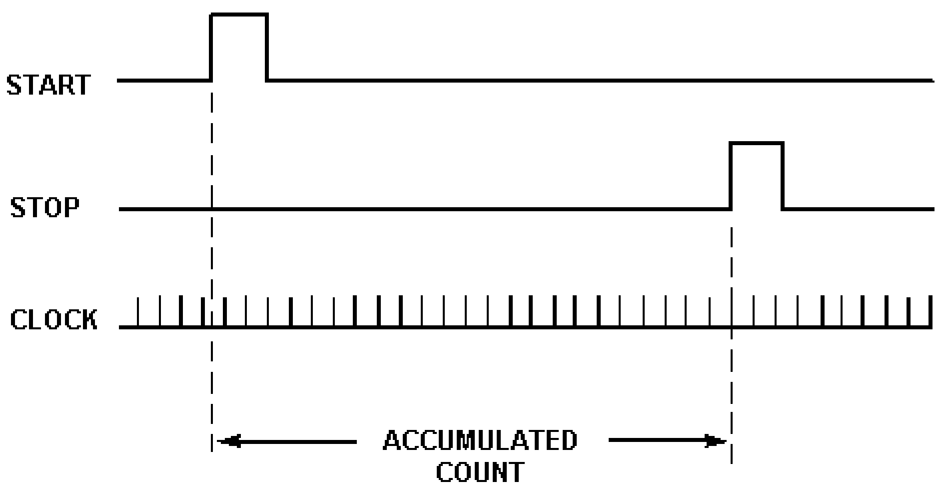


Figure 5-15.—Clock pulses.

Resolution

The resolution of the measurement is determined by the frequency of the counted clock (for example, a 10-MHz clock provides 100 ns resolution [see figure 5-10, FREQUENCY RESOLUTION, N selection switch]). The input amplifier, main gate, and DCAs (elements of the time interval counter) must operate at speeds consistent with the clock frequency; otherwise the instrument's resolution would be meaningless. Clock frequencies of 1, 10, and 100 MHz, and other 10n frequencies, are preferred, since the accumulated count, with the appropriate placement of the decimal point, gives a direct readout of the time interval. This explains why the conventional time interval counter is presently limited to 10 nanoseconds, a clock frequency of 100 MHz. One GHz is beyond reach, and a clock frequency of 200 MHz would require some arithmetic processing of the accumulated count in the DCAs to enable time to be displayed directly.

Time Interval Averaging

The time interval averaging technique is based on the fact that if the ±1 count error is truly random, it can be reduced by averaging a number of measurements. The words "truly random" are significant. For time interval averaging to work, the time interval must (1) be repetitive and (2) have a repetition frequency that is asynchronous to the instrument's clock. Under these conditions, the resolution of the measurement is:

$$\text{resolution} = \frac{\pm 1 \text{ count}}{\sqrt{N}}$$

$$\text{where } N = \text{number of time intervals averaged}$$

With averaging, resolution of a time interval measurement is limited only by the noise inherent in the instrument. The 5328A can obtain 10-picoseconds resolution. Most time interval averaging has one severe limitation: The clock period limits the minimum measurable time interval. With the FUNCTION selector switch in the T.I. AVG AB position, synchronizers are used to remove this limitation. These synchronizers enable the 5328A to measure intervals as short as 100 picoseconds.

Referring to figure 5-16, note that the input waveshape shows a repetitive time interval, which is asynchronous to the square wave clock. When these signals are applied to the main gate, with no synchronizers, an output similar to the third waveform results. Since the DCAs are designed to count at the clock frequency and are unable to accept pulses of shorter duration than the clock, the resulting counts accumulated in the DCAs will be in error, as shown in the fourth waveform. This problem is alleviated by the synchronizers, which are designed to detect leading edges of the clock pulses that occur while the gate is open. They detect and reconstruct the leading edges, making the pulses applied to the DCAs the same duration as the clock, as shown in the fifth waveform. Occasionally, when the input time interval repetition is synchronous with the internal clock, time interval averaging cannot be performed.

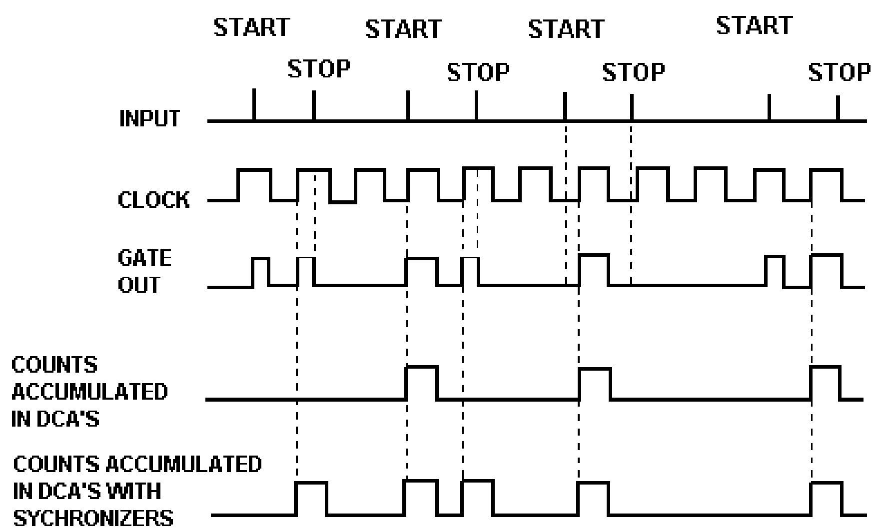


Figure 5-16.—Synchronizer operation with time interval averaging.

This ends our discussion on electronic frequency counters. Now, we'll study an area of electronics test equipment that is becoming more widespread and important each day—the testing of electronic logic components. A test instrument of value for any technician who works on digital equipment is the LOGIC PROBE, which is an integrated circuit-testing device.

INTEGRATED CIRCUIT-TESTING DEVICES

Digital integrated circuits are relatively easy to troubleshoot and test because of the limited numbers of input and output combinations involved in circuits. The two-state conditions in logic circuits are often referred to as (1) low or high, (2) on or off, or (3) one or zero (1 or 0).

Other terminology may also be used. Any particular integrated circuit (IC) can be tested by simply comparing it to a known good one. The LOGIC PROBE is a device that can be of great value in troubleshooting digital integrated logic circuits.

The ideal logic probe has the following characteristics:

- It will detect a steady logic level.
- It will detect a train of logic levels.
- It will detect an open circuit.
- It will detect a high-speed transient pulse.
- It will have over-voltage protection.
- It will be small, light, and easy to handle.

The use of a suitable logic probe can greatly simplify your troubleshooting of logic levels through digital integrated logic circuitry. It can immediately show you whether a specific point in the circuit is low, high, open, or pulsing. Some probes have a feature that detects and displays high-speed transient pulses as small as 10 nanoseconds wide. These probes are usually connected directly to the power supply of the device being tested, although a few have internal batteries.

Most IC failures show up in a circuit as a constant high or low level. Because of this, logic probes provide a quick, inexpensive way of locating the fault. They can also display the single, short-duration pulse that is hard to detect on an oscilloscope. Figure 5-17 shows a basic logic probe.

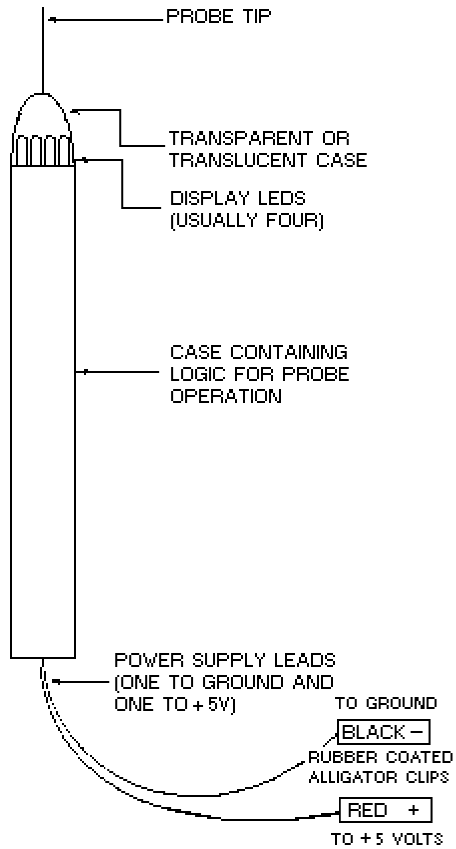


Figure 5-17.—Basic logic probe.

The logic probe can be powered from the supply of the circuit under test or from a regulated dc power supply. If a separate power supply is used, the ground points of the power supply and circuit under test should be connected together.

The display LED (light-emitting diode) near the probe tip provides an immediate indication of the logic state existing in the circuit under test. The LED will provide any of four indications: (1) off, (2) dim (about one-half brilliance), (3) bright (full brilliance), and (4) flashing on and off. The LED is normally in the dim state and must be driven to one of the other three states by voltage levels at the probe tip. The LED is usually bright for inputs above the logic "1" threshold and off for inputs below the logic "0" threshold. The LED is usually dim for voltages between the logic "1" and logic "0" thresholds and for open circuits.

Q-7. The LED lamps of a typical logic probe are normally in what state?

Another logic circuit analysis technique is useful with the logic probe. This technique is to run the circuit under test at its normal clock (timing) rate while monitoring for various control signals, such as RESET, START, STOP, SHIFT, TRANSFER, or CLOCK. Questions such as "Is the counter operating?" are quickly resolved by noting if the probe indicator is flashing on and off, indicating that pulse train activity is present.

This ends our discussion on logic probes. Now, we'll study another piece of electronics test equipment that is used in evaluating integrated circuits, the HUNTRON TRACKER 2000.

HUNTRON TRACKER 2000

The logic probe we just discussed is but one specialized tool used to isolate problems to the component level. Another device you can use is the Huntron Tracker 2000. It is a very versatile electronic troubleshooting tool that is used to evaluate suspect components and/or locate defective components on de-energized circuit cards quickly and safely without requiring the removal of component leads. The unit provides a built-in display that allows you to visually analyze the component under test conditions.

CAUTION

Before connecting the Huntron Tracker 2000, you must first secure all power, then discharge all high-voltage capacitors.

PHYSICAL FEATURES

Because the Tracker 2000 has so many controls and indicators, it would impractical to cover each within this chapter. We will therefore concentrate our discussion only to the externally accessible features. To find information on internal controls and indicators, you should review the most current technical manual with up-to-date changes entered for the unit being used.

Front Panel

The front-panel (figure 5-18) design allows you to easily select the desired function. All the push buttons are the momentary action type and have light-emitting diode (LED) indicators embedded in them to show the functions that are active, by lighting up when active. A detailed description of each item on the front panel is provided in table 5-3.

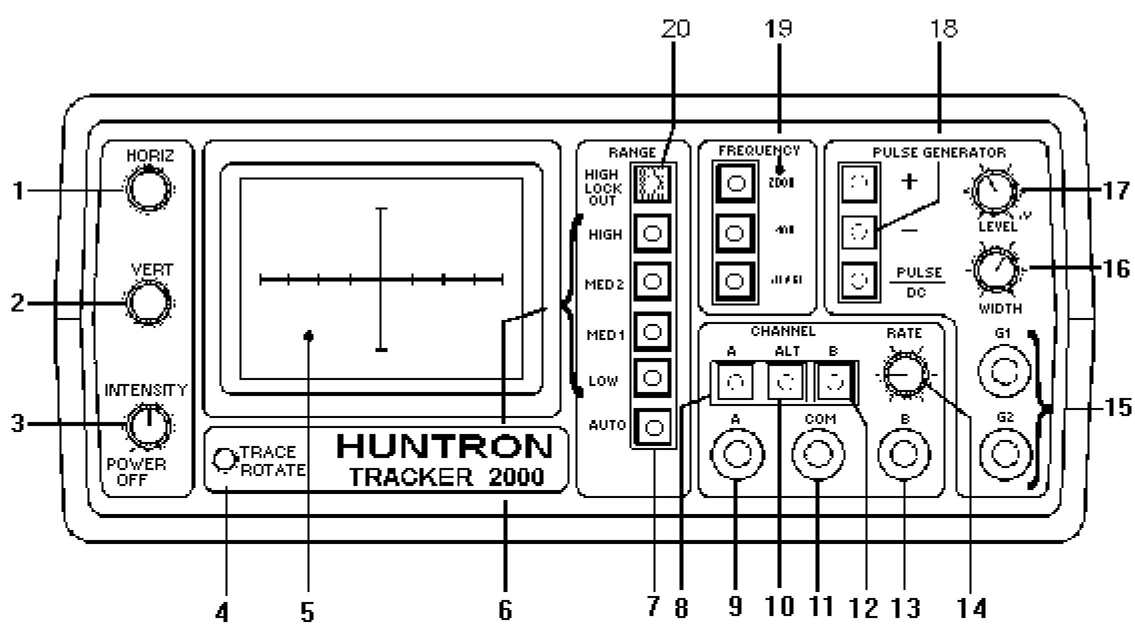


Figure 5-18.—Front Panel.

Table 5-3.—Front-Panel Controls and Connectors

NO.	NAME	DESCRIPTION
1	HORIZ control	When adjusted this controls the horizontal position of the CRT display.
2	VERT control	When adjusted, this controls the vertical position of the CRT display.
3	POWER on/off and INTENSITY control switch	When this is rotated clockwise, the power is turned on. Further adjustment of the switch controls the intensity of the CRT display. When it is rotated fully counterclockwise, the power is turned off.
4	TRACE ROTATE control	When adjusted, this controls the trace rotation of the CRT.
5	CRT display	This displays the signatures produced by the unit.
6	RANGE selectors	These are push buttons that are used to select one of four impedance ranges: LOW, MEDIUM 1, MEDIUM 2, and HIGH.
7	AUTO selector	This push button, when selected, initiates automatic scanning of the four ranges from low to high. The speed of the scanning is determined by the RATE control (item 14)
8	CHANNEL A selector	This push button, when selected, causes channel A to be displayed on the CRT.
9	CHANNEL A test plug	This is a fused test lead connector that is active when channel A is selected. All test lead connectors accept standard banana plugs.
10	ALT selector	When selected, this push button causes the unit to alternate between channel A and channel B. The speed of this is determined by the RATE control (item 14).
11	COM test plug	This test lead connector is the instrument common to and the common reference point for channel A and channel B.
12	CHANNEL B selector	This push button, when selected, causes channel B to be displayed on the CRT.
13	CHANNEL B test plug	This is a fused test lead connector that is active when channel B is selected.
14	RATE control	This controls the channel alteration and/or the range scanning.
15	G1 and G2 plugs	These are used for the pulse generator output test leads.
16	WIDTH control	This controls the duty cycle of the pulse generator.
17	LEVEL control	This controls the amplitude of the internal pulse generator.
18	PULSE GENERATOR selectors	These push buttons are used to select the output modes of the pulse generator: positive (+), negative (–) or PULSE/DC.
19	FREQUENCY selectors	These push buttons are used to select one of the three test signal frequencies: 50/60 Hz, 400 Hz, 2000 Hz.

Cathode-Ray Tube (CRT) Display

The CRT display (figure 5-19) is used to view the signature of the component under test. The display has a graticule consisting of a horizontal and vertical axes. The horizontal axis is used to represent voltage with the vertical axis being used to represent current. The axes divide the display into four quadrants. Each quadrant displays a different portion of the signature for the component under test.

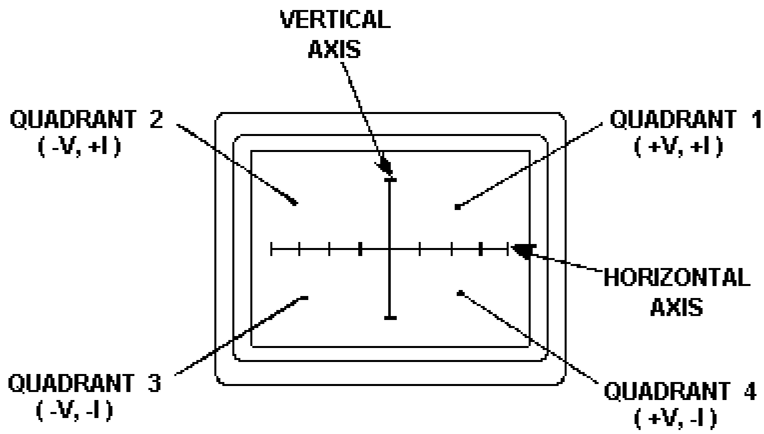


Figure 5-19.—CRT display.

- Quadrant 1 displays positive voltage and positive current.
- Quadrant 2 displays negative voltage and positive current.
- Quadrant 3 displays negative voltage and negative current.
- Quadrant 4 displays positive voltage and negative current.

Q-8. On the CRT display, what information is displayed in Quadrant 4?

The horizontal axis (see figure 5-19) is divided into eight equal divisions, allowing the technician to estimate the voltage at which changes occur in the signature for the component being tested. The associated approximate horizontal sensitivities for each range are:

- High = 15.0 Volts/Div.
- Medium 2 = 5.0 Volts/Div.
- Medium 1 = 3.75 Volts/Div.
- Low = 2.5 Volts/Div.

Back Panel

The back panel (figure 5-20) provides three additional controls and connectors. One is the accessory output connector (ACC), which provides a clock signal and power for the Huntron Switcher Model HSR410. Next, is the power cord connector used to provide the required ac line voltage used to power the unit. The last item found on the back panel is the FOCUS control; this controls the focus for the front-panel CRT display.

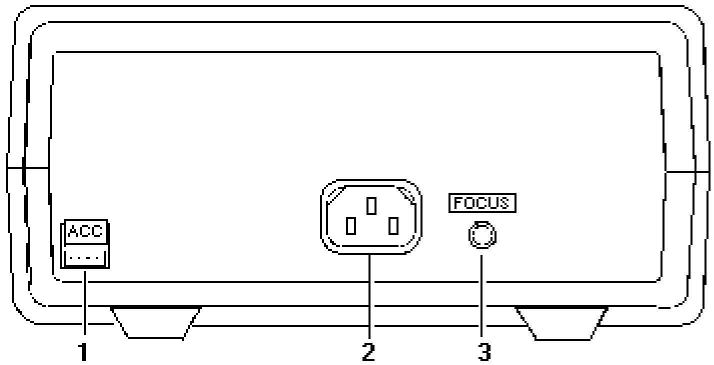


Figure 5-20.—Back panel.

OPERATION

The following sections explain how to use most of the front- and back-panel controls and connectors. Each control and connector has already been briefly described. To review this information, refer to figure 5-18 and table 5-3.

Initial Setup

First, turn the POWER/INTENSITY knob located on the front panel to the clockwise, on, position. Under normal conditions, the unit will come on with the following LEDs illuminated: power, channel A, 50/60 Hz, low range, and pulse/DC.

Focusing the CRT is very critical for the technician to be able to properly analyze the signature being displayed. This is done by first turning the INTENSITY control to a level comfortable for the eye, and then adjusting the FOCUS control (back panel) for the narrowest possible trace.

Aligning the trace will help determine, during troubleshooting, which quadrant the portion of the signature is in during a change. With a short circuit applied to channel A (connect a cable between jack A and COM), adjust the TRACE ROTATE control until the trace is as close to parallel as possible to the vertical axis. Then, adjust the HORIZ (horizontal) control until the vertical trace is as close to even with the vertical axis as possible.

Q-9. When aligning the trace with a short applied to channel A, which control should be adjusted to bring the trace parallel to the vertical axis?

With an Open applied to channel A (nothing connected to A and COM test plug), adjust the VERT (vertical) control until the horizontal trace is as close to even with the horizontal axis as possible. Once set, these adjustments should not need readjusting during the unit's operation. However, remember that each time the unit is used, this process will need to be repeated.

Range Selection

Four impedance ranges (LOW, MED 1, MED 2, and HIGH) can be selected on the Tracker 2000. These ranges will become active when the appropriate front-panel button is pressed. To obtain the most useful signature display when troubleshooting a component, you should always start by selecting one of the two medium ranges (medium 1 or medium 2). If the display appears to be an open (horizontal trace), then the next higher range should be selected to get a more accurate signature for analysis. If the display appears to be a short (vertical trace), then the next lower range should be selected.

The AUTO feature, when selected, will allow the unit to automatically scan through the four ranges at a speed set by the technician using the RATE control knob. This allows the user to keep his or her hands free to hold test leads while still being able to observe the component under test signature for analysis. The HIGH LOCKOUT, when selected by the technician, prevents the unit from functioning in the HIGH range in either the manual or AUTO mode.

Channel Selection

You can select two channels by pressing the channel A (test probes connected to A jack and COM) or channel B (test probes connected to B and COM jacks) push button on the front panel. When using a single channel, you should plug the red probe into the corresponding channel test jack, and plug the black test lead into the common test jack. When testing a component, you should connect the red probe to the positive terminal and the black probe to the negative terminal of the component under test. Following this procedure every time will ensure that the signature for the component under test will be displayed in the correct quadrants of the CRT display.

The ALT (alternate) mode provides automatic switching back and forth between channel A and channel B. This allows you to easily compare two components or the same test points on two circuit boards. You select the ALT mode by pressing the ALT push button on the front panel. The rate of switching between channels A and B can be varied by adjusting the RATE control knob on the front panel. You will find that the ALT mode feature is very useful for comparing a known good component with the same type of component that is of unknown quality.

Figure 5-21 shows a typical way of connecting the unit to a known good circuit board and a board under test. This test mode uses the supplied common test leads to connect two equivalent points on the boards to the common test jack. Note that the black probe is now being used in the channel B jack rather than the COM jack. When the technician uses the ALT and AUTO features together, each channel is displayed before the range selection will change. Figure 5-22 shows the sequence of these changes.

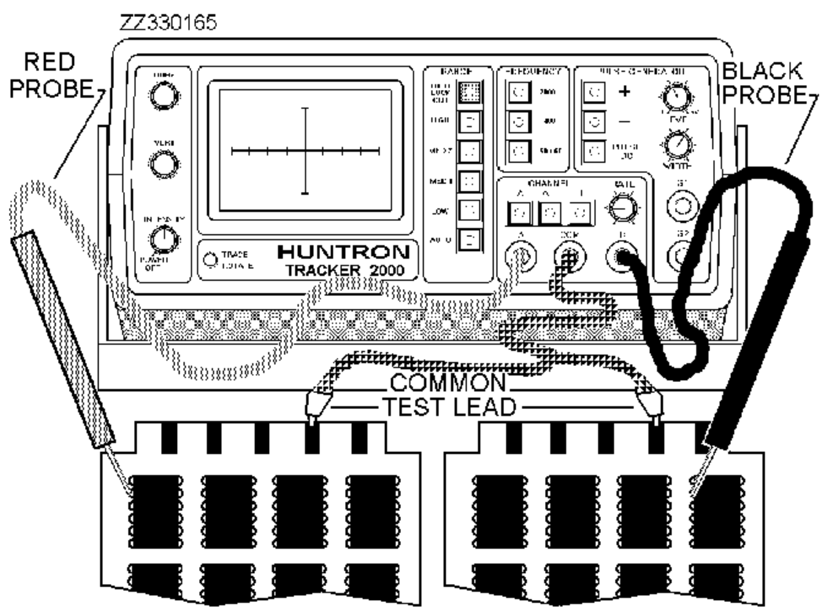


Figure 5-21.—ALT (alternate) mode setup.

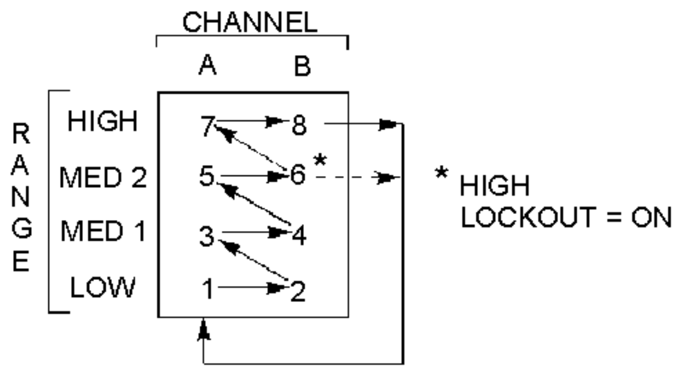


Figure 5-22.—AUTO/ALT sequence.

Frequency Selection

There are three test signal frequencies (50/60 Hz, 400 Hz, and 2000 Hz) that can be selected and then provided by pressing the appropriate front-panel push button. During most troubleshooting evolutions, the 50/60 Hz test signal is the best to start with. The 400 Hz and 2000 hz frequencies are generally used to view small amounts of capacitance or large amounts of inductance.

Pulse Generator

A built-in pulse generator is also provided with the Tracker 2000. It allows the technician to do dynamic, in-circuit testing of certain devices in their active mode. In addition to using the red and black probes, you can connect the output of the pulse generator to the control input of the device to be tested with one of the blue micro clips provided with the unit. The pulse generator has two outputs (G1 and G2 jacks) so that three devices can also be tested in the ALT (alternate) mode. Figure 5-23 shows a way of connecting the unit in the ALT mode using the pulse generator.

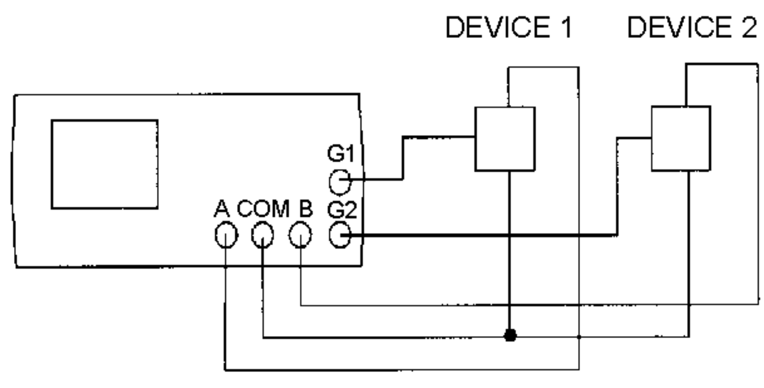


Figure 5-23.—Pulse generator comparison mode.

There are a variety of output waveforms available using the pulse generator selection buttons, as shown in figure 5-24. First, the technician must select the PULSE mode or DC mode using the PULSE/DC button located on the front panel. In the PULSE mode, the PULSE/DC LED flashes at a slow rate. While in the DC mode, this LED will be continuously on.

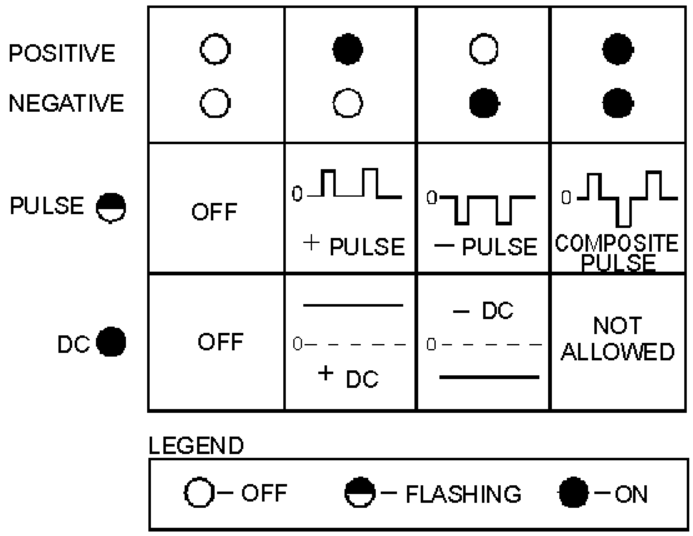


Figure 5-24.—Pulse generator selector chart.

Next, the technician needs to select the polarity of the output desired by using the positive (+) and/or negative (–) push buttons. All three buttons only function in a push-on/push-off mode and will only interact with each other to avoid the NOT ALLOWED state found in figure 5-24.

Once the specific output type has been selected, the output desired by the technician is set using the LEVEL and WIDTH controls. The LEVEL control is used to vary the magnitude of the output amplitude from zero to 5 volts (peak or DC). During the PULSE mode, the WIDTH control will adjust the cycle of the pulse output from low to 50 percent maximum (square wave).

The start of a pulse will be triggered by the appropriate zero crossing of the test signal, which results in the pulse frequency being equal to the selected test signal frequency.

The end of the pulse is determined by the WIDTH control setting, which determines the cycle length. The WIDTH control, however, has no effect when the DC mode is selected.

FUNCTIONAL OVERVIEW

There are six major sections in the Tracker 2000, as shown in the block diagram in figure 5-25. The control logic section controls the selection of the channel, frequency, impedance range, and pulse generator mode according to the front-panel buttons pushed by the technician. The oscillator provides the test signal that is used by the signal and pulse generator sections. In the signal section, the test jacks are driven by the test signal, while signal conditioners monitor the jacks and produce the horizontal and vertical signals used by the CRT section to produce a component signature on the display. The pulse generator provides an added source for testing three additional terminal devices. The power supply produces all the required voltages needed to operate the Tracker 2000.

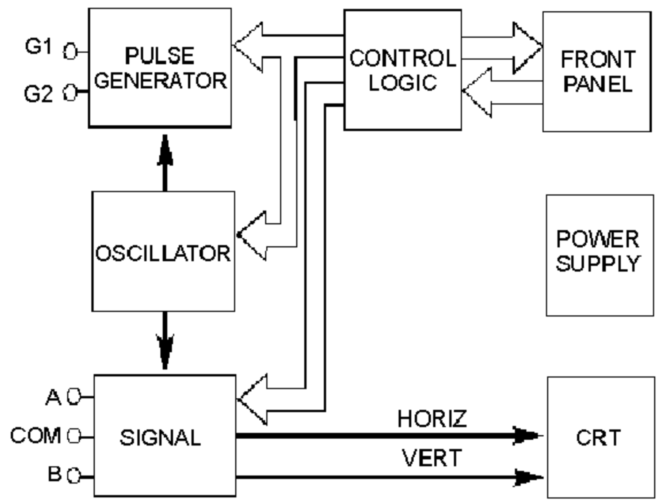


Figure 5-25.—Tracker 2000 block diagram.

Control Logic

The control logic senses which button is pushed on the front panel. Since the buttons are the momentary action type, the logic must remember what button was pushed, turn on the LED indicator within the button, and activate the appropriate configuration of the oscillator, signal, and pulse generator sections.

After a button is pushed, the unit will remain in that configuration until another selection is made or the power is secured. The HIGH LOCKOUT and PULSE GENERATOR push buttons are the only ones that require repetitive pressing to be turned on or off.

The channel relay is controlled by CHANNEL buttons A, ALT, and B. The relay is a single-pole, double-throw type and is de-energized for channel A and energized for channel B. If CHANNEL A is already selected and the CHANNEL B button is pressed, CHANNEL A will be canceled and CHANNEL B selected. When the ALT button is pressed, another control line is set, which enables an internal clock to toggle the channel relay on and off, causing the unit to alternate between channels A and B.

The internal clock is controlled by the RATE control knob on the front panel. When the ALT mode is active, the LEDs within the A and B buttons flash alternately, and the ALT LED is on continuously. Pressing either of the channel buttons will cancel the ALT mode, and the selected channel will then become active.

The FREQUENCY buttons (50/60, 400, and 2000 Hz) directly control the operation of the oscillator and the pulse generator. The RANGE buttons (LOW, MED 1, MED 2, and HIGH) control four relays in the signal section that select the appropriate terminal characteristics for each impedance range (table 5-4).

Table 5-4.—Terminal Characteristics for Impedance ranges

RANGE	OPEN CIRCUIT VOLTAGE (Vp)	SHORT CIRCUIT CURRENT (mAr)
High	60	0.57
Medium 2	20	0.53
Medium 1	15	8.5
Low	10	132

You can select the four ranges manually by pressing the RANGE button, or you can scan them automatically by using the AUTO function. When AUTO is activated, the control logic will follow the sequence LOW, MED 1, MED 2, and HIGH over and over if HIGH LOCKOUT is off. The current active range is always indicated by the LED for the range selected. The AUTO mode will stay active until you select a particular range by pressing its associated button. While AUTO is active, the AUTO LED is continuously on.

The speed at which the ranges are scanned is controlled by the front-panel RATE control knob. This allows you to adjust the time each range is displayed for signature analysis on the CRT display. If AUTO and ALT (alternate) are active at the same time, the RATE control affects the speed of both functions with ALT having priority. This is done so that the two channels can be compared to each other within one range before the next range is selected (figure 5-26).

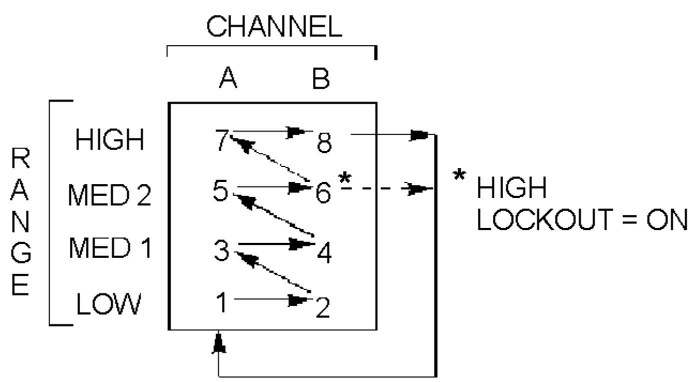


Figure 5-26.—Range scanning sequence with AUTO and ALT active.

The HIGH LOCKOUT function disables the HIGH range and limits the maximum test signal to 20 volts peak vice 60 volts peak. When you select the manual mode (AUTO off), activating the HIGH LOCKOUT prevents the HIGH range from being selected. If the HIGH range is active when the HIGH LOCKOUT is pressed, the HIGH range is canceled and the next lower range (MED 2) will be selected and become active. When you select the AUTO mode, the RANGE sequence with the HIGH LOCKOUT active will start with LOW, and sequence through MED 1, MED 2, back to LOW, and continue until the AUTO mode is stopped.

The PULSE GENERATOR buttons (positive [+], negative [–], and PULSE/DC) toggle control lines that change the polarity and output type of the pulse generator.

Oscillator

The oscillator produces a constant amplitude, low distortion sine wave test signal. The frequency of the test signal is programmable between one variable frequency (50/60 Hz), and two fixed frequencies (400 Hz and 2000 Hz). The variable frequency depends on the input power line used for the Tracker 2000; a 50-Hz line produces a 50-Hz test signal, and a 60-Hz line produces a 60-Hz test signal. If a 400-Hz power line is used, an 80-Hz test signal is provided. This versatility is built in to ensure you will always have low, medium, and high frequencies to work with.

Signal Section

The signal section is considered to be the heart of the Tracker 2000. In this section, the signal from the oscillator is applied across two terminals of a device being tested via the front-panel jacks.

The test signal causes a current to flow through the device and a voltage drop to occur across its terminals. The current flow causes a vertical deflection of the trace on the CRT display, while the voltage drop across the device causes a horizontal deflection of the trace on the CRT display. By combining these, the current-voltage signature of the device being tested is displayed on the CRT.

The test signal appears at the front-panel test jacks as though it is being originated by a voltage source (V_s) with a series output impedance (Z_s). An equivalent circuit of the signal section is shown in figure 5-27. As you can see, the figure also shows how the terminal voltage affects the horizontal deflection plates of the CRT, and how the current through the terminals affects the vertical deflection plates through current sensing point I.

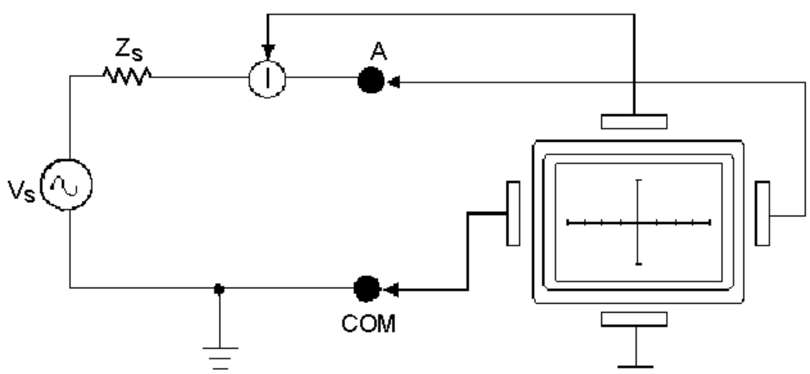


Figure 5-27.—Signal section equivalent circuit.

An open circuit has zero current flowing through the terminals and maximum voltage drop across the terminals. In all ranges, this is represented by a straight horizontal trace from left to right on the CRT display, as seen in figure 5-28 view A. When a short occurs, maximum current flows through the terminals, and the voltage drop is considered to be zero. This occurs in all ranges and is represented by a straight vertical trace from top to bottom of the CRT display, as seen in figure 5-28 view B.

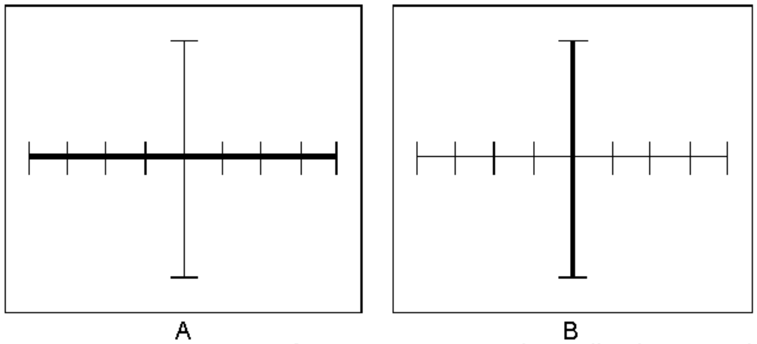


Figure 5-28.—Open and short circuit display.

Pulse Generator

The pulse generator provides dynamic testing for certain types of devices by driving the control input of the device under test. The normal two-terminal mode of testing can be considered a static test, since devices with three or more terminals are not tested in their active mode. However, with the pulse generator, an in-circuit active test is possible.

In the PULSE mode, this circuit uses the zero crossing of the test signal to trigger the start of the pulse. When positive (+) is enabled, a positive-going zero crossing triggers a positive pulse. When negative (−) is enabled, a negative-going zero crossing triggers a negative pulse. If both are enabled, then both positive and negative pulses are produced on alternate crossings (composite pulses). Once a pulse is triggered, its duration is set by the WIDTH control knob. Figure 5-29 shows the waveforms for three pulse polarity types at various settings of the WIDTH control.

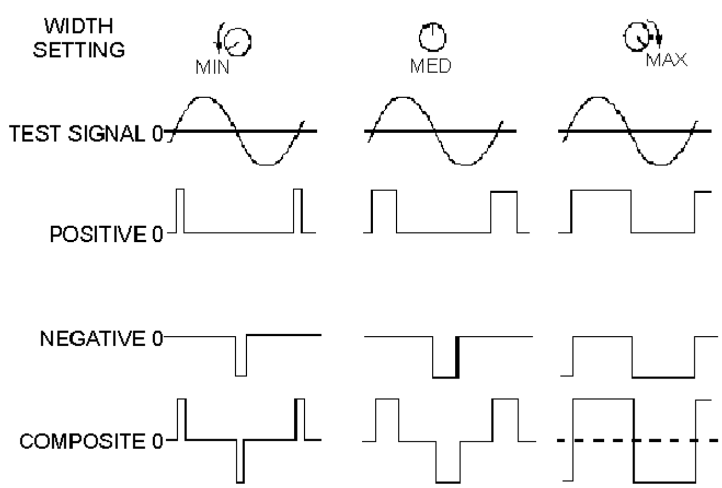


Figure 5-29.—Pulse generator waveforms.

The LEVEL control adjusts the peak of each pulse from zero to 5 volts with the polarity dependent on the pulse polarity selected. When an open circuit is present, a maximum output of 5 volts peak-to-peak is present with either positive or negative selected, and 10 volts peak-to-peak when the composite pulse is active.

Q-10. What minimum/maximum voltage level can be attained in the pulse generator section by adjusting the LEVEL control?

In the DC mode, a zero-to-5-volts DC level is produced at G1 and G2 on the front panel. The polarity of the level is controlled by the positive and negative buttons. By pressing the positive button, you enable a positive DC output and disable the negative button. By pressing the negative button, you enable a negative DC voltage only if the positive button is in the off position. The WIDTH control knob has no effect in the DC mode of operation.

CRT Display

The CRT deflection drivers boost the low-level outputs from the signal section to the higher voltage levels needed by the deflection plates in the CRT. The HORIZ (horizontal) and VERT (vertical) controls on the front panel adjust the position of the CRT trace. The TRACE ROTATE control on the front panel is used to adjust the short circuit vertical trace to be parallel with the vertical axis on the CRT graticule.

Three other controls (INTENSITY, FOCUS, and astigmatism) are used to adjust the proper brightness and clarity of the trace. The front-panel INTENSITY control is the primary way to adjust the visual quality of the trace. FOCUS is located on the back panel and is used as your trimming adjustment. Astigmatism is an internal adjustment and is set at the factory.

Power Supply

The power supply is an ac-line-operated power supply that is turned on and off by the POWER/INTENSITY knob located on the front panel. Once power is turned on, the power supply provides 12 V dc (nominal) and ±5 V dc (regulated) for normal circuit operational use in the oscillator, pulse generator, signal, and control logic sections of the Tracker 2000.

The other outputs from the power supply are provided to the CRT display section. The CRT is provided with a filament voltage of 6.3 V_{rms}, +180 V dc for the deflection driver circuits, and a regulated −1320 V dc for the CRT acceleration voltage.

COMPONENT TESTING

Now that you have a basic understanding of how the Tracker 2000 works, we will show you a few examples of different components with values and their associated displays. Because of the large number of different values that can be given to any component, this section will present only a few.

Testing Resistors

A resistance across the test probes will cause the trace of the Tracker 2000 to rotate in a counterclockwise direction around its center axis from an open circuit position. The degree of rotation is directly related to the resistance value. The higher the value, the less rotation will be observed.

LOW RANGE.—The low range is designed to test for resistance values between 1Ω and 1K. Figure 5-30 shows the effect of resistance on the angle of rotation in low range. A 1Ω resistor causes almost 90º of rotation, and a 50Ω resistor produces a 45º rotation. A 400Ω resistor causes a very small rotation angle. Resistors lower than 1Ω will appear as a short circuit (vertical trace), and resistance values above 400Ω will look like an open circuit (horizontal trace).

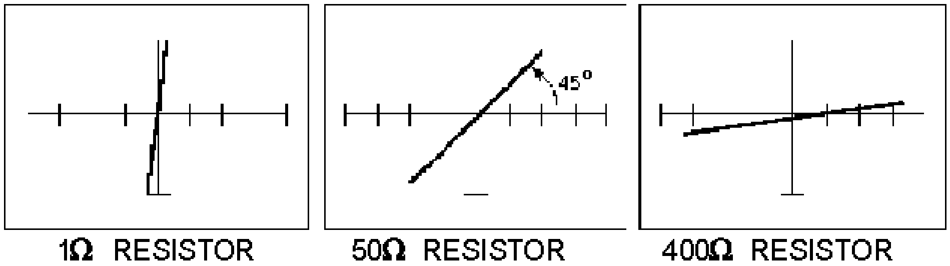


Figure 5-30.—Effects of resistance on the rotation angle in low range.

MEDIUM 1 RANGE.—The medium 1 range is designed to test for resistance values between 50Ω and 10KΩ. Figure 5-31 shows the signatures for a 50Ω resistor, a 1KΩ resistor, and a 10KΩ resistor using the medium 1 range. Resistors that are smaller than 50Ω display a signature that is almost a vertical line. A 1KΩ resistor causes a change in the angle of rotation of 45º, whereas the display of a 10KΩ resistor shows only a slight rotation. Resistance values under test higher than 10KΩ produce a signature with such a small rotation angle that it almost appears to be a horizontal line.

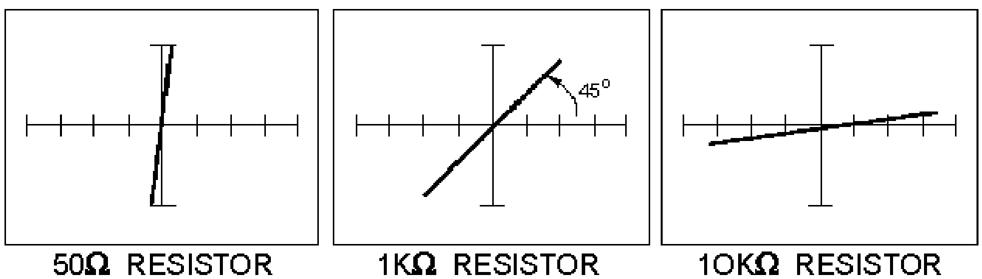


Figure 5-31.—Effects of resistance on the rotation angle in medium 1 range.

Q-11. Medium 1 range is designed to check what resistance values?

MEDIUM 2 RANGE.—The medium 2 range is designed to test for resistance values between 1KΩ and 200KΩ Figure 5-32 shows the signatures for a 1KΩ resistor, a 15KΩ resistor, and a 200KΩ resistor in the medium 2 range. Resistance values that are smaller than 1KΩ will appear to be almost a vertical line. A 15KΩ resistor causes a change in the angle of rotation of 45º, whereas the display for a 200KΩ resistor shows only a slight rotation. When resistance values being tested are higher than 200KΩ, the displayed signature that they produce will have such a small rotation that it appears to be almost a horizontal line.

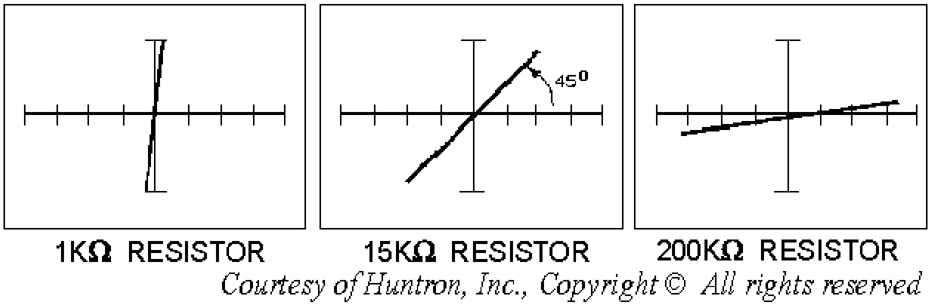


Figure 5-32.—Effects of resistance on the rotation angle in medium 2 range.

HIGH RANGE.—The high range is designed to test resistance values between 3KΩ and 1MΩ. Figure 5-33 shows the signatures that would be displayed for a 3KΩ resistor, a 50KΩ resistor, and a 1MΩ resistor using the high range. Resistors that are smaller than 3KΩ will appear to be almost a vertical line. A 50KΩ resistor will cause a change in the angle of rotation of 45º, whereas the display for a 1MΩ resistor shows only a slight rotation that is very close to a horizontal line. Resistance values higher than 1MΩ will produce such a small rotation that it appears to be a horizontal line.

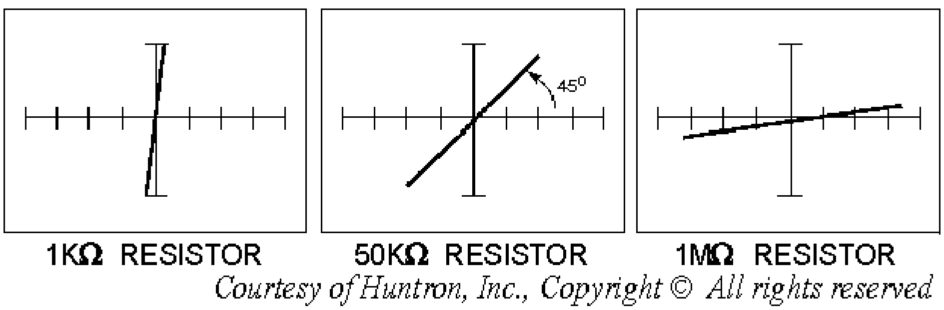


Figure 5-33.—Effects of resistance on the rotation angle in high range.

Testing Capacitors

When you test capacitors, the signature will be displayed as an ellipse. The size and shape of the ellipse depend on the capacitor value, test signal frequency, and the selected impedance range. Figure 5-34 shows the signature of a 0.22µF capacitor in each of the 12 combinations of range and frequency. As you review this figure, you will notice that the signature appears to be an open circuit in the low range at 60 Hz; while in the high range at 2000 Hz, the signature appears to be a short. Between these, the signatures displayed are a variety of ellipsoids, which demonstrates that certain range and frequency combinations are better than others for examining a capacitor.

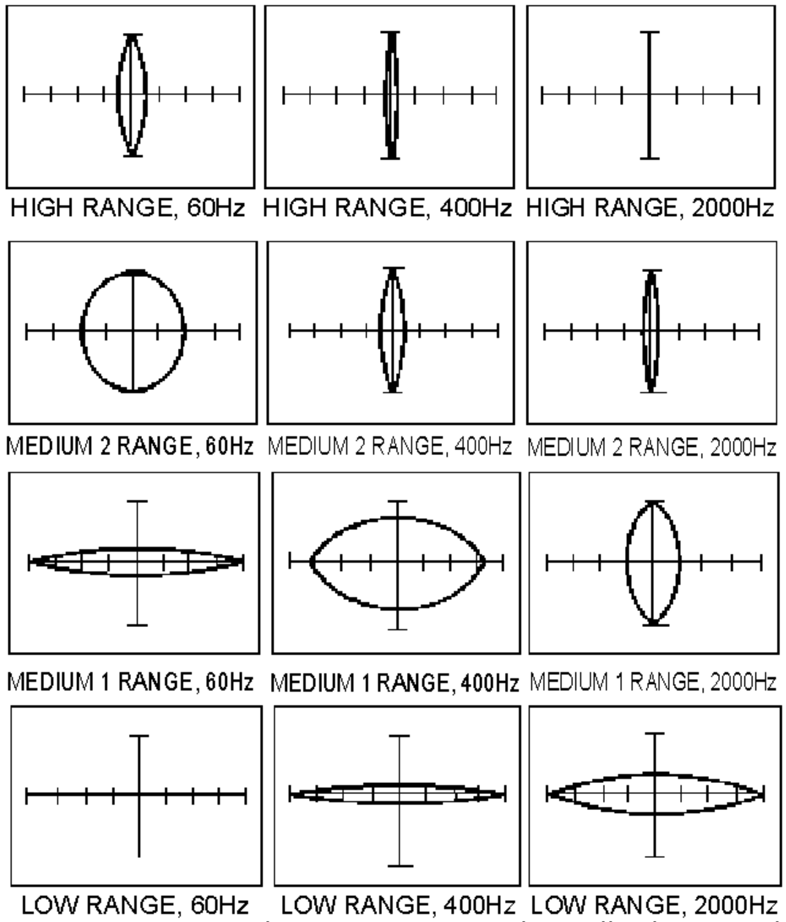


Figure 5-34.—Signature of a 0.22µF capacitor.

Table 5-5 lists the range of capacitance covered by each of the 12 range and frequency combinations for the Tracker 2000. The lowest value of capacitance in each combination will produce a signature of a very narrow horizontal ellipsoid. Capacitors with less of a value than these will appear to be an open. The uppermost value of capacitance in each combination will produce a very narrow vertical ellipsoid signature. Capacitors of greater value than these will appear as a vertical line signature of a short circuit.

Table 5-5.—Min/Max Capacitance Values

RANGE	TEST FREQUENCY		
	50/60 Hz	400 Hz	2000 Hz
HIGH	.001µF-1µF	500pF-.1µF	100pF-.02µF
MEDIUM 1	.01µF-2µF	.001µ-.5µF	200pF-.05µF
MEDIUM 2	.2µF-50µF	.02µF-5µF	.005µ-1µF
LOW	5µF-2000µF	.5mF-100µF	.2mF-25µF

Testing Inductors

Inductors, like capacitors, produce an elliptical signature on the Tracker 2000. Figure 5-35 shows you the signatures produced in each of the 12 range and frequency combinations by a 250mH inductor.

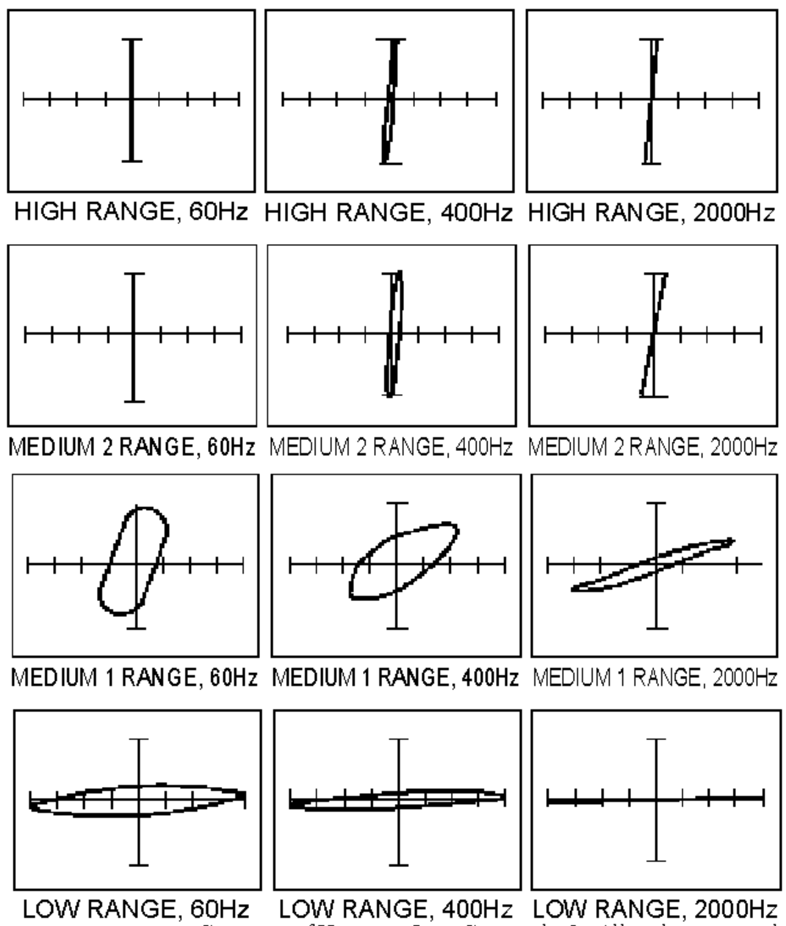


Figure 5-35.—Signatures of a 250mH inductor.

Ferrite inductors can also be checked using this unit; however, the signature produced will be different. Ferrite inductors operate well at high frequencies, but saturate at low frequencies. Figure 5-36 shows the signature of a 490mH ferrite inductor tested at 60 Hz. In low and medium 1, you can see that the signature shows distortion. However, in medium 2 and high ranges, the impedance of the inductor is low compared with the internal impedance of the Tracker 2000, so the signatures are a split vertical trace. Figures 5-37 and 5-38 show the same 490mH inductor being tested at 400 Hz and 2000 Hz.

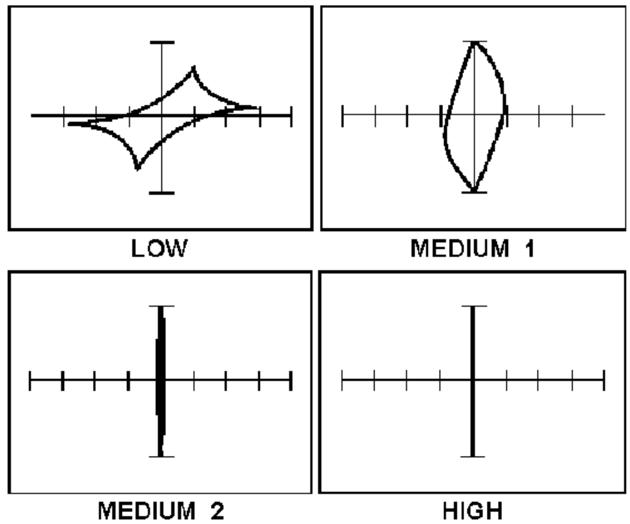


Figure 5-36.—Signatures of a 490mH ferrite inductor tested at 60 Hz.

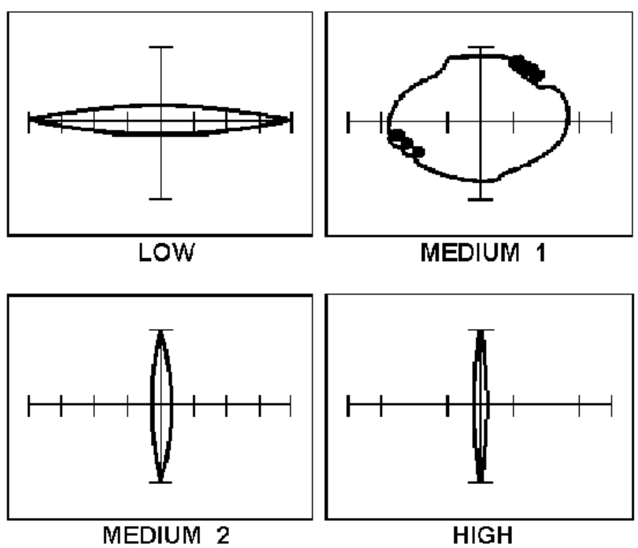


Figure 5-37.—Signature of a 490mH ferrite inductor at 400 Hz.

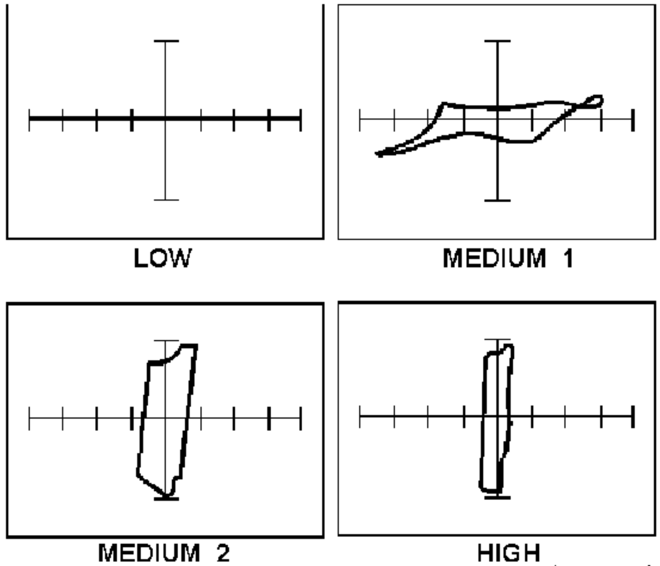


Figure 5-38.—Signature of a 490mH ferrite inductor at 2000 Hz.

SUMMARY

Now that we have completed this chapter, let's review the more important points.

The type of meter used to measure power is the **POWER METER**.

One type of wattmeter is the AN/URM-120. It is an **IN-LINE TYPE WATTMETER**.

The principal function of a **SIGNAL GENERATOR** is to produce an alternating voltage of the desired frequency and amplitude, which has the necessary modulation for the test or measurement concerned.

There are basically two types of signal generators: **AF** and **RF FREQUENCY GENERATORS**.

The instrument used to determine the frequency of a signal is the **FREQUENCY COUNTER**.

An instrument of great value to a technician in troubleshooting digital integrated logic circuits is the **LOGIC PROBE**.

The **HUNTRON TRACKER 2000** is a versatile piece of test equipment that is used to compare known good devices against those of unknown quality or troubleshoot to the component level after power is disconnected to the device under test.

ANSWERS TO QUESTIONS Q1. THROUGH Q11.

A-1. Load.

A-2. An unbalance in the metering bridge.

A-3. Attenuator.

A-4. Oscillator circuit, modulator, and output control circuit.

A-5. To produce an af (or video) signal that can be superimposed on the rf signal produced by the oscillator.

A-6. 1 MHz and 10 MHz.

A-7. DIM.

A-8. Positive voltage (+V) and negative current (–I).

A-9. TRACE ROTATE control.

A-10. 0 to 5 volts.

A-11. 50Ω to 10KΩ.

CHAPTER 6

THE OSCILLOSCOPE AND SPECTRUM ANALYZER

LEARNING OBJECTIVES

Upon completing this chapter, you should be able to:

1. Describe the purpose of the CRT used in the oscilloscope.
2. Explain the operation of an oscilloscope.
3. Describe the purpose of the controls and indicators found on an oscilloscope.
4. Describe the proper procedure for using a dual-trace oscilloscope.
5. Describe the accessory probes available for use with a dual-trace oscilloscope.
6. Explain the operation of the spectrum analyzer.
7. Describe the purpose of the controls and indicators found on the spectrum analyzer.

INTRODUCTION

One of the most widely used pieces of electronic test equipment is the OSCILLOSCOPE. An oscilloscope is used to show the shape of a video pulse appearing at a selected equipment test point. Although some oscilloscopes are better than others in accurately showing video pulses, all function in fundamentally the same way. If you learn how one oscilloscope operates, you will be able to learn others.

As you will learn in this chapter, there are many different types of oscilloscopes—varying in complexity from the simple to the complex. Before we get into our discussion of the dual-trace oscilloscope, we will first present a general overview of basic single-trace oscilloscope operation. Shortly, we will see how oscilloscopes use a CATHODE-RAY TUBE (CRT) in which controlled electron beams are used to present a visible pattern of graphical data on a fluorescent screen.

Another piece of test equipment used is the SPECTRUM ANALYZER. This test equipment is used to sweep over a band of frequencies to determine what frequencies are being produced by a specific circuit under test, and then the amplitude of each frequency component. An accurate interpretation of the display will allow you to determine the efficiency of the equipment being tested.

CATHODE-RAY TUBES

A detailed discussion of CATHODE-RAY TUBES (CRTs) is presented in NEETS, *Module 6, Electronic Emission, Tubes, and Power Supplies*. Before continuing with your study of CRTs in this section, you may want to review chapter 2 of that module.

Cathode-ray tubes used in oscilloscopes consist of an ELECTRON GUN, a DEFLECTION SYSTEM, and a FLUORESCENT SCREEN. All of these elements are enclosed in the evacuated space inside the glass CRT. The electron gun generates electrons and focuses them into a narrow beam. The deflection system moves the beam horizontally and vertically across the screen. The screen is coated with a phosphorous material that glows when struck by the electrons. Figure 6-1 shows the construction of a CRT.

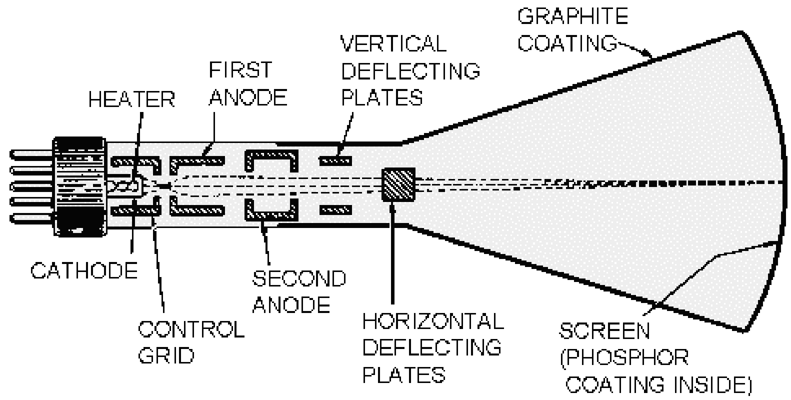


Figure 6-1.—Construction of a CRT.

ELECTRON GUN

The ELECTRON GUN consists of a HEATER and a CATHODE to generate electrons, a CONTROL GRID to control brightness by controlling electron flow, and two ANODES (FIRST and SECOND). The main purpose of the first (FOCUSING) anode is to focus the electrons into a narrow beam on the screen. The second (ACCELERATING) anode accelerates the electrons as they pass. The control grid is cylindrical and has a small opening in a baffle at one end. The anodes consist of two cylinders that contain baffles (or plates) with small holes in their centers.

Q-1. What element controls the number of electrons striking the screen?

Q-2. What element is controlled to focus the beam?

Cathode and Control Grid

As in most conventional electron tubes, the cathode is indirectly heated and emits a cloud of electrons. The control grid is a hollow metal tube placed over the cathode. A small opening is located in the center of a baffle at the end opposite the cathode. The control grid is maintained at a negative potential with respect to the cathode to keep the electrons bunched together.

A high positive potential on the anodes pulls electrons through the hole in the grid. Because the grid is near the cathode, it can control the number of electrons that are emitted. As in an ordinary electron tube, the negative voltage of the grid can be varied either to control electron flow or stop it completely. The brightness (intensity) of the image on the fluorescent screen is determined by the number of electrons striking the screen. This is controlled by the voltage on the control grid.

Electrostatic Lenses and Focusing

The electron beam is focused by two ELECTROSTATIC FIELDS that exist between the control grid and first anode and between the first and second anodes.

Figure 6-2 shows you how electrons move through the electron gun. The electrostatic field areas are often referred to as LENSES because the fields bend electron streams in the same manner that optical lenses bend light rays. The first electrostatic lens cause the electrons to cross at the first focal point within the field. The second lens bend the spreading streams and return them to a new, second focal point at the CRT.

Q-3. Why are the electrostatic fields between the electron gun elements called lenses?

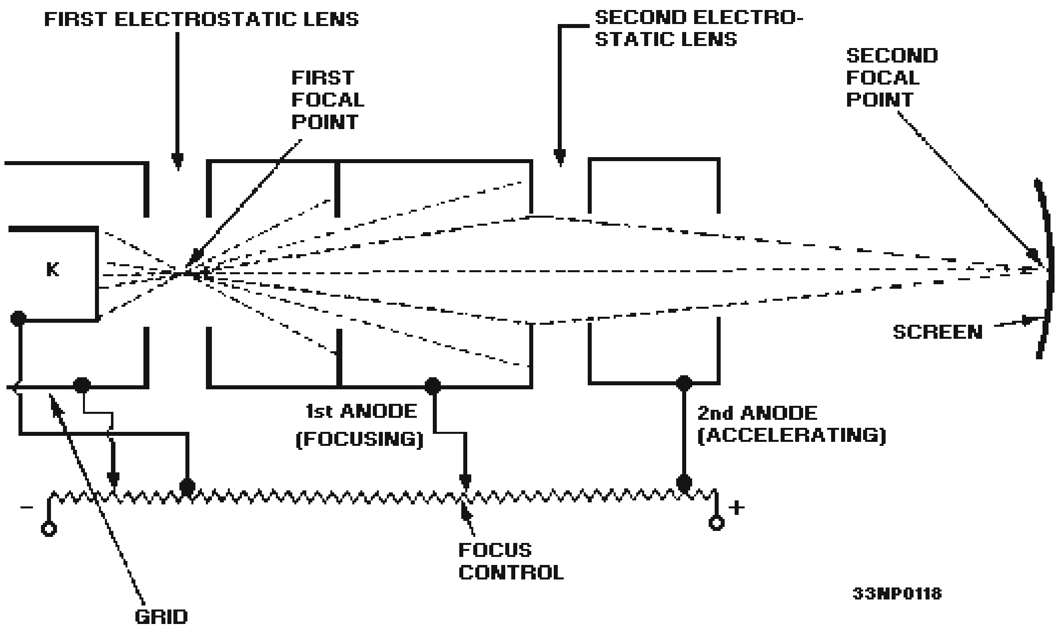


Figure 6-2.—Formation of an electron beam.

Figure 6-2 also shows the relative voltage relationships on the electron-gun elements. The cathode (K) is at a fixed positive voltage with respect to ground. The grid is at a variable negative voltage with respect to the cathode. A fixed positive voltage of several thousand volts is connected to the second (accelerating) anode. The potential of the first (focusing) anode is less positive than the potential of the second anode. The first anode can be varied to place the focal point of the electron beam on the screen of the tube. Control-grid potential is established at the proper level to allow the correct number of electrons through the gun for the desired image intensity.

Q-4. What is the function of the second anode?

ELECTRON BEAM-DEFLECTION SYSTEM

The electron beam is developed, focused, and accelerated by the electron gun. The beam appears on the screen of the CRT as a small, bright dot. If the beam is left in one position, the electrons will soon burn away the illuminating coating in that one area. To be of any use, the beam must be able to move. As you have studied, an electrostatic field can bend the path of a moving electron.

As you have seen in the previous illustrations, the beam of electrons passes through an electrostatic field between two plates. You should remember that electrons are negatively charged and that they will be deflected in the direction of the electric force (from negative to positive). This deflection causes the electrons to follow a curved path while in the electrostatic field.

When the electrons leave the electrostatic field, they will take a straight path to the screen at the angle at which they left the field. Because they were all deflected equally, the electrons will be traveling toward the same spot. Of course, the proper voltages must exist on the anodes to produce the electrostatic field. Changing these voltages changes the focal point of the beam and causes the electron beam to strike the CRT at a different point.

Factors Influencing Deflection

The ANGLE OF DEFLECTION (the angle the outgoing electron beam makes with the CRT center line axis between the plates) depends on the following factors:

- Length of the deflection field;
- Spacing between the deflection plates;
- The difference of potential between the plates; and
- The accelerating voltage on the second anode.

LENGTH OF DEFLECTION FIELD.—As shown in figure 6-3, a long field (long deflection plates) has more time to exert its deflecting forces on an electron beam than does a shorter field (short deflection plates). Therefore, the longer deflection plates can bend the beam to a greater deflection angle.

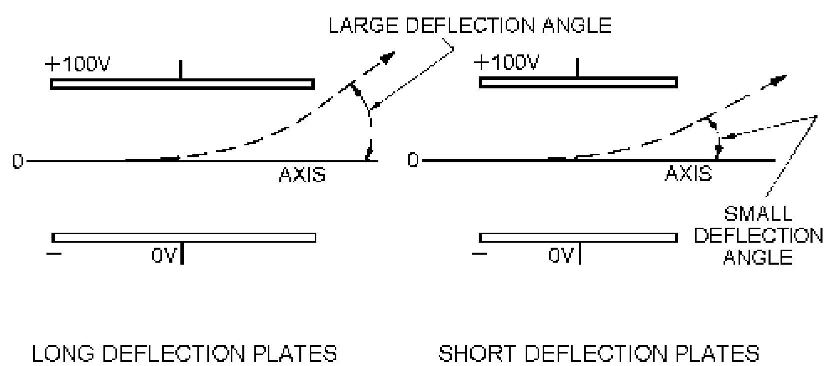


Figure 6-3.—Factors influencing length of field.

Q-5. What effect do longer deflection plates have on the electron beam?

SPACING BETWEEN PLATES.—As shown in figure 6-4, the closer together the plates, the more effect the electric force has on the deflection angle of the electron beam.

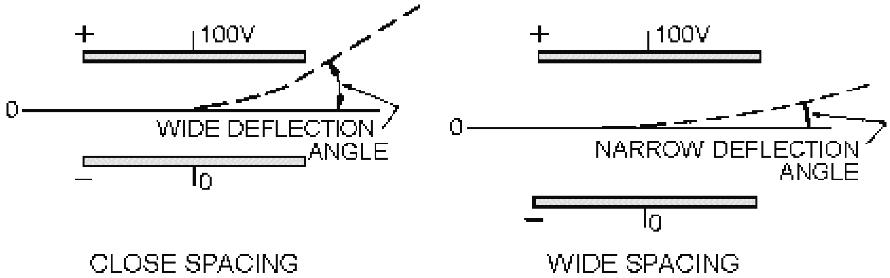


Figure 6-4.—Spacing between plates.

Q-6. What effect does closer spacing of plates have on the electron beam?

DIFFERENCE OF POTENTIAL.— The potential on the plates (figure 6-5) can be varied to cause a wider or narrower deflection angle. The greater the potential, the wider the deflection angle.

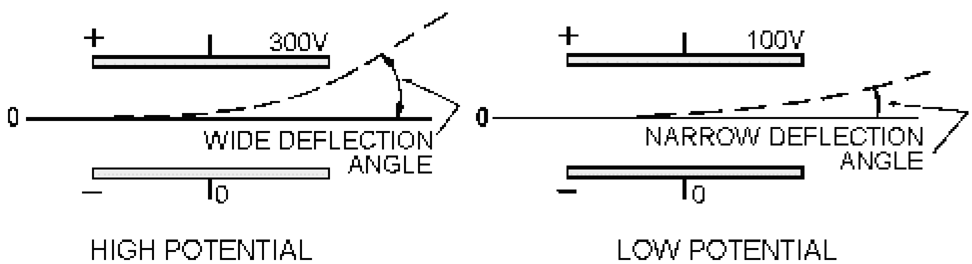


Figure 6-5.—Differences of potential.

Q-7. Is the deflection angle greater with higher or lower potential on the plates?

BEAM ACCELERATION.—The faster the electrons are moving, the smaller their deflection angle will be, as shown in figure 6-6.

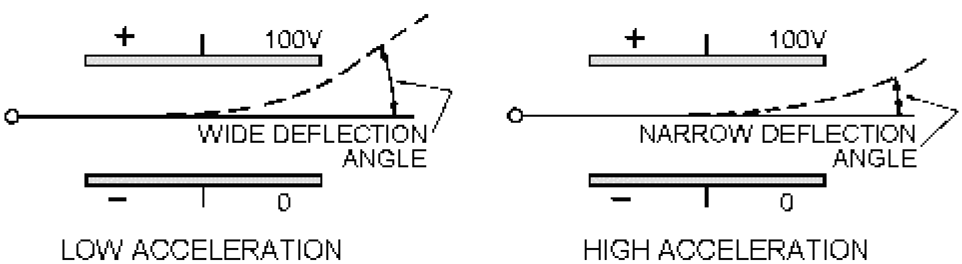


Figure 6-6.—Beam acceleration.

Q-8. Is the deflection angle greater when the beam is moving faster or slower?

Vertical and Horizontal Plates

If two sets of deflection plates are placed at right angles to each other inside a CRT (figure 6-7), the electron beam can be controlled in any direction. By varying the potential of the vertical-deflection plates, you can make the spot (beam) on the face of the tube move vertically. The distance the beam moves will be proportional to the change in potential difference between the plates. Changing the potential difference between the horizontal-deflection plates will cause the beam to move a given distance from one side to the other. Directions other than up-down and left-right are achieved by a combination of horizontal and vertical movement.

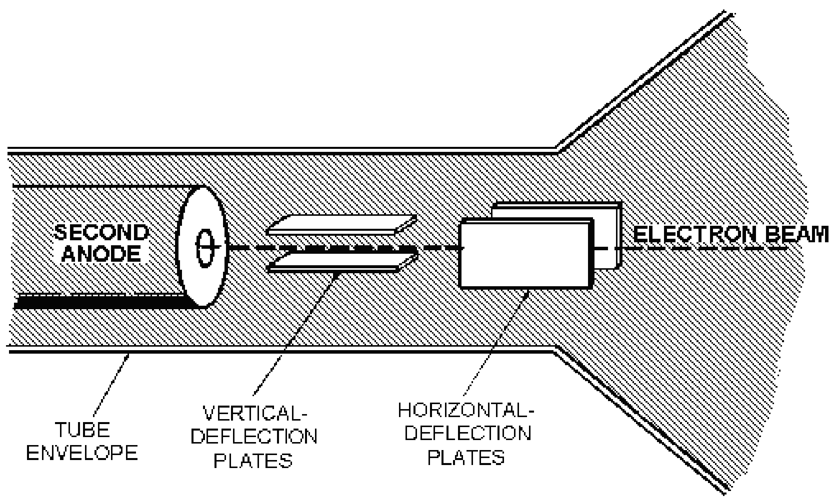


Figure 6-7.—Deflection plate arrangement.

As shown in figure 6-8, position X of the beam is in the center. It can be moved to position Y by going up 2 units and then right 2 units. Movement of the beam is the result of the simultaneous action of both sets of deflection plates. The electrostatic field between the vertical plates moves the electrons up an amount proportional to 2 units on the screen. As the beam passes between the horizontal plates, it moves to the right an amount proportional to 2 units on the screen.

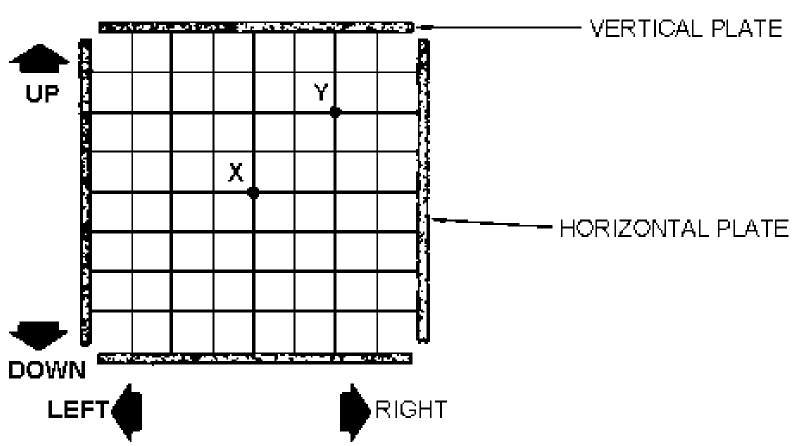


Figure 6-8.—Beam movement on the CRT.

If the amount of deflection from the left and down occurred so that each set of plates acted at the same time, the picture would be like the one in view A of figure 6-9. For example, if the vertical plates moved the beam downward (starting from point X) at the rate of 3 units per second and the horizontal plates moved it to the left at the rate of 1 unit per second, both movements would have been completed in 1 second at point Y. The result would be a straight line.

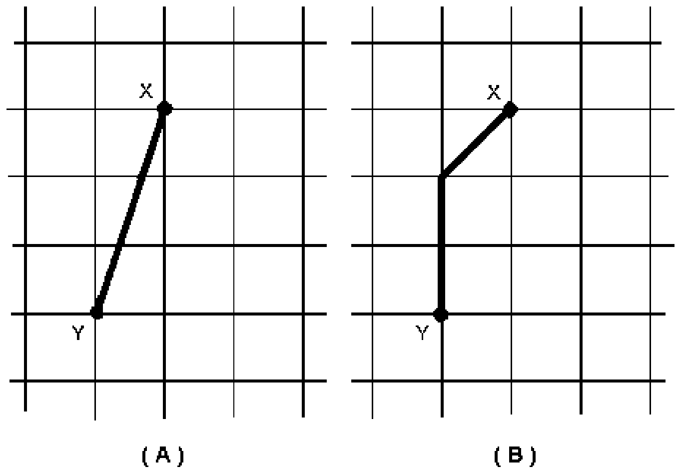


Figure 6-9.—Deflection of the beam.

In view B, the potentials on the vertical and horizontal plates change at the same rate. In the same time period, say 1 second, both plates move the beam 1 unit. The horizontal plates have completed their task at the end of 1 second, but the vertical plates have moved the beam only one-third of the required distance. In this case, the picture in view B would appear on the screen.

Beam-deflection Plate Action

Recall from your study of chapter 2 of this module that waveforms are described in terms of amplitude versus time. You have just seen how the movement of the CRT beam depends on both potential (amplitude) and time.

Q-9. Waveforms are described in terms of what two functions?

VERTICAL-DEFLECTION PLATES.—We will use figure 6-10 to explain the action of the vertical-deflection plates in signal amplitude measurements. As this discussion begins, remember that vertical-deflection plates are used to show amplitude of a signal, and horizontal-deflection plates are used to show time and/or frequency relationships.

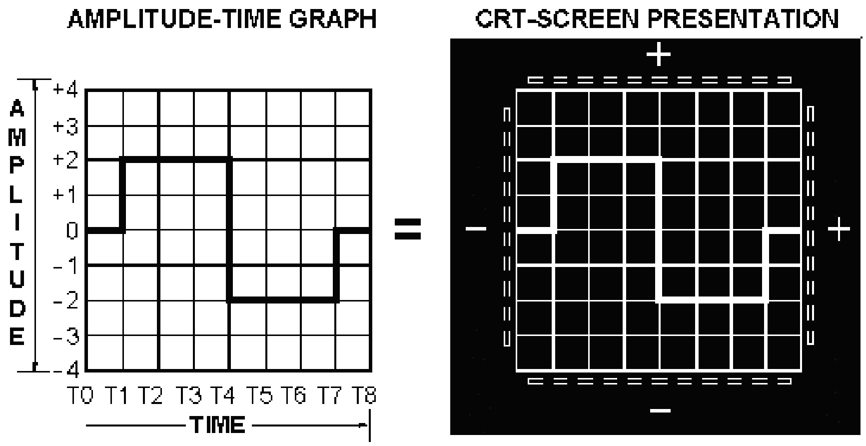


Figure 6-10.—Amplitude versus time.

1. From T0 to T1, the vertical plates maintain their static difference in potential and the beam stays at 0 units; the T0 to T1 change causes an increasing potential difference in the horizontal plates, and the beam moves 1 unit to the right.

2. At T1, a positive potential difference change in the vertical plates occurs, which causes the beam to move up (instantaneously) 2 units. This vertical (amplitude) beam location is maintained from T1 to T4; horizontal beam movement continues moving to the right as 3 units of time pass.

3. At T4, an instantaneous negative change in potential of 4 units in amplitude occurs, and the beam moves from +2 to −2 units.

4. From T4 to T7, the beam remains at −2 units. During this time period, the beam continues moving horizontally to the right, indicating the passage of time.

5. At T7, a positive increase of amplitude occurs, and the beam moves vertically from −2 to 0 units. From T7 to T8, no change occurs in vertical beam movement; however, horizontal movement continues with time.

The vertical-plate potential difference follows the voltage of the waveform. The horizontal-plate potential follows the passage of time. Together, they produce the image (trace) produced on the screen by the moving beam.

Q-10. The vertical-deflection plates are used to reproduce what function?

Q-11. The horizontal-deflection plates are used to produce what function?

HORIZONTAL-DEFLECTION PLATES.—Now let's look at horizontal-deflection action. Assume that the resistance of the potentiometer shown in figure 6-11 is spread evenly along its length. When the arm of the potentiometer is at the middle position, the same potential exists on each plate. Since there is zero potential difference between the plates, an electrostatic field is not moved downward at a uniform rate; the right plate will become more positive than the left (you are looking down through the top of the CRT). The electron beam will move to the right from screen point 0 through points 1, 2, 3, and 4 in equal time intervals.

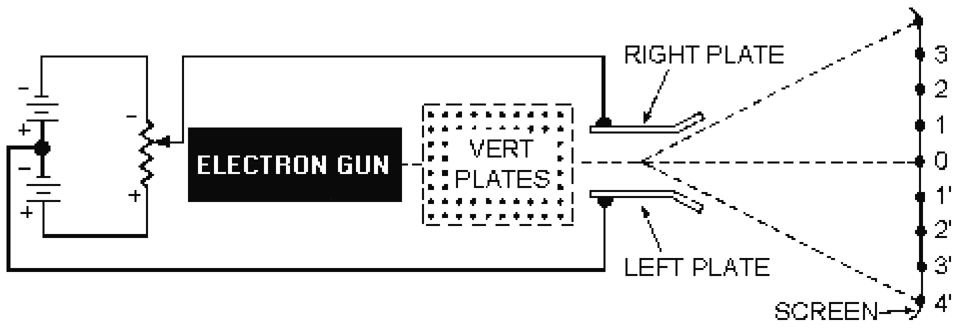


Figure 6-11.—Horizontal plates (top view).

If the potentiometer arm is moved at the same rate in the opposite direction, the right plate will decrease in positive potential until the beam returns to the 0 position. At that point, the potential difference between the plates is again zero. Moving the arm toward the other end of the resistance causes the left plate to become more positive than the right, and the beam moves from screen points 0 through 4. If the movement of the potentiometer arm is at a uniform (linear) rate, the beam moves at a uniform rate.

Notice that the ends of the deflection plates are bent outward to permit wide-angle deflection of the beam. The vertical plates are bent up and down in the same manner.

Q-12. Why are the ends of the deflection plates bent outward?

For ease of explanation, the manual movement of the potentiometer arm is satisfactory to introduce you to horizontal beam movement. However, in the oscilloscope this is not how horizontal deflection is accomplished. Beam movement voltages are produced much faster by sawtooth circuitry. You may want to review the sawtooth generation section in NEETS, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuitry* before continuing. Nearly all oscilloscopes with electrostatic deflection apply a sawtooth voltage to the horizontal plates to produce horizontal deflection of the beam, as shown in figure 6-12.

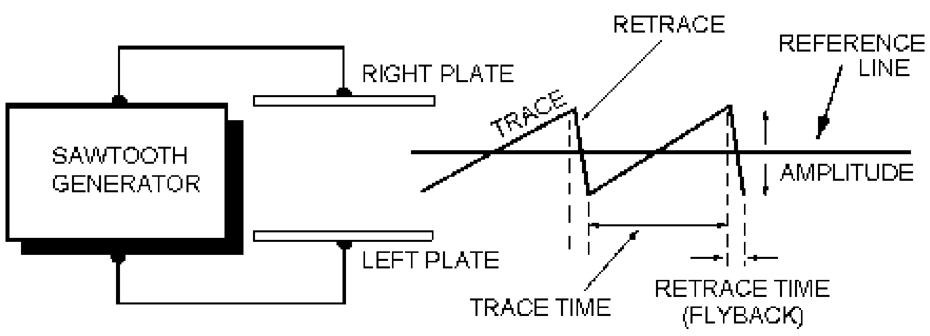


Figure 6-12.—Sawtooth generator.

In the figure, the sawtooth generator replaces the potentiometer and is connected to both horizontal plates of the CRT. At the reference line, the potential on both plates is equal. Below the line, the left plate is more positive and the right plate is less positive. This causes the beam to move left. Above the line, the right plate is made more positive than the left and the beam moves to the right. The waveform amplitude causes a uniform movement of the beam across the screen (called TRACE). RETRACE time, shown at the trailing edge of the waveform, quickly deflects the beam back to the starting point.

CRT GRATICULE

A GRATICULE was used in our previous discussion in figure 6-10. It is simply a calibrated scale (made of clear plastic) of amplitude versus time that is placed on the face of the CRT.

The graticule can be used to determine the voltage of waveforms because the DEFLECTION SENSITIVITY of a CRT is uniform throughout the vertical plane of the screen. Deflection sensitivity states the number of inches, centimeters, or millimeters a beam will be deflected for each volt of potential difference applied to the deflection plates. It is directly proportional to the physical length of the deflection plates and their distance from the screen and inversely proportional to the distance between the plates and to the second-anode voltage. Deflection sensitivity is a constant that is dependent on the construction of the tube.

Deflection sensitivity for a given CRT might typically be 0.2 millimeters per volt. This means the spot on the screen will be deflected 0.2 millimeters (about 0.008 inch) when a difference of 1 volt exists between the plates. Sometimes the reciprocal of deflection sensitivity (called DEFLECTION FACTOR) is given. The deflection factor for the example given would be 125 volts per inch (1/0.008).

Q-13. What term is used to describe the reciprocal of deflection sensitivity of a scope?

In the above example, 125 volts applied between one set of plates would deflect the beam 1 inch on the screen. This means that the deflection caused by small signals would likely not be observed. For this reason, the deflection plates are connected to amplifiers that magnify the signals applied to the vertical input of the scope.

Assume, for example, that a peak-to-peak value of a known voltage applied to the oscilloscope indicates that each inch marking on the graticule is equal to 60 volts. Each of the 10 subdivisions will, therefore, equal a value of 6 volts. Most oscilloscopes have ATTENUATOR controls to decrease or GAIN controls to increase the strength of a signal before it is placed on the deflection plates. Attenuator and gain settings must not be disturbed after the calibration has been made. For maximum accuracy, you should recalibrate the graticule each time a voltage is to be measured.

CRT DESIGNATIONS

Cathode-ray tubes are identified by a tube number, such as 2AP1, 2BP4, or 5AP1A. The first number identifies the diameter of the tube face. Typical diameters are 2 inches, 5 inches, and 7 inches. The first letter designates the order in which a tube of a given diameter was registered. The letter-digit combination indicates the type of phosphor (glowing material) used on the inside of the screen. Phosphor P1, which is used in most oscilloscopes, produces a green light at medium PERSISTENCE. Persistence refers to the length of time the phosphor glows after the electron beam is removed. P4 provides a white light and has a short persistence. If a letter appears at the end, it signifies the number of the modification after the original design.

OSCILLOSCOPE CONTROL COMPONENTS

Although the CRT is a highly versatile device, it cannot operate without control circuits. The type of control circuits required depends on the purpose of the equipment in which the CRT is used.

There are many different types of oscilloscopes. They vary from relatively simple test instruments to highly accurate laboratory models. Although oscilloscopes have different types of circuits, most can be divided into the basic sections shown in figure 6-13: (1) a CRT, (2) a group of control circuits that control the waveform fed to the CRT, (3) a power supply, (4) sweep circuitry, and (5) deflection circuitry.

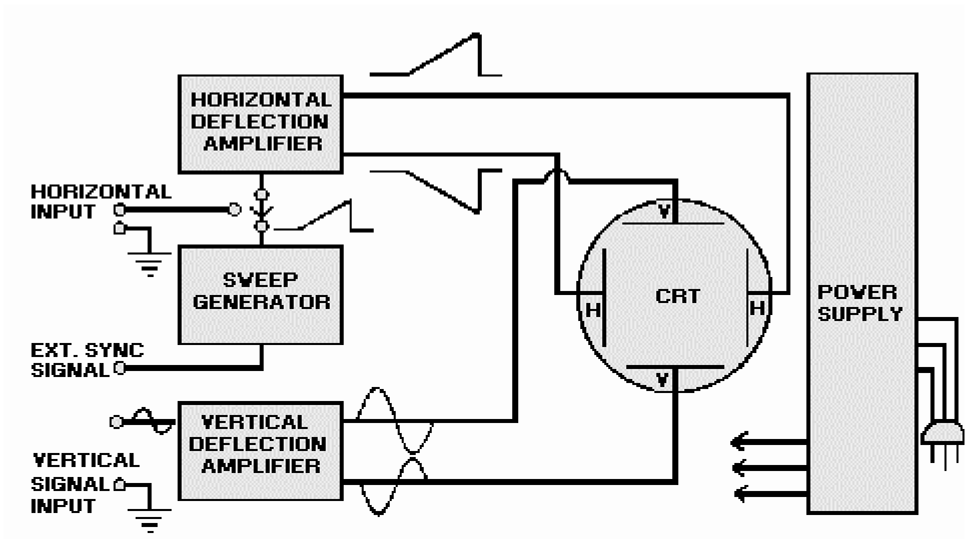


Figure 6-13.—Block diagram of an oscilloscope.

Q-14. List the circuits that all oscilloscopes have in common.

Figure 6-14 is a drawing of the front panel of a dual-trace, general-purpose oscilloscope. Oscilloscopes vary greatly in the number of controls and connectors. Usually, the more controls and connectors, the more versatile the instrument. Regardless of the number, all oscilloscopes have similar controls and connectors. Once you learn the fundamental operation of these common controls, you can move with relative ease from one model of oscilloscope to another. Occasionally, controls that serve similar functions will be labeled differently from one model to another. However, you will find that most controls are logically grouped and that their names usually indicate their function.

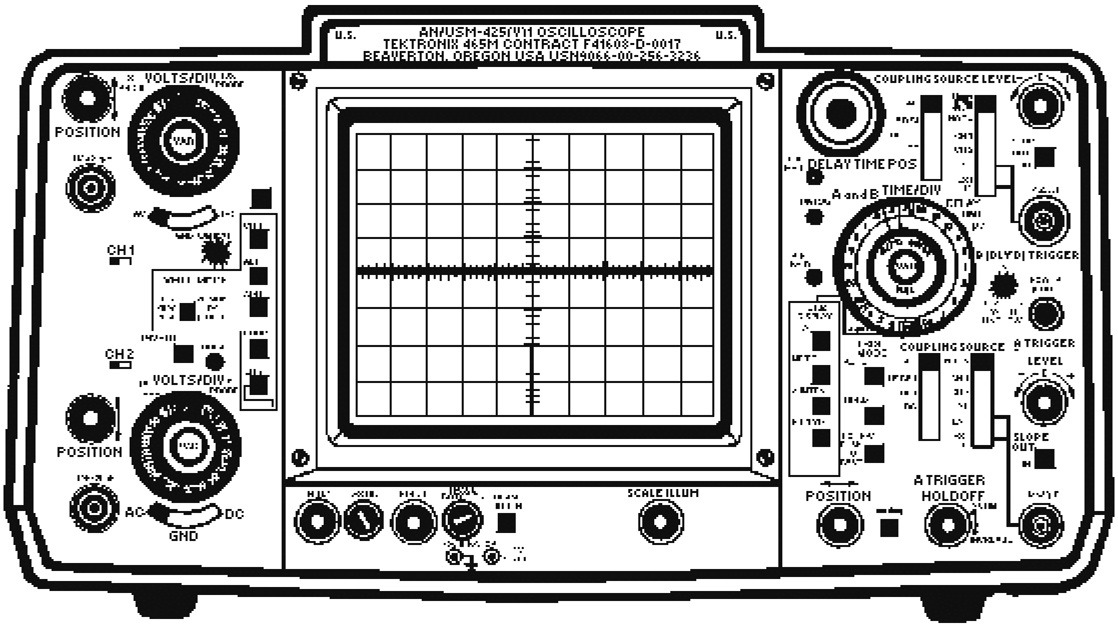


Figure 6-14.—Dual-trace oscilloscope.

The oscilloscope in figure 6-14 is called DUAL-TRACE because it can accept and display two vertical signal inputs at the same time—usually for comparison of the two signals or one signal and a reference signal. This scope can also accept just one input. In this case, it is used as a SINGLE-TRACE OSCILLOSCOPE. For the following discussion, we will consider this to be a single-trace oscilloscope. The oscilloscope in the figure is commonly used in the fleet. You are likely to use this one (model AN/USM-425) or one very similar to it. Let's now look at the front panel controls.

COMPONENTS USED TO DISPLAY THE WAVEFORM

The CRT DISPLAY SCREEN is used to display the signal (figure 6-15). It allows you to make accurate measurements using the vertical and horizontal graticules, as discussed earlier.

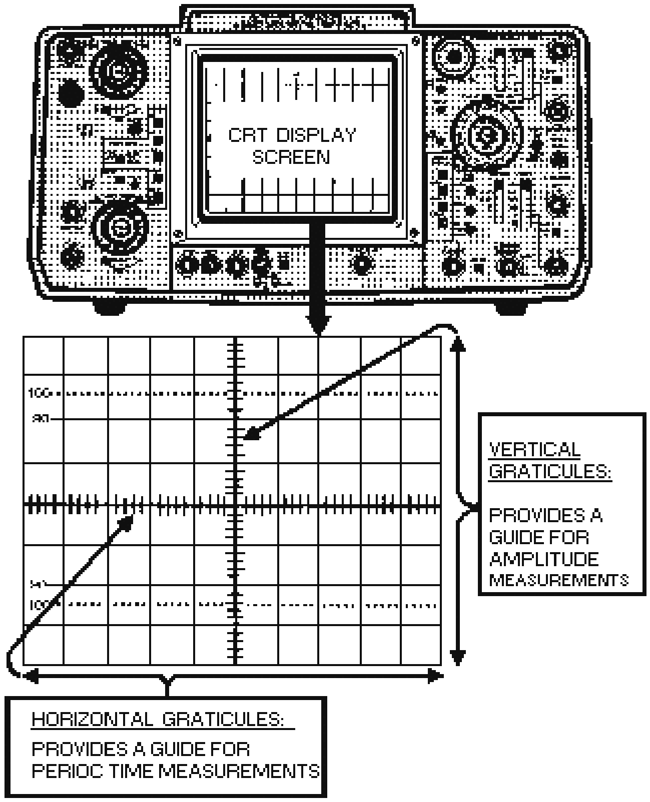


Figure 6-15.—CRT display and graticule.

COMPONENTS USED TO ADJUST CRT DISPLAY QUALITY

The controls in figure 6-16 allow you to adjust for a clear signal display. They also allow you to adjust the display position and magnify the horizontal trace by a factor of 10 (X10). Keep in mind that the controls may be labeled differently from one model to another, depending on the manufacturer. Refer to figure 6-16 as you study the control descriptions in the next paragraphs.

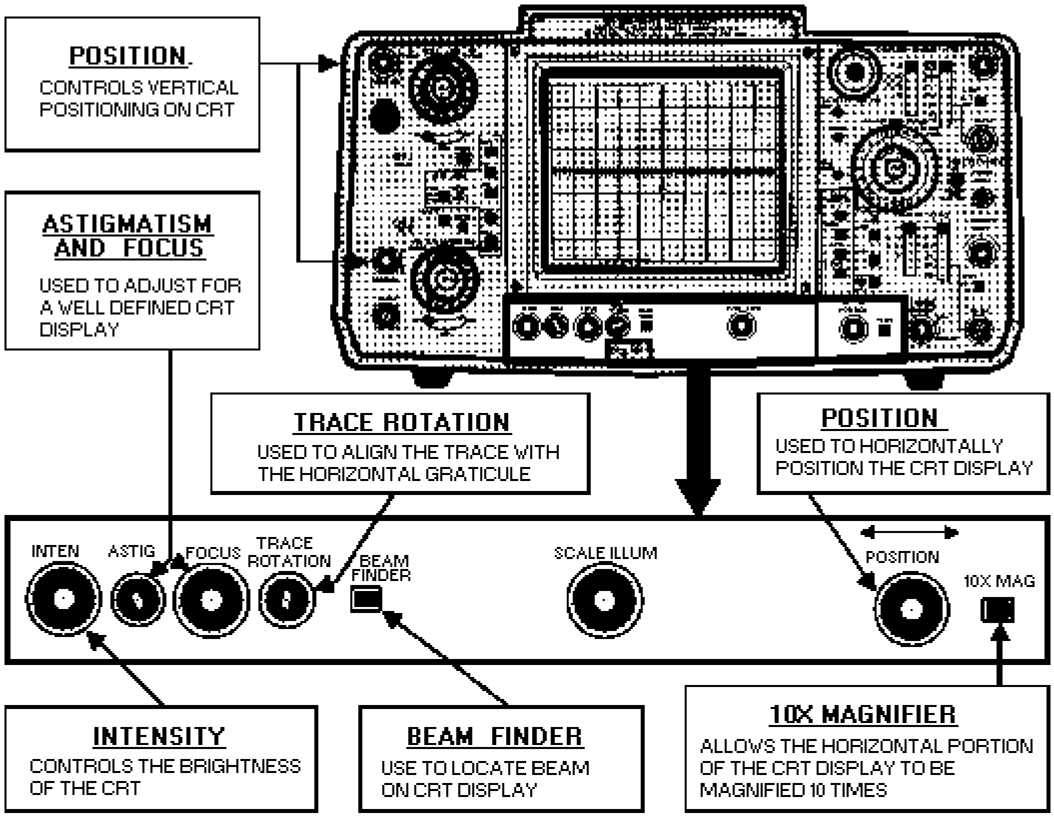


Figure 6-16.—Quality adjustment for CRT display.

INTEN (Intensity) Control

The INTEN (intensity) control (sometimes called BRIGHTNESS) adjusts the brightness of the beam on the CRT. The control is rotated in a clockwise direction to increase the intensity of the beam and should be adjusted to a minimum brightness level that is comfortable for viewing.

FOCUS and ASTIG (Astigmatism) Controls

The FOCUS control adjusts the beam size. The ASTIG (astigmatism) control adjusts the beam shape. The FOCUS and ASTIG controls are adjusted together to produce a small, clearly defined circular dot. When displaying a line trace, you will use these same controls to produce a well-defined line. Figure 6-17, view A, shows an out-of-focus beam dot. View B shows the beam in focus. Views C and D show out-of-focus and in-focus traces, respectively.

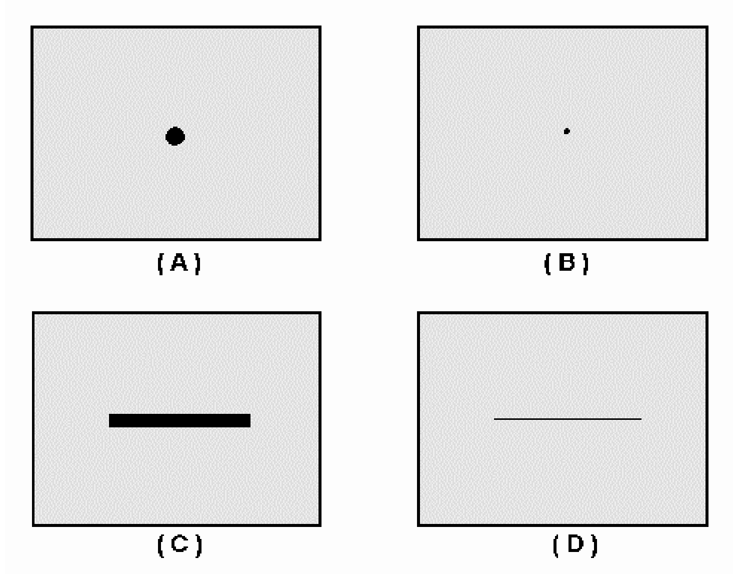


Figure 6-17.—Effects of FOCUS and ASTIG (astigmatism) controls.

TRACE ROTATION Control

The TRACE ROTATION control (figure 6-16) allows for minor adjustments of the horizontal portion of the trace so that you can align it with the horizontal lines on the graticule.

BEAM FINDER Control

Occasionally, the trace will actually be located off the CRT (up or down or to the left or right) because of the orientation of the deflection plates. When pushed, the BEAM FINDER (figure 6-16) pulls the beam onto the screen so that you can use the horizontal and vertical POSITION controls to center the spot.

Horizontal and Vertical POSITION Controls

The horizontal and vertical POSITION controls (figure 6-16) are used to position the trace. Because the graticule is often drawn to represent a graph, some oscilloscopes have the positioning controls labeled to correspond to the X and Y axes of the graph. The X axis represents horizontal movement; the Y axis represents the vertical movement. Figure 6-18 shows the effects of positioning controls on the trace.

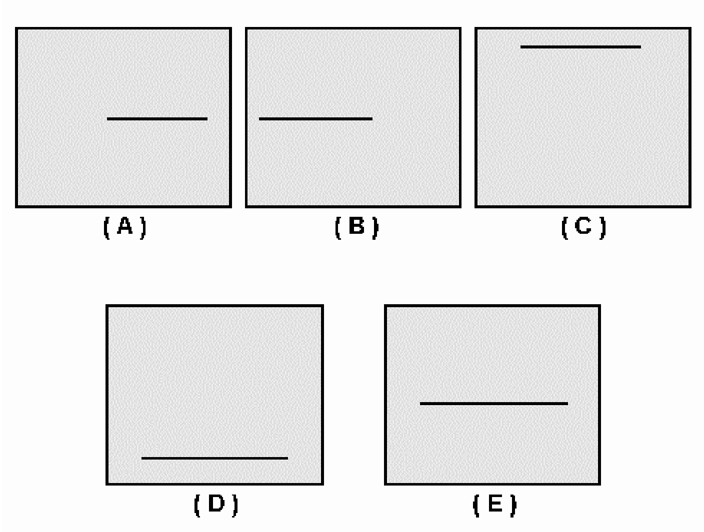


Figure 6-18.—Effects of horizontal and vertical controls.

In view A, the horizontal control has been adjusted to move the trace too far to the right; in view B, the trace has been moved too far to the left. In view C, the vertical POSITION control (discussed later) has been adjusted to move the trace too close to the top; in view D, the trace has been moved too close to the bottom. View E (figure 6-18) shows the trace properly positioned.

10X MAG (Magnifier) Switch

The 10X MAG (magnifier) switch (figure 6-16) allows you to magnify the displayed signal by a factor of 10 in the horizontal direction. This ability is important when you need to expand the signal to evaluate it carefully.

COMPONENTS USED TO DETERMINE THE AMPLITUDE OF A SIGNAL

We will now discuss the dual-trace components of the scope. You will use these components to determine the amplitude of a signal. Notice in figure 6-19 that the highlighted section at the upper left of the scope looks just the same as the section at the lower left of the scope. This reveals the dual-trace capability section of the scope. The upper left section is the CH (channel) 1 input and is the same as the CH 2 input at the lower left. An input to both inputs at the same time will produce two independent traces on the CRT and use the dual-trace capability of the scope.

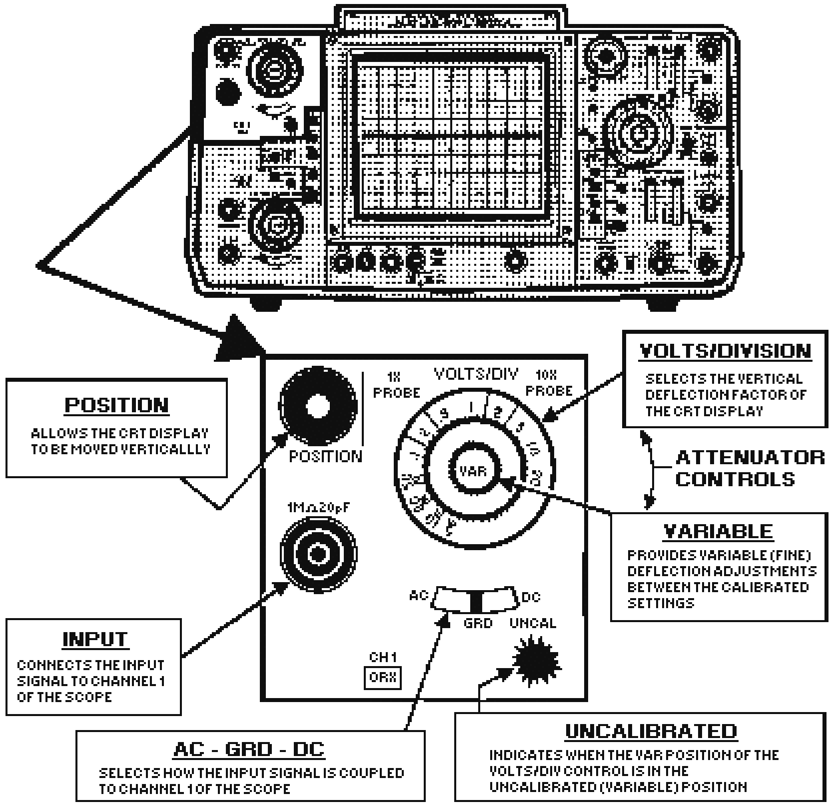


Figure 6-19.—Components that determine amplitude.

For purposes of this introductory discussion, we will present only CH (channel) 1. You should realize that the information presented also applies to CH 2.

Vertical POSITION Control

The vertical POSITION control allows you to move the beam position up or down, as discussed earlier.

Input Connector

The vertical input (or signal input) jack connects the signal to be examined to the vertical-deflection amplifier. Some oscilloscopes may have two input jacks, one labeled AC and the other labeled DC. Other models may have a single input jack with an associated switch, such as the AC GRD DC switch in figure 6-19. This switch is used to select the ac or dc connection. In the DC position, the signal is connected directly to the vertical-deflection amplifier; in the AC position, the signal is first fed through a capacitor. Figure 6-20 shows the schematic of one arrangement.

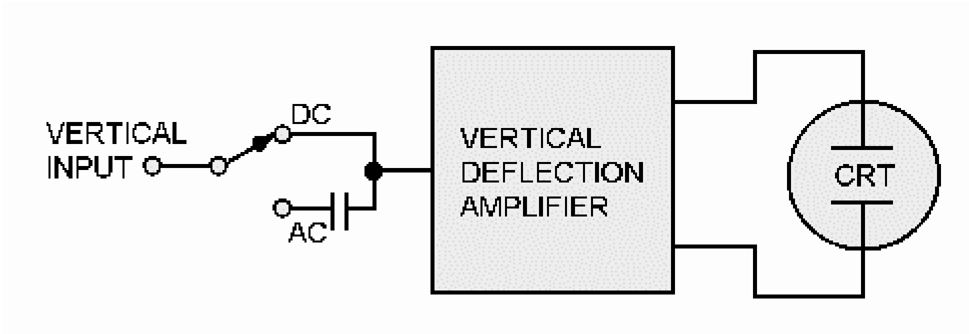


Figure 6-20.—Vertical input arrangement.

The VERTICAL-DEFLECTION AMPLIFIER increases the amplitude of the input signal level required for the deflection of the CRT beam. The deflection amplifier must not have any other effect on the signal, such as changing the shape (called DISTORTION). Figure 6-21 shows the results of distortion occurring in a deflection amplifier.

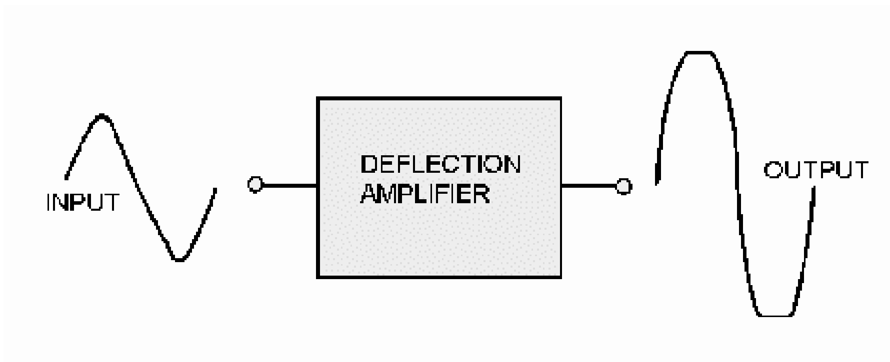


Figure 6-21.—Deflection amplifier distortion.

Attenuator Control

An amplifier can handle only a limited range of input amplitudes before it begins to distort the signal. Signal distortion is prevented in oscilloscopes by the incorporation of circuitry that permits adjustment of the input signal amplitude to a level that prevents distortion from occurring. This adjustment is called the ATTENUATOR control in some scopes (VOLTS/DIV and VAR in figure 6-19). This control extends the usefulness of the oscilloscope by enabling it to handle a wide range of signal amplitudes.

The attenuator usually consists of two controls. One is a multiposition (VOLTS/DIV) control, and the other is a variable (VAR) potentiometer. Each position of the control may be marked either as to the amount of voltage required to deflect the beam a unit distance, such as VOLTS/DIV, or as to the amount of attenuation (called the DEFLECTION FACTOR) given to the signal, such as 100, 10, or 1.

Suppose the .5 VOLTS/DIV position were selected. In this position, the beam would deflect vertically 1 division for every 0.5 volts of applied signal. If a sine wave occupied 4 divisions peak-to-peak, its amplitude would be 2 volts peak-to-peak (4×0.5), as shown in figure 6-22.

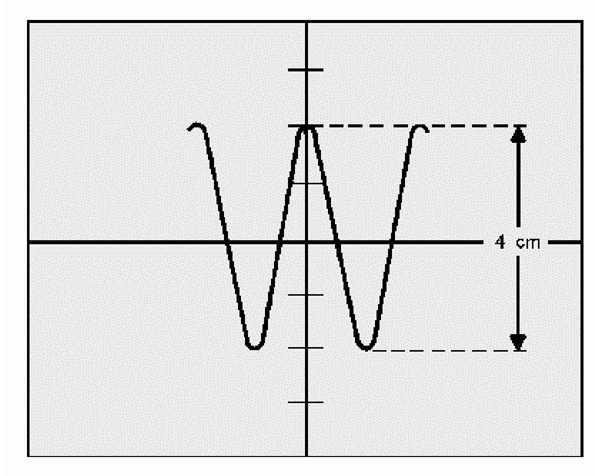


Figure 6-22.—Sine wave attenuation.

The vertical attenuator control (VOLTS/DIV in figure 6-19) provides a means of adjusting the input signal level to the amplifiers by steps. These steps are sequenced from low to high deflection factors. The potentiometer control (VAR in figure 6-19) provides a means of fine, or variable, control between steps. This control may be mounted separately, or it may be mounted on the attenuator control. When the control is mounted separately, it is often marked as FINE GAIN or simply GAIN. When mounted on the attenuator control, it is usually marked VARIABLE or VAR.

The variable control adds attenuation to the step that is selected. Since accurately calibrating a potentiometer is difficult, the variable control is either left unmarked or the front panel is marked off in some convenient units, such as 1-10 and 1-100. The attenuator control, however, can be accurately calibrated. To do this, you turn off the variable control to remove it from the attenuator circuit. This position is usually marked CAL (calibrate) on the panel, or an associated light indicates if the VAR control is on or off. In figure 6-19, the light called UNCAL indicates the VAR control is in the uncalibrated position.

COMPONENTS USED TO SELECT THE VERTICAL OPERATING MODE

As we discussed earlier, channel 1 is being used to discuss basic operating procedures for the oscilloscope. Figure 6-23 shows how the vertical mode of operation is selected. The VERT MODE section contains push-button switches that enable you to select channel 1, channel 2, and several other vertical modes of operation. For the present discussion, note only that CH 1 is selected by these switches.

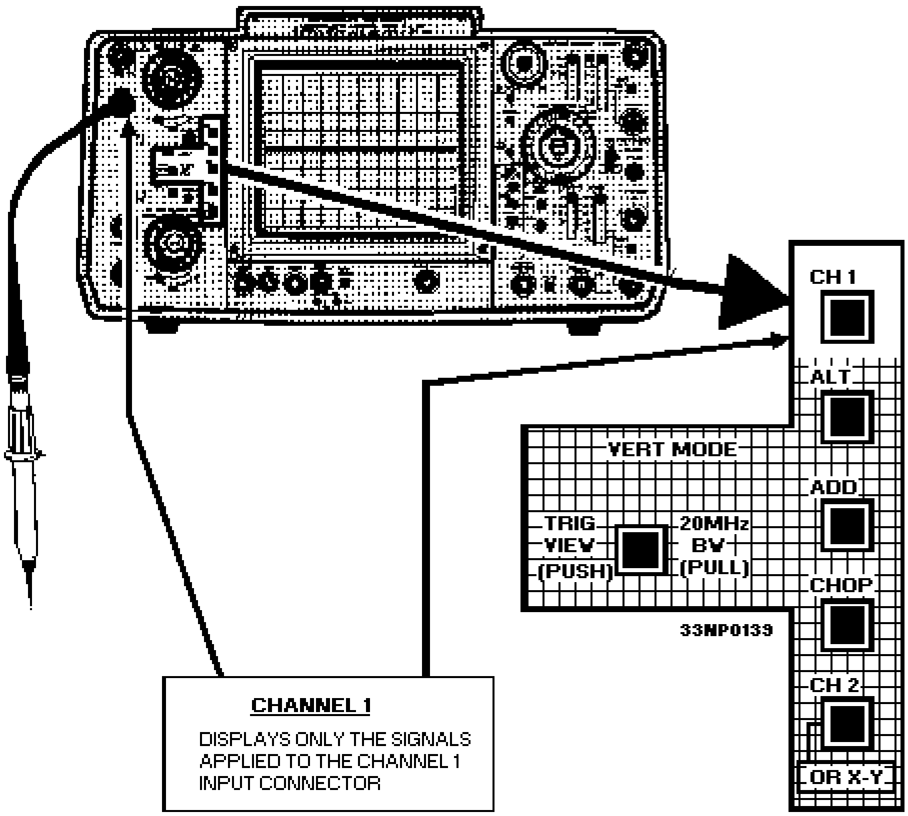


Figure 6-23.—Vertical-deflection controls.

COMPONENTS USED TO DETERMINE PERIOD TIME OF THE DISPLAY

The TIME/DIV (figure 6-24) controls on the scope determine the period time of the displayed waveform. As we discussed earlier, the sweep generator develops the sawtooth waveform that is applied to the horizontal-deflection plates of the CRT. This sawtooth voltage causes the beam to move across the screen. This trace (sometimes called SWEEP) sets the frequency of the TIME BASE of the oscilloscope. The frequency of the time base is variable, which enables the oscilloscope to accept a wide range of input frequencies. Again, two controls are used (figure 6-24). One is a multiposition switch (TIME/DIV) that changes the frequency of the sweep generator in steps. The second control is a potentiometer (VAR) that varies the frequency between steps. Each step on the TIME/DIV control is calibrated. The front panel has markings that group the numbers into microseconds and milliseconds.

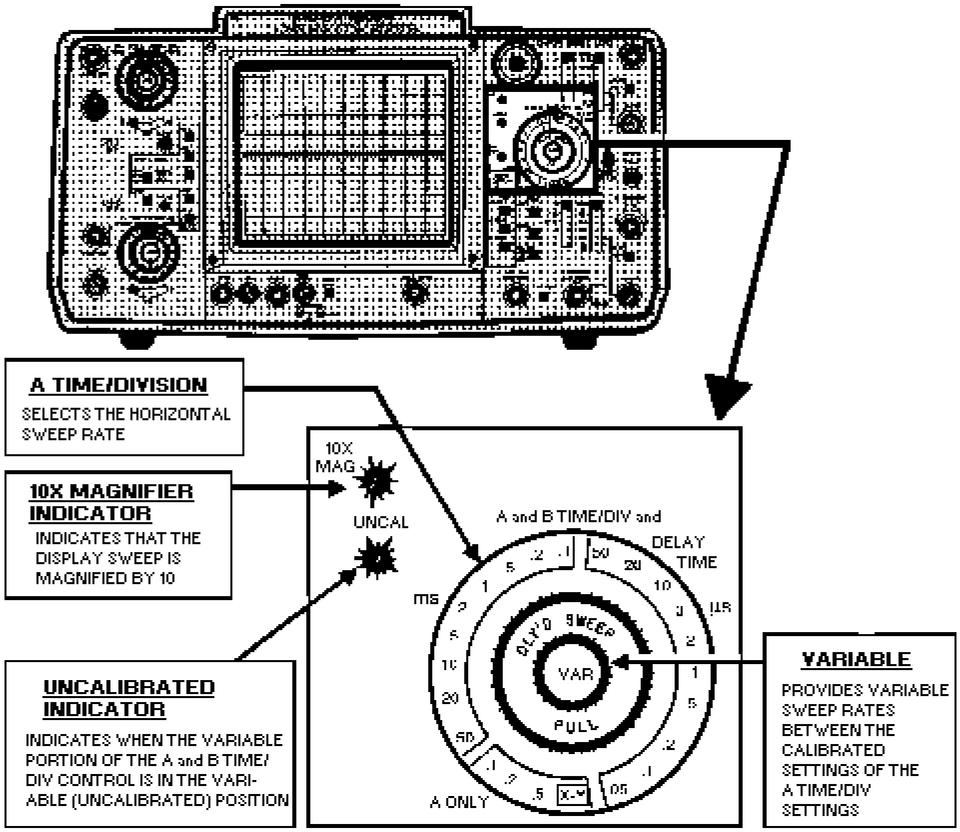


Figure 6-24.—Period time of the waveform (TIME/DIV).

The potentiometer is labeled VAR, and the panel has an UNCAL indicator that lights when the VAR control is in the variable position. When you desire to accurately measure the time of one cycle of an input signal, turn the VAR control to the CAL position and turn the TIME/DIV switch to select an appropriate time base. Suppose you choose the 10-microsecond position to display two cycles of an input signal, as shown in figure 6-25. One cycle occupies 3 centimeters (small divisions) along the horizontal axis. Each cm has a value of 10 microseconds. Therefore, the time for one cycle equals 30 microseconds (3×10). Recall that the frequency for a signal may be found by using the following procedure:

$$
\begin{aligned}
f &= \frac{1}{\text{time (t)}} \\
&= \frac{1}{30 \times 10^{-6}} \\
&= 33.33 \text{ kHz}
\end{aligned}
$$

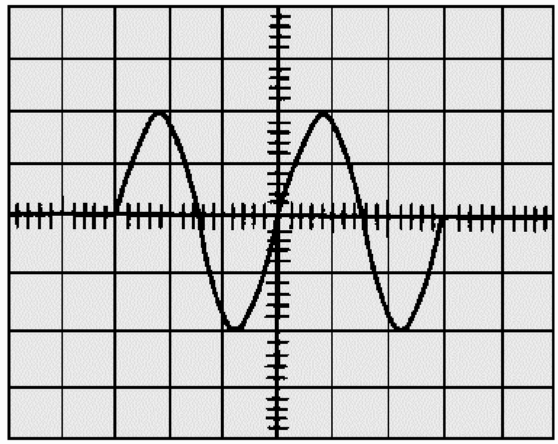


Figure 6-25.—Time measurement of a waveform (TIME/DIV).

In selecting a time base, you should select one that is lower in frequency than the input signal. If the input signal requires 5 milliseconds to complete one cycle and the sawtooth is set for 0.5 milliseconds per centimeter with a 10-centimeter-wide graticule, then approximately one cycle will be displayed. If the time base is set for 1 millisecond per centimeter, approximately two cycles will be displayed. If the time base is set at a frequency higher than the input frequency, only a portion of the input signal will be displayed.

In the basic oscilloscope, the sweep generator runs continuously (FREE-RUNNING); in more elaborate oscilloscopes, it is normally turned off. In the oscilloscope we're using as an example, the sweep generator can be triggered by the input signal or by a signal from some other source. (Triggering will be discussed later in this chapter.) This type of oscilloscope is called a triggered oscilloscope. The triggered oscilloscope permits more accurate time measurements to be made and provides a more stable presentation than the nontriggered-type oscilloscope.

On some oscilloscopes, you will find a 10 times (10X) magnification control. As previously mentioned, this allows the displayed sweep to be magnified by a factor of 10.

Q-15. When you select the time base to display a signal, should the time base be the same, higher, or lower than the input signal?

COMPONENTS USED TO PROVIDE A STABLE DISPLAY

The triggering and level controls are used to synchronize the sweep generator with the input signal. This provides a stationary waveform display. If the input signal and horizontal sweep generator are unsynchronized, the pattern tends to jitter, making observations difficult.

The A TRIGGER controls at the lower right of the scope (figure 6-26) are used to control the stability of the oscilloscope CRT display. They are provided to permit you to select the source, polarity, and amplitude of the trigger signal. These controls, labeled A TRIGGER, LEVEL, SOURCE, and SLOPE, are described in the following paragraphs.

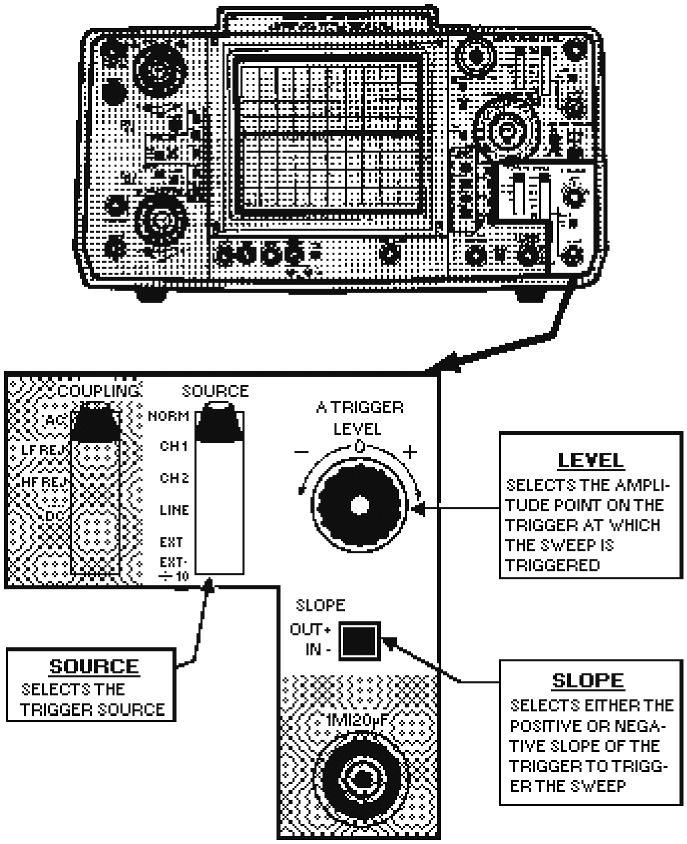


Figure 6-26.—Components that control stability.

SOURCE Control

The SOURCE control allows you to select the appropriate source of triggering. You can select input signals from channel 1 or 2, the line (60 hertz), or an external input.

TRIGGER LEVEL/SLOPE Controls

The LEVEL control allows you to select the amplitude point of the trigger signal at which the sweep is triggered. The SLOPE lets you select the negative or positive slope of the trigger signal at which the sweep is triggered.

The TRIGGER LEVEL (mounted with the TRIGGER SLOPE on some scopes) determines the voltage level required to trigger the sweep. For example, in the TRIGGER modes, the trigger is obtained from the signal to be displayed. The setting of the LEVEL control determines the amplitude point of the input waveform that will be displayed at the start of the sweep.

Figure 6-27 shows some of the displays for a channel that can be obtained for different TRIGGER LEVEL and TRIGGER SLOPE settings. The level is zero and the slope is positive in view A; view B also shows a zero level but a negative slope selection. View C shows the effects of a positive trigger level setting and positive trigger slope setting; view D displays a negative trigger level setting with a positive trigger slope setting. Views E and F have negative slope settings. The difference is that view E has a positive trigger level setting, whereas F has a negative trigger level setting.

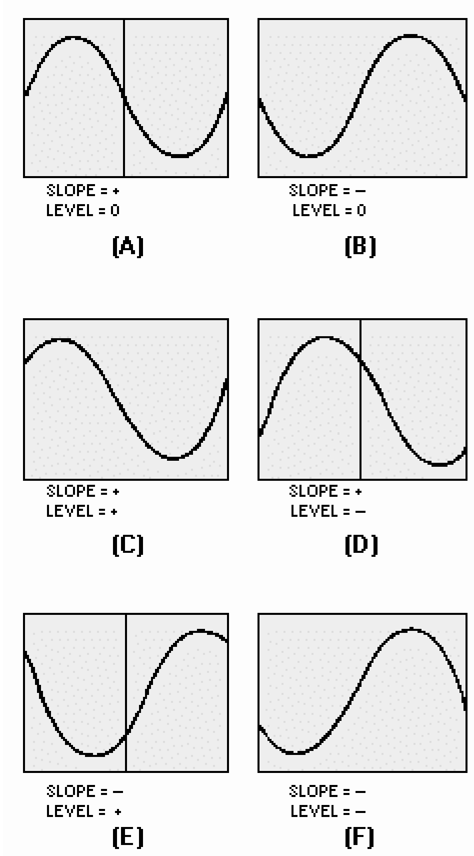


Figure 6-27.—Effects of SLOPE and TRIGGER LEVEL controls.

In most scopes, an automatic function of the trigger circuitry allows a free-running trace without a trigger signal. However, when a trigger signal is applied, the circuit reverts to the triggered mode of operation and the sweep is no longer free running. This action provides a trace when no signal is applied.

Synchronization is also used to cause a free-running condition without a trigger signal. Synchronization is not the same as triggering. TRIGGERING refers to a specific action or event that initiates an operation. Without this event, the operation would not occur. In the case of the triggered sweep, the sweep will not be started until a trigger is applied. Each succeeding sweep must have a trigger before a sweep commences. SYNCHRONIZATION, however, means that an operation or event is brought into step with a second operation.

A sweep circuit that uses synchronization instead of triggering will cause a previously free-running sweep to be locked in step with the synchronizing signal. The TRIGGER LEVEL control setting can be increased until synchronization occurs; but, until that time, an unstable pattern will appear on the CRT face.

COUPLING Section

The COUPLING section allows you to select from four positions: AC, LF REJ, HF REJ, and DC. The AC position incorporates a coupling capacitor to block any dc component. The LF and HF REJ positions reject low- and high-frequency components, respectively. The DC position provides direct coupling to the trigger circuits. This is useful when you wish to view only the LF or HF component of a signal.

COMPONENTS USED TO SELECT SCOPE TRIGGERING

The TRIG MODE section in figure 6-28 allows for automatic triggering or normal triggering. In AUTO (automatic), the triggering will be free-running in the absence of a proper trigger input or will trigger on the input signal at frequencies above 20 hertz. In NORM (normal), the vertical channel input will trigger the sweep.

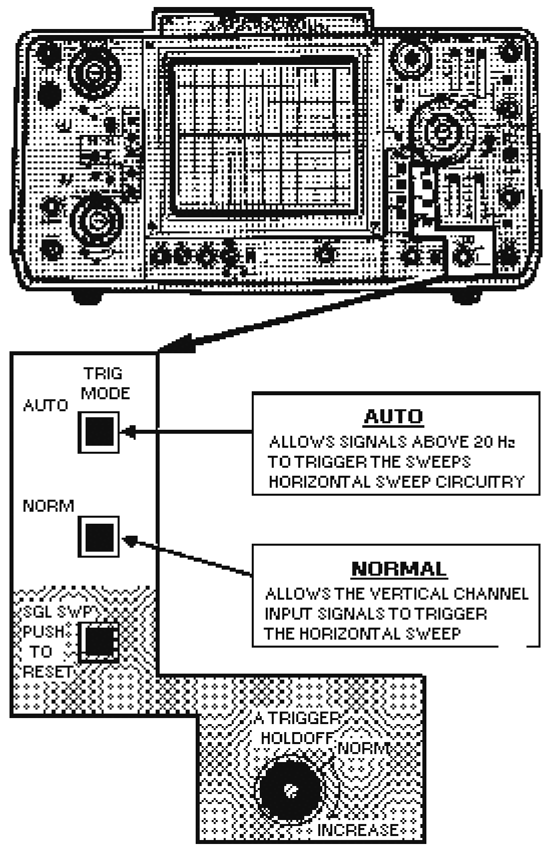


Figure 6-28.—Components to select triggering.

COMPONENTS USED TO SELECT HORIZONTAL-DEFLECTION MODE

For the present, notice only that the HORIZ DISPLAY (horizontal display) in figure 6-29 can be controlled by the TIME/DIV switch. Other switches in this section will be explained later in this chapter.

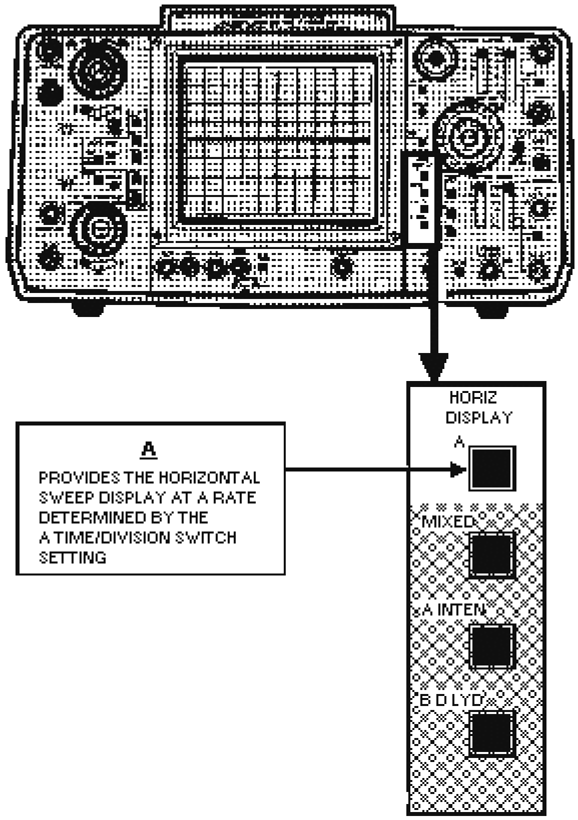


Figure 6-29.—Components to select mode of horizontal deflection.

COMPONENTS USED TO CALIBRATE THE PROBE OF THE SCOPE

In figure 6-30, you can see the components used to calibrate the test probe on the scope. A 1-volt, 2-kilohertz square wave signal is provided for you to adjust the probe for an accurate square wave and to check the vertical gain of the scope. You adjust the probe with a screwdriver, as shown in the figure.

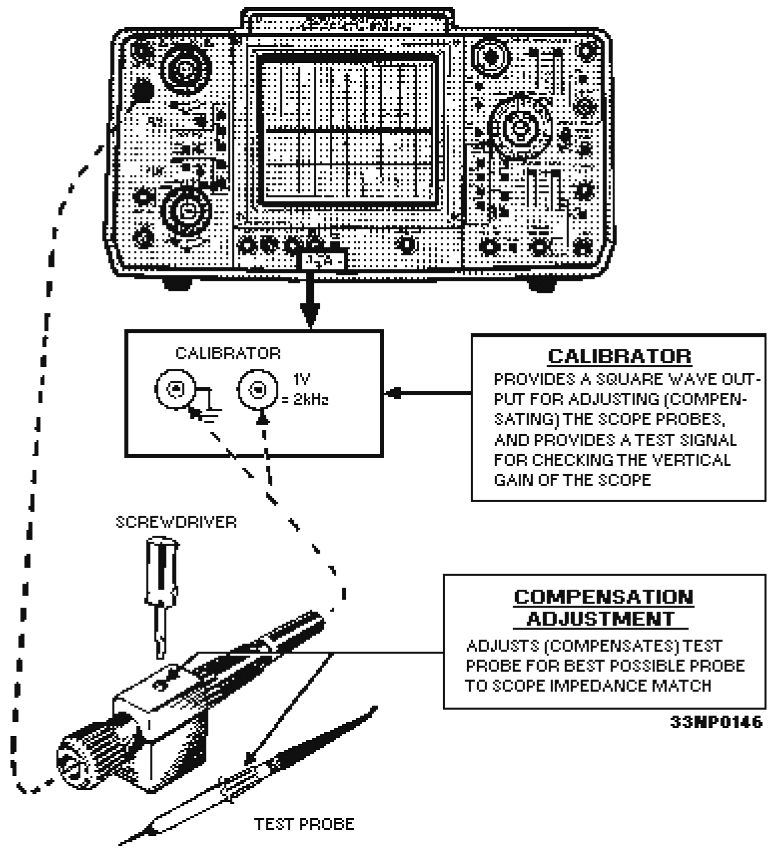


Figure 6-30.—Components to calibrate probe.

SIMILARITIES AMONG OSCILLOSCOPES

The oscilloscope you use may differ in some respects from the one just covered. Controls and circuits may be identified by different names. Many of the circuits will be designed differently. However, all the functions will be fundamentally the same. Before using an oscilloscope, you should carefully study the operator's manual that comes with it.

USING THE OSCILLOSCOPE

An oscilloscope can be used for several different types of measurements, such as time, phase, frequency, and amplitude of observed waveforms. Earlier in this chapter, you learned that the oscilloscope is most often used to study the shapes of waveforms when the performance of equipment is being checked. The patterns on the scope are compared with the signals that should appear at test points (according to the technical manual for the equipment under test). You can then determine if the equipment is operating according to peak performance standards.

Q-16. Oscilloscopes are used to measure what quantities?

TURNING ON THE SCOPE

Before turning on the scope, make sure it is plugged into the proper power source. This may seem obvious, but many technicians have turned all knobs on the front panel out of adjustment before they noticed that the power cord was not plugged in. On some scopes, the POWER switch is part of the INTEN (intensity) control. Turn or pull the knob until you hear a click or a panel light comes on (figure 6-31). Let the scope warm up for a few minutes so that voltages in all of the circuits become stabilized.

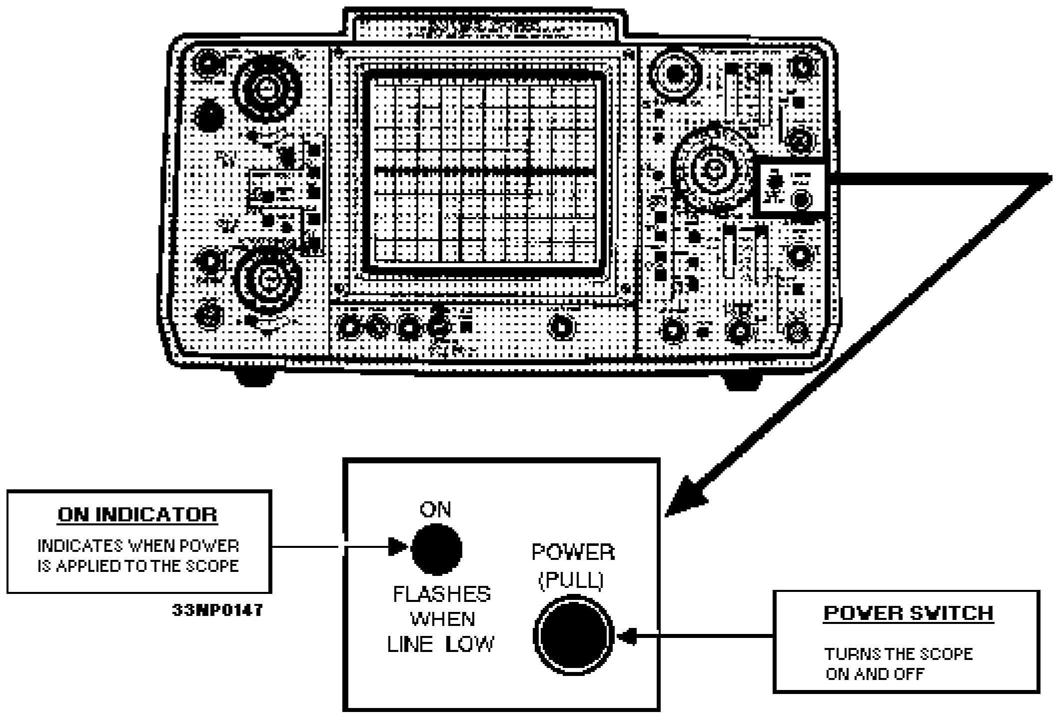


Figure 6-31.—Components to energize scope.

OBTAINING A PATTERN ON THE SCREEN

When adjusting a pattern onto the screen, adjust the INTEN (intensity) and FOCUS controls for a bright, sharp line. If other control settings are such that a dot instead of a line appears, turn down the intensity to prevent burning a hole in the screen coating. Because of the different speeds at which the beam travels across the screen, brightness and sharpness will vary at various frequency settings. For this reason, you may have to adjust the INTEN and FOCUS controls occasionally while taking readings.

NUMBER OF CYCLES ON THE SCREEN

Because distortion may exist at the beginning and end of a sweep, it is better to place two or three cycles of the waveform on the screen instead of just one, as shown in figure 6-32.

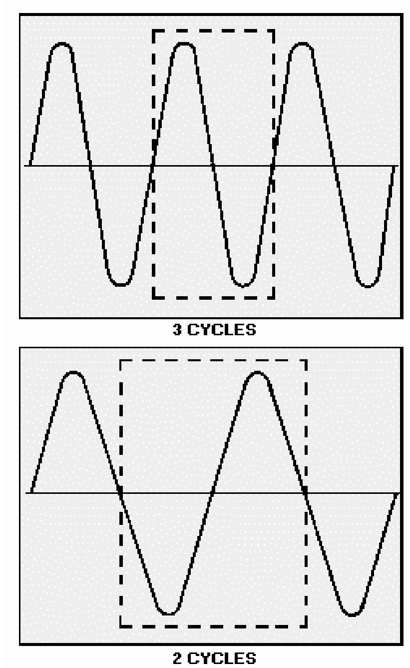


Figure 6-32.—Proper signal presentation.

The center cycle of three cycles provides you with an undistorted waveform in its correct phase. The center of a two-cycle presentation will appear inverted, but will be undistorted. To place waveforms on the CRT in this manner, you must understand the relationship between horizontal and vertical frequencies. The relationship between the frequencies of the waveform on the vertical plates and the sawtooth on the horizontal plates determines the number of cycles on the screen, as shown in figure 6-33.

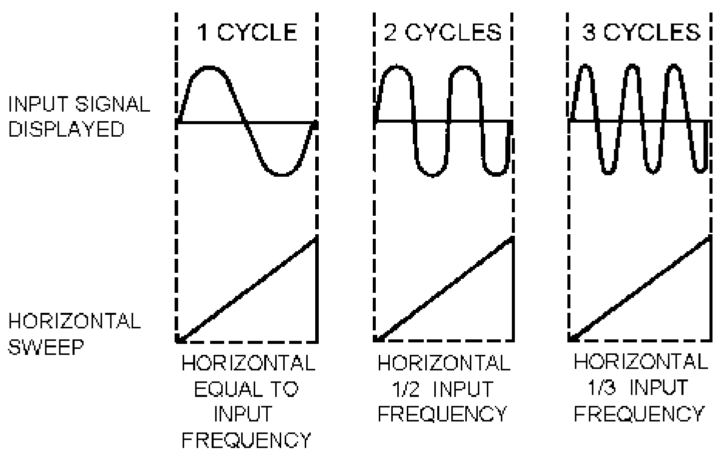


Figure 6-33.—Vertical versus horizontal relationship.

The horizontal sweep frequency of the scope should always be kept lower than, or equal to, the waveform frequency; it should never be higher. If the sweep frequency were higher, only a portion of the waveform would be presented on the screen.

If, for example, three cycles of the waveform were to be displayed on the screen, the sweep frequency would be set to one-third the frequency of the input signal. If the input frequency were 12,000 hertz, the sweep frequency would be set at 4,000 hertz for a three-cycle scope presentation. For two cycles, the sweep frequency would be set at 6,000 hertz. If a single cycle were desired, the setting would be the same as the input frequency, 12,000 hertz.

DUAL-TRACE OPERATION

The information presented in the previous sections served as a general overview of basic single-trace oscilloscope operation using one channel and operating controls. Now, you will be introduced to DUAL-TRACE operation.

Dual-trace operation allows you to view two independent signal sources as a dual display on a single CRT. This operation allows an accurate means of making amplitude, phase, time displacement, or frequency comparisons and measurements between two signals.

A dual-trace oscilloscope should not be confused with a dual-beam oscilloscope. Dual-beam oscilloscopes produce two separate electron beams on a single scope, which can be individually or jointly controlled. Dual-trace refers to a single beam in a CRT that is shared by two channels.

Q-17. Scopes that produce two channels on a single CRT with a single beam are referred to as what types of scopes?

Components Used to Select Vertical-Deflection Operating Mode

The VERT MODE controls (figure 6-34) allow you to select the operating mode of the scope for vertical deflection.

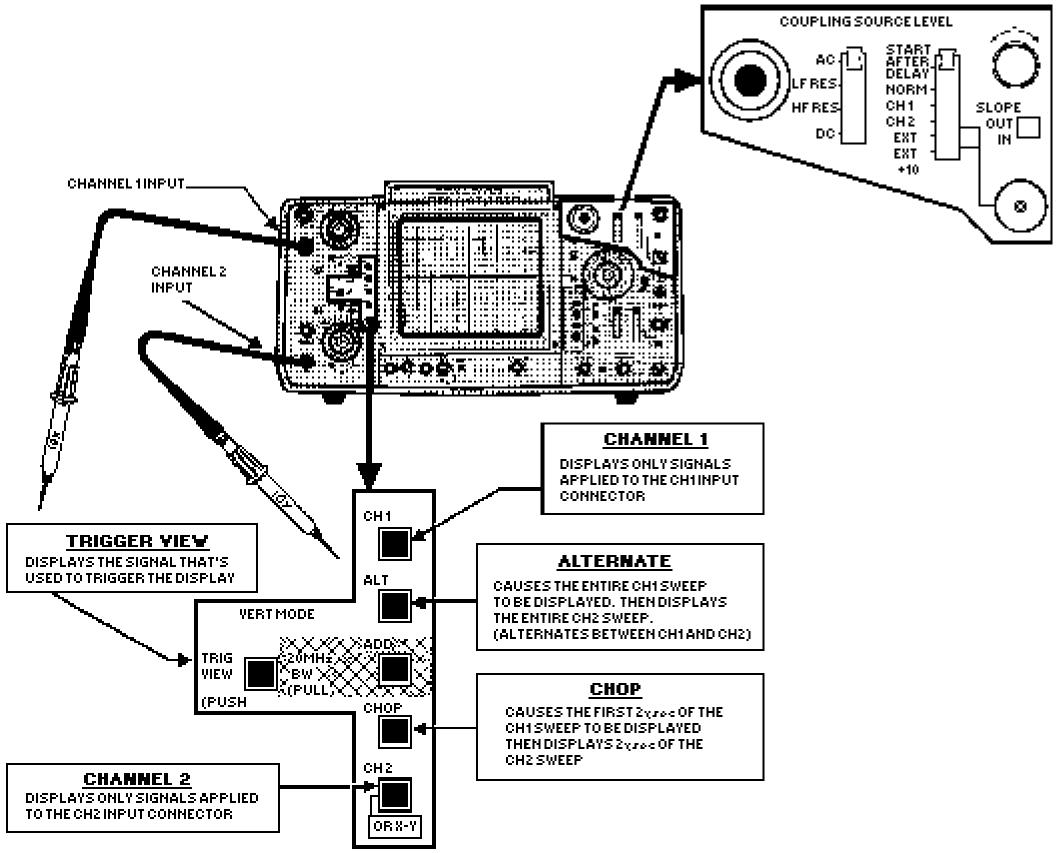


Figure 6-34.—Components to select vertical operating mode.

CH 1 AND CH 2.—These controls allow you to display signals applied to either channel 1 or channel 2, as discussed earlier.

TRIGGER VIEW.—The TRIG VIEW allows you to display the signal that is actually used to trigger the display. (Triggering was discussed earlier.)

ALT.—The ALT (alternate) mode (figure 6-35) of obtaining a dual trace uses the techniques of GATING between sweeps. This control allows the signal applied to channel 1 to be displayed in its entirety; then, channel 2 is displayed in its entirety. This method of display is continued alternately between the two channels. At slow speeds, one trace begins to fade while the other channel is being gated. Consequently, the ALT mode is not used for slow sweep speeds. The CHOP mode, shown in figure 6-36 (explained next), will not produce a satisfactory dual sweep at high speeds. The ALT mode is deficient at low speeds. Therefore, both are used on dual-trace oscilloscopes to complement each other and give the scope a more dynamic range of operation.

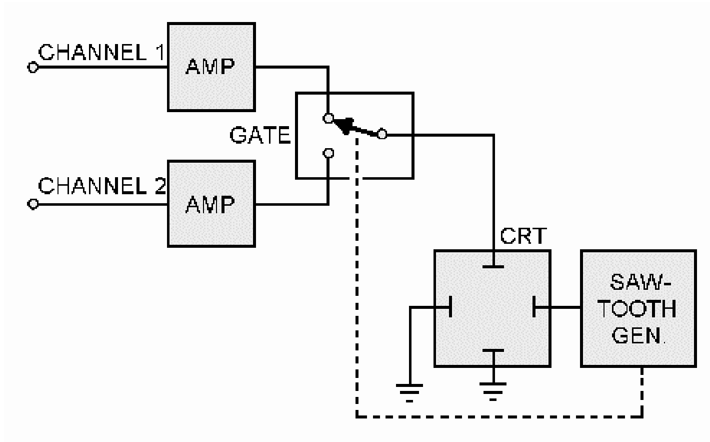


Figure 6-35.—ALT (alternate) mode.

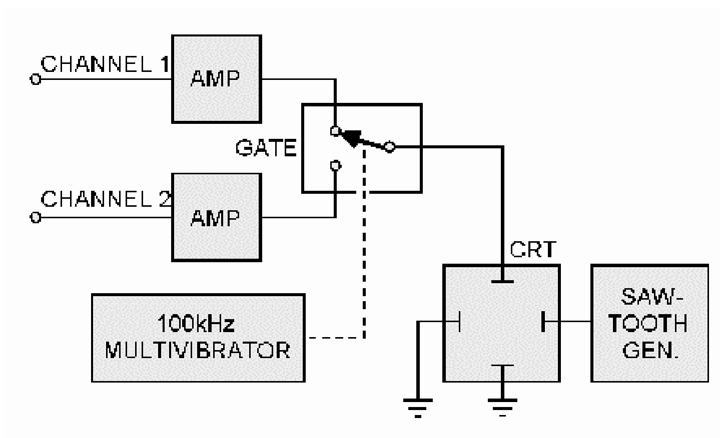


Figure 6-36.—CHOP mode.

The output dc voltage references on each of the amplifiers are independently adjustable. Therefore, the beam will be deflected by different amounts on each channel if the voltage reference is different at each amplifier output. The output voltage from each amplifier is applied to the deflection plates through the gate. The gate is actually an electronic switch. In this application, it is commonly referred to as a BEAM SWITCH.

Switching is controlled by a high-frequency multivibrator in the CHOP mode. That is, the gate selects the output of one channel and then the other at a high-frequency rate (1200 kilohertz in most oscilloscopes). Because the switching time is very short in a good-quality oscilloscope, the resultant display is two sets of horizontally dashed lines, as shown in figure 6-37, view A.

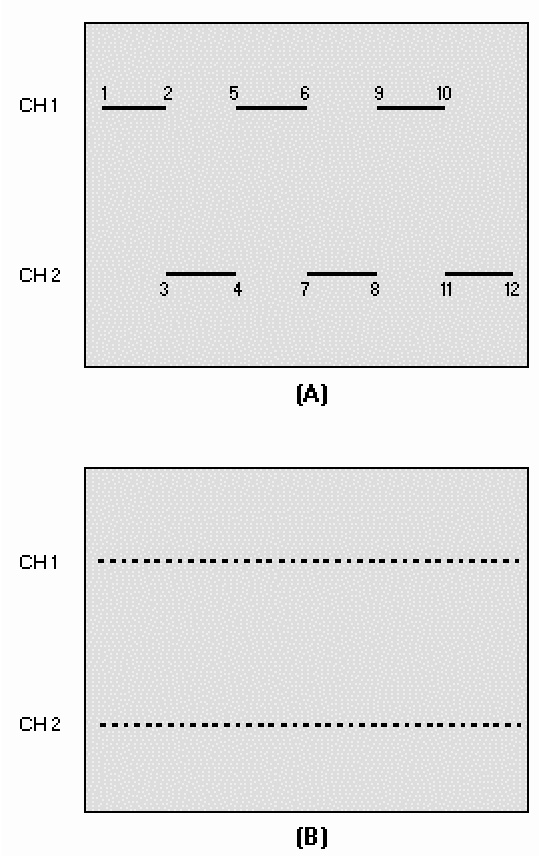


Figure 6-37.—Displaying CHOP mode.

Dashed line CH 1 is the output of one channel, while line CH 2 is the output of the other. The trace moves from left to right because of the sawtooth waveform applied to the horizontal plates. A more detailed analysis shows that the beam moves from CH 1 to CH 2 while the gate is connected to the output from one channel. Then, when the gate samples the output of the CH 2 during time 3 to 4, the beam is at a different vertical LOCATION. (This is assuming that CH 2 is at a different voltage reference.) The beam continues in the sequence 5 to 6, 7 to 8, 9 to 10, and 11 to 12 through the rest of one horizontal sweep.

When the chopping frequency is much higher than the horizontal sweep frequency, the number of dashes will be very large. For example, if the chopping occurs at 100 kilohertz and the sweep frequency is 1 kilohertz, each horizontal line would then appear as a series of closely spaced dots, as shown in figure 6-37 view B. As the sweep frequency becomes lower compared to the chopping frequency, the display will show apparently continuous traces; therefore, the CHOP mode is used at low sweep rates.

When signals are applied to the channel amplifiers (view A of figure 6-38), the outputs are changed according to the triggering signal (view B). The resultant pattern (view C) on the screen provides a time-base presentation of the signals of each channel.

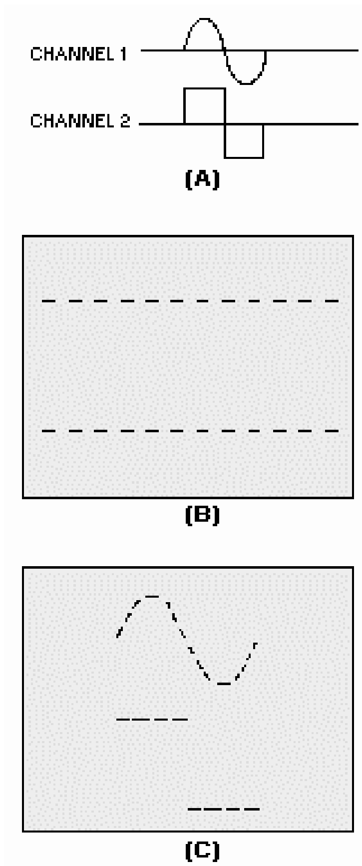


Figure 6-38.—Dual-channel display in CHOP mode.

ADD.—The ADD switch (shown earlier in figure 6-34) algebraically adds the two signals of channels 1 and 2 together for display.

Other Dual-Trace Oscilloscope Controls

Most dual-trace oscilloscopes have both an A and B time base for horizontal sweep control. Notice in the upper right corner on our example scope (figure 6-34) the COUPLING, SOURCE LEVEL, and SLOPE controls. These serve the same function as did those same controls in the A time-base section of the scope. The B time base is selected using the same A and B TIME/DIV control (pull out outer knob).

The use of the B time base is controlled by the HORIZ DISPLAY section discussed earlier in the A time-base section. However, inexperienced technicians generally do not use A and B time bases together in the MIXED, A INTEN (intensified), and B D'LYD (delayed) settings. These controls are fully explained in the applicable technical manuals; therefore, we will not discuss the controls in this chapter. Figure 6-39 is a block diagram of a basic dual-trace oscilloscope without the power supplies.

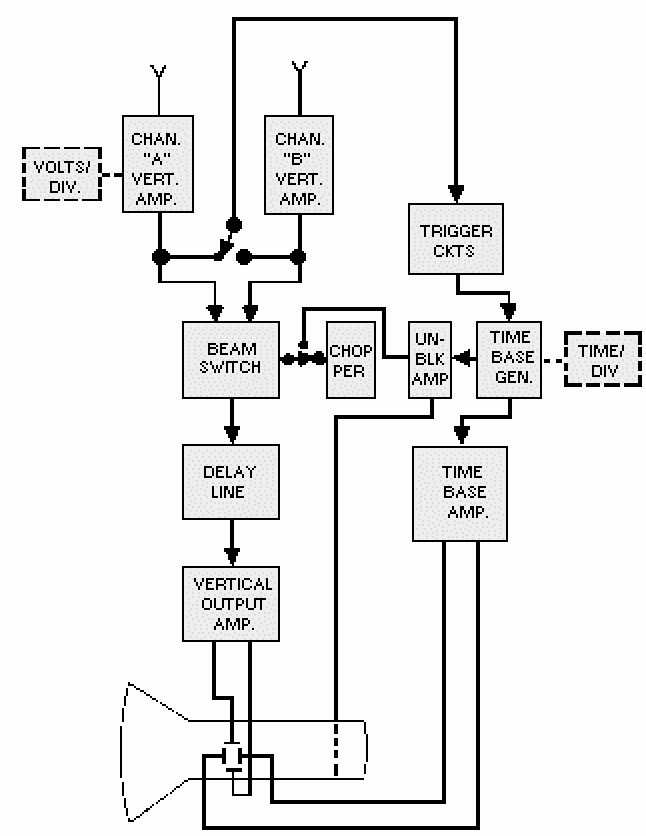


Figure 6-39.—Basic dual-trace oscilloscope block diagram.

ACCESSORIES

The basic dual-trace oscilloscope has one gun assembly and two vertical channels. However, there are many variations. The horizontal sweep channels vary somewhat from equipment to equipment. Some have one time-base circuit and others have two. These two are interdependent in some oscilloscopes and in others they are independently controlled. Also, most modern general-purpose oscilloscopes are constructed of modules. That is, most of the vertical circuitry is contained in a removable plug-in unit, and most of the horizontal circuitry is contained in another plug-in unit.

The main frame of the oscilloscope is often adapted for many other special applications by the design of a variety of plug-in assemblies. This modular feature provides much greater versatility than in a single-trace oscilloscope. For instance, to analyze the characteristics of a transistor, you can replace the dual-trace, plug-in module with a semiconductor curve-tracer plug-in module.

Other plug-in modules available with some oscilloscopes are high-gain, wide-bandwidth amplifiers; differential amplifiers; spectrum analyzers; physiological monitors; and other specialized units. Therefore, the dual-trace capability is a function of the type of plug-in unit that is used with some oscilloscopes.

To get maximum usefulness from an oscilloscope, you must have a means of connecting the desired signal to the oscilloscope input. Aside from cable connections between any equipment output and the oscilloscope input, a variety of probes are available to assist in monitoring signals at almost any point in a circuit. The more common types include 1-TO-1 PROBES, ATTENUATION PROBES, and CURRENT PROBES. Each of these probes may be supplied with several different tips to allow measurement of signals on any type of test point. Figure 6-40 shows some of the more common probe tips.

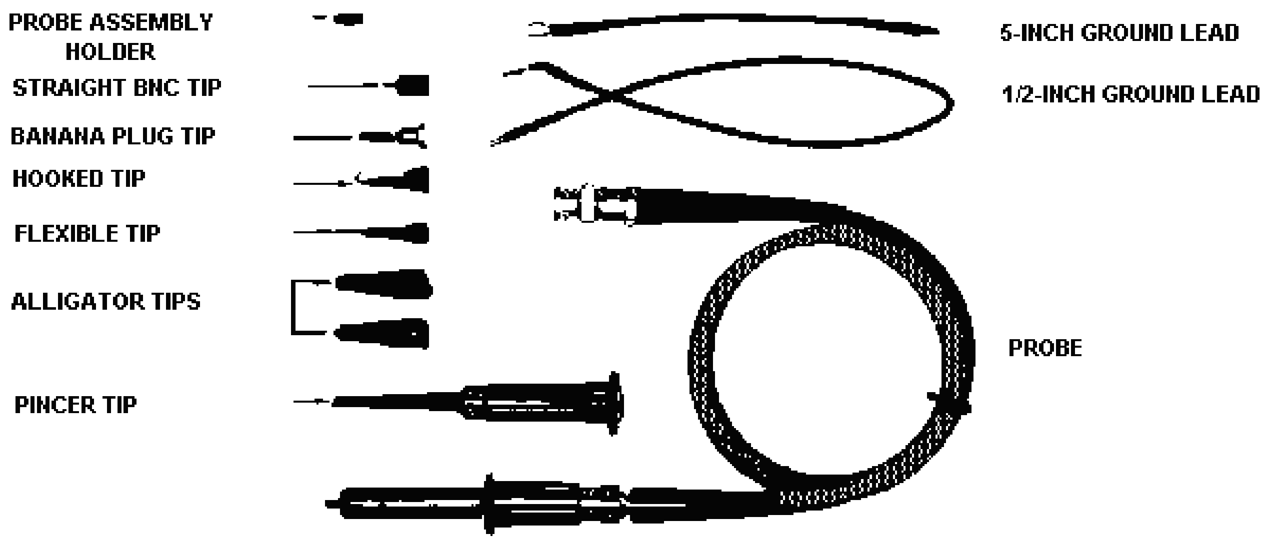


Figure 6-40.—Common probe tips.

In choosing the probe to use for a particular measurement, you must consider such factors as circuit loading, signal amplitude, and scope sensitivity.

The 1-to-1 probe offers little or no attenuation of the signal under test and is, therefore, useful for measuring low-level signals. However, circuit loading with the 1-to-1 probe may be a problem. The impedance at the probe tip is the same as the input impedance of the oscilloscope.

An attenuator probe has an internal high-value resistor in series with the probe tip. This gives the probe a higher input impedance than that of the oscilloscope. Because of the higher input impedance, the probe can measure high-amplitude signals that would overdrive the vertical amplifier if connected directly to the oscilloscope. Figure 6-41 shows a schematic representation of a basic attenuation probe. The 9-megohm resistor in the probe and the 1-megohm input resistor of the oscilloscope form a 10-to-1 voltage divider.

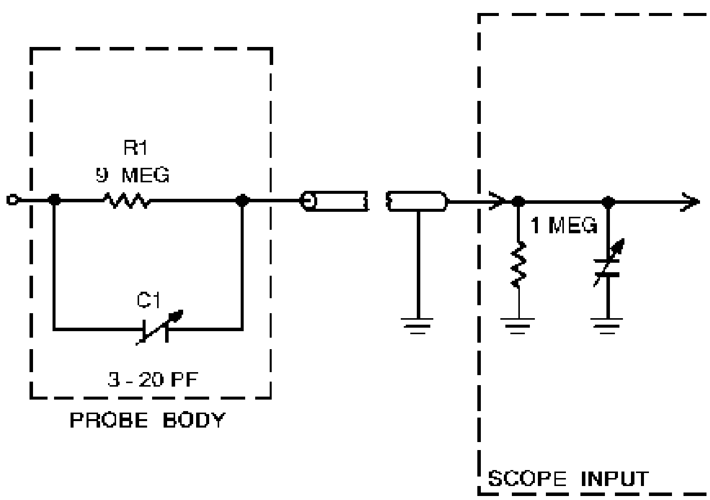


Figure 6-41.—Basic attenuation probe.

Since the probe resistor is in series, the oscilloscope input resistance is 10 megohms when the probe is used. Thus, using the attenuator probe with the oscilloscope causes less circuit loading than using a 1-to-1 probe.

Before using an attenuator probe for measurement of high-frequency signals or for fast-rising waveforms, you must adjust the probe compensating capacitor (C1) according to instructions in the applicable technical manual. Some probes will have an IMPEDANCE EQUALIZER in the end of the cable that attaches to the oscilloscope. The impedance equalizer, when adjusted according to manufacturer's instructions, assures proper impedance matching between the probe and oscilloscope. An improperly adjusted impedance equalizer will result in erroneous measurements, especially when you are measuring high frequencies or fast-rising signals.

More information on oscilloscope hook-ups can be found in Electronics Information Maintenance Books (EIMB), *Test Methods and Practices*.

Special current probes have been designed to use the electromagnetic fields produced by a current as it travels through a conductor. This type of probe is clamped around a conductor without disconnecting it from the circuit. The current probe is electrically insulated from the conductor, but the magnetic fields about the conductor induce a potential in the current probe that is proportional to the current through the conductor. Thus, the vertical deflection of the oscilloscope display will be directly proportional to the current through the conductor.

SPECTRUM ANALYZER

The spectrum analyzer is used to examine the frequency spectrum of radar transmissions, local oscillators, test sets, and any other equipment operating within its testable frequency range. With experience, you will be able to determine definite areas of malfunctioning components within equipment. Successful spectrum analysis depends on the proper operation of a spectrum analyzer and your ability to correctly interpret the displayed frequencies. Although there are many types of spectrum analyzers, we will use the Tektronix, Model 492 for our discussion.

The spectrum analyzer accepts an electrical input signal and displays the frequency and amplitude of the signal on a CRT. On the vertical, or Y, axis, the amplitude is plotted. The frequency would then be found on the horizontal, or X, axis. The overall pattern of this display (figure 6-42) indicates the proportion of power present at the various frequencies within the spectrum (fundamental frequency with sideband frequencies).

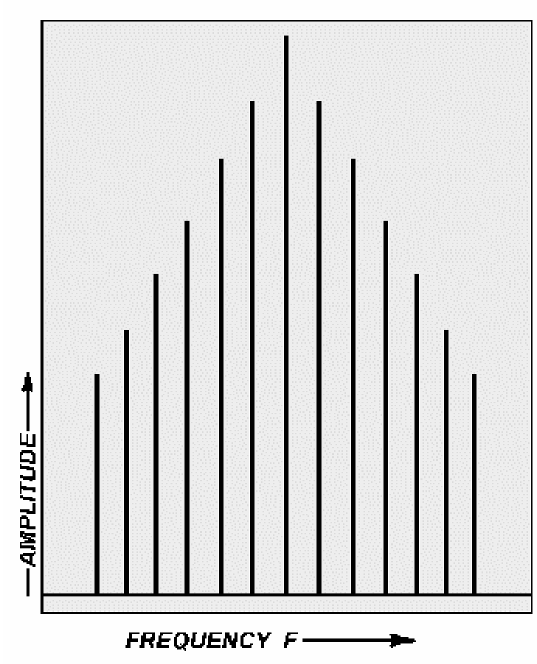


Figure 6-42.—Spectrum analyzer pattern.

BASIC FUNCTIONAL DESCRIPTION

The model 492 analyzer can be divided into six basic sections, as follows:

- Converter section;
- Intermediate frequency (IF) section;
- Display section;
- Frequency control section;
- Digital control section; and
- Power and cooling section.

Converter Section

The converter section actually consists of three frequency converters, made up of a mixer, local oscillator (LO), and required filters. Only one frequency can be converted at a time and pass through the filters to reach the next converter. The analysis frequency can, however, be changed by altering the frequency of the LO and adjusting the FREQUENCY control knob.

FIRST CONVERTER.—The first (front end) converter changes the input signal to a usable IF signal that will either be 829 MHz or 2072 MHz. The IF signal to be produced is dependent on which measurement band selection is currently being used. The 829 MHz IF signal will be selected for bands 2 through 4, while the 2072 MHz IF signal is selected for bands 1 and 5 through 11.

Q-18. The first converter is also known by what other name?

SECOND CONVERTER.—The second converter actually contains two converters. Only one of these two converters in this section is ever operational, and selected as a result of the measurement band currently being used. The selected converter will convert the frequency received from the first converter to a usable (110 MHz) IF signal, which is then provided to the third converter.

THIRD CONVERTER.—This converter takes the 110 MHz IF signal, amplifies it, and then converts it to the final IF of 10 MHz. This signal, in turn, is then passed on to the IF section.

IF Section

The IF section receives the final IF signal and uses it to establish the system resolution by using selective filtering. System resolution is selected under microcomputer control among five bandwidths (1 MHz, 100 kHz, 10 kHz, 1 kHz, and 100 Hz). The gain for all bands are then leveled and logarithmically amplified. This is done so that each division of signal change on the CRT display remains equal in change to every other division on the CRT. For example, in the 10-dB-per-division mode, each division of change is equal to a 10 dB difference, regardless of whether the signal appears at the top or bottom of the CRT. The signal needed to produce the video output to the display section is then detected and provided.

Display Section

The display section provides a representative display of the input signal on the CRT. It accomplishes this by performing the following functions:

- Receives the video signals from the IF section and processes these signals to adjust the vertical drive of the CRT;
- Receives the sweep voltages and processes these signals to produce the horizontal CRT drive plate voltage;
- Receives character data information and generates CRT plate drive signals to display alpha and numeric characters on the CRT;
- Receives control levels from the front panel beam controls and generates unblanking signals to control display presence, brightness, and focus.

The vertical deflection of the beam is increased as the output of the amplitude detector increases. The horizontal position is controlled by the frequency control section and is the frequency analyzed at that instant. The beam sweeps from left to right, low to high frequencies during its analysis. During this analysis, any time a signal is discovered, a vertical deflection will show the strength of the signal at the horizontal position that is the frequency. This results in a display of amplitude as a function of frequency.

Frequency Control Section

The frequency control section accomplishes the tuning of the first and second LOs within the converter section. The frequency immediately being analyzed is controlled by the current frequencies of the LOs. To analyze another frequency, you must change an LO frequency to allow the new frequency to be converted to a 10 MHz signal by the converter section. Periodically, the unit sweeps and analyzes a frequency range centered on the frequency set by the FREQUENCY knob. Adjusting the FREQUENCY knob will cause the LOs to be tuned to the new frequency. Only the LOs of the first two converters can be changed to vary the frequency being analyzed.

Digital Control Section

All the internal functions are controlled from the front panel through the use of a built-in microcomputer. The microcomputer uses an internal bus to receive or produce all communication or control to any section of the analyzer.

Power and Cooling Section

The main power supply provides almost all the regulated voltages required to operate the unit. The display section provides the high voltage necessary for CRT operation.

The cooling system allows fresh cool air to be routed to all sections of the unit in proportion to the heat that is generated by each section.

SPECTRUM ANALYZER FRONT PANEL CONTROLS, INDICATORS, AND CONNECTORS

This section will describe the function of the front panel controls, indicators, and connectors. For a complete description of each function, refer to table 6-1 while reviewing the front panel in figure 6-43. The numbers located in column 1 of table 6-1 equate to the same numbers found on the front panel of figure 6-43. Because most operational functions of this spectrum analyzer are microprocessor-controlled, they are switch-selected rather than adjusted.

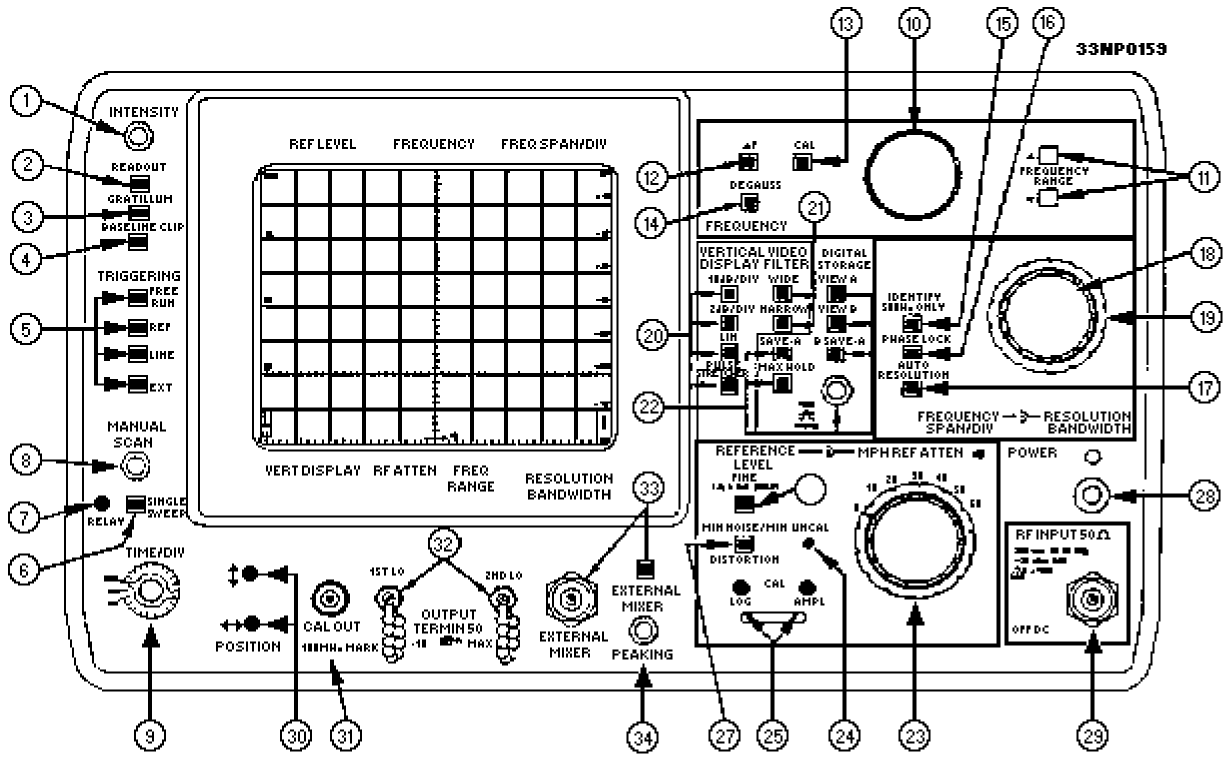


Figure 6-43.—Spectrum analyzer front panel controls, indicators, and connectors.

Table 6-1.—Description of Front Panel Controls, Indicators, and Connectors

ITEM	FUNCTION	DESCRIPTION
1	INTENSITY	This knob controls the brightness of the CRT trace and the CRT readout display. The focus is electronically adjusted.
2	READOUT	This push button switches the readout display on and off. All spectrum analyzer parameters are displayed except TIME/DIV. The brightness for this display is proportional to the trace brightness and can be readjusted on internal controls only by a qualified technician.
3	GRATILLUM	This push button switches the graticule light on and off.
4	BASELINE CLIP	This push button, when activated, clips (subdues) the intensity at the baseline.
5	TRIGGERING	This area allows one of four triggering modes to be selected by push buttons that illuminate when active. When any of these four are selected, the others are canceled.
5a	FREE RUN	When activated, the sweep is free-running without regard to trigger signals.
5b	INT	When activated, the sweep is triggered by any signal at the left edge of the display with an amplitude of 1.0 divisions of the graticule or more.
5c	LINE	When activated, a sample of the ac power line voltage is used to trigger the sweep.
5d	EXT	When selected, the sweep is triggered by an external signal (applied through the back panel IN HORZ/TRIG connector) between a minimum and maximum of 0.5 and 50 volt peak.
6	SINGLE SWEEP	This push button, plus a ready indicator (No. 7), provides the single sweep operation. When this operation is selected, one sweep is initiated after the sweep circuit has been triggered. Pushing this button does not cancel the other trigger modes. When single sweep is first selected, the present sweep is aborted, but the sweep circuit is not yet armed. An additional push is required to initially arm the sweep. The button must be pushed again to rearm the sweep circuit each time the sweep has run. To cancel single sweep, you must select one of the four trigger mode selections.
7	READY	When single sweep is selected, this indicator lights while the sweep circuit is armed and ready for a trigger signal. The indicator stays lit until the sweep is complete.
8	MANUAL SCAN	When the TIME/DIV (No. 9c) selector is in the MNL position, this control will manually vary the CRT beam across the full horizontal axis of the display.
9	TIME/DIV	Is used to select sweep rates from 5 μsec/div to 20 μsec/div. This switch also selects AUTO, EXT, and MNL modes.
9a	AUTO	In this position, the sweep rate is selected by the microcomputer to maintain a calibrated display for any FREQ SPAN/DIV, RESOLUTION, and VIDEO FILTER combination.
9b	EXT	When selected, this control allows an external input source to be used with the sweep rates.
9c	MNL	When selected, this control is used in conjunction with No. 8 (see MANUAL SCAN, No. 8).
10	FREQUENCY	This control is manually turned to allow you to tune to the center frequency.
11	FREQUENCY RANGE (band)	These two push buttons are used to shift the center frequency up or down. Frequency range on the band is displayed on the CRT readout.
12	F	This control is used for measuring the frequency difference between signals. When selected, the frequency readout goes to zero. It will then read out the deviation from this reference to the next frequency desired as the FREQUENCY knob is adjusted.

Table 6-1.—Description of Front Panel Controls, Indicators, and Connectors—Continued

ITEM	FUNCTION	DESCRIPTION
13	CAL	When this is activated, the frequency readout can be calibrated to center the center frequency by adjusting the FREQUENCY control for the correct reading. When accomplished, you should deactivate the CAL mode.
14	DEGAUSS	When this button is pressed, current through the local oscillator system is reduced to zero in order to minimize magnetism build-up around the LOs. This is done to enhance the center frequency display and amplitude accuracy. You should do this after every significant frequency change and before calibrating the center frequency.
15	IDENTIFY 500 kHz ONLY	The signal identify feature can become functional only when the FREQ SPAN/DIV is set to 500 kHz. When activated (button lit), true signals will change in amplitude on every sweep. Images and spurious response signals will shift horizontally or go completely off the CRT display. To ensure that the signal is changing amplitude every sweep, you should decrease the sweep rate so that each sweep can be analyzed.
16	PHASE LOCK	When this control is activated (button lit), it will reduce residual FM when narrow spans are selected. In narrow spans, the phase lock can be turned off or back on by pressing the button. Switching the PHASE LOCK off may cause the signal to shift position. In narrow spans, the signal could shift off the display; however, it will usually return to its phase locked position after a few moments. The microcomputer automatically selects PHASE LOCK for a span/division of 50 kHz or below in bands 1 through 3, 100 kHz or below for band 3, and 200 kHz for bands 5 and above.
17	AUTO RESOLUTION	This push button, when activated, will automatically select the bandwidth for FREQ SPAN/DIV, TIME/DIV, and VIDEO FILTER. The internal microcomputer selects the bandwidth to maintain a calibrated display. This can be checked by changing the FREQ SPAN/DIV and observing the bandwidth change on the display.
18	FREQ SPAN/DIV	This is a continuous detent control that selects the frequency span/div. The span/div currently selected is displayed on the CRT. The range of the span/div selection is dependent on the frequency band selected: <u>BAND</u> / <u>NARROW SPAN</u> / <u>WIDE SPAN</u> 1-3 (0-7.1GHz) / 10kHz/Div / 200MHz/Div 4-5 (5.4-21GHz) / 50 kHz/Div / 500 MHz/Div 6 (18-26GHz) / 50 kHz/Div / 1 GHz/Div 7-8 (26-60GHz) / 100 kHz/Div / 2 GHz/Div 9 (60-90GHz) / 200 kHz/Div / 2 GHz/Div 10 (90-140GHz) / 500 kHz/Div / 5 GHz/Div 11 (140-220GHz) / 500 kHz/Div / 10 GHz/Div Two additional bands are provided: full band (max span) and 0 Hz span. When max span is selected, the span displayed is the full band. When zero span is selected, time/div is read out instead of span/div.
19	RESOLUTION BANDWIDTH	This is also a continuous detent control that selects the resolution bandwidth. The bandwidth is shown on the CRT display. The range of adjustment is from 1 kHz to 1 MHz in decade steps. When you change the resolution bandwidth with this control, it will deactivate the AUTO RESOLUTION.
20	VERTICAL DISPLAY	These four push buttons select the display mode. The scale factor can be seen on the CRT display.
20a	10dB/DIV	When this is activated, the dynamic range of the display is calibrated to 80 dB, with each major graticule representing 10 dB.

Table 6-1.—Description of Front Panel Controls, Indicators, and Connectors—Continued

ITEM	FUNCTION	DESCRIPTION
20b	2dB/DIV	When activated, this will increase the resolution so that each major graticule division represents 2 dB.
20c	LIN	When activated, this selects a linear display between zero volts (bottom graticule line) and the reference level (top graticule line) scaled in volts/division (see REFERENCE LEVEL, No. 23a).
20d	PULSE STRETCHER	When selected, this increases the fall time of the pulse signals so that very narrow pulses in a line spectrum display can be observed.
21	VIDEO FILTER	One of two (NARROW OR WIDE) filters can be activated to reduce video bandwidth and high-frequency components for display noise averaging. The narrow filter is approximately 1/300th of the selected resolution bandwidth with the wide filter being 1/30th the bandwidth. Activating either one will cancel the other. To disable, completely switch filter off.
22	DIGITAL STORAGE	Five push buttons and ON control operate the digital storage functions. With none of the push buttons activated, the display will not be stored.
22a	VIEW A, VIEW B	When either or both of these push buttons are selected, the push button illuminates, and the contents of memory A and/or memory B are displayed. With Save A mode off, data in a memory is interlaced with data from B memory.
22b	B-SAVE A	When activated, the differential (arithmetic difference) of data in B memory and the saved data in memory A are displayed. SAVE A mode is activated and SAVE A button will be lit.
22c	MAX HOLD	When activated, the digital storage memory retains the maximum signal amplitude at each memory location. This permits visual monitoring of signal frequency and amplitude at each memory location over an indefinite period of time. This feature is used to measure drift, stability, and record peak amplitude.
22d	PEAK/AVERAGE	This control selects the amplitude at which the vertical display is either peak detected or averaged. Video signals above the level set by the control (shown by a horizontal line or cursor) are peak detected and stored while video signals below the cursor are digitally averaged and stored.
23	MIN RF ATTEN	This control is used to set the minimum amount of RF attenuation. Changing RF LEVEL will not decrease RF attenuation below that set by the MIN RF ATTEN selector.
23a	REFERENCE LEVEL	This is a continuous control that requests the microcomputer to change the reference level one step for each detent. In the 10 dB/DIV vertical-display mode, the steps are 1 dB or 0.25 dB if the FINE mode (No. 26) is selected.
23b	MIN RF ATTEN DB	This selects the lowest value of attenuation allowed: Actual RF attenuation is set by the microcomputer according to the logarithm selected by the MIN NOISE/MIN DISTORTION (No. 27) button. If RF attenuation is increased by changing MIN RF ATTEN, the microcomputer automatically changes IF gain to maintain the current reference level.
24	UNCAL	This indicator lights when the display amplitude is no longer calibrated (selecting a sweep rate that is not compatible with the frequency span/div and resolution bandwidth).
25	LOG and AMPL CAL	These adjustments calibrate the dynamic range of the display. The LOG calibrates any logarithm gain dB/Div, and the AMPL calibrates the reference level of the top graticule line at the top of the display.
26	FINE	When activated, the REFERENCE LEVEL (No. 23a) switches in 1 dB increments for 10 dB/Div display mode, 0.25 dB for 2 dB/Div, and volts 1 dB for LIN display mode.

Table 6-1.—Description of Front Panel Controls, Indicators, and Connectors—Continued

ITEM	FUNCTION	DESCRIPTION
27	MIN NOISE/MIN DISTORTION	This selects one of two logarithms used to control attenuator and IF gain. MIN NOISE (button illuminated) reduces the noise level by reducing attenuation and IF gain 10 dB. MIN DISTORTION (button not illuminated) reduces distortion to its minimum. To observe any changes, the RF attenuation displayed on the CRT readout must be 10 dB higher than that set by the MIN RF ATTEN selector.
28	POWER	This is a pull switch that turns power on when extended.
29	RF INPUT	This is a 50 ohm coaxial input jack used to input signals of 21GHz or below. The maximum nondestructive input signal level that can be applied to this input is +13 dBm or 30 mW. Signals above 10 dB may cause signal compression.
30	POSITION	These controls are used to position the display on the horizontal and vertical axes.
31	CAL OUT	This is an output jack that has a calibrated 20 dBm 100 MHz signal, with frequency markers spaced 100 MHz apart. The calibrated 100 MHz marker is used as a reference for calibrating the reference level and log scale. The combination of 100 MHz markers is used to check span and frequency readout accuracy.
32	OUTPUT 1ST AND 2ND LO	These jacks provide access to the output of the respective LOs. The jacks must have 50 ohm terminators installed when not connected to an external device.
33	EXTERNAL MIXER	When the EXTERNAL MIXER button is activated, bias is provided out the EXTERNAL MIXER port for external waveguide mixers. The IF output from the EXTERNAL MIXER is then applied through the EXTERNAL MIXER port to the second converter for use.
34	PEAKING	This control varies the mixer bias for external mixers in the EXTERNAL MIXER mode. This control should be adjusted for maximum signal amplitude.

NORMAL INDICATIONS UPON POWER ON

With power applied (power knob pulled out), the spectrum analyzer will automatically (upon microcomputer control) go into the following conditions. If you do not find these indications, there is a probably a problem with the unit.

- Vertical display: 10 dB/div;
- Frequency: 0.00 MHz;
- REF level: +30 dB;
- RF attenuation: 60 dB;
- Frequency range: 0.0 to 1.8 GHz;
- Auto resolution: 1 MHz;
- Resolution bandwidth: 1 MHz;
- Freq Span/Div: Max;
- Triggering: Free run;
- Readout: On;
- Digital storage: View A/View B On;

- All other indicators off or inactive.

SUMMARY

Now that we have completed this chapter, we will briefly review the more important points covered.

A **CATHODE-RAY TUBE (CRT)** is used in an oscilloscope to display the waveforms.

The CRT used in oscilloscopes consists of an **ELECTRON GUN**, a **DEFLECTION SYSTEM**, and a **FLUORESCENT SCREEN**.

The **ELECTRON BEAM** in an oscilloscope is allowed to be controlled in any direction by means of **HORIZONTAL-** and **VERTICAL-DEFLECTION PLATES**.

VERTICAL-DEFLECTION PLATES are used to show **AMPLITUDE** of a signal.

HORIZONTAL-DEFLECTION PLATES are used to show **TIME** and/or **FREQUENCY** relationship.

A **GRATICULE** is a calibrated scale of **AMPLITUDE VERSUS TIME** that is placed on the face of the CRT.

A **DUAL-TRACE OSCILLOSCOPE** is designed to accept two vertical inputs at the same time. It uses a single beam of electrons shared by two channels.

The **SPECTRUM ANALYZER** accepts an electrical input signal and displays the signal's frequency and amplitude on a CRT display.

ANSWERS TO QUESTIONS Q1. THROUGH Q18.

A-1. Control grid.

A-2. The first anode.

A-3. Because they bend electron streams in much the same manner that optical lenses bend light rays.

A-4. It accelerates the electrons emerging from the first anode.

A-5. A greater deflection angle.

A-6. A greater deflection angle.

A-7. Higher potential.

A-8. Slower beam.

A-9. Amplitude and time.

A-10. Amplitude.

A-11. Time and/or frequency relationships.

A-12. To permit wide-angle deflection of the beam.

A-13. Deflection factor.

A-14. A CRT, a group of control circuits, power supply, sweep circuitry, and deflection circuitry.

A-15. Lower.

A-16. Amplitude, phase, time, and frequency.

A-17. Dual-trace oscilloscopes.

A-18. Front end.

APPENDIX I

GLOSSARY

ABSORPTION WAVEMETER—An instrument used to measure audio frequencies.

AMMETER—A meter used to measure current.

BACK RESISTANCE—The larger resistance value observed when you are checking the forward resistance of a semiconductor.

BAND PASS FILTER—A tuned circuit that passes only a specific frequency.

BAND REJECT FILTER—A tuned circuit that does not pass a specified band of frequencies.

BARRETTER—A type of bolometer characterized by an increase in resistance as the dissipated power rises.

BEAT FREQUENCY—The difference between the oscillator frequency and the unknown audio frequency.

BEL—The unit that expresses the logarithmic ratio between the input and output of any given component, circuit, or system.

BOLOMETER—A loading device that undergoes changes in resistance as changes in dissipated power occur.

CAVITY WAVEMETER—An instrument used to measure microwave frequencies.

CONTINUITY—An uninterrupted, complete path for current flow.

CORRECTIVE MAINTENANCE—Used to isolate equipment failures. Includes replacement of defective parts to return equipment to proper performance.

DAMPING—The process of smoothing oscillations of the meter pointer.

D'ARSONVAL METER MOVEMENT—The permanent-magnet moving coil movement used in most meters.

dBm—An abbreviation used to represent power levels above or below a 1-milliwatt reference.

DUMMY ANTENNA—See DUMMY LOAD.

DUMMY LOAD—A resistor used to replace the normal load, which is specifically designed to have low reactance and possess the ability to dissipate required amounts of power.

ELECTRODYNOMETER METER MOVEMENT—A meter movement using fixed field coils and a moving coil; usually used in wattmeters.

EXTERNALLY EXCITED METER—A term used to describe meters that get their power from the circuit to which they are connected.

FORWARD RESISTANCE—The smaller resistance value observed when you are checking the forward resistance of a semiconductor.

FREQUENCY METER—An instrument used to measure the rate at which ac voltages are generated.

GALVANOMETER—A meter used to measure small values of current by electromagnetic or electrodynamic means.

GENERAL PURPOSE ELECTRONIC TEST EQUIPMENT (GPETE)—Test equipment that has the capability, without modification, to generate, modify, or measure a range of electronic functions required to test two or more equipments or systems of basically different designs.

INDUCTANCE BRIDGE—An ac bridge circuit used to measure an unknown value of inductance.

MAINTENANCE—Work done to correct, reduce, or counteract wear, failure, and damage to equipment.

MEASURE (Metrology Automated System for Uniform Recall and Reporting)—A Navy standardized system designed to provide the recall, scheduling, and documenting of test equipment into calibration facilities.

MECHANICAL-ROTATION FREQUENCY—The speed in revolutions per minute of armatures in electric motors and engine-driven generators; blade speed in turbines.

MEGGER—A meter used to measure insulation resistance.

METER MOVEMENT—The part of the meter that moves to indicate some electrical value.

METER SHUNT—A resistor placed in parallel with the meter terminals; used to provide increased range capability.

OHMMETER—A meter used to measure resistance.

POWER FACTOR—An indication of the various losses of a capacitor, such as current leakage and dielectric absorption.

PREVENTIVE MAINTENANCE—Consists of mechanical, electrical, and electronic checks; used to determine whether or not equipment is operating properly.

SCAT CODES—A four-digit subcategory code used to identify the functional measurement parameters that can be satisfied by any one of many pieces of test equipment.

SELF-EXCITED METER—A term used to describe meters that operate from their own power sources.

SENSITIVITY—(1) For an ammeter, the amount of current that will cause full-scale deflection of the meter. (2) For a voltmeter, the ratio of the voltmeter resistance divided by the full-scale reading of the meter; expressed in ohms per volt.

SCLSIS (Ship Configuration and Logistics Support Information System)—This encompasses the automated data processing system and all practices and procedures used for identification and status accounting of ship's configuration, logistics, technical data reviews, and equipment configuration audits.

SHORT CIRCUIT—An unintentional current path between two components in a circuit or between a component and ground; usually caused by a circuit malfunction.

SPECIAL PURPOSE ELECTRONIC TEST EQUIPMENT (SPETE)—Test equipment that is specifically designed to generate, modify, or measure a range of electronic functions of a specific or peculiar nature on a single system or equipment.

STROBOSCOPE—An instrument that allows viewing of rotating or reciprocating objects by producing the optical effect of a slowing down or stopping motion.

SYMPTOM ELABORATION—Using built-in indicating instruments or other aids to define an equipment malfunction.

SYMPTOM RECOGNITION—Recognition of a situation in equipment operation that is not normal.

TACHOMETER—An instrument that measures the rate at which a shaft is turning.

TEST EQUIPMENT INDEX—The Navy guide used to assist in identifying portable electrical/electronic test equipment required for support of prime electrical/electronic, IC, weapons, and reactor instrumentation systems.

TEST POINTS—Locations in equipment that are accessible to the technician's probes where operating voltages or signals can be monitored.

THERMISTOR—A type of bolometer characterized by a decrease in resistance as the dissipated power increases.

TROUBLESHOOTING—A procedure used to evaluate equipment performance and repair equipment when it fails to operate properly.

TUNED CIRCUIT—A circuit that is used as a filter which passes or rejects specific frequencies.

VOLTMETER—A meter that is used to measure voltage.

WATT—The unit of electrical power that is the product of voltage and current.

WATTMETER—An electrodynamometer type meter used to measure electrical power.

WAVEFORM ANALYSIS—Observation of displays of voltage and current variations with respect to time or by harmonic analysis of complex signals.

WAVEMETER—Calibrated resonant circuits that are used to measure frequency.

WHEATSTONE BRIDGE—An ac bridge circuit used to measure unknown values of resistance, inductance, or capacitance.

APPENDIX II

REFERENCES USED TO DEVELOP THIS TRAMAN

NOTE: Although the following references were current when this TRAMAN was published, their continued currency cannot be assured. Therefore, you need to ensure that you are studying the latest revision.

CHAPTER 1

EIMB, *Test Equipment*, NAVSEA 0967-LP-000-0040, Naval Sea Systems Command, Washington, DC, 1981.

EIMB, *Test Methods and Practices*, NAVSEA 0967-LP-000-0130, Naval Sea Systems Command, Washington, DC, 1980.

Electronic Test Equipment Calibration Program Indoctrination Handbook, NAVMAT P-9491, Naval Material Command, Washington, DC, 1979.

Metrology Automated System for Uniform Recall and Reporting (MEASURE) Users Manual, OP 43P6A, Chief of Naval Operations, Washington, DC, 1984.

Navy Electricity and Electronics Training Series, Module 3, *Introduction to Circuit Protection, Control, and Measurement,* NAVEDTRA B72-03-00-93, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1993.

Portable Electrical/Electronic Test Equipment (PEETE) Index NAVSEA ST-AA-IDX-010-PEETE, Naval Sea Systems Command, Washington, DC, 1986.

Ships Configuration and Logistics Support Information System (*SCLSIS*), S0752-AA-SPN-010/SCLSIS, Naval Sea Systems Command, Washington, DC, 1988.

Standard General Purpose Electronic Test Equipment, MIL-STD 1364(Series) (NAVY), Naval Sea Systems Command, Washington, DC, 1991.

Stowage Guide for Portable Electrical/Electronic Test Equipment (PEETE) NAVSEA ST000-AB-GYD-010/PEETE, Naval Sea Systems Command, Washington, DC, 1987.

CHAPTER 2

Electronics Technician 3 and 2, NAVEDTRA 80197-1, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1987.

Fire Controlman Third Class, NAVEDTRA 10276-1, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1988.

Navy Electricity and Electronics Training Series, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*, NAVEDTRA 172-09-00-83, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1983.*

Navy Electricity and Electronics Training Series, Module 7, *Introduction to Solid-State Devices and Power Supplies*, NAVEDTRA B72-07-00-92, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1992.

Navy Electricity and Electronics Training Series, Module 21, *Test Methods and Practices*, NAVEDTRA B72-21-00-87, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1987.

Navy Electricity and Electronics Training Series, Module 2, *Introduction to Alternating Current and Transformers*, NAVEDTRA 172-02-00-91, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1991.

Radioman 3 & 2, NAVEDTRA 10228-H1, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1991.

CHAPTER 3

Navy Electricity and Electronics Training Series, Module 5, *Introduction to Generators and Motors*, NAVEDTRA B72-05-00-94, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1994.

Navy Electricity and Electronics Training Series, Module 1, *Introduction to Matter, Energy, and Direct Current*, NAVEDTRA B72-01-00-92, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1992.

Navy Electricity and Electronics Training Series, Module 3, *Introduction to Circuit Protection, Control, and Measurement*, NAVEDTRA B72-03-00-93, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1993.

CHAPTER 4

8000A Digital Multimeter, NAVSEA 0969-LP-279-9010, Naval Sea Systems Command, Washington, DC.

Navy Electricity and Electronics Training Series, Module 7, *Introduction to Solid-State Devices and Power Supplies*, NAVEDTRA B72-07-00-92, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1992.

Navy Electricity and Electronics Training Series, Module 13, *Introduction to Number Systems and Logic Circuits*, NAVEDTRA B72-13-00-86, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1986.*

Operation and Maintenance Instructions, Volt-Ohm-Milliammeter 260 Series, NAVSEA 0969-LP-286-1010, Naval Sea Systems Command, Washington, DC, 1977.

CHAPTER 5

EIMB, *Test Methods and Practices*, NAVSEA 0967-LP-000-0130, Naval Sea Systems Command, Washington, DC, 1980.

Huntron Tracker 2000 Operation and Maintenance Manual, Mill Creek, WA, 1985.

Navy Electricity and Electronics Training Series, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*, NAVEDTRA 172-10-00-83, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1983.*

Navy Electricity and Electronics Training Series, Module 12, *Modulation Principles*, NAVEDTRA 172-12-00-83, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1983.*

CHAPTER 6

EIMB, *Test Methods and Practices*, NAVSEA 0967-LP-000-0130, Naval Sea Systems Command, Washington, DC, 1980.

EW Technician AN/SLQ-32A(V)3 ECM Maintenance Trainee Guide, A-102-0216 (Vol 1), Chief of Naval Technical Training, Pensacola, FL, 1992.

Navy Electricity and Electronics Training Series, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*, NAVEDTRA 172-09-00-83, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1983.*

Navy Electricity and Electronics Training Series, Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies*, NAVEDTRA B72-06-00-92, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1992.

Tektronix 492/492P Spectrum Analyzer Instruction Manual, Beaverton, OR, 1981.

*Effective 1 September 1986, the Naval Education and Training Program Development Center became the Naval Education and Training Program Management Support Activity. Effective 1 October 1996, the name was changed to Naval Education and Training Professional Development and Technology Center.

MODULE 16 INDEX

D

E

F

G

H

I

J

M

O

U

V

W

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Test Equipment Administration and USE," pages 1-1 through 1-33. Chapter 2, "Miscellaneous Measurements," pages 2-1 through 2-27.

1-1. What system is currently used by U.S. military services to identify electronic equipment with standardized nomenclature?

1. Joint Electronic Type Designation System (JETDS)
2. Joint Electronics Type Category System (JETCS)
3. Army-Navy (AN) System
4. Navy Tactical Data Systems (NTDS)

1-2. Which of the following categories of test equipment is/are designed to test, without modification, a range of electronic parameters needed for two or more systems that are different in design?

1. GPETE
2. SPETE
3. Both 1 and 2 above
4. Installed

1-3. For what purpose was the ETE classification board established?

1. To control inventory limits
2. To control the increase of nonstandard GPETE
3. To control the increase of nonstandard SPETE
4. To become final approval authority for SPETE

1-4. Which of the following systems provides an inventory of test equipment actually located in the fleet?

1. 3M
2. SCLSIS
3. STEED
4. SPETREL

1-5. You can determine if a piece of test equipment has been calibrated by checking which of the following documents?

1. The test equipment logbook
2. A directive from the EMO
3. The completed maintenance action form for the instrument
4. A tag or label attached to the instrument

1-6. What label is used to identify a test instrument that is within tolerance on all parameters?

1. INACTIVE
2. CALIBRATED
3. CALIBRATED—Refer to report
4. ORGANIZATION LEVEL CALIBRATED

1-7. Which of the following conditions warrants the use of one of the SPECIAL CALIBRATION labels and/or tags?

1. Calibration deviates from the usual tolerances
2. The instrument has more than one calibration interval
3. The instrument is too large to move and requires in-place calibration
4. Each of the above

1-8. Under what circumstances should the USER CALIBRATION label be affixed to a piece of test equipment?

1. When a certain calibration schedule is assigned
2. When user calibration is required before, during, or after use
3. When the equipment is out for calibration
4. Each of the above

1-9. Which of the following statements must appear on the CALIBRATION NOT REQUIRED label affixed to an instrument?

1. By what authority the label was affixed
2. The reason no calibration is required
3. The date the label was affixed
4. The METRL page number

1-10. Which of the following labels is attached to an unusable instrument, and may have an additional tag attached?

1. INACTIVE
2. REJECTED
3. CALIBRATED
4. SPECIAL CALIBRATION

1-11. A test instrument that has plug-in modules and/or easily accessible potentiometer or controls which affect the calibration of the instrument should have which of the following labels attached?

1. CALIBRATED
2. CALIBRATION NOT REQUIRED
3. CALIBRATED—REFER TO REPORT
4. CALIBRATION VOID IF SEAL BROKEN

1-12. A replacement part needed by an IMA to repair a piece of test equipment should be ordered by the activity sending the inoperative equipment for repair and calibration. For which of the following reasons is this a good practice?

1. IMAs are not allowed to order repair parts
2. The ship has more repair funds than the IMA
3. The ship is usually able to obtain the parts more quickly
4. It is more likely the technician on the ship will obtain the correct part

1-13. When maintenance personnel are not authorized to make repairs to a piece of test equipment, what items, if any, must be sent to the calibration repair facility with the equipment?

1. The unit and its power cord
2. All the accessories
3. Standards used to calibrate
4. None

1-14. In what type of environment should test equipment be stowed?

1. Dry
2. Dark and damp
3. High humidity and low temperature
4. High temperature and high humidity

1-15. In a stowage space aboard ship, what device(s) should be used to hold the test equipment in place?

1. Set clasp springs
2. Tie down cord
3. Steel straps
4. Stretch seat-belt-type straps

1-16. What system is used to provide for a standardized recall and scheduling of test equipment into calibration facilities?

1. MDCC
2. SCLSIS
3. METER
4. MEASURE

1-17. The meter card is used to provide what information concerning test equipment?

1. Changes
2. Additions
3. Deletions
4. All of the above

1-18. Which of the following actions would be classified as preventive maintenance?

1. Purchasing a new piece of test equipment
2. Isolating an equipment failure to the component level
3. Aligning a servo assembly after a repair
4. Replacing a defective transistor

1-19. Which of the following actions would be regarded as part of corrective maintenance?

1. Routine lubrication of a radar pedestal
2. Mechanical inspection of a bearing assembly in a motor housing
3. Alignment of a servo assembly after a repair
4. Cleaning a filter in accordance with a maintenance requirement card

1-20. Troubleshooting electrical and electronic equipment includes which of the following actions?

1. Fault isolation
2. Equipment repair
3. Equipment performance evaluation
4. Each of the above

1-21. The initial operating conditions of newly installed equipment are referred to as

1. alignment data
2. manufacturer's specifications
3. baseline operating characteristics
4. expected operation characteristics

1-22. When working on energized equipment, you should follow which of the following practices?

1. Work alone
2. Work with both hands
3. Insulate yourself from ground
4. Wear rubber gloves at all times

1-23. When measuring 300 volts or more, you should first take what step?

1. Turn off equipment power
2. Ground all components capable of retaining an electrical charge
3. Short-circuit all components capable of retaining an electrical charge
4. Connect the meter leads to the points to be measured

1-24. Which of the following insulating materials is suitable for covering a grounded metal work bench?

1. Dry insulating material that contains no holes or conductors
2. Dry canvas that has holes in it
3. Dry phenolic material that has conductors embedded in it
4. Damp plywood

1-25. Prior to working on a circuit, you use a shorting probe discharge which of the following types of components?

1. Capacitors only
2. Cathode-ray tubes only
3. Capacitors and cathode-ray tubes
4. Inductors

1-26. If a 28 volt 6 ampere fuse blows, the proper procedure is to replace it with which of the following devices?

1. A larger fuse until the cause of the overload has been determined
2. A fuse of the same voltage and current rating
3. A fuse rated 20 percent lower than the blown fuse
4. A copper strap until the cause of the overload is determined

1-27. Before electrical equipment is overhauled or repaired, what general safety precaution, if any, should be followed?

1. The fuse for the associated circuits should be replaced with circuit breakers
2. The main supply switches should temporarily be shorted out
3. The power switches should be secured open and tagged out of service
4. None

1-28. After work on equipment is complete, who should remove any attached tags?

1. The job inspector
2. The repair crew leader
3. Any member of the repair crew
4. The person who signed and attached the tag

1-29. What is the purpose of the grounding cable attached to the frame of a generator aboard ship?

1. Create a potential difference between the frame and the ship
2. Conduct power to the generator under emergency conditions
3. Ensure equipment is at same ground potential as the ship
4. Break the circuit between the frame and the power supply under emergency conditions

1-30. Which of the following steps should you take to help ensure that metal-case test equipments are safe to use?

1. Energize the instrument to test the ground
2. Ensure the equipment is grounded
3. Insulate the metal case from ground
4. Connect all metal cases to a common ungrounded lead

1-31. Which of the following precautions should you observe when using measuring instruments?

1. Avoid strong magnetic fields
2. Avoid excessive current
3. Avoid mechanical shock
4. Each of the above

1-32. The meter movement in a voltmeter can be easily damaged by excessive current if you do not follow certain procedures. When setting up the meter to read voltage, the RANGE SWITCH should first be set to (a) what relative range and then changed to (b) what relative range?

1. (a) Highest
   (b) Closest to the voltage to be read
2. (a) Highest
   (b) Lower than the voltage to be read
3. (a) Lowest
   (b) Lower than the voltage to be read
4. (a) Lowest
   (b) Closest to the voltage to be read

A.	Place one hand in your pocket or behind your back.
B.	Turn on the power.
C.	Connect the meter ground to the equipment ground.
D.	Place the positive meter lead on the test point; select for positive or negative polarity.

Figure 1A.—Procedures

IN ANSWERING QUESTION 1-33, REFER TO THE PROCEDURES IN FIGURE 1A.

1-33. When you measure voltages less than 300 volts, in what order should you complete the task?

1. A B C D
2. B C D A
3. C A D B
4. D C B A

IN ANSWERING QUESTIONS 1-34 THROUGH 1-37, SELECT FROM THE MEASUREMENT COLUMN BELOW THE ANSWER THAT MATCHES THE SITUATION BEING DESCRIBED.

MEASUREMENT

1. Current
2. Inductance
3. Resistance
4. Capacitance

1-34. This measurement is rarely taken in preventive or corrective maintenance or testing because unsoldering is usually required. Ohm's law is normally applied to determine this value.

1-35. This is a valuable aid in locating faults during corrective maintenance, but cannot be made with power applied. Many technical manuals contain charts that indicate the test points for this measurement.

1-36. This measurement provides an indication of dielectric strength and is used to determine the power factor.

1-37. This measurement is seldom taken during troubleshooting. It can be taken using a bridge or another instrument that is primarily designed to measure another quantity; however, a conversion chart is required.

1-38. The power factor is an indication of the losses caused by which of the following conditions?

1. Excessive voltage
2. Dielectric absorption
3. Current leakage
4. Both 2 and 3 above

1-39. The Wheatstone bridge can be used for precision measurements of which of the following quantities?

1. Voltage
2. Current
3. Impedance
4. Resistance

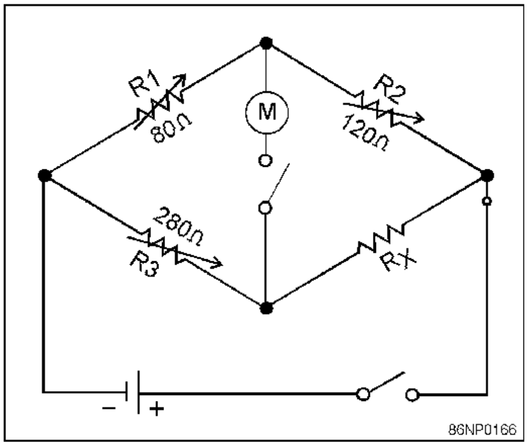


Figure 1B.—Dc resistance bridge.

IN ANSWERING QUESTION 1-40, REFER TO FIGURE 1B.

1-40. In the dc resistance bridge, what is the value of Rx?

1. 42 ohms
2. 400 ohms
3. 420 ohms
4. 4,200 ohms

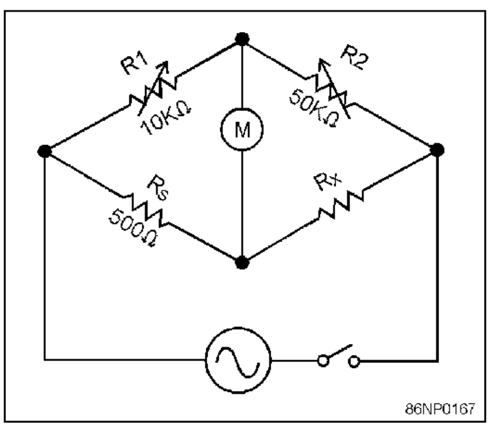


Figure 1C.—Ac resistance bridge.

IN ANSWERING QUESTION 1-41, REFER TO FIGURE 1C.

1-41. In the ac resistance bridge, what is the value of R_x?

1. 25 ohms
2. 250 ohms
3. 2,500 ohms
4. 25,000 ohms

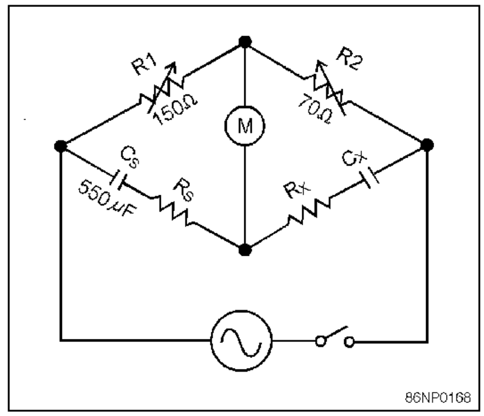


Figure 1D.—Capacitance bridge.

IN ANSWERING QUESTION 1-42, REFER TO FIGURE 1D.

1-42. In the capacitance bridge, what is the value of C_x?

1. 25 microfarads
2. 117 microfarads
3. 256 microfarads
4. 1,178 microfarads

1-43. Dc power is stated in which of the following units?

1. Watts
2. Farads
3. Amperes
4. Henries

1-44. Power in an audio-frequency circuit is stated in which of the following units?

1. Decibels (dB) only
2. Decibels referenced to 1 milliwatt (dBm) only
3. Both dB and dBm
4. Volt units (Vu)

1-45. The bel is a unit of measurement used with voltage, current, or power that compares which of the following circuit values?

1. The input to the output
2. The output to a reference
3. The voltage to power
4. The current to power

1-46. What is the relationship between the values of the bel and the decibel?

1. The bel is twice the decibel
2. The decibel is twice the bel
3. The bel is 1/10 the decibel
4. The decibel is 1/10 the bel

1-47. What is the corresponding increase in dBm each time power is doubled?

1. +1 dB
2. +2 dB
3. +3 dB
4. +10 dB

1-48. A thermocouple ammeter is used to measure which of the following quantities?

1. Rf current
2. Af current
3. Motor current
4. Generator current

1-49. A bolometer is a power meter that measures power in certain frequency ranges. Which of the following methods is/are used by the bolometer to measure power values?

1. A barretter detects increases in power when its resistance increases
2. A thermistor detects increases in power when its resistance decreases
3. Both 1 and 2 above
4. Power is measured directly

THIS SPACE LEFT BLANK INTENTIONALLY.

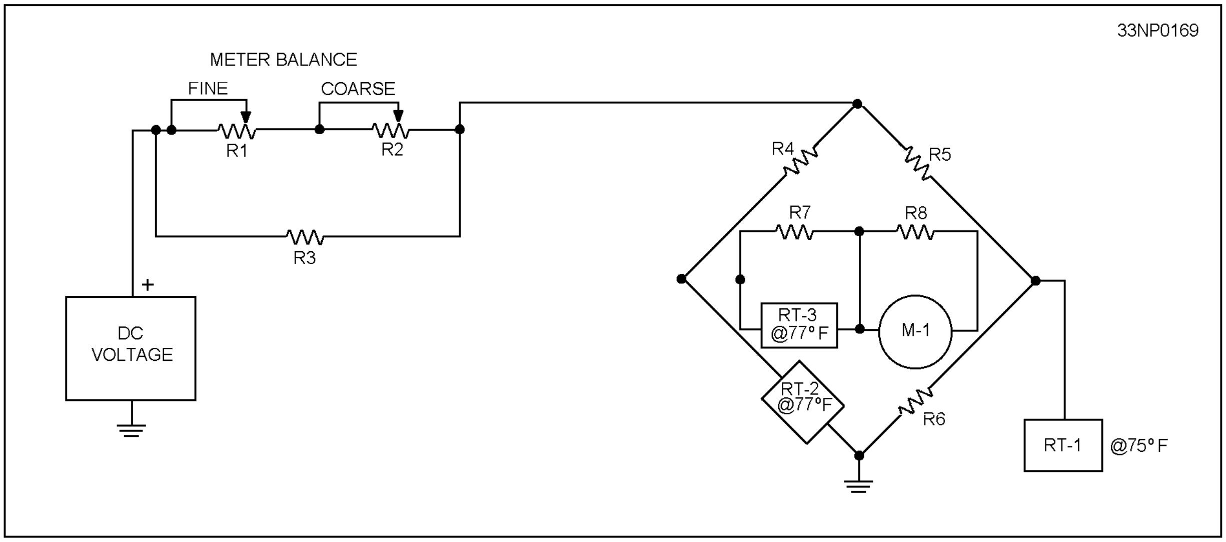


Figure 1E.—Thermistor bridge.

IN ANSWERING QUESTION 1-50, REFER TO FIGURE 1E.

1-50. In the thermistor bridge, what is the purpose of RT-1 and RT-3?

1. To compensate for power used by RT-2
2. To compensate for temperature changes outside the waveguide
3. To control the amount of rf energy applied to RT-2
4. To control the temperature of the waveguide

1-51. To measure shaft rotation rate on an engine, you should use, which, if any, of the following instruments?

1. An ammeter
2. A bolometer
3. A tachometer
4. None of the above

1-52. In the centrifugal tachometer, what component restricts the action on the lower collar that is produced by centrifugal force?

1. The spring
2. The pointer
3. The upper collar
4. The lower collar

1-53. What is the usual speed range (in feet per minute) of a chronometric tachometer?

1. 0 to 30
2. 0 to 300
3. 0 to 3,000
4. 0 to 30,000

1-54. You are measuring the speed of a fan blade by using a stroboscopic tachometer. Setting the flash at a rate 5 rpm SLOWER than the fan speed will cause the blades to appear to move (a) in what relative direction and (b) at what speed?

1. (a) Forward (b) 5 rpm
2. (a) Forward (b) 10 rpm
3. (a) Backward (b) 10 rpm
4. (a) Backward (b) 5 rpm

1-55. The flashing rate of a strobotac is controlled by which of the following circuits?

1. An electronic pulse generator
2. A frequency divider
3. A power supply
4. An amplifier

1-56. The flashing rate of the strobotron tube affects its life expectancy. What is the range (in hours) of life expectancy of the strobotron tube?

1. 15 to 24
2. 25 to 49
3. 50 to 99
4. 100 to 250

1-57. The vibrating-reed frequency meter is a delicate instrument and should not be subjected to vibrations, such as those associated with motor-generators or their associated control panels.

1. True
2. False

1-58. When using the vibrating-reed frequency meter, you take the reading in which of the following ways?

1. Read the digital readout
2. Read the dial indication
3. Read the mechanical setting
4. Read the reed that vibrates the most

1-59. Bandpass filters and band reject filters are tuned circuits that either pass or reject specific frequencies. In these filters, (a) what type offers very high impedance to currents at its resonant frequency, and (b) what type offers a very low impedance to currents at its resonant frequency?

1. (a) Parallel-tuned (b) series-tuned
2. (a) Parallel-tuned (b) parallel-tuned
3. (a) Series-tuned (b) parallel-tuned
4. (a) Series-tuned (b) series-tuned

1-60. When you are zero beating an unknown frequency with a frequency provided by a calibrated, high-precision oscillator within a heterodyne frequency meter, what will be the indication when the two frequencies are matched?

1. One dot of light on the screen will be superimposed on the other
2. One vertical line on the screen will be superimposed on the other
3. The two tones in the headset will achieve the same pitch, at which time a series of clicks will begin
4. The tone in the headset will decrease in pitch and be replaced by clicks that will become slow or nonexistent

THIS SPACE LEFT BLANK INTENTIONALLY.

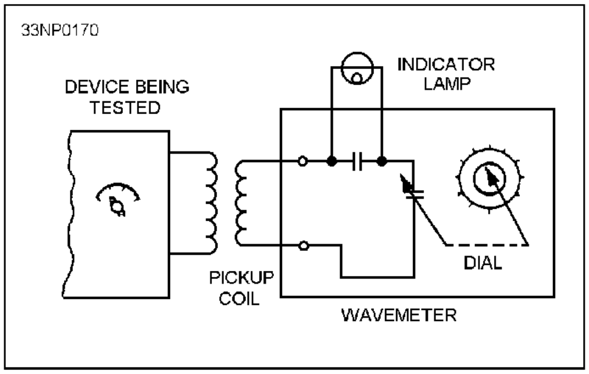


Figure 1F.—Absorption wavemeter circuit.

IN ANSWERING QUESTION 1-61, REFER TO FIGURE 1F.

1-61. When you are using an absorption wavemeter to measure frequency, the greatest accuracy may be obtained by loosely coupling the pickup coil so that the indicator lamp burns (a) with what degree of brilliance (b) under what resonance condition?

1. (a) Maximum brilliance
   (b) when tuned to the resonant frequency
2. (a) Maximum brilliance
   (b) when not tuned to the resonant frequency
3. (a) Dimly
   (b) when not tuned to the resonant frequency
4. (a) Dimly
   (b) when tuned to the resonant frequency

1-62. Which of the following instruments should be used to accurately measure a frequency in the shf range?

1. Absorption wavemeter
2. Resonant, cavity-type wavemeter
3. Resonant, coaxial-line-type wavemeter
4. Both 2 and 3 above

1-63. In a cavity wavemeter, moving the plunger farther into the cavity space causes which of the following changes to (a) the cavity size and (b) the resonant frequency of the cavity?

1. (a) Decrease (b) increase
2. (a) Decrease (b) decrease
3. (a) Increase (b) decrease
4. (a) Increase (b) increase

1-64. For which of the following purposes is a cathode-ray oscilloscope used?

1. To measure microwave energy
2. To visually analyze waveforms
3. To provide frequency modulation
4. To locate stray radio interference

1-65. The synchroscope contains which of the following circuits?

1. Retrace blanking circuit
2. A wide band amplifier
3. A trigger sweep
4. All of the above

THIS SPACE LEFT BLANK INTENTIONALLY.

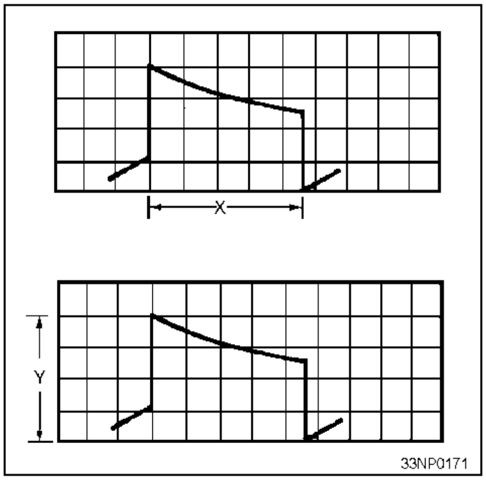


Figure 1G.—Waveform display.

IN ANSWERING QUESTION 1-66, REFER TO FIGURE 1G. MEASUREMENT X REPRESENTS THE HORIZONTAL DISTANCE OF THE WAVEFORM; MEASUREMENT Y REPRESENTS THE VERTICAL DISTANCE OF THE WAVEFORM.

1-66. In the figure, (a) time is represented by what measurement, and (b) amplitude is represented by what measurement?

1. (a) X (b) X
2. (a) X (b) Y
3. (a) Y (b) Y
4. (a) Y (b) X

1-67. An oscilloscope measures voltage and displays waveforms. It can be used to measure currents, temperatures, speeds, and accelerations if they are first converted to

1. heat
2. light
3. voltages
4. pressures

1-68. The spectrum analyzer is used to display which of the following quantities?

1. Amplitude within each frequency component in a circuit
2. Proportions of power within each frequency component in the spectrum
3. Frequencies produced in a circuit
4. Each of the above

1-69. While testing a semiconductor diode, you determine that the forward resistance value is 60 ohms. You should consider the diode good if the backward resistance is at least which of the following values?

1. 6 ohms
2. 60 ohms
3. 600 ohms
4. 6,000 ohms

1-70. When you are using an oscilloscope to test a crystal diode, what is shown by (a) the vertical deflection and (b) the horizontal deflection?

1. (a) Crystal current
   (b) voltage applied to the diode
2. (a) Crystal current
   (b) power developed in the diode
3. (a) Crystal voltage
   (b) power developed in the diode
4. (a) Crystal voltage
   (b) voltage applied to the diode

1-71. When you are using the oscilloscope to test the Zener diode, what is represented by (a) vertical deflection and (b) horizontal deflection?

1. (a) Zener current
   (b) Zener power
2. (a) Zener current
   (b) applied voltage
3. (a) Applied voltage
   (b) Zener power
4. (a) Applied voltage
   (b) Zener current

1-72. When troubleshooting transistorized circuits, you should first check the condition of which of the following circuits?

1. Counters
2. Amplifiers
3. Oscillators
4. Power supplies

1-73. Which of the following instruments is used to check transistors for collector leakage current and current gain?

1. Ohmmeter
2. Voltmeter
3. Wheatstone bridge
4. Semiconductor test set

1-74. When making base-to-emitter bias voltage checks on a transistor, you should read which of the following voltage ranges?

1. 5 to 20 microvolts
2. 50 to 200 millivolts
3. 5 to 20 volts
4. 50 to 200 volts

1-75. When making resistance measurements on a transistor with an ohmmeter, you should allow what maximum current in the transistor?

1. milliampere
2. milliampere
3. microampere
4. microampere

ASSIGNMENT 2

Textbook assignment: Chapter 3, "Basic Meters," pages 3-1 through 3-34. Chapter 4, "Common Test Equipment," pages 4-1 through 4-10.

2-1. What are the two basic components of a galvanometer?

1. A movable permanent magnet and a movable coil
2. A stationary permanent magnet and a stationary coil
3. A stationary permanent magnet and a movable coil
4. A stationary coil and a movable coil

2-2. The coil in a galvanometer rotates to allow measurement of current. Which of the following actions causes this reaction?

1. Current flowing in opposite directions through two coils
2. Tension of the hairspring and the magnetism produced by the permanent magnet
3. Magnetism produced by current flowing in the movable coil and the tension of the hairspring
4. Magnetism of the permanent magnet and magnetism produced by current in the movable coil

2-3. In a galvanometer, the phosphor bronze ribbon serves which of the following functions?

1. Provides a conduction path from the circuit being tested to the coil
2. Restores the coil to its original position
3. Allows the coil to twist
4. Each of the above

2-4. In galvanometers, which of the following components is/are used to indicate the value of the current being measured?

1. Pointer
2. Light and mirror
3. Both 1 and 2 above
4. Digital readout

2-5. The phosphor bronze ribbon in the galvanometer serves the same purpose as what component in the D'Arsonval meter?

1. Hairspring
2. Movable coil
3. Fixed iron core
4. Permanent magnet

2-6. The direction in which the D'Arsonval meter pointer deflects depends on what characteristic of the current applied to the coil?

1. Phase
2. Polarity
3. Frequency
4. Amplitude

2-7. The D'Arsonval meter movement is damped to prevent which of the following conditions?

1. Oscillating readings
2. Inconsistent readings
3. Consistently low readings
4. Consistently high readings

2-8. The weight of the rotating coil assembly and the type of bearings used in the D'Arsonval meter are factors that affect which of the following characteristics of the meter?

1. The accuracy and the linearity of the meter scales
2. The amount of restraining force required of the hairspring
3. The maximum current that can be measured
4. The sensitivity

2-9. For a meter to read linearly, its face is divided into equal segments. What meter feature makes this possible?

1. The curved poles of the permanent magnet
2. The jeweled bearings in the meter movement
3. An additional coil placed in the meter circuit
4. A long, lightweight meter pointer

2-10. What is the purpose of a shunt in a dc ammeter?

1. To decrease the sensitivity of the meter
2. To increase the linearity of the meter movement
3. To increase the current range of the meter
4. To decrease meter damping

2-11. A particular D'Arsonval meter has a full-scale current reading of 1 milliampere. A full-scale reading of 100 milliamperes may be achieved by using which of the following components?

1. A low-value resistance placed in series with the meter terminals
2. A high-value resistance placed in series with the meter terminals
3. A movable coil composed of large-diameter wire
4. A resistance of proper value placed in parallel with the meter terminals

2-12. To measure 10 milliamperes on a 1-milliampere D'Arsonval meter movement, a shunt resistance is added that will carry 9 milliamperes. What maximum value of current will pass through the meter movement?

1. 1 milliampere
2. 3 milliamperes
3. 6 milliamperes
4. 9 milliamperes

2-13. In a meter movement, shunt strips with a zero temperature coefficient are used instead of regular carbon resistance for which of the following reasons?

1. Because regular carbon resistances cause interfering magnetic fields
2. Because regular carbon resistances are too large to be used
3. Because regular carbon resistances are not able to handle the current changes
4. Because regular carbon resistances are affected by heat due to current and cause readings to vary

2-14. One consideration in choosing the value of a meter shunt resistance is that the meter readings should be in the midscale range. Which of the following factors is another consideration?

1. Meter switching is easier for midscale deflection
2. Meter shielding against magnetic interference is greatest near midscale
3. Minimum loading effect will be experienced near midscale
4. The meter is protected from unexpected surge currents

2-15. For which of the following current ranges would you likely use a meter that contains internal shunt?

1. 1 to 10 amperes
2. 11 to 30 amperes
3. 31 to 50 amperes
4. All of the above

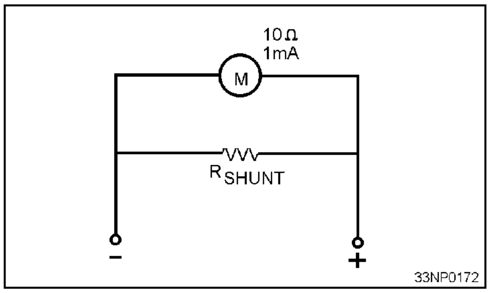


Figure 2A.—Shunt ammeter.

IN ANSWERING QUESTIONS 2-16 THROUGH 2-19, REFER TO FIGURE 2A. THE METER IN THE CIRCUIT IS DESIGNED FOR MAXIMUM OF .001 AMPERE AND HAS AN INTERNAL RESISTANCE OF 10 OHMS. YOU ARE FIGURING THE SHUNT RESISTANCE NECESSARY TO MEASURE 5 AMPERES.

2-16. What is the voltage drop across the meter coil?

1. .01 volt
2. .005 volt
3. .0001 volt
4. .0005 volt

2-17. What is the voltage drop across the shunt resistance?

1. .01 volt
2. .005 volt
3. .0001 volt
4. .0005 volt

2-18. When the meter is deflected full scale and is measuring 5 amperes, what is the maximum value of current flow through the shunt resistance?

1. 0.010 ampere
2. 0.490 ampere
3. 4.999 ampere
4. 5 amperes

2-19. What is the approximate value of the shunt resistance?

1. .001 ohm
2. .002 ohm
3. .003 ohm
4. .004 ohm

2-20. Simple range-switching arrangements for current meters are less satisfactory than other methods of range switching for which of the following reasons?

1. Meter damage can occur when line current momentarily flows through the meter
2. Resistance in the contacts may cause inaccurate readings
3. Both 1 and 2 above
4. Resistor damage may occur

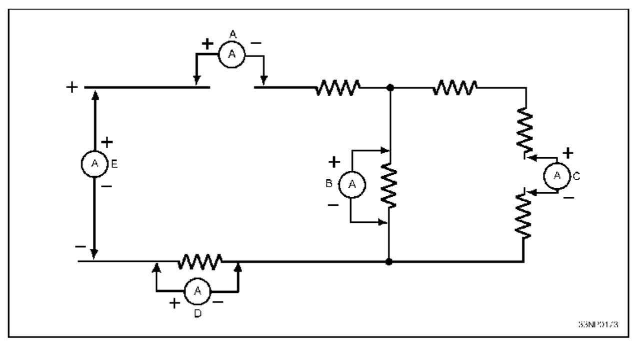


Figure 2B.—Ammeter connections.

IN ANSWERING QUESTION 2-21, REFER TO FIGURE 2B.

2-21. In the figure, five ammeters are connected to the circuit resistors. Of those five, which one(s) is/are connected correctly?

1. E only
2. A and E
3. A and C
4. A, B, and D

2-22. What will be the probable result of connecting an ammeter (or milliammeter) in PARALLEL with a source of voltage or a circuit component?

1. A burned-out meter that will provide no useful readings
2. A higher than normal meter reading
3. A lower than normal meter reading
4. A normal meter reading

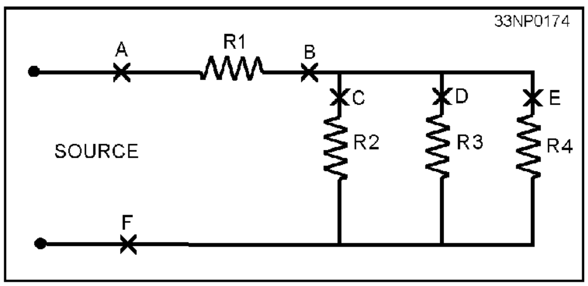


Figure 2C.—Series-parallel circuit.

IN ANSWERING QUESTIONS 2-23 AND 2-24, REFER TO FIGURE 2C.

2-23. To measure total circuit current, you should break the circuit to connect an ammeter at which of the following points?

1. A only
2. B or C
3. C, D, or E
4. A, B, or F

2-24. To measure the current flow through resistor R2 only, you should break which of the following points to connect the ammeter?

1. A
2. B
3. C
4. F

2-25. Meter sensitivity is determined by the amount of current required by the ammeter coil to provide full-scale deflection of the pointer. An ammeter coil requiring which of the following current values provides the greatest sensitivity?

1. 1 milliampere
2. 10 milliamperes
3. 100 microamperes
4. 500 microamperes

2-26. Circuits in which low-sensitivity ammeters are used are said to be "loaded." Which of the following statements describes the cause of circuit loading?

1. The ammeter circuit draws NO current from the circuit being tested
2. The ammeter circuit draws MINIMUM current from the circuit being tested
3. The ammeter circuit draws EXCESSIVE current from the circuit being tested
4. The ammeter circuit INDUCES current into the circuit being tested

2-27. In which of the following electronic circuits does the use of a meter with low sensitivity have the greatest loading effect?

1. High-power circuits
2. Low-current circuits
3. High-current circuits
4. High-voltage circuits

2-28. A basic D'Arsonval meter is used to measure voltage by connecting its meter coil to (a) what type of component in (b) what circuit arrangement?

1. (a) Multiplier resistor
   (b) series
2. (a) Multiplier resistor
   (b) parallel
3. (a) Capacitor
   (b) parallel
4. (a) Capacitor
   (b) series

2-29. In a voltmeter, the D'Arsonval meter movement is caused to move by what electrical action?

1. Power
2. Voltage
3. Current
4. Conductance

2-30. In a voltmeter, the meter scale is calibrated in which of the following categories?

1. Power
2. Voltage
3. Current
4. Conductance

2-31. To figure the total value of series resistance needed to extend the range of a voltmeter, you need to know the value of current to cause full-scale deflection of the meter and what other value?

1. Minimum applied voltage
2. Maximum applied voltage
3. Maximum applied current
4. Minimum applied current

2-32. Your voltmeter has four ranges: 1V, 10V, 100V, and 1,000V. To measure an unknown voltage in an amplifier, you should first select what range?

1. 1 V
2. 10 V
3. 100 V
4. 1,000 V

2-33. Which of the following types of circuits are most affected by the loading effect of voltmeters?

1. Low-voltage
2. Low-resistance
3. High-resistance
4. High-current

2-34. A voltmeter with a 10-microampere meter movement has a sensitivity of how many maximum ohms per volt?

1. 1,000
2. 10,000
3. 100,000
4. 1,000,000

2-35. A megger is widely used for which of the following purposes?

1. To make voltage checks
2. To make continuity checks
3. To measure insulation resistance
4. To measure resistance of components

2-36. Before you can take an accurate resistance measurement with an ohmmeter, what meter adjustment must you complete?

1. Zero voltage
2. Zero resistance
3. Maximum voltage
4. Maximum resistance

2-37. When the leads of an ohmmeter are placed across a resistor, that resistor adds to the internal series coil resistance of the meter. The pointer is deflected to the left of its full-scale position, giving a reading in ohms for the resistor being tested. Which of the following reasons explains why the pointer moves to less than full-scale?

1. Because voltage in the meter circuit is greater than full-scale voltage
2. Because current in the meter circuit is greater than full-scale current
3. Because current in the meter circuit is less than full-scale current
4. Because voltage in the meter circuit is less than full-scale current

2-38. With the ohmmeter range switch set at R X 100, the pointer of the meter indicates 850. What is the actual value of the resistor?

1. 8.5 kilohms
2. 85 kilohms
3. 850 kilohms
4. 8.5 megohms

RANGE SCALE		RESISTOR VALUE
A.	R×1	50 kilohms
B.	R×10	500 kilohms
C.	R×100	5 megohms

Figure 2D.—Range scales and resistance values.

IN ANSWERING QUESTION 2-39, REFER TO FIGURE 2D.

2-39. You are measuring resistors using range settings as shown in the figure. What condition listed in the figure, if any, will cause the greatest amount of current to move through the ohmmeter coil circuit?

1. A
2. B
3. C
4. None; they all allow the same amount of current

2-40. Which of the following locations on the meter scale provides the most accurate reading for resistance?

1. To the far left side of the scale
2. To the far right side of the scale
3. Halfway between the left side and center of the scale
4. Near the center of the scale

2-41. An ordinary ohmmeter is unsuitable for measuring insulation resistance for which of the following reasons?

1. Voltage is present in the conductors attached to the insulating materials being measured
2. Insulation resistance values are too great for an ohmmeter to measure
3. Ohmmeter current will damage insulation material
4. The accuracy of an ohmmeter is too low to measure insulation resistance

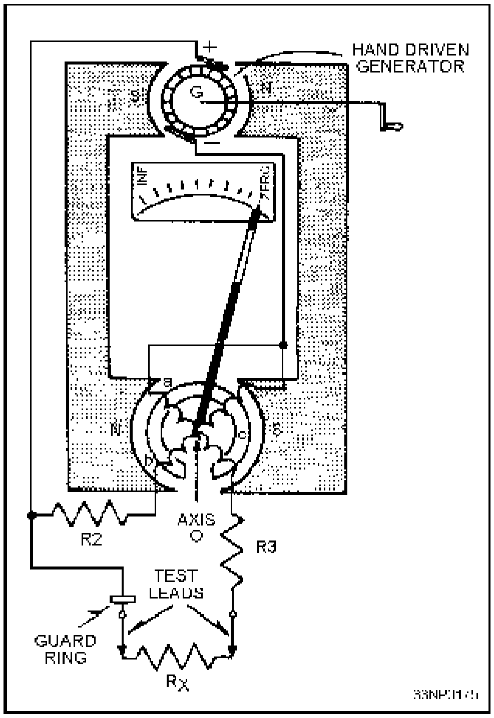


Figure 2E.—Megger circuit.

IN ANSWERING QUESTION 2-42, REFER TO FIGURE 2E.

2-42. In the megger circuit, what component prevents leakage current from affecting ohmmeter measurements?

1. Guard ring
2. Generator
3. Resistor R3
4. Coil

2-43. When a megger is used to measure an unknown resistance, what circuit action causes the pointer to come to rest at the correct resistance value?

1. The current flow in coil A
2. The current flow in coil B
3. The interaction between the currents in coils A and B
4. The interaction between the restraining springs of the meter and the current in coil A

2-44. A megger uses high voltage to check the insulation leakage in the megohm range. What is the source of this voltage?

1. Batteries
2. The ship's ac power
3. The ship's dc power
4. A hand-driven dc generator

2-45. Meggers with which of the following maximum voltage ratings are commonly found in the fleet?

1. 500 volts
2. 700 volts
3. 1,000 volts
4. 2,500 volts

2-46. When a megger is used to measure the resistance of an electrical cable, what does a reading of infinity indicate?

1. The meter is faulty
2. The cable is shorted
3. The cable is grounded
4. The resistance is too large to measure

2-47. A megger is prevented from exceeding its rated output voltage by which of the following actions?

1. Battery discharge limits the voltage
2. Tension in the cable
3. Friction clutch slippage
4. Current leaks through internal insulation

2-48. When the crank of a 500-volt megger is rotated faster than its designed rate, what maximum output voltage does it produce?

1. 100 volts
2. 500 volts
3. 520 volts
4. 550 volts

2-49. The galvanometer-type meter movement differs from the electrodynamometer meter movement in that the electrodynamometer type uses which of the following components to produce the magnetic field?

1. Two sets of coils
2. Two permanent magnets
3. One fixed and one movable coil
4. One movable coil and one permanent magnet

2-50. The fixed coils in the electrodynamometer-type movement are wound with heavy wire to enable the instrument to measure which of the following values?

1. Rf currents
2. High voltage
3. Large currents
4. High resistance

2-51. An advantage that the electrodynamometer has over the standard galvanometer in measuring ac is that the electrodynamometer requires

1. no rectifying device
2. a less complicated rectifying device
3. less current to obtain a full-scale deflection
4. fewer multiplier resistors to cover the measurement range

THIS SPACE LEFT BLANK INTENTIONALLY.

2-52. When power is being measured with a wattmeter, why is it important to strictly observe the safe rating limits for current and voltage?

1. Because the meter pointer will likely hit against the upper limit of the dial and be damaged
2. Because the pointer does not give an indication to alert the user when the internal coils are overloaded
3. Because a false reading may be obtained
4. Because the meter pointer will likely hit against the lower limit of the dial and be damaged

2-53. The reading of a wattmeter is dependent upon which of the following circuit characteristics?

1. Current
2. Voltage
3. Power factor
4. All of the above

2-54. A continuity test is performed on a piece of electronic equipment to discover what kind of fault?

1. High voltage
2. Low voltage
3. Open circuits
4. Changes in component values

2-55. Which of the following meters is recommended for circuit continuity tests?

1. A megger
2. An ammeter
3. A voltmeter
4. An ohmmeter

2-56. When preparing to use an ohmmeter to test a circuit for grounds, you should first take which of the following actions?

1. Energize the circuit
2. Disconnect all intentional grounds
3. Measure the circuit voltage at the power source
4. Connect all intentional grounds

2-57. When preparing to use a voltmeter to measure voltage in a circuit, you should first take which of the following actions?

1. Set the meter to the lowest voltage range
2. Remove the suspected component from the circuit
3. Check the voltage from the power source to ensure it is correct
4. Check the current flow through the circuit with an ammeter

2-58. It is important to set a voltmeter on its highest range scale before taking a measurement for which of the following reasons?

1. To protect the meter from damage
2. To decrease the effects of input impedance
3. To increase the sensitivity of the measurement
4. To protect the equipment being tested from damage

THIS SPACE LEFT BLANK INTENTIONALLY.

2-59. An accurate measurement of a particular resistor in a set of parallel resistors may be obtained by performing which of the following procedural steps?

1. Connecting the ohmmeter leads across the resistor while in place
2. Disconnecting the resistor from the set before taking the measurement
3. Grounding the resistance set before taking the measurement
4. Using the highest ohmmeter range

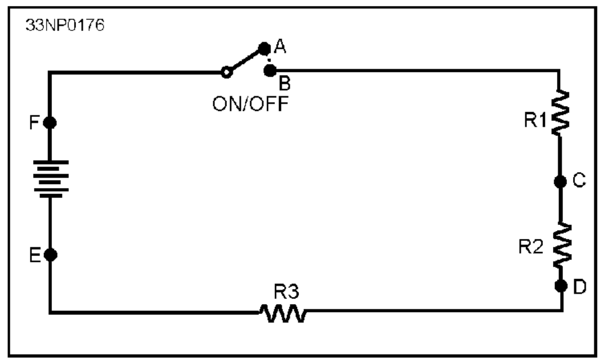


Figure 2F.—Circuit.

IN ANSWERING QUESTION 2-60, REFER TO FIGURE 2F.

2-60. To measure R2 with an ohmmeter, (a) in what position should the ON/OFF switch be placed, and (b) where should the test probes be placed?

1. (a) Off (b) at B and C
2. (a) Off (b) at C and D
3. (a) On (b) at C and D
4. (a) On (b) at B and C

2-61. In what arrangement is an ammeter connected to a circuit?

1. In series
2. In parallel
3. In a series-parallel combination
4. In a parallel-series combination

2-62. When you are measuring voltage using a voltmeter, where should you stand to view the meter reading?

1. To the right of the meter only
2. To the left of the meter only
3. To the right or left of the meter, depending on your handedness
4. Directly in front of the meter

2-63. A multimeter is used to measure which of the following electrical properties?

1. Voltage
2. Current
3. Resistance
4. Each of the above

2-64. Which of the following characteristics is an advantage of a volt-ohm meter?

1. It replaces three separate meters
2. There are no calibrations to be made
3. It is the most accurate meter available
4. It cannot be damaged

A. SELECT RANGE SCALE
B. SHORT ENDS OF PROBES TOGETHER
C. ZERO THE METER USING THE ZERO ADJUST CONTROL

Figure 2G.—Zeroadjust steps.

IN ANSWERING QUESTION 2-65, REFER TO FIGURE 2G.

2-65. Before you measure resistance, it is important that you calibrate (zero) the ohmmeter. In what order should the actions in the figure be performed?

1. A, B, and C
2. B, C, and A
3. C, B, and A
4. B, A, and C

2-66. On an ohmmeter, which of the following switches allows selection of ac or dc readings?

1. ZERO-OHMS
2. FUNCTION
3. RANGE
4. RESET

2-67. Which of the following actions MUST be taken before resistance measurements are made in a circuit?

1. All semiconductor devices must be removed from the circuit
2. Expected measurements must be recorded
3. The circuit must be completely de-energized
4. The high range of the ohmmeter must be selected

2-68. The power required to operate a basic ohmmeter comes from which of the following sources?

1. Batteries
2. An ac power supply
3. Both 1 and 2 above
4. A hand crank

THIS SPACE LEFT BLANK INTENTIONALLY.

	POINTER LOCATION
1.	EXTREME RIGHT
2.	EXTREME LEFT
3.	MIDSCALE

Figure 2H.—Ohmeter pointer location

IN ANSWERING QUESTIONS 2-69 THROUGH 2-72, REFER TO FIGURE 2H AND SELECT THE OHMMETER POINTER LOCATION THAT RESULTS FROM THE CONDITION GIVEN.

2-69. Both test leads are touching the metal chassis of a piece of equipment.

2-70. The test leads are shorted together.

2-71. The test leads are separated from each other and not touching anything else.

2-72. An accurate resistance reading is being made.

2-73. The function of the ZERO OHM control on a multimeter is to compensate for which of the following conditions?

1. Meter battery aging
2. Large values of resistance in the circuit to be measured
3. Inter-electrode capacitance in the circuit to be measured
4. Stray voltages in the circuit under test

2-74. When using a multimeter to measure an output voltage, you should ensure that the dc voltage component does not exceed what maximum voltage?

1. 100 volts
2. 200 volts
3. 300 volts
4. 400 volts

2-75. When measuring unknown currents, you should determine the range scale that is appropriate in what way?

1. Start with the expected scale
2. Start with the lowest scale and work up
3. Start with the highest scale and work down
4. Use the highest scale only

ASSIGNMENT 3

Textbook assignment: Chapter 4, "Common Test Equipment," pages 4-11 through 4-28. Chapter 5, "Special Application Test Equipment," pages 5-1 through 5-40. Chapter 6, "The Oscilloscope and Spectrum Analyzer," pages 6-1 through 6-46.

3-1. Compared to the VOM, the electronic digital multimeter has which of the following advantages?

1. It has higher input impedance
2. It can be read directly without using a scale
3. It has little or no loading effect on the circuit under test
4. All of the above

3-2. Digital multimeters can be divided into three functional sections: (1) signal conditioning, (2) analog-to-digital, and what third section?

1. Display
2. Amplifier
3. Oscillator
4. Power supply

3-3. What section in the electronic multimeter decodes the digital (binary) information from the a/d converter section?

1. Display
2. Analog ic
3. Digital ic
4. Signal conditioning

3-4. The ac/dc differential voltmeter is capable of performing which of the following functions?

1. Voltage readings as an electronic voltmeter
2. Voltage readings as a precision potentiometer
3. Voltage variation readings about some known value
4. Each of the above

3-5. In a differential voltmeter, an adjustable reference voltage can be produced in increments as small as how many microvolts?

1. 1
2. 2
3. 5
4. 10

3-6. When you are taking ac or dc voltage readings, the differential voltmeter will not load the circuit under test.

1. True
2. False

3-7. Which of the following transistor parameters are measured with the transistor tester?

1. Collector leakage and cutoff current
2. Collector leakage and maximum power dissipation
3. Collector leakage and current gain
4. Emitter leakage and power gain

3-8. When you are disconnecting a transistor from the transistor tester, in what position should the POLARITY switch be placed?

1. ON
2. OFF
3. PNP
4. NPN

3-9. When you are testing a transistor with a transistor tester, for the reading to have ±15 percent accuracy, the resistance from emitter to base must be what minimum value?

1. 100 ohms
2. 200 ohms
3. 300 ohms
4. 50 ohms

3-10. When you are testing an in-circuit diode with the transistor tester, the meter pointer deflects below the midscale point. What does this indicate?

1. The diode is normal
2. The diode is faulty
3. The circuit impedance is 8 ohms
4. The POLARITY switch is in the PNP position

3-11. The RCL bridge measures an unknown resistance by balancing the resistance value of the unknown component with that of known components inside the test set. What type of circuit is used in this method of measurement?

1. O'Neill bridge
2. Wheatstone bridge
3. Colpitts oscillator
4. D'Arsonval movement

3-12. The RCL bridge is used to measure which of the following quantities?

1. Resistance, capacitance, and inductance
2. Capacitor quality
3. The turns ratio of transformers
4. Each of the above

3-13. To make an inductance measurement on the model 250DE, you must first adjust the DET GAIN control to what position?

1. 1
2. 2
3. 3
4. 4

3-14. The direct-measuring power meter is used for which of the following types of measurements?

1. Incident power
2. Reflected power
3. Both 1 and 2 above
4. Average power

3-15. When you are selecting forward or reverse power measurements, what component(s) of the rf wattmeter is/are restricted to a 180°; rotation range?

1. The coupler-detector
2. The POWER RANGE knob
3. The u-type connector
4. Both 2 and 3 above

3-16. Of the following available rf wattmeter power ranges, which one should you select when measuring an unknown power?

1. 10 watts
2. 100 watts
3. 500 watts
4. 1000 watts

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 3-17 AND 3-18, ASSUME THAT YOU HAVE OBTAINED THE FOLLOWING RF POWER MEASUREMENTS:

— INCIDENT POWER, 144 WATTS
— REFLECTED POWER, 1 WATT

3-17. What is the approximate standing wave ratio?

1. .85 to 1
2. 1 to 1
3. .08 to 1
4. 1.2 to 1

3-18. What total amount of power is absorbed by the load?

1. 1 watt
2. 13 watts
3. 143 watts
4. 173 watts

3-19. To take power readings, you connect an in-line rf power meter in what configuration in the transmission line?

1. Series
2. Parallel
3. Series-parallel
4. Horizontal

3-20. Which of the following test instruments produces a standard of measurement of ac energy for testing electronics equipment?

1. Power meter
2. Oscilloscope
3. Signal generator
4. Frequency counter

3-21. Signal generators are equipped with attenuators for which of the following purposes?

1. To regulate the output frequency
2. To regulate the output signal level
3. To determine the modulating frequency
4. To set the level of internal modulation

3-22. Which of the following signal generators should you select to test audio equipment?

1. Af signal generator
2. Fm signal generator
3. Rf signal generator
4. Video signal generator

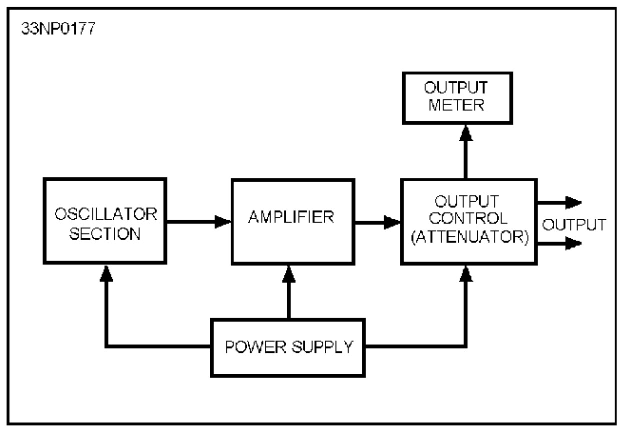


Figure 3A.—Af signal generator block diagram.

IN ANSWERING QUESTION 3-23, REFER TO FIGURE 3A.

3-23. What section of the signal generator regulates the output to the equipment under test?

1. The amplifier
2. The attenuator
3. The oscillator
4. The output meter

3-24. Typical rf signal generators use which of the following methods for modulation?

1. Internal modulation
2. External modulation
3. Both 1 and 2 above
4. Power supply modulation

3-25. The modulating circuit in an rf signal generator serves which of the following purposes?

1. It produces an fm signal that can be superimposed on the rf signal
2. It produces a video signal that can be superimposed on the rf signal
3. It produces an audio signal that can be superimposed on the rf signal
4. Both 2 and 3 above

3-26. The output level meter of a signal generator reads 0. 10 with the attenuator set to 080. 0. What is the actual output in microvolts?

1. 00.08
2. 00.80
3. 08.00
4. 80.00

3-27. Frequency counters serve which of the following purposes?

1. Measure voltage in a circuit
2. Measure frequencies in a circuit
3. Produce voltages to test a circuit
4. Produce frequencies to power a circuit

3-28. In which of the following ways is a logic probe useful to a technician?

1. It detects steady logic levels
2. It detects a train of logic levels
3. Both 1 and 2 above
4. It displays the shape of high-speed transients

3-29. In logic probes, which of the following devices is used as an indicator?

1. A dial
2. An LED
3. A scale
4. A pointer

3-30. The Tracker 2000 is used to isolate defective components on de-energized circuits only.

1. True
2. False

3-31. The VERT control adjustment controls what position of the signal on the CRT display?

1. Horizontal
2. Vertical
3. Trace rotation left
4. Trace rotation right

3-32. The axes of the CRT display on the Tracker 2000 are divided into what number of quadrants?

1. 1
2. 2
3. 3
4. 4

3-33. Quadrant 2 on the Tracker 2000 CRT is used to display what signal information?

1. Positive voltage negative current
2. Positive voltage and positive current
3. Negative voltage and negative current
4. Negative voltage and positive current

3-34. When you are adjusting the level control on the Tracker 2000, the peak of each pulse will be from 0 volts to what maximum voltage level?

1. 1
2. 5
3. 3
4. 9

3-35. The medium 1 range on the Tracker 2000 is designed to test resistance values between what (a) minimum and (b) maximum values?

1. (a) 5 ohms (b) 1 kilohm
2. (a) 50 ohms (b) 10 kilohms
3. (a) 500 ohms (b) 100 kilohms
4. (a) 50 kilohms (b) 10 megohms

3-36. What is the principal use of the oscilloscope?

1. To measure microwave energy
2. To visually examine waveforms
3. To measure in-line power supply currents
4. To locate sources of radio interference

3-37. Cathode-ray tubes used in oscilloscopes contain which of the following components?

1. An electron gun
2. A deflection system
3. A fluorescent screen
4. All of the above

3-38. In a basic oscilloscope, what is the purpose of the deflection system?

1. To filter harmonic frequencies
2. To deflect harmonic frequencies
3. To position the beam on the screen
4. To deflect synchronous side effects

3-39. If the electron beam is left in one position on the CRT for long periods, damage is likely to occur to what component(s)?

1. Illuminating coating
2. Deflection plates
3. Signal generator
4. Electron gun

3-40. Of the following factors, which one(s) control(s) the angle of deflection of the electron beam in the CRT?

1. Difference of potential between plates
2. Length of deflection field
3. Beam acceleration
4. All of the above

3-41. In an oscilloscope, which of the following waveform characteristics are represented by (a) vertical deflection and (b) horizontal deflection?

1. (a) Amplitude (b) frequency
2. (a) Power (b) amplitude
3. (a) Power (b) frequency
4. (a) Time (b) amplitude

3-42. In oscilloscopes using electrostatic CRTs, what type of signal is used to produce horizontal beam movement?

1. Dc
2. Sine wave
3. Square wave
4. Sawtooth wave

3-43. The length of time the phosphor coating on the CRT remains bright after the electron beam is removed depends on which of the following characteristics?

1. Persistence of the coating
2. Amplitude of the applied signal
3. Frequency of the applied signal
4. Synchronization frequency of the oscilloscope

3-44. An oscilloscope that can display two vertical input signals at the same time is said to be what type?

1. Two-function
2. Dual-trace
3. Single-function
4. Single-trace

3-45. The FOCUS control on the front of an oscilloscope is used to adjust what characteristic on the CRT display?

1. Beam size
2. Beam location
3. Trace position
4. Beam brilliance

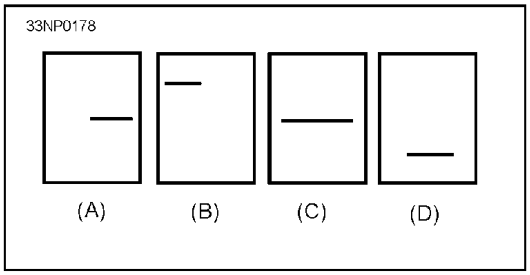


Figure 3B.—CRT traces.

IN ANSWERING QUESTIONS 3-46 THROUGH 3-48, REFER TO FIGURE 3B.

3-46. What trace in the figure should you correct by adjusting ONLY the HORIZONTAL POSITION control?

1. A
2. B
3. C
4. D

3-47. What trace in the figure should you correct by adjusting ONLY the VERTICAL POSITION control?

1. A
2. B
3. C
4. D

3-48. In the figure, what trace would be corrected by adjusting both the HORIZONTAL POSITION and VERTICAL POSITION controls?

1. A
2. B
3. C
4. D

3-49. What is the purpose of the deflection amplifiers in a cathode-ray oscilloscope?

1. To isolate the input signal from the vertical deflection plates
2. To increase the amplitude of the signal applied to the vertical deflection plates
3. To eliminate distortion of the CRT beam
4. To position the beam on the CRT

3-50. What control on the front panel of an oscilloscope limits the input signal amplitude and allows the oscilloscope to be used with a wide range of signals?

1. TIME/CM
2. TIME BASE
3. ATTENUATOR
4. TRIGGER

3-51. What is the purpose of a variable potentiometer mounted on the VOLTS/DIV control of an oscilloscope?

1. To provide definite step control of the input signal
2. To provide fine control of the input signal
3. To provide fine control of the output signal
4. To provide definite step control of the output signal

3-52. The time base of an oscilloscope is variable to enable the instrument to

1. measure low- and high-amplitude signal voltages
2. operate over a wide range of input frequencies
3. make accurate measurements of signal amplitudes
4. accurately position the presentation on the CRT

3-53. The triggered oscilloscope has which of the following advantages over a basic oscilloscope?

1. Better low-voltage input- handling capability
2. Lower distortion of input signal
3. Both 1 and 2 above
4. Improved presentation stability

3-54. The TRIGGER and LEVEL controls are used to synchronize the sweep generator with what signal?

1. Input
2. Output
3. Vertical deflection rate
4. Horizontal deflection rate

3-55. The setting of what control establishes the amplitude point of the input waveform at which the displayed sweep will begin?

1. TRIGGER LEVEL
2. TRIGGER SLOPE
3. AMPLITUDE
4. A delayed by B

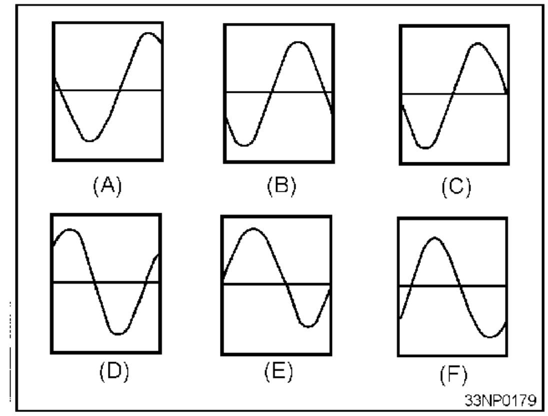


Figure 3C.—Oscilloscope displays.

IN ANSWERING QUESTIONS 3-56 THROUGH 3-60, REFER TO FIGURE 3C. SELECT THE OSCILLOSCOPE DISPLAY THAT CORRESPONDS TO THE SETTING OF THE TRIGGER LEVEL AND SLOPE CONTROLS IN THE QUESTIONS.

3-56. TRIGGER LEVEL set to zero; SLOPE set to negative.

1. A
2. C
3. E
4. F

3-57. TRIGGER LEVEL set to positive; SLOPE set to negative.

1. A
2. D
3. E
4. F

3-58. TRIGGER LEVEL set to negative; SLOPE set to positive.

1. B
2. C
3. D
4. F

3-59. TRIGGER LEVEL set to negative; SLOPE set to negative.

1. A
2. B
3. C
4. D

3-60. TRIGGER LEVEL set to positive; SLOPE set to positive.

1. A
2. C
3. D
4. F

3-61. Which of the following electrical quantities can be measured using an oscilloscope?

1. Current
2. Frequency
3. Inductance
4. Capacitance

3-62. What effect does sweep frequency that is higher than the incoming signal frequency have on the displayed signal?

1. It produces a jittery view of the incoming signal
2. It produces a multiple view of the incoming signal
3. It produces an exact view of the incoming signal
4. It produces less than a complete view of the incoming signal

3-63. A dual-trace oscilloscope differs from a dual-beam oscilloscope in that the dual-trace device uses (a) what number of electron beams and (b) what number of channels?

1. (a) 1 (b) 1
2. (a) 2 (b) 1
3. (a) 2 (b) 2
4. (a) 1 (b) 2

3-64. What two modes are used to obtain the dual trace on an oscilloscope?

1. Chop and beam
2. Chop and alternate
3. Slow sweep and beam
4. Alternate and slow sweep

3-65. Of the following oscilloscope switch settings, which ones will provide desirable (a) slow sweep speeds and (b) high sweep speeds?

1. (a) CHOP
   (b) ALTERNATE
2. (a) SLOW SWEEP
   (b) ALTERNATE
3. (a) BEAM SWITCH
   (b) SLOW SWEEP
4. (a) ALTERNATE
   (b) CHOP

3-66. In a dual-trace oscilloscope, the gate that controls both sweeps is controlled by a multivibrator that operates at which of the following maximum frequencies?

1. 50 kHz
2. 100 kHz
3. 500 kHz
4. 1,200 kHz

3-67. What is the basic internal configuration of the typical dual-trace oscilloscope?

1. One gun assembly with two vertical channels
2. Two gun assemblies with two horizontal channels
3. One gun assembly with two vertical and horizontal channels
4. Two gun assemblies with one vertical and horizontal channel

3-68. The horizontal sweep channels of the dual-trace oscilloscopes have which of the following time base circuit configurations?

1. One time base circuit
2. Two interdependent time base circuits
3. Two independently controlled time base circuits
4. Variations of all the above configurations

3-69. Which of the following types of probes are commonly available for use with the oscilloscope?

1. Current
2. One-to-one
3. Attenuation
4. All of the above

3-70. On the spectrum analyzer, what information can be found on the (a) x axis and (b) y axis?

1. (a) Voltage (b) current
2. (a) Frequency (b) amplitude
3. (a) Amplitude (b) frequency
4. (a) Current (b) voltage

3-71. The intensity knob on the spectrum analyzer controls what display function(s)?

1. Focus
2. Brightness of the CRT trace
3. Brightness of the readout display
4. Both 2 and 3 above

3-72. On the spectrum analyzer, the RF input jack is used to accept what maximum frequency input signals?

1. 21 GHz or below
2. 21 GHz or above
3. 31 GHz or below
4. 31 GHz or above

3-73. The POSITION control on the spectrum analyzer is used to adjust which of the following CRT display features?

1. Vertical position only
2. Horizontal position only
3. Vertical and horizontal position
4. Beam focus

NONRESIDENT
TRAINING
COURSE
SEPTEMBER 1998

Navy Electricity and Electronics Training Series

Module 17—Radio-Frequency Communications Principles

NAVEDTRA 14189

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Although the words “he,” “him,” and “his” are used sparingly in this course to enhance communication, they are not intended to be gender driven or to affront or discriminate against anyone.

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: To introduce the student to the subject of Radio-Frequency Communications Principles who needs such a background in accomplishing daily work and/or in preparing for further study.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and either the occupational or naval standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards*, NAVPERS 18068.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
EWC Gary L. Holloway

NAVSUP Logistics Tracking Number
0504-LP-026-8420

Sailor’s Creed

“I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country’s Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all.”

TABLE OF CONTENTS

CREDITS

The figures listed below and included in this edition of NEETS, Module 17, *Radio-Frequency Communications Principles*, were provided by Martin Marietta DSCS III Program. Permission to use these illustrations is gratefully acknowledged.

SOURCE	FIGURE
Martin Marietta DSCS III Program (former company name: General Electric Company)	4-15 4-16

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7*, Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 7 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

THIS PAGE LEFT BLANK INTENTIONALLY.

Student Comments

Course Title: *NEETS Module 17*
Radio-Frequency Communications Principles

NAVEDTRA: 14189 **Date**: ______

We need some information about you:

Rate/Rank and Name: ______ SSN: ______ Command/Unit ______

Street Address: ______ City: ______ State/FPO: ______ Zip ______

Your comments, suggestions, etc.:

Privacy Act Statement: Under authority of Title 5, USC 301, information regarding your military status is requested in processing your comments and in preparing a reply. This information will not be divulged without written authorization to anyone other than those within DOD for official use in determining performance.

NETPDTC 1550/41 (Rev 4-00)

CHAPTER 1

INTRODUCTION TO RADIO-FREQUENCY COMMUNICATIONS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the OCC/ECC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completion of this chapter, you will be able to:

1. Define electrical telecommunications.
2. Describe the use of radiotelegraph, radiotelephone, teletypewriter, and facsimile.
3. Define and describe the interrelationships of the system, set, group, unit, assembly, subassembly, part, and reference designations.
4. State the frequency ranges of the various frequency bands and describe the most common uses of those bands by the Navy.
5. Describe a strategic communications link.
6. Describe a tactical communications link.
7. Describe the five basic communications modes of operation.
8. Describe a switched communications network.
9. Describe the purpose of the two Navy-only networks.

INTRODUCTION TO NAVAL TELECOMMUNICATIONS

When the wireless (radiotelegraph) was invented, the Navy saw a possible use for it. It could be used for communications from shore stations to ships along the coast. In 1899, the first official naval radio message was sent from ship to shore. It only traveled a distance of 20 miles but that was a start. The next advance was in 1916 when the Navy first used radiotelephone between ships. Three years later the first airborne radio was used to communicate with a ground station. In the early years, communications was not the best because of poor tuning techniques. Receivers often did not pick up the signal. This problem was almost eliminated in 1931 when the first superheterodyne receivers were installed in the fleet. In 1944, another important event took place. The first successful radio teletypewriter transmissions between ships were completed. The first successful use of radiophoto (facsimile) occurred in 1945 with the transmission of the surrender document signing that ended World War II. Naval communications has grown tremendously in size and complexity since then.

The fleets of our modern Navy travel faster and are spread over greater areas of ocean than any seagoing force of the past. Commanders and their subordinates throughout the Department of the Navy use the facilities of naval communications as a primary method of communicating.

Naval communications relies on top performance from all of its assigned personnel. Reliable, secure, and timely transmission and receipt of information, based on wartime requirements, is the ultimate goal.

Previous modules have discussed electronic components or circuitry in individual units. In this chapter we will tie up some loose ends for you and discuss radio-frequency communications.

We will cover the considerations involved in receiving or transmitting a radio-frequency signal between two or more geographic locations. Let's start by defining telecommunications.

TELECOMMUNICATIONS refers to communications over a distance and includes any transmission, emission, or reception of signs, signals, writings, images, or sounds. Intelligence produced by visual means, oral means, wire, radio, or other electromagnetic systems are also included. Electrical, visual, and sound telecommunications are all used in the Navy. In this chapter we will talk only about electrical types of telecommunications.

ELECTRICAL

The types of electrical communications are radio and wire. Radio uses electromagnetic waves to transmit and receive intelligence. The waves are not guided by a physical path between sender and receiver. Wire uses conductors to carry these waves. Radio is the most important method the Navy has of communicating between widely separated forces. The transmission methods we will be discussing are radiotelegraph, radiotelephone, teletypewriter, and facsimile.

Radiotelegraph

Radiotelegraph transmissions are referred to as continuous wave (cw) telegraphy. Cw is a manual or automatic system of transmitting signals using a wave of radio-frequency (rf) energy. The radio operator separates a continuously transmitted wave into dots and dashes based on the Morse code. This is accomplished by opening and closing a telegraphic hand key.

Radiotelegraphy was the first means of radio communications that had military and commercial importance. Radiotelegraph still is used as a means of communication to, from, and among widely separated units of the Navy.

Relative slow speed of transmission and the requirement for experienced operators are the major disadvantages of radiotelegraph. The main advantage is reliability. A thinking person at both sending and receiving stations provides a capability of being understood not present in automated systems.

Radiotelephone

Radiotelephone is one of the most useful military communications methods. Because of its directness, convenience, and ease of operation, radiotelephone is used by ships, aircraft, and shore stations. It has many applications and is used for ship-to-shore, shore-to-ship, ship-to-ship, air-to-ship, ship-to-air, air-to-ground, and ground-to-air communications. Modern means of operation make it possible to communicate around the world by radiotelephone. One of the most important uses of radiotelephone is short-range tactical communications. This method permits tactical commanders to communicate directly with other ships. Little delay results while a message is prepared for transmission, and acknowledgments can be returned instantly. Radiotelephone equipment for tactical use usually is operated on frequencies that are high enough to have line-of-sight characteristics; that is, the waves do not follow the curvature of the earth. As you know, these characteristics limit the usual range of radiotelephone from 20 to 25 miles. This is important because it reduces the chances of the enemy intercepting the message. Radiotelephone procedures can be learned easily by persons with no other training in communications.

Radiotelephone has some disadvantages. You may find transmissions unreadable because of static, enemy interference, or high local noise level caused by shouts, gunfire, and bomb or shell bursts. Wave propagation characteristics of radiotelephone frequencies sometimes are unpredictable, and tactical transmissions may be heard from great distances. Most radiotelephone messages are in plain language, and if information is to be kept from the enemy, users must keep their messages short, stick to the proper procedures, and be careful of what they say.

Q1. What are the two types of electrical communications?

Q2. What is the main advantage of radiotelegraph communications?

Q3. Why is radiotelephone one of the most useful methods of military communications?

Q4. What are the disadvantages of radiotelephone communications?

Teletypewriter

Teletypewriter (tty) signals may be transmitted by either landline (wire), cable, or radio. The landline tty is used both by the military services and by commercial communication companies. The Navy uses radio teletypewriter (rtty) mainly for high-speed automatic communications across ocean areas. The tty unit is equipped with a keyboard similar to a typewriter. When the operator presses a key, a sequence of signals is transmitted. At receiving stations, the signals are fed into terminal equipment that translates the sequences of signals into letters, figures, and symbols and types the messages automatically.

The rtty mode of transmission and reception is rapidly becoming more efficient and reliable for communications between ships and from ship-to-shore. Ships copy what is known as "fleet broadcast" messages on rtty. The speed at which message traffic is transmitted on rtty circuits depends on the equipment in use. Normal speed of operation is 100 words per minute, but it may be faster or slower. You may find high-speed equipment, capable of printing a line or even a page at a time, in some communications centers. The use of rtty has brought about a considerable savings in manpower.

Facsimile

Facsimile (fax) is the process used to transmit photographs, charts, and other graphic information electronically. The image to be transmitted is scanned by a photoelectric cell. Electrical changes in the cell output, corresponding to the light and dark areas being scanned, are transmitted to the receiver. At the receiver, the signal operates a recorder that reproduces the picture. The fax signals may be transmitted either by landline or radio.

Facsimile transmissions suffer distortion from all of the common sources of interference experienced with ordinary radiotelegraph and radio teletypewriter. Certain characteristics of TIF transmission make it less susceptible to complete loss of intelligence. For example, picture quality will be downgraded by any noise bursts, since facsimile recording is a continuous recording of signals coming from a receiver. However, because the machine scans material at the rate of about 100 lines per inch, each line is only 1/100th of an inch high. So you can see, if a noise burst interfers with the signal, it will distort a line only 1/100th of an inch high, leaving the image still readable. Under similar circumstances on a conventional rtty circuit, such distortion could cause a portion of the page copy to be unreadable.

Facsimile transmission is not intended to be a replacement for teletypewriter and other general methods of transmission. It is an important communications supplement and provides a means of handling certain types of graphic and pictorial intelligence by swift communications methods. It is widely used by the Navy weather information services and ship and station weather centers to obtain the latest weather maps. Chances are the photo you saw in the newspaper was transmitted by facsimile.

Q5. What is the main use of a radio teletypewriter?

Q6. What is facsimile?

SYSTEM INTRODUCTION

Until recently, RADIO COMMUNICATIONS brought to mind either telegraphy (cw), voice (AM), or possibly radio teletypewriter (rtty) communications. Today, radio communications has become a highly sophisticated field of electronics. Even small Navy ships have the capability to "come up" on the commonly used ship-to-ship, ship-to-air, and ship-to-shore communications circuits. These circuit operations are accomplished through the use of compatible and flexible communications systems.

A communications system (as you will see later in this chapter) consists of two or more equipment sets (sets will be explained a little later). Communications systems follow the system subdivision shown in figure 1-1. Systems are arranged and interconnected to perform a circuit operation that cannot be performed by any single piece of equipment. Navy communications systems vary from the simple to the very complex, depending upon the circuit operations involved. Because a Navy ship must use every inch of available space, the communications equipment may be spread over several portions of the ship, for instance, receivers in one location, transmitters in another, and terminal equipment in another. The equipment must be installed in such a manner that it is flexible and can be used interchangeably with other installed communications equipment. Consequently, large numbers of sets which make up the shipboard communications system are installed and are capable of operating separately and simultaneously. Flexibility is provided through a complex arrangement of interconnections. These allow the physically separated equipment to be selectively switched (patched) by you into different circuit configurations.

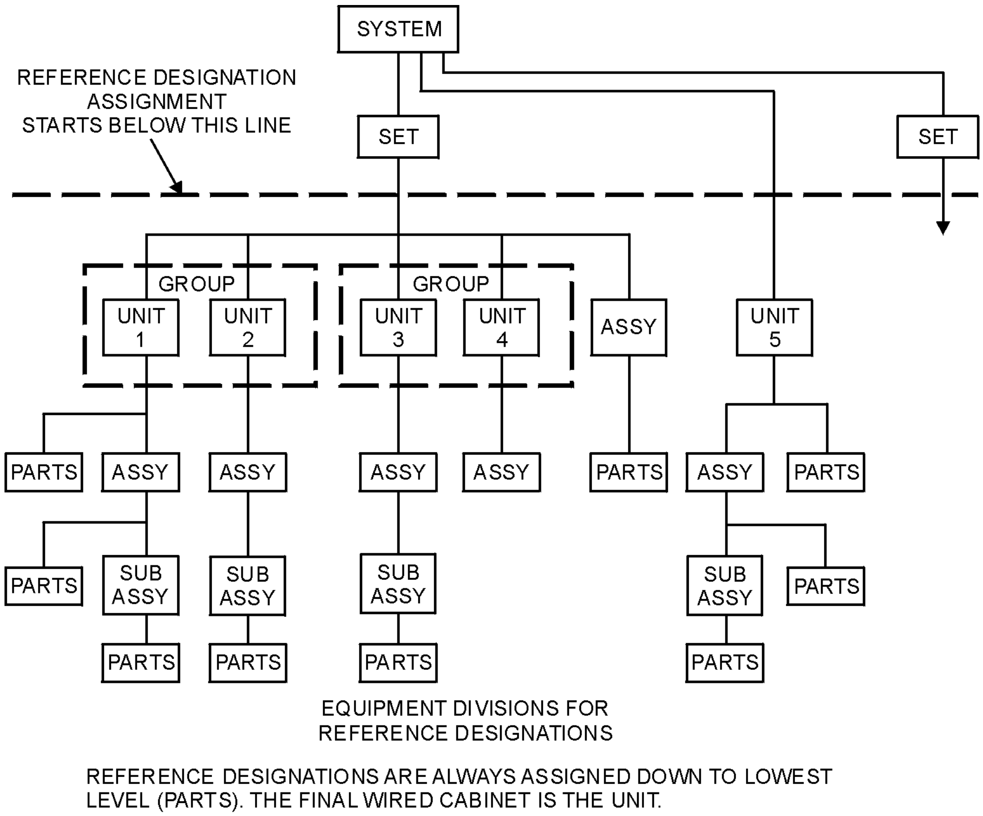


Figure 1-1.—System subdivision.

As naval electronics has grown in capability and complexity, an orderly plan of equipment designations has been adopted. The largest designator, *system*, describes equipment that work together for a specific function. For example, the radar system of a ship includes every item of electronics equipment used in or with a radar on board that ship. The smallest designator, *part*, describes one single piece, such as a bolt or a resistor.

System

A SYSTEM is a combination of *sets, units, assemblies, subassemblies*, and *parts* joined together to perform a specific operational function or several functions. Examples are communications systems, radar systems, or navigation systems. You will find the majority of troubleshooting done in the Navy is system oriented. Figure 1-2 is a pictorial view of a typical communications system containing the components necessary for transmission and reception of voice, telegraphy, and teletypewriter signals. Figure 1-3 is a block diagram of the same communications system with the arrows showing the direction of signal flow.

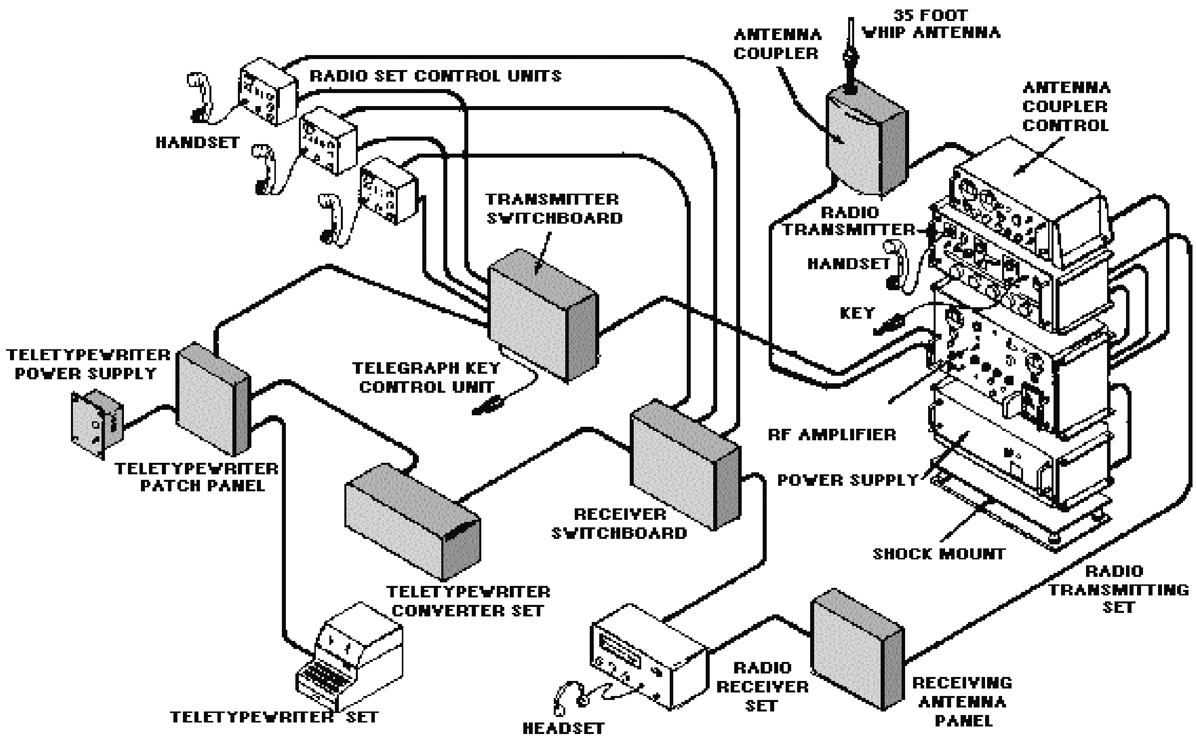


Figure 1-2.—Communications system pictorial view.

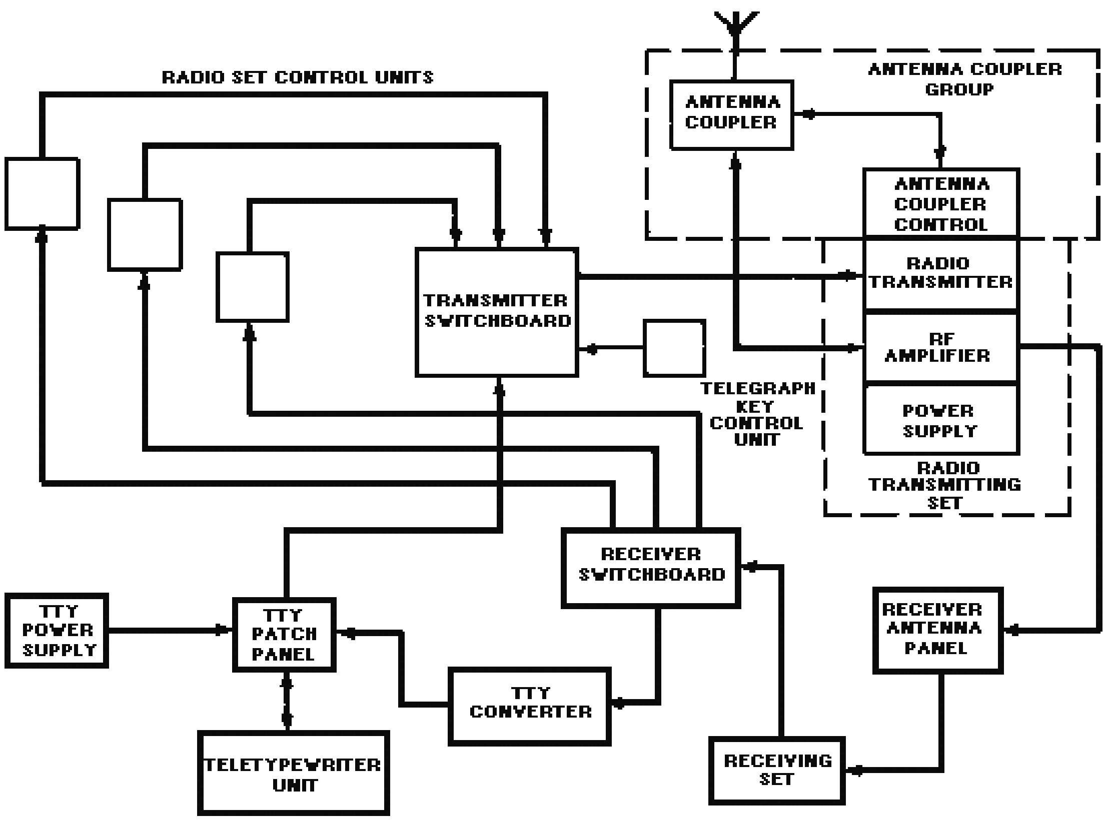


Figure 1-3.—Communications system block diagram.

Set

A SET consists of a unit or units and the assemblies, subassemblies, and parts connected or associated together to perform a specific function. A good example of this is a radio receiving set or a radio transmitting set.

Figure 1-4 is a block diagram of a radio transmitting set. It consists of a radio-frequency amplifier unit (1), a radio transmitter unit (2), a power supply unit (3), and an antenna coupler group.

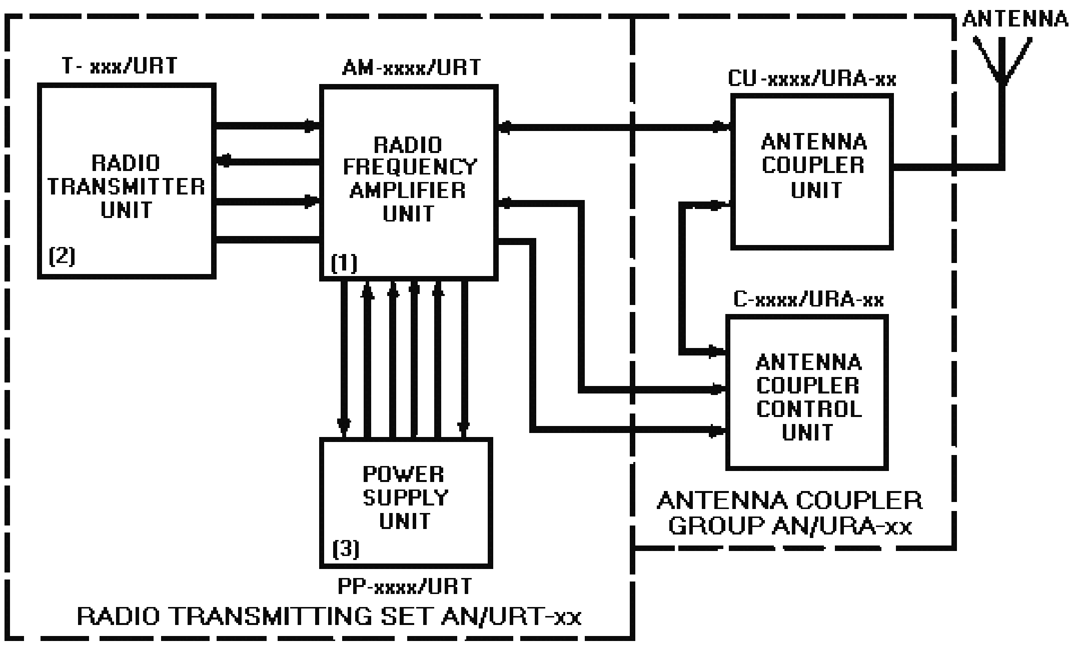


Figure 1-4.—Radio transmitting set.

Group

A GROUP is a collection of units, assemblies, subassemblies, and parts. It is a subdivision of a set or system, but it is not capable of performing a complete operational function. The coupler requires power and signals from the radiofrequency amplifier unit for operation. An example is the antenna coupler group in figure 1-4.

Unit

A UNIT is an assembly or any combination of parts, subassemblies, and assemblies mounted together. A unit is normally capable of independent operation in a variety of situations. An example of a unit might be a power supply.

Assembly

An ASSEMBLY is a number of parts or subassemblies, or any combination thereof, joined together to perform a specific function. Figure 1-5 shows a unit (2) with its six assemblies. The assembly (A6) contains six subassemblies.

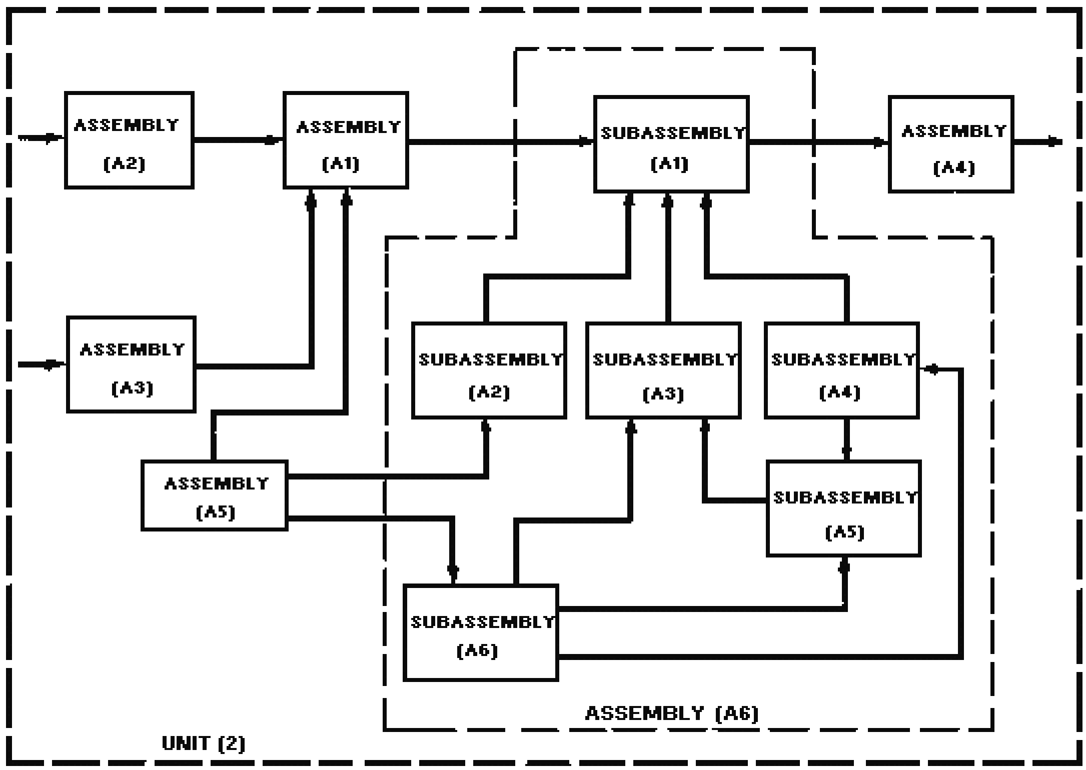


Figure 1-5.—Unit and assembly.

Subassembly

A SUBASSEMBLY consists of two or more parts that form a portion of an assembly or a unit. It is replaceable as a whole, but some of its parts are individually replaceable.

The distinction between an assembly and a subassembly is not always exact; an assembly in one application may be a subassembly in another when it forms a portion of an assembly. Figure 1-6 shows a printed circuit board subassembly and some of the parts which may be mounted on it.

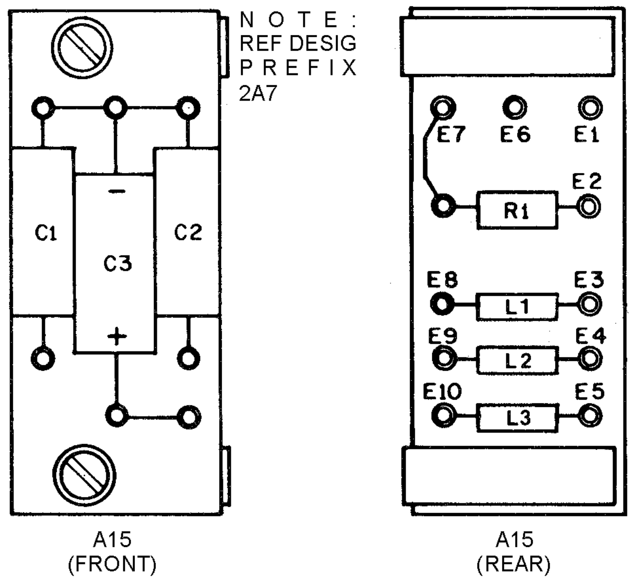


Figure 1-6.—Typical subassembly.

Part

A PART is one component or two or more components joined together. A part is not normally subject to disassembly without destruction. Resistors, capacitors, and transistors are examples of parts.

Reference Designations.

Reference designations consist of letters, numbers, or both and are used for identification purposes. Reference designations can be used in several different ways. One important way you will use them is as a cross-reference for locating supply stock numbers. When you know the reference designator, ordering the correct replacement for a failed component is easy. You will also use them frequently in corrective and preventive maintenance. Reference designators will help you to locate test points and adjustments; they will also help you to move back and forth between various technical manuals, schematics, tables, or other references.

Each set within a system is assigned an AN nomenclature. Each unit, assembly, subassembly, and part of a set has an assigned reference designation. Systems, sets, and groups have no reference designation. The unit is the highest level assigned a reference designator.

Each unit is assigned an identifying number. This number begins with the number 1 and runs consecutively for all units of a system or a set. Let's look back at the radio transmitting set AN/URT-xx with the unit numbers 1, 2, 3 on figure 1-4. You should note that these units may also have an AN nomenclature, such as T-xxx/URT. The T indicates the equipment is a transmitter. The xxx would be replaced by 3 digits that indicate the model number.

By examining the reference designator of a unit, you will be able to determine in which group, if any, the unit is contained. Let's look at a complete reference designator for a unit. A good example for us to break down is the reference designator 2A2A3C1 on figure 1-7.

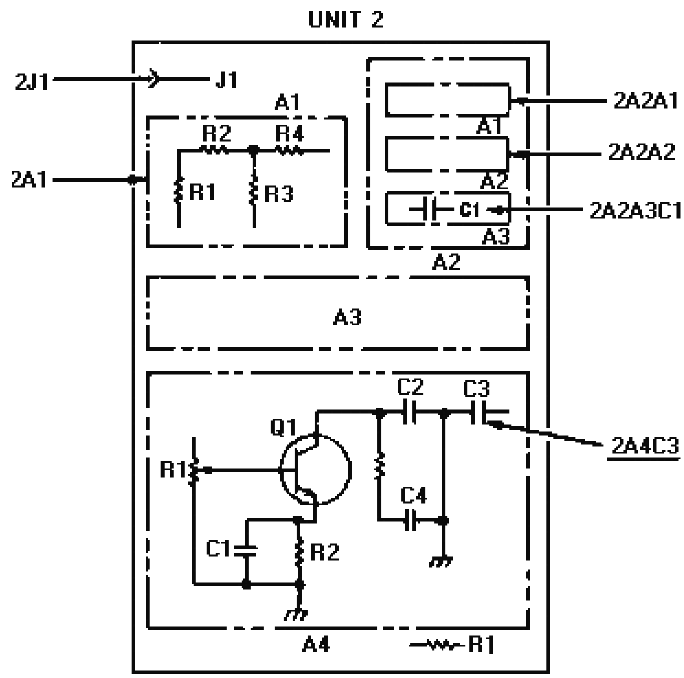


Figure 1-7.—Reference designations.

The first indicator, 2, is numeric and refers to unit 2. The next indicator, A2, is alphanumeric and refers to assembly A2. The next indicator, A3, is also alphanumeric and refers to subassembly A3. The last indicator, C1, like the two previous, is alphanumeric and refers to the part C1. We have just located capacitor C1 on subassembly A3, which is on assembly A2, which is in unit 2 of the equipment.

Reference designations may be expanded or reduced to as many levels as required to identify a particular part. Let's look at a couple of examples on our figure. The designator 2J1 identifies jack J1, which is mounted directly on unit 2. The designator 2A4C3 identifies capacitor C3, which is on assembly A4 in unit 2.

Partial reference designations are used to save space on diagrams. For example, refer back to figure 1-6. Partial reference designations are placed near the parts on subassembly A15, and a note indicates the reference designation prefix is added. Capacitor C3 on subassembly A15 has the complete reference designation 2A7A15C3.

Q7. A system is subdivided into what levels?

Q8. In the example 1A6CR3, what is the assembly designator?

NAVY FREQUENCY BAND USE

Rapid growth in the quantity and complexity of communications equipment and increased worldwide international requirements for radio frequencies have placed large demands upon the radio-frequency spectrum. These demands include military and civilian applications such as communications, location and ranging, identification, standard time and frequency transmission, and industrial, medical, and other scientific uses.

The allocation, assignment, and protection of all frequencies used by any component of the Navy are the responsibility of Commander Naval Telecommunications Command (COMNAVTELCOM). Table

1-1 shows the radio-frequency spectrum broken down into nine bands used by the military. Propagation of radio waves varies widely at different frequencies. Frequencies and equipment are chosen to meet the communications application desired. We will discuss the radio-frequency spectrum in the following paragraphs.

Table 1-1.—Radio-Frequency Spectrum

FREQUENCY	DESCRIPTION
30 GHZ - 300 GHz	extremely high frequency (EHF)
3 GHZ - 30 GHz	super high frequency (SHF)
300 MHZ - 3 GHz	ultra high frequency (UHF)
30 MHZ - 300 MHz	very high frequency (VHF)
3 MHZ - 30 MHz	high frequency (HF)
300 KHZ - 3 MHz	medium frequency (MF)
30 KHZ - 300 KHz	low frequency (LF)
3 KHZ - 30 KHz	very low frequency (VLF)
300 HZ - 3 KHz	voice frequency
Up to 300 Hz	extremely low frequency (ELF)

Extremely Low-Frequency Communications

The purpose of the EXTREMELY LOW-FREQUENCY (elf) communications system is to send short "phonetic letter spelled out" (PLSO) messages from operating authorities in the continental United States (CONUS) to submarines operating at normal mission speeds and depths. Elf has the ability to penetrate ocean depths to several hundred feet with little signal loss. This ability allows submarines to be operated well below the immediate surface and enhances submarine survivability by making detection more difficult.

This is a one-way communications system from the operating authority to submarines at sea. The large size of elf transmitters and antennas makes elf transmission from submarines impractical.

Very-Low-Frequency Communications

The communications commitments of the Navy now cover the face of the earth. New sea frontiers to the north have opened a four-million-square-mile, ice-covered ocean of strategic importance. Our Navy must maintain control of the operating forces in an ever expanding coverage area. This additional area requires changes in communications capacity, range, and reliability. Additional needs have been particularly great in the North Atlantic and the newly opened Arctic Ocean. High-frequency circuits are too unreliable in these areas because of local atmospheric disturbances.

VERY-LOW-FREQUENCY (vlf) transmissions provide a highly reliable path for communications in these northern latitudes as well as over and under all oceans and seas of the world. At present, practically all Navy vlf transmitters are used for fleet communications or navigation. The vlf transmission is normally considered a broadcast, that is, one-way transmission, no reply required. The vlf transmitter normally transmits single-channel rtty.

Vlf is currently used for communications to large numbers of satellites and as a backup to shortwave communications blacked out by nuclear activity. Our Navy depends on vlf for crucial communications during hostilities.

Secondary applications of the vlf range include worldwide transmission of standard frequency and time signals. Standard frequency and time signals with high accuracy over long distances have become increasingly important in many fields of science. It is essential for tracking space vehicles, worldwide clock synchronization and oscillator calibration, international comparisons of atomic frequency standards, radio navigational aids, astronomy, national standardizing laboratories, and communications systems.

A vlf broadcast of standard time and frequency signals has more than adequate precision for the operation of synchronous cryptographic devices, decoding devices, and single-sideband transmissions.

Low-Frequency Communications

The LOW-FREQUENCY (lf) band occupies only a very small part of the radio-frequency spectrum. This small band of frequencies has been used for communications since the advent of radio.

Low-frequency transmitting installations are characterized by their large physical size and by their high construction and maintenance costs. Another disadvantage is low-frequency signal reception being seriously hampered by atmospheric noise, particularly at low geographical latitudes. Over the years, propagation factors peculiar to the low-frequency band have resulted in their continued use for radio communications. Low-frequency waves are not so seriously affected during periods of ionospheric disturbance when communications at the high frequencies are disrupted. Because of this, the Navy has a particular interest in the application of low frequencies at northern latitudes.

The Navy's requirement to provide the best possible communications to the fleet requires operation on all frequency bands. Constant research is being done to improve existing capabilities and to use new systems and developments as they become operationally reliable.

In the past, the fleet broadcast system provided ships at sea with low-frequency communications via cw telegraph transmissions. As technology advanced, the system was converted to single-channel radio teletypewriter transmission. Today If communications is used to provide eight channels of frequency-division multiplex rtty traffic on each transmission of the fleet multichannel broadcast system.

Medium-Frequency Communications

The MEDIUM-FREQUENCY (mf) band of the radio-frequency spectrum includes the international distress frequencies (500 kilohertz and approximately 484 kilohertz). Some ships have mf equipment. If desired the distress frequencies may be monitored. When this is done the transmitter usually is kept in the standby position. Ashore, the mf receiver and transmitter equipment configuration is usually affiliated with search and rescue organizations, which are generally located near the coast.

Only the upper and lower ends of the mf band have naval use because of the commercial broadcast band (AM) extending from 535 to 1,605 kilohertz. Frequencies in the lower portion of the mf band (300 to 500 kilohertz) are used primarily for ground-wave transmission for moderately long distances over water and for moderate to short distances over land. Transmission in the upper mf band is generally limited to short-haul communications (400 miles or less).

High-Frequency Communications

The Navy began using HIGH FREQUENCIES for radio communications around World War I when only a few communications systems were operated on frequencies near 3 megahertz. When we look at the extensive present-day use of high frequencies for long-distance communications, the fact that those Navy systems were intended for very short-range communications of a few miles seems curious. The general belief at the time was that frequencies above 1.5 megahertz were useless for communications purposes.

One of the prominent features of high-frequency, long-distance communications is the variable nature of the propagation medium. (You studied this in NEETS, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*). Successful transmission of hf signals over a long distance is dependent upon refraction of radio waves by layers of the ionosphere. The height and density of these layers is formed mainly by ultraviolet radiation from the sun. They vary significantly with the time of day, season of the year, and the eleven-year cycle of sunspot activity. Because of these variations, you must generally use more than a single frequency, sometimes up to four or five, to maintain communications on a circuit.

In spite of the difficulties we encounter with hf propagation, the economic and technical advantages of using high frequencies have led to rapid expansion of the use of the hf band. Because the number of users has increased, the hf spectrum is approaching saturation.

The hf band is shared by many domestic and foreign users, and only portions scattered throughout the band are allocated to the military services. Like other agencies, Navy requirements have grown; the capacity of the Navy's assigned portion of the hf spectrum has become severely taxed. The use of single-sideband equipment and the application of independent sideband techniques have increased the capacity, but not enough to catch up with the demand. Some predict that satellite communications will eventually relieve congestion in the hf band and that, for some types of service, it will replace hf for long-distance communications. We will present more information to you concerning satellite communications in chapter 3. Even with new technology the hf spectrum most likely will continue to be in high demand for some time.

Naval communications within the hf band can be grouped into four general types of services: point-to-point, ship-to-shore, ground-to-air, and fleet broadcast. All but the fleet broadcast are normally operated with two-way communications. Some of these services involve ships and aircraft that present special problems because of their physical characteristics and mobility. Generally, the less than optimum hf performance of this shipboard equipment is at least partially offset by powerful transmitters and sensitive receiving systems at the shore terminals.

POINT-TO-POINT.—Point-to-point systems are established to communicate over long-distance trunks or links between fixed terminals. A trunk is normally a message circuit between two points that are both switching centers or individual message distribution points. A link is a transmitter-receiver system connecting two locations.

Generally, enough real estate is acquired at the terminals to permit the use of large, high-gain antennas aimed at opposite terminals of each link. This increases the effective radiated power and the sensitivity of the receiving system; it also reduces susceptibility of a circuit to interference.

With the path length and direction fixed, other propagation factors are simplified and highly reliable communications can be achieved.

SHIP-TO-SHORE.—This application of the hf band is more difficult than point-to-point since the ship is moving and constantly changing its position. In ship-to-shore the path length and direction are variable. Aboard ship, limited space and other restrictions prohibit installation of large, efficient hf antennas. Because of the mobility of ships, shipboard antennas are designed to be as nearly omnidirectional as possible.

Our problems are not as severe at the shore terminal where we have sufficient space for efficient omnidirectional antennas or arrays designed for coverage of large areas of the earth. At shore stations, rotatable, high-gain antennas or fixed, point-to-point antennas are used. For example, a rhombic antenna ashore may work well for long-haul, ship-to-shore communications when the ship is within range of the antenna.

Several frequencies are usually assigned for each circuit. Therefore, a frequency can be selected that best matches the propagation path conditions between the shore terminal and the ship.

GROUND-TO-AIR.—The use of hf radio for ground-to-air communications is similar to ship-to-shore. The only exception is an aircraft moves more rapidly than a ship. All major circuit improvements must be made at the ground station. For example, higher powered transmitters, lower noise receivers, and more efficient antennas must be used on the ground.

FLEET BROADCASTS.—As the name implies, this service involves broadcast area coverage from shore-based transmitters to ships at sea. Messages to be sent to ships are delivered by various means to the proper broadcast station. They are then broadcast for shipboard reception. To overcome propagation problems, naval communicators send the messages on several frequencies at once. This is known as frequency-diversity transmission. This type of transmission allows the ship to choose the best frequency for reception. Space-diversity with physically separated receive antennas also helps to overcome this problem.

Very-High-Frequency and Above Communications

Frequencies above 30 megahertz are not normally refracted by the atmosphere and ground-wave range is minimal. This normally limits our use of this frequency spectrum to line of sight. The exception to this is increased range through the use of tropospheric scatter techniques. Some communications using vhf and above frequencies use a technique called forward propagation by tropospheric scatter (fpts). This method will be discussed in more detail in chapter 5.

Certain atmospheric and ionospheric conditions can also cause the normal line-of-sight range to be extended. Frequencies at the lower end of this band are capable of overcoming the shielding effects of hills and structures to some degree; but as the frequency is increased, the problem becomes more pronounced. Reception is notably free from atmospheric and man-made static. (The VERY-HIGH-FREQUENCY (vhf) and ULTRAHIGH-FREQUENCY (uhf) bands are known as line-of-sight transmission bands.) Because this is line-of-sight communications, the transmitting antenna is in a direct line with the receiving antenna and not over the horizon. The line-of-sight characteristic makes the vhf band ideal for amphibious operations (beach landing from sea craft) and the uhf well suited for tactical voice transmissions (maneuvering of ships traveling together). The SUPERHIGH-FREQUENCY (shf) band is used for radar and satellite communications, whereas the EXTREMELY HIGH-FREQUENCY (ehf) band is used only in the experimental stage.

Q9. he majority of vlf transmitters are used for what purpose?

Q10. Today the Navy uses lf communications as a segment of what operational system?

Q11. Why does the Navy only use the upper and lower ends of the mf band?

Q12. What are the four general types of communications services in the hf band?

Q13. A message transmitted on several frequencies at the same time is an example of what type of transmission?

Q14. Physically separating receive antennas is an example of what technique?

Q15. When using frequencies above 30 megahertz, you are normally limited to using what range?

COMMUNICATIONS FUNDAMENTALS

Now that we have learned the Navy's fundamental use of the various frequency bands, we should look at the types of communications links and their modes of operation. The Navy uses many modes of operation; the type used is based upon the function of the circuit or network. These modes (or functions) are combined to form a communications link. We will also discuss some of the actual networks the Navy uses on a daily basis.

COMMUNICATIONS LINKS

A complex of links forms a major communications system. The naval communications system is broken down into strategic and tactical groups.

Strategic

Strategic communications are generally world-wide in nature. They are operated on a common-user (Navy, Army, Department of Defense, and so on) or special-purpose basis. A strategic system may be confined within a specified area or limited to a specific type of traffic, but the configuration is designed so that combined operations with other strategic systems are possible. As an example, we will look at the automatic voice network, automatic digital network, and the defense special security communications system later in this chapter.

Tactical

Tactical communications are usually limited to a specific area of operations and are used to direct or report the movement of specific forces. Some tactical networks are used only for operational traffic; others may be used for operational and administrative traffic. For instance, the task force, task-group, and air-control networks are ordinarily used for operational traffic. Ship-to-shore networks and broadcast networks serve both types of traffic.

Modes of Operation

Communications links have numerous modes of operation. In our discussion, a mode of operation is identified as a link or path between two or more points that is capable of providing one or more channels for the transmission of intelligence. Let's take a look at the five most common modes of operation.

SIMPLEX.—The simplex (splx) mode uses a single channel or frequency to exchange information between two or more terminals. Communications is in one direction only.

HALF DUPLEX.—The half-duplex (hdx) mode has one-way flow of information between terminals. Technical arrangements often permit transmission in either direction, but not simultaneously. This term must be qualified to show s/o (send only), r/o (receive only), or s/r (send or receive).

SEMIDUPLEX.—The semiduplex (sdx) uses an arrangement of equipment where one terminal is simplex configured and the other uses two channels or frequencies in full duplex. A clarifying example is a ship in a simplex mode terminated full duplex with a shore station. The ship may send or receive but not do both at the same time.

FULL DUPLEX.—The full-duplex (fdx) mode is a method of operation in which telecommunications between stations takes place simultaneously in both directions using two separate frequencies. In other words, a ship may send and receive different messages at the same time. The term "full duplex" is synonymous with "duplex."

BROADCAST.—Broadcast (bc) is the type of operation in which one station transmits information on one or more channels directed to more than one station and/or unit. The broadcast system has no provision for receipt or reply; however, special arrangements may require the receiving station to reply or receipt for the message at a later time by other means. Broadcasts are the primary means of delivering messages to the fleet. Since Navy units copying broadcasts are not required to receipt for messages received, they can maintain radio silence while still receiving essential messages.

Message traffic is normally sent to the fleet by three methods: broadcast, intercept, and receipt. The first two are "do not answer" methods; the third, as its name implies, requires a receipt from the addressee (addee) for each message. Broadcast and intercept methods allow the fleet to preserve radio silence, which is a great advantage from the standpoint of security. By the intercept method, a shore radio station transmits messages to another shore station that repeats them back. Ships intercept and copy all of this message traffic.

Broadcast is preferable to intercept chiefly because it is faster. It is the method by which nearly all fleet traffic is handled. It uses radiotelegraph, radiotelephone, radio teletypewriter, and facsimile.

There is some similarity between civilian and naval broadcasts. Just as commercial stations in the broadcast band transmit programs to radio receivers in the homes in their communities, Navy communications stations broadcast messages to fleet units in their particular geographic areas. The resemblance between Navy and commercial stations ceases there. Information broadcast by naval communications stations is contained in chronologically numbered messages addressed to the ships. The messages are copied by the fleet units, which check the serial numbers to ensure they have a complete file. This checks and balances system ensures the ship has not missed any of the broadcast message traffic.

Fleet broadcasts follow regular schedules. Messages are placed on the schedules in order of precedence. If a message of higher precedence is given to a transmitter station while a lower precedence message is being transmitted, the latter message may be interrupted to transmit the message of higher precedence. All ships copy all messages appearing on the broadcast schedule they are guarding.

Messages are normally transmitted on several frequencies to make sure they are received. This gives the receiving station the choice of frequency selection when considering time of day and atmospheric conditions for best reception.

Q16. The naval communications system is made up of what two groups of communications?

Q17. What are the five most prominent communications modes of operation?

SWITCHED NETWORKS

The defense communications system (DCS) is composed of all worldwide, long-haul, government-owned and leased point-to-point circuits, trunks, terminals, switching centers, control facilities, and tributaries of military departments and other defense activities. In essence the DCS combines into a single system all the elements that make up the naval communications system and the Army and Air Force equivalent.

The switched networks discussed in this section, automatic voice network, automatic secure voice communications, automatic digital network, and the defense special security communications system, are part of the DCS and are managed by the Defense Communications Agency (DCA). You should not confuse these DCS networks with the HICOM (high-command communications network) and NORATS (Navy operational radio and telephone switchboard) networks. We will discuss both of these Navy-only networks later in this chapter.

Automatic Voice Network (AUTOVON)

The DCS AUTOVON offers rapid, direct interconnection of DOD and certain other government installations through worldwide telephone exchanges. AUTOVON is a worldwide, general-purpose direct dialing telephone system. The goal of the AUTOVON system is to complete connections between two points anywhere in the world in about two seconds and to complete regular connections with push-button speed.

The AUTOVON system is made up of several installations comparable in function to commercial telephone exchanges. An installation is referred to as an AUTOVON switch, or simply switch. Within individual areas we have local command, control, and administrative voice communications systems. These systems connect into the worldwide AUTOVON through manually operated telephone switchboards or automatic dial exchanges by using direct in and out dialing.

Normal AUTOVON service allows your station to call other stations on a worldwide basis for day-to-day communications by using the telephone.

Automatic Secure Voice Communications (AUTOSEVOCOM)

Another close relative to the AUTOVON system is the AUTOSEVOCOM a worldwide, switched telephone network. It provides authorized users with a means for exchanging classified information over communications security (COMSEC) circuitry or over other approved circuitry. The system consists of both manual and automated networks within a single system.

For subscribers to the AUTOSEVOCOM network, telephone directories containing subscriber listings, general instructions for placing calls, and trouble-reporting procedures are provided.

Automatic Digital Network (AUTODIN)

The DCS AUTODIN is a fully automatic, digital system. The system converts word messages to digital form for transmission.

AUTODIN is used to furnish instantaneous, error-free, and secure communications around the world to several thousand directly connected subscriber terminals. Daily capacity of the system is about five-million average-length messages.

AUTODIN switching centers are interconnected through a network of high-frequency radio channels, submarine cables, microwave and tropospheric channels, and a variety of wire lines.

The whole concept of AUTODIN is to reduce manual handling of messages to a minimum by the use of automated equipment. This system has reduced message delivery times and delay anywhere in the world to a matter of seconds rather than minutes or hours.

Defense Special Security Communications System (DSSCS)

The defense special security communications system (DSSCS) was established for the purpose of integrating the critical intelligence communications (CRITICOMM) and the special intelligence communications (SPINTCOMM) networks into a single automated communications network. In effect, the integration of DSSCS subscribers into AUTODIN provides two separate systems within AUTODIN-one system for special intelligence (SI) message traffic and the other for the AUTODIN regular message traffic.

NAVY-ONLY NETWORKS

Some networks are used by the Navy only. As mentioned previously, these are the high command communications network (HICOM) and the Navy operational radio and telephone switchboard (NORATS) networks. Let's look at some of their functions and purposes.

High Command Communications Network (HICOM)

The HICOM network provides a voice link between the Chief of Naval Operations (CNO) and all subordinate commands ashore, afloat, and airborne. CNO is the master control station and each fleet commander in chief has an area network control station. All naval communications stations are members.

In cases where a fleet unit is suffering communications difficulties with normal channels, HICOM is used on a not-to-interfere basis to restore communications. All naval communications stations are required to guard HICOM for their respective area networks and use this system.

Navy Operational Radio and Telephone Switchboard (NORATS)

The NORATS meets our need for a connection between Navy tactical voice systems of the operating forces and the various fixed telephone services ashore. This system extends tactical voice to shore-based operational commands. NORATS provides a connecting point in the fleet center of each communication station. This point allows us to connect or patch all ship-to-shore voice circuits and designated local shore telephone systems and extensions. A combined HICOM/NORATS console exists at many naval communications stations.

Q18. What four switched networks are part of the defense communications system?

Q19. What two elements support only designated Navy requirements?

SUMMARY

Now that you have completed this chapter, a short review of what you have learned is in order. The following summary will refresh your memory of radio-frequency communications terms.

TELECOMMUNICATIONS refers to transmission, emission, or reception of signs, signals, writings, images, or sounds. This is done by visual, oral, wire, radio, or other electromagnetic means.

RADIO COMMUNICATIONS is the term describing teletypewriter, voice, telegraphic, and facsimile communications.

SYSTEM is a combination of sets, units, assemblies, subassemblies, and parts joined together to form a specific operational function or several functions.

SET is a unit or units and the assemblies, subassemblies, and parts connected or associated together to perform a specific function.

GROUP is a collection of units, assemblies, subassemblies, and parts. It is a subdivision of a set or system but is not capable of performing a complete operational function.

UNIT is an assembly or any combination of parts, subassemblies, and assemblies mounted together. Normally capable of independent operation.

ASSEMBLY is a number of parts or subassemblies, or any combination thereof, joined together to perform a specific function.

SUBASSEMBLY consists of two or more parts that form a portion of an assembly or a unit.

PART is one component or two or more components joined together. It is not normally subject to disassembly without destruction.

EXTREMELY LOW FREQUENCY is the band of frequencies up to 300 hertz.

VERY LOW FREQUENCY is the band of frequencies from 3 kilohertz to 30 kilohertz.

LOW FREQUENCY is the band of frequencies from 30 kilohertz to 300 kilohertz.

MEDIUM FREQUENCY is the band of frequencies from 300 kilohertz to 3 megahertz.

HIGH FREQUENCY is the band of frequencies from 3 megahertz to 30 megahertz.

VERY HIGH FREQUENCY is the band of frequencies from 30 megahertz to 300 megahertz.

ULTRAHIGH FREQUENCY is the band of frequencies from 300 megahertz to 3 gigahertz.

SUPERHIGH FREQUENCY is the band of frequencies from 3 gigahertz to 30 gigahertz.

EXTREMELY HIGH FREQUENCY is the band of frequencies from 30 gigahertz to 300 gigahertz.

ANSWERS TO QUESTIONS Q1. THROUGH Q19.

A1. Radio and wire.

A2. Reliability.

A3. It is direct, convenient and easy to use.

A4. Static, enemy interference or a high local noise level.

A5. High speed automatic communications across ocean areas.

A6. The process used to transmit photographs, charts and other graphic information electronically.

A7. Set, group, unit, assembly, subassembly, and part.

A8. A6.

A9. Fleet communications or navigation.

A10. Fleet Multichannel Broadcast System.

A11. Due to the commercial broadcast (AM) band.

A12. Point-to-point, ship-to-shore, ground-to-air, and fleet broadcast.

A13. Frequency-diversity.

A14. Space-diversity.

A15. Line of sight.

A16. Strategic and tactical.

A17. Simplex, half-duplex, semiduplex, duplex, and broadcast.

A18. AUTOVON, AUTOSEVOCOM, AUTODIN, and DSSCS.

A19. HICOM and NORATS.

CHAPTER 2

INTRODUCTION TO COMMUNICATIONS THEORY

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Describe the four basic types of transmitters.
2. Describe the two basic types of single-sideband circuits.
3. Describe the three basic types of teletypewriter circuits.
4. List the four primary functions of a basic receiver.
5. Describe the four primary functions of a basic receiver.
6. State the four characteristics of a basic receiver.
7. Evaluate the four characteristics of a basic receiver.
8. Describe the fundamental heterodyning process.
9. Describe the basic difference between an AM and an fm receiver.
10. Describe single-sideband suppressed carrier communications.
11. State the purpose of carrier reinsertion and how it is used in single-sideband communications.
12. Describe the basic theory and functions of receiver control circuits.
13. Describe the basic frequency synthesis process.
14. Describe the basic audio reproduction process.

INTRODUCTION

In the previous chapter you learned the fundamentals of U.S. naval telecommunications and communications. Now, let's look at the equipment and systems that are used to communicate in the Navy. The fundamental equipment used to communicate are the transmitter and receiver.

Transmitters and receivers must each perform two basic functions. The transmitter must generate a radio frequency signal of sufficient power at the desired frequency. It must have some means of varying (or modulating) the basic frequency so that it can carry an intelligible signal. The receiver must select the desired frequency you want to receive and reject all unwanted frequencies. In addition, receivers must be able to amplify the weak incoming signal to overcome the losses the signal suffers in its journey through space.

Representative transmitters and their fundamental features are described for you in this module.

TRANSMITTER FUNDAMENTALS

Basic communication transmitters include continuous wave (cw), amplitude modulated (AM), frequency modulated (fm), and single sideband (ssb) types. A basic description of each of these transmitters is given in this chapter.

CONTINUOUS WAVE TRANSMITTER

The continuous wave is used principally for radiotelegraphy; that is, for the transmission of short or long pulses of rf energy to form the dots and dashes of the Morse code characters. This type of transmission is sometimes referred to as interrupted continuous wave. Cw transmission was the first type of radio communication used, and it is still used extensively for long-range communications. Two of the advantages of cw transmission are a narrow bandwidth, which requires less output power, and a degree of intelligibility that is high even under severe noise conditions. (For example, when the receiver is in the vicinity of rotating machinery or thunderstorms.)

A cw transmitter requires four essential components. These are a *generator*, *amplifier*, *keyer*, and *antenna*. We have to generate rf oscillations and have a means of amplifying these oscillations. We also need a method of turning the rf output on and off (keying) in accordance with the intelligence to be transmitted and an antenna to radiate the keyed output of the transmitter.

Let's take a look at the block diagram of a cw transmitter and its power supply in figure 2-1. The oscillator generates the rf carrier at a preset frequency and maintains it within close tolerances. The oscillator may be a self-excited type, such as an electron-coupled oscillator, or a quartz crystal type, which uses a crystal cut to vibrate at a certain frequency when electrically excited. In both types, voltage and current delivered by the oscillator are weak. The oscillator outputs must be amplified many times to be radiated any distance.

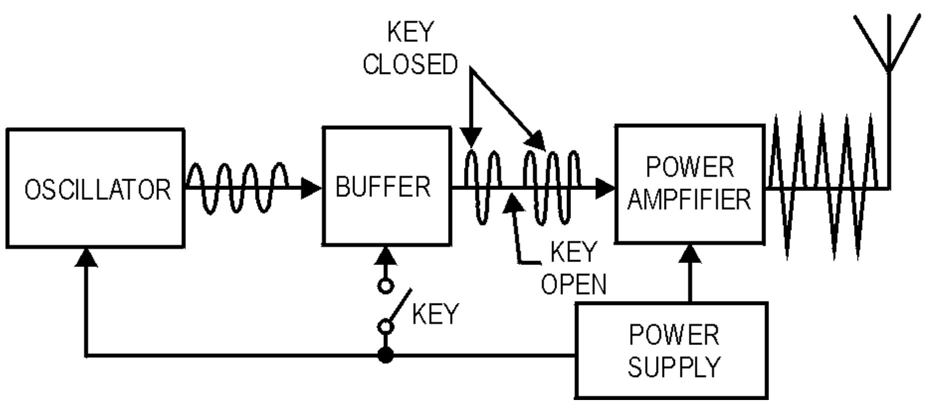


Figure 2-1.—Cw transmitter block diagram.

The buffer stage or first intermediate power amplifier stage (referred to as the ipa) is a voltage amplifier that increases the amplitude of the oscillator signal to a level that drives the power amplifier (pa). You will find the signal delivered by the buffer varies with the type of transmitter and may be hundreds or thousands of volts.

The buffer serves two other purposes. One is to isolate the oscillator from the amplifier stages. Without a buffer, changes in the amplifier caused by keying or variations in source voltage would vary the load of the oscillator and cause it to change frequency. It may also be used as a frequency multiplier, which is explained later in this text.

As you can see in the figure, a key is used to turn the buffer on and off. When the key is closed, the rf carrier passes through the buffer stage; when the key is open (buffer is turned off), the rf carrier is prevented from getting through.

The final stage of a transmitter is the power amplifier (referred to as the pa). In chapter 3 of NEETS, Module 1, *Introduction to Matter, Energy, and Direct Current*, you learned that power is the product of current and voltage (P = IE). In the power amplifier a large amount of rf current and voltage is made available for radiation by the antenna.

The power amplifier of a high-power transmitter may require far more driving power than can be supplied by an oscillator and its buffer stage. One or more low-power intermediate amplifiers are used between the buffer and the final amplifier that feeds the antenna. The main difference between many low- and high-power transmitters is in the number of intermediate power-amplifier stages used.

Figure 2-2 is a block diagram of the input and output powers for each stage of a typical medium-power transmitter. You should be able to see that the power output of a transmitter can be increased by adding amplifier stages capable of delivering the power required. In our example, the .5 watt output of the buffer is amplified in the first intermediate amplifier by a factor of 10, (this is a times 10 [× 10] amplifier) giving us an input of 5 watts to the second intermediate amplifier. You can see in this example the second intermediate amplifier multiplies the 5 watt input to it by a factor of 5 (× 5) and gives us a 25 watt input to our power (final) amplifier. The final amplifier multiplies its input by a factor of 20 (× 20) and gives us 500 watts of power out to the antenna.

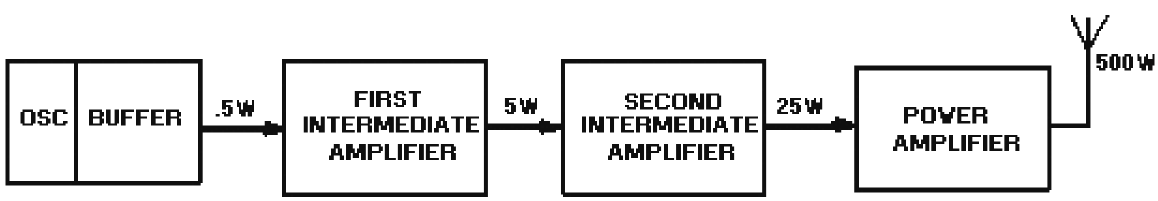


Figure 2-2.—Intermediate amplifiers increase transmitter power.

Q1. What are the four basic transmitter types?

Q2. What is the function of the oscillator in a cw transmitter?

Q3. What is the final stage of a transmitter?

AMPLITUDE MODULATED TRANSMITTER

In AM transmitters, the instantaneous amplitude of the rf output signal is varied in proportion to the modulating signal. The modulating signal may consist of many frequencies of various amplitudes and phases, such as the signals making up your own speech pattern.

Figure 2-3 gives you an idea of what the block diagram of a simple AM transmitter looks like. The oscillator, buffer amplifier, and power amplifier serve the same purpose as those in the cw transmitter. The microphone converts the audio frequency (af) input (a person’s voice) into corresponding electrical energy. The driver amplifies the audio, and the modulator further amplifies the audio signal to the amplitude necessary to fully modulate the carrier. The output of the modulator is applied to the power amplifier. The pa combines the rf carrier and the modulating signal in the power amplifier to produce the amplitude-modulated signal output for transmission. In the absence of a modulating signal, a continuous rf carrier is radiated by the antenna.

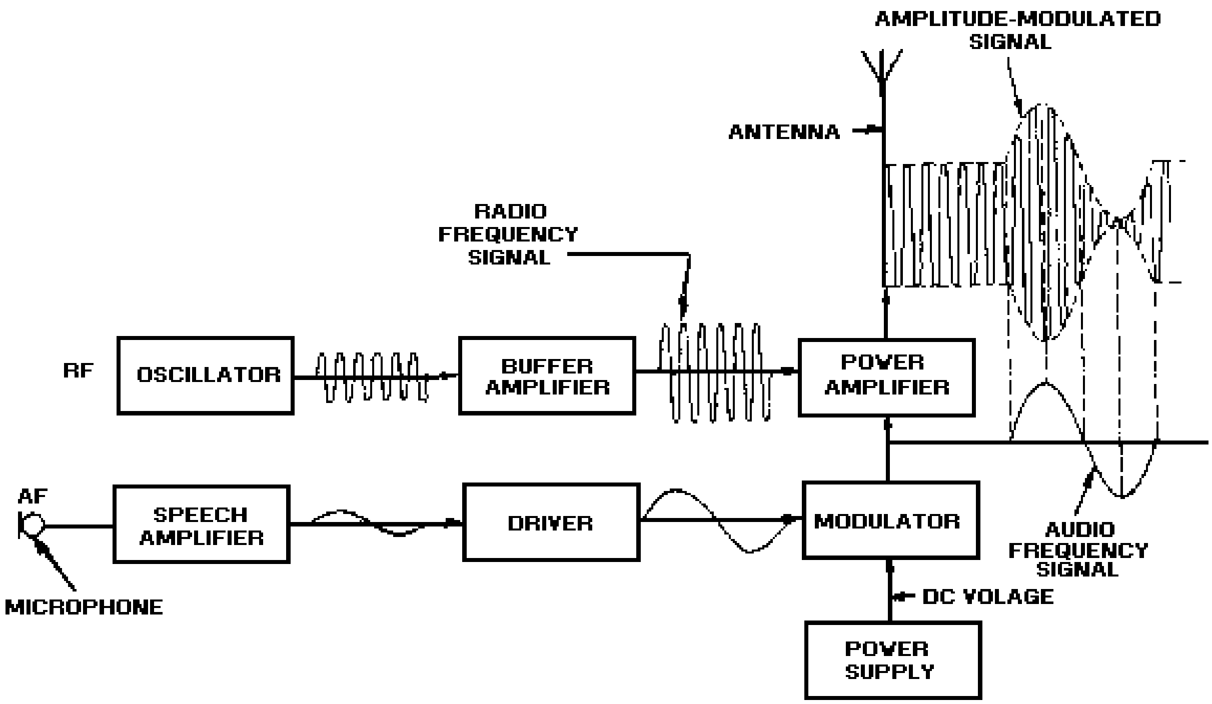


Figure 2-3.—AM radiotelephone transmitter block diagram.

FREQUENCY MODULATED TRANSMITTER

In frequency modulation (fm) the modulating signal combines with the carrier to cause the frequency of the resultant wave to vary with the instantaneous amplitude of the modulating signal.

Figure 2-4 shows you the block diagram of a frequency-modulated transmitter. The modulating signal applied to a varicap causes the reactance to vary. The varicap is connected across the tank circuit of the oscillator. With no modulation, the oscillator generates a steady center frequency. With modulation applied, the varicap causes the frequency of the oscillator to vary around the center frequency in accordance with the modulating signal. The oscillator output is then fed to a frequency multiplier to increase the frequency and then to a power amplifier to increase the amplitude to the desired level for transmission.

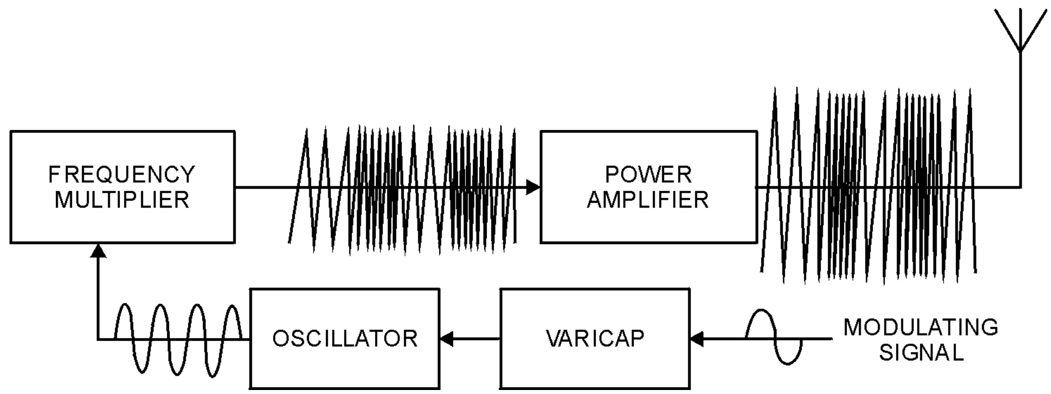


Figure 2-4.—Fm transmitter block diagram.

Harmonics

True harmonics are always exact multiples of the basic or fundamental frequency generated by an oscillator and are created in amplifiers and their associated circuits. Even harmonics are 2, 4, 6, and so on, times the fundamental; odd harmonics are 3, 5, 7, and so on, times the fundamental. If an oscillator has a fundamental frequency of 2,500 kilohertz, the harmonically related frequencies are

5,000………………	second harmonic
7,500………………	third harmonic
10,000………………	fourth harmonic
12,500………………	fifth harmonic

You should note that the basic frequency and the first harmonic are one and the same.

The series ascends indefinitely until the intensity is too weak to be detected. In general, the energy in frequencies above the third harmonic is too weak to be significant.

In some electronics books, and later in this chapter, you will find the term SUBHARMONIC used. It refers to a sine wave quantity (for example, an oscillator output) that has a frequency that is a submultiple of the frequency of some other sine wave quantity it helped make. For example, a wave that is half the fundamental frequency of another wave is called the second subharmonic of that wave; one with a third of the fundamental frequency is called a third subharmonic; and so forth.

Q4. What purpose does a microphone perform in an AM transmitter?

Q5. In an fm transmitter, when does an oscillator generate only a steady frequency?

Q6. What is a harmonic?

Q7. If the fundamental frequency is 200 megahertz, what is the third harmonic?

Frequency Multiplication

Designing and building a stable crystal oscillator is difficult. As operating frequencies increase, the crystal must be ground so thin that it often cracks while vibrating. You will find that you can get around this problem by operating the oscillators in most transmitters at comparatively low frequencies, sometimes as low as 1/100 (.01) of the output frequency. You raise the oscillator frequency to the required output frequency by passing it through one or more frequency multipliers. Frequency multipliers are special power amplifiers that multiply the input frequency. Stages that multiply the frequency by 2 are called doublers; those that multiply by 3 are triplers; and those multiplying by 4 are quadruplers.

You will find the main difference between low-frequency and high-frequency transmitters is the number of frequency-multiplying stages used. Figure 2-5 shows the block diagram of the frequency-multiplying stages of a typical Navy uhf/vhf transmitter. The oscillator in this transmitter is tunable from 18 megahertz to 32 megahertz. You have multiplier stages that increase the oscillator frequency by a factor of 12 through successive multiplications of 2, 2, and 3.

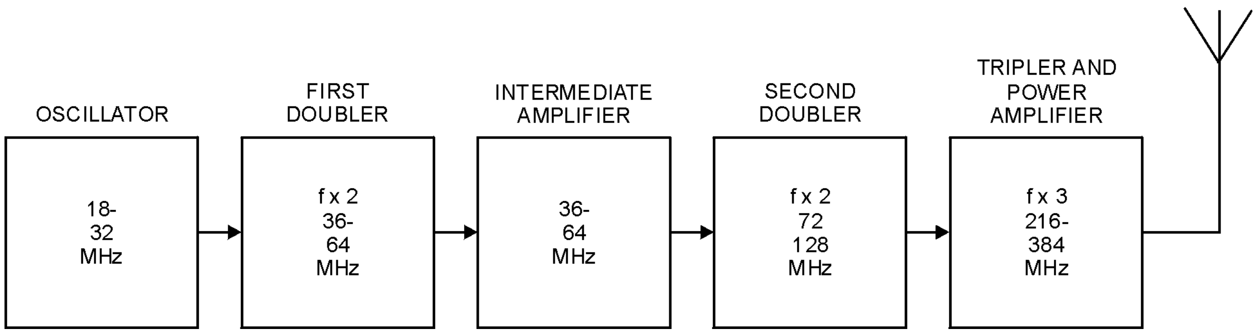


Figure 2-5.—Frequency multiplying stages of a typical vhf/uhf transmitter.

Figure 2-6 is a block diagram of an fm transmitter showing waveforms found at various test points. In high-power applications you often find one or more intermediate amplifiers added between the second doubler and the final power amplifier.

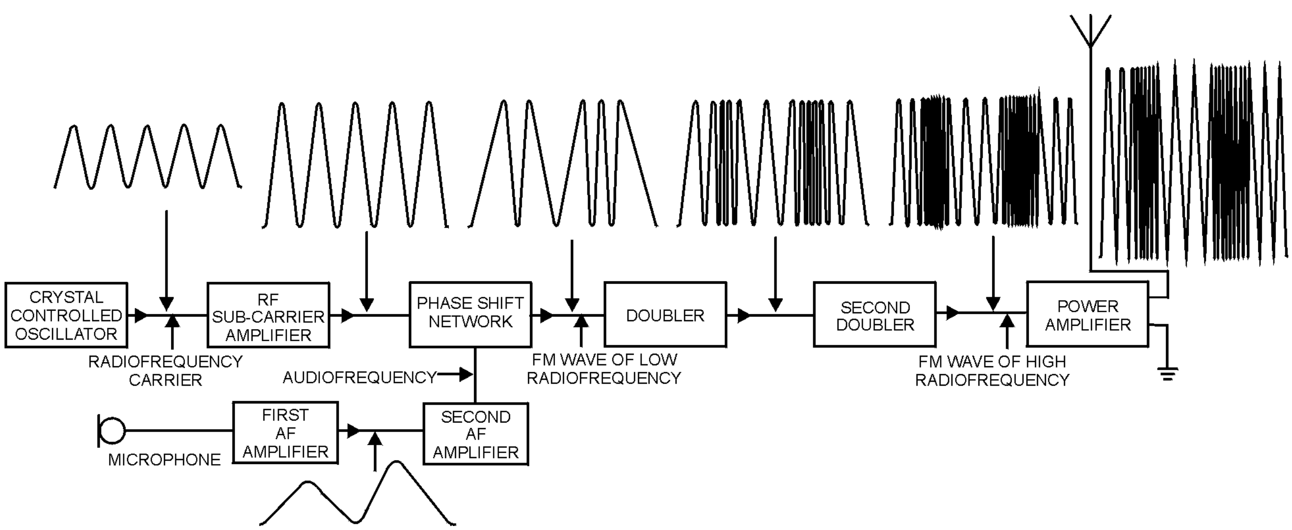


Figure 2-6.—Block diagram of an fm transmitter and waveforms.

SINGLE-SIDEBAND TRANSMITTER

You should remember the properties of modulation envelopes from your study of NEETS, Module 12, *Modulation Principles*. A carrier that has been modulated by voice or music is accompanied by two identical sidebands, each carrying the same intelligence. In amplitude-modulated (AM) transmitters, the carrier and both sidebands are transmitted. In a single-sideband transmitter (ssb), only one of the sidebands, the upper or the lower, is transmitted while the remaining sideband and the carrier are suppressed. SUPPRESSION is the elimination of the undesired portions of the signal.

Figure 2-7 is the block diagram of a single-sideband transmitter. You can see the audio amplifier increases the amplitude of the incoming signal to a level adequate to operate the ssb generator. Usually the audio amplifier is just a voltage amplifier.

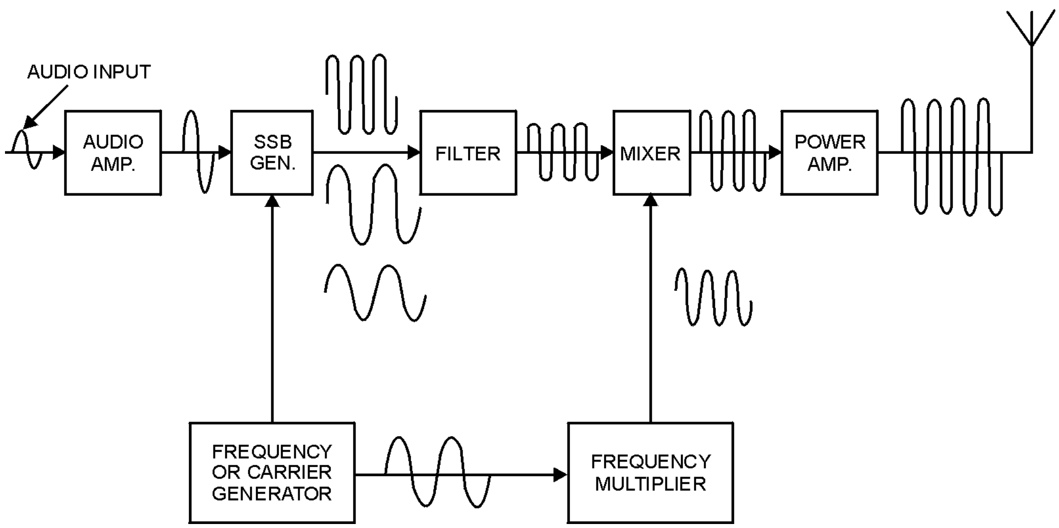


Figure 2-7.—Ssb transmitter block diagram.

The ssb generator (modulator) combines its audio input and its carrier input to produce the two sidebands. The two sidebands are then fed to a filter that selects the desired sideband and suppresses the other one. By eliminating the carrier and one of the sidebands, intelligence is transmitted at a savings in power and frequency bandwidth.

In most cases ssb generators operate at very low frequencies when compared with the normally transmitted frequencies. For that reason, we must convert (or translate) the filter output to the desired frequency. This is the purpose of the mixer stage. A second output is obtained from the frequency generator and fed to a frequency multiplier to obtain a higher carrier frequency for the mixer stage. The output from the mixer is fed to a linear power amplifier to build up the level of the signal for transmission.

Suppressed Carrier

In ssb the carrier is suppressed (or eliminated) at the transmitter, and the sideband frequencies produced by the carrier are reduced to a minimum. You will probably find this reduction (or elimination) is the most difficult aspect in the understanding of ssb. In a single-sideband suppressed carrier, no carrier is present in the transmitted signal. It is eliminated after modulation is accomplished and is reinserted at the receiver during the demodulation process. All rf energy appearing at the transmitter output is concentrated in the sideband energy as "talk power."

After the carrier is eliminated, the upper and lower sidebands remain. If one of the two sidebands is filtered out before it reaches the power amplifier stage of the transmitter, the same intelligence can be transmitted on the remaining sideband. All power is then transmitted in one sideband, instead of being divided between the carrier and both sidebands, as it is in conventional AM. This provision gives you an increase in power for the wanted sideband. You should note in figure 2-8 that the bandwidth required for the ssb suppressed carrier, view B, is approximately half that needed for conventional AM, view A. This enables us to place more signals in a smaller portion of the frequency spectrum and permits a narrower receiver bandpass.

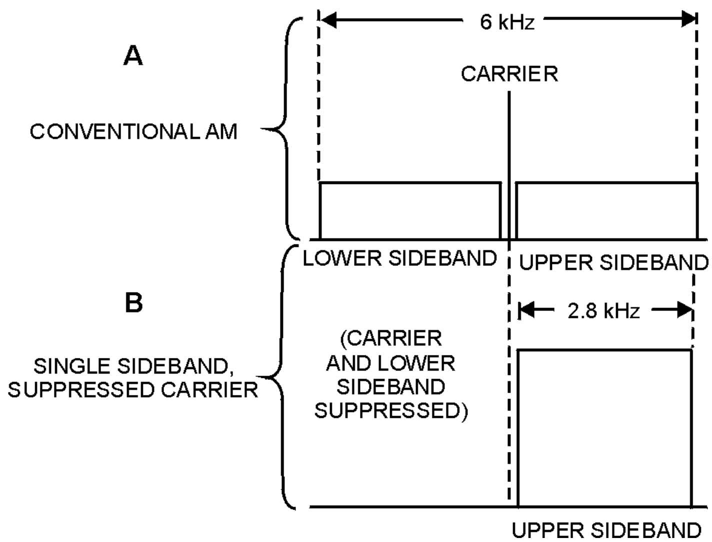


Figure 2-8.—Comparison of bandwidths of conventional AM and ssb voice channels.

Applications

Single-sideband transmission is the most common communications mode used today. Some of the ssb applications used in naval communications are described for you in the following paragraphs.

SSB VOICE CIRCUITS.—The high command (HICOM) network uses ssb as a means of communications between fleet commanders; and fleet commanders use it for communications between their subordinates and adjacent commands.

Ssb is generally used whenever special voice communications circuits are necessary between shore activities or between ships and shore activities because it is less susceptible to atmospheric interference than amplitude modulation.

SSB TELETYPEWRITER CIRCUITS.—With few exceptions, you will find ssb used on all long-haul (great distance) teletypewriter circuits, which includes ship-to-shore. Most of these systems are covered circuits; that is, an electronic cryptographic device on both ends of the circuit automatically encrypts and decrypts message traffic. These devices are used on point-to-point, ship-to-shore, ship-to-ship, and broadcast circuits.

Point-to-Point.—Most point-to-point, long-haul circuits between naval communications stations quickly use up the available frequency spectrum that ssb provides. Independent sideband (isb) transmission is normally used to compensate for the deficiency. Isb is used extensively in naval communications to expand our traffic capabilities. You will find there is a similarity between ssb and isb. Isb uses outputs from two sideband generators; it suppresses both carriers and then filters out an upper sideband from one and a lower sideband from the other. We then combine the two remaining sidebands and transmit an envelope with upper and lower sidebands that contain different intelligence. Isb can be used with MULTIPLEXING (a method for simultaneous transmission of two or more signals over a common carrier wave) to transmit a lot of intelligence on one circuit. Independent sideband and multiplexing will be discussed in more detail in chapter 3.

Ship-to-Shore.—Many ships handle enough message traffic to justify ship-to-shore teletypewriter circuits. Depending on traffic load, these circuits may contain from one to four (minimum) teletypewriter circuits on one sideband circuit. If the traffic load warrants more than one teletypewriter circuit, we usually use time division multiplex or frequency division multiplex (mux) equipment. This equipment is capable of handling many incoming and outgoing circuits. One circuit normally is used as an ORDERWIRE CIRCUIT for operator-to-operator service messages and for making frequency changes when necessary. The remaining circuits are available for handling official message traffic.

Q8. Why are frequency multipliers used?

Q9. What are two advantages of ssb transmission?

Q10. What is the purpose of an order-wire circuit?

Ship-to-Ship.—Ship-to-ship ssb teletypewriter circuits are in wide use today. Their main application is with task force or task group networks or with several ships in company. By using this type of network, ships can send their outgoing messages to another ship that relays traffic ashore. You can see this procedure saves manpower and circuit time, prevents individual ships from overcrowding ship-to-shore circuits, and conserves the frequency spectrum. Depending on the number and types of ships in company, the guard can be shifted to other ships from time to time. A major advantage of these circuits is that electronic cryptographic devices can be used to send classified messages without need for manual encryption. These circuits are used for incoming as well as outgoing traffic, and they can use either hf or uhf communications equipment.

RECEIVER FUNDAMENTALS

An AM receiver processes amplitude-modulated signals received by its antenna. It delivers an output that is a reproduction of the signal that originally modulated the rf carrier at the transmitter. The signal can then be applied to some reproducing device, such as a loudspeaker, or to a terminal device, such as a teletypewriter. Actual AM receivers vary widely in complexity. Some are very simple; others contain a large number of complex circuits.

FUNCTIONS

Whatever its degree of sophistication, a receiver must perform certain basic functions to be useful. These functions, in order of their performance, are *reception, selection, detection,* and *reproduction*.

Reception

Reception occurs when a transmitted electromagnetic wave passes through the receiver antenna and induces a voltage in the antenna.

Selection

Selection is the ability of the receiver to select a particular frequency of a station from all other station frequencies appearing at the antenna of the receiver.

Detection

Detection is the action of separating the low (audio) frequency intelligence from the high (radio) frequency carrier. A detector circuit is used to accomplish this action.

Reproduction

Reproduction is the action of converting the electrical signals to sound waves, which can then be interpreted by your ear as speech, music, and the like. An example of this might be the stereo speakers in your car.

RECEIVER CHARACTERISTICS

Sensitivity, *noise*, *selectivity*, and *fidelity* are important receiver characteristics. These characteristics will be useful to you when performing receiver tests. They can help you to determine whether a receiver is working or not or in comparing one receiver to another.

Sensitivity

The ability of a receiver to reproduce weak signals is a function of the sensitivity of a receiver. The weaker a signal that can be applied to a receiver and still produce a certain value of signal output, the better the sensitivity rating. Sensitivity of a receiver is measured under standardized conditions. It is expressed in terms of the signal voltage, usually in the microvolts that must be applied to the antenna input terminals to give an established level of the output. The output may be an ac or dc voltage measured at the detector output or a power measurement (measured in decibels or watts) at the loudspeaker or headphone terminals.

Noise

All receivers generate a certain amount of noise, which you must take into account when measuring sensitivity. Receiver noise may originate from the atmosphere (lightning) or from internal components (transistors, tubes). Noise is the limiting factor of sensitivity. You will find sensitivity is the value of input carrier voltage (in microvolts) that must be applied from the signal generator to the receiver input to develop a specified output power.

Selectivity

Selectivity is the degree of distinction made by the receiver between the desired signal and unwanted signals. You will find the better the ability of the receiver to reject unwanted signals, the better its selectivity. The degree of selection is determined by the sharpness of resonance to which the frequency-determining circuits have been engineered and tuned. You usually measure selectivity by taking a series of sensitivity readings. As you take the readings, you step the input signal along a band of frequencies above and below the circuit resonance of the receiver; for example, 100 kilohertz below to 100 kilohertz above the tuned frequency. As you approach the tuned frequency, the input level required to maintain a given output level will fall. As you pass the tuned frequency, the required input level will rise. Input voltage levels are then compared with frequency. They can be plotted on paper or you might view them on an oscilloscope. They would appear in the form of a response curve. The steepness of the response curve at the tuned frequency indicates the selectivity of the receiver.

Fidelity

The fidelity of a receiver is its ability to accurately reproduce, in its output, the signal that appears at its input. You will usually find the broader the band passed by frequency selection circuits, the greater your fidelity. You may measure fidelity by modulating an input frequency with a series of audio frequencies; you then plot the output measurements at each step against the audio input frequencies. The resulting curve will show the limits of reproduction.

You should remember that good selectivity requires that a receiver pass a narrow frequency band. Good fidelity requires that the receiver pass a broader band to amplify the outermost frequencies of the sidebands. Receivers you find in general use are a compromise between good selectivity and high fidelity.

Q11. What four basic functions must a receiver perform?

Q12. What are the four basic receiver characteristics?

SUPERHETERODYNE RECEIVER

The superheterodyne is the type receiver most familiar to you. You probably see one daily in your home in the form of an AM and/or fm radio. We will discuss the basic workings of both AM and fm types and their differences.

Amplitude Modulation Receiver

Figure 2-9 shows a block diagram with waveforms of a typical AM superheterodyne receiver developed to overcome the disadvantages of earlier type receivers. Let's assume you are tuning the receiver. When doing this you are actually changing the frequency to which the rf amplifier is tuned. The rf carrier comes in from the antenna and is applied to the rf amplifier. The output of the amplifier is an amplified carrier and is sent to the mixer. The mixer also receives an input from the local oscillator. These two signals are beat together to obtain the IF through the process of heterodyning. (Heterodyning will be further discussed later in this chapter and was covered in NEETS, Module 12, *Modulation Principles.*) At this time you should note the dotted lines connecting the local oscillator, rf amplifier, and the mixer. This is used on block diagrams and schematics to indicate GANGED TUNING. Ganged tuning is the process used to tune two or more circuits with a single control. In our example, when you change the frequency of the receiver all three stages change by the same amount. There is a fixed difference in frequency between the local oscillator and the rf amplifier at all times. This difference in frequency is the IF. This fixed difference and ganged tuning ensures a constant IF over the frequency range of the receiver.

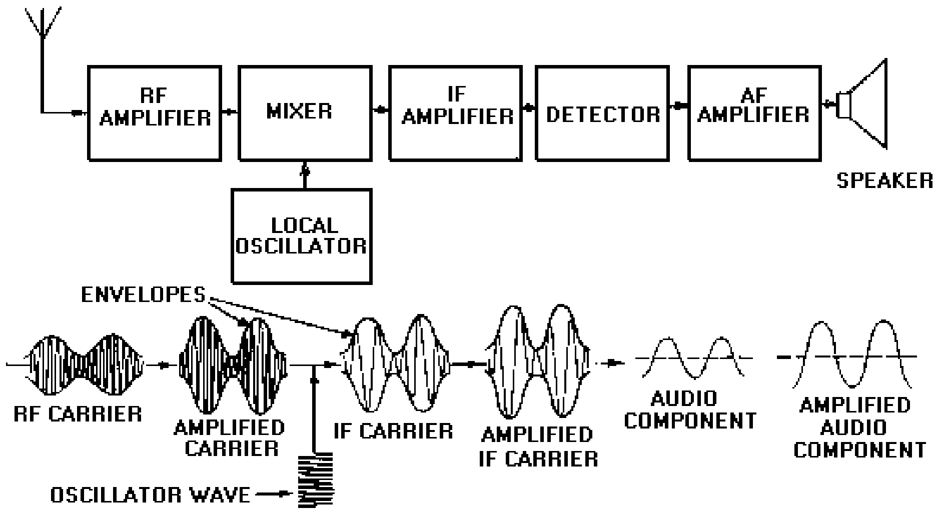


Figure 2-9.—AM superheterodyne receiver and waveforms.

The IF carrier is applied to the IF amplifier. The amplified IF carrier is then sent to the detector. The output of the detector is the audio component of the input signal. This audio component is then passed through an audio frequency amplifier. The amplified audio component is sent to a speaker for reproduction. This allows you to hear the signal.

You should note that a superheterodyne receiver may have more than one frequency-converting stage and as many amplifiers as needed to obtain the desired power output. (Additional amplifiers are not shown.)

HETERODYNING.—As you know the intermediate frequency is developed by a process called heterodyning. This action takes place in the mixer stage (sometimes called a converter or first detector). Heterodyning is the combining of the incoming signal with the local oscillator signal. When heterodyning the incoming signal and the local oscillator signal in the mixer stage, four frequencies are produced. They are the two basic input frequencies and the sum and the difference of those two frequencies. The amplifier that follows (IF amplifier) will be tuned to the difference frequency. This difference frequency is known as the intermediate frequency (IF). A typical value of IF for an AM communications receiver is 455 kilohertz. The difference frequency is a lower frequency than either the rf input or oscillator frequencies. This lower frequency gives slightly better gain but does increase the chances of image frequency interference. Image frequencies will be discussed later in this chapter.

DETECTION.—Once the IF stages have amplified the intermediate frequency to a sufficient level, it is fed to the detector. When the mixer is referred to as the first detector, this stage would be called the second detector. The detector extracts the modulating audio signal. The detector stage consists of a rectifying device and filter, which respond only to the amplitude variations of the IF signal. This develops an output voltage varying at an audio-frequency rate. The output from the detector is further amplified in the audio amplifier and is used to drive a speaker or earphones.

Frequency Modulated Receiver

The function of a frequency-modulated receiver is the same as that of an AM superheterodyne receiver. You will find some important differences in component construction and circuit design caused by differences in the modulating technique. Figure 2-10 is a block diagram showing waveforms of a typical fm superheterodyne receiver. Comparison of block diagrams in figures 2-9 and 2-10 shows that in both AM and fm receivers, the amplitude of the incoming signal is increased in the rf stages. The mixer combines the incoming rf with the local oscillator signal to produce the intermediate frequency, which is then amplified by one or more IF amplifier stages. You should note that the fm receiver has a wide-band IF amplifier. The bandwidth for any type of modulation must be wide enough to receive and pass all the side-frequency components of the modulated signal without distortion. The IF amplifier in an fm receiver must have a broader bandpass than an AM receiver.

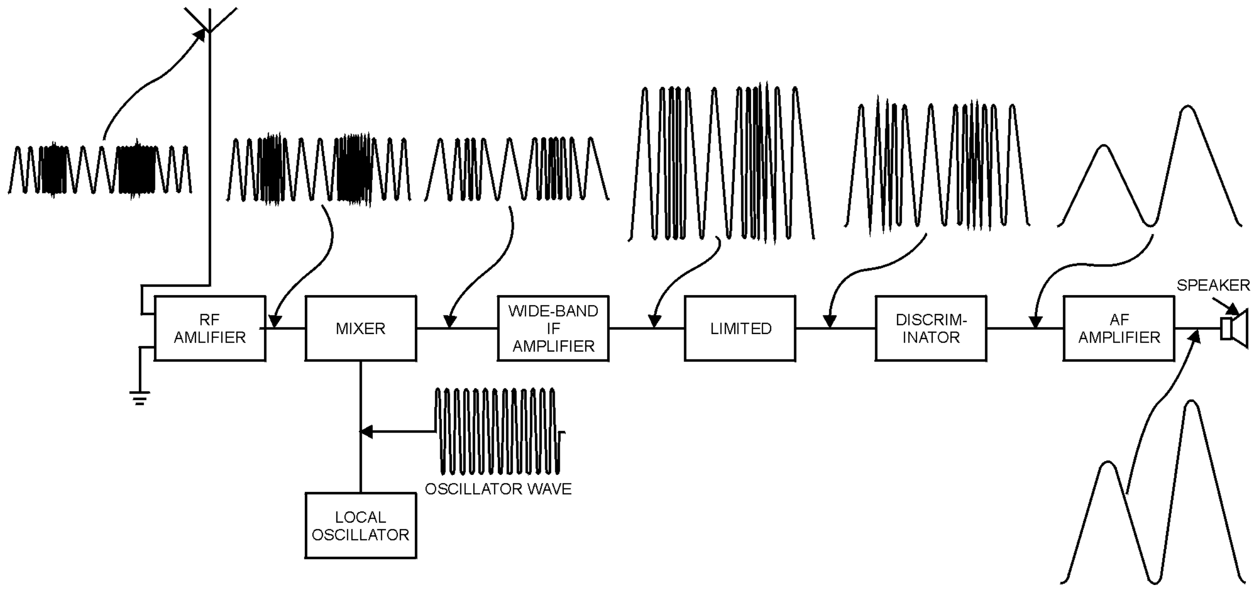


Figure 2-10.—Block diagram of an fm receiver and waveforms.

Sidebands created by fm differ from the AM system. You should recall that the AM system consists of a single set of side frequencies for each radio-frequency signal modulated. An fm signal inherently occupies a wider bandwidth than AM because the number of extra sidebands that occur in an fm transmission is directly related to the amplitude and frequency of the audio signal.

You should observe that only two fundamental sections of the fm receiver are electrically different from the AM receiver. These are the discriminator (detector) and the limiter.

Beyond the IF stage, the two receivers have a marked difference. AM demodulation involves the detection of variations in the amplitude of the signal; fm demodulation is the process of detecting variations in the frequency of the signal. In fm receivers a DISCRIMINATOR is a circuit designed to respond to frequency shift variations. A discriminator is preceded by a LIMITER circuit, which limits all signals to the same amplitude level to minimize noise interference. The audio frequency component is then extracted by the discriminator, amplified in the af amplifier, and used to drive the speaker.

ADVANTAGES.—In normal reception, fm signals are almost totally absent of static while AM signals are subject to cracking noises and whistles. Fm followed AM in development and has the advantage of operating at a higher frequency where a greater amount of frequencies are available. Fm signals provide much more realistic sound reproduction because of an increase in the number of sidebands. This increase in the number of sidebands allows more of the original audio signal to be transmitted and, therefore, a greater range of frequencies for you to hear.

As you can see, fm requires a wide bandpass to transmit signals. Each transmitting station must be assigned a wide band in the fm frequency spectrum. During fm transmissions, the number of significant sidebands that must be transmitted to obtain the desired fidelity is related to the deviation (change in carrier' frequency) divided by the highest audio frequency to be used. At this point you may want to review chapter 2 of NEETS, Module 12, *Modulation Principles*. For example, if the deviation is 40 kilohertz and the highest audio frequency is 10 kilohertz, the modulation index is figured as shown below:

$$\frac{40 \text{ kilohertz}}{10 \text{ kilohertz}} = 4$$

In this example, a modulation index of 4 equates to 14 significant sidebands. Because the audio frequency is 10 kilohertz and there are 14 side-bands, the bandwidth must accommodate a 140-kilohertz signal. You can see this is considerably wider than the 10-to-15-kilohertz bandpass used in AM transmitting.

FREQUENCY CONVERSION.—Frequency conversion is accomplished by using the heterodyne principle of beating two frequencies together to get an intermediate frequency. So far, you have only become familiar with single conversion; however, some receivers use double or triple conversion. These methods are sometimes referred to as double or triple heterodyning. Receivers using double or triple conversion are very selective and suppress IMAGE SIGNALS to yield sharp signal discrimination. (Image signals are undesired, modulated carrier signals that differ by twice the intermediate frequency from the frequency to which the superheterodyne receiver is tuned.) Double and triple conversion receivers also have better adjacent channel selectivity than can be realized in single conversion sets.

In military communications receivers you may sacrifice fidelity to improve selectivity. This is permitted because intelligence (voice, teletypewriter) can be carried on a fairly narrow band of frequencies. Entertainment receivers, on the other hand, must reproduce a wider band of frequencies to achieve their high-fidelity objective.

Q13. What frequency conversion principle is used to develop the IF?

Q14. What is the function of the detector?

Q15. What is the major disadvantage of an fm signal as compared to an AM signal?

SINGLE-SIDEBAND

You know from studying the single-sideband transmitter material in this chapter you may transmit only one sideband of an AM signal and retain the information transmitted. Now you will see how a single-sideband signal is received.

Advantages

Figure 2-11 illustrates the transmitted signal for both AM and ssb. Ssb communications has several advantages. When you eliminate the carrier and one sideband, all of the transmitted power is concentrated in the other sideband. Also, an ssb signal occupies a smaller portion of the frequency spectrum in comparison to the AM signal. This gives us two advantages, narrower receiver bandpass and the ability to place more signals in a small portion of the frequency spectrum.

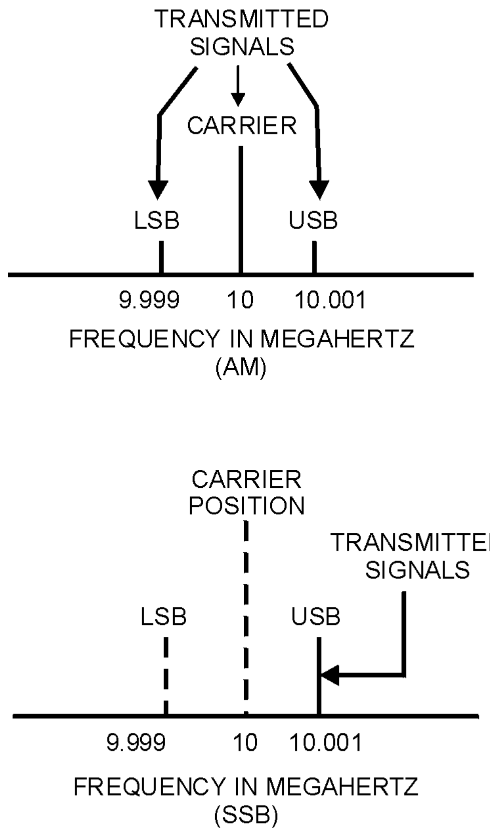


Figure 2-11.—Comparison of AM and ssb transmitted signals.

Ssb communications systems have some drawbacks. The process of producing an ssb signal is somewhat more complicated than simple amplitude modulation, and frequency stability is much more critical in ssb communication. While we don't have the annoyance of heterodyning from adjacent signals, a weak ssb signal is sometimes completely masked or hidden from the receiving station by a stronger signal. Also, a carrier of proper frequency and amplitude must be reinserted at the receiver because of the direct relationship between the carrier and sidebands.

Figure 2-12 is a block diagram of a basic ssb receiver. It is not significantly different from a conventional superheterodyne AM receiver. However, a special type of detector and a carrier reinsertion oscillator must be used. The carrier reinsertion oscillator must furnish a carrier to the detector circuit. The carrier must be at a frequency which corresponds almost exactly to the position of the carrier used in producing the original signal.

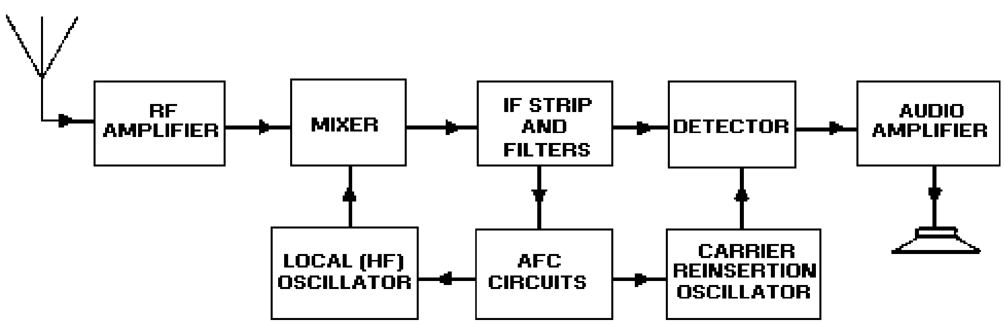


Figure 2-12.—Basic ssb receiver.

Rf amplifier sections of ssb receivers serve several purposes. Ssb signals may exist in a small portion of the frequency spectrum; therefore, filters are used to supply the selectivity necessary to adequately receive only one of them. These filters help you to reject noise and other interference.

Ssb receiver oscillators must be extremely stable. In some types of ssb data transmission, a frequency stability of ±2 hertz is required. For simple voice communications, a deviation of ±50 hertz may be tolerable.

These receivers often employ additional circuits that enhance frequency stability, improve image rejection, and provide automatic gain control (agc). However, the circuits contained in this block diagram are in all single-sideband receivers.

Carrier Reinsertion

The need for frequency stability in ssb operations is extremely critical. Even a small deviation from the correct value in local oscillator frequency will cause the IF produced by the mixer to be displaced from its correct value. In AM reception this is not too damaging, since the carrier and sidebands are all present and will all be displaced an equal amount. Therefore, the relative positions of carrier and sidebands will be retained. However, in ssb reception there is no carrier, and only one sideband is present in the incoming signal.

The carrier reinsertion oscillator frequency is set to the IF frequency that would have resulted had the carrier been present. For example, assume that a transmitter with a suppressed carrier frequency of 3 megahertz is radiating an upper sideband signal. Also assume that the intelligence consists of a 1-kilohertz tone. The transmitted sideband frequency will be 3,001 kilohertz. If the receiver has a 500-kilohertz IF, the correct local oscillator frequency is 3,500 kilohertz. The output of the mixer to the IF stages will be the difference frequency, 499 kilohertz. Therefore, the carrier reinsertion oscillator frequency will be 500 kilohertz, which will maintain the frequency relationship of the carrier to the sideband at 1 kilohertz.

Recall that 1 kilohertz is the modulating signal. If the local oscillator frequency should drift to 3,500.5 kilohertz, the IF output of the mixer will become 499.5 kilohertz. The carrier reinsertion oscillator, however, will still be operating at 500 kilohertz. This will result in an incorrect audio output of 500 hertz rather than the correct original 1-kilohertz tone. Suppose the intelligence transmitted was a complex signal, such as speech. You would then find the signal unintelligible because of the displacement of the side frequencies caused by the local oscillator deviation. The local oscillator and carrier reinsertion oscillator must be extremely stable.

Q16. What two components give a ssb receiver its advantages over an AM superheterodyne receiver?

RECEIVER CONTROL CIRCUITS

This section deals with circuits that control receiver functions. We will explain how some of the basic manual and automatic receiver control functions work.

Manual Gain Control (mgc)

You learned previously that high sensitivity is one of the desirable characteristics of a good receiver. In some cases high sensitivity may be undesirable. For example, let's suppose the signals received from a nearby station are strong enough to overload the rf sections of your receiver. This may cause the audio output to become distorted to the point of complete loss of intelligibility. To overcome this problem, you can use manual gain control of the rf section. By using the manual gain control, you can adjust the receiver for maximum sensitivity and amplify weak input signals. When you receive a strong input signal, the rf gain may be reduced to prevent overloading. A typical manual gain control circuit for a receiver is illustrated in figure 2-13. Let's go through the basic circuitry.

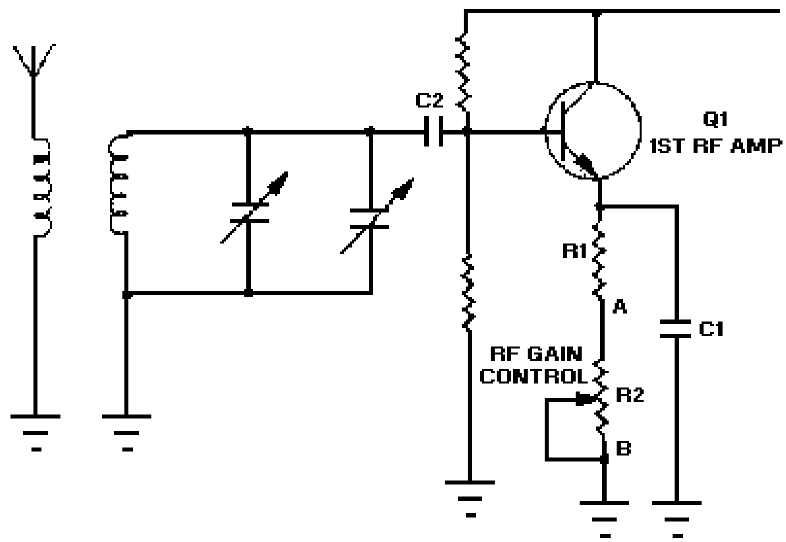


Figure 2-13.—Typical rf gain control.

C1 is an emitter bypass capacitor. Resistors R1 and R2 develop emitter bias for the amplifier. C2 provides dc isolation between the tank and the base of transistor Q1. You should recall from your studies of NEETS, Module 7, *Introduction to Solid-State Devices and Power Supplies*, and Module 8, *Introduction to Amplifiers*, that amplifier gain may be varied by changing bias. Potentiometer R2, the rf gain control, is nothing more than a manual bias adjustment. When the wiper arm of R2 is set at point B, minimum forward bias is applied to the transistor. This causes the amplifier to operate closer to cutoff and reduces gain. When you move the control toward point A, the opposite occurs. R1 limits the maximum conduction of Q1 when R2 is short circuited. You may run into an alternate biasing method when the transistor is operated near saturation. In that case, a large change in gain would again be a function of bias.

Manual Volume Control (mvc)

Figure 2-14 shows the circuitry for a common method of controlling volume in a superheterodyne receiver. C1 and R1 form an input signal coupling circuit and are also the means of controlling the level applied to the audio amplifier. R1, R2, and R3 develop forward bias and set the operating point for the transistor amplifier. R4 is the collector load resistor for Q1, and C3 is the output coupling capacitor. Potentiometer R1 in the circuit shown causes the input impedance of the stage to remain fairly constant. The signal from the preceding stage is felt across R1. By adjusting R1, you can change the input level to Q1 and vary the output amplitude.

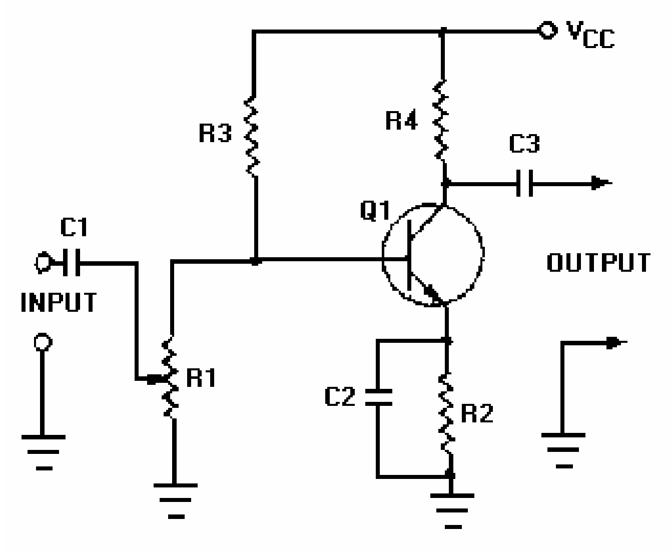


Figure 2-14.—Typical manual gain/volume control.

Automatic Gain/Volume Control (agc/avc)

Output volume variations of a receiver often result from variations in the input signal strength. Changes in input signal strength occur when we change stations or when we experience fading because of changing atmospheric conditions. The function of an AUTOMATIC GAIN CONTROL, also referred to as an AUTOMATIC VOLUME CONTROL, is to limit unwanted variations in the output of the receiver caused by variations in strength of the received signal input. A receiver without agc would require continuous manual readjustment to compensate for received signal changes so that it could maintain a constant output level.

Signals from stations operating at the same power level may not reach the receiver antenna with the same power. This is because of differences in transmission distances, carrier frequencies, atmospheric conditions, and obstructions between the transmitter and receiver antennas.

You might draw the conclusion that an agc network is not necessary when the receiver is operating on a single station. However, this is not true; atmospheric conditions may cause the signal strength to vary (fade in and out), or the antenna may receive components of the signal which have traveled along different paths. For example, one component may travel directly from the antenna, and another may have been reflected from a distant object. The two signals will sometimes be in phase and at other times be out of phase, thus tending to reinforce or cancel each other. The result is a variation in signal strength at the receiver antenna. This variation in signal strength is often referred to as FADING. The effect of fading in the output signal voltage of an rf stage is best demonstrated by the following example: An rf amplifier connected to a receiving antenna has a voltage gain of 100. If the antenna receives an input signal of 10 microvolts, the output voltage is computed as follows:

$$100 \times 10 \text{ microvolts} = 1{,}000 \text{ microvolts}$$

$$\text{or}$$

$$1 \text{ millivolt.}$$

With the output voltage equal to 1 millivolt, and if fading is to be avoided, the output voltage must remain at 1 millivolt. However, if a reflected signal is received that is approximately one-half the strength (5 microvolts) of the original and is in phase with the original signal, the total input signal to the receiving antenna will increase to 15 microvolts. To maintain the desired 1 millivolt of output signal, you must somehow reduce the gain of the rf amplifier. With an input of 15 microvolts and a desired output of 1 millivolt (1,000 microvolts), the gain of the amplifier must be reduced to:

$$\frac{1{,}000 \times 10^{-6} \text{ volts}}{15 \times 10^{-6} \text{ volts}} = 66.7$$

When the 10-microvolt original signal and the 5-microvolt reflected signal are out of phase with each other, the signal strength at the receiving antenna will decrease to 5 microvolts. If we want to maintain our original 1,000-microvolt output signal, the voltage gain of the amplifier must be increased as follows:

$$\frac{1{,}000 \times 10^{-6} \text{ volts}}{5 \times 10^{-6} \text{ volts}} = 200$$

A variation of amplifier gain, similar to the example, is necessary if we are going to compensate for input signal strength variations. The required amplifier gain variations can be accomplished automatically by the addition of an agc circuit within the receiver. Let's take a look at the methods and circuits used to produce agc and the manner in which agc (avc) controls receiver gain.

CIRCUITRY.—Figure 2-15 is a block diagram representing agc feedback to preceding stages. The detector circuit has a dc component in the output that is directly proportional to the average amplitude of the modulated carrier. The agc circuitry uses this dc component by filtering the detector output to remove the audio and IF components and by applying a portion of the dc component to the preceding stages. This agc voltage controls the amplification of any or all of the stages preceding the detector stage. Solid-state receivers may use either positive or negative voltage for agc. The type of transistors used and the elements to which the control voltage is applied determine which type we will have.

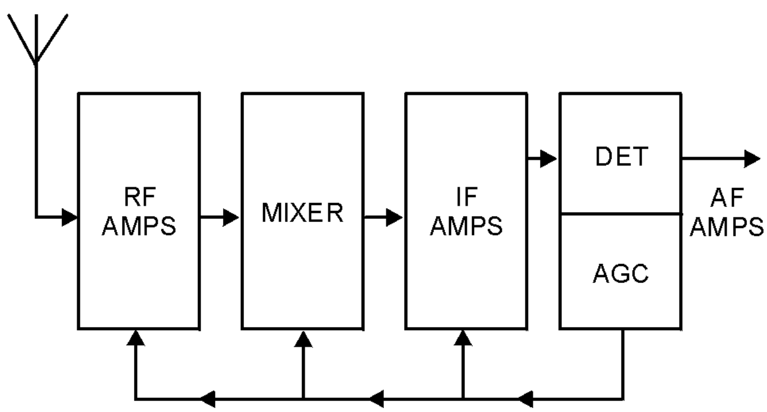


Figure 2-15.—Block diagram showing agc application.

The circuit shown in figure 2-16 produces a positive agc voltage. Transformer T1, diode CR1, capacitor C1, and resistor R1 comprise a series diode detector. The agc network is made up of R2 and C2. With normal detector operation and the positive (+) potential shown at the input, CR1 conducts. Conduction of the diode will cause a charging current (shown by the dashed line) to flow through agc capacitor C2 and agc resistor R2. This charging current develops a voltage across C2. When the potential across T1 reverses, the diode will be reverse biased and will not conduct. When this happens, the charging current ceases and C2 begins to discharge. The discharge path for C2 is shown by the solid arrows. The discharge path time constant of C2, R1, and R2 is chosen to be longer than the period (1/f) of the lowest audio frequency present in the output of the detector. Because of the longer time constant, C2 will not discharge much between peaks of the modulating signal, and the voltage across C2 will be essentially a dc voltage. This voltage is proportional to the average signal amplitude. Now, if the signal strength varies, C2 will either increase or decrease its charge, depending on whether the signal increases or decreases. Since the charge on the agc capacitor responds only to changes in the average signal level, instantaneous variations in the signal will not affect the agc voltage.

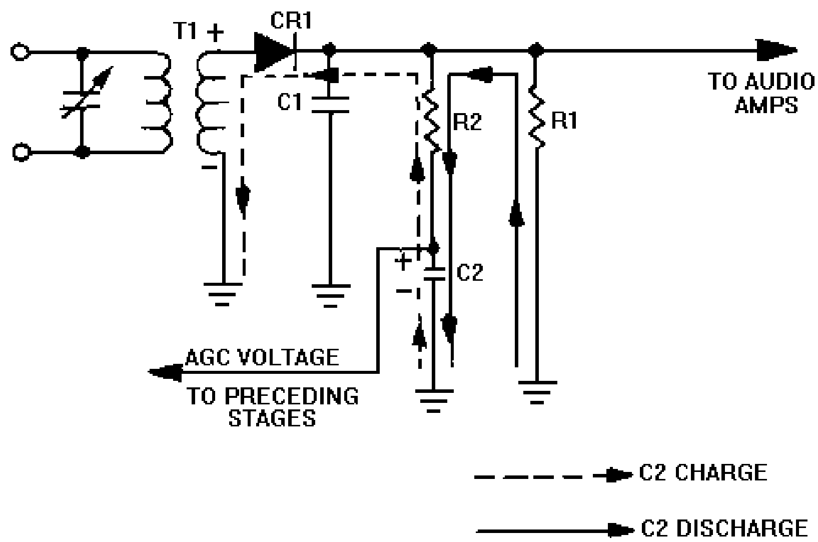


Figure 2-16.—Series diode detector and simple agc circuit.

You should remember that, depending on transistor types, the receiver may require either a positive or a negative agc voltage. A negative agc voltage could be easily obtained by reversing CR1. Once the values for R2 and C2 have been selected, the voltage divider action of the components is fixed, and the circuit operates automatically without further adjustment. If the average amplitude of the signal increases, the charge on C2 will also increase. If the signal amplitude decreases so does the charge on C2.

The agc voltages in a receiver provide controlled degenerative feedback. By adjusting the operating point of an amplifier, you can control the gain. Under no-signal conditions, bias of the rf and IF amplifiers is developed by standard means, such as self bias. With an applied signal, an agc voltage is developed, which in conjunction with normal biasing methods develops the operating bias for the amplifiers.

TRANSISTOR AMPLIFIER GAIN.—You have seen how a dc voltage that is obtained at the output of the agc network is proportional to, and will reflect, the average variations of the average signal level. Now all we have to do is use this agc voltage to control the amplification of one or more of the preceding amplifiers. Figure 2-17 illustrates a common-emitter amplifier with agc applied to the base element. A change in the agc voltage will change the operating point of the transistor and the dc emitter current. In this circuit, R1 and R4 form a voltage divider and establish no-signal (forward) bias on the base. Since a pnp transistor is used, the base has a negative potential. The agc voltage from the detector is positive with respect to ground and is fed to the base through dropping resistor R2. You will find when the dc output of the detector increases (because of an increase in the average signal level) the agc voltage will become more positive. This increased positive potential is applied to the base of Q1, which decreases the forward bias of Q1 and decreases the gain of the amplifier. Agc, in this application, works with controlled degenerative feedback. Use of an npn transistor, in the same configuration, would require the agc voltage to possess a negative potential.

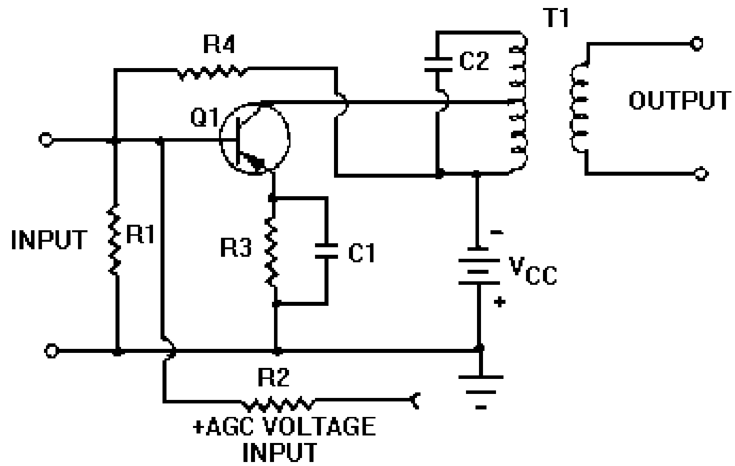


Figure 2-17.—Common emitter amplifier with agc.

FORWARD AND REVERSE AGC.—When we use an agc voltage to cause degeneration by driving the amplifiers toward cutoff, it is referred to as REVERSE agc. Figure 2-18 shows the type of agc circuitry normally used with this method. A second method that uses agc is an application called FORWARD agc. In the case of forward agc, you'll find the amplifier is driven toward the saturation region of its characteristic curve. (Sometimes referred to as an energy diagram.)

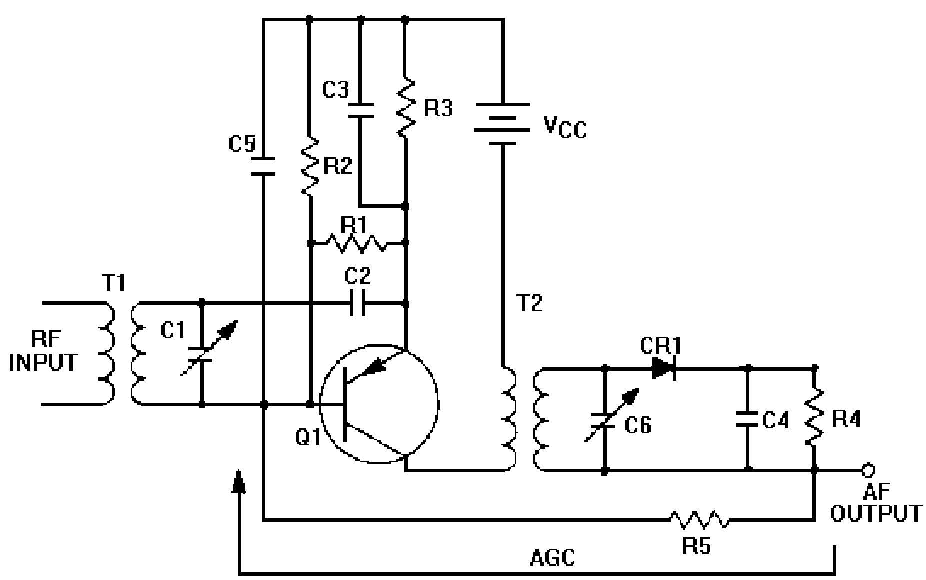


Figure 2-18.—Reverse agc.

Let’s look at our example. Assume the agc voltage is negative. Under this condition, Q1 is self biased; under no-signal conditions, it is operating well up on its characteristic curve. When a signal is applied, negative agc voltage is developed in the detector circuit and fed back to Q1, which increases forward bias. Any increase in signal level causes an increase in agc voltage. An increase in agc voltage increases conduction, which in turn drives the transistor to or near saturation. As the transistor approaches saturation, its gain is correspondingly reduced.

On the other hand, if the input signal level decreases, the negative agc voltage decreases. The forward bias is then reduced, and the transistor operates on a lower portion of its characteristic curve where gain is higher.

Forward agc provides you with better signal-handling capabilities; however, reverse agc is simpler to use, causes less loading of the tuned circuits, and produces smaller variations in input and output capacitance.

Q17. What does manual gain control do to strong and weak signals, respectively?

Q18. What is the purpose of agc/avc in a receiver?

Delayed Automatic Gain Control

The disadvantage of automatic gain control, attenuating even the weak signal, is overcome by the use of delayed automatic gain control (dagc). Let's take a look at the typical dagc circuitry in figure 2-19. This type of system develops no agc feedback until an established received signal strength is attained. For signals weaker than this value, no agc is developed. For sufficiently strong signals, the delayed agc circuit operates essentially the same as ordinary agc.

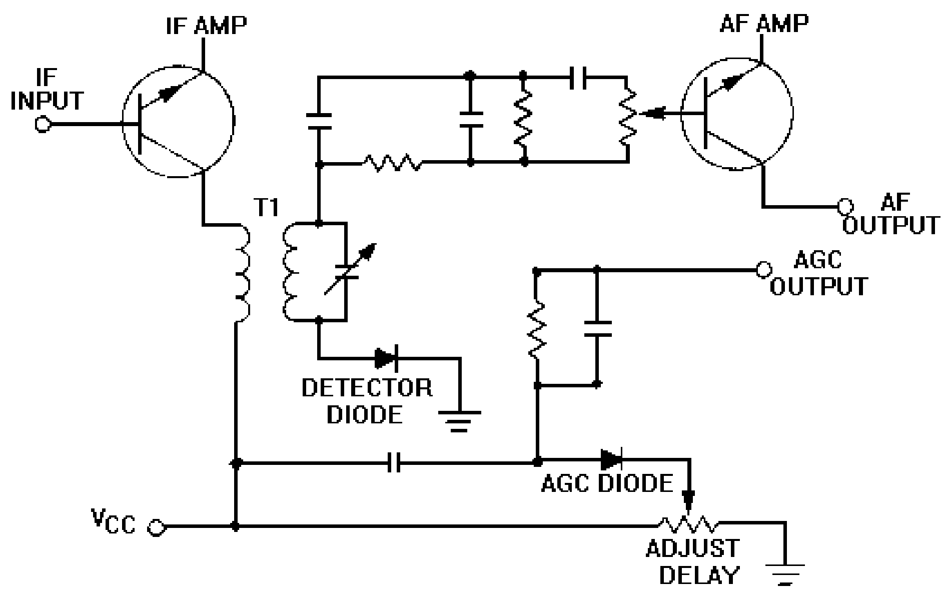


Figure 2-19.—Delayed agc action.

Our circuit uses two separate diodes; one is the detector diode and the other the agc diode. The agc diode is connected to the primary of the last IF transformer and the detector diode to its secondary. A positive bias is applied to the cathode of the agc diode. This keeps it from conducting until a prearranged signal level has been reached. The adjust delay control allows manual control of the agc diode bias. Manual control allows you to select the signal level at which agc is applied. If mostly weak stations are to be received, the setting should be high (no agc until the signal level is high). However, you should set it as low as possible to prevent overloading of the last IF amplifier by stronger signals.

Finally, you must have two diodes to obtain delayed agc. If only one diode were used, the agc would be developed from the detector diode, and there would be no delayed action. Or, if a signal diode were biased to provide the delaying action desired, no signal would pass to the audio amplifier until the bias was exceeded by the input signal.

Beat-Frequency Oscillator

The beat-frequency oscillator (bfo) is necessary when you want to receive cw signals. Cw signals are not modulated with an audio component, you remember, so we must provide one. The action of the rf amplifier, mixer, local oscillator, and IF amplifier is the same for both cw and AM; but the cw signal reaches the detector as a single frequency signal with no sideband components. To produce an af output, you must heterodyne (beat) any cw signal with an rf signal of the proper frequency. This separate signal is obtained from an oscillator known as a beat-frequency oscillator.

Figure 2-20 is a block diagram of a superheterodyne receiver capable of receiving and demodulating a cw signal. The bfo heterodynes at the detector and produces an af output. The detector (second detector) is used primarily because the mixer (first detector) is normally used as the source of agc.

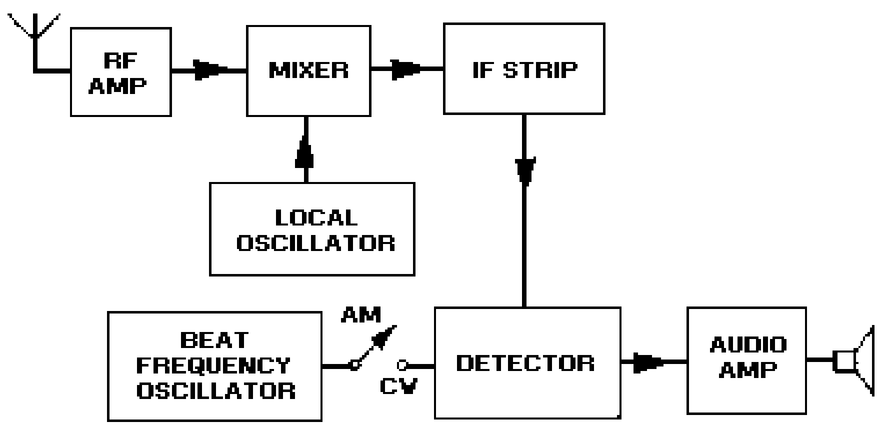


Figure 2-20.—Placement of the beat frequency oscillator.

If the intermediate frequency is 455 kilohertz and the bfo is tuned to 456 kilohertz or 454 kilohertz, the difference frequency of 1 kilohertz is heard in the output. Generally, you will tune the bfo from the front panel of a receiver. When you vary the bfo control, you are varying the output frequency of the bfo and will hear changes in the tone of the output audio signal.

Squelch

The sensitivity of a receiver is maximum when no signal is being received. This condition occurs, for example, when a receiver is being tuned between stations. At this time background noise is picked up by the antenna, and you will hear noise greatly amplified. This noise is highly annoying and occurs because receiver gain is maximum without a signal. You can often overcome this problem by using a circuit called a SQUELCH, NOISE SILENCER, NOISE SUPPRESSOR, or NOISE LIMITER. All of these noise type circuits just clip the peaks of the noise spikes. *Squelch* will actually eliminate noise. Figure 2-21 is a typical circuit of this type. The circuit cuts off receiver output when no input signal is being received. It accomplishes this by blocking either the detector or audio amplifier when no signal is present. Let's take a look at the theory involved in this process.

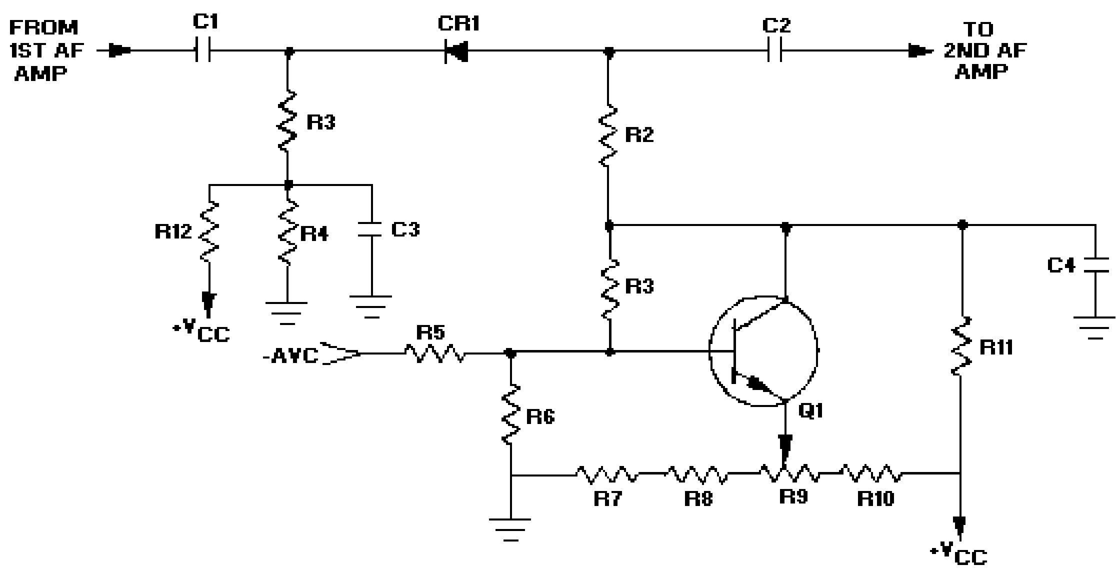


Figure 2-21.—Squelch circuit.

The squelch diode CR1 connects the output of the first af stage to the input of the second. Amplifier Q1 serves as the control transistor for the circuit. The anode and cathode voltages of CR1 are normally biased positive with respect to ground.

With no input signal, R9 is adjusted until Q1 draws enough collector current to reduce its collector voltage and the anode voltage of CR1 to a value below the voltage on the cathode of CR1. At this point the anode voltage of the squelch diode is negative with respect to its cathode, and conduction ceases. Audio output is now reduced to zero and the receiver is silent.

The base of Q1 is connected to the automatic volume control (avc) line. Anytime a signal enters the receiver, a negative avc voltage is applied to the base of Q1. This reduces the collector current and increases the collector voltage, which in turn increases the anode voltage of CR1 until the anode becomes positive with respect to the cathode. Once again diode CR1 will conduct, and the signal will be passed to the second af amplifier. Diode CR1 is effectively a switch controlled by the avc voltage.

Q19. What is a disadvantage of agc?

Q20. What is the main difference between agc and dagc?

Q21. What is the function of the bfo?

Q22. What is the purpose of a squelch circuit?

Audio Tone

The tone of the sound reproduced in the audio section of a receiver depends on several factors. The frequency response of the audio amplifiers determines the degree of amplification provided to different frequencies in the sound spectrum. The size and quality of any loudspeaker used will determine its response to various frequencies. Response of the human ear is the final judge of tonal quality, and that varies with the individual.

Because of these variables, some form of tone control is sometimes used in Navy receivers. Treble tones are defined as the audio frequencies above approximately 3,000 hertz and bass tones are those below approximately 300 hertz. Although several methods of tone control can be used, we are only going to mention the attenuation method. With this method, a decrease in the intensity of one tone can produce an apparent increase in the intensity of another tone. As an example, let's look at tones of 400 and 4,000 hertz produced by a speaker with the same intensity. If we reduce the intensity of the 4,000-hertz tone, the 400-hertz tone will appear to be louder, even though its intensity has not actually changed. You should see from this example that bass emphasis can be accomplished by attenuating treble tones.

The simplest type of tone control is illustrated in figure 2-22. Fixed capacitor C1 parallels the primary winding of the output transformer, effectively shunting the higher frequencies to ground. The size of C1 determines the lowest frequency to be affected. When you select the BASS position of the tone control, C1 is connected and improves bass response by de-emphasizing the treble tones. You can often use this circuit to improve the output of a small speaker with poor treble response.

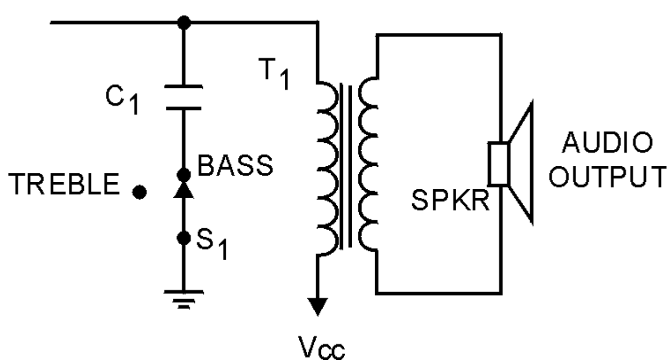


Figure 2-22.—Fixed capacitor tone control.

A continuously variable tone control is illustrated in figure 2-23. Tone control R1 and bypass capacitor C1 act as a variable RC filter. With the wiper arm of R1 in the upper (BASS) position, C1 bypasses the higher frequencies to ground and provides better bass response. When the wiper arm of R1 is in the lower (TREBLE) position, the resistance of R1 is placed in series with C1, which reduces the shunting effect of C1 to high frequencies and improves the treble response. This method gives you the advantage of smooth, continuous tone control at all points between maximum bass and maximum treble response.

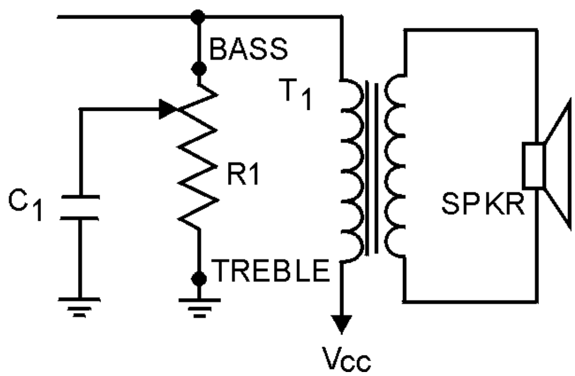


Figure 2-23.—Variable tone control using an RC filter.

A switch-type, variable tone control is illustrated in figure 2-24. With this method we use a three-position switch to provide fixed degrees of tone control. When tone control S1 is in the BASS position, capacitor C1 bypasses the high frequencies and provides bass emphasis. With S1 in the normal (NORM) position, C2 acts as the bypass, and a moderate amount of high-frequency attenuation is accomplished. This position provides balanced bass and treble response. When S1 is in the TREBLE position, C3 acts as the bypass and provides minimum high-frequency attenuation and maximum treble emphasis. As a rule of thumb, you can figure the capacitance of C2 is approximately five times the value of C3, and C1 is approximately ten times the value of C3. For example with C3 at .001 microfarads, C2 would be .005 microfarads, and C1 would be .01 microfarads.

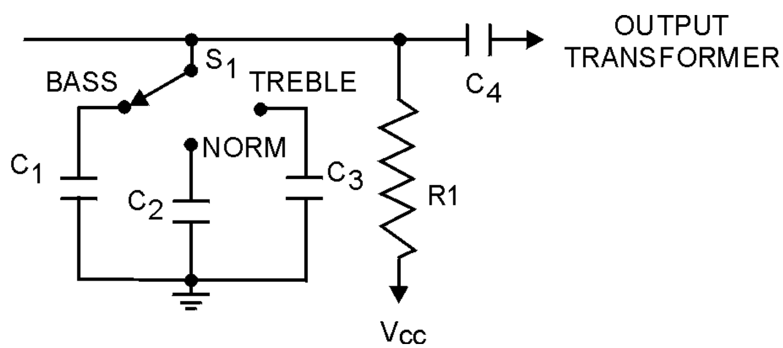


Figure 2-24.—Switch tone control.

Crystal Filters

A quartz crystal, used as a selective filter in the IF section of a communications receiver, is one of the most effective methods of achieving maximum selectivity. It is especially useful when the channel is crowded and considerable noise (both external and internal) is present.

One possible circuit arrangement is shown in figure 2-25. Let's look at the theory involved in understanding this circuit. You can see a crystal in one leg of the bridge circuit. The secondary of the input transformer (T1) is balanced to ground through the center tap connection. The crystal acts as a high Q series resonant circuit. It allows signals within the immediate vicinity of resonance to pass through the crystal to the output coil (L3). The desired signal appears between the center tap of L3 and ground.

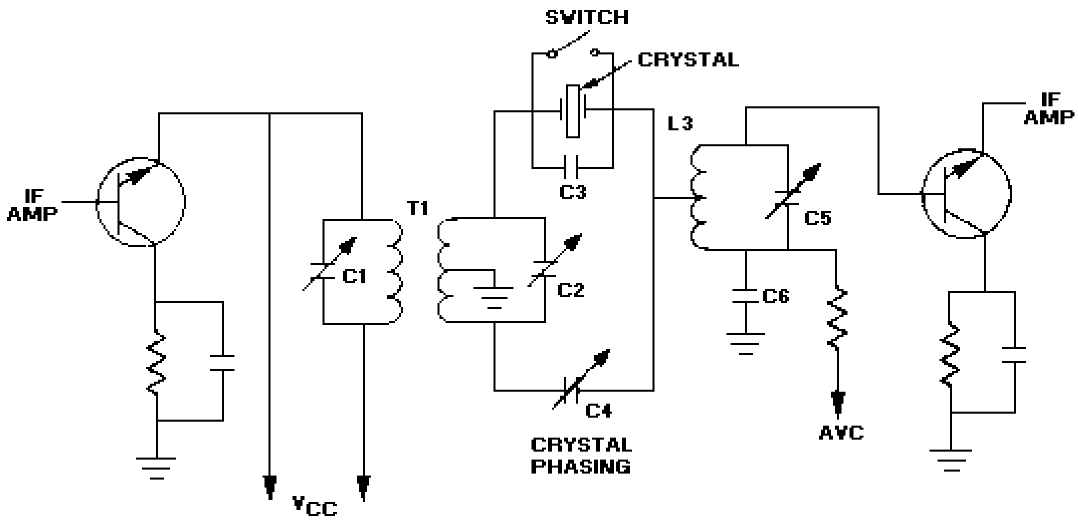


Figure 2-25.—Crystal filter used in the IF section of a superheterodyne receiver.

The capacitance that exists between the crystal holder plates (C3) may bypass unwanted signals around the crystal; therefore, some method must be provided to balance out this capacitance. You accomplish circuit balancing by applying out-of-phase voltages to neutralize any undesired signal voltages. The tap on L3 permits proper impedance matching and is the balanced input to the next stage.

Q23. What does a tone control circuit in a receiver do to the audio signal?

Q24. What is the function of a crystal filter in a receiver?

Automatic Frequency Control

Afc circuits are used in situations where you must accurately control the frequency of an oscillator by some external signal. Basically, this type circuit does two things: It senses the difference between the actual oscillator frequency and the frequency that is desired and produces a control voltage proportional to the difference; it also uses the control voltage to change the oscillator to the desired frequency. Afc circuits are used to control the frequency of sinusoidal oscillators and nonsinusoidal oscillators. Only sinusoidal afc circuits will be covered here. Afc circuits are used in radio receivers, fm transmitters, and frequency synthesizers to maintain frequency stability. Figure 2-26 is a block diagram illustrating afc operation in a receiver. Let's run through the applicable parts of this block diagram.

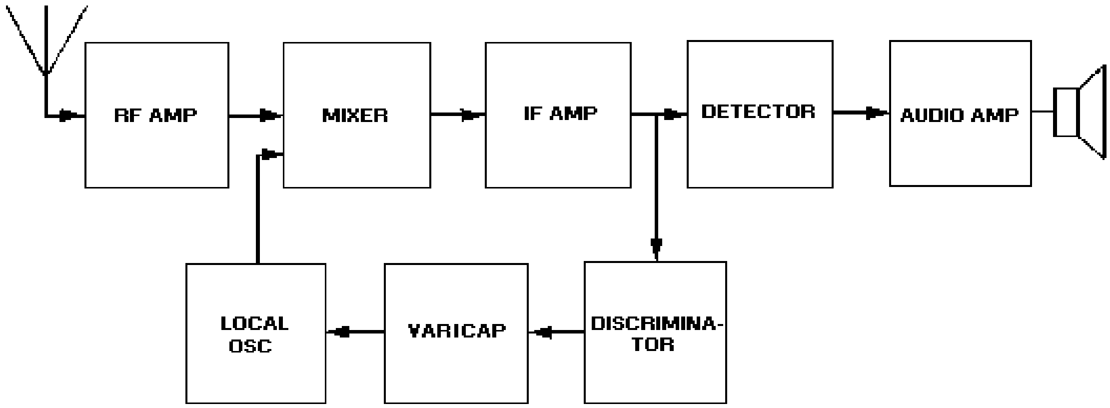


Figure 2-26.—Block diagram of receiver showing automatic frequency control.

The frequency discriminator controls the varicap in this receiver. A varicap is used to keep the IF stable. You may want to review varicap theory in chapter 3 of NEETS, Module 7, *Introduction to Solid-State Devices and Power Supplies* at this point. The varicap application here produces an apparent reactance, which is included in the oscillator frequency control circuitry. For example, let's assume the IF is 455 kilohertz and the local oscillator (lo) is tracking below the incoming station. When the lo output decreases slightly in frequency, the IF will rise. This causes the output of the discriminator to increase the capacitive reactance of the varicap, which increases the oscillator frequency to the desired value. Now let's assume the lo output increases. The IF will then decrease. This causes the discriminator output to decrease the capacitive reactance of the varicap. This will cause the oscillator frequency to decrease.

Figure 2-27 shows another widely used type of afc and its circuitry. This type is commonly referred to as a BALANCED-PHASE DETECTOR or PHASE-DISCRIMINATOR. This circuit uses fixed capacitors and the varying conductance of the diodes to achieve a variable reactance. As you have seen in the block diagram, an afc circuit requires two sections, a frequency detector and a variable reactance. Our detector output is a dc control voltage proportional to the amount of frequency change. This dc voltage is applied directly to the oscillator. The phase inverter input signals are discriminated IF outputs fed to the two diodes 180 degrees out of phase.

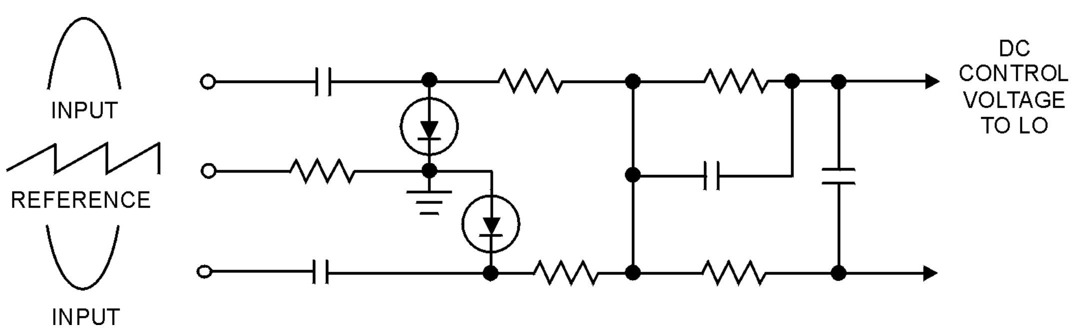


Figure 2-27.—Automatic frequency control (phase discriminator).

A reference voltage is also applied to both diodes. The diodes are biased to conduct only during the peak portions of the input signals. Any change in oscillator frequency will alter the phase relationship between the sawtooth reference voltage and the incoming signals. If this happens, one diode will conduct more than the other and produce a control signal. This system remains unbalanced at all times because any change in frequency is instantaneously corrected. The network between the diodes and oscillator is essentially a low-pass filter. This filter prevents discriminator pulses from reaching the oscillator.

FREQUENCY SYNTHESIS

In present day communications systems, long term accuracy of one part in a million is required from many of the frequency generators (local oscillators) used in communications equipment. Variable frequency oscillators cannot practically achieve this high degree of stability. Therefore, a system known as FREQUENCY SYNTHESIS has been developed to meet the stringent demands for stability. This system uses circuitry that produces a signal frequency through a heterodyning and frequency selection process. This signal is not harmonically related to any of the signals used in the heterodyning process. It is also not related to the selected crystal frequency. This makes the signal unique.

Figure 2-28 is a multiple crystal, frequency synthesizer that produces desired output frequencies by mixing frequencies from several crystal oscillators. Each oscillator uses ten or more crystals to control its operating frequency. This provides for a large number of output frequency combinations. Figure 2-29 is a practical frequency synthesizer in which the harmonics and subharmonics of a single standard oscillator are combined to provide a wide multichoice of output signals. Each of these signals is harmonically related to a subharmonic of the standard oscillator. You will find the primary difficulty encountered in the frequency synthesizers is the presence of spurious signals generated in the "combining mixers." Extensive filtering and extremely careful selection of operating frequencies are required for even the simplest circuits. Spurious frequency problems increase and channel spacing decreases as the range of operating frequencies increases.

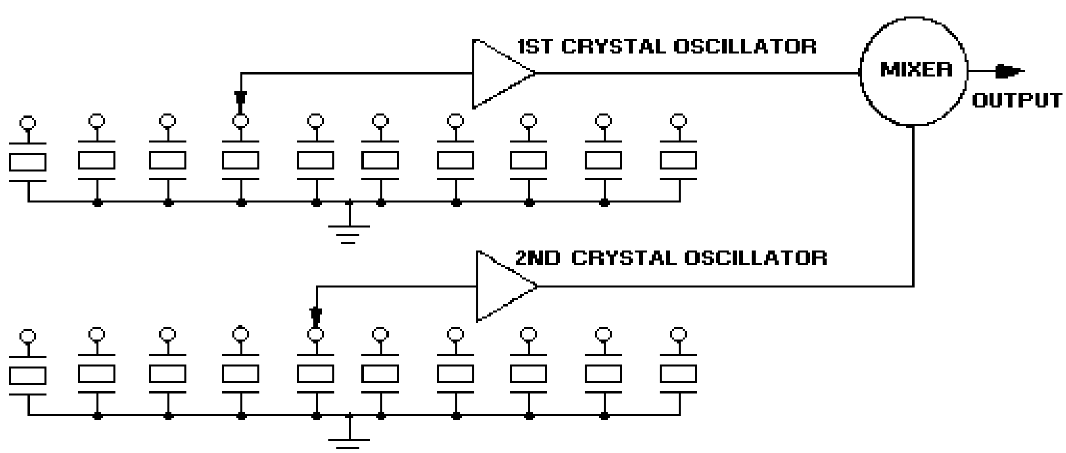


Figure 2-28.—Multiple crystal frequency synthesizer.

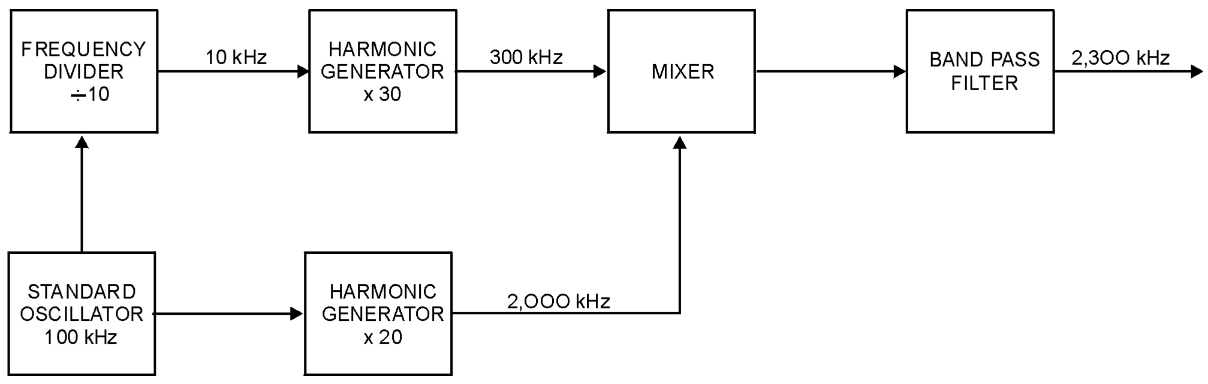


Figure 2-29.—Simple crystal frequency synthesizer.

Q25. What is the primary function of an afc circuit?

Q26. What is frequency synthesis?

AUDIO REPRODUCTION DEVICES

The purpose of audio reproduction devices, such as loudspeakers and headphones, is to convert electrical audio signals to sound power. Figure 2-30 shows you a diagram of a loudspeaker called the PERMANENT MAGNET SPEAKER. This speaker consists of a permanent magnet mounted on soft iron pole pieces, a voice coil that acts as an electromagnet, and a loudspeaker cone connected to the voice coil. The audio signal has been previously amplified (in terms of both voltage and power) and is applied to the voice coil. The voice coil is mounted on the center portion of the soft iron pole pieces in an air gap so that it is mechanically free to move. It is also connected to the loudspeaker cone; as it moves, the cone will also move. When audio currents flow through the voice coil, the coil moves back and forth proportionally to the applied ac current. As the cone (diaphragm) is attached to the voice coil, it also moves in accordance with the signal currents; in so doing, it periodically compresses and rarefies the air, which produces sound waves.

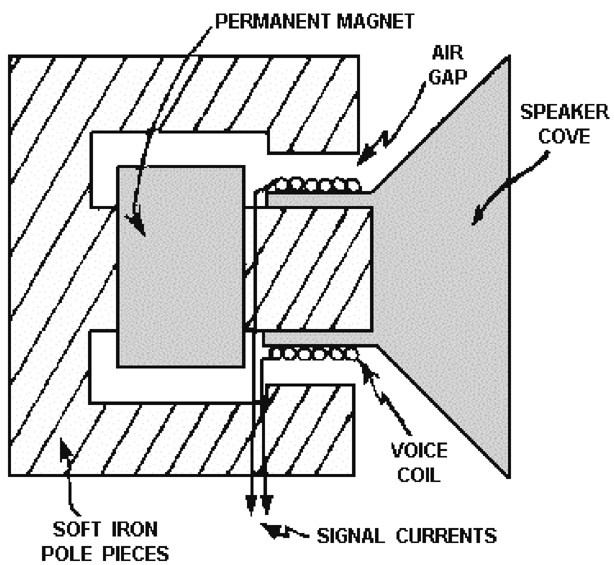


Figure 2-30.—Permanent magnet speaker.

Most speakers of the above type receive their input by means of transformer coupling. This is necessary because of the normal, low impedance of the voice coil. You will find the standard impedance values for this type speaker are 4, 8, 16, and 32 ohms. Other impedance values may be obtained, but those listed are the most common.

While permanent magnet speakers perform reasonably well in the audio range, they have limitations. Most Navy speakers reproduce low audio frequencies quite well, mid-band frequencies fairly well, and high frequencies quite poorly. Let's see why. When the speaker is constructed, only a limited number of turns may be built into the voice coil. This gives us a fixed inductance. At low frequencies, the inductive reactance of the voice coil is relatively low, and large audio currents flow. This provides a strong magnetic field around the voice coil and a strong interaction with the field of the permanent magnet. Low frequency response is excellent. At midband frequencies, inductive reactance increases and less current flows in the voice coil. This produces less magnetic field and less interaction. Midband response is still acceptable in a properly designed speaker. At high audio frequencies inductive reactance is quite high, and little current flows in the voice coil. This results in a greatly reduced voice coil field and little interaction with the permanent magnetic field. Also at high frequencies the interwinding capacitance of the voice coil tends to shunt some of the high audio frequencies, which further reduces the high frequency response.

Frequency response of most permanent magnet speakers falls off at the higher audio frequencies. This problem is normally overcome either by the use of an expensive, specially designed speaker, or through the use of two speakers, one of which is designed to operate well at the higher audio frequencies (tweeter) and one at the lower frequencies (woofer).

As shown in figure 2-31, an electromagnet may be used in place of a permanent magnet to form an electromagnetic dynamic speaker. When we do this, sufficient dc power must be available to energize the field electromagnet. The operation otherwise is much the same as that of the permanent magnet type. This type of speaker is seldom used in Navy equipment.

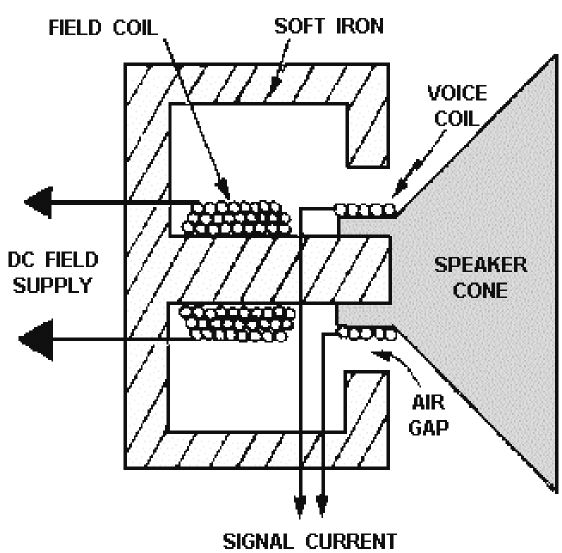


Figure 2-31.—Electromagnetic speaker.

Figure 2-32 shows a diagram of typical headphones used with Navy equipment. The device consists of a permanent magnet and two small electromagnets through which the signal currents pass. A soft iron diaphragm is used to convert the electrical effects of the device into sound power. When no signal currents are present, the permanent magnet exerts a steady pull on the soft iron diaphragm. Signal current flowing through the coils mounted on the soft iron pole pieces develops a magnetomotive force that either adds to or subtracts from the field of the permanent magnet. The diaphragm thus moves in or out according to the resultant field. Sound waves are then reproduced that have an amplitude and frequency (within the mechanical capability of the reproducer) similar to the amplitude and frequency of the signal currents.

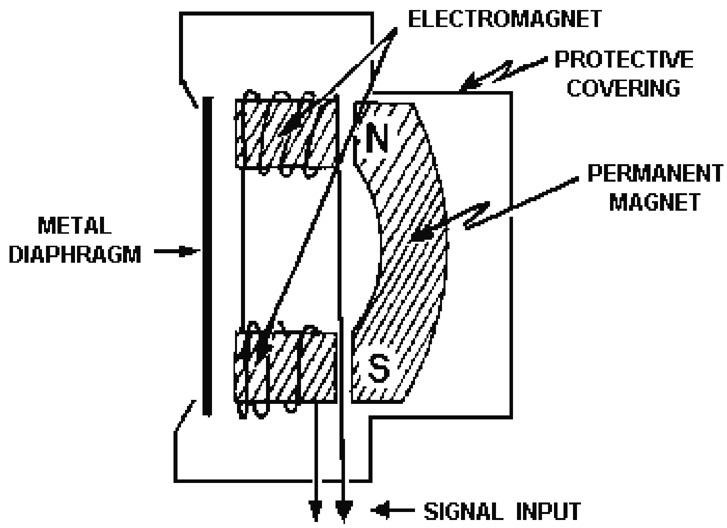


Figure 2-32.—Headphone.

As compared to permanent magnet speakers, standard headphones are considered to be high-impedance devices. Headphone electromagnets are normally wound with many turns of small wire, which provide the larger impedance. Because of the physically small size and inflexibility of the metal diaphragm, the headphones often give poor response to the lower audio frequencies. In the voice range of audio, most standard issue headphones are adequate.

SUMMARY

In this chapter you learned transmitter and receiver fundamentals. We also discussed modes of operation and special controls circuits. Let's review some of these areas.

A **HARMONIC** is an exact multiple of the fundamental frequency. Even harmonics are 2, 4, and so on, times the fundamental. Odd are 3, 5, and so on, times the fundamental frequency.

A **SUBHARMONIC** is an exact submultiple of the fundamental frequency. Even subharmonics are one-half, one-quarter, and so on. Odd subharmonics are one-third, one-fifth, and so on, of the fundamental frequency.

SUPPRESSION is the process of eliminating an undesired portion of a signal.

MULTIPLEXING is a method for simultaneous transmission of two or more signals over a common carrier wave.

An **ORDER-WIRE CIRCUIT** is a circuit between operators used for operations control and coordination.

RECEPTION is when an electromagnetic wave passes through a receiver antenna and induces a voltage in that antenna.

DETECTION is the separation of low-frequency (audio) intelligence from the high (radio) frequency carrier.

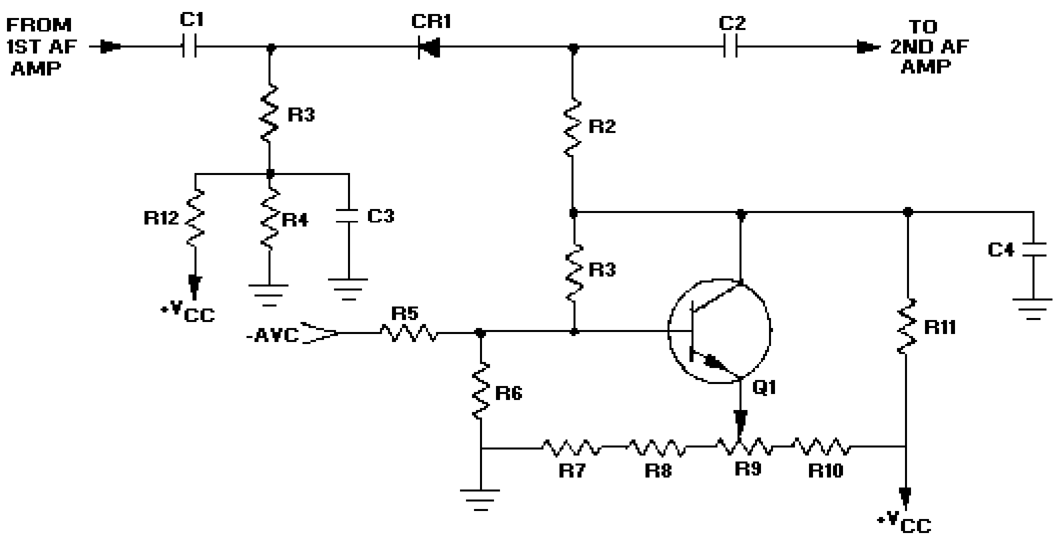
REPRODUCTION is the process of converting electrical signals to sound waves. This sound is speech, music, and so on.

SENSITIVITY of a receiver is the ability to reproduce weak signals. The greater the receiver sensitivity, the weaker the signal that will be reproduced.

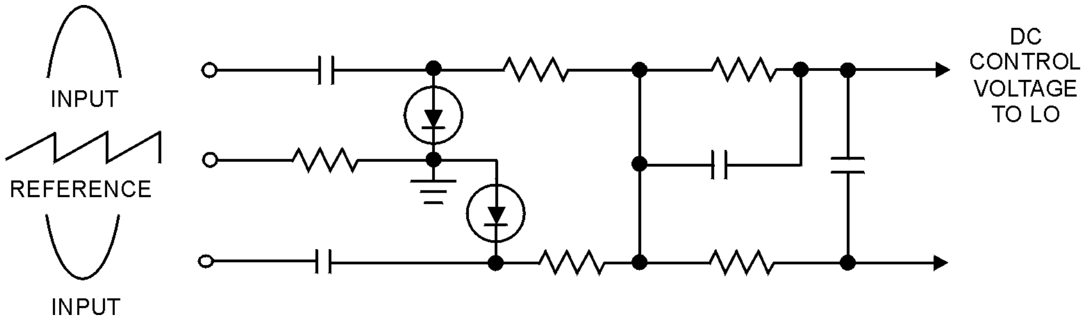
Receiver **SELECTIVITY** is the ability to select the desired signal and reject unwanted signals.

NOISE SILENCER, NOISE SUPPRESSOR, or **NOISE LIMITER**, are circuits that clip the peaks of the noise spikes in a receiver.

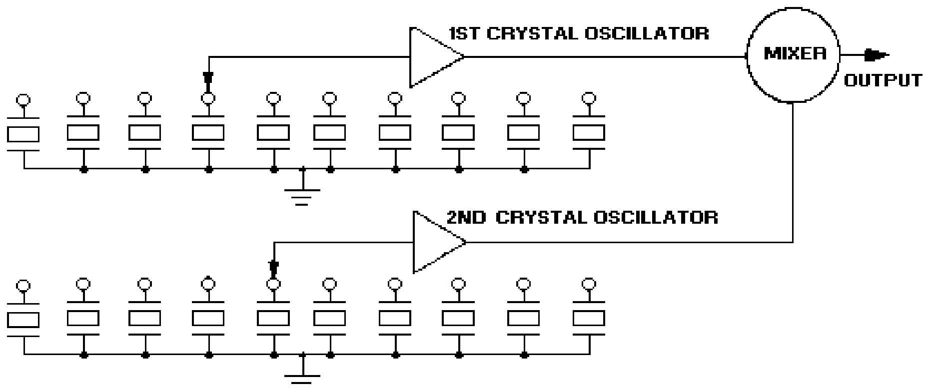
SQUELCH is a circuit that cuts off the output of a receiver when there is no input.



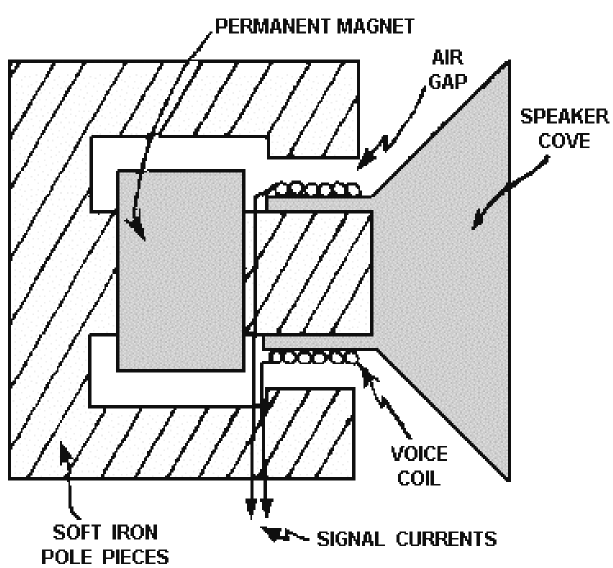
A **BALANCED-PHASE DETECTOR** or **PHASE-DISCRIMINATOR** is a circuit that controls the oscillator frequency (afc).



FREQUENCY SYNTHESIS is a signal-producing process through heterodyning and frequency selection.



A **PERMANENT MAGNET SPEAKER** is one with a permanent magnet mounted on soft iron pole pieces.



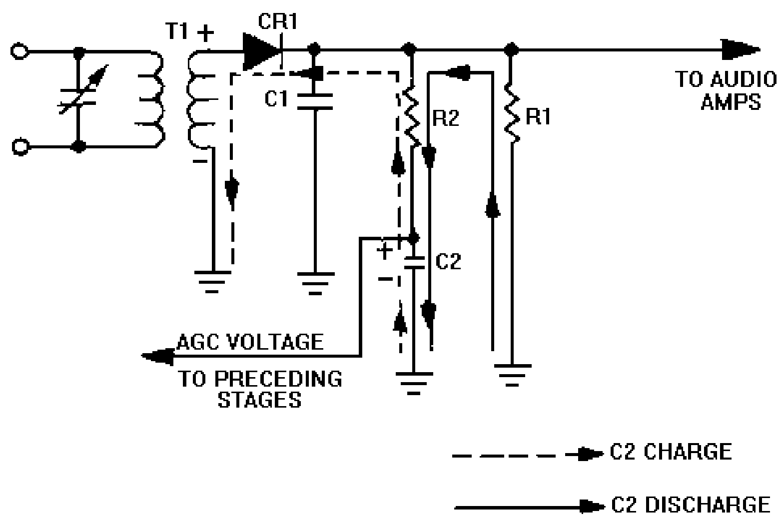
The **FIDELITY** of a receiver is the ability to accurately reproduce at its output the signal at its input.

GANGED TUNING is the process used to tune two or more circuits with a single control.

HETERODYNING is the mixing of the incoming signal with the local oscillator frequency. This produces the two fundamentals and the sum and difference frequencies.

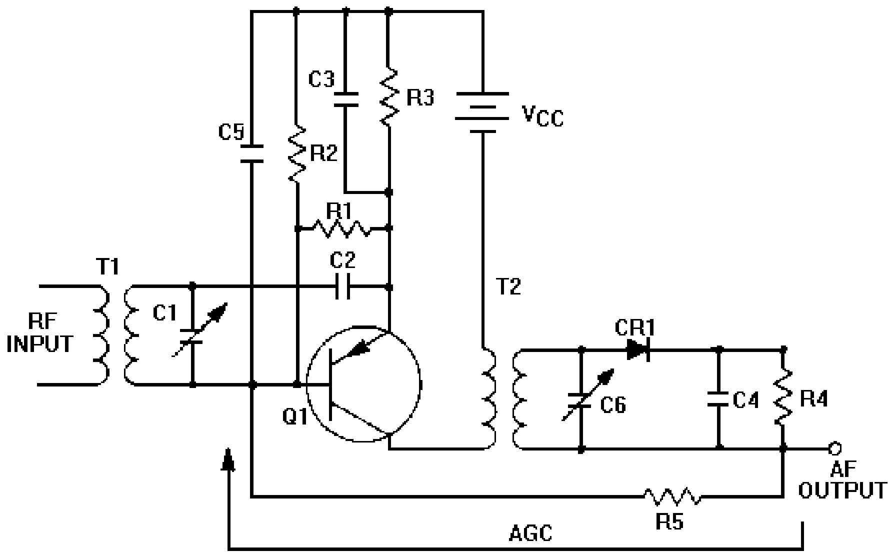
An **IMAGE FREQUENCY** is an undesired frequency capable of producing the desired frequency through heterodyning.

AUTOMATIC VOLUME/GAIN CONTROL is a circuit used to limit variations in the output signal strength of a receiver.



FADING is the variations in signal strength at the antenna of a receiver.

REVERSE AGC is when an amplifier is driven toward cutoff.



FORWARD AGC is when an amplifier is driven toward saturation.

A **BEAT-FREQUENCY OSCILLATOR** is an additional oscillator used in a receiver when the receiver is receiving a cw signal and provides an audible tone.

ANSWERS TO QUESTIONS Q1. THROUGH Q26.

A1. Am, fm, cw, ssb.

A2. It generates an rf carrier at a given frequency within required limits.

A3. Power amplifier.

A4. It converts audio (sound) into electrical energy.

A5. When no modulation is present.

A6. It is an exact multiple of the basic or fundamental frequency.

A7. 600 megahertz.

A8. To obtain higher carrier frequencies.

A9. It saves power and frequency bandwidth.

A10. For operator-to-operator service messages and frequency changes.

A11. Reception, selection, detection, and reproduction.

A12. Sensitivity, noise, selectivity, and fidelity.

A13. Heterodyning.

A14. To extract the modulating audio signal.

A15. Wide bandpass.

A16. A special type of detector and a carrier reinsertion oscillator.

A17. Attenuates the strong and amplifies the weak.

A18. To limit unwanted variations in the output.

A19. Weak signals produce bias, which could result in no usable receiver output.

A20. Dagc does not attenuate weak signals.

A21. It is heterodyned with the rf to produce an audio frequency.

A22. It eliminates noise when no signal is being received.

A23. It controls the amount of bass and treble response.

A24. It is used to achieve maximum selectivity.

A25. It is used to accurately control the frequency of the oscillator.

A26. The process of selecting and/or heterodyning frequencies to produce a signal frequency.

CHAPTER 3

FUNDAMENTAL SYSTEMS EQUIPMENT

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. State the function of a radio communications handset, a radio set control, and a transfer switchboard.
2. Describe the functions and interrelationships of a radio transmitter.
3. Describe the functions of receive and transmit multicouplers.
4. Describe the differences between the codes used for manual telegraphy and teletypewriter transmissions.
5. Describe the two basic modes of teletypewriter operation.
6. Describe the two types of teletypewriter dc circuits.
7. State the two types of radio teletypewriter shift systems and describe their basic differences.
8. Describe the functions and interrelationships of radio-frequency-carrier shift send and receive systems.
9. Describe the signal flow in an audio-frequency-tone shift system.
10. State the function of the tone terminal set in an audio-frequency-tone shift system.
11. Describe the basic multiplexing process.
12. Describe the three operations performed by a facsimile system.
13. Describe the functions and interrelationships of facsimile equipment.
14. Describe the countermeasures that can be used to eliminate compromising emanations.

EQUIPMENT PURPOSES

A communications system is a collection of equipment used together to do a specific job. You may see this equipment used to send or receive voice communications, or both, or to send, receive, or send and receive teletypewriter information.

Figure 3-1 is a basic block diagram of a voice system. You can see how this equipment is interconnected to form a basic communications system. We are going to look at several of the equipment blocks in detail.

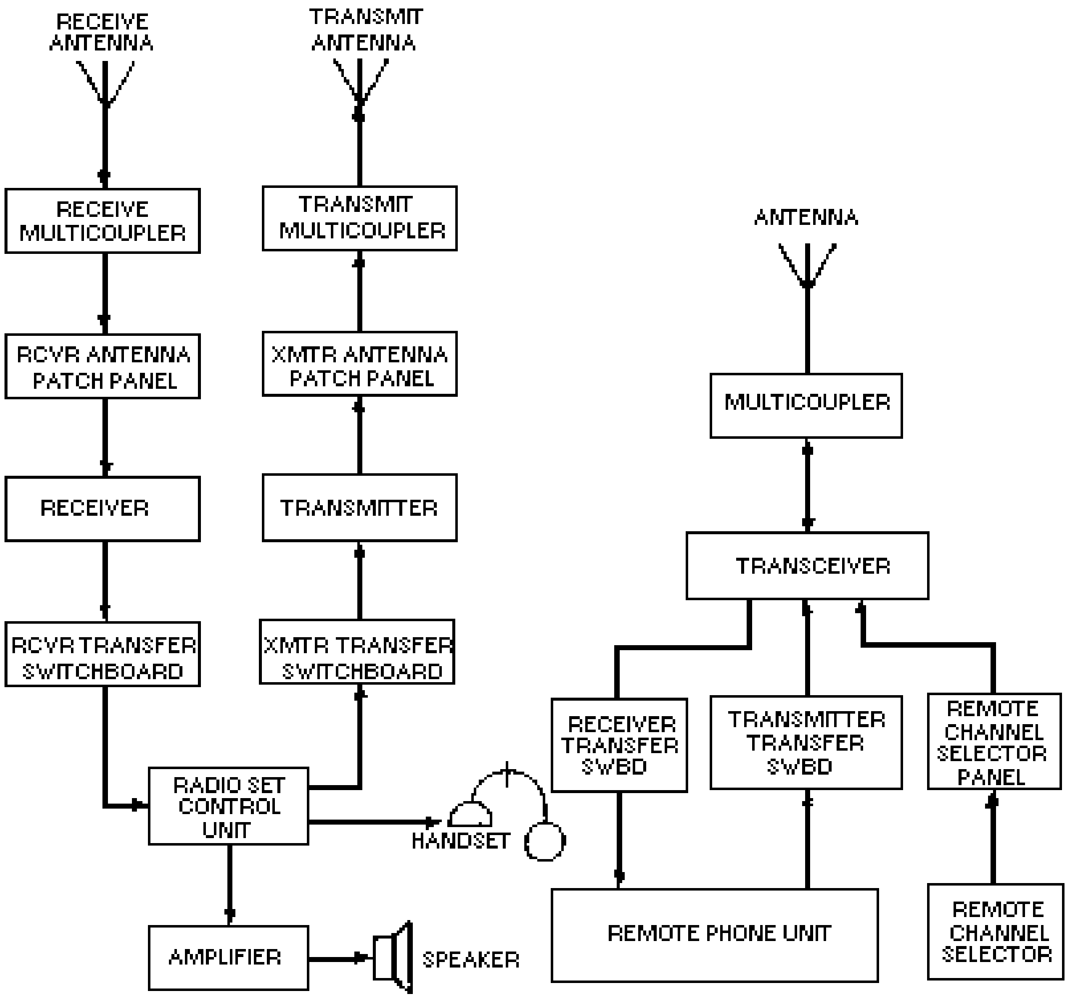


Figure 3-1.—Voice system.

HANDSET

The handset converts acoustical energy (your voice) to electrical energy for use in modulating a radio transmitter. It also converts electrical energy to acoustical energy for reproduction of a received signal. When the push-to-talk button is depressed on the handset, the dc keying circuit to the transmitter is closed, placing the transmitter on the air.

Handsets are normally connected to a radio set control unit.

RADIO SET CONTROL UNIT

The radio set control unit shown in figure 3-2 provides a capability to remotely control some radiophone transmitter functions and the receiver output. Some of the controls are used for turning the transmitter on and off. Others are used for voice modulating the transmission (or keying when cw operation is desired). You can even control the audio output level of the receiver and silence the receiver when transmitting.

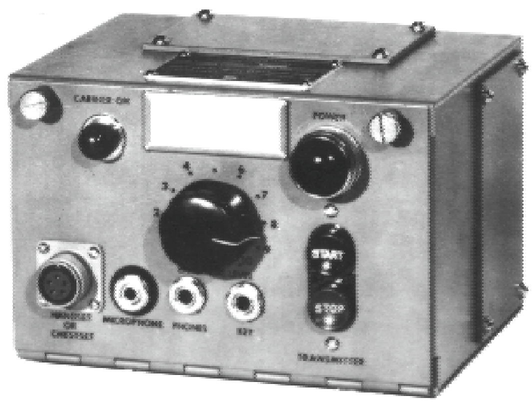


Figure 3-2.—Radio set control unit.

Under standard operating conditions up to four of these units can be used in parallel with a single transmitter and receiver group to provide additional operating positions. This setup is often found aboard ship where a transmitter and/or receiver is controlled and operated from several locations such as the bridge or the combat information center.

TRANSFER SWITCHBOARDS

A transmitter transfer switchboard provides the capability to transfer remote control station functions and signals to transmitters. Figure 3-3 is a representative transfer switchboard that provides the capability for selectively transferring any one, or all, of ten remote control station functions and signals to any one of six transmitters. The cabinet has ten rotary switches arranged in two vertical rows of five each. Each switch has eight positions. The circuitry is arranged so that you cannot parallel transmitter control circuits; that is, you cannot connect more than one transmitter to any remote control location.

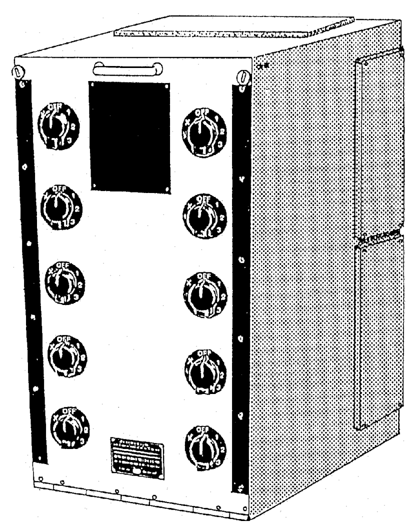


Figure 3-3.—Transmitter transfer switchboard.

Each switch operating knob corresponds to a remote control station. Each switch position (1 through 6) corresponds to a transmitter. One switch position, X, provides for transfer of all circuits to additional transmitter transfer switchboards when more than six transmitters are installed in the system. When the rotary switch is placed in the OFF position the remote control station is removed from the system.

Let's look at an example of one transfer switchboard application. When remote control station number two is to have control of transmitter number three, the switch knob designated number two is rotated until its pointer indicates position three on its dial plate.

The receiver transfer switchboard permits the operator to transfer the audio output from a receiver to a remote control station audio circuit. A representative receiver transfer switchboard is shown in figure 3-4. This switchboard contains ten seven-position switches. Each switch is connected to a remote control station, and each switch position (one through five) is connected to a receiver.

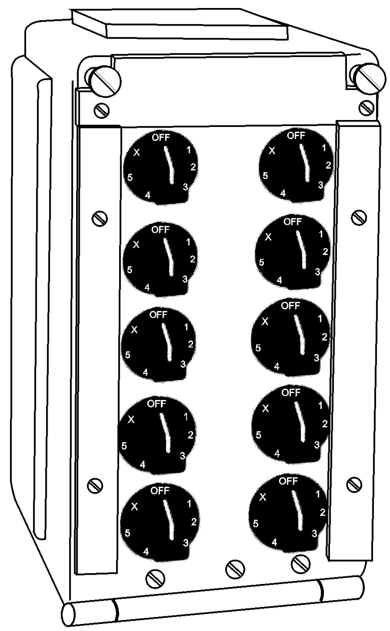


Figure 3-4.—Receiver transfer switchboard.

The X position on each switch allows transfer of circuits to additional switchboards just like with the transmitter transfer switchboard.

Q1. What are the basic functions of a handset?

Q2. What capability does a transmitter transfer switchboard provide?

Q3. What function does a receiver transfer switchboard perform?

TRANSMITTERS

You learned earlier that transmitters may be simple with low power (milliwatts) capabilities. These may be used to send voice messages a short distance. You may also use highly sophisticated units that use thousands (even millions) of watts of power to send many channels of data (for example voice, teletypewriter, television, telemetry) simultaneously over long distances. Let's look at a complete transmitter set.

Radio Transmitting Set

The applications, configurations, and components you will become familiar with here are typical of most general purpose transmitter systems used in the Navy. A specific transmitter is used only for ease of illustration and example.

We will be discussing a 1,000 watt, single-sideband radio transmitting set that is available to the Navy in any one of four setups. The normal configuration has a transmitter capable of voice, continuous wave, and radio teletypewriter transmissions in the 2- to 30-megahertz frequency range. Exact spacing and number of channels available within the frequency spectrum, modes of operation, and frequency range depend on the model of equipment and how it is configured for use. Stack or rack mounting is used in a ship or shore permanent installation with accessory equipment (for example an rf amplifier, coupler control unit, or power supply) to form a complete communications system. One of three different three-phase primary power sources can be used (depending on whether the transmitter is land, air, or shore based) to provide operating power to the set. Combinations available are 115 volts, 400 hertz; or 208/440 volts, 60 hertz.

General Description

Figure 3-5 shows the major units of this set. They are the radio transmitter, the radio frequency amplifier, the power supply, and the electrical equipment shock mount base. An antenna coupler group (consisting of a coupler and coupler control unit) is normally used to match the impedance of the system to a 50-ohm transmission line. If you want to operate with any 50-ohm antenna system, terminating connections are available.

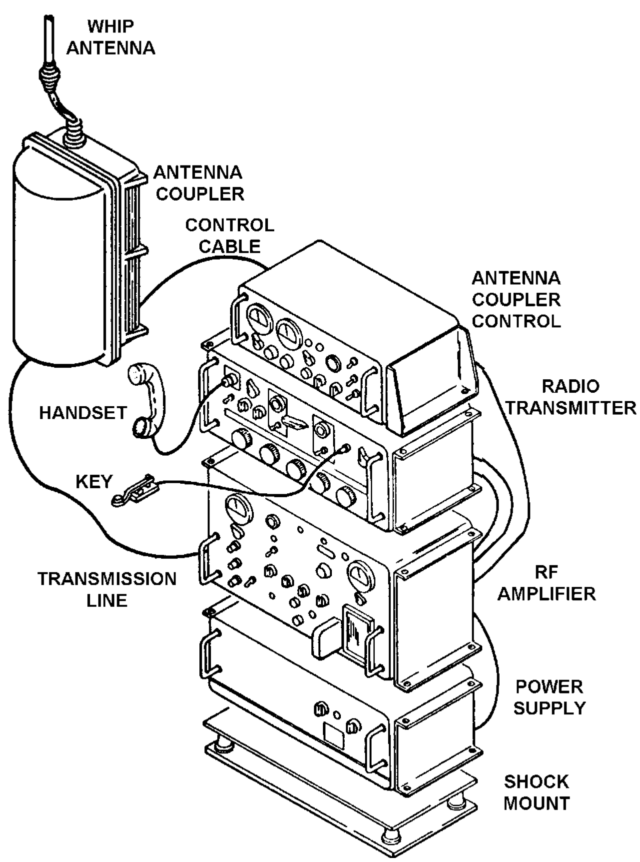


Figure 3-5.—Radio transmitting set.

The transmitter unit provides an upper sideband (usb), lower sideband (lsb), independent sideband (isb), cw, fsk, or compatible AM signal. The output of the transmitter has enough power to drive the radio frequency amplifier.

Depending on the model, the transmitter tunes across the frequency range in 100- or 500-hertz increments. Digital circuitry is used to accomplish this process. Transmitter outputs are also applied to the rf amplifier to automatically tune it to the correct frequency. We will go through a detailed breakdown of the transmitter unit later in this chapter.

RADIO FREQUENCY AMPLIFIER.—The rf amplifier unit is a two-stage linear power amplifier that produces an output of 1,000 watts with a nominal input of 100 milliwatts. Nineteen frequency bands are used to cover the operating frequency range. The operating band is automatically selected by digital coding generated by the transmitter. The code controls two motor-driven band switch assemblies. Automatic control circuits protect the unit against overload and compensate for variations in system gain, mode of operation, and loading.

All low voltages required for operation (except two of the relay control voltages) are internally produced. The high voltages required in the amplifier stages are produced by the associated power supply (when using 60 hertz primary power) or the optional internally mounted power supply (when using 400 hertz primary power).

Let's take a look at figure 3-6 to see all the operating controls and indicators located on the front panel. Some controls are used only for initial setup and are protected by a hinged access cover. All connections are made at the rear of the case. The amplifiers and the associated interstage broadband transformer assemblies are cooled by forced ventilation. Cooling air is drawn through a filter on the front panel and exhausted through a port on the rear of the case. You should always take particular care to clean or replace any filter in electronic equipment as a regular part of your preventive maintenance program.

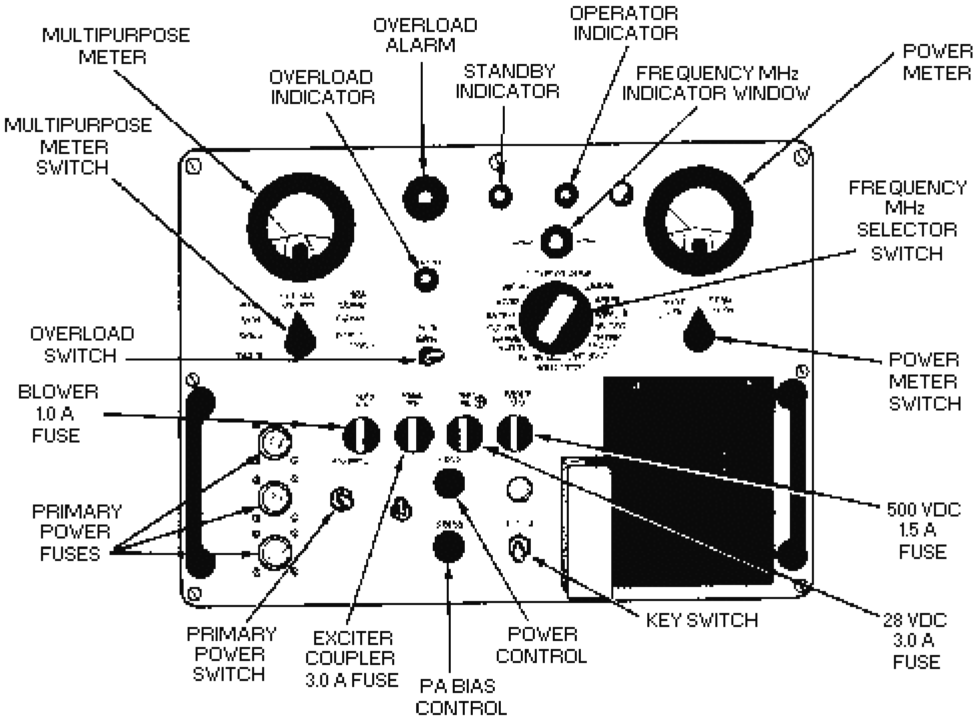


Figure 3-6.—Rf amplifier unit.

POWER SUPPLIES.—One power supply produces operating voltages for the amplifier when operating from a 60-hertz power source. All components of the power supply, except the power transformers, are mounted on a chassis and panel assembly that is hinge-mounted to a metal case. The power transformers are constructed as part of the case and there are no operating controls.

The other power supply produces operating voltages for the rf amplifier when a 400-hertz, three-phase, 115-volt primary power source is used.

ANTENNA COUPLER GROUP.—The antenna coupler group is an automatic antenna tuning system. However, the equipment design includes provisions for manual or semiautomatic tuning. This makes the system adaptable for use with other radio transmitters. The manual tuning capability is useful when a failure occurs in the automatic tuning circuitry. Tuning can also be accomplished without the use of rf power (SILENT TUNING). This method is useful in installations where radio silence must be maintained except for brief transmission periods.

The antenna coupler matches the impedance of a 15-, 25-, 28-, or 35-foot whip antenna to a 50-ohm transmission line at any frequency in the 2- to 30-megahertz range. Control signals from the associated antenna coupler control unit automatically tune the matching network in less than five seconds. During manual and silent operation, tuning is accomplished by the operator with the controls mounted on the antenna coupler control unit. A low power (not to exceed 250 watts) cw signal is required for tuning. Once tuned, the coupler is capable of handling 1,000 watts peak envelope power (pep).

The coupler is enclosed in an aluminum, airtight, pressurized case. Six mounting feet enable the unit to be attached to the mast of a ship at the base of a whip antenna. The coupler is pressurized with dry nitrogen to aid internal heat transfer and to prevent corona and arcing. All components of the coupler are secured to a chassis that is mounted to the case so that an air duct exists between the chassis plate and the case. An internal fan circulates the nitrogen over and through the heat-producing elements and then through the air duct. While passing through the air duct, the nitrogen loses its heat to the bottom of the case. This heat is then transferred by convection through fins on the bottom of the case and by conduction through the mounting feet.

Figure 3-7 shows the antenna coupler control unit. This unit provides the power and control signals required to tune the coupler. Control signals are either automatically produced by the coupler control when a tune cycle is initiated or manually produced with the front panel controls.

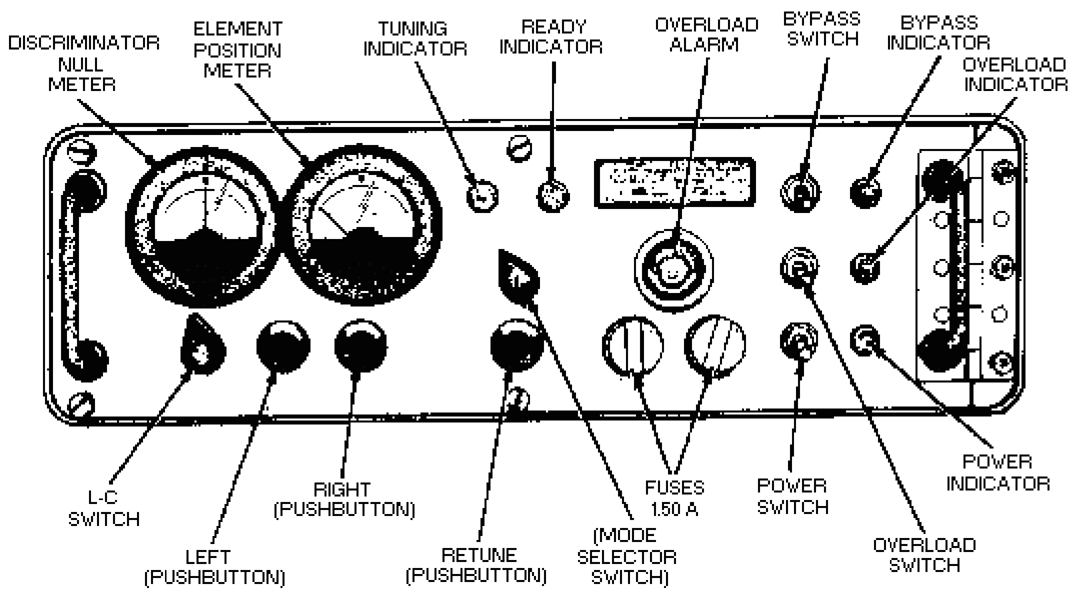


Figure 3-7.—Antenna coupler control unit.

All dc operating voltages are produced from a 115-volt, 48- to 63- or 350- to 450-hertz, single-phase primary power source. Meter and protection circuits are used to give you complete control of the coupler from the remotely positioned coupler control unit.

Q4. If the rf amplifier discussed has an 80 milliwatt input, what would be the maximum output?

Q5. What are the tuning modes for the coupler group discussed?

Q6. What is the purpose of an antenna coupler?

Q7. Why is the coupler pressurized with nitrogen?

RADIO TRANSMITTER.—Figure 3-8 shows the front panel of the radio transmitter unit. The radio transmitter accepts audio or coded intelligence and uses it to modulate one of 280,000 possible operating radio frequencies in the 2.0- to 29.999-megahertz frequency range. Tuning is accomplished digitally by means of five control knobs and a switch located on the front panel. The transmitter has a normal rf output level of at least 100 milliwatts and is designed to be used with an associated rf power amplifier.

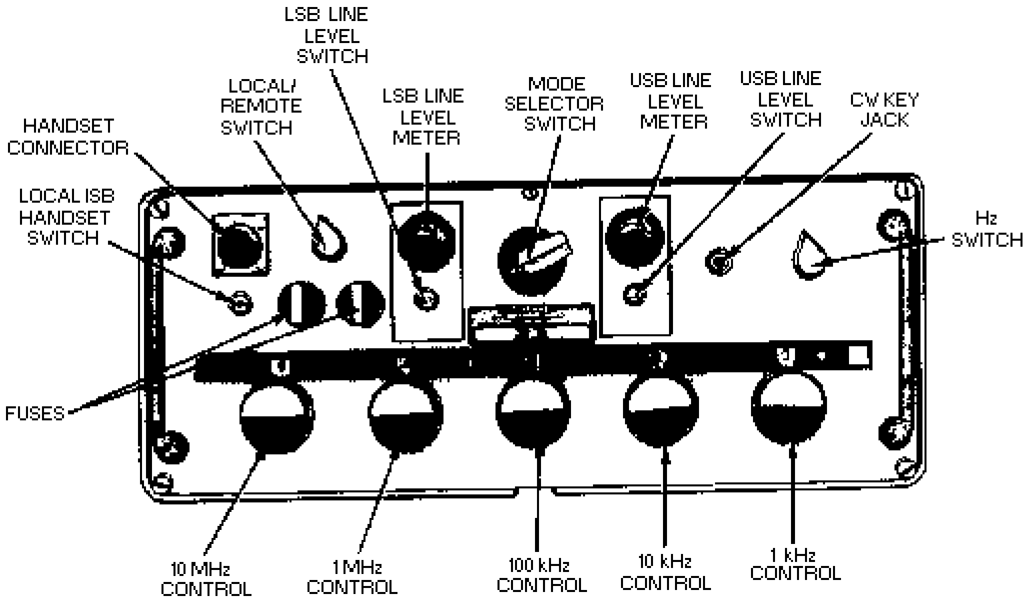


Figure 3-8.—Radio transmitter unit.

When the AM and ssb transmit modes of operation are used, the output from a handset is applied to the transmitter. The voice signals are amplified and used to modulate a 500-kilohertz local carrier that produces a 500-kilohertz IF. The resulting double sideband signal is filtered in the AM mode, amplified, and converted by a triple-conversion process to the desired rf operating frequency. The rf signal is amplified to a nominal 100 milliwatt level. In cw operation, the 500-kilohertz local carrier is inserted directly into the IF amplifiers. The signal is further processed in the same manner as the voice signals in the AM or ssb modes of operation. In fsk operation, the loop current is converted to audio frequencies representing marks and spaces. These audio signals are applied to the audio circuits of the transmitter. Thereafter, these signals are processed in the same manner as the voice signals in AM or ssb modes of operation. A typical radio transmitting set block diagram is shown in figure 3-9.

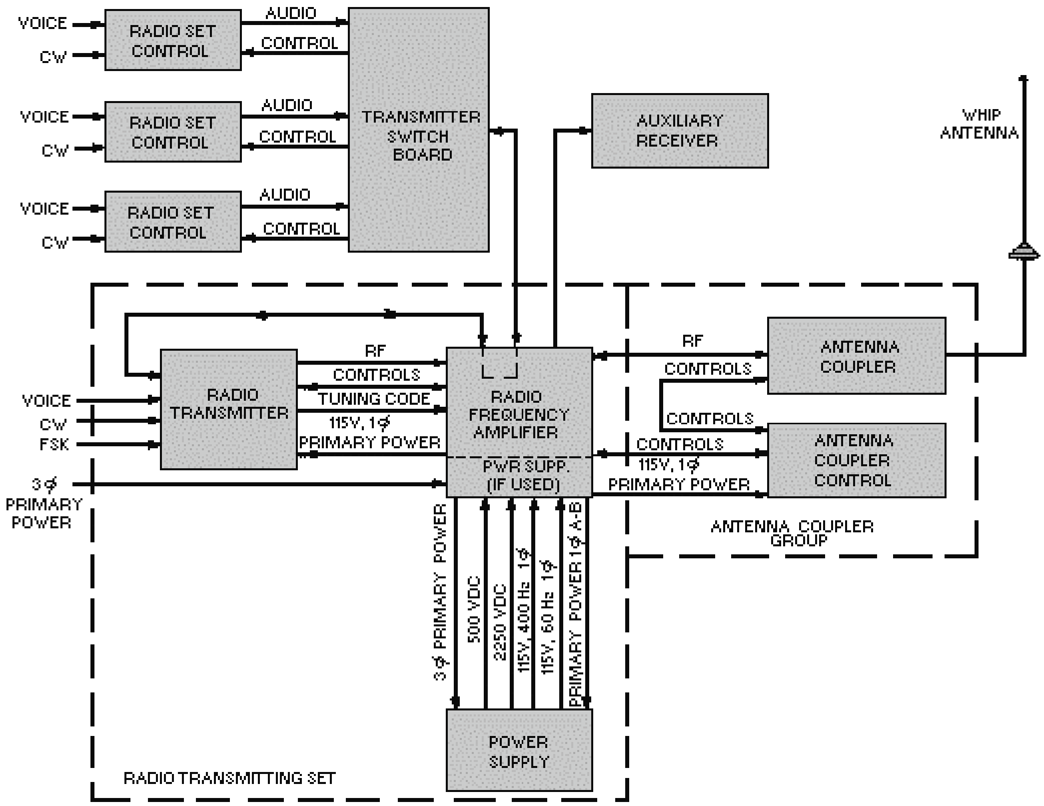


Figure 3-9.—Typical radio transmitting set block diagram.

RECEIVERS

The receiver we will discuss is a triple-conversion superheterodyne, tunable from 2 to 30 megahertz. Triple conversion uses three IF frequencies to give better adjacent-channel selectivity and greater image-frequency suppression. Figure 3-10 shows the front panel of this receiver where tuning is done digitally by five controls and a switch. A display window directly above each control provides a digital readout of the frequency setting. The displayed frequency can be changed in 1-kilohertz increments. The front panel switch allows the operating frequency to be changed in 100- or 500-hertz increments depending on the model. This will provide you with 280,000 discrete frequencies locked to a very accurate frequency standard. You can continuously tune each 1,000-hertz increment by selecting the VERNIER position of the hertz switch. When using the vernier, the full accuracy of the frequency standard is sacrificed. The receiver demodulates and provides audio outputs for the lsb, usb, isb, AM, cw, and fsk types of received signals.

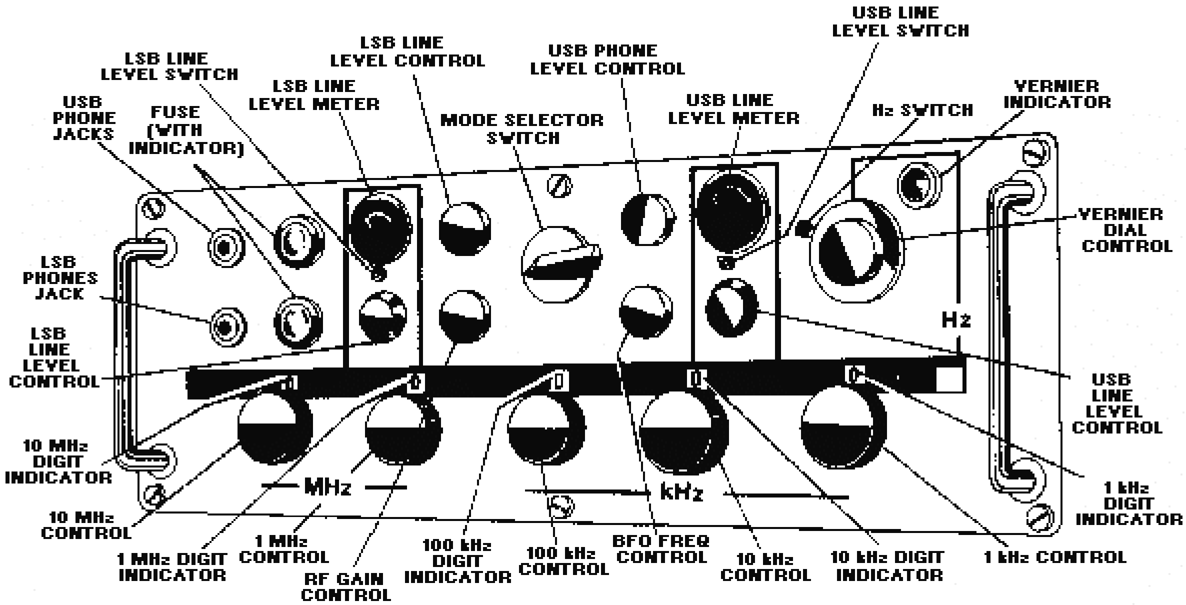


Figure 3-10.—Typical radio receivers

Q8. What are the transmitter operating modes?

Q9. What type of tuning does the receiver use?

ANTENNA DISTRIBUTION SYSTEMS

Receiving antenna distribution systems operate at low power levels and are built to fit a standard 19-inch rack. Each piece of distribution equipment is fitted with termination or patch fittings designed for ease of connecting and disconnecting. A basic patch panel is shown in figure 3-11. Even a fundamental distribution system has several antenna transmission lines and several receivers. Normally a patch panel consists of two basic patch panels. One panel is used to terminate the antenna transmission lines and the other the lines leading to the receivers. Any antenna can be patched to any receiver through the use of patch cords.

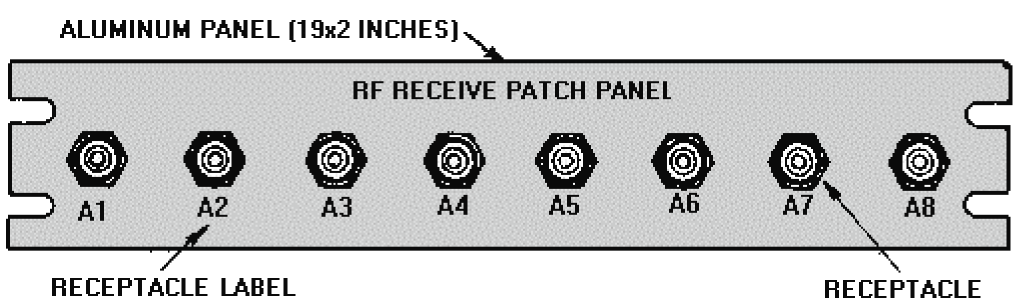


Figure 3-11.—Basic rf receive patch panel.

Many distribution systems are more complex. A complex distribution system to cover most situations is illustrated in figure 3-12. In this system you can patch four antennas to four receivers, or you can patch one antenna to more than one receiver via the multicouplers (multicouplers are covered later in this chapter). You can also patch rf and audio from one compartment to another. A frequency standard is connected (through a distribution amplifier not shown) to the receivers.

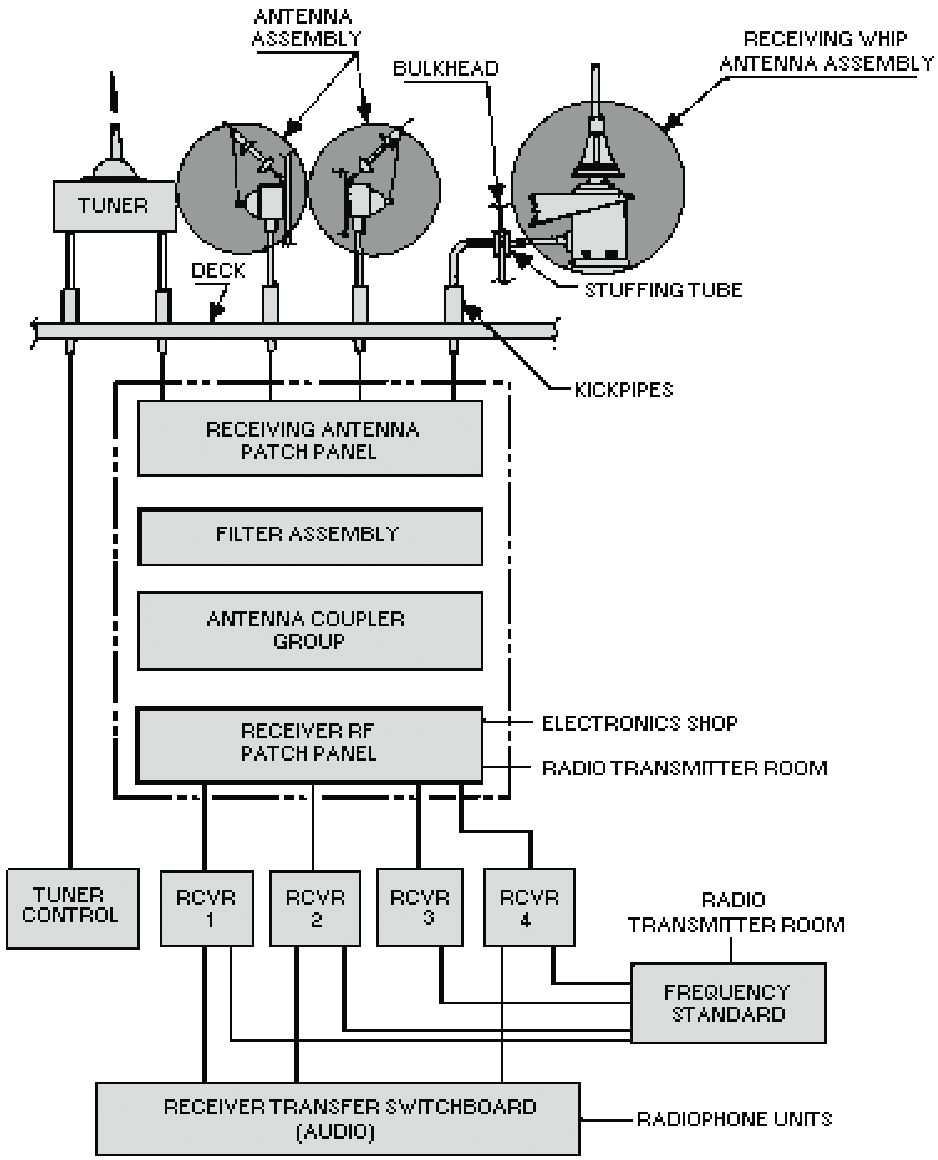


Figure 3-12.—Complex distribution system.

Transmitting antenna distribution systems perform the same functions as receiving systems. However, because of the higher power levels, design and fabrication problems are more difficult. The ideal design would be to have all the transmission lines designed for the highest power level. But because high-power patch cords are expensive, large, and difficult to handle, this approach is seldom followed.

In practice, the basic patch panel we just looked at in figure 3-11 is practical for low power levels. Another type of transmitter patch panel is shown in figure 3-13.

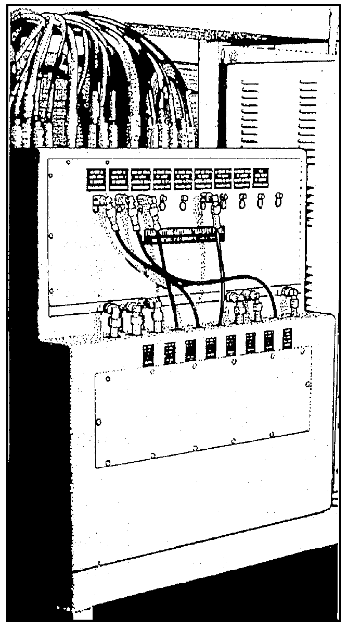


Figure 3-13.—Transmitting antenna patch panel.

This type of transmitting antenna patch panel is interlocked with the transmitter so that no open jack connection can be energized and no energized patch cord can be removed. This provides you with a greater degree of personnel and equipment safety.

Receive Multicoupler

Figure 3-14 is a filter assembly multicoupler that provides seven radio frequency channels in the 14-kilohertz to 32-megahertz range. Any or all of these channels may be used independently of any of the other channels, or they may operate simultaneously. You can make connections to the receiver by means of coaxial patch cords, which are short lengths of cable with plugs attached to each end.

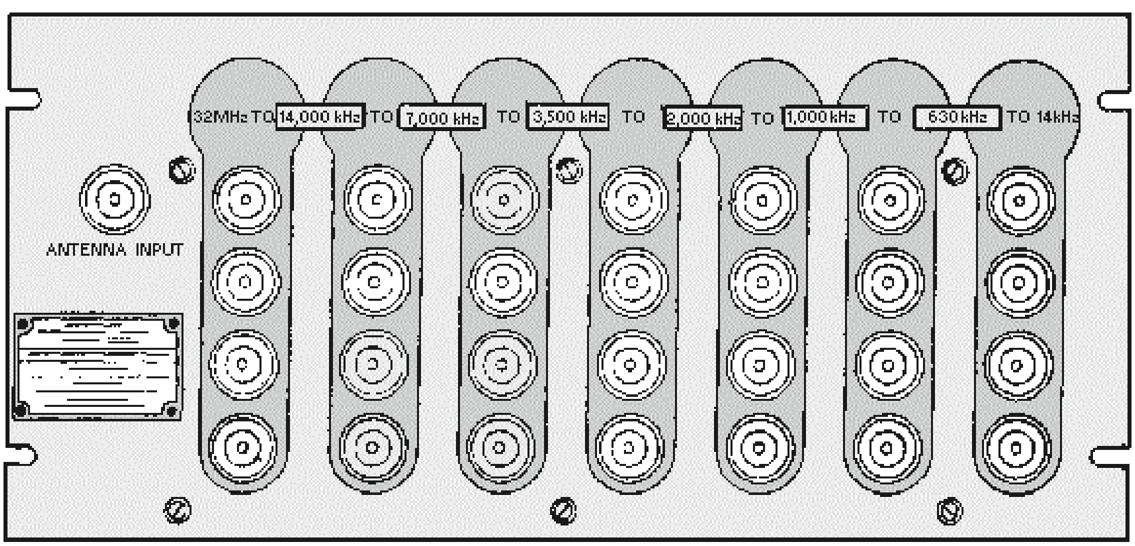


Figure 3-14.—Electrical filter assembly.

A set of nine plug-in type filter assemblies is furnished with the equipment and covers the entire vlf, lf, mf, and hf bands. Only seven of the assemblies may be installed at one time, and you have the option of selecting those you need to cover the most used frequency bands.

Figure 3-12 illustrates how the filter assembly is used in combination with other units to pass an rf signal from an antenna to one or more receivers.

Transmit Multicouplers

Most multicouplers for the hf range are designed for use with either transmitters or receivers, although some are used with both. There are a large number of channels in a multicoupler so that many transmitters can be used at the same time on one antenna. This is especially true in the 2- to 12-megahertz range.

Figure 3-15 shows you an antenna coupler group designed primarily for shipboard use. Each coupler group permits several transmitters to operate simultaneously into a single, associated, broadband antenna. You can see this reduces the total number of antennas required in the limited space aboard ship.

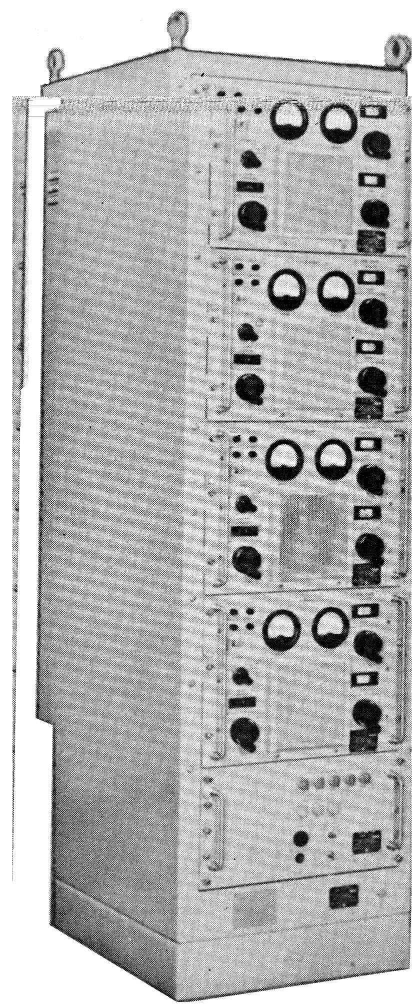


Figure 3-15.—Antenna coupler group.

These antenna coupler groups provide a coupling path of prescribed efficiency between each transmitter and its associated antenna. They also provide isolation between transmitters, tunable bandpass filters, and matching networks.

TELETYPEWRITER AND FACSIMILE EQUIPMENT

In previous areas we have discussed different methods of voice communications. At times, however, the message is too long for practical transmission by voice. To get information or an idea across to another person far away, you may also need a chart, map, or photograph. Teletypewriter (tty) and facsimile equipment allow us to do just that, with ease. Let's see how this is done.

BASIC PRINCIPLES

To give you an idea of how intelligence is sent via teletypewriter, let's take a look at the manual telegraph circuit. This circuit, shown in figure 3-16, includes a telegraph key, a source of power (battery), a sounder, and a movable sounder armature. If the key is closed, current flows through the circuit and the armature is attracted to the sounder by magnetism. When the key is opened, the armature is retracted by a spring. With these two electrical conditions of the circuit, intelligence can be transmitted by means of a teletypewriter code. These two conditions of the circuit are referred to as MARKING and SPACING. The marking condition occurs when the circuit is closed and a current flows; the spacing condition occurs when it is open and no current flows.

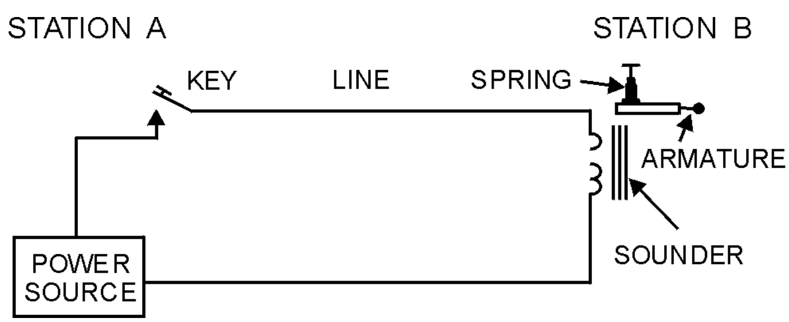


Figure 3-16.—Manual telegraph circuit.

If the key at station A is replaced by a transmitting teletypewriter and the sounder arrangement at station B is replaced by a receiving teletypewriter, the basic teletypewriter circuit (loop) shown in figure 3-17 is formed.

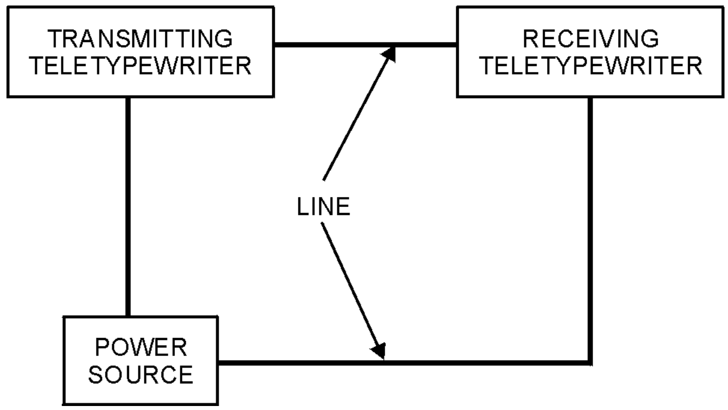


Figure 3-17.—Simple teletypewriter circuit.

If a teletypewriter signal could be drawn on paper, it would resemble figure 3-18. This is the code combination for the letter R. Shaded areas show intervals during which the circuit is closed, and the blank areas show the intervals during which the circuit is open. The signal has a total of seven units. Five of these are numbered and are called INTELLIGENCE units. The first and last units of the signal are labeled START and STOP. They are named after their functions: the first starts the signal, and the last stops it. These are a part of every teletypewriter code signal: the START unit is always spacing, and the STOP unit is always marking.

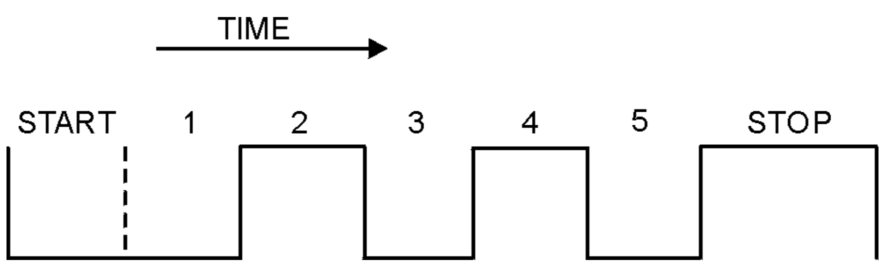


Figure 3-18.—Mark and space signals.

The teletypewriter signal is theoretically a perfect signal. The time between each unit remains the same during transmission of the signal. The shift from mark to space (and vice versa) is called a TRANSITION. A transition occurs at the beginning and end of each unit when it shifts from mark to space or space to mark; a character may have two, four, or six transitions.

When figuring the time duration of a signal character, no allowance for transition time is made since the transition is instantaneous and is considered to have zero time duration. The time duration for each unit is measured in milliseconds.

Q10. What is the function of an antenna patch panel?

Q11. What are the functions of a multicoupler?

Q12. What are the terms used to describe an open or closed telegraph circuit?

Q13. How many units are in a tty signal and what are they?

Codes

Two of the codes the Navy uses are found in manual telegraphy and in teletypewriter operation. One is very easy to understand while the other is more complex. Let's look at these two types and how they work.

MANUAL TELEGRAPHY.—In manual telegraphy, the most widely used code is the Morse code. In this code, two distinctive signal elements are employed-the dot and the dash. The difference between a dot and a dash is its duration, a dash being three times as long as a dot. Each character is made up of a number of dots and/or dashes. The dot and dash elements making up any character are separated from each other by a time interval equal to the duration of one dot. The time interval between the characters for each word is equal to the duration of three dots. The interval between words is equal to seven dots. (A signal-man uses the Morse code to send visual flashing-light messages. The radioman uses the Morse code to send messages electrically.)

TELETYPEWRITER MESSAGE TRANSMISSION.—In teletypewriter operation, the code group for each character is of uniform length. Since the Morse code is an uneven length code, it cannot be used in teletypewriter operation without additional code converters.

The FIVE-UNIT (five-level) CODE has been the most commonly used in modern printing telegraphy and is universally used in teletypewriter operation. This is also known as the Baudot code. The mechanical sending device in the teletypewriter divides the sending time for each character into five short code elements (impulses) of equal duration. The five-unit code is an example of what is called an even length or constant length code (one in which the number of signal elements for a character is the same for every character and the duration of each element is constant). In the five-unit code, each character consists of a combination of five signal elements; each element may be either a mark or a space. A total of thirty-two combinations of signal elements are possible with this arrangement.

The thirty-two possible combinations available from the five-unit code are insufficient to handle the alphabet and numbers since twenty-six combinations are required for the letters of the English alphabet alone. This leaves only six combinations for numerals, symbols, or nonprinting functions. This number of combinations is obviously inadequate; therefore, two of the thirty-two combinations are used as shift signals. The shift signals are often referred to as case-shift signals (one case is a letter shift, and the other a figure shift.) These two shift signals permit the remaining code combination to be used as letter-shift signals for letters and as figure-shift signals for numerals, function signs, and so forth. When a letter shift is transmitted, it sets the receiving instrument in a condition to recognize any letter signal combination. It will recognize letter combinations until a figure shift is received. Then the receiving instrument sets itself in a condition to recognize any figure signal combination received. The interpretation of a signal combination is determined by the previous shift signal. This plan enables 30 of the 32 available combinations to have two meanings.

Q14. There are not enough combinations of the five-unit code to handle the alphabet, symbols and so forth. What is used to increase the number of available code combinations?

Modes of Operation

The two basic modes of teletypewriter operation are ASYNCHRONOUS (start-stop) and SYNCHRONOUS. The most common mode used in teletypewriter operation is the start-stop mode. Synchronous operation is used more in high-speed data systems. Let's examine their differences.

ASYNCHRONOUS.—In the start-stop mode of operation, the receiving device is allowed to run for only one character. It is then stopped to await the reception of a start signal indicating the next character is about to start. In this manner any difference in speed between the transmitting and receiving devices can accumulate only during the duration of one character. However, you should note that a penalty must be paid for this advantage. The length of each character must be increased to include a unit (element) to start the receiving device and another to stop it.

The start unit precedes the first intelligence unit and is always a space signal. Its purpose is to start the receiving machine. The stop unit follows the last code unit and is always a mark signal. Its purpose is to stop the receiving machine in preparation for receiving the next character. The start unit must be equal to at least one unit of the code. The standard mode uses a stop unit that is 1.42 times the length of one intelligence unit. It is common practice to refer to a code unit as an element and to use the terms interchangeably. You will also hear duration of a unit referred to as the unit interval.

The length of time required to transmit the entire character is called the CHARACTER INTERVAL. Character interval becomes very important in some transmissions because certain items of equipment are character length conscious or code conscious. Stop unit intervals of various lengths are used or produced by various equipment (1.0, 1.27, 1.5, 1.96, 2.0, and so forth). Basically, the only difference between them is the length of time required to transmit one character.

SYNCHRONOUS.—Synchronous teletypewriter operation does not in all cases have to rely upon elements of the transmitted character to maintain proper position in relation to the receiving device. External timing signals may be used that allow the start and stop elements to be discarded. You will then see only the elements necessary to convey a character.

Synchronous systems have certain advantages over asynchronous systems. The amount of time taken to transmit stop and start elements is made available for information transmission rather than for synchronizing purposes. Only the intelligence elements are transmitted. In start-stop signaling, the ability of the receiving device to select the proper line signal condition is dependent upon signal quality. For example, suppose the stop-to-start transition arrives before it should; then, because of atmospherics, all subsequent selection positions in that character will appear earlier in time in each code element. A synchronous system has a higher capability for accepting distorted signals because it does not depend on a start-stop system for synchronization.

Modulation Rate

Several terms are used to refer to teletypewriter modulation rates or signaling speeds. These include BAUD RATE, BITS PER SECOND, and WORDS PER MINUTE. Baud is the only term that is technically accurate. The other terms are either approximations or require explanation.

The word baud by definition is a unit of modulation rate. You will sometimes see it used to refer to a signal element, but this reference is technically incorrect. Baud rate is the reciprocal of the time in seconds of the shortest signal element. To find the modulation rate of a signal in bauds, you must divide the number 1 by the time duration of the shortest unit interval present in the signal. For example, 22 milliseconds (.022 seconds) is the time interval of the shortest unit in the five-unit code at 60 words per minute. To find the number of bauds corresponding to 60 words per minute, divide 1 by .022. Rounding off the result of the division gives us the number 45.5, which is the baud equivalent of 60 words per minute. Each increase in words per minute will correspondingly decrease the signal unit time interval. (The defense communications system standard speed for teletypewriter operation is 100 words per minute or 75 baud.)

Words per minute is used only when speaking in general terms for an approximation of speed. The term *100 words per minute* means 100 five letter words with a space between them can be transmitted in a 60-second period. However, you can obtain this nominal words-per-minute rate in several systems by varying either modulation rate or the individual character interval (length). For this reason, the modulation rate (baud) method of reference rather than words per minute is used.

Formula for baud rate and words per minute are as follows

$$\text{Baud} = \frac{1}{\text{unit interval (in seconds)}}$$

$$\text{Words per minute} = \frac{\text{Baud rate}}{\text{unit code} \times 0.1}$$

BIT is an acronym for the words *binary digit*. In binary signals, a bit is equivalent to a signal element. Because of the influence of computer and data processing upon our language, modulation rate is sometimes expressed in bits per second. When you understand all signal elements being transmitted are of equal length, then the modulation rate expressed in bits per second is the same as the modulation rate expressed in baud.

Dc Circuits

You were told the two conditions *mark* and *space* may be represented by any convenient means. The two most common are NEUTRAL and POLAR operation. In neutral, current flow represents the mark, and no current flow represents the space; in polar operation, current impulses of one polarity represent mark, and impulses of the opposite polarity of equal magnitude represent the space.

NEUTRAL.—Neutral circuits make use of the presence or absence of current flow to convey information. A neutral teletypewriter circuit is composed of a transmitting device, a battery source to supply current, a variable resistor to control the amount of current, a receiving device, and a line for the transmission medium.

POLAR.—Polar operation differs from neutral operation in two ways. Current is always present in the polar system, and it is either positive or negative. A polar teletypewriter circuit contains the same items as a neutral circuit plus an additional "battery" source. The battery referred to here is not an actual battery but is a solid-state dc power supply. It provides variable current to the teletypewriters. The reason for having an extra battery source is because polar circuits use positive battery for marks and negative battery for spaces.

You will find in polar operation that the distortion of a signal is almost impossible through low line currents, high reactance, or random patching of signal circuits or equipment. In polar signaling when you experience a complete loss of current (a reading of zero on a milliammeter), you know you have line or equipment trouble; whereas the same condition with neutral signaling may indicate a steady space is being transmitted. This gives us a condition called RUNNING OPEN. Under this condition, the teletypewriter appears to be running because the machines is decoding the constant space as the Baudot character blank and the type hammer continually strikes the type box but there is no printing or type box movement across the page.

Q15. What are the two teletypewriter modes of operation?

Q16. Define baud.

Q17. Define bit.

Q18. What are the two types of dc operations used to represent mark and space conditions?

BASIC SYSTEMS

When two ttys are connected by communications wire or cable (over short or long distances), the exchange of information between them is direct. When the teletypewriters are not physically joined, exchange of information is more involved. Direct-current mark and space intervals cannot be sent through the air. The gap between the machines must be bridged by radio using a radio transmitter and receiver. The transmitter produces a radio frequency carrier wave to carry the mark and space intelligence. A KEYER is needed to change the dc pulses from the tty into corresponding mark and space modulation for the carrier wave in the transmitter. The radio receiver and a CONVERTER are required to change the radio frequency signal back to dc pulses.

Radio Teletypewriter Systems

The Navy uses two basic radio teletypewriter (ratt) systems. These are the TONE-MODULATED SYSTEM, referred to as audio-frequency tone shift (afts), and the CARRIER-FREQUENCY SHIFT SYSTEM, referred to as radio-frequency-carrier shift (rfcs). The rfcs system is also called frequency-shift keying (fsk).

Figure 3-19 shows a modulated carrier wave with audio tone impulses impressed on the radio-frequency carrier wave. These correspond to dc mark and space signals.

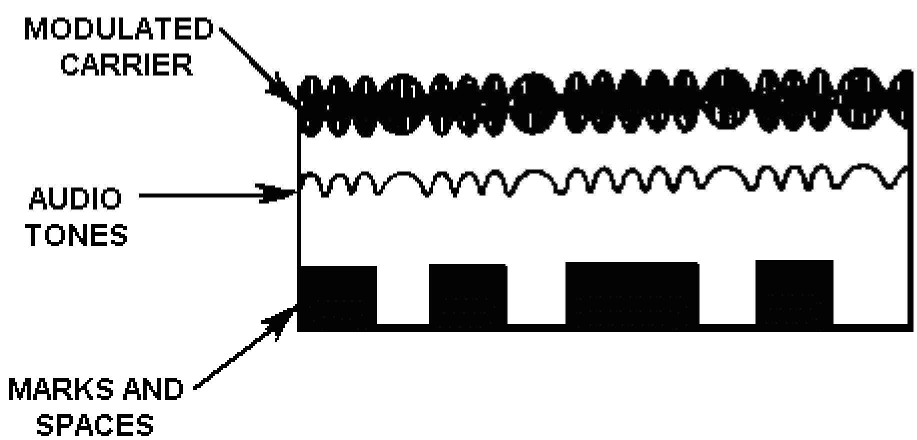


Figure 3-19.—Modulated carrier wave with audio tone for mark and space.

We can best explain the rfcs signal by comparing it to the on-off cw signal. Cw signals are essentially a constant frequency with no variations along the frequency axis. Figure 3-20, view A, is an example. The complete intelligence is carried as variations in the signal amplitude. Figure 3-20, view B, shows the same signal as a shift in frequency between the mark and space.

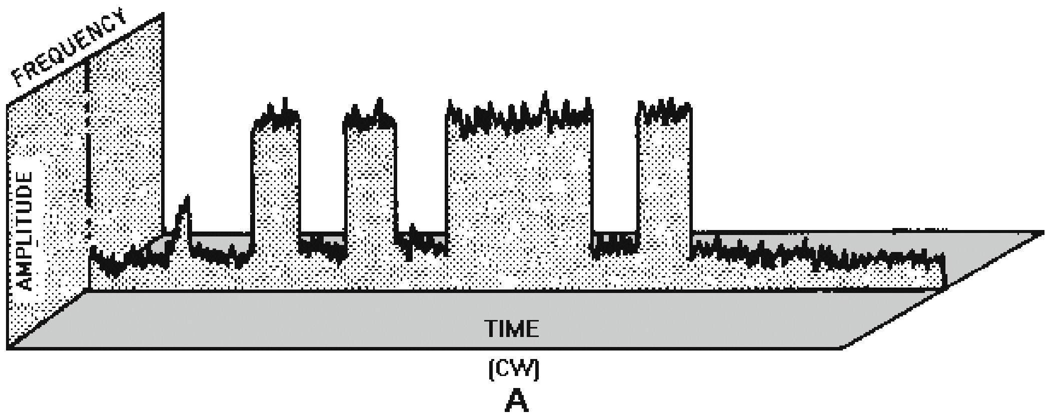


Figure 3-20A.—Cw compared to an rfcs teletypewriter signal.

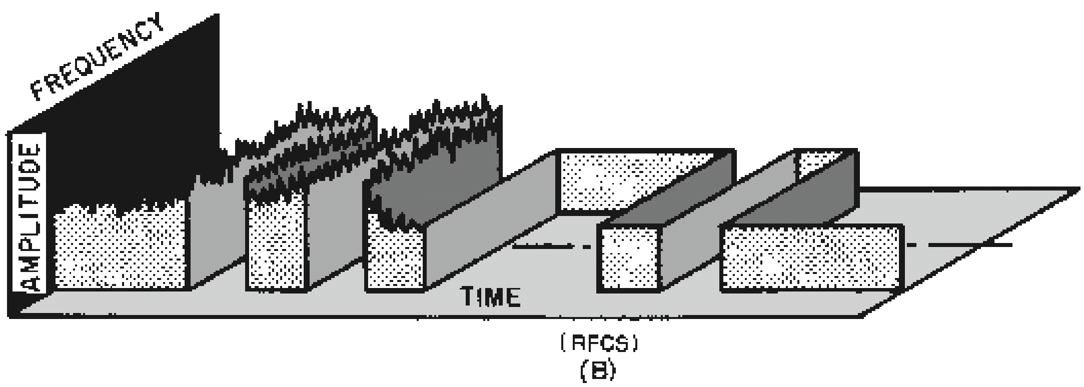


Figure 3-20B.—Cw compared to an rfcs teletypewriter signal.

AUDIO FREQUENCY TONE SHIFT.—Tone-modulated (afts) systems use amplitude modulation to change dc mark and space impulses into audio electrical impulses.

A basic tone-modulated system is shown in figure 3-21. Conversion to audio tones is accomplished by an audio oscillator in the tone converter. Rapid varying of the tone, according to the characters transmitted from the teletypewriter equipment, amplitude modulates the carrier wave in the transmitter. The receiver receives the modulated signal and separates the audio signal from the carrier. This process of separating the modulated signal is known as detection or demodulation.

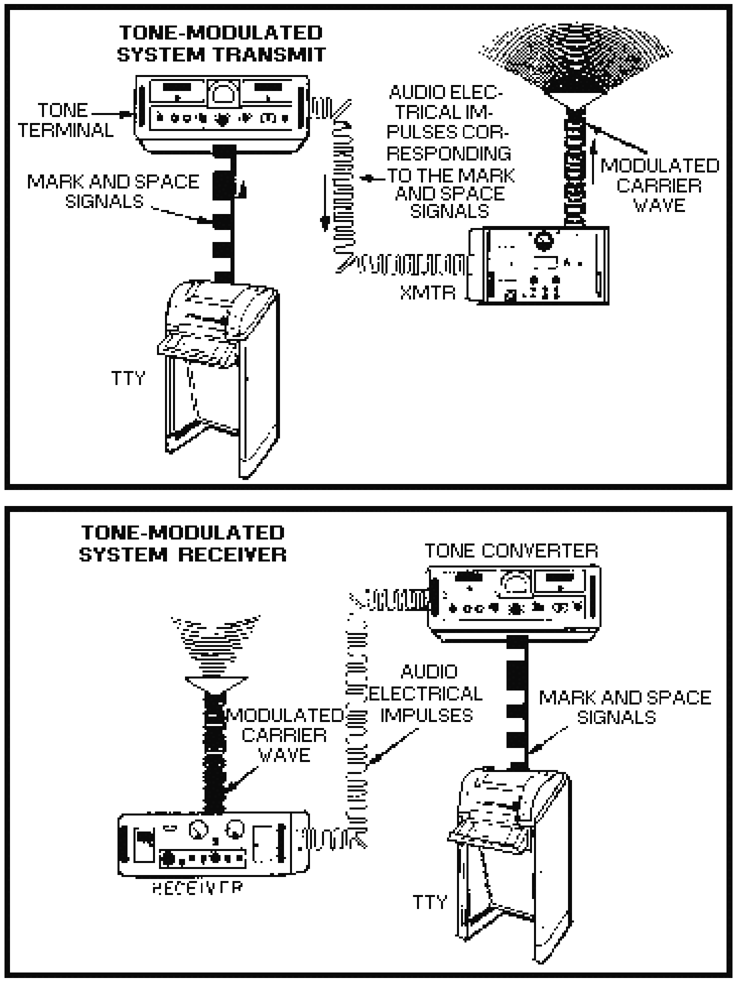


Figure 3-21.—Basic tone modulated (afts) system.

RADIO-FREQUENCY-CARRIER SHIFT.—For frequency-shift (fsk) systems, the transmitter provides a source of radio-frequency excitation. Figure 3-22 illustrates a basic frequency-shift keyed system. In modern systems, the keyer is built into the transmitter. The keyer shifts the signal box below or above the assigned frequency to correspond with the mark or space required to transmit tty characters. Normally the keyer is adjusted for an 850-hertz spread, 425 hertz above and 425 hertz below the assigned frequency. A spacing impulse will be 425 hertz above the operating frequency, and a marking impulse will appear 425 hertz below.

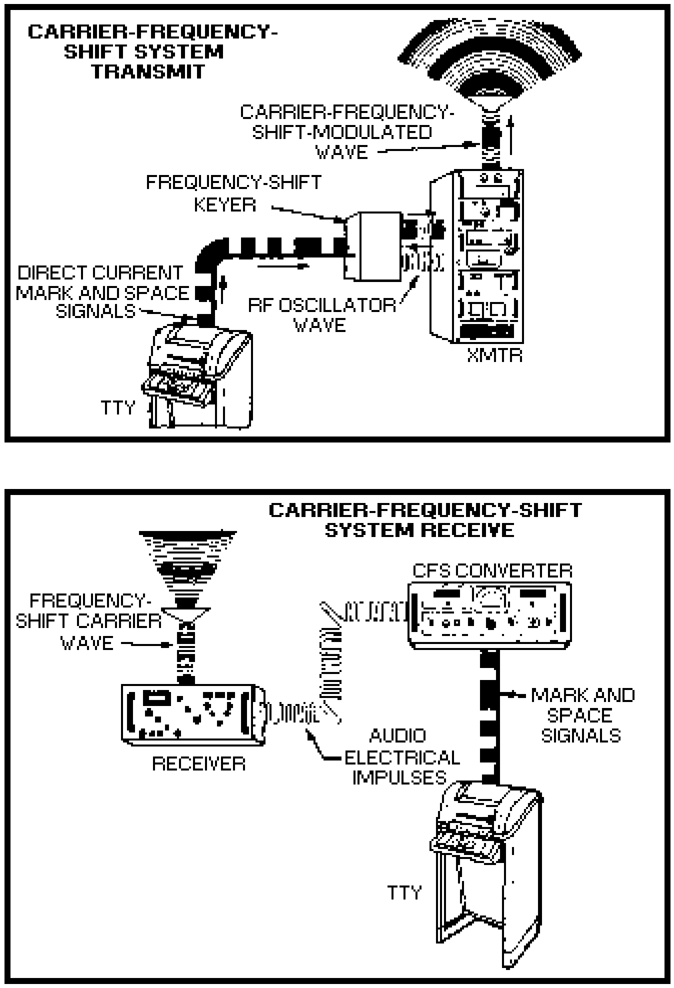


Figure 3-22.—Basic radio-frequency-carrier shift system (rfcs).

In both the tone-modulated system and the carrier-frequency shift system, all tty signals pass through the tty panel that controls the looping current in all the circuits. Looping current is the current supplied by the tty battery. The tty panel integrates the tone-modulated and the carrier-frequency shift systems. It provides every possible interconnection of available tty equipment. With this configuration maximum operational flexibility is achieved with the least amount of circuitry and equipment.

Q19. What is the function of a keyer?

Q20. What is the function of a converter?

Q21. Basically describe an afts system.

Q22. Basically describe an rfcs system.

Rfcs Send System

Figure 3-23 shows an rfcs teletypewriter transmit communications system. You should refer to this figure frequently while reading the functional descriptions of the equipment shown.

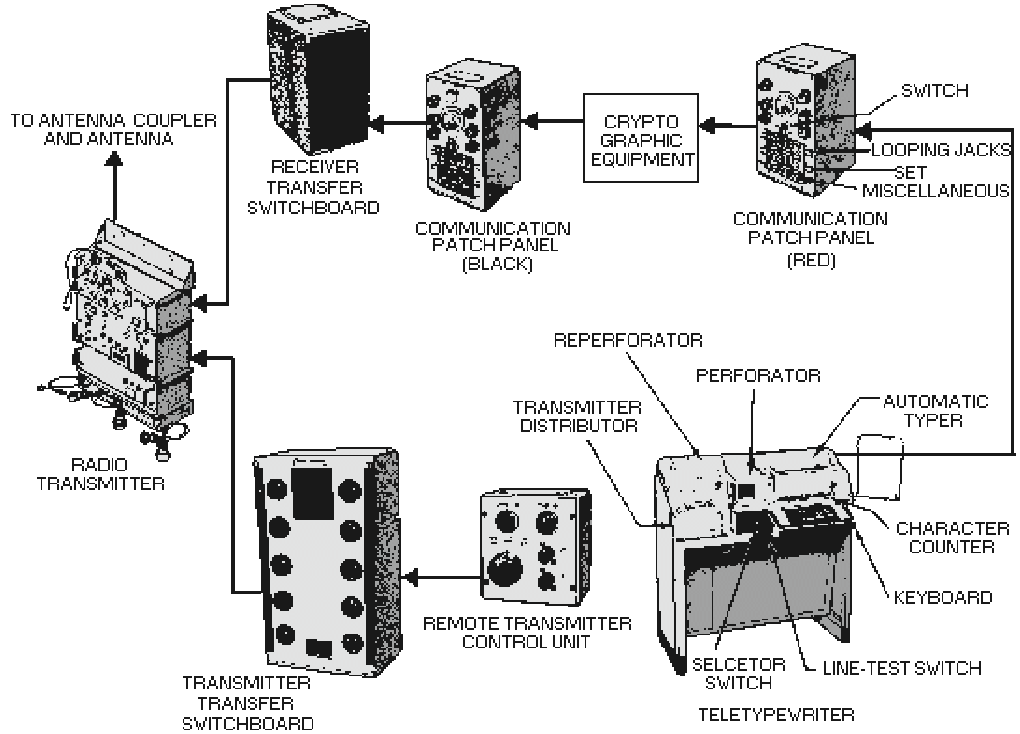


Figure 3-23.—Rfcs transmit (send) system.

TELETYPEWRITER SETS.—Most of the teletypewriter sets used by the Navy belong to one family of tty equipment. This equipment features various weights and sizes, quiet operation, and high operating speeds. They present relatively few maintenance problems. Because of this they are well suited for severe shipboard conditions of roll, vibration, and shock.

These teletypewriters operate at various speeds. Conversion from one speed to another is usually only a matter of changing the gears that are located within the equipment.

Teletypewriters may be send/receive units or receive only units. They may be designed as floor models, table models, or rack and wall-mounted sets. The teletypewriter shown is a send/receive floor model.

The teletypewriter receives messages and prints them on page-size copy paper. In addition, it can receive and record messages on perforated tape. You can use the keyboard or perforated tape to send messages. Page print monitoring is available with both methods. The set shown can prepare perforated and printed tape for separate transmission. It does this with or without simultaneous transmission and page-print monitoring. The combinations of services available are extensive.

The tty set may include a CABINET, KEYBOARD, PAGE PRINTER, TYPING PERFORATOR, TRANSMITTER DISTRIBUTOR, TYPING REPERFORATOR, power distribution panels, and a power supply.

In operation, the components are linked by electrical or mechanical connections. You are given a wide range of possibilities for sending, receiving, or storing tty messages. All equipment components are housed within the cabinet. Transmission signals are initiated through the keyboard (kybd) or through the transmitter distributor (td). Signals received or local transmissions can be monitored on the page printer. The typing perforator and typing reperforator are devices for preparing tapes on which locally initiated or incoming tty messages can be stored for future transmission through the td.

COMMUNICATION PATCHING PANELS.—Ttys are provided flexibility by jacks that are used to terminate all ttys and associated equipment. The jacks are wired in communications patching panels, usually referred to as tty patch panels. You are able to connect any combination of equipment electrically by means of patch cords.

The plugs on the patch cords are inserted into the jacks at the front of the panel. These plugs have three different parts. They are the tip, ring, and sleeve. The tip carries the intelligence signal while the ring carries the synchronizing (step) or timing signals. The sleeve carries an alarm signal that indicates (both visually and audibly) a problem to the operator. The problem may be equipment failure, loss of loop current, or improper patching. Commonly used combinations of equipment are often wired together within the panel (called normal-through). Individual pieces of equipment are wired on jacks to allow you to use them alone or in combination.

Tty patch panels also furnish a central point for connecting the dc voltage supply into the tty circuits. One source of supply can be used for all circuits passing through a particular panel.

RED and BLACK are used on patch panels to identify whether that panel is used for passing secure or nonsecure information. Red indicates that secure (encrypted) information is being passed through the panel. Black indicates that nonsecure (unencrypted) information is being passed. Patch panels through which secure information is passed are indicated by a red sign on the front that has inch high white block letters that say "RED PATCH PANEL." Panels through which nonsecure information is passed are indicated by two black signs on the front with inch high white block letters. One sign says "BLACK PATCH PANEL" and the other "UNCLAS ONLY."

Each panel contains six channels. Each channel has its own series circuit of looping jacks, set jacks, and a rheostat for adjusting line current. The number of looping and set jacks in each channel varies with the panel model. Each panel includes a meter and rotary selector switch for measuring the line current in any channel. There are six miscellaneous jacks. Any tty equipment not regularly assigned to a channel, may be connected to one of these jacks.

If the desired tty equipment is wired in the same looping channel as the radio adapter used, no patching is required. But, if the desired tty is not wired in the same looping channel as the keyer or converter, it must be patched. For example, let's put a tty on channel 1 and a converter on channel 3. If you want to receive, you must insert one end of the patch cord in the set jack for channel 1 and the other end in either one of the two looping jacks of channel 3.

In any switching operation between the plugs and jacks of a tty panel, the cord plug must be pulled from the looping jack before you remove the other plug from the set (machine) jack. Pulling the plug from the set jack first opens the circuits to the channel, causing all tty messages in the channel to be interrupted.

WARNING

Removing the set (machine) jack before the looping jack exposes a dangerous dc voltage on the exposed plug.

Q23. Most Navy tty sets operate at what speeds?

Q24. A receive tty set provides outputs in what formats?

Q25. What does the color red indicate on a tty patch panel?

CRYPTOGRAPHIC EQUIPMENT.—Cryptographic equipment is used to ENCRYPT and DECRYPT tty messages that require security handling. (Encrypting is the method used to code a transmitted message; decrypting is used to decode a received message.) To code or decode any message, the send and receive cryptographic equipment must be compatible.

REMOTE TRANSMITTER CONTROL UNIT.—The remote transmitter-control unit is mounted close to the kybd and permits remote control of the transmitter (xmtr). It has a transmitter power on-off switch, a power-on indicator lamp, a carrier-on indicator lamp, and a three-position rotary selector switch. For rfcs operation you set the switch to CFS SEND to transmit and to CFS REC to receive. Use the TONE S/R position for both transmitting and receiving afts signals.

An audio frequency tone-shift system will be discussed later in this chapter.

TRANSMITTER TRANSFER SWITCHBOARD.—The transmitter transfer switchboard is used in this system to connect the remote transmitter control unit to the radio transmitter.

RADIO TRANSMITTER.—The radio transmitter transmits the tty signal. You should be careful when tuning the transmitter for rfcs operation. The carrier frequency setting is critical and must be properly set to ensure a correct output from the transmitter.

Q26. What are the functions of cryptographic equipment?

Rfcs Receive System

Figure 3-24 shows the rfcs receive system used to receive the transmitted signal and translate it back to a usable output. You should look at this figure while studying the units in this section.

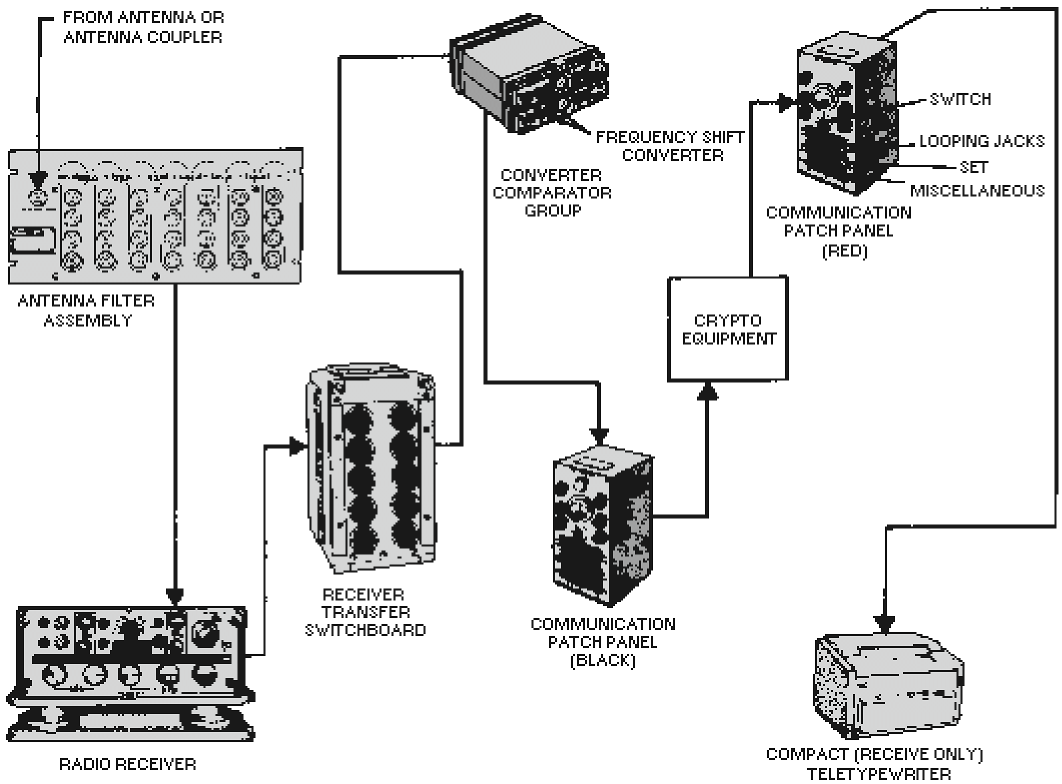


Figure 3-24.—Rfcs receive system.

ANTENNA FILTER ASSEMBLY.—The antenna filter assembly is connected to the antenna and receives the rf signal from the antenna. It filters out any unwanted rf signals and allows the desired band of frequencies to pass.

RADIO RECEIVER.—The radio receiver takes the rf signal passed on by the antenna filter and translates it to an audio signal.

RECEIVER TRANSFER SWITCHBOARD.—The receiver transfer switchboard is used to tie the receiver to any converter unit connected to it. This allows you a wide selection of equipment for connection to the same receiver.

CONVERTER-COMPARATOR GROUP.—The converter-comparator group is used with receivers in either space or frequency diversity operation. When diversity operation is not required, each converter can be used separately with a single receiver.

Each converter has its own COMPARATOR circuitry. This built-in design feature results in a considerable reduction in size from older units. The comparator was located in a separate chassis in the older units. Size has been further reduced through the use of microelectronics.

Figure 3-25 shows the basic method we use to convert a frequency-shift rf signal into a signal that controls the dc loop of a tty. The frequency shifts of the af output from the receiver are converted into dc pulses by the af discriminator. The dc pulses are then fed into the keyer. The keyer opens and closes the dc loop of the tty according to the mark and space characters received.

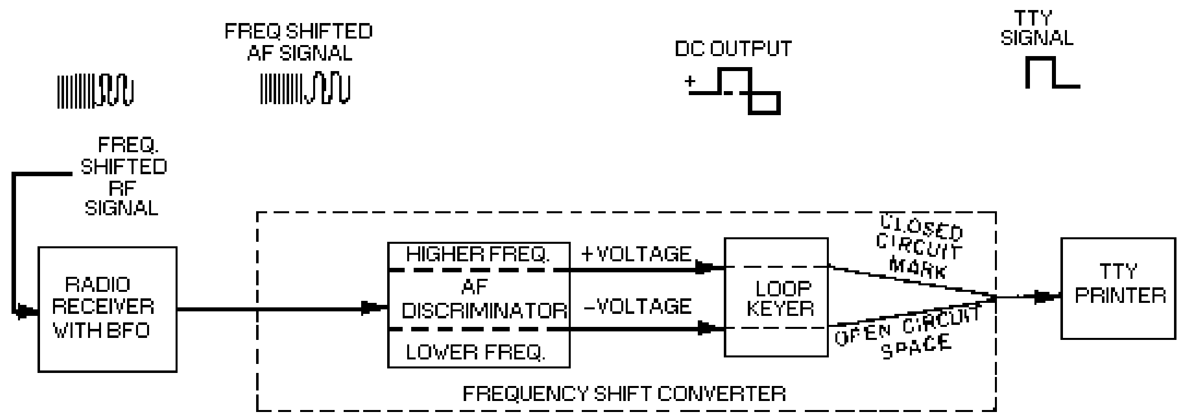


Figure 3-25.—Frequency shift receiving system simplified block diagram.

In diversity operation the comparator section of the converter-comparator group (shown in figure 3-24) compares the strength of the signals from two receivers. Signals from each converter are fed into a comparator circuit that compares the signals. This comparison is displayed on a crt on the front of the equipment. The comparison is in the form of LISSAJOUS PATTERNS. A lissajous pattern is a combined, simultaneous display of the amplitude and phase relationships of two input signals. One signal is applied to the vertical and the other to the horizontal deflection circuits. Lissajous patterns have many applications in electronics. They have operational uses as well as uses in corrective and preventive maintenance. Further coverage on lissajous patterns can be found in NEETS, Module 19, *The Technician's Handbook*. Figure 3-26 shows several typical lissajous monitoring patterns for the converter-comparator group. Once we have a correctly tuned signal, the comparator feeds it to the communication patching panel for patching to the tty. Now let's refer back to figure 3-24 while we discuss the rest of the units in the system.

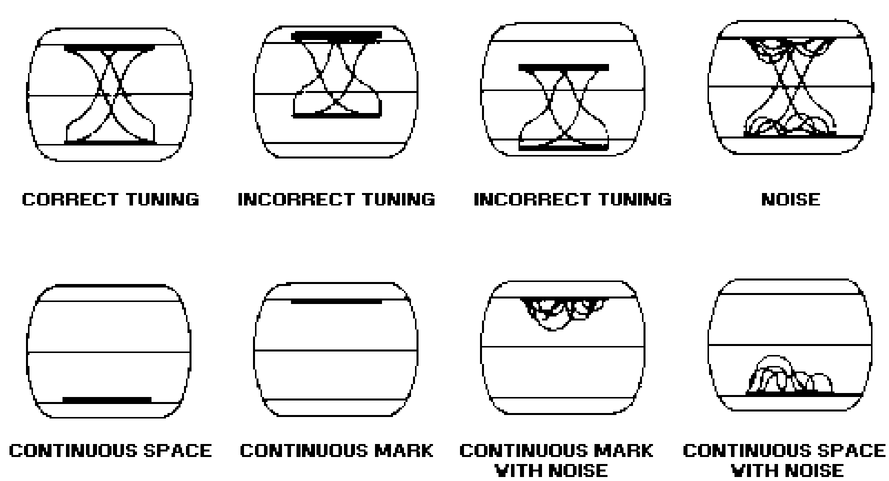


Figure 3-26.—Typical lissajous monitoring patterns.

COMMUNICATION PATCH PANEL.—The communication patch panel serves the same functions on the receive side of the rfcs system as it did on the transmit side. It routes the dc signal to the proper cryptographic equipment. It also routes the decoded teletypewriter signal from the cryptographic equipment to the selected tty.

CRYPTOGRAPHIC EQUIPMENT.—The cryptographic equipment converts the transmitted coded signal to a decoded signal that can be printed out in its original state.

TELETYPEWRITER.—The tty equipment is used to convert the dc signal received from the communication patch panel to a printed copy of the original transmitted message. The tty shown is used only for receive and does not have the ability to transmit.

Q27. What are the functions of a converter-comparator group?

Afts System

Figure 3-27 is a simplified block diagram of a HALF-DUPLEX (send or receive) uhf, audio-frequency-tone shift system. A half-duplex communications circuit permits two-way communications between stations. Communications can be in either direction but not simultaneously. The term half-duplex is qualified by adding send only, receive only, or send or receive. Let's use the block diagram to trace a signal through the system.

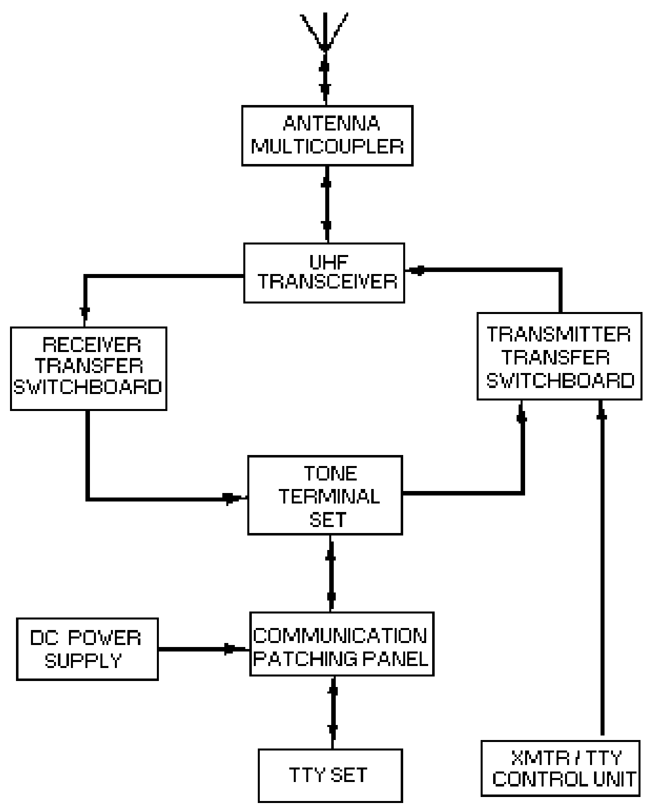


Figure 3-27.—Half-duplex afts teletypewriter system.

SIGNAL FLOW.—On the transmit side, dc signals from the tty set are fed to the communication patching panel. From the panel they are patched to the tone terminal set. The tone terminal set converts the dc signals into audio tone-shift signals. These signals are then patched to the transmitter section of the transceiver through the transmitter transfer switchboard. The audio tone-shift signals modulate the rf carrier generated by the transmitter (xmtr). The rf tone-modulated signals are then radiated by the antenna.

On the receive side, the rf tone-modulated signals are received at the antenna. You then patch the signal via the multicoupler to the receiver section of the transceiver. Demodulation takes place at this point. The resulting audio tone-shift signals are then patched through the receiver transfer switchboard. The signals now go from the switchboard to the tone terminal set, where they are converted back to dc signals. The dc signals are then patched through the communication patching panel to the tty for printing.

TONE TERMINAL SET.—In tone modulation transmission, the tty pulses are converted into corresponding audio tones. These tones amplitude modulate the rf carrier in the transmitter. Conversion to audio tones is accomplished by an audio oscillator in the tone converter.

An internal relay in the tone converter closes the control line to the transmitter. This keys the transmitter on the air when the operator begins typing a message. The transmitter remains keyed until after the message has been transmitted.

On the receive side, the tone converter accepts the mark and space tones coming in from a receiver and converts them into signals suitable to operate a relay in the converter. The make and break contacts of the relay are connected in the local tty dc loop circuit. This causes the teletypewriter to print in unison with the mark and space signals from the distant tty.

Multiplexing Equipment

The number of communications networks in operation throughout any given area is increasing. As a result, all areas of the rf spectrum have become highly congested.

The maximum number of intelligible transmissions taking place in the radio spectrum is being increased through the use of MULTIPLEXING. Multiplexing is the simultaneous transmission of a number of intelligible signals (messages) in either or both directions using only a single rf carrier. You may use two methods of multiplexing. These are TIME-DIVISION and FREQUENCY-DIVISION.

TIME-DIVISION.—With AM voice and tone communications, we want to transmit and receive for 360 degrees of each sine wave. However, an audio signal may be transmitted and received satisfactorily by periodically sampling the signal. The sampling process yields a received signal like the one shown in figure 3-28. There is no limit to the maximum number of samples that may be made, but you must sample at least twice per cycle of audio to get satisfactory results. In practical systems, 2.4 samples per cycle are usually taken. This concept of sampling forms the basis for time-division multiplex (tdm) operation.

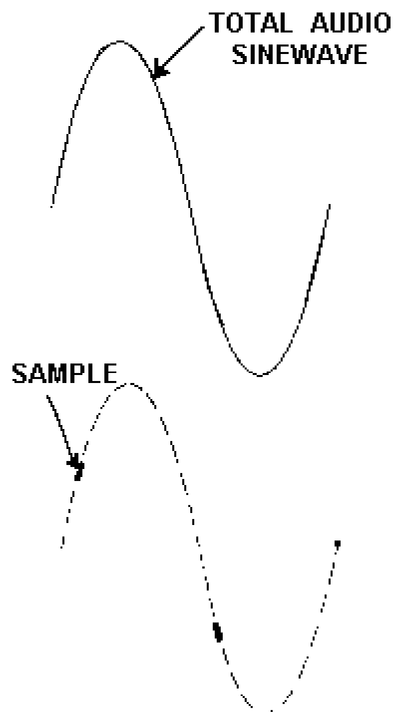


Figure 3-28.—Components of a sine wave.

Figure 3-29, view A, illustrates, the fundamental principle of tdm. Let's look at an example. Assume that a 3,000-hertz tone is applied to each of the six channels in the transmitter. Assume also that the rotating switch turns fast enough to sample, in turn, each of the six channels 2.4 times during each cycle of the 3,000-hertz tone. The speed of rotation of the switch must then be $2.4 \times 3,000$ or 7,200 rotations per second. This is the optimum sampling for a practical system.

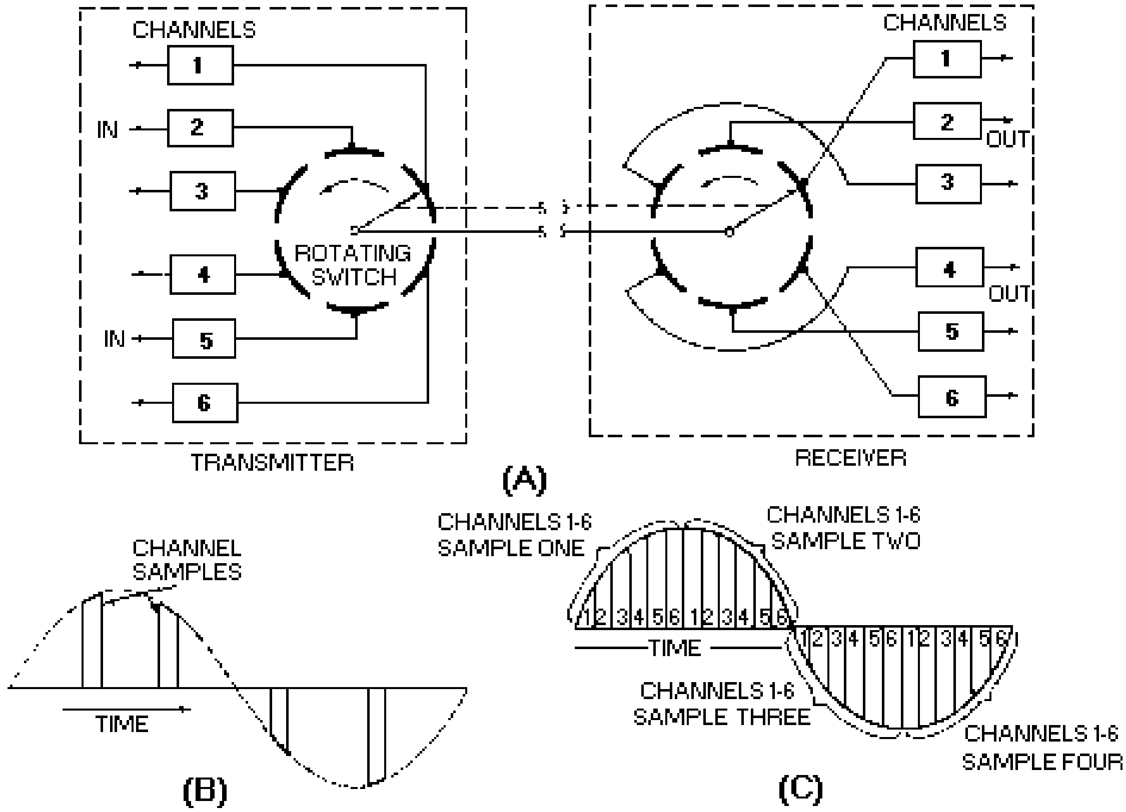


Figure 3-29.—Fundamental principle of time-division multiplexing.

When the transmitter and receiver switches are synchronized, the signals will be fed in the proper sequence to the receiver channels. The samples from transmitter channel one will be fed to receiver channel one. In this way, many channels of audio are combined to form a single output (multiplexed) chain. Time spacing occurs between the components of the separate channels. The chain is transmitted (via wire or radio path) to distant demultiplexing receivers. Each receiving channel functions to select and reconstruct only the information included in the originally transmitted channel.

In most present day applications, electronic switching is used as the sampling component. The main advantage to electronic sampling is the longer life of an electronic switch when compared to an electromechanical switch. We use a mechanical system in our example to make this concept easier for you to see.

Now let's look at figure 3-29, view B, where channel one is shown sampled four times. (This is the output of channel one in our transmitter.) Figure 3-29, view C, shows all six channels being sampled four times during each cycle. (This is the output of the rotating switch in our transmitter.) What you see here is a continuous, time-sharing waveform.

More than six channels (perhaps 24 or more) may be used. As we increase the number of channels, the width of each sample segment must be reduced. The problem with reducing the width of the pulse is that the bandwidth (bw) necessary for transmission is greatly increased. Decreasing the pulse width decreases the minimum required rise time of the sampling pulse and increases the required bandwidth. When you increase the number of channels, you increase the bw. The bw is also affected by the shape of the sampling pulse and the method used to vary the pulse.

Common methods of time-division multiplexing include PULSE AMPLITUDE MODULATION (pam), PULSE WIDTH or PULSE DURATION MODULATION (pwm or pdm), PULSE POSITION MODULATION (ppm), and PULSE CODE MODULATION (pcm). We have been studying an example of pulse amplitude modulation. (These methods of tdm were discussed in NEETS, Module 12, *Modulation Principles*.)

FREQUENCY DIVISION.—Frequency division multiplexing (fdm), unlike tdm, transmits and receives for the full 360 degrees of a sine wave. Fdm used presently by the Navy may be divided into two categories. One category is used for voice communications and the other for tty communications.

The normal voice speaking range is from 100 to 3,500 hertz. During single channel AM voice communications, the audio frequency amplitude modulates a single rf (carrier frequency). However, in voice fdm, each voice frequency modulates a separate frequency lower than the carrier frequency (subcarrier frequency). If these subcarrier frequencies are separated by 3,500 hertz or more, they may be combined in a composite signal. This signal modulates the carrier frequency without causing excessive interference.

In figure 3-30, the output of channel one is the voice frequency range of 100 to 3,500 hertz. The output of channel two is the combination of a different voice frequency with a subcarrier frequency of 4,000 hertz. The output of channel three is another voice frequency. This voice frequency combined with a subcarrier frequency of 8,000 hertz gives you an output frequency range of 8,100 to 11,500 hertz. The overall bw for the composite modulation package shown is 100 to 15,500 hertz. Each separate channel occupies its own band of frequencies. The composite signal is used to modulate the carrier frequency of the transmitter.

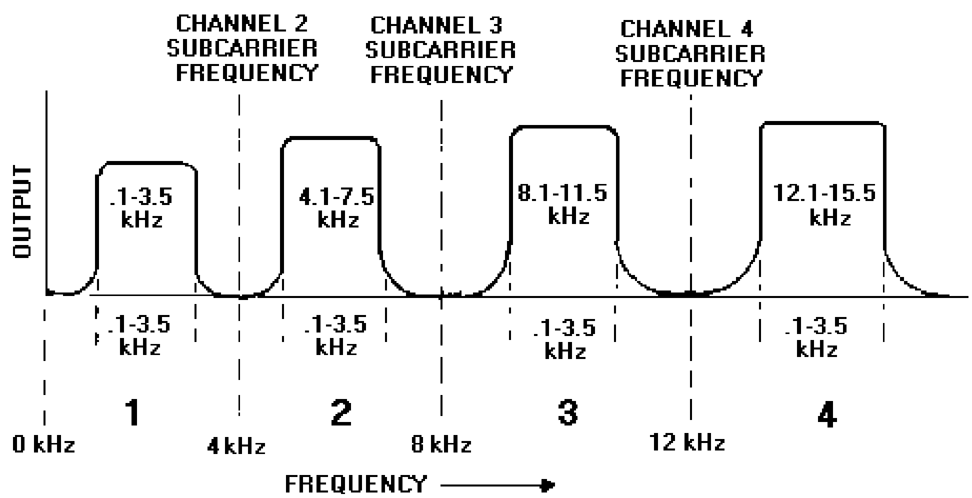


Figure 3-30.—Block diagram of a frequency-division multiplexing system.

Multichannel broadcast and ship/shore terminations use tty fdm. With this system, each channel of the composite tone package of the broadcast is assigned an audio frequency. By multiplexing tty circuits, up to 16 circuits may be carried in any one of the 3,000 hertz multiplexed channels described above. Don't confuse the two types of multiplexing. In the first case, 3,000 hertz audio channels have been combined. In the second case, a number of dc tty circuits are converted to tone keying and combined in a single 3,000-hertz audio channel. Figure 3-31 illustrates a 16-channel, tty-multiplexing system. The output of the dc pulsed circuits is converted to audio keying. Each channel has a separate audio center frequency. Channel frequencies range from 425 hertz for the lowest channel to 2,975 hertz for the highest channel. A mark in an individual tty loop keys an audio tone 42.5 hertz below the center frequency. A space in the input signal keys an audio tone 42.5 hertz above the center frequency. Let's look at an example. The mark and space frequencies for channel one are calculated as 382.5 hertz and 467.5 hertz, respectively (425 ± 42.5). Combining these keyed tones into a composite signal results in a tone package within a standard 3,000-hertz bandwidth. By occupying no more than 3,000 hertz of the audio spectrum, the output signal is suitable for transmission via radio or landline.

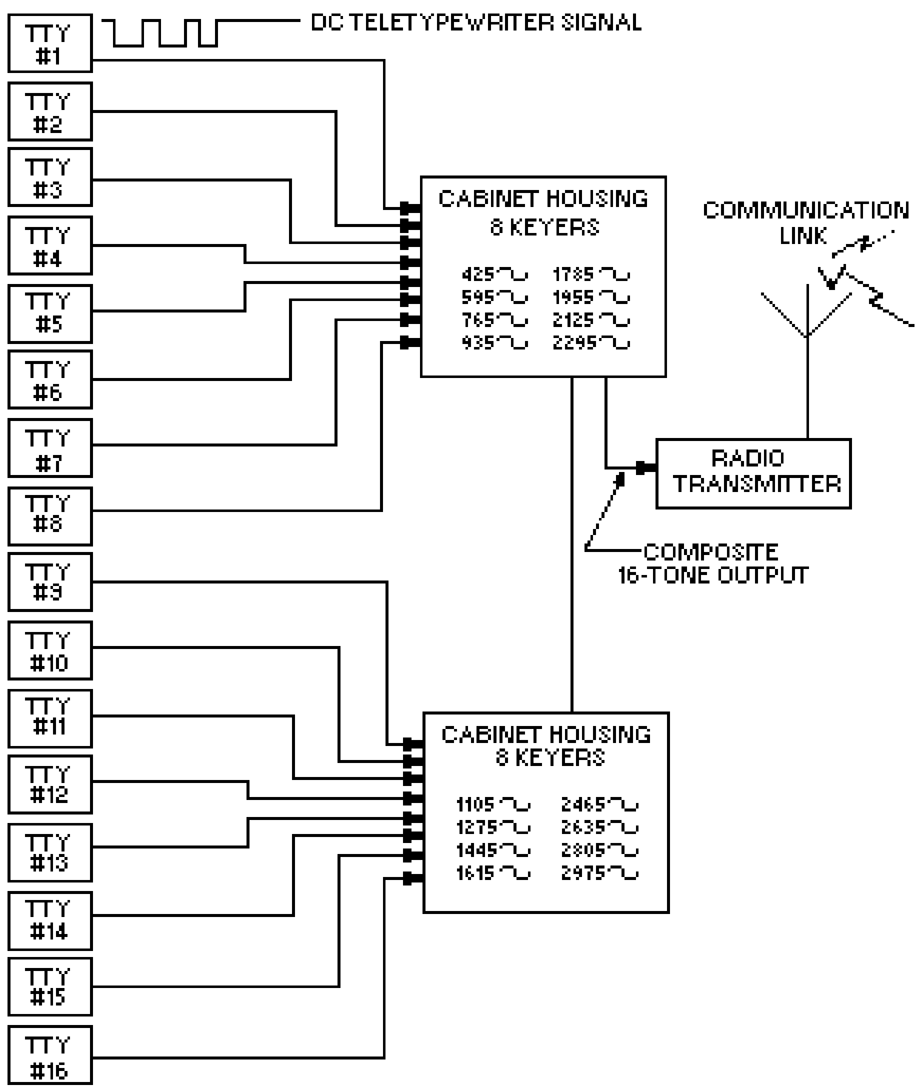


Figure 3-31.—Block diagram of modulator units.

Q28. What is the function of a tone terminal set?

Q29. What are the two types of multiplexing?

Q30. What is the purpose of multiplexing?

Facsimile

FACSIMILE (fax) is a method of transmitting still images over an electrical communications system. The images, called "pictures" or "copy" in fax terminology, may be weather maps, photographs, sketches, typewritten or printed text, or handwriting. Figure 3-32 shows a facsimile transceiver. You must realize that the still image serving as the fax copy or picture cannot be transmitted instantly in its entirety. Three distinct operations are performed. These are (1) scanning, (2) transmitting, and (3) recording or receiving.

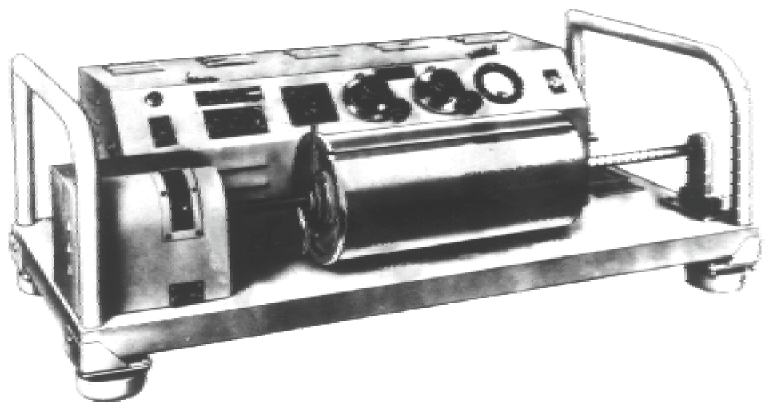


Figure 3-32.—Facsimile transceiver.

Scanning consists of subdividing the picture in an orderly manner into a large number of segments. This process is accomplished in the fax transmitter by a scanning drum and phototube arrangement.

The picture you want to transmit is mounted on a cylindrical scanning drum. This drum rotates at a constant speed and at the same time moves longitudinally along a shaft. Light from an exciter lamp illuminates a small segment of the moving picture and is reflected by the picture through an aperture to a phototube. During picture transmission, the light crosses every segment of the picture as the drum slowly spirals past the fixed lighted area.

The amount of light reflected back to the phototube is a measure of the lightness or darkness of the segment of the picture being scanned. The phototube changes the varying amounts of light into electrical signals. These are used to amplitude modulate the constant frequency output of a local oscillator. The modulated signal is then amplified and sent to the radio circuits.

Signals received by the fax receiver are amplified and actuate a recording mechanism. This recorder makes a permanent recording (segment by segment) on paper. The paper is attached to a receiver drum similar to the one in the fax transmitter. The receiver drum rotates synchronously with the transmitter drum. Synchronization of the receiver and transmitter is done to reduce distortion. Synchronization is obtained by driving both receiver and transmitter drums with synchronous motors operating at the same speed. Drum rotation continues until the original picture is reproduced. The recording mechanism may reproduce the picture photographically by using a modulated light source shining on photographic paper or film. It may also reproduce directly by burning a white protective coating from specially prepared black recording paper.

The receiver drum is FRAMED with respect to the transmitter drum by a series of phasing pulses that are transmitted just before transmission. The pulses operate a clutch mechanism that starts the scanning drum in the receiver. This ensures proper phasing with respect to the starting position of the scanning drum in the transmitter.

Figure 3-33 is a block diagram of the equipment necessary for radio facsimile operation. View A shows the receiving system. This system consists of a standard radio receiver, a frequency-shift converter, and a facsimile recorder. View B shows two systems for transmitting TIF signals. The upper row of blocks is for carrier-frequency shift transmission. This system consists of a facsimile transceiver, a keyer adapter, a frequency shift keyer and a transmitter capable of fsk emission. The lower row of blocks is for audio-frequency shift transmission and uses a fax transceiver, a radio modulator, and an AM transmitter.

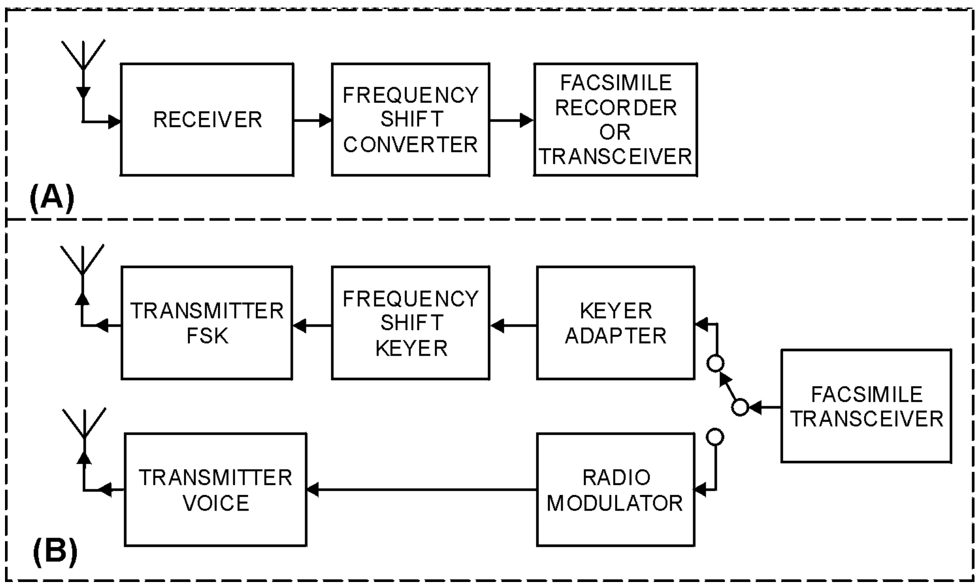


Figure 3-33.—Radio facsimile systems.

SECURITY, QUALITY MONITORING, AND SAFETY

Security, quality monitoring, and safety are important areas that you must be aware of. If the fundamentals are followed, you will see higher quality communications. You will also help meet the communications goals of the Navy. Let's find out what these fundamentals are and what they will do for you.

TEMPEST

Compromising emanations (ce) are, generally referred to as TEMPEST. These signals may be unintentional, data-related, or intelligence-bearing. If intercepted or analyzed, these signals could disclose classified information. TEMPEST problems are associated with material transmitted, received, handled, or otherwise processed by electrical information processing equipment or systems. Any electrical information processing device may cause problems. Even your electric typewriter or a large, complex data processor may emit interceptable compromising emanations. Some countermeasures taken to ensure against TEMPEST problems are listed below:

- Design of equipment in which ce is suppressed
- Approved installation criteria that limits interaction between classified and unclassified signal lines, power lines, grounds, equipment, and systems
- Low level keying and signaling
- Shielded enclosures for equipment installations
- Proper shipboard grounding of equipment, including ground straps

TRANSMISSION SECURITY

Transmission security includes all measures designed to protect transmission from interception, traffic analysis, and imitative deception. Every means of transmission is subject to interception. In radio transmission, it should be assumed that all transmissions are intercepted.

Speed Versus Security

Three fundamental requirements of a military communications system are *reliability, security*, and *speed*. Reliability is always first. Security and speed are next in importance and, depending on the stage of an operation, are interchangeable. During the planning phase, security is more important than speed. During the execution phase, speed sometimes passes security in importance.

Radio Transmission Security

When a message is transmitted by radio, the originator may know some of those who are receiving it, but will never know *all* of those who are receiving the message. You must assume that an enemy receives every transmission. Property prepared messages using modern cryptographic systems may prevent an enemy from understanding a message. However, they can still learn a lot. For example, as time for a planned operation approaches, the number of messages transmitted increases. An enemy then knows that something will occur soon, and their forces are alerted. Strict radio silence is the main defense against radio intelligence.

The amount of radio traffic is not the only indicator used by an enemy. Statistical studies of message headings, receipts, acknowledgments, relays, routing instructions, and services are also used by an enemy. Communications experts can often learn much about an opponent from these studies. Direction finders are another aid the enemy can use to determine where messages originate.

Radiotelephone Security

Radiotelephone networks are operated so frequently that many operators tend to be careless. There are too many instances of interception of vhf and uhf transmissions at distances of many thousands of miles. You may have occasion to work on or around this type of equipment. If you are ever required to bring any transmitter on the air for any purpose, you must be familiar with and use all the correct procedures.

Q31. The transmission of still images over an electrical communications system is known as what?

Q32. The term TEMPEST refers to what?

Q33. What are the three fundamental requirements of a military communications system?

Q34. Which of the above requirements is most important?

SHIPBOARD COMMUNICATIONS SYSTEMS QUALITY MONITORING (QMCS)

In recent years the volume of shipboard communications has increased greatly. This expansion has led to the shipboard installation of sophisticated equipment. Factors such as frequency accuracy and dc signal distortion are critical to the operation of communications systems. These systems demand precise initial lineup and monitoring to ensure satisfactory operations are maintained. System degradation is often caused by many small contributing factors. When these factors are added together, the system becomes unusable.

Scheduled Maintenance

When you perform scheduled, logical checks that ensure continuous, optimum performance of shipboard communications systems, you are doing SCHEDULED MAINTENANCE. In many cases this maintenance prevents outages before they occur. Some of the scheduled checks will include the following:

- Transmitter/receiver frequency
- Transmitter power out
- Receiver sensitivity/bandwidth
- Primary power (voltage, current, cycles)

Electromagnetic Interference (emi)

Many complex electronic systems are installed aboard Navy ships. In modern ships, complex systems with higher power and greater sensitivity are being crowded into a restricted and corrosive area. Figure 3-34 is a Spruance class destroyer with its crowded (compact) communications environment. The ability of these systems to perform their individual functions without interference is known as ELECTROMAGNETIC COMPATIBILITY (emc). Emc is concerned with the structure of the ship and its electrical and electronic system. Compact environment is a major limitation to the effectiveness of a total ship system concept.

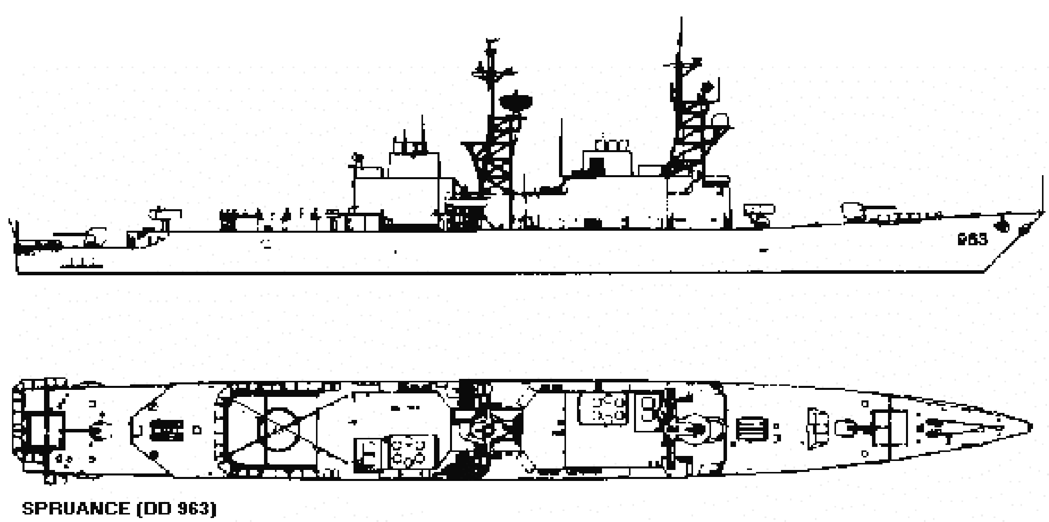


Figure 3-34.—Total ship.

Operation of a total ship system in this unique shipboard environment presents a challenge to all concerned. You must always consider the effects that motion, temperature variations, and exposure to adverse elements will have on the performance of the total ship system. This is particularly true on those system components that are mounted topside.

On board ship, you will find much attention is given to keeping the topside cosmetically and mechanically shipshape. It is equally important to keep it electronically shipshape. Minor mechanical problems, such as loose connections, broken bond straps, or rusty junctions can cause serious communications problems. These sources of electromagnetic radiations reduce receiver performance and are known as ELECTROMAGNETIC INTERFERENCE (emi). Sources of emi can be divided into the following broad categories:

- Functional. Emi can originate from any source designed to generate electromagnetic energy and which may create interference as a normal part of its operation. The interference may be unintentional or caused by other on board or adjacent platform systems. This interference also may be intentional or caused by electronic countermeasures (ECM).
- Incidental. Emi can originate from man-made sources. These are sources not designed specifically to generate electromagnetic energy but which do in fact cause interference. Examples of incidental emi sources include power lines, motors, and switches.
- Natural. Emi can be caused by natural phenomena, such as electrical storms, rain particles, and solar and interstellar radiation. It is recognized by the following audible noises:
  - Intermittent impulses of high intensity that are caused by nearby electrical storms
  - Steady rattling or cracking caused by distant electrical storms
  - Continuous noise of precipitation static caused by electrically charged rain drops
  - A steady hiss at high frequencies caused by interstellar noise
- Hull-generated. Emi can be caused by the interaction of radiated signals with elements of the hull and rigging of a ship. (The functional signals themselves do not cause interference.)

The following are two general methods by which emi is transmitted:

Conduction. Undesired energy from one equipment is coupled to interconnecting cables or components of another equipment. This energy is conducted via the wiring in the shielded enclosure that protects sensitive circuits. You will find proper design, adequate isolation, and shielding of cables and equipment can control this problem.

Radiation. Energy is beamed directly from the transmitting antenna, or source, to the victim receiving antenna. When this interference is picked up by a receiver, you have two solutions. Interfering energy can be eliminated at the source or you can filter, or blank it out at the victim equipment. Filtering is far less desirable. Interference may be on the same frequency as the desired signal and will not be eliminated without affecting the reception of all desired signals.

Most unprotected shipboard receivers are susceptible to emi over a frequency range much wider than their normal bandpass. Off-frequency rejection rarely excludes strong, adjacent-channel signals. These signals enter the receiver and degrade receiver performance by being processed along with the desired, tuned signal. Usually, the presence of emi will be apparent to you. It has a bad effect. Upon the desired signal quality, such as that in CROSS-MODULATION where a spurious response occurs when the carrier of a desired signal intermodulates with the carrier of an undesired signal. Extremely strong, off-frequency signals may even burn out the sensitive front-end stages of a receiver. Emi also can degrade overall receiver performance in a less noticeable way. It does this by desensitizing the receiver front end. The noise level is raised and effectively lowers the signal to noise ratio and thus the sensitivity. This causes a decrease in desired signal amplification. For these reasons, shipboard receive systems are designed to include protective circuitry between the antenna and receiver to filter out off-frequency signals. This prevents or limits interference, desensitization, or burnout. Depending upon the system, these protective devices may include filters, multicouplers, preselectors, and so forth. These devices can minimize interference caused by inadequate frequency separation or poor physical isolation between transmit and receive antennas.

Q35. What is the purpose of QMCS?

Q36. What is emi?

Q37. What are the two emi transmission methods?

ELECTROMAGNETIC RADIATION

Radio-frequency (rf) transmitting systems with high-power transmitting tubes and high-gain antennas have increased the possibility of injury to personnel working in the vicinity.

An electromagnetic radiation hazard exists when electronic equipment generates a strong enough electromagnetic field to fall in a category listed below:

- Causes harmful or injurious effects to humans and wildlife
- Induces or otherwise couples currents and/or voltages of magnitudes large enough to initiate electroexplosive devices or other sensitive explosive components of weapons systems, ordnance, or other explosive devices
- Creates sparks large enough to ignite flammable mixtures or materials that must be handled in the affected areas

These hazardous situations can be caused by a transmitter or antenna installation. These generate electromagnetic radiation in the vicinity of personnel, ordnance, or fueling operations in excess of established safe levels. Sometimes the existing electromagnetic radiation levels increase to a hazardous level. When personnel, ordnance, or fueling evolutions are located in an area that can be illuminated by electromagnetic radiation, hazardous situations may occur.

Electromagnetic radiation is hazardous to personnel in two ways. It can cause rf burns; and it can cause biological, thermal, and neurological effects to personnel (RADHAZ). Because of the differences in characteristics and safety precautions required for each of the two types, they will be discussed separately.

An rf burn hazard is a hazardous condition caused by the existence of radio frequency (rf) voltages in places where they are not intended to be. Any ship with high-power hf transmitters is susceptible. Potentially hazardous voltages have been found in many areas. Some of these areas are lifelines, vertical ladders, ASROC launchers, gun mounts, rigging for underway replenishment, and boat davits. Another of these areas is on aircraft tied down on carrier and helicopter flight decks.

Whether or not an induced voltage creates an rf burn hazard depends on whether personnel will come into contact with the object being energized. Generally, only the voltage between an object and the deck is important. The rf burn occurs when a person comes into contact with a source of rf voltage in a manner that allows rf current to flow through the area of contact. Resistance of the skin to the current flow at the areas of contact causes heat. The effect of the heat on a person at the point of contact ranges from noticeable warmth to a painful burn.

The most useful and widespread technique in the reduction of rf burn hazards is the proper bonding and grounding of all metallic objects in the rf radiation field.

In some cases, the rf burn hazard can be eliminated only through the use of restrictive operating procedures. These procedures govern the simultaneous use of transmitting and cargo equipment. Techniques such as operation of transmitters at reduced power and the prohibition of simultaneous use of certain combinations of antennas, frequencies, and cargo handling equipment are used.

Figure 3-35 shows typical rf radiation hazard warning signs.

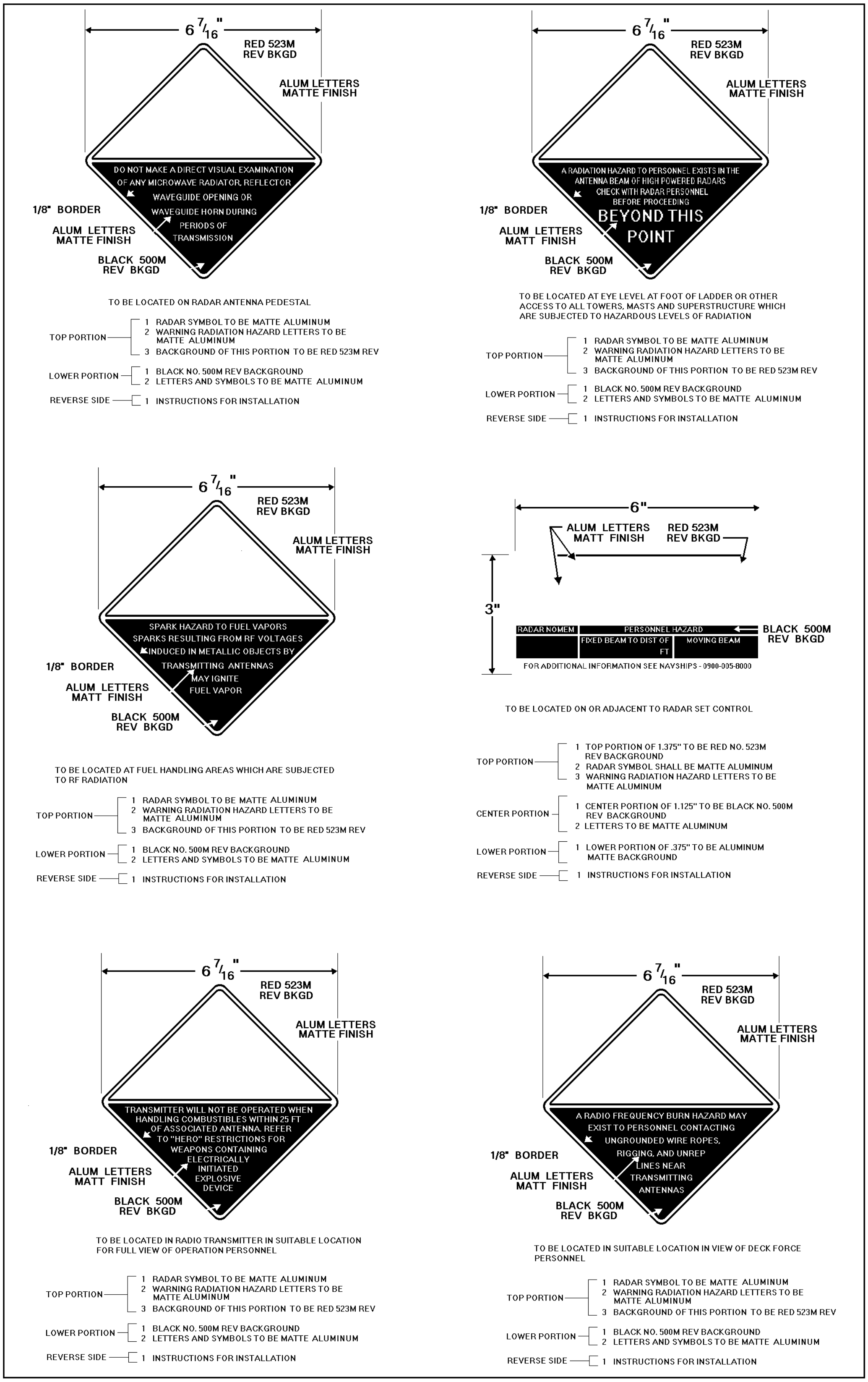


Figure 3-35.—Typical rf radiation hazard warning signs.

Most studies on the subject of radiation hazards (RADHAZ) have emphasized the impact of electromagnetic radiation on man. Man is singled out because of the biological, thermal, and neurological effects that occur in human organs and other biological tissues. Certain organs of the body are considered to be more susceptible than others to the effects of electromagnetic radiation. Presently available information and experience indicate that the eyes and testes are the most vulnerable body organs. The overwhelming danger to date appears to be the hazard from thermal effects, which are a function of intensity of radiation and frequency. This is particularly true in the range of 1 to 3 gigahertz. Thermal effects appear to taper off in severity outside this range.

When the body is irradiated by energy from a point source, the total body surface is usually not exposed. The larger the area exposed and the larger the radiation power density, the higher the body temperature rise and the greater the hazard. Microwave radiation from a radar source will "cook" you internally, just as a microwave oven cooks a chicken.

An injury of great concern is that to the lens of the eye. Exposure of the lens to high-intensity microwaves may cause cataracts. Current medical evidence indicates that a significant temperature elevation of the lens is required for cataract formation. If exposure is limited to 10 milliwatts per centimeter squared, the lens temperature is not elevated to levels at which cataracts occur.

In addition to thermal effects, nonionizing radiation is known to produce nonthermal effects. An association of a biological hazard with the nonthermal effects has not been demonstrated.

A peculiar effect experienced by some personnel is the sensation of sound when they are exposed to pulsed microwave fields. This occurs at levels below stated hazard limits and is not, by itself, considered dangerous.

Q38. Electromagnetic radiation is hazardous to personnel in what two ways?

Q39. What is the most useful and widespread technique to reduce rf burn hazards?

SUMMARY

Now that you have completed this chapter, a short review of what you have learned is in order. The following summary will refresh your memory of basic systems equipment, its principles, terms, and typical circuitry required for you to understand this concept.

A **RADIO SET CONTROL UNIT** is used to remotely control certain transmitter and receiver functions.

TRANSMITTER TRANSFER SWITCHBOARDS selectively transfer remote control station functions and signals to transmitters.

RECEIVER TRANSFER SWITCHBOARDS transfer receiver audio outputs to remote control station audio circuits.

A **TRANSMITTER** generates an rf carrier, modulates it with intelligence, amplifies it, and applies it to an antenna.

An **ANTENNA COUPLER** is a device used for impedance matching between an antenna and a transmitter or receiver.

A **RECEIVER** receives electromagnetic energy (rf) and may convert it to a visible or audible form.

MULTICOUPLERS patch several receivers or transmitters to one antenna. They also filter out harmonics and spurious responses, and provide impedance matching.

MARKING is when a circuit is closed and current flows in teletypewriter operation.

SPACING is when a circuit is open and no current flows in teletypewriter operation.

INTELLIGENCE is any signal that conveys information (voice, teletypewriter, facsimile).

A **START** unit is the first unit of a teletypewriter signal. It is always a space.

A **STOP** unit is the last unit of a teletypewriter signal. It is always a mark.

A **TRANSITION** is the time it takes to shift from a mark to a space condition or from a space to a mark condition.

A **CODE** in teletypewriter operation is a combination of mark and space conditions representing symbols, figures, or letters.

NONSYNCHRONOUS teletypewriter operation is when both transmitter and receiver do not operate continuously.

SYNCHRONOUS teletypewriter operation is when both transmitter and receiver operate continuously.

WORDS-PER-MINUTE is an approximate rate of speed. It means the number of five letter words with a space between them that can be transmitted or received in a one-minute period.

BAUD is a measurement of speed based on the number of code elements or units per second.

BITS-PER-SECOND is an acronym of the words binary digit. One bit is equal to one signal unit or element.

NEUTRAL teletypewriter operation is where current flow represents a mark and no flow represents a space.

POLAR teletypewriter operation is where current flow of one polarity represents a mark and current of the opposite polarity is a space.

RUNNING OPEN is the teletypewriter condition where the type hammer constantly strikes the type box but does not print or move across the page.

A **KEYER** is a device that changes dc pulses to mark and space modulation for teletypewriter transmissions.

A **CONVERTER** changes an audio signal back to dc pulses during teletypewriter reception.

AUDIO FREQUENCY TONE SHIFT systems use amplitude modulation to change dc mark and space impulses into audio impulses.

RADIO FREQUENCY CARRIER SHIFT systems use a keyer to shift a radio frequency signal above or below an assigned frequency. These shifts correspond to marks and spaces.

A **TELETYPEWRITER** is a machine that can transmit and or receive letters, numbers, or symbols. It may have a keyboard similar to a typewriter.

A **PERFORATOR** is a device that stores a teletypewriter message on a paper tape by punching Baudot coded messages into it.

TRANSMITTER DISTRIBUTOR is a device that reads Baudot code from paper tape and allows a message to be sent or a message to be printed on a page printer.

A **REPERFORATOR** stores an incoming tty signal on paper tape.

A **PAGE PRINTER** prints teletypewriter characters one at a time in a full-page format. This is usually a high-speed printer.

RED is the reference color of equipment that passes classified information. It normally refers to patch panels.

BLACK is the reference color of equipment that passes unclassified information. It normally refers to patch panels.

A **PATCH PANEL** is used to tie a receiver or transmitter to its associated equipment.

A **COMPARATOR** compares incoming signals and selects the strongest to be fed to a teletypewriter through a patch panel. This is used in diversity operation

A **LISSAJOUS PATTERN** is a combined, simultaneous display of the amplitude and phase relationships of two input signals on a crt.

A **TONE-TERMINAL** set converts tty dc pulses into audio tones for modulation of a transmitter in audio-frequency tone shift transmissions.

MULTIPLEXING is the process of transmitting a number of intelligence signals simultaneously over a single rf carrier.

TIME-DIVISION multiplexing is the process that periodically samples several intelligence signals. This can be a received signal or one to be transmitted.

FREQUENCY-DIVISION multiplexing transmits and receives the full 360 degrees of each sine wave.

FACSIMILE is the method for transmitting and receiving still images. These images can be maps, photographs, and handwritten or printed text.

SCANNING is the process of subdividing a picture in an orderly manner into segments. This is used in facsimile transmission.

FRAMING is the process of synchronizing a facsimile receiver to a transmitter. This allows proper picture reproduction.

TEMPEST is a term normally used to describe compromising emanations. These emanations are unintentionally radiated signals that could disclose classified information.

ELECTROMAGNETIC INTERFERENCE is a term used to describe the degradation of a receiver or system by externally produced rf energy.

ANSWERS TO QUESTIONS Q1. THROUGH Q39.

A1. To convert energy electrical/acoustic to acoustic/electrical and to key/unkey a transmitter. Also it mutes a receiver when transmitting.

A2. Transferring remote control functions and signals to transmitters.

A3. Transfers receiver audio outputs to remote control stations.

A4. 800 watts.

A5. Automatic, semiautomatic, and manual.

A6. It matches the impedance of an antenna to that of a transmission line at any desired frequency.

A7. To aid in heat transfer and prevent corona and arcing.

A8. Lsb, usb, isb, AM, cw, fsk.

A9. Digital.

A10. To connect an antenna/transmission line to a receiver/transmitter.

A11. Patching and filtering and permits the multiple use of receivers and/or transmitters on a single antenna.

A12. Space and mark.

A13. Intelligence (5), start (1), stop (1).

A14. Shift signals.

A15. Synchronous and nonsynchronous.

A16. A unit of modulation rate.

A17. Binary digit.

A18. Neutral and polar.

A19. Converts dc to corresponding mark and space modulation.

A20. Converts the audio signal to dc pulses.

A21. Uses AM to change dc to audio.

A22. A keyer provides rf excitation, which can be shifted above or below the assigned frequency.

A23. 60, 75, or 100 wpm.

A24. Page-size copy paper and perforated tape.

A25. It handles classified information.

A26. To code or decode messages.

A27. *The comparator compares the signal strengths from the receivers and the converter converts the frequency-shift rf signal into a tty set dc loop control signal.*

A28. *It converts dc to audio or vice versa.*

A29. *Time-division and frequency-division.*

A30. *It allows simultaneous transmission of multiple signals on a single transmission path.*

A31. *Facsimile.*

A32. *Compromising emanations.*

A33. *Reliability, security, and speed.*

A34. *Reliability.*

A35. *To ensure continuous, optimum performance of communications systems.*

A36. *Electromagnetic interference.*

A37. *Conduction and radiation.*

A38. *Rf burns and biological, thermal, and neurological effects.*

A39. *Proper bonding and grounding.*

CHAPTER 4

INTRODUCTION TO SATELLITE COMMUNICATIONS

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Describe the basic operation of the two types of satellites.
2. Describe the basic components of an operational satellite system.
3. Describe the function of earth terminal equipment.
4. Describe the basic signal flow of a typical shipboard receive-only system.
5. Describe the basic signal flow of a typical shipboard transceiver system.
6. Describe the advantages of satellite communications in terms of capacity, reliability, vulnerability, and flexibility.
7. Describe the limitations of satellites in terms of power, receiver sensitivity, and availability.

HISTORY OF SATELLITE COMMUNICATIONS

The first artificial satellite was placed in orbit by the Russians in 1957. That satellite, called *Sputnik*, signaled the beginning of an era.

The United States, who was behind the Russians, made an all-out effort to catch up, and launched *Score* in 1958. That was the first satellite with the primary purpose of communications.

The first regular satellite communications service was used by the Navy in 1960. The moon was used to bounce teletypewriter signals between Hawaii and Washington, D.C. During the early 1960s, the Navy used the moon as a medium for passing messages between ships at sea and shore stations. This method of communications proved reliable when other methods failed.

Military satellite communications technology was at a low level until 1965. At that time high quality voice transmissions were conducted between a satellite and two earth stations. That was the stepping stone to the Initial Defense Communications Satellite Program (IDCSP), which will be covered later in this chapter.

Experience with satellite communications has demonstrated that satellite systems can satisfy many military requirements. They are reliable, survivable, secure, and a cost effective method of telecommunications. You can easily see that satellites are the ideal, if not often the only, solution to problems of communicating with highly mobile forces. Satellites, if properly used, provide much needed options to large, fixed-ground installations.

For the past fifty years, the Navy has used high-frequency (hf) transmissions as the principal method of sending messages. In the 1970s, the hf spectrum was overcrowded and "free" frequencies were at a premium. Hf jamming and electronic countermeasures (ECM) techniques became highly sophisticated during that period. As a result the need for new and advanced long-range transmission methods became apparent.

Communications via satellite is a natural outgrowth of modern technology and of the continuing demand for greater capacity and higher quality in communications.

In the past, the various military branches have had the resources to support their communications needs. Predicted usage indicates that large-scale improvements will have to be made to satisfy future needs of the Department of Defense. These needs will require greater capacity for long-haul communications to previously inaccessible areas. Satellite communications has the most promise for satisfying these future requirements.

DEFENSE COMMUNICATIONS SATELLITE PROGRAM (DCSP)

The Defense Communications Satellite Program (DCSP) was initiated by the Secretary of Defense in 1962. Phase I of the program was given the title Initial Defense Communications Satellite Program (IDCSP). The first satellite launch occurred in June 1966 when seven experimental satellites were placed into orbit. The final launch of this program consisted of eight satellites and occurred in June 1968.

DEFENSE SATELLITE COMMUNICATIONS SYSTEM (DSCS) PHASE II

The Phase II Defense Satellite Communications System (DSCP Phase II) has changed from an all-analog communications system to an all-digital communications system. The performance capability provided by the Phase II DSCS is limited by equipment availability. Extensive digital traffic capability has become common. You can credit this to the availability of digital modems (modulator/demodulator) and broadband equipment. Overall performance of the Phase II DSCS is a great improvement over the capabilities provided by Phase I DSCS. The Phase II satellites provide a great increase in effective radiated power and rf bandwidths. You will find these satellite configurations use wide coverage and narrow beam antennas. They provide an extensive range of communications services and capabilities. (This will be further discussed later, in this chapter.)

FUNDAMENTAL SATELLITE COMMUNICATIONS SYSTEM

A satellite communications system uses satellites to relay radio transmissions between earth terminals. The two types of communications satellites you will study are ACTIVE and PASSIVE. A passive satellite only reflects received radio signals back to earth. An active satellite acts as a REPEATER; it amplifies signals received and then retransmits them back to earth. This increases signal strength at the receiving terminal to a higher level than would be available from a passive satellite.

A typical operational link involves an active satellite and two or more earth terminals. One station transmits to the satellite on a frequency called the UP-LINK frequency. The satellite then amplifies the signal, converts it to the DOWN-LINK frequency, and transmits it back to earth. The signal is next picked up by the receiving terminal. Figure 4-1 shows a satellite handling several combinations of links simultaneously.

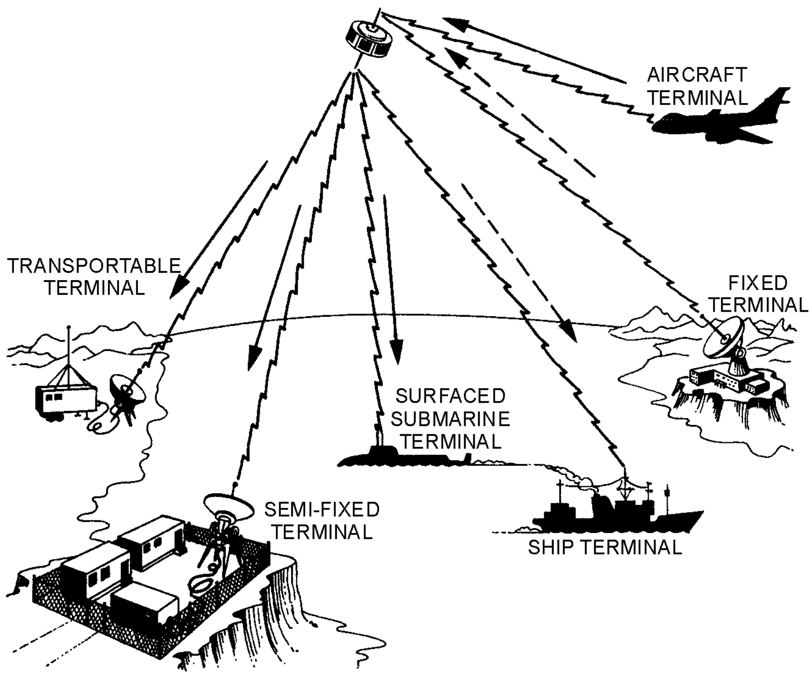


Figure 4-1.—Satellite communications system.

DESCRIPTION OF COMMUNICATIONS SATELLITE SYSTEM

The basic design of a satellite communications system depends to a great degree upon the characteristics of the orbit of the satellite. In general terms, an orbit is either elliptical or circular in shape. A special type of orbit is a SYNCHRONOUS ORBIT. In this type you will find the period (time required for one revolution) of the orbit the same as that of the earth. An orbit that is not synchronous is called ASYNCHRONOUS. A period of orbit that approaches that of the earth is called NEAR SYNCHRONOUS (subsynchronous). Orbits are discussed in more detail later in this chapter.

In addition to the fundamental components shown in figure 4-1, the design of the overall system determines the complexity of the various components and the manner in which the system operates. Current satellites are capable of handling many teletypewriter (tty) and voice circuits at the same time.

Orbit Descriptions

Orbits generally are described according to the physical shape of the orbit and the angle of inclination of the plane of the orbit. These terms aye discussed in the following paragraphs:

PHYSICAL SHAPE.—All satellites orbit the earth in elliptical orbits. (A circle is a special case of an ellipse.) The shape of the orbit is determined by the initial launch parameters and the later deployment techniques used.

PERIGEE and APOGEE are two, of the three parameters used to describe orbital data of a satellite. These are shown on figure 4-2. Perigee is the point in the orbit nearest to the center of the earth. Apogee is the point in the orbit the greatest distance from the center of the earth. Both distances are expressed in nautical miles.

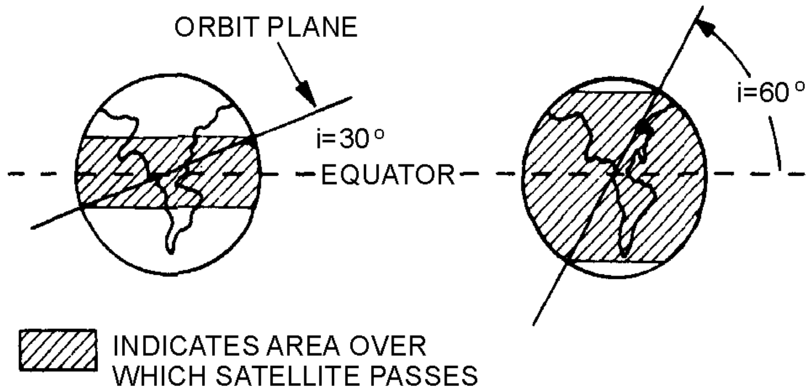


Figure 4-2.—Elliptical satellite orbit.

ANGLE OF INCLINATION.—The ANGLE OF INCLINATION (angle between the equatorial plane of the earth and the orbital plane of the satellite) is the third parameter used to describe the orbit data of a satellite. Figure 4-3 depicts the angle of inclination between the equatorial plane and the orbital plane. Most satellites orbit the earth in orbital planes that do not coincide with the equatorial plane of the earth. A satellite orbiting in any plane not identical with the equatorial plane is in an INCLINED ORBIT.

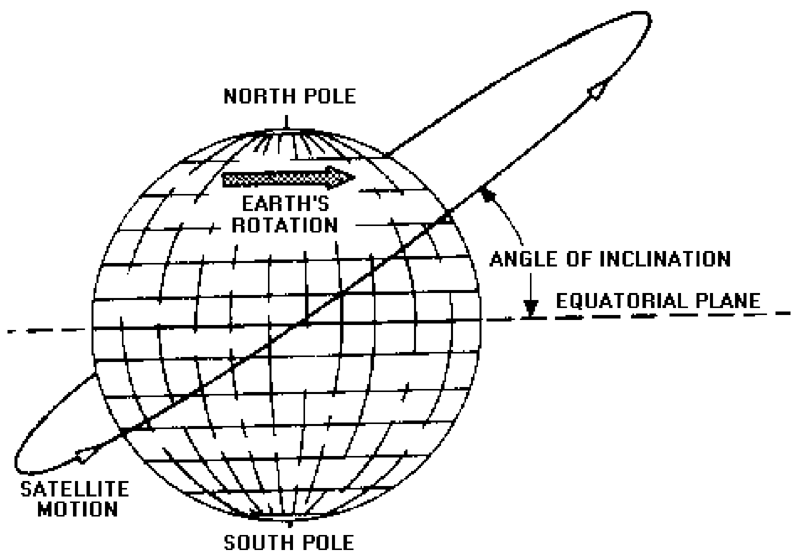


Figure 4-3.—Inclined satellite orbit.

The inclination of the orbit determines the area covered by the path of the satellite. As shown in figure 4-4, the greater the inclination, the greater the amount of surface area covered by the satellite.

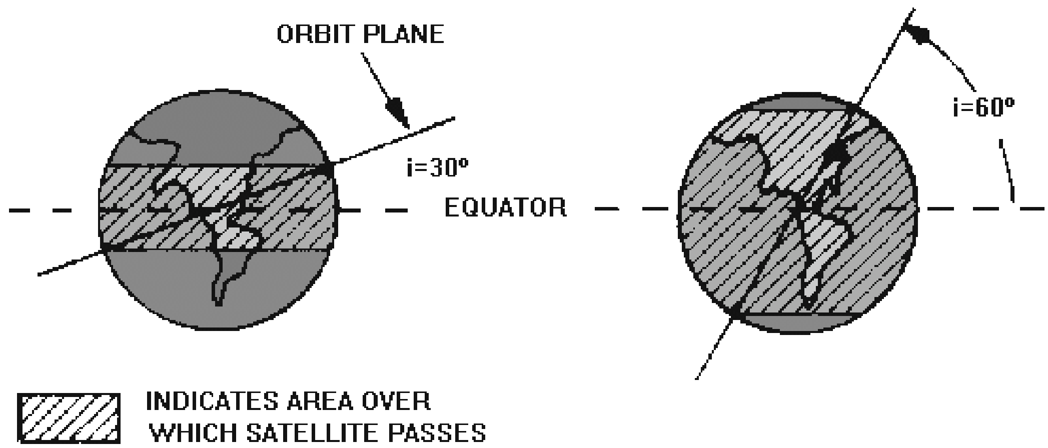


Figure 4-4.—Effect of orbit plane inclination on satellite coverage.

SPECIAL TYPES OF INCLINED ORBITS.—A satellite orbiting in a plane that coincides with the equatorial plane of the earth is in an EQUATORIAL ORBIT. A satellite orbiting in an inclined orbit with an angle of inclination of 90 degrees or near 90 degrees is in a POLAR ORBIT.

SPECIAL TYPES OF CIRCULAR ORBITS.—We stated previously that a circular orbit is a special type of elliptical orbit. You should realize a circular orbit is one in which the major and minor axis distances are equal or approximately equal. Mean height above earth, instead of perigee and apogee, is used in describing a circular orbit. While we are discussing circular orbits, you should look at some of the terms mentioned earlier in this chapter. A satellite in a circular orbit at a height of approximately 19,300 nautical miles above the earth is in a synchronous orbit. At this altitude the period of rotation of the satellite is 24 hours, the same as the rotation period of the earth. In other words, the orbit of the satellite is in sync with the rotational motion of the earth. Although inclined and polar synchronous orbits are possible, the term synchronous usually refers to a synchronous equatorial orbit. In this type of orbit, satellites appear to hover motionlessly in the sky. Figure 4-5 shows how one of these satellites can provide coverage to almost half the surface of the earth.

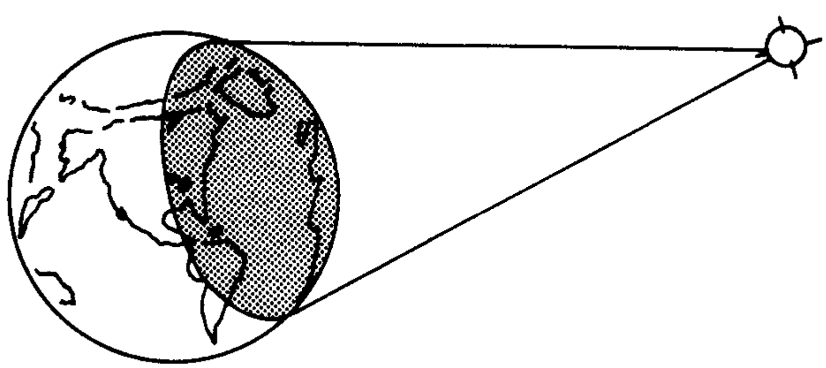


Figure 4-5.—Illumination from a synchronous satellite.

Three of these satellites can provide coverage over most of the earth (except for the extreme north and south polar regions). A polar projection of the global coverage of a three-satellite system is shown in figure 4-6.

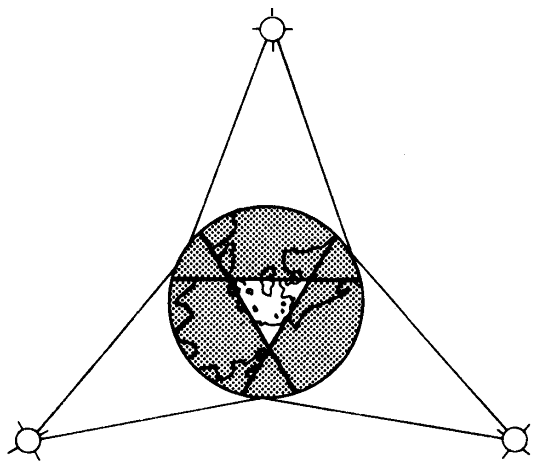


Figure 4-6.—Worldwide synchronous satellite system viewed from above the North Pole.

A satellite in a circular orbit at other than 19,300 nautical miles above the earth is in a near-synchronous orbit. If the orbit is lower than 19,300 nautical miles, the period of orbit of the satellite is less than the period of orbit of the earth. The satellite then appears to be moving slowly around the earth from west to east. (This type of orbit is also called subsynchronous.) If the orbit is higher than 19,300 nautical miles, the period of orbit of the satellite is greater than the period of orbit of the earth. The satellite then appears to be moving slowly around the earth from east to west. Although inclined and polar near-synchronous orbits are possible, near synchronous implies an equatorial orbit.

A satellite in a circular orbit from approximately 2,000 miles to 12,000 miles above the earth is considered to be in a MEDIUM ALTITUDE ORBIT. The period of a medium altitude satellite is considerably less than that of the earth. When you look at this altitude satellite, it appears to move rather quickly across the sky from west to east.

Q1. What are the two types of communications satellites?

Q2. A typical satellite communications operational link consists of a satellite and what two other components?

Q3. A satellite in a synchronous orbit can cover how much of the surface of the earth?

Q4. What areas of the earth are not normally covered by satellites?

SATELLITE CHARACTERISTICS

Early communications satellites were limited in size to the diameter of the final stage of the rocket that was used for launching. Weight was determined by the thrust of the rocket motors and the maximum weight the rocket could lift into orbit.

As early as June 1960, two satellites were successfully placed in orbit by the same launch vehicle. With the development of multilaunch capability, added flexibility became available. We then had choices as to the size, weight, and number of satellites to be included in each launch.

Using our multilaunch capabilities, the Defense Satellite Communications System (DSCS) has placed larger and heavier satellites in synchronous equatorial orbits. Figure 4-7 is a drawing of a DSCS satellite. It shows each pair of transmit and receive dish antennas. As you can see, a large area of the earth can be covered using only one satellite.

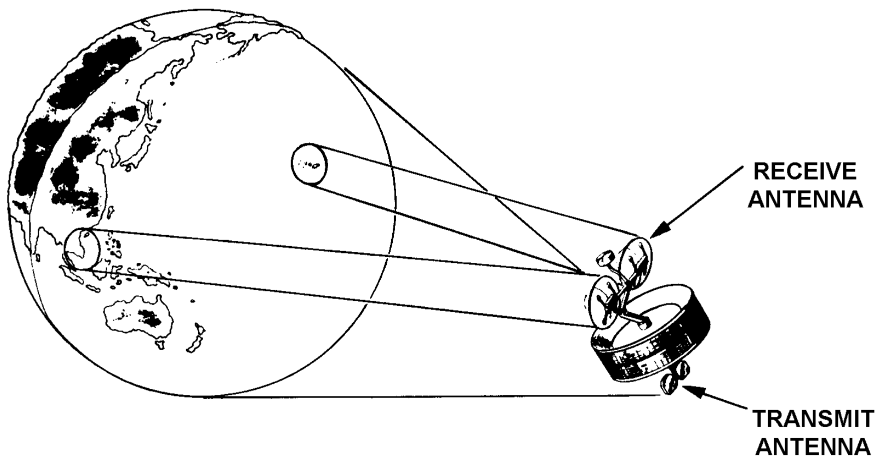


Figure 4-7.—DSCS satellite.

Satellite Power Sources

Early communications satellites were severely limited by the lack of suitable power sources. This severely limited the output power of the satellite transmitter. The only source of power available within early weight restrictions was a very inefficient panel of solar cells without battery backup. A major disadvantage of this type of power source is that the satellite has no power when it is in ECLIPSE (not in view of the sun). For continuous communications, this outage is unacceptable.

A combination of solar cells and storage batteries is a better prime power source. This is a practical choice, even though the result is far from an ideal power source. About ten percent of the energy of the sunlight that strikes the solar cells is converted to electrical power. This low rate is sometimes decreased even further. You find this when the solar cells are bombarded by high-energy particles that are sometimes found in space.

Early satellites had over 8,500 solar cells mounted on the surface of the satellite, which supplied about 42 watts of power. No battery backup was provided in these satellites.

Newer communications satellites have about 32,000 solar cells mounted on the surface of the satellite, and they supply about 520 watts. A nickel cadmium battery is used for backup power during eclipses.

Nuclear power sources have been used in space for special purposes, but their use stops there. Technology has not progressed sufficiently for nuclear power sources to be used as a power source.

Satellite Orientation

Satellite orientation in space is important for continuous solar cell and antenna orientation. Since the primary source of power in most satellites is from solar cells, a maximum number of the solar cells must be exposed to the sun at all times. The satellite antenna must also be pointed at the appropriate earth terminals. Our communications satellites use what is termed spin stabilization to meet these important requirements.

Spin stabilization operates on the principle that direction of the spin axis of a rotating body tends to remain fixed in space. An example of spin stabilization is the effect of the rotation of the earth in keeping its axis fixed in space. A satellite that has a spin axis parallel to the axis of the earth will maintain this position since both axes are fixed in space. Figure 4-8 illustrates the use of this principle. It depicts an equatorial orbit satellite used to keep a doughnut-shaped antenna pattern pointing toward the earth.

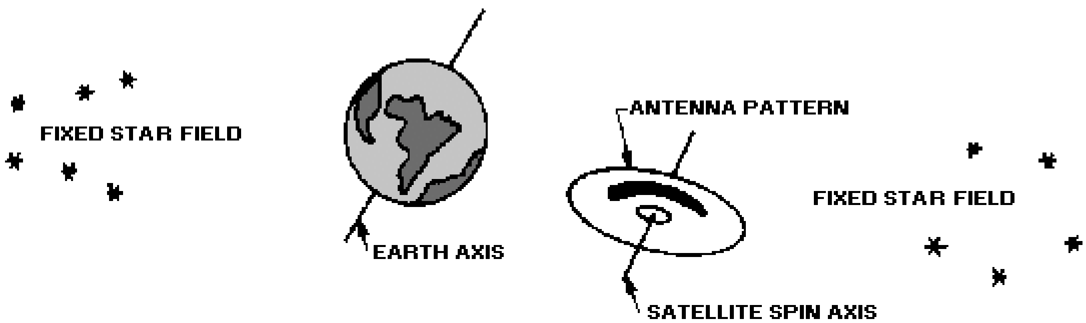


Figure 4-8.—Spin-stabilized satellite antenna pattern.

Once the system is in motion, spin stabilization requires virtually no additional energy. A spin-stabilized satellite is usually constructed like a flywheel. Its heavier equipment is mounted in the same plane and as close to the outside surface as possible.

After reaching its orbit, the radial jets are pulsed to start the satellite spinning. The satellite spin axis is orientated to the axis of the earth by means of small axial jets. Velocity jets are used to place the satellite in orbit position and provide velocity correction. Figure 4-9 is an example of spin stabilization.

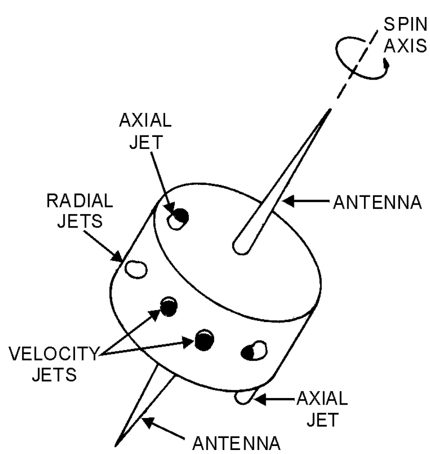


Figure 4-9.—Spin-stabilized satellite controls.

Solar cells are installed around the outside surface of a spin-stabilized satellite. This gives you a large number of solar cells exposed to the sun at all times (except when the satellite is in eclipse). The use of omnidirectional antennas causes a small part of the total radiated energy to be directed toward the earth at all times.

Ominidirectional antennas radiate only a small amount of energy toward the earth. Many techniques have been tried to achieve an earth-oriented antenna system. One system developed uses spin stabilization for orientation of the satellite. It uses a stationary inner platform for mounting remote controlled antennas. The satellite is constructed in two parts with both parts having radial jets. The inner platform contains the communications antennas and the communications package. After the satellite is stabilized in space, inner radial jets spin the inner platform. The inner platform is stationary with respect to earth and is oriented to such a position that the communications antennas point continuously toward the earth. This arrangement allows the use of high-gain directional antennas that concentrate the majority of the radiated energy in the direction of the earth.

The latest versions of communications satellites use a stationary platform with four high-gain antennas. Two steerable narrow beam antennas are used for communications between and within regions of high traffic density. Two horn antennas provide for earth communications between facilities outside the narrow beam coverage. The antenna arrangement for these types of communications satellites is shown in figure 4-7.

Q5. What was the major operational limitation of early communications satellites?

Q6. Satellite orientation in space is important for what two reasons?

EARTH TERMINAL CHARACTERISTICS

Communications satellite earth terminals are usually located in areas remote from the actual users of these communications. This is necessary to minimize rf interference to the satellite.

Locating the terminals in these remote locations requires interconnecting communications links. Links permit communications flow to and from the users of the satellite systems. Interconnect links are usually connected via telephone cables or microwave radio links with normal terminal equipment.

Earth satellite communications terminals generally have a single, large antenna; a highly sensitive receiver; a powerful transmitter; multiplex equipment; modulating-demodulating equipment; and telemetry equipment. Each of these elements will be discussed later in this chapter.

Antennas

Earth terminal antennas are highly directional, high-gain antennas capable of transmitting and receiving signals simultaneously. Generally, large, high-gain, parabolic antennas are used.

Generally speaking, three sizes of parabolic-type antennas are currently in use at earth terminal sites. These are a parabolic antenna sixty feet in diameter, a parabolic antenna forty feet in diameter, and a cluster of four parabolic antennas, each ten feet in diameter. These four in combination are equal to a parabolic antenna eighteen feet in diameter. They are shown in figures 4-10, 4-11, and 4-12, respectively.

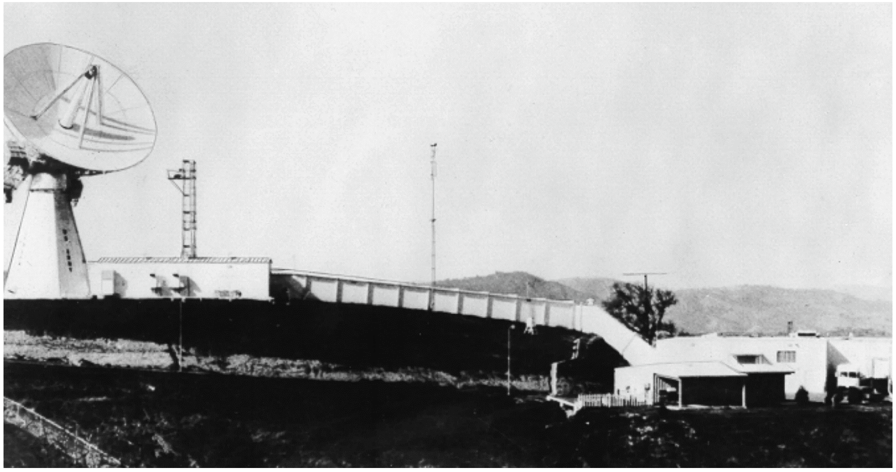


Figure 4-10.—Typical satellite earth terminal with sixty-foot antenna.

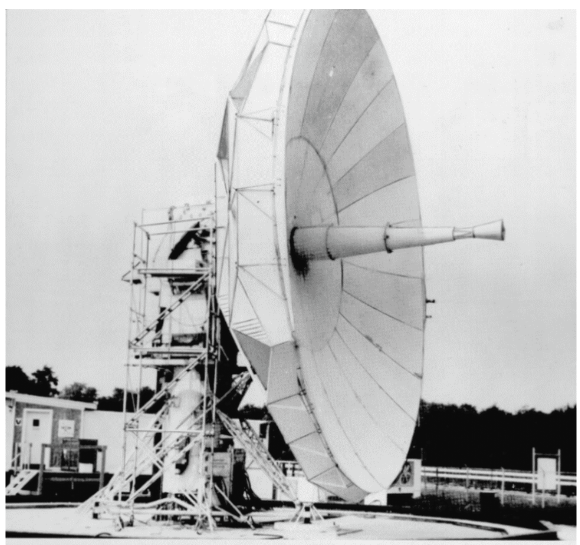


Figure 4-11.—Forty-foot antenna and pedestal.

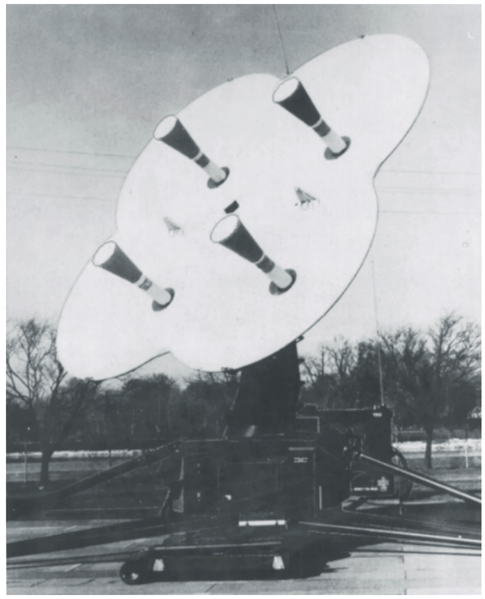


Figure 4-12.—Parabolic antenna cluster.

Receivers

All satellite communications earth terminals are equipped with specially designed, highly sensitive receivers. These receivers are designed to overcome down-link power losses and to permit extraction of the desired communications information from the weak received signal. The terminals currently in use have specially designed preamplifiers mounted directly behind the antennas.

Transmitters

All earth terminal transmitters generate high-power signals for transmission to the communications satellites. High-powered transmitters and highly directional, high-gain antennas are combined in this configuration. This is necessary to overcome up-link limitations and to ensure that the signals received by the satellite are strong enough to be detected by the satellite. Each transmitter has an exciter/modulator and a power amplifier. The modulator accepts the input signal from the terminal equipment and modulates an IF carrier. The exciter translates the IF signal to the up-link frequency and amplifies it to the level required by the power amplifier.

Transmitters used in earth terminals have output power capabilities that vary from 10 watts to 20 kilowatts, depending on the type used and the operational requirements.

Telemetry Equipment

Telemetry equipment is included in all communications satellite systems. This permits monitoring of the operating conditions within the satellite. Telemetry can be used also for remote control of satellite operations, such as energizing axial jets for changing the spin axis of the satellite.

Q7. What type of antennas are generally used at earth terminals?

Q8. Why do earth terminals require highly sensitive receivers?

Q9. What is the range of earth terminal transmitter output power?

SHIPBOARD RECEIVE-ONLY EQUIPMENT SYSTEMS

The purpose of a shipboard receive-only system is to receive fleet multichannel teletypewriter broadcasts, which, as you recall from chapter 1, require no receipt. These are transmitted from a ground station and relayed to naval vessels by satellite.

Figure 4-13 is a typical shipboard receive-only system. In this system the transmitted carrier may be frequency modulated (fm) or phase-shift-key (psk) modulated for tty operation. The receiving antennas for this system are positioned about the ship. They are arranged in a manner (normally one in each quadrant of the ship) that at no time allows the line-of-sight to be blocked between the relay satellite and one or more of the antennas. Incoming signals pass from the antennas to an amplifier-converter. Each amplifier-converter routes an IF signal on one line of a twin axial cable that connects it to the combiner-demodulator. An operating power and local-oscillator signal are coupled from the combiner-demodulator to each amplifier-converter on the other line of the cable used for the IF signal. Because of signal path variations, shading, and reflections, the incoming signals are subject to random phase and amplitude variations. The combiner operation performed within the combiner-demodulator removes the phase variations from each input signal. It then measures the amplitudes of the signals for optimum combining and sums the signals. After being combined, the signal is demodulated and coupled from a receiver transfer switchboard to a telegraph demultiplex terminal.

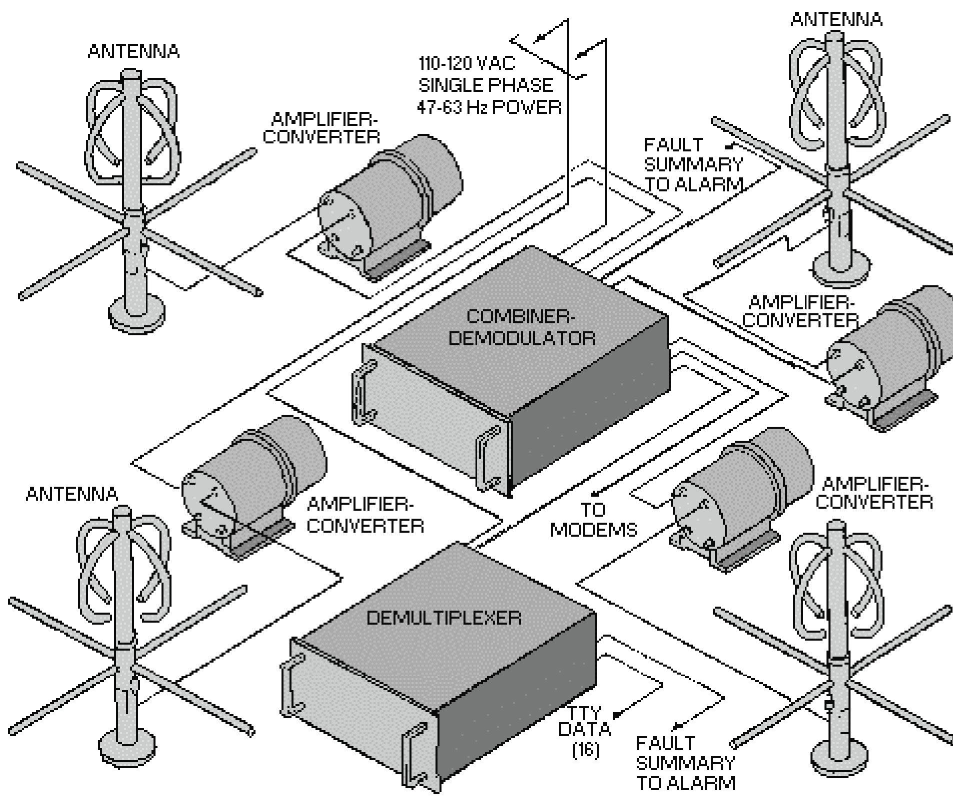


Figure 4-13.—Typical shipboard receive only system.

Q10. What is the function of shipboard receive-only equipment?

Q11. What types of modulation are shipboard receive-only equipment designed to receive?

SATELLITE ACQUISITION AND TRACKING

An essential operation in communicating by satellite is the acquisition (locating) of the satellite by the earth terminal antenna and the subsequent tracking of the satellite. Initial acquisition depends upon an exact knowledge of the position of the satellite. In combination with the geographic location of the earth terminal, knowing the position of the satellite enables you to compute accurate antenna pointing information. The degree of difficulty in locating and tracking a satellite is determined largely by what type orbit the satellite is in.

The locating and tracking of a synchronous satellite is relatively simple. This is because the satellite appears to be stationary. Locating a near-synchronous satellite is also relatively simple because of the slow relative motion of the satellite However, the movement of a near-synchronous satellite is enough that accurate tracking is required to keep the narrow beam antenna pointed toward the satellite. Satellites in medium altitude circular orbits or in elliptical orbits are more difficult to acquire and to track because of the rapid changes in position.

Orbital Prediction

To acquire and track a satellite in space, the earth terminal antennas must be provided with very accurate pointing information. Antenna pointing information is based upon the orbital prediction of the satellite. This information is derived from an EPHEMERIS table. This table provides the coordinates of a satellite or a celestial body at specific times during a given period. After you know the ephemeris data of a satellite, you can predict for any given location the apparent track of the satellite as viewed from that location.

The constants defining an orbit are initially obtained by the process of tracking. At the time of launch, the rocket is tracked by radar from lift-off to orbit and then until it passes out of sight. Tracking data obtained in this way is sufficient for making rough predictions of the orbit. These predictions are made rapidly with a computer and sent to tracking stations all over the world. These other tracking stations watch for the satellite during its first trip and record additional data. During the first week of orbiting, tracking stations all around the world are obtaining progressively more accurate data concerning the Satellite. This data is put into a computer where corrections of earlier estimates of the orbit are made.

Once the initial predictions are complete and the satellite link becomes operational, very little change in these calculations is made. The orbits of a satellite will change slightly over a period of time; however, these changes are so gradual that predictions will be accurate enough to be used for weeks or even months without further corrections. When the orbits are known precisely, an ephemeris can be calculated for each satellite of the system.

Antenna Pointing

Antenna pointing instructions for each satellite must be computed separately for each ground station location. A satellite that bears due south of station A at an elevation of 25 degrees may simultaneously bear due southeast of station B at an elevation of 30 degrees. Antenna pointing instructions are determined by taking into consideration the orbital prediction and the latitude and longitude of each ground station.

To establish radio contact with a satellite, the ground station needs to know the bearing and elevation of a satellite. This allows the antenna to be properly pointed.

Acquisition

The acquisition of satellite signals by a ground station equipped with large antennas and operated at microwave frequencies places severe requirements on the system. Several factors must be considered. These factors are discussed below:

SPATIAL-TIME FACTOR.—Very accurate antenna pointing information is available to earth terminals from the satellite control facility located in Sunnyvale, California. Because of equipment limitations, a small search about the predicted location of the satellite must often be conducted to make initial contact. Either a manual or automatic scan is made around a small area close to the point where the satellite appearance is predicted.

FREQUENCY CONTROL.—The frequency of a radio signal received from a satellite is not generally the exact assigned down-link frequency. This variation depends upon the type of orbit of the satellite. The greatest frequency variations in signals from satellites occur in medium altitude circular or elliptical orbits. The smallest frequency variations occur in signals from satellites in near-synchronous or synchronous orbits.

Tracking

When a particular satellite has been acquired, the earth terminal antenna will track that satellite for as long as it is used as a communications relay. Several methods of tracking are in actual use; however, we will explain PROGRAMMED TRACKING and AUTOMATIC TRACKING.

PROGRAMMED TRACKING.—In programmed tracking the known orbital parameters of the satellite are fed into computation equipment to generate antenna pointing angles. The antenna pointing angles are fed as commands to the antenna positioning servomechanisms. (You may want to review servos in NEETS, Module 15, *Principles of Synchros, Servos, and Gyros*.) These point the antenna in the required direction. The amount of data and computations involved in using programmed tracking is extensive. These are a result of the antenna mount flexing and atmospheric and ionospheric bending of radio waves. Because of these uncertainties, programmed tracking is not used extensively.

AUTOMATIC TRACKING.—In automatic tracking, the equipment generates antenna pointing information by comparing the direction of the antenna axis with the direction from which an actual satellite signal is received. Automatic tracking systems track the apparent position of a satellite. The direction of arrival of the radio signal and the real position of the satellite is not required. The automatic tracking system uses a servomechanism to move the antenna. Once the satellite has been located, the servomechanism generates its own pointing data. This eliminates the requirement for continuous data input and computation.

SATELLITE OUTAGE TIME.—The satellite outage time specifications allow for stewing (moving) the earth terminal antennas, acquiring the satellite signal, and checking for circuit continuity at HAND OVER. (Hand over is the period of time for one earth terminal to yield control to another as a satellite moves out of its area of coverage.) This hand over period represents an outage time. If the control terminal is unable to hand over to another terminal within a specified time, other arrangements are made. For example, control may be retained or transferred to another terminal within the coverage area. There are several reasons why a terminal may be unable to assume control on time; these reasons may combine to increase the outage time. The difference of drift velocities of the satellites leads to bunching within a coverage area. This causes gaps in coverage and increases outage times. When two or more satellites simultaneously occupy the same space of the terminal antennas, they will interfere with each other. This prevents reliable communications. Other factors leading to increased outage times are SATELLITE-SUN CONJUNCTION (increased noise while the satellite passes near the sun), SATELLITE ECLIPSE (absence of power from solar cells), and satellite failures. The distribution of outage times is a complicated function of time and earth-station locations. With careful coverage coordination, maximum communications effectiveness is obtained.

Q12. Why is satellite acquisition and tracking important?

ROLE OF SATELLITE COMMUNICATIONS

In the context of a worldwide military communications network, satellite communications systems are very important. Satellite communications links add capacity to existing communications capabilities and provide additional alternate routings for communications traffic. Satellite links, as one of several kinds of long-distance links, interconnect switching centers located strategically around the world. They are part of the defense communication systems (DCS) network. One important aspect of the satellite communications network is that it continues in operation under conditions that sometimes render other methods of communications inoperable. Because of this, satellites make a significant contribution to improved reliability of Navy communications.

ADVANTAGES OF SATELLITE COMMUNICATIONS

Satellite communications have unique advantages over conventional long distance transmissions. Satellite links are unaffected by the propagation variations that interfere with hf radio. They are also free from the high attenuation of wire or cable facilities and are capable of spanning long distances. The numerous repeater stations required for line-of-sight or troposcatter links are no longer needed. They furnish the reliability and flexibility of service that is needed to support a military operation.

Capacity

The present military communications satellite system is capable of communications between backpack, airborne, and shipboard terminals. The system is capable of handling thousands of communications channels.

Reliability

Communications satellite frequencies are not dependent upon reflection or refraction and are affected only slightly by atmospheric phenomena. The reliability of satellite communications systems is limited only by the equipment reliability and the skill of operating and maintenance personnel.

Vulnerability

Destruction of an orbiting vehicle by an enemy is possible. However, destruction of a single communications satellite would be quite difficult and expensive. The cost would be excessive compared to the tactical advantage gained. It would be particularly difficult to destroy an entire multiple-satellite system such as the twenty-six random-orbit satellite system currently in use. The earth terminals offer a more attractive target for physical destruction. These can be protected by the same measures that are taken to protect other vital installations.

A high degree of freedom from jamming damage is provided by the highly directional antennas at the earth terminals. The wide bandwidth system that can accommodate sophisticated anti-jam modulation techniques also lessens vulnerability.

Flexibility

Most operational military satellite earth terminals are housed in transportable vans. These can be loaded into cargo planes and flown to remote areas. With trained crews these terminals can be put into operation in a matter of hours. Worldwide communications can be established quickly to remote areas nearly anywhere in the free world.

SATELLITE LIMITATIONS

Limitations of a satellite communications system are determined by the technical characteristics of the satellite and its orbital parameters. Active communications satellite systems are limited by two things. Satellite transmitter power on the down links and receiver sensitivity on the up links. Some early communications satellites have been limited by low-gain antennas.

Power

The amount of power available in an active satellite is limited by the weight restrictions imposed on the satellite. Early communications satellites were limited to a few hundred pounds because of launch-vehicle payload restraints. The only feasible power source is the inefficient solar cell. (Total power generation in the earlier satellites was less than 50 watts.) As you can see, the rf power output is severely limited; therefore, a relatively weak signal is transmitted by the satellite on the down link. The weak transmitted signal is often reduced by propagation losses. This results in a very weak signal being available at the earth terminals. The level of signals received from a satellite is comparable to the combination of external atmospheric noise and internal noise of standard receivers. Special techniques must be used to extract the desired information from the received signal. Large, high-gain antennas and special types of preamplifiers solve this problem but add complexity and size to the earth terminal. (The smallest terminal in the defense communication systems network has effectively an 18-foot antenna and weighs 19,500 pounds.) Development of more efficient power sources and relaxation of weight restrictions have permitted improved satellite performance and increased capacity.

Receiver Sensitivity

Powerful transmitters with highly directional antennas are used at earth stations. Even with these large transmitters, a lot of signal loss occurs at the satellite. The satellite antenna receives only a small amount of the transmitted signal power. A relatively weak signal is received at the satellite receiver. This presents little problem as the strength of the signal received on the up link is not as critical as that received on the down link. The down-link signal is critical because the signal transmitted from the satellite is very low in power. Development of high-gain antennas and highly sensitive receivers have helped to solve the down-link problem.

Availability

The availability of a satellite to act as a relay station between two earth terminals depends on the locations of the earth terminals and the orbit of the satellite. All satellites, except those in a synchronous orbit, will be in view of any given pair of earth stations only part of the time. The length of time that a nonsynchronous satellite in a circular orbit will be in the ZONE OF MUTUAL VISIBILITY (the satellite can be seen from both terminals) depends upon the height at which the satellite is circling. Elliptical orbits cause the satellite zone of mutual visibility between any two earth terminals to vary from orbit to orbit. These times of mutual visibility are predictable. Figure 4-14 illustrates the zone of mutual visibility.

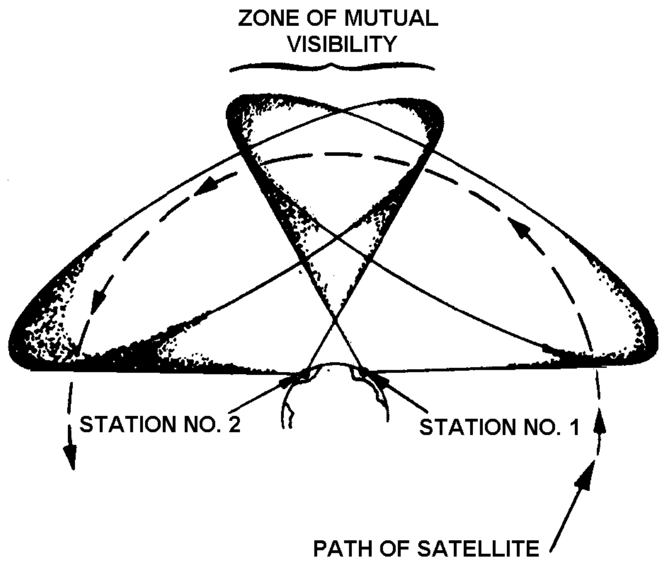


Figure 4-14.—Zone of mutual visibility.

Q13. What are the two limitations to an active satellite communications system?

FUTURE SATELLITE COMMUNICATIONS

Satellite communications are becoming well established in the Navy. In October 1983 the Department of the Navy established the Naval Space Command, which assumed operational responsibility for Navy space systems plus coordination responsibility with other operational activities. Most ships have satellite communications capability. New systems have been installed on ships and are fully compatible with other electronic systems and equipment. Communications via satellite has increased existing Navy communications capabilities for the command and control of naval forces. Satellite communications has not replaced all existing means of radio communications. However, it is a major step in modernizing Navy communications. It has relieved the Navy of its total dependence on hf radio transmission and reduced the need for many hf ground stations overseas.

A recent step in the advancement of satellite communications was the start of the DSCS Phase III. The first Phase III satellite was launched into orbit by the space shuttle in the summer of 1984. Seven satellites will be placed in space during this phase. Figure 4-15 shows a Phase III satellite being tested in a simulated space environment, Figure 4-16 shows the Phase III satellite as it appears in space. Phase III will develop the use of 40-watt, solid-state amplifiers to replace the currently used traveling-wave tube (twt). It will also be used to develop new type filters. These filters will provide increased channel bandwidth, which will provide additional communications capacity.

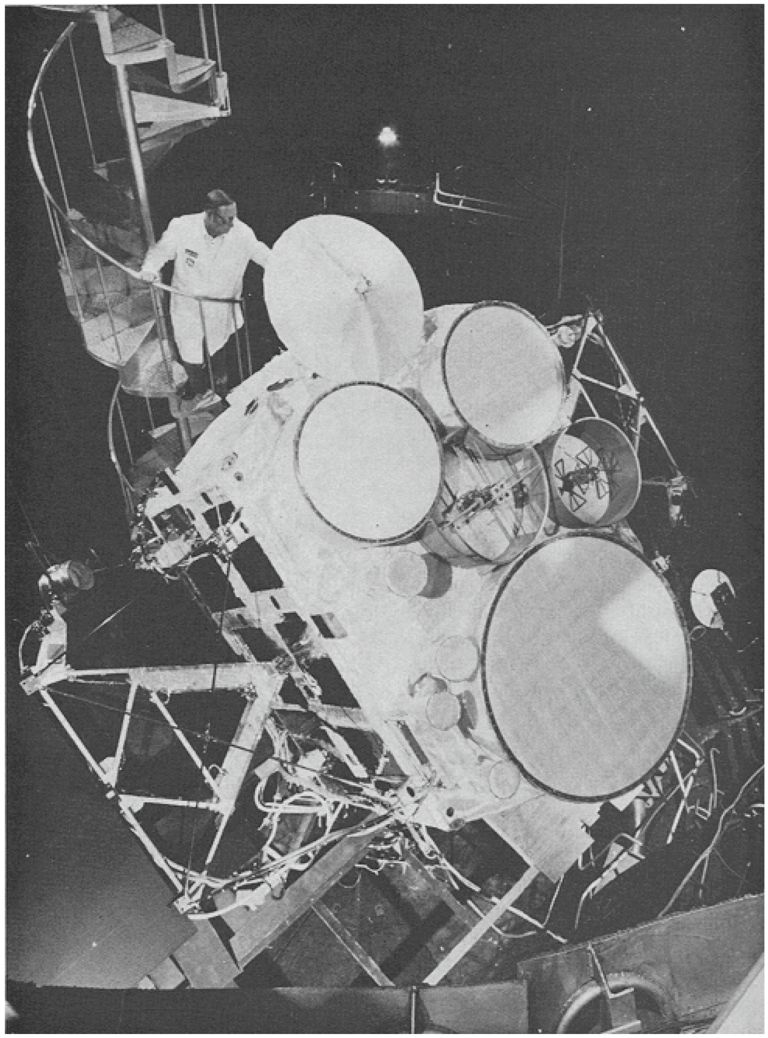


Figure 4-15.—DSCS Phase III satellite being tested.

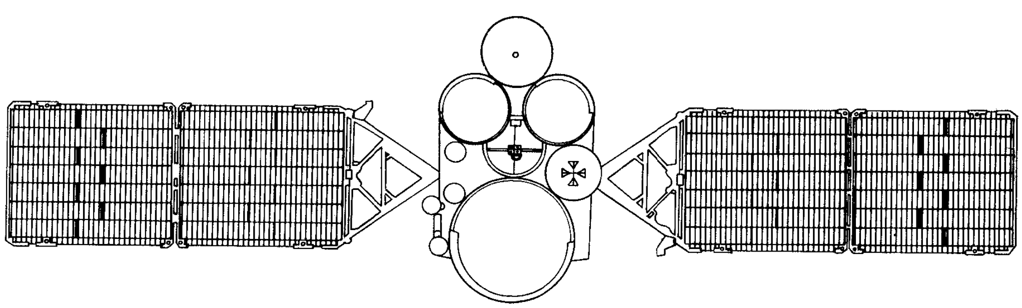


Figure 4-16.—DSCS Phase III satellite as it appears in space.

The survivability of reliable communications for the command and control of our strategic nuclear forces is paramount. Space systems perform many missions more effectively than earthbound systems.

Spaceborne communications increase the effectiveness of military operations. The Department of Defense is engaged in the development of new communications techniques and systems, including some that are space based. As the use of space continues its march forward, vital new opportunities for national defense will continue to materialize. This will improve the survivability of our strategic communications against nuclear and electronic attack.

More information on satellite communications can be found in Navy publication NTP 2, *Navy Satellite Operations*. This publication was written to concisely explain the role of the Navy in the Defense Communications Satellite Program. It also issues procedures for effective, coordinated use of available satellite resources.

SUMMARY

Now that you have completed this chapter, a short review of what you have learned will be helpful. The following review will refresh your memory of satellite communications, equipment, and theory.

A **PASSIVE SATELLITE** is one that reflects radio signals back to earth.

An **ACTIVE SATELLITE** is one that amplifies the received signal and retransmits it back to earth.

REPEATER is another name for an active satellite.

The **UP LINK** is the frequency used to transmit a signal from earth to a satellite.

The **DOWN LINK** is the frequency used to transmit an amplified signal from the satellite back to earth.

A **SYNCHRONOUS ORBIT** is one in which the satellite moves or rotates at the same speed as the earth.

An **ASYNCHRONOUS ORBIT** is one where the satellite does not rotate or move at the same speed as the earth.

A **NEAR SYNCHRONOUS ORBIT** is one in which the satellite rotates close to but not exactly at the same speed as the earth.

PERIGEE is the point in the orbit of a satellite closest to the earth.

APOGEE is the point in the orbit of a satellite the greatest distance from the earth.

The **ANGLE OF INCLINATION** is the angular difference between the equatorial plane of the earth and the plane of orbit of the satellite.

INCLINED ORBITS are orbits where there is some amount of inclination. These include equatorial and polar orbits.

An **EQUATORIAL ORBIT** is an orbit that occurs when the plane of a satellite coincides with the plane of the earth at the equator.

A **POLAR ORBIT** is an orbit that has an angle of inclination of or near 90 degrees.

A **MEDIUM ALTITUDE ORBIT** is an orbit from 2,000 to 12,000 miles above the earth. The rotation rate of the earth and satellite are quite different, and the satellite moves quickly across the sky.

An **ECLIPSE** is when the satellite is not in view or in direct line of sight with the sun. This happens when the earth is between them.

An **EPHEMERIS** is a table showing the precalculated position of a satellite at any given time.

PROGRAMMED TRACKING uses known satellite orbital parameters to generate antenna pointing angles.

AUTOMATIC TRACKING is done by the equipment comparing the direction of the antenna axis and the direction of the received signal.

HAND OVER is the period of time for one earth terminal to yield control to another as a satellite moves out of its area of coverage.

SATELLITE-SUN CONJUNCTION is when the satellite and sun are close together and the noise from the sun prevents or hampers communications.

A **SATELLITE ECLIPSE** is an eclipse where the rays of the sun don't reach the satellite. This prevents recharging of the solar cells of the satellite and decreases the power to the transmitter.

The **ZONE OF MUTUAL VISIBILITY** is where the satellite can be seen by both the up- and down-link earth terminals.

ANSWERS TO QUESTIONS Q1. THROUGH Q13.

A1. Passive and active.

A2. Earth terminals.

A3. Approximately one-half.

A4. The extreme polar regions.

A5. The lack of suitable power sources.

A6. To allow maximum solar cell exposure to the sun and satellite antenna exposure to earth terminals.

A7. Large, high-gain parabolic antennas.

A8. To overcome satellite transmitter low power and permit extraction of the desired information from the received signal.

A9. Up to 20 kilowatts.

A10. To receive fleet multichannel tty broadcasts.

A11. Fm or psk.

A12. To ensure earth terminal antennas are always pointed towards the satellite.

A13. Satellite down-link transmitter power and up-link receiver sensitivity.

CHAPTER 5

INTRODUCTION TO MISCELLANEOUS COMMUNICATIONS SYSTEMS AND EQUIPMENT

LEARNING OBJECTIVES

Upon completion of this chapter you will be able to:

1. Describe the basic operation of communications systems that operate at medium frequencies and below.
2. Describe the basic microwave line-of-sight communications system.
3. Describe the basic tropospheric scatter communications system.
4. Describe the objective/purpose of the naval tactical data system (NTDS).
5. Describe the naval tactical data system (NTDS) data transmission subsystems in terms of links.
6. Explain the various applications of portable communications equipment.
7. Define the term laser.
8. Describe the basic theory of operation of lasers
9. Describe the possible applications of lasers in communications.

INTRODUCTION

In the previous four chapters we've looked at communications equipment and systems that were used in several frequency ranges. Some have had many applications. In this chapter you will look at systems used in some portions of the rf spectrum that have not been covered in detail. We will also discuss the naval tactical data system (NTDS), which operates in the high-frequency and ultrahigh-frequency regions. Various portable communications equipments used in the military and an introduction to the laser and its uses in communications are included. Some of the applications presented are fairly new to the military community.

SYSTEMS

As discussed in chapter 1, the frequency range from elf to shf is from below 300 hertz up to 30 gigahertz. The first area we will cover is the lower frequency bands (medium frequency [mf] and below). You will then get a look at the microwave region and the high-frequency and ultrahigh-frequency range as it pertains to the naval tactical data system (NTDS).

MEDIUM FREQUENCY AND BELOW

Most of the receivers and transmitters that you will see used in the mf portions of the rf spectrum and below are very similar in design. In chapter 1 we discussed the operational uses of the equipment; now let's look at the equipment itself.

Equipment items covered in this and other chapters are meant to be merely representative of equipment that may be encountered in naval communications. No attempt will be made to include all of the possible equipment or equipment configurations.

Transmit Equipment

You should realize the transmitters used in bands of medium frequency and below are similar to those you studied in chapter 2. In other words, a transmitter used in one frequency range is basically the same as one used in another range. However, there are some differences. Two of the differences are component size and the use of a technique called DOUBLING UP.

The components used in bands of medium frequency and below are much larger physically than the ones previously discussed. This is because of the higher operating voltage and current levels required to produce the very high-powered rf outputs needed for the uses covered in chapter 1. A given resistor used in an hf application may be rated at 1/2 watt, whereas the same resistor used in a lower frequency application would probably be rated in tens or even hundreds of watts.

A block diagram of a doubled-up transmitter is shown in figure 5-1. Remember, bands of medium frequencies and below are used almost exclusively for broadcast and are on the air continuously. Doubling up increases reliability. As you can see, two transmitters are located in the same equipment cabinet. This allows you to quickly transfer circuits if one should fail. This dual installation also allows both amplifiers to be used together to double the output power. When you use this application, you sacrifice the doubling-up capability of only the power amplifier. All the other components are still available as backups. Let's go through figure 5-1 and describe the block functions.

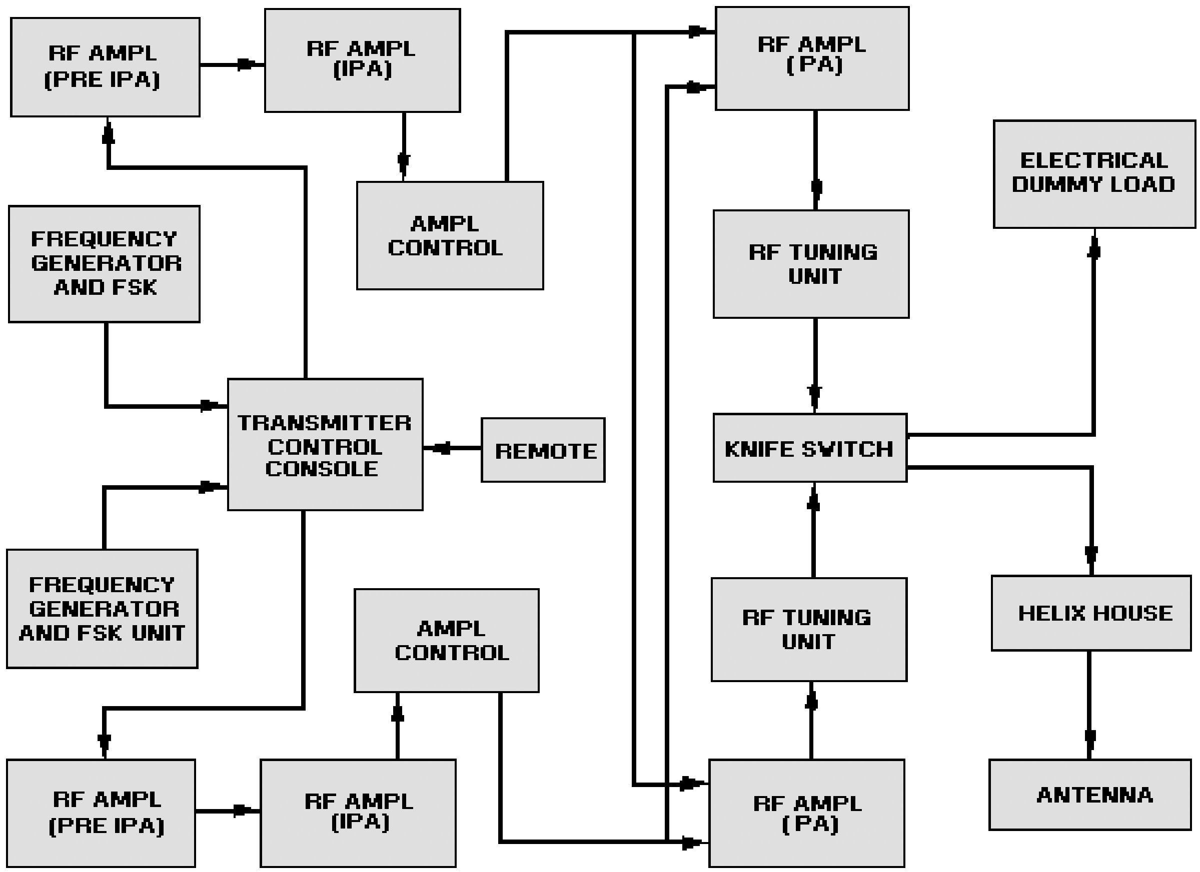


Figure 5-1.—Doubled-up transmitter block diagram.

The frequency generator part of the frequency generator and fsk block is an oscillator. It provides the carrier frequencies for the cw mode. The fsk part is a FREQUENCY SYNTHESIZER (a frequency source of high accuracy). It makes both the mark and space frequencies from a very stable clock oscillator. The keying pulses determine which fsk frequency the keyer chooses to transmit. This signal is then sent to the transmitter control console where it is distributed to the first rf amplifier. This amplifier is referred to as the preliminary intermediate-power amplifier (pre-ipa). The pre-ipa uses linear, untuned, push-pull, rf amplifiers to provide amplified rf to drive other rf amplifiers. The pre-ipa output goes to the intermediate power amplifier (ipa).

The ipa receives the pre-ipa output, amplifies the signal, and drives other selected power amplifiers. The ipa is a single-stage, untuned, linear, push-pull, rf circuit that uses water and forced-air cooled tubes.

Signals are then sent through the amplifier control, where they are used for signal monitoring purposes before being applied to the final rf amplifier (pa). The pa amplifies the signal to the final desired power level. The pa also contains variometers (variable inductors) for coupling. This coupled output is fed to the rf tuning unit.

The rf tuning unit consists of variable oil-filled capacitors and a fixed inductor for frequency tuning. The signal is then sent to a knife switch. This switch simply routes the signal to the DUMMY LOAD or the antenna by way of the HELIX HOUSE. (A dummy load is a nonradiating device the absorbs the rf and has the impedance characteristics of the antenna.) The dummy load is impedance matched to the pa. It allows testing of the pa without putting a signal on the air. When the equipment is in an operating mode, the dummy load is not used. The helix house is a small building physically separated from the transmitter location. It contains antenna loading, coupling, and tuning circuits. The main components consist of a HELIX (large coil) and variable inductors. The signal is fed from the helix directly to the antenna. Sometimes two antennas are used.

Antenna designs vary with the amount and type of land available, desired signal coverage, and bandwidth requirements. Figure 5-2 shows a simplified transmit antenna. The Navy uses TOP-HAT (flat-top) capacitive loading with one or more radiating elements. Typical top hat antennas consist of two or more lengths of wire parallel to each other and to the ground, each fed at or near its mid point. The lengths of wire are usually supported by vertical towers. These antennas may take many shapes. The matching network shown is in the helix house. Figure 5-3 shows the installation at the naval communications unit in Cutler, Maine. The Navy has several of these types of installations. They are used primarily for fleet broadcasts and have power outputs in the .25- to 2-megahertz range. You should notice the transmitter, the location of the helix houses, and the dual antennas. You should also notice the transmission line tunnel. It is underground and over a half-mile long. Figure 5-4, view (A) and view (B), shows another antenna configuration. This array of monopoles (quarter-wave, vertically polarized stubs) is referred to as a TRIATIC antenna. A triatic antenna is a special form of a rhombic-arranged monopole array. This type of array is designed to transmit from a particular location. Triatics are all basically the same but have some design differences at each site. The physical differences compensate for differences in terrain. Now that we have looked at the transmit side, let's look at the receive side.

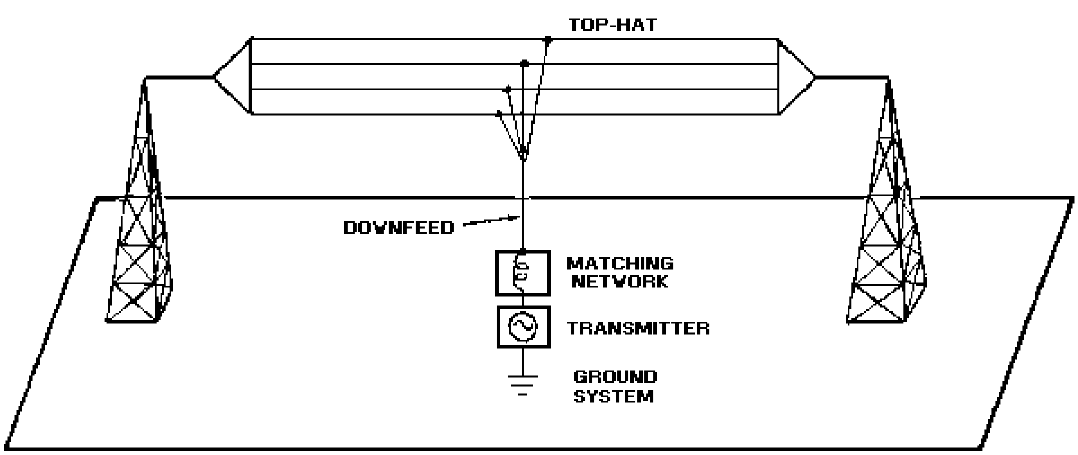


Figure 5-2.—Simplified vlf transmitting antenna.

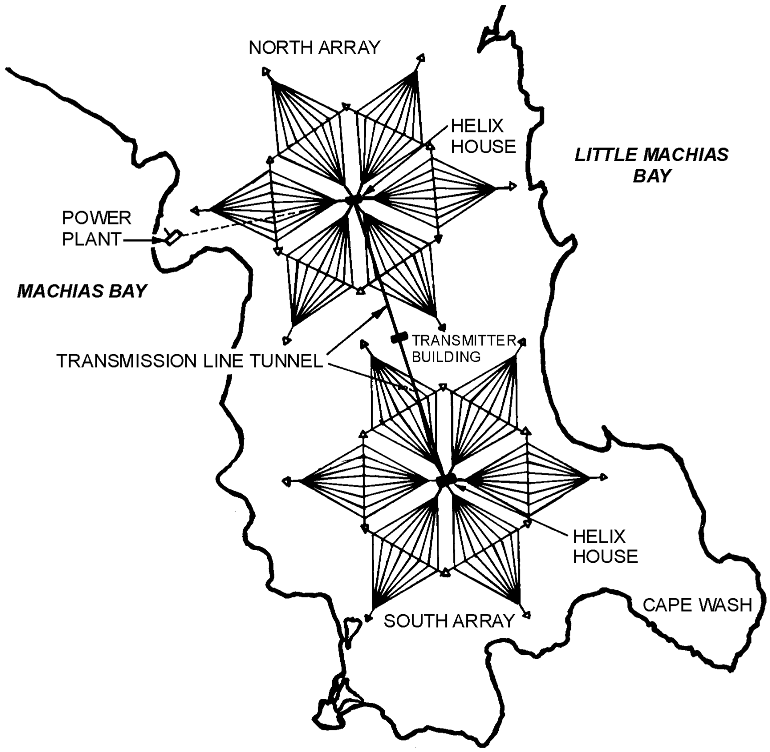


Figure 5-3.—Cutler, Maine antenna installation.

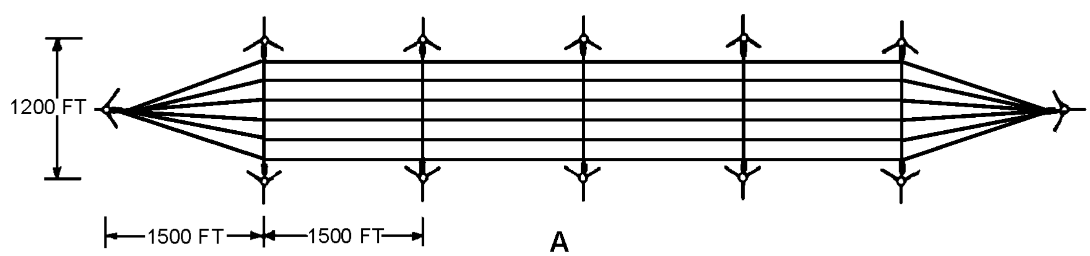


Figure 5-4A.—Triatic type antenna.

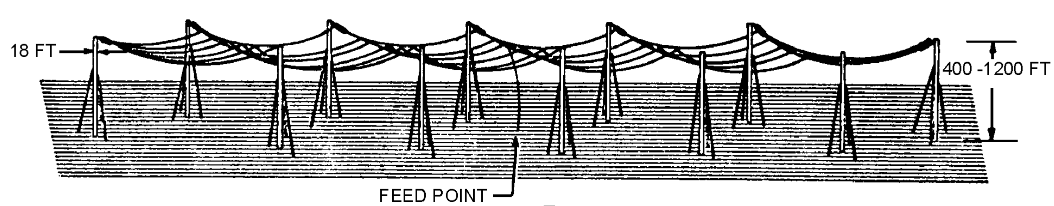


Figure 5-4B.—Triatic type antenna.

Receive Equipment

The receiver you will study here is fundamentally the same as those we covered in chapter 2. A receiver used in this frequency range is about the same electrically as one used in any other range. Figure 5-5 shows the receiver we will discuss. It is a highly sensitive, special purpose receiver because it is capable of splitting-out multiplex signals for detection and reproduction. This receiver covers the frequency range of 3 kilohertz to 810 kilohertz in five bands. It will receive most types of signals, including AM, cw, ssb, fm, and fsk. All operator controls are on the front panel, and a speaker and headset jack permit monitoring.

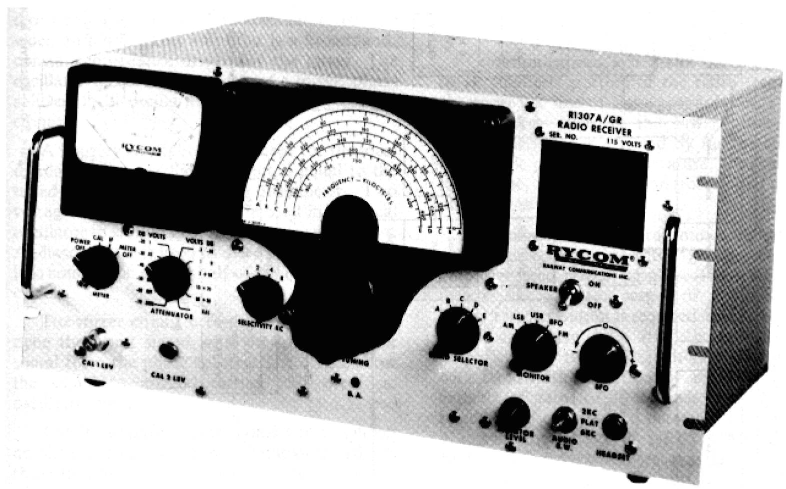


Figure 5-5.—Typical vlf to mf receiver.

Our receiver has five basic stages excluding the power supply. With the exception of a video amplifier in place of an rf amplifier, the circuits perform the functions normally associated with a typical receiver. Figure 5-6 is a block diagram showing the signal paths of the receiver. The input stage consists of a low-pass filter, an attenuator, a calibration oscillator, and a video amplifier. The low-pass filter passes input frequencies below 900 kilohertz. These frequencies are passed to the attenuator, which sets the signal to the proper level to drive the mixer. This minimizes noise and distortion. The calibration oscillator produces a 250-kilohertz output. It is used to calibrate the receiver level and to check for tuning dial accuracy. The input signal is direct-coupled from the attenuator to the video amplifier. This amplifier is a broadband, constant-impedance driver for the mixer. The oscillator-mixer stage consists of a mixer, phase splitter, local oscillator, and frequency control circuits.

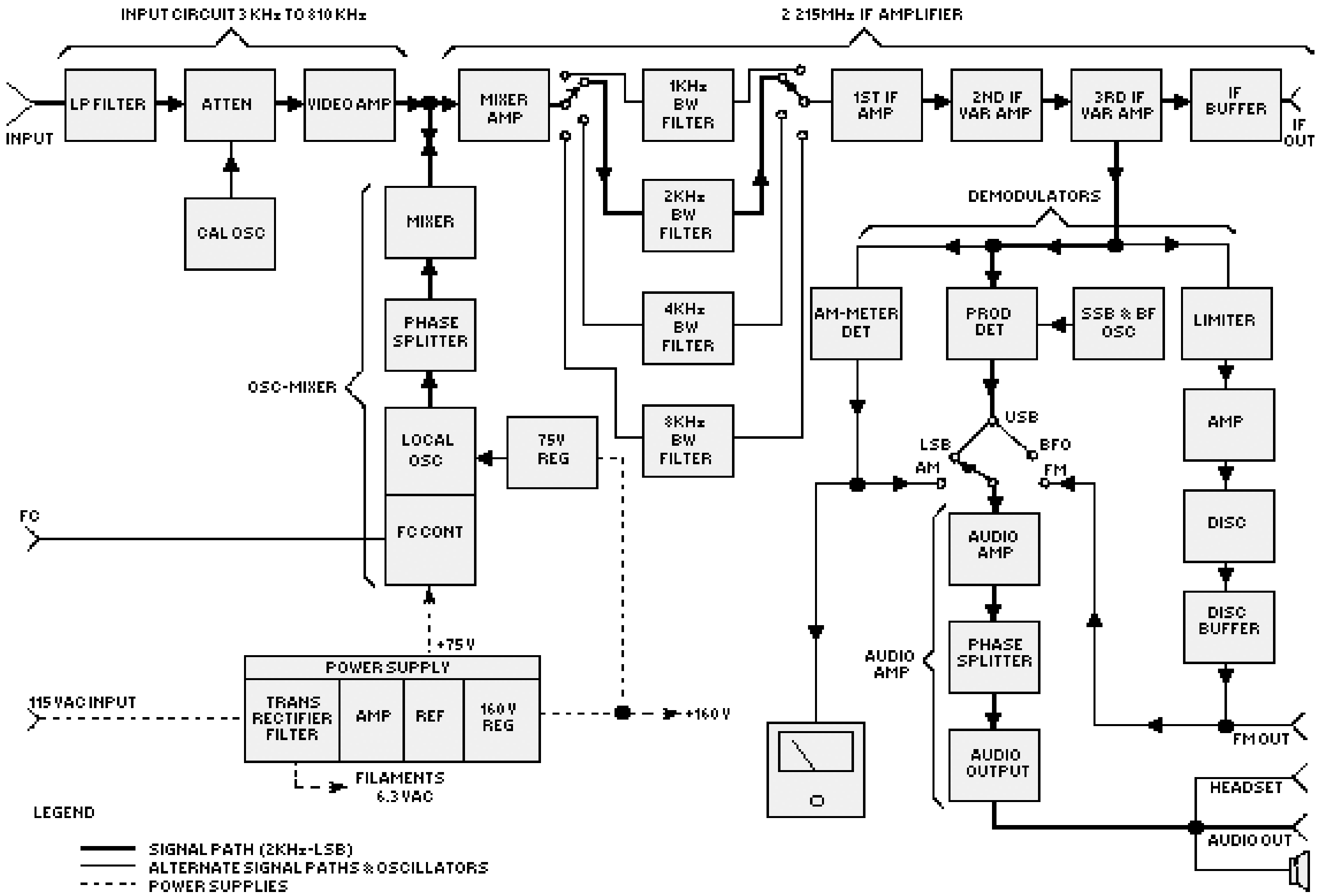


Figure 5-6.—Receiver block diagram.

A Hartley configuration is used for the local oscillator. The oscillator output is equal to the tuned frequency plus 2.215 megahertz. Two voltage-variable capacitors are used in the local oscillator to stabilize small frequency variations. A phase splitter is used to drive the mixer diodes into conduction during half of the local oscillator cycle.

The mixer circuit uses the diodes to heterodyne the input signal with the local oscillator signal from the phase splitter. The diodes short the signal to ground during half the local oscillator cycle.

The IF amplifier stages consist of the mixer amplifier, four selectable bandwidth filters, three IF amplifiers, and an IF buffer amplifier.

The output of the mixer is directly coupled to the mixer amplifier. The IF signal is then directed through one of four bandwidth filters to the first IF amplifier. The signal proceeds to the second and third IF amplifiers for amplification before demodulation. An IF buffer amplifier is used to pass the IF to the IF OUT jack and to isolate this jack from the rest of the circuitry.

Three demodulators are used in this receiver. They are the AM detector, product detector, and fm detector. The AM detector is used to demodulate AM signals. The product detector demodulates ssb, cw, and fsk signals, and the fm detector demodulates fm signals only. An output from the fm detector is provided to the FM OUT jack. This fm output may be used for recording or detailed analysis.

The output from the selected demodulator is amplified by the audio amplifier and presented simultaneously to the HEADSET jack, AUDIO OUT terminals, and the speaker.

You should note that this receiver, as with most others, requires no other special equipment. It uses a standard df loop or a whip antenna. If it is installed in a submarine, a trailed, (towed) long-wire antenna may be used.

MICROWAVE

Communications systems in the 1 gigahertz to 10 gigahertz portion of the radio frequency spectrum use line-of-sight propagation. Propagation takes place in the lower atmosphere (troposphere). It is affected by factors such as barometric pressure, temperature, water vapor, turbulence, and stratification (forming of atmospheric layers).

A typical microwave transmitter includes an exciter group, a modulator group, a power amplifier, and power supplies. The transmitter usually has a power output of about 1 watt. When a higher output is required (about 5 watts), a traveling-wave tube (twt) is used as the amplifier. (A twt is a high-gain, low-noise, wide-bandwidth microwave amplifier. It is capable of gains of 40 decibels or more, with bandwidths of over an octave. The twt was discussed in chapter 2 of NEETS, Module 11, *Microwave Principles*.) A typical microwave receiver contains an rf-IF group, local oscillator, demodulator, and amplifier. Both transmitters and receivers contain special circuits because of the high operating frequencies and critical frequency stability requirements.

Line-of-Sight System

A line-of-sight (los) microwave system consists of one or more point-to-point hops as shown in figure 5-7. Each hop is designed so that it can be integrated into a worldwide communications network. Los systems have many characteristics. In these systems, propagation is only affected by changes in the troposphere. The distance between microwave system hop points ranges from 50 to 150 kilometers (31 to 95 statute miles). These systems are capable of handling up to 600 4-kilohertz voice channels and can also transmit television. These signals can usually be transmitted with less than 10 watts of power. Both the transmit and receive antennas are horn-driven paraboloids that provide high gain and narrow beam widths. In some applications, as shown in figure 5-8, plane reflectors are used with the paraboloids. These systems are very reliable. They are designed to operate over 99 percent of the time. These systems are well adapted to multichannel communications and closed circuit television.

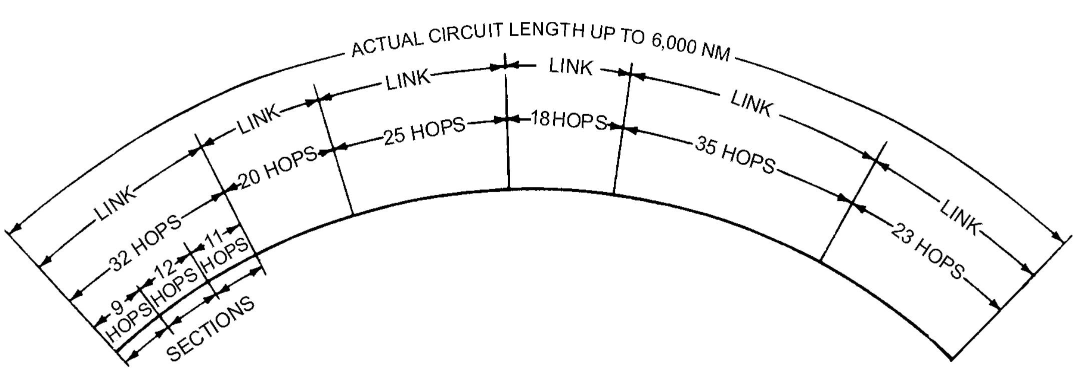


Figure 5-7.—Typical hop-link and section allocation.

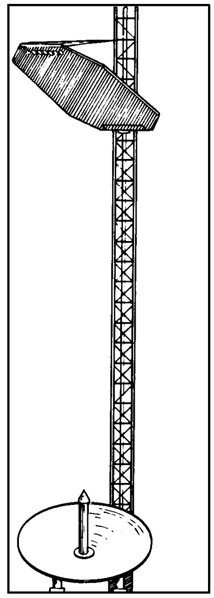


Figure 5-8.—Parabolic antenna and passive reflector combination.

Now let us take a look at another system. It is called the tropospheric-scatter microwave system. But first, you may want to review tropospheric propagation in NEETS, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas.*

Tropospheric Scatter System

A tropospheric-scatter (tropo-scatter) microwave system gets results similar to those of the line-of-sight system. It does it in a different way. The los system uses towers to relay information.

The tropo system uses the turbulence in the layer between the troposphere and the stratosphere to bounce signals back to earth. This method provides several hops and communications beyond los. The propagation reliability and communications capability is the same. The transmission range is up to 800 kilometers (500 statute miles). Transmitter output power may be up to 75 kilowatts depending on the operational requirements. The antennas are horn-driven paraboloids and may be as large as 50 to 60 feet in diameter. Figure 5-9 shows a typical tropospheric-scanner antenna. Remember that hf has a hop distance (skywave) of about 1,400 miles; the distance of one hop for a line-of-sight system is between 31 and 95 miles. The tropospheric-scatter system conveniently fills the gap between these distances.

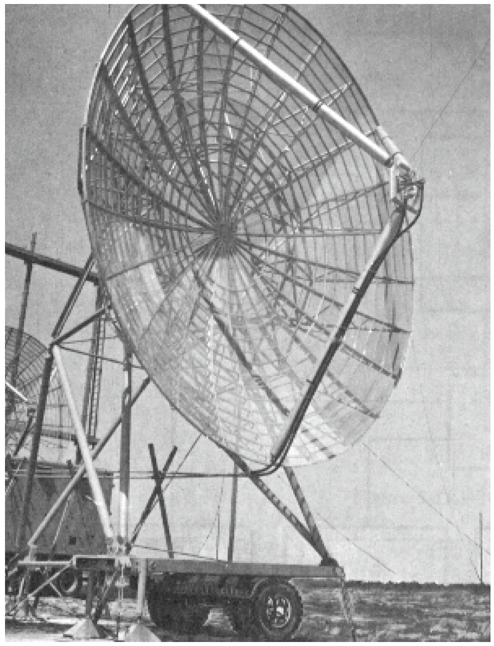


Figure 5-9.—Mobile 30-foot tropospheric-scanner antenna.

Both of these systems are used ashore. You're now going to get a look at a shipboard data information exchange system.

Q1. What is a dummy load?

Q2. What is the function of a product detector?

Q3. What is the frequency range of the mf band?

Q4. Microwave systems use what portion of the atmosphere?

Q5. What is the voice channel capacity of an los communications system?

Q6. What is the one-hop transmission range of a tropospheric-scatter system?

NAVAL TACTICAL DATA SYSTEM

In recent years, the Navy has introduced several new highly technical and effective combat weapons systems. However, these weapons systems did not solve the basic combat command problems that confront our Navy. In combat, a fleet continues to be involved in close-range offense and defense. During close-range combat, the shipboard combat information center (CIC) is involved in complex tactical situations. These situations require intelligent and highly important decisions. Each decision has to be made in a short period of time. You will find the speed at which these combat situations must be solved is inconceivable to someone thinking in terms of typical CIC operations of the recent past. Therefore, the NTDS was developed by the U.S. Navy as a command tool for commanders in tactical combat situations.

Objectives

The naval tactical data system (NTDS) is based on the interaction of humans and machines. The NTDS helps coordinate fleet air defense, antisubmarine warfare, and surface defense operations. Through automation, the NTDS provides commanders with a broad picture of the current tactical situation. It also assists them in directing their operations in time to intercept and destroy all potential enemy threats. The use of digital computers and digital data processing techniques reduces reaction time and increases force effectiveness.

NTDS uses a variety of equipment. This equipment includes transmitters, receivers, cryptographic equipment, high-speed digital computers, magnetic tapes, disks, and a variety of displays. Figure 5-10 shows the NTDS equipment grouping and how it interfaces with the weapons and sensor systems of a ship. Figure 5-11 shows the NTDS system inputs and outputs. As you can see, large amounts and various types of information are provided to or taken from the NTDS. Now that you have seen the types of information associated with the NTDS, let's look at how this information is transmitted and used.

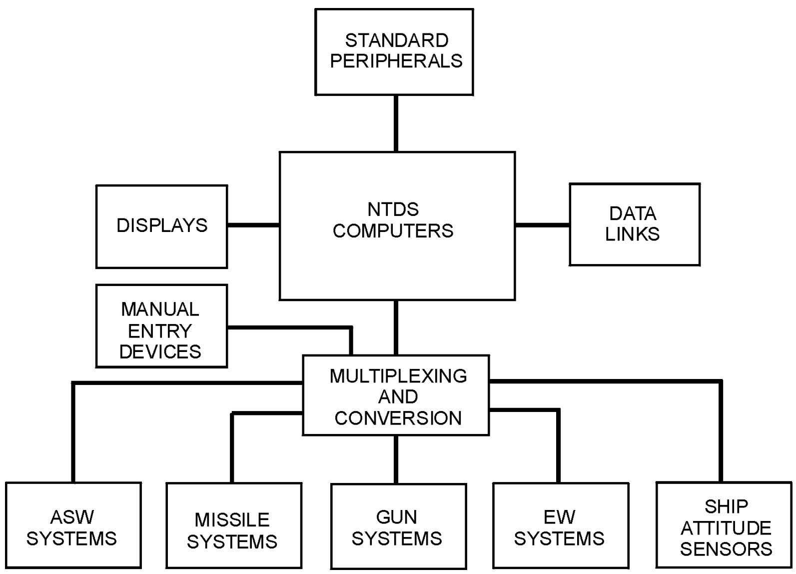


Figure 5-10.—NTDS equipment grouping.

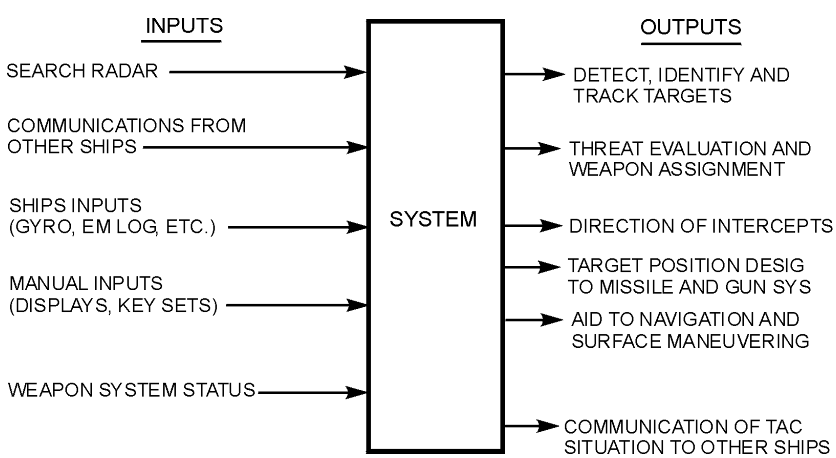


Figure 5-11.—NTDS system inputs/outputs.

NTDS Data Transmission Subsystems

NTDS uses three separate data transmission links to maintain tactical data communications between tactical units. Figure 5-12 illustrates these links. Each link is able to transfer data rapidly to other ships, aircraft, and shore facilities without the delay of human interface (link 14 receive is an exception to this). The data processing subsystem formats the messages for each of the data links. These messages are based on shipboard inputs (from displays, sensors, and other data links).

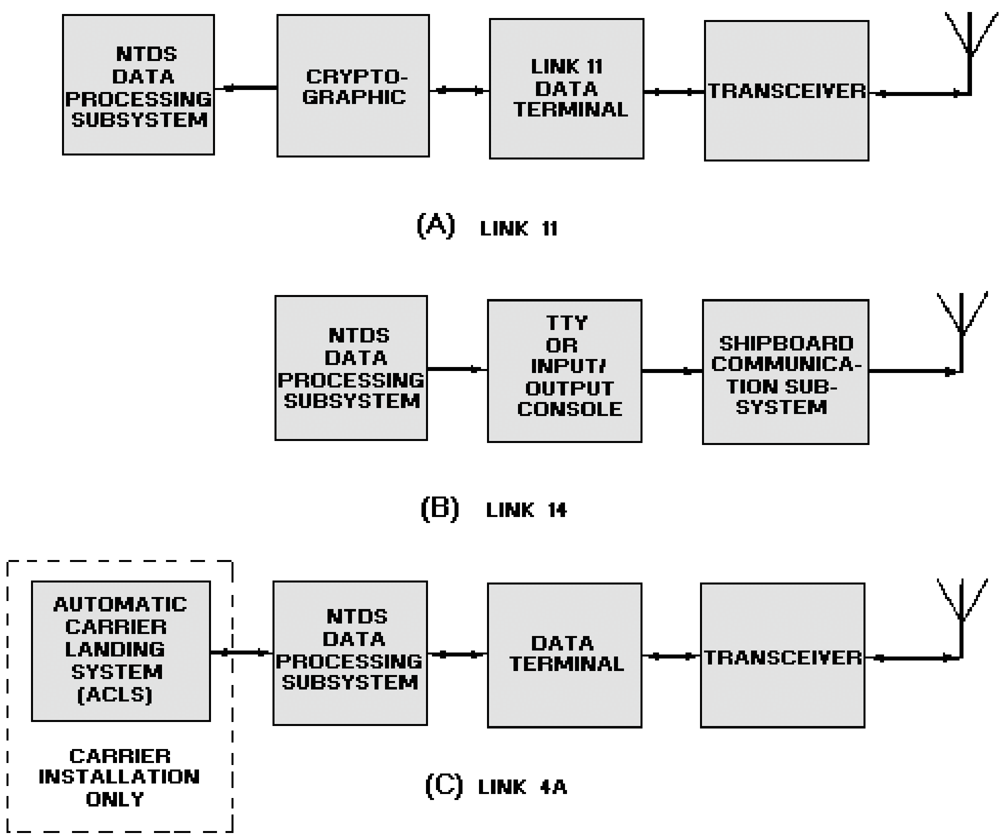


Figure 5-12.—NTDS communications links.

The automatic data communications links provide the operational commander with a high-speed, accurate mode of tactical communications. Link 11 provides high-speed, computer-to-computer transfer of tactical information, command orders, and unit status to all tactical data systems. View A shows you this configuration. The type of tactical information currently transferred is surface, subsurface, air, and EW track information. Data is provided on friendly, hostile, and unknown identity tracks. This broadcast originates through console button actions by the console operators.

Link 14 provides a means of transmitting track information to those units not capable of participating in the link 11 network. View B shows this network. This is a one-way broadcast of information.

Link 4A permits the computer to take control of the autopilot in an equipped aircraft. Also this link can control a plane under other situations. It may control a flight out to a strike area and return it to base without the need for pilot action. The pilot also has the option of overriding the link. The pilot may use the visual display to aid in understanding the intercept controller, or to totally ignore the link 4A transmission. View C shows this link used in conjunction with the automatic carrier landing system (ACLS).

Figure 5-13 is a drawing of an intersystem communications employment diagram. It shows the overall possibilities and flexibility of the NTDS. The new terms shown are defined below:

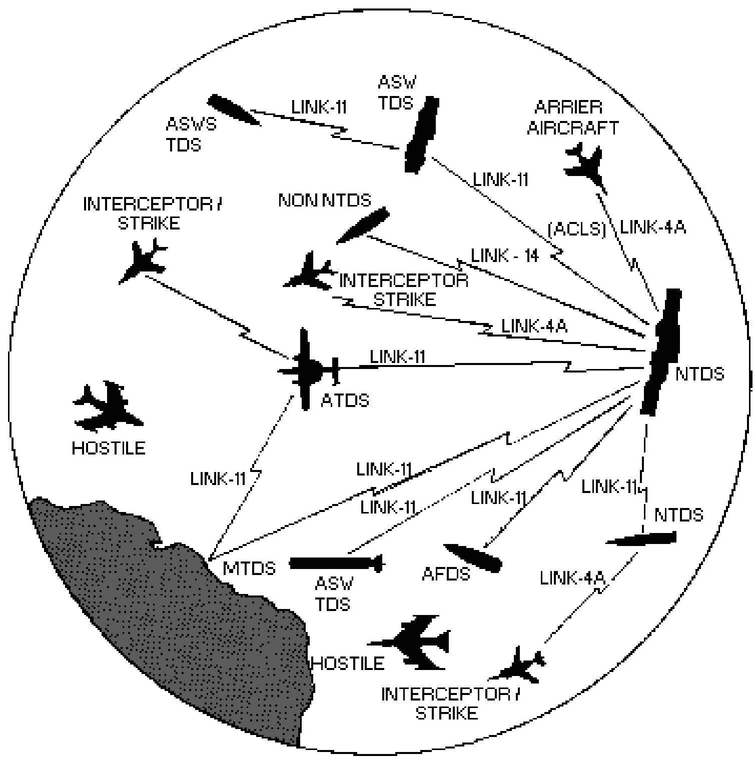


Figure 5-13.—Intersystem communications employment.

- MTDS-Marine tactical data system
- AFDS-Amphibious flagship data system
- ATDS-Airborne tactical data system
- ASWTDS-Antisubmarine warfare tactical data system

Now that we've looked at a complex and stationary system, let's study some fundamental portable equipment.

Q7. What is the primary advantage of NTDS over conventional systems?

Q8. What are the three NTDS data transmission subsystems?

PORTABLE EQUIPMENT

Portable and pack radio sets must be lightweight, compact, and self-contained. Usually, these sets are powered by a battery or a hand generator, have low output power, and are either transceivers or transmitter-receivers. A transceiver consists of a transmitter and a receiver that share common circuits and are housed in the same case or cabinet. A transmitter-receiver is the combination of two separate pieces of equipment that are used together. Navy ships carry a variety of these radio sets for emergency and amphibious communications. The numbers and types of equipment vary according to the individual ship.

EMERGENCY EQUIPMENT

One piece of emergency equipment is shown in figure 5-14. It is a rugged emergency transmitter carried aboard ships and aircraft for use in lifeboats and life rafts. The transmitter operates on the international distress frequency (500 kilohertz) and the survival craft communications frequency (8,364 kilohertz).

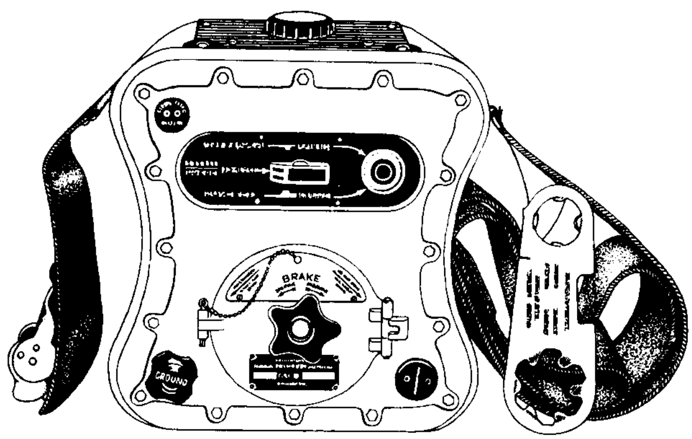


Figure 5-14.—Typical emergency lifeboat transmitter.

The complete radio transmitter, including the power supply, is contained in an aluminum cabinet that is airtight and waterproof. The cabinet is shaped to fit between the legs of the operator and has a strap for securing it in the operating position.

The only operating controls are a three-position selector switch and a push-button telegraph key. A hand crank screws into a socket in the top of the cabinet. The generator, automatic keying, and automatic frequency changing are all operated by turning the hand crank. While the hand crank is being turned, the set automatically transmits the distress signal SOS in Morse code. The code consists of six groups of SOS followed by a 20-second dash. It is transmitted alternately on 500 kilohertz and 8,364 kilohertz. The frequency automatically changes every 50 seconds. These signals are intended for reception by two groups of stations, each having distinct rescue functions. Direction-finding stations cooperating in long-range rescue operations normally use 8,364 kilohertz, whereas aircraft or ships locally engaged in search and rescue missions use 500 kilohertz.

Besides the automatic feature, you can key the transmitter manually on 500 kilohertz only. This is done by means of a push-button telegraph key.

Additional items (not shown) packaged with the transmitter include the antenna, a box kite, and balloons for supporting the antenna. Hydrogen-generating chemicals for inflating the balloon and a signal lamp that can be powered by the hand-crank generator are also included.

The equipment floats and is painted brilliant orange-yellow to provide good visibility against dark backgrounds.

A transceiver is shown in figure 5-15. It is portable, battery powered, and has two channels. It provides homing information and two-way voice communications between life rafts and searching ships and aircraft. This transceiver is a microminiature, solid-state, hand-held radio that operates on the 121.5-megahertz and the 243-megahertz guard channels. The transceiver has four operating controls. These are the volume (VOL) control, the two-position FREQUENCY SELECTOR, the PUSH-TO-TALK/ CODE button, and the three-position MODE switch.

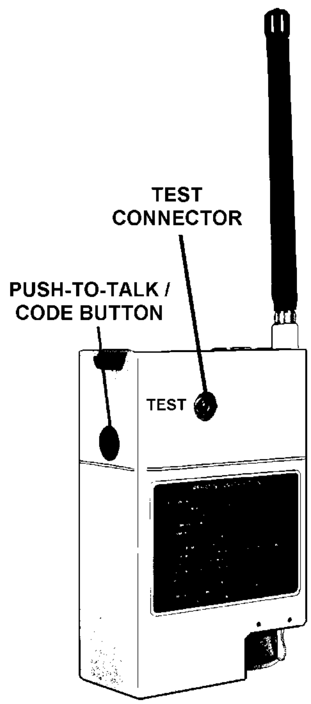


Figure 5-15.—Emergency portable transceiver.

The antenna is a rubber covered, omnidirectional, flexible whip antenna that is 7.74 inches long. The batteries supplied with the radio set are lithium D cells. Each cell is fused to protect against damage from external short circuits. Two cells are installed in the transceiver and four are packaged as spares.

OPERATIONAL EQUIPMENT

An operational transceiver is shown in figure 5-16. It is watertight, lightweight, portable, and operates in the vhf and uhf range. You can use any of 1,750 channels, spaced 200 kilohertz apart, in the 225-400 megahertz range. Its mode of operation is AM voice and it supplies an average output power of 3 watts. It was designed mainly for manpack (backpack) use, but it may also be used at a fixed station or in vehicles when certain accessories are added. When not in use, the equipment is disassembled and stowed in a special aluminum case similar to an ordinary suitcase.

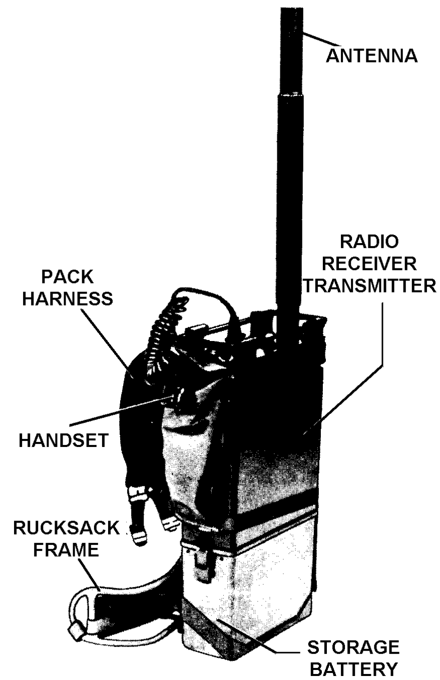
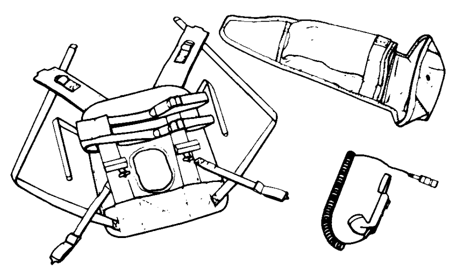
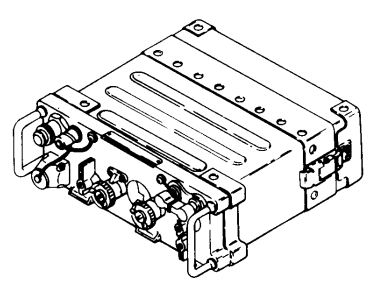


Figure 5-16.—Typical vhf/uhf backpack transceiver.

Figure 5-17 shows a typical vhf miniaturized manpack radio set. View A shows the pack frame, the handset, and the canvas accessory pouch. The pouch contains two antennas (one flexible steel band-type whip and one collapsible rigid whip). The handset fits in the pouch when not in use. View B shows the transmitter-receiver.



A



B

Figure 5-17.—Vhf receiver-transmitter.

Now that you have learned about portable equipment, let's look at one of the newest areas of communications. You are going to learn the fundamentals of how a laser works and how it may be used in the field of communications.

Q9. What are the three main design considerations of portable equipment?

LASERS

The word LASER is an acronym for light amplification by stimulated emission of radiation. The laser is widely used in industry, and experimental work is being done with it in communications. You will find a laser is similar to uhf and microwave power sources and could replace either of them in point-to-point communications.

THEORY OF OPERATION

Lasers take energy at (or near) the visible light spectrum and convert it to a very narrow and intense beam in the same region. A close relative of the laser is the light emitting diode (LED). The LED takes dc or low frequency ac power and converts the energy into visible light.

The principle of the laser is somewhat similar to that of a very high-Q cavity resonator. Chapter 1 of NEETS, Module 11, *Microwave Principles*, explains cavity resonators. The laser is shock-excited by a spark transmitter. This transmitter is called a spark transmitter because it uses the discharge of a capacitor through an inductor and a spark gap as a source of rf. While the input energy of the laser covers a wide band of frequencies, the output is on one frequency. Energy outputs of the laser are either INCOHERENT or COHERENT. For example, if you turn on a transmitter with no modulation, you will get coherent radiation. When you connect a noise source to an antenna, the result is incoherent radiation.

Lasers can be either cw or pulsed. Actually, lasers are little different from conventional oscillators. However, the way lasers convert energy from one form to another is quite different. In conventional oscillators, dc power from the collector is converted to rf energy. The frequency is for the most part independent of the molecular or atomic structure of the generator. This is not true for the laser. Laser conversion takes place directly within the molecular structure of a crystal or gas. The external circuits have little effect on actual output frequency. The fact that the light from an LED is always the same color results from similar conditions. In a laser, incoherent light excites the electrons in the atoms to higher energy levels than they normally would have. The new energy states are unstable and the electrons drop down to lower energy levels. Energy is then released in the form of light.

Figure 5-18 shows the ends of the crystal or glass tube laser with light waves reflecting back and forth between two mirrored surfaces. One mirror is only partially reflective, and light energy is transmitted through it to form the light beam. You will find that power sources for lasers include flash tubes or, in the case of diode-type lasers, dc power supplies.

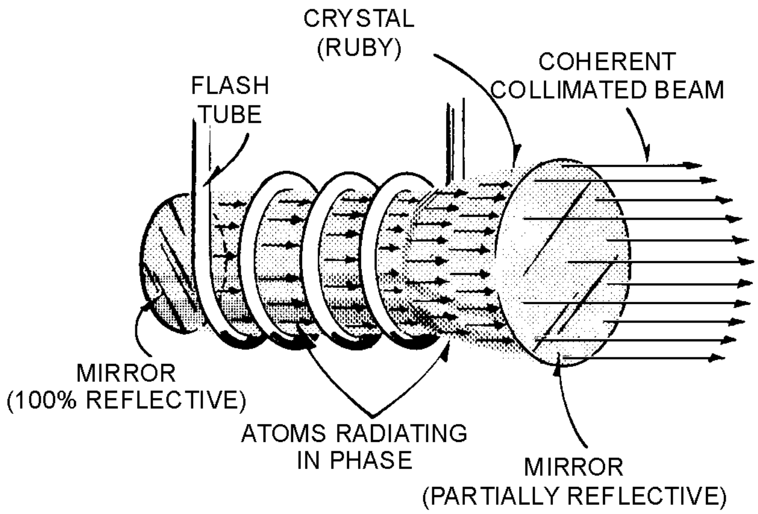


Figure 5-18.—Simple laser.

COMMUNICATIONS APPLICATIONS

High-energy lasers have very small power losses. As a result, a laser with a 2- or 3-inch initial beam can be used to communicate directly with a distant planet. However, from a communications standpoint, they do have some drawbacks. While laser energy is capable of being formed into a very narrow beam, it is still subject to the same difficulties as any other form of light transmission. Fog and other adverse weather conditions can absorb the light. Small atmospheric temperature variations can cause deflection and scattering. An optical waveguide was designed to help overcome this problem. It consists of a thin dielectric fiber clothed by another dielectric coating several wavelengths thick. It has been successfully used to move the beam over considerable distances and around bends. The use of optical waveguides is known as FIBER OPTICS. A number of fibers can be paralleled to reduce the attenuation through the waveguide. This technique has resulted in an inexpensive telephone system with a bandwidth greater than that of previous methods. Researchers are attempting to develop a laser that will operate in the blue-green portion of the visible spectrum. Water offers little attenuation to the blue-green band of frequencies. Because of this, blue-green communications lasers could possibly penetrate the ocean to great depths.

This could give us a very secure undersea communications link.

Q10. Lasers operate in what portion of the frequency spectrum?

Q11. What are the two types of lasers?

Q12. What are the effects of adverse weather on the laser beam?

SUMMARY

Now that you have completed this chapter, a review of what you have learned is in order. The following summary will refresh your memory of new terms.

DOUBLING UP is a type of two-equipment installation where one unit can be substituted for another in the event of failure.

FREQUENCY SYNTHESIZER is a frequency source of high accuracy.

DUMMY LOAD is a nonradiating device that absorbs the rf and has the impedance characteristics of the antenna.

HELIX HOUSE is a building at a transmitter site that contains antenna loading, coupling, and tuning circuits.

A **HELIX** is a large coil of wire. It acts as a coil and is used with variable inductors for impedance matching of high-power transmitters.

TOP-HAT antennas are center-fed and capacitively loaded.

TRIATIC is a special type of monopole antenna array.

LASER is an acronym for light amplification by stimulated emission of radiation.

COHERENT refers to radiation on one frequency or nearly so.

INCOHERENT refers to radiation on a broad band of frequencies.

FIBER OPTICS are conductors or optical waveguides that readily pass light.

MTDS is an abbreviation for the marine tactical data system.

AFDS is an abbreviation for the amphibious flagship data system.

ATDS is an abbreviation for the airborne tactical data system.

ASWTDS is an abbreviation for the antisubmarine warfare tactical data system.

ANSWERS TO QUESTIONS Q1. THROUGH Q12.

A1. An impedance-matched device capable of absorbing all of a transmitters power.

A2. It demodulates ssb, cw, and fsk signals.

A3. 300 kilohertz to 3 megahertz.

A4. Troposphere.

A5. Up to 600 4-kilohertz channels.

A6. Up to 800 kilometers (500 statute miles).

A7. Speed.

A8. Links 4A, 11, and 14.

A9. They must be lightweight, compact, and self-contained.

A10. At or near visible light.

A11. Cw or pulsed.

A12. It absorbs it.

APPENDIX I

GLOSSARY

ACTIVE SATELLITE—A satellite that amplifies the received signal and retransmits it back to earth.

AFDS—An abbreviation for the amphibious flagship data system.

ANGLE OF INCLINATION—The angular difference between the equatorial plane of the earth and the plane of orbit of the satellite.

ANTENNA COUPLER—A device used for impedance matching between an antenna and a transmitter or receiver.

APOGEE—The point in the orbit of a satellite the greatest distance from the earth.

ASSEMBLY—A number of parts or subassemblies, or any combination thereof, joined together to perform a specific function.

ASWTDS—An abbreviation for the antisubmarine warfare tactical data system.

ASYNCHRONOUS ORBIT—One where the satellite does not rotate or move at the same speed as the earth.

ATDS—An abbreviation for the airborne tactical data system.

AUDIO FREQUENCY TONE SHIFT—A system that uses amplitude modulation to change dc mark and space impulses into audio impulses.

AUTOMATIC TRACKING—Tracking done by the equipment comparing the direction of the antenna axis and the direction of the received signal.

AUTOMATIC VOLUME/GAIN CONTROL—A circuit used to limit variations in the output signal strength of a receiver.

BALANCED PHASE DETECTOR—A circuit that controls the oscillator frequency (afc).

BAUD—A measurement of speed based on the number of code elements or units per second.

BEAT-FREQUENCY OSCILLATOR—An additional oscillator used in a receiver when receiving a cw signal. It provides an audible tone.

BITS-PER-SECOND—Bit is an acronym of the words binary digit. One bit is equal to one signal unit or element.

BLACK—The reference color of equipment that passes unclassified information. It normally refers to patch panels.

CODE—In teletypewriter operation, code is a combination of mark and space conditions representing symbols, figures, or letters.

COHERENT—Radiation on one frequency.

COMPARATOR—An equipment that compares incoming signals and selects the strongest to be fed to a teletypewriter through a patch panel. This is used in diversity operation.

CONVERTER—An equipment that changes the audio output of a receiver to dc pulses. These pulses are fed to a tty to indicate marks and spaces.

DETECTION—The separation of low-frequency (audio) intelligence from the high (radio) frequency carrier.

DOUBLING UP—This is a type of two-equipment installation where one unit can be substituted for another in the event of failure.

DOWN LINK—The frequency used to transmit an amplified signal from the satellite back to earth.

DUMMY LOAD—A nonradiating device that absorbs the rf and has the impedance characteristics of the antenna.

ECLIPSE—This occurs when the satellite is not in view or in direct line of sight with the sun. This happens when the earth is between them.

ELECTROMAGNETIC INTERFERENCE—A term used to describe the degradation of a receiver or system.

EPHEMERIS—A table showing the precalculated position of a satellite at any given time.

EQUATORIAL ORBIT—An orbit that occurs when the plane of a satellite coincides with the plane of the earth at the equator.

EXTREMELY HIGH FREQUENCY—The band of frequencies from 30 gigahertz to 300 gigahertz.

EXTREMELY LOW FREQUENCY—The band of frequencies up to 300 hertz.

FACSIMILE—The method for transmitting and receiving still images. These images can be maps, photographs, and handwritten or printed text.

FADING—The variations in signal strength at the antenna of a receiver.

FIBER OPTICS—Conductors or optical waveguides that readily pass light.

FIDELITY—The ability of a receiver to accurately reproduce, at its output, the signal at its input.

FORWARD AGC—The type of agc that causes an amplifier to be driven towards saturation.

FRAMING—The process of synchronizing a facsimile receiver to a transmitter. This allows proper picture reproduction.

FREQUENCY-DIVISION MULTIPLEXING—Multiplexing that transmits and receives the full 360 degrees of each sine wave.

FREQUENCY SYNTHESIS—A process that uses hetrodyning and frequency selection to produce a signal.

FREQUENCY SYNTHESIZER—A frequency source of high accuracy.

GANGED TUNING—The process used to tune two or more circuits with a single control.

GROUP—A collection of units, assemblies, subassemblies, and parts. It is a subdivision of a set or system but is not capable of performing a complete operational function.

HAND OVER—The operation where one earth terminal yields control to another as a satellite moves out of its area of coverage.

HARMONIC—An exact multiple of the fundamental frequency. Even harmonics are 2, 4, and so on, times the fundamental. Odd are 3, 5, and so on, times the fundamental frequency.

HELIX—A large coil of wire. It acts as a coil and is used with variable inductors for impedance matching of high-power transmitters.

HELIX HOUSE—A building at a transmitter site that contains antenna loading, coupling, and tuning circuits.

HETERODYNING—The mixing of the incoming signal with the local oscillator frequency. This produces the two fundamentals and the sum and difference frequencies.

HIGH FREQUENCY—The band of frequencies from 3 megahertz to 30 megahertz.

IMAGE FREQUENCY—An undesired frequency capable of producing the desired frequency through heterodyning.

INCLINED ORBIT—Orbits where there is some amount of inclination. These include equatorial and polar orbits.

INCOHERENT—This refers to radiation on a broad band of frequencies.

INTELLIGENCE—Any signal that conveys information (voice, teletypewriter, facsimile).

KEYER—A device that changes dc pulses to mark and space modulation for teletypewriter transmissions.

LASER—An acronym for light amplification by stimulated emission of radiation.

LISSAJOUS PATTERN—A combined, simultaneous display of the amplitude and phase relationships of two input signals on a CRT.

LOW FREQUENCY—The band of frequencies from 30 kilohertz to 300 kilohertz.

MARKING—The state where a circuit is closed and current flows in teletypewriter operation.

MEDIUM ALTITUDE ORBIT—An orbit from 2,000 to 12,000 miles above the earth. The rotation rate of the earth and satellite are quite different, and the satellite moves quickly across the sky.

MEDIUM FREQUENCY—The band of frequencies from 300 kilohertz to 3 megahertz.

MTDS—An abbreviation for the marine tactical data system.

MULTICOUPLERS—Couplers patch receivers or transmitters to antennas. They also filter out harmonics and spurious responses, and impedance-match the equipment.

MULTIPLEXING—A method for simultaneous transmission of two or more signals over a common carrier wave.

NEAR SYNCHRONOUS ORBIT—An orbit in which the satellite rotates close to but not exactly at the same speed as the earth.

NEUTRAL—The teletypewriter operation where current flow represents a mark and no flow represents a space.

NOISE SILENCER, NOISE SUPPRESSOR, OR NOISE LIMITER—Circuits that clip the peaks of the noise spikes in a receiver.

NONSYNCHRONOUS—The teletypewriter operation where both transmitter and receiver do not operate continuously

ORDER-WIRE CIRCUIT—A circuit between operators used for operations control and coordination.

PAGE PRINTER—A high-speed printer that prints teletypewriter characters one at a time in a full-page format.

PASSIVE SATELLITE—A satellite that reflects radio signals back to earth.

PATCH PANEL—A panel used to tie a receiver or transmitter to its associated equipment.

PART—A part is one component or two or more components joined together. It is not normally subject to disassembly without destruction.

PERFORATOR—A device that stores a teletypewriter message on a paper tape. It may be stored for later transmission.

PERIGEE—The point in the orbit of a satellite closest to the earth.

PERMANENT MAGNET SPEAKER—A speaker with a permanent magnet mounted on soft iron pole pieces.

POLAR—The teletypewriter operation where current flow of one polarity represents a mark and current of the opposite polarity represents a space.

POLAR ORBIT—An orbit that has an angle of inclination of or near 90 degrees.

PROGRAMMED TRACKING—The method that uses known satellite orbital parameters to generate antenna pointing angles.

RADIO COMMUNICATIONS—The term describing teletypewriter, voice, telegraphic, and facsimile communications.

RADIO FREQUENCY CARRIER SHIFT—The system that uses a keyer to shift a radio frequency signal above or below an assigned frequency. These shifts correspond to marks and spaces.

RADIO SET CONTROL UNIT—Equipment used to remotely control certain transmitter and receiver functions.

RECEIVER—Equipment that converts electromagnetic energy into a visible or an audible form.

RECEIVER TRANSFER SWITCHBOARD—Equipment used to transfer receiver audio outputs to remote control station audio circuits.

RECEPTION—The instant when an electromagnetic wave passes through a receiver antenna and induces a voltage in that antenna.

RED—The reference color of equipment that passes classified information. It normally refers to patch panels.

REPEATER—Another name for an active satellite.

REPERFORATOR—Equipment that converts the incoming tty signal and stores it on paper tape.

REPRODUCTION—The process of converting electrical signals to sound waves. This sound is speech, music, and so on.

REVERSE AGC—The type of agc that causes an amplifier to be driven toward cut-off.

RUNNING OPEN—The teletypewriter condition where the type hammer constantly strikes the type box but does not print or move across the page.

SATELLITE ECLIPSE—An eclipse where the rays of the sun don't reach the satellite. This prevents recharging of the solar cells of the satellite and decreases the power to the transmitter.

SATELLITE-SUN CONJUNCTION—A period when the satellite and sun are close together and the noise from the sun prevents or hampers communications.

SCANNING—The process of subdividing a picture in an orderly manner into segments. This is used in facsimile transmission.

SELECTIVITY—The ability of a receiver to select the desired signal and reject unwanted signals.

SENSITIVITY—The ability of a receiver to reproduce very weak signals. The greater the receiver sensitivity, the weaker the signal that will be reproduced.

SET—A unit or units and the assemblies, subassemblies, and parts connected or associated together to perform a specific function.

SPACING—The condition in teletypewriter operation where a circuit is open and no current flows.

SQUELCH—A circuit that cuts off the output of a receiver when there is no input.

START—The first unit of a teletypewriter signal. It is always a space.

STOP—The last unit of a teletypewriter signal. It is always a mark.

SUBASSEMBLY—Consists of two or more parts that form a portion of an assembly or a unit.

SUBHARMONIC—An exact submultiple of the fundamental frequency. Even subharmonics are one-half, one-quarter, and so on. Odd subharmonics are one-third, one-fifth, and so on of the fundamental frequency.

SUPERHIGH FREQUENCY—The band of frequencies from 3 gigahertz to 30 gigahertz.

SUPPRESSION—The process of eliminating an undesired portion of a signal.

SYNCHRONOUS—A type of teletypewriter operation where both transmitter and receiver operate continuously.

SYNCHRONOUS ORBIT—An orbit in which the satellite moves or rotates at the same speed as the earth.

SYSTEM—A combination of sets, units, assemblies, subassemblies, and parts joined together to form a specific operational function or several functions.

TELECOMMUNICATIONS—The transmission, emission, or reception of signs, signals, writings, images, or sounds. This is done by visual, oral, wire, radio, or other electromagnetic means.

TELETYPEWRITER—A machine that can transmit and or receive letters, numbers, or symbols. It may have a keyboard similar to a typewriter.

TEMPEST—A term normally used to describe compromising emanations. These emanations are unintentionally radiated signals that could disclose classified information.

TIME-DIVISION MULTIPLEXING—The process that periodically samples the full 360 degrees of each sine wave. The sample can be of a received signal or of a signal to be transmitted.

TONE-TERMINAL SET—Equipment that converts tty dc pulses into audio tones for modulation of a transmitter in audio-frequency-tone shift transmissions.

TOP-HAT—An antenna that is center-fed and capacitively loaded.

TRANSITION—The time it takes to shift from a mark to a space condition or from a space to a mark condition.

TRANSMITTER—Equipment that generates and amplifies an rf carrier, modulates the rf carrier with intelligence, and radiates the signal into space.

TRANSMITTER DISTRIBUTOR—A device that reads Baudot code from paper tape and allows a message to be printed on a page printer.

TRANSMITTER TRANSFER SWITCHBOARD—Equipment that selectively transfers remote control station functions and signals to transmitters.

TRIATIC—A special type of monopole antenna array.

ULTRAHIGH FREQUENCY—The band of frequencies from 300 megahertz to 3 gigahertz.

UNIT—An assembly or any combination of parts, subassemblies, and assemblies mounted together. Normally capable of independent operation.

UP LINK—The frequency used to transmit a signal from earth to a satellite.

VERY HIGH FREQUENCY—The band of frequencies from 30 megahertz to 300 megahertz.

VERY LOW FREQUENCY—The band of frequencies from 3 kilohertz to 30 kilohertz.

WORDS-PER-MINUTE—An approximate rate of speed. It means the number of five letter words with a space between them that can be transmitted or received in a one-minute period.

ZONE OF MUTUAL VISIBILITY—The area where the satellite can be seen by both the up- and down-link earth terminals.

APPENDIX II

REFERENCE LIST

CHAPTER ONE

ELF Communications System Program, NAVELEXINST 5430.26, October 1983.

Shipboard Electronics Materials Officer, NAVEDTRA 10478-A, 1982.

Electronics Technician, 3 & 2, Vol. 3, NAVEDTRA 10198, 1979.

Cryptologic Technician M 3 & 2, NAVEDTRA 10232-C, 1978.

Radioman 3 & 2, NAVEDTRA 10228-G, 1978.

Howeth, L. S., *History of Communications-Electronics in the United States Navy,* Bureau of Ships and Of fice of Naval History, Washington, D.C., 1963.

CHAPTER TWO

Electronics Technician 3 & 2, Vol. 3, NAVEDTRA 10198, 1979.

Cryptologic Technician M 3 & 2, NAVEDTRA 10232-C, 1978.

Radioman 3 & 2, NAVEDTRA 10228-G, 1978.

Cryptologic Collection Equipments, NAVEDTRA 10251, 1977.

Basic Electronics, Vol. 1, NAVPERS 10087-C, 1970.

Basic Electronics, Vol. 2, NAVPERS 10087-C1, 1979.

Fundamentals of Single Sideband, NAVELEX 0967-LP-222-2010, 1958.

CHAPTER THREE

Shipboard Electronics Material Officer, NAVEDTRA 10478-A, 1982.

Electronics Technician 3 & 2, Vol. 3, NAVEDTRA 10198, 1979.

Cryptologic Technician M 3 & 2, NAVEDTRA 10232-C, 1978.

The Electronics Material Officer's Guide to Shipboard Electromagnetic Interference Control, NAVSEASYSCOM STD-407-5287556, 1983.

CHAPTER FOUR

Shipboard Electronics Material Officer, NAVEDTRA 10478-A, 1982.

Electronics Technician 3 & 2, Vol. 3, NAVEDTRA 10198, 1979.

Cryptologic Technician M 3 & 2, NAVEDTRA 10232-C, 1978.

Radioman 3 & 2, NAVEDTRA 10228-G, 1978.

Satellite Communications Systems, United States Air Force, KEO-90008, 1972.

Fleet Satellite Communications System Description, Naval Ocean Systems Center, January 1978.

CHAPTER FIVE

Shipboard Electronics Material Officer, NAVEDTRA 10478-A, 1982.

Electronics Technician 3 & 2, Vol. 3, NAVEDTRA 10198, 1979.

Microwave and Tropo-Scatter Communications Systems, NAVELEX 0101,112, 1972.

Vlf, Lf and Mf Communications Systems, NAVELEX 0101,113, 1972.

Design Handbook for Optical Fiber Systems, U.S. Army Communications-Electronics, CCC-CED-XES-78-0l, 1978.

Cooper, Robert S., "SPACE Another Dimension of Defense," *Defense 84,* August 1984, pp. 12-19.

McGraw-Hill, Vol. 3, *Encyclopedia of Science and Technology*, 1971.

McGraw-Hill, Vol. 12, *Encyclopedia of Science and Technology*, 1971.

MODULE 17 INDEX

G

H

I

L

M

N

O

P

R

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Introduction to Radio-Frequency Communications," pages 1-1 through 1-20. Chapter 2, "Introduction to Communications Theory," pages 2-1 through 2-37.

1-1. For naval communications to be effective, all involved must be top performers. Reliable, secure, and timely receiving and transmitting of information is the goal. Which of the following requirement standards are used to determine whether or not this goal has been met?

1. Wartime
2. Peacetime
3. Cold war less 10 percent
4. Nuclear war less 25 percent

1-2. What are the two types of electrical communications?

1. Radio and wire
2. Television and wire
3. Telegraph and radio
4. Television and radio

1-3. Which of the following terms includes intelligence produced by wire, radio, visual means, oral means or electromagnetic systems?

1. Telecommunications
2. Radiotelegraph
3. Electrolysis
4. Photocopy

1-4. Radiotelegraph (cw) is valuable when communicating to, from, and among widely separated naval units. What is the main advantage of the cw mode?

1. Cost
2. Speed
3. Security
4. Reliability

1-5. Tactical communications is usually considered as line-of-sight. What maximum distance is normally within the line-of-sight range?

1. 5 miles
2. 25 miles
3. 50 miles
4. 250 miles

1-6. You want to transmit high-speed automatic page or tape copy across an ocean area. Which of the following methods should you choose?

1. Facsimile
2. Radiotelegraph
3. Radiotelephone
4. Radioteletypewriter

1-7. What method is normally used to transmit graphs electronically?

1. Facsimile
2. Radio teletypewriter
3. Frequency-shift keying
4. Audio-frequency-tone shifting

1-8. When you use subdivisions to assign reference designators to equipment, what is the designator of (a) the largest and (b) the smallest?

1. (a) System (b) set
2. (a) Set (b) unit
3. (a) Unit (b) assembly
4. (a) System (b) part

1-9. When using the reference designator 1A6A4J6, what level does the number 1 refer to?

1. Subassembly
2. Assembly
3. Group
4. Unit

1-10. What is the total number of frequency bands the military is currently using for communications?

1. 5
2. 7
3. 9
4. 11

1-11. Extremely low-frequency transmissions are primarily directed at which of the following users?

1. Shore installations
2. Surface ships
3. Submarines
4. Aircraft

1-12. Vlf transmitters are used primarily for which of the following purposes?

1. Navigation and fleet communications
2. Frequency standards and time signals
3. Astronomy and oscillator calibration
4. Aircraft control and space vehicle tracking

1-13. Navy use of the low-frequency band is mainly for which, if any, of the following broadcasts?

1. Ship to shore
2. Fleet multichannel
3. Space vehicle telemetry
4. None of the above

1-14. For which of the following reasons does the Navy use only the upper and lower ends of the mf band?

1. They are the most reliable
2. They produce the best propagation
3. The commercial fm band occupies the middle
4. The commercial AM band occupies the middle

1-15. Hf communications over long-distance trunks, or links between fixed terminals are examples of which of the following types of systems?

1. Fleet broadcast
2. Point-to-point
3. Ground-to-air
4. Ship-to-shore

1-16. Sending a message on several frequencies at once is an example of which of the following transmission types?

1. Time-diversity
2. Phase-diversity
3. Distance-diversity
4. Frequency-diversity

1-17. What type of diversity uses physically separated transmit or receive antennas to improve communications?

1. Time
2. Phase
3. Space
4. Frequency

1-18. Normally the transmission range of vhf is limited to line of sight. What technique is used to increase this range?

1. Tropospheric scatter
2. Atmospheric diversity
3. Ionospheric maneuvering
4. Each of the above

1-19. A complex of links make up a major communications system. The naval communications system is further broken down into what two groups?

1. Strategic and local
2. World-wide and local
3. Strategic and tactical
4. Tactical and world-wide

1-20. Communications links have many modes of operation. One terminal in a link has its equipment setup in simplex. The other terminal is using two channels or frequencies in a configuration that allows sending and receiving of different messages at the same time. These two terminals working together make up what mode of operation?

1. Full duplex
2. Half duplex
3. Quasiduplex
4. Semiduplex

1-21. What communications link mode of operation provides telecommunications capability between stations at the same time in both directions?

1. Half duplex
2. Semiduplex
3. Broadcast
4. Duplex

1-22. Aid in restoring downed fleet communications channels is furnished on a not-to-interfere basis by which of the following networks?

1. AUTOSEVOCOM
2. NORATS
3. HICOM
4. DSSCS

1-23. Which of the following switched networks extends tactical voice to shore-based operational commands?

1. NORATS
2. AUTOVON
3. AUTODIN
4. AUTOSEVOCOM

1-24. Of the following transmitter types, which are used for basic communications?

1. Cw, AM, fm, and ssb
2. Fsk, cw, AM, and tty
3. Cw, ssb, voice and fm
4. Voice, tty, fsk, and AM

1-25. Cw transmissions have narrow bandwidths and a high degree of intelligibility under severe noise conditions. What is the primary Navy use for cw?

1. Radioteletypewriter
2. Radiotelegraphy
3. Facsimile
4. Voice

1-26. A cw transmitter must contain which of the following components?

1. A traveling-wave tube
2. A demodulator
3. A combiner
4. A keyer

1-27. In a cw transmitter, a buffer stage performs which of the following functions?

1. Current divider
2. Voltage divider
3. Current amplifier
4. Voltage amplifier

1-28. There are differences between low- and high-power transmitters. The main difference is the high power transmitter contains a larger number of which of the following types of amplifiers?

1. Oscillator
2. Final power
3. Intermediate power
4. Intermediate frequency

1-29. In an AM transmitter, audio frequencies are converted into corresponding electrical energy by which of the following components?

1. An oscillator
2. A microphone
3. A modulator
4. A headset

1-30. In an fm transmitter, a varicap performs which of the following functions?

1. It amplifies the outgoing signal
2. It varies the oscillator frequency
3. It demodulates the outgoing signal
4. It multiplies the oscillator frequency

1-31. If an oscillator has a fundamental frequency of 3,550 megahertz, what is the frequency of the third harmonic?

1. 5,325 megahertz
2. 7,100 megahertz
3. 9,875 megahertz
4. 10,650 megahertz

1-32. If the fundamental frequency of an rf carrier is 1,000 kilohertz, what is the frequency of the fourth subharmonic?

1. 500 kilohertz
2. 333 kilohertz
3. 250 kilohertz
4. 200 kilohertz

1-33. Oscillator output frequencies are raised to usable values by frequency multipliers. To raise an oscillator frequency from 20 megahertz to 120 megahertz, what combination of frequency multipliers would be used?

1. Two doublers
2. A doubler and a tripler
3. A doubler and a quadruplet
4. A tripler and a quadruplet

1-34. When an AM signal leaves the antenna of a transmitter, which of the following frequency components does the signal contain?

1. The carrier
2. The upper sideband
3. The lower sideband
4. All of the above

1-35. In a single-sideband transmitter, selection of the desired sideband and suppression of the other is done by which of the following components?

1. Mixer
2. Filter
3. Detector
4. Oscillator

1-36. When compared to a conventional AM signal, an ssb signal provides which of the following advantages?

1. Improved frequency stability
2. Increased receiver gain
3. Reduced distortion
4. Reduced bandwidth

1-37. For ship-to-shore teletypewriter circuits, which of the following types of multiplexing is/are used?

1. Time and/or phase
2. Time and/or frequency
3. Phase and/or modulation
4. Frequency and/or modulation

1-38. Operators of transmitters and receivers use a circuit to coordinate the service of messages and to make frequency changes. What is the name of this circuit?

1. Order-wire circuit
2. Documentation circuit
3. Synchronization circuit
4. Operator-eyes-only circuit

1-39. A transmitted electromagnetic wave enters an antenna, induces a voltage into it, and passes that voltage to a receiver. What is this chain of events called?

1. Reproduction
2. Selection
3. Detection
4. Reception

1-40. When a receiver picks one frequency out from all other frequencies, it's performing which of the following basic functions?

1. Selection
2. Reception
3. Detection
4. Reproduction

1-41. When a receiver separates the audio frequencies from the radio-frequency carrier it is performing which of the following basic functions?

1. Reception
2. Selection
3. Detection
4. Reproduction

1-42. The receiver action of converting electrical energy to a usable format, such as sound, is an example of which of the following basic functions?

1. Reception
2. Selection
3. Detection
4. Reproduction

1-43. Which of the following measurements provides an indication of the ability of a receiver to reproduce weak signals?

1. Bandwidth
2. Sensitivity
3. Selectivity
4. Frequency response

1-44. Overall sensitivity of a receiver is limited by which of the following factors?

1. Noise
2. Bandwidth
3. Output power
4. Frequency response

1-45. How is a receiver's ability to reject unwanted signals and receive desired signals determined?

1. Noise
2. Fidelity
3. Selectivity
4. Sensitivity

1-46. When high fidelity is your prime consideration you should select a receiver that has been designed with which of the following features?

1. High gain
2. High output power
3. Broadband frequency selection circuits
4. Narrowband frequency selection circuits

1-47. The IF frequency in a receiver is produced by which of the following methods?

1. Modulation
2. Heterodyning
3. Frequency synthesis
4. Frequency multiplication

1-48. The process of heterodyning takes place in which of the following receiver circuits?

1. Mixer
2. Comparator
3. Oscillator
4. Second IF amplifier

1-49. Of the following frequencies, which one is a typical value of IF for AM communications receivers?

1. 455 kilohertz
2. 554 kilohertz
3. 455 megahertz
4. 554 megahertz

1-50. Two or more circuits within a receiver are varied by a single control through the use of which of the following processes?

1. Ganged tuning
2. Frequency synthesis
3. Automatic gain control
4. Automatic frequency control

1-51. There are electrical differences between AM and fm receivers. An fm receiver contains which of the following circuits?

1. Comparator
2. Discriminator
3. Limiter
4. Both 2 and 3 above

1-52. An fm signal has which of the following characteristics when compared to an AM signal?

1. More noise
2. Less static
3. A higher power output
4. A lower operating frequency

1-53. Ssb transmissions have which of the following characteristics when compared to AM transmissions?

1. Wide bandpass frequencies
2. Concentrated power
3. Less modulation
4. High fidelity

1-54. Single sideband receivers use a special oscillator. The output of that oscillator is fed directly to the detector circuit. What type of oscillator is used?

1. Local
2. High frequency
3. Variable frequency
4. Carrier reinsertion

1-55. A transmitter has a suppressed carrier frequency of 4 megahertz and is radiating only an upper sideband signal. When the intelligence is a 1-kilohertz tone, which of the following sideband frequencies will be transmitted?

1. 3,999 kilohertz
2. 4,000 kilohertz
3. 4,001 kilohertz
4. Both 2 and 3 above

1-56. Manual gain lets you adjust a receiver for maximum sensitivity and amplify weak input signals. Which of the following internal sections of the receiver are varied by this control?

1. Oscillator
2. Audio frequency
3. Radio frequency
4. Intermediate frequency

1-57. Manual volume control of a receiver internally varies the input to which of the following circuits?

1. Detector
2. Audio amplifier
3. Frequency converter
4. Radio frequency amplifier

1-58. Changes in receiver input strength due to changing atmospheric conditions is described by which of the following terms?

1. Gain
2. Fading
3. Ducting
4. Trapping

1-59. The rf amplifier connected to your receiving antenna has a voltage gain of 240. When the antenna is receiving a signal of 8 microvolts, what will be the maximum output voltage of the rf amplifier?

1. 1.92 millivolts
2. 2.91 microvolts
3. 30 microvolts
4. 33 millivolts

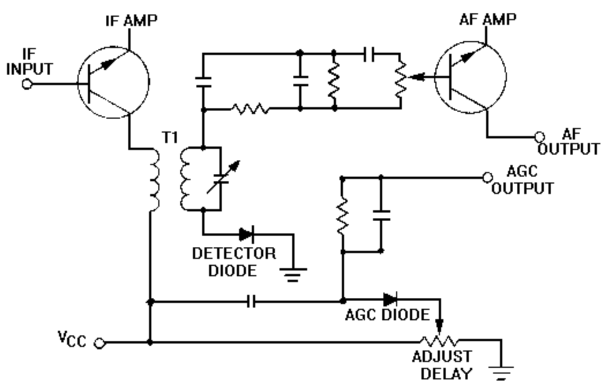


Figure 1A.

IN ANSWERING QUESTIONS 1-60 AND 1-61, REFER TO FIGURE 1A.

1-60. If the agc diode shorts, which of the following actions would result?

1. There would be no aqc
2. There would be no delay
3. The agc diode would be reverse biased
4. The polarity of the agc would be reversed

1-61. The amount of agc feedback depends on attaining an established received signal strength. If the established signal strength is set for 50 microvolts, and the input signal measures 44 microvolts, approximately which, if any, of the following values of agc is developed?

1. 44 microvolts
2. 50 microvolts
3. 94 microvolts
4. None of the above

1-62. To automatically compensate for input signal strength variations within a receiver, which of the following types of circuits are added?

1. Afc
2. Nfc
3. Agc
4. Nsu

1-63. AgC circuitry within a receiver uses a portion of which of the following detector voltage components as a feedback signal to preceding stages?

1. Dc
2. Ac
3. IF
4. Audio

1-64. Which of the following types of agc voltage drives an amplifier toward cutoff?

1. Saturation
2. Delayed
3. Reverse
4. Forward

1-65. What is the purpose of the squelch circuit in a receiver?

1. To attenuate very strong signals in order to prevent their overdriving the remaining stages in the receiver
2. To suppress receiver noise output when no input signal is being received
3. To suppress the electronic "whine" of the rf amplification stage
4. To reject signals of other than the desired frequency should the receiver drift off frequency

1-66. How does a receiver accomplish the squelch function?

1. By proportional blocking of the rf amplifier stage output
2. By blocking the detector or audio amplifier when there is no signal
3. By switching an inductive/capacitive filter into the output of the last rf amplifier
4. By switching in a matched-frequency stage which passes matching frequencies and rejects all others

1-67. A quartz crystal filter is used in a communications receiver to improve which of the following characteristics?

1. Fidelity
2. Sensitivity
3. Selectivity
4. Reproduction

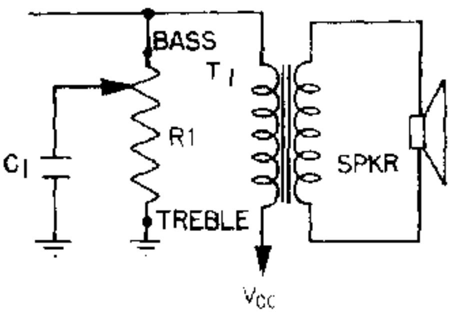


Figure 1B.

IN ANSWERING QUESTION 1-68, REFER TO FIGURE 1B.

1-68. When the wiper of R1 is placed in the full treble position, which of the following actions occur?

1. Bass response is improved
2. High frequency shunting is reduced
3. Higher frequencies are shunted to ground
4. The capacitor and resistor are placed in parallel

1-69. Automatic frequency control circuits are used in a receiver for which of the following purposes?

1. To adjust IF amplifier gain
2. To correct for oscillator frequency drift
3. To extend the frequency range of the receiver
4. To automatically tune the receiver to the desired frequency

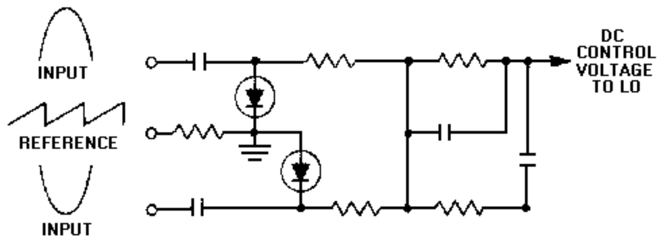


Figure 1C.

IN ANSWERING QUESTIONS 1-70 AND 1-71, REFER TO FIGURE 1C.

1-70. What is the phase relationship between the input signals fed to the diodes?

1. In phase
2. 90 degrees out of phase
3. 180 degrees out of phase
4. 270 degrees out of phase

1-71. A change in oscillator frequency will change which of the following input relationships between the sawtooth reference voltage and the incoming signal?

1. Phase
2. Voltage
3. Current
4. Amplitude

1-72. How is the long term stability and accuracy required of modern communications receivers attained?

1. Through the use of a single, crystal-controlled oscillator, as the local oscillator
2. Through the use of an electron-coupled oscillator, as the local oscillator
3. Through a process of automatic frequency control
4. Through a process known as frequency synthesis

1-73. When using the frequency synthesis process, a signal of the desired accuracy and stability is produced by which, if any, of the following methods?

1. Automatic frequency control, that is, by sensing the difference between the oscillator frequency and the desired frequency and automatically compensating for this difference
2. Using a crystal-controlled oscillator to produce a stable high frequency, and through the process of frequency division selecting a subharmonic of this frequency as the desired frequency
3. The heterodyning and selection of frequencies which are not harmonically related to each other
4. None of the above

1-74. Permanent magnet speakers respond quite well to which of the following audio frequency ranges?

1. Low
2. High
3. Mid band
4. Each of the above

1-75. For which of the following reasons do most standard Navy headphones respond poorly to low frequencies?

1. Small diaphragm size
2. Diaphragm inflexibility
3. Both 1 and 2 above
4. Input signal filtering

ASSIGNMENT 2

Textbook assignment: Chapter 3, "Fundamental Systems Equipment," pages 3-1 through 3-47.

2-1. In a basic communications voice system, which of the following functions occur in the handset during the receive process?

1. Correction voltages are fed to the local oscillator
2. Phasing voltages are received from the master oscillator
3. Voice energy is transformed into electronic impulses
4. Electrical energy is converted to acoustical energy

2-2. Radio set control units are often used aboard ship to remotely control transmitters and receivers. Under standard operating conditions what is the maximum number of units that can be paralleled with a single transmitter and receiver group?

1. One
2. Two
3. Eight
4. Four

2-3. When you are using a transmitter transfer switchboard, what is the maximum number of transmitters that may be connected to a single remote control station?

1. One
2. Two
3. Eight
4. Four

THIS SPACE LEFT BLANK INTENTIONALLY.

2-4. You are using a radio set control unit to remotely control 1 of 8 transmitters. Operating knob number 1 on transmitter transfer switchboard number 1 is used to select transmitters 1 through 6. If you want to control transmitter number 7, what switch position must be selected?

1. 1
2. Any of 1-6
3. X
4. OFF

2-5. On a receiver transfer switchboard, each switch position is connected to what maximum number of receivers?

1. One
2. Two
3. Three
4. Four

IN ANSWERING QUESTIONS 2-6 THROUGH 2-13, REFER TO THE RADIO TRANSMITTING SET DISCUSSED IN CHAPTER 3.

2-6. The impedance matching of equipment to transmission line is accomplished by which of the following equipment groups?

1. A radio transmitter
2. An antenna coupler
3. An rf amplifier
4. A power supply

2-7. The output power of the transmitter unit drives which of the following units?

1. The rf amplifier
2. The power supply
3. The antenna coupler
4. The antenna coupler control

2-8. The rf amplifier receives digital tuning information that is generated in which of the following units?

1. The receiver
2. The transmitter
3. The coupler control
4. The radio set control

2-9. In some installations tuning must be done without the use of rf power, as rf is suppressed except during brief transmission periods. Under these conditions, which of the following methods of tuning should be used?

1. Silent
2. Anti-jam
3. Automatic
4. Semiautomatic

2-10. Operator controlled tuning of the antenna coupler group by the use of front panel controls is possible during which of the following modes of operation?

1. Silent and manual
2. Manual and automatic
3. Semiautomatic and silent
4. Automatic and semiautomatic

2-11. Once tuned, the antenna coupler is able to handle which of the following maximum amounts of power?

1. 500 watts average
2. 2,000 watts average
3. 250 watts peak envelope
4. 1,000 watts peak envelope

2-12. The antenna coupler is pressurized with dry nitrogen for which of the following reasons?

1. To prevent corona
2. To prevent arcing
3. To aid in internal heat transfer
4. Each of the above

2-13. You are using the cw mode of the radio transmitter unit. The 500-kilohertz local carrier is directly inserted into which of the following circuits?

1. The rf amplifiers
2. The IF amplifiers
3. The modulator
4. The detector

IN ANSWERING QUESTIONS 2-14 THROUGH 2-16, REFER TO THE RADIO RECEIVER DISCUSSED IN CHAPTER 3.

2-14. Adjacent-channel selectivity and image-frequency suppression have been improved in this receiver by the addition of which of the following features?

1. Digital tuning
2. Triple conversion
3. Front panel readout
4. Very accurate frequency standard

2-15. The full accuracy of the frequency standard is sacrificed when which of the following types of tuning is used?

1. Vernier
2. Automatic
3. 1-kilohertz incremental
4. 100- or 500-hertz incremental

2-16. The receiver demodulates and provides audio outputs for which of the following types of received signals?

1. Fm, AM, and fsk
2. AM, cw, and isb
3. Cw, fm, and isb
4. Lsb, usb, and fm

2-17. A receiving antenna patch panel serves which of the following functions?

1. Terminates lines leading to receivers
2. Terminates incoming antenna transmission lines
3. Both 1 and 2 above
4. Physically connects transmitters to receivers

2-18. Transmitting antenna patch panels are interlocked with the transmitter for which of the following reasons?

1. Safety
2. Ease of operation
3. Ease of maintenance
4. Both 2 and 3 above

2-19. A transmit multicoupler provides which of the following functions?

1. Receiver isolation
2. Transmitter tuning
3. Additional amplification
4. Isolation between transmitters

2-20. In a manual telegraph circuit, the only two conditions are marking and spacing. Marking is characterized by which of the following descriptions?

1. The key is open
2. Current is flowing
3. Current is not flowing
4. The armature is retracted by a spring

2-21. A teletypewriter code signal consists of 7 units. Of the following functions, which describes the middle 5 units?

1. Provides channel data
2. Carries the intelligence
3. Signals stop information
4. Signals start information

2-22. Which of the following terms describe the time between a space and mark or mark and space condition in a teletypewriter?

1. Movement
2. Variation
3. Transition
4. Character interval

2-23. The time interval between words when using the Morse code is equal to which of the following durations?

1. 1 dot
2. 7 dots
3. 3 dashes
4. 5 dashes

2-24. When you are using the five-unit code in teletypewriter operation, what is the maximum number of combinations available that will print letters, figures, function signs, and numerals?

1. 30
2. 32
3. 60
4. 74

2-25. When you are using the teletypewriter five-unit code, which of the following signals are used to increase the printing capacity of the equipment?

1. Inverter
2. Combiner
3. Case-shift
4. Type-adjust

2-26. Which, if any, of the following modes of teletypewriter operation is more often used in high-speed data systems?

1. Start-stop
2. Synchronous
3. Asynchronous
4. None of the above

2-27. In teletypewriter operation, what term defines the length of time required to transmit one letter, figure, function sign or numeral?

1. Baud rate
2. Bit speed
3. Code length
4. Character interval

2-28. Synchronous teletypewriter systems are characterized by which of the following features when compared to asynchronous systems?

1. Internal timing signals are always used
2. Only the start-stop element must be transmitted
3. Only the intelligence elements must be transmitted
4. Signal quality determines receiver line signal condition

2-29. When you are referring to the unit of teletypewriter signaling speed, the reciprocal of the time (in seconds) of the shortest signal element is described by which of the following terms?

1. Unit code
2. Baud rate
3. Bits per second
4. Words per minute

2-30. The teletypewriter condition where current flow represents a mark and no current flow represents a space occurs in which of the following types of operation?

1. Polar
2. Arctic
3. Biased
4. Neutral

2-31. You are using neutral keying and the teletypewriter type hammer continually strikes the type box but there is no printing or type box movement across the page. What is the name of this condition?

1. Debugging
2. Running open
3. Baudot blanking
4. Decoding at random

2-32. Of the following equipment which one changes teletypewriter dc pulses to mark and space modulation for the transmitter carrier wave?

1. A comparator
2. A modulator
3. A converter
4. A keyer

2-33. To change an rf signal to do pulses for teletypewriter operation, you must use a receiver and what other piece of equipment?

1. A keyer
2. A converter
3. A comparator
4. A demodulator

2-34. A tone-modulated radio teletypewriter system uses what modulation method to change dc mark and space impulses into audio electrical impulses?

1. Amplitude
2. Frequency
3. Phase
4. Pulse

2-35. In a basic tone-modulated radio teletypewriter system, separation of the audio signal from the carrier is accomplished by what process?

1. Conversion
2. Modulation
3. Selection
4. Detection

2-36. In a radio-frequency-carrier shift system, what equipment is the source of radio-frequency excitation voltages?

1. The inverter
2. The converter
3. The comparator
4. The transmitter keyer

2-37. The keyer in a radio-frequency-carrier shift system is normally adjusted for which of the following maximum frequency spreads?

1. 425 hertz
2. 500 hertz
3. 750 hertz
4. 850 hertz

2-38. Of the following teletypewriter equipment, which one is used to store incoming teletypewriter messages on tapes for future transmission on a transmitter distributor?

1. A keyboard
2. A page printer
3. A typing reperforator
4. A communication patching panel

2-39. Teletypewriter patch panels perform which of the following functions?

1. They provide a means for connecting the teletypewriter equipment in various combinations
2. They provide a means for permanently connecting commonly used combinations of equipment
3. They provide a central point for connecting the dc supply voltage to the teletypewriter circuits
4. Each of the above

2-40. You are working with a teletypewriter patch panel. What color signifies that secure information is being passed?

1. Red
2. Gray
3. Black
4. Green

2-41. In any switching operation between plugs and jacks of a teletypewriter panel, if the cord plug is pulled from the set (machine) jack before the plug is removed from the looping jack, which of the following conditions will occur?

1. A dangerous dc voltage will be produced on the exposed plug
2. All teletypewriter messages in the channel will be interrupted
3. Both 1 and 2 above
4. Classified information will be Compromised

2-42. Cryptographic equipment performs which, if any, of the following functions?

1. Encodes and decodes messages
2. Reduces mean-time between messages
3. Acts as an additional power amplifier
4. None of the above

2-43. In the radio-frequency-carrier shift system, translation of an rf signal to an audio signal is done by which of the following equipment?

1. A converter
2. A comparator
3. A radio receiver
4. An antenna filter

2-44. A comparator compares signal strength during which of the following types of receiver operation?

1. Single
2. Space diversity
3. Frequency diversity
4. Both 2 and 3 above

2-45. In an afts transmit system, the conversion of dc signals into audio tone-shift signals is done by which of the following pieces of equipment?

1. Cryptographic
2. Tone terminal set
3. Converter/comparator
4. Modulator/demodulator

2-46. The process of simultaneous transmission of several intelligible signals on the same frequency during the same period of time is called

1. duplexing
2. simplexing
3. complexing
4. multiplexing

2-47. What are the two methods of multiplexing?

1. Time-division and frequency-multiplication
2. Time-division and frequency-division
3. Time-multiplication and frequency-multiplication
4. Time-multiplication and frequency-division

2-48. In time-division multiplexing, assume that a 4,000-hertz tone is applied to each of six channels in a telegraph transmitter and that each channel is to be sampled at a rate of 2. 5 times during each cycle of the 4,000-hertz tone. At what speed, in revolutions per second, must the rotating switch turn to accomplish this sampling rate?

1. 3,000
2. 4,000
3. 7,200
4. 10,000

2-49. In time-division multiplexing, what drawback is encountered if an excessive number of frequency channels is used?

1. Static is increased
2. Bandwidth is increased
3. Switching becomes unreliable
4. Reception becomes unintelligible

2-50. How many times per cycle is a practical time-division multiplex system optimumly sampled?

1. 1.5
2. 2.0
3. 2.4
4. 3.1

2-51. Frequency-division multiplexing systems transmit and receive during a maximum of how many degrees of a sinewave?

1. 90
2. 180
3. 270
4. 360

2-52. By using frequency-division multiplexing, tty circuits may carry a maximum of how many single, 3,000-hertz channels?

1. 12
2. 16
3. 18
4. 24

2-53. In a 16-channel tty-multiplexing system, the maximum difference between a mark and a space, for any give channel, is how many hertz?

1. 85
2. 382.5
3. 425
4. 467.5

2-54. Weather charts and photographs are examples of materials transmitted by

1. aw telegraphy
2. FAX (facsimile)
3. landline teletypewriter
4. rttv (radio teletypewriter)

2-55. Which of the following facsimile transceiver operations consists of subdividing the picture in an orderly manner and into a large number of segments?

1. Scanning
2. Recording
3. Receiving
4. Transmitting

2-56. The scanning operation is accomplished in the facsimile transmitter by a

1. scanning drum and a phototube arrangement
2. scanning drum and aperture tube
3. spiral drum and amplifier
4. phototube amplifier

2-57. The purpose of the phototube in facsimile equipment is to

1. illuminate a segment of the picture
2. produce the carrier signal for the exciter lamp
3. maintain the output voltage at a predetermined fixed value
4. transform varying amounts of light into electrical signals

2-58. Which of the following means is used to synchronize the receiving drum with the transmitting drum in a radio facsimile system?

1. The drums are mechanically linked
2. Each drum is started by an accurate clock
3. Both drums are operated by synchronous motors
4. Both drums are stepped around a precise number of steps by a relay

2-59. In a facsimile system, which of the following methods is used to accomplish framing at the receiver unit?

1. A synchronous clutch mechanism is actuated
2. A primary time/frequency standard is switched in
3. A series of phasing pulses are transmitted prior to image transmission
4. A synchronous motor at both the transmitter and receiver is engaged

2-60. Of the following terms, which one is primarily concerned with compromising emanations?

1. Tempest
2. Radiation hazard
3. Quality monitoring
4. Electromagnetic interference

2-61. Of the following fundamental requirements of a military communications system, which one, if any, is most important?

1. Speed
2. Security
3. Reliability
4. None of the above

2-62. What one assumption may be made regarding all military radio transmissions?

1. They are secure
2. They have been encrypted
3. They have been decrypted
4. They have been intercepted

2-63. Scheduled maintenance in support of QMCS is designed to alert you to which of the following problems?

1. Safety hazards
2. Equipment failure
3. System degradation
4. Improper operating procedures

2-64. Of the following terms, which one is defined as the ability of an electronic system to perform its individual functions without interference?

1. Electronic countermeasures
2. Electromagnetic interference
3. Electromagnetic compatibility
4. Electronic counter-countermeasures

2-65. The sources of electromagnetic radiations that reduce receiver performance are known by which of the following terms?

1. Electronic countermeasures
2. Electromagnetic interference
3. Electromagnetic compatibility
4. Electronic counter-countermeasures

2-66. Which of the following categories of electromagnetic interference includes interference generated by electrically charged raindrops?

1. Natural
2. Functional
3. Incidental
4. Hull-generated

2-67. Cross modulation is a form of emi where the desired carrier intermodulates with an undesired signal. Which of the following devices should minimize this interference?

1. Filters
2. Preselectors
3. Both 1 and 2 above
4. Preamplifiers

2-68. Shipboard receive systems are designed to include protective circuitry between the antenna and the receiver that prevent which of the following problems?

1. Degradation of overall receiver performance by processing of off-frequency signals
2. Decrease of desired signal amplification
3. Burn out of front-end stages
4. Each of the above

2-69. Of the following body organs, which are considered the most vulnerable to radiation hazards (RADHAZ)?

1. Eyes and testes
2. Heart and lungs
3. Liver and spleen
4. Kidneys and brain

2-70. Which of the following methods of reducing rf burn hazards is the most useful and widespread technique used?

1. Operate receivers only
2. Vary the operating frequency
3. Bond and ground all metallic objects
4. Operate transmitters only at low power

2-71. The greatest hazard from thermal effects appears to come from equipment operated in which of the following frequency ranges?

1. 1 to 3 gigahertz
2. 2 to 30 megahertz
3. 30 to 300 kilohertz
4. 225 to 500 megahertz

2-72. You have been working on a piece of equipment and your eyes have been exposed to high-intensity microwaves. Which of the following types of eye problems may occur?

1. Detached retina
2. Conjunctivitis
3. Cataracts
4. Glaucoma

ASSIGNMENT 3

Textbook assignment: Chapter 4, "Introduction to Satellite Communications," pages 4-1 through 4-21.
Chapter 5, "Introduction to Miscellaneous Systems and Equipment," pages 5-1 through 5-20.

3-1. What artificial satellite is credited with starting the era of space technology?

1. Vela
2. Midas
3. Score
4. Sputnik

3-2. There are two types of communications satellites. What type acts as a repeater for the signal?

1. Active
2. Passive
3. Reflecting
4. Retransmitting

3-3. Transmission of information to a satellite is done on what frequency?

1. Up-link
2. Down-link
3. Transponder
4. Termination

3-4. An earth terminal receives signals on what frequency?

1. Up-link
2. Transmit
3. Down-link
4. Termination

3-5. Of the following orbit parameters, which one describes the basic orbit shape of a communications satellite?

1. Inclined or polar
2. Polar or equatorial
3. Elliptical or circular
4. Synchronous or nonsynchronous

3-6. When the period of an orbit is identical to that of the earth, the orbit is

1. synchronous
2. asynchronous
3. subsynchronous
4. near-synchronous

3-7. What is the parameter in the orbit of a satellite that refers to the point nearest the center of the earth?

1. Apogee
2. Perigee
3. Altitude
4. Inclination

3-8. As a reference, perigee and apogee are measured in which of the following units?

1. Nautical miles
2. Statute miles
3. Light years
4. Kilometers

3-9. A satellite which has a flight path that does not coincide with the equatorial plane of the earth is said to be in what type if orbit?

1. A circular
2. An inclined
3. An elliptical
4. An equatorial

3-10. A satellite orbiting with an angle of inclination of approximately 90 degrees describes which of the following types of orbit?

1. Polar
2. Equatorial
3. Synchronous
4. Asynchronous

3-11. In order to cover most of the earth except the polar regions, what is the minimum number of satellites that mast be orbited?

1. 5
2. 6
3. 3
4. 4

3-12. There were many limitations that caused problems on the first communications satellites. Of the following problems, which one was considered the most severe?

1. The excessive size
2. The excessive weight
3. The too low orbit altitude
4. The lack of a suitable power source

3-13. Which of the following power sources is/are considered a practical choice for satellites?

1. Solar cells only
2. Storage batteries only
3. A combination of solar cells and storage batteries
4. Sunlight and leclanche cells

3-14. What development in satellite communications improved back-up power during eclipses?

1. The installation of a battery back-up
2. The installation of a nuclear power source
3. The continuous exposure of solar cells to the sun
4. The increase in solar cells mounted on the surface of the satellite

3-15. Why is satellite orientation in space so important?

1. Because it is a necessity for back-up power
2. To meet the requirements of spin stabilization
3. To ensure that sunlight converging on the solar cells is converted to electrical power
4. Because it is essential for maximum solar cell exposure to the sun and satellite antenna visibility to earth terminals

3-16. Why are communications satellite earth terminals generally located in areas remote from the actual users?

1. To minimize cost
2. To minimize jamming
3. To minimize rf interference
4. To allow for future expansion

3-17. Which of the following characteristics is a requirement for a satellite earth terminal antenna?

1. It must be omnidirectional
2. It must be of the Franklin collinear type
3. It must have low gain and be highly directional
4. It must be capable of transmitting and receiving signals simultaneously

3-18. One earth terminal antenna uses a cluster of four 10-foot parabolic antennas. This array is effectively a total of how many feet in diameter?

1. 18
2. 20
3. 32
4. 40

3-19. Why do satellite earth terminals require highly sensitive receivers?

1. To overcome the down-link power losses
2. To permit extraction of the desired communications information from the received signal
3. Both 1 and 2 above
4. Because of the signal scatter effect of the antennas

3-20. Which of the following functions is performed by the exciter stage of an earth terminal transmitter?

1. Modulation of the IF carrier
2. Translation of the IF signal to the up-link frequency
3. Amplification of the IF signal to the level required by the receiver
4. Conversion of the down-link frequency to an IF

3-21. Telemetery equipment used in satellite communications systems performs which of the following functions?

1. They monitor the operating conditions within the satellite
2. They provide local control for satellite operations
3. They furnish high-capacity wide-band tty trunks
4. They measure ambient weather conditions

3-22. A typical shipboard receive-only satellite system uses which of the following types of modulation?

1. Pulsed or amplitude
2. Pulsed or frequency
3. Amplitude or phase-shift-key
4. Frequency or phase-shift-key

3-23. Locating a near-synchronous satellite is rather simple for which of the following reasons?

1. It is stationary
2. It is moving north to south
3. It has a slow relative motion
4. It has a fast relative motion

3-24. What is the name of the table that provides coordinates of a satellite at specific times?

1. Bearing location
2. Longitudinal
3. Propagation
4. Ephemeris

3-25. To establish radio contact with a satellite, an earth terminal must know which of the following satellite data?

1. Attitude
2. Elevation
3. Operating speed
4. Angle of inclination

3-26. Satellite down-link frequency variations occur most often from satellites in which of the following orbits?

1. Low altitude elliptical
2. Medium altitude circular
3. High altitude synchronous
4. Superhigh altitude near-synchronous

3-27. Of the following terms, which one describes the period of time required for one earth terminal to yield control of a satellite to another?

1. Slewing
2. Hand over
3. Control shift
4. Terminal continuity

3-28. When compared to hf communications, which of the following advantages are unique to satellite communications links?

1. They are more reliable and flexible
2. They are unaffected by propagation variations affecting hf
3. They do not require repeater stations or troposcatter links
4. Each of the above

3-29. Which of the following factors limits the reliability of active satellite communications system?

1. The reflection or refraction of signals
2. The reliability of the equipment used
3. The skill of the operating and maintenance personnel
4. Both 2 and 3 above

3-30. An increase of invulnerability to jamming of satellite communications systems is seen through the use of which of the following features?

1. Narrow bandwidths
2. Low transmitter output power
3. Antijamming modulation techniques
4. Omnidirectional earth terminal Antennas

3-31. Which of the following statements best describes the advantage of satellite communications in terms of flexibility?

1. The antenna group of any earth terminal can be mounted on the weather deck of a ship
2. Certain earth terminals are housed in vans and can be transported to remote areas
3. Military satellite communications are capable of handling hundreds of voice channels
4. A high degree of protection from jamming is afforded by the highly Directional antennas at earth terminals

3-32. A satellite communications system is limited by which of the following characteristics?

1. The attitude of the satellite repeater
2. The technical design of the satellite
3. The immobility of the satellite
4. The mobility of the satellite

3-33. Active communications satellite systems have two major limitations. What are they?

1. Complex preamplifiers and high gain antennas
2. Up-link transmitter power and earth terminal antenna size
3. Down-link transmitter power and uplink receiver sensitivity
4. Down-link receiver sensitivity and external atmospheric noise

3-34. The rf power output of a satellite communications system is severely limited due to which of the following factors?

1. A lack of adequate jamming capabilities
2. An inefficient solar-cell package aboard the satellite
3. An unstable satellite orientation with respect to the horizon
4. A requirement for large antenna-farm earth-terminal systems

3-35. The availability of a satellite to act as a relay station between two earth terminals depends upon which of the following considerations?

1. The mobility of the satellite
2. The location of the earth terminals
3. The operating frequencies of the satellite
4. The electronic design of the earth terminals

3-36. What determines the length of time that a nonsynchronous satellite in a circular orbit will be in the zone of mutual visibility?

1. Height of the orbit
2. Earth terminal antenna size
3. Down-link transmitter power
4. Up-link receiver sensitivity

3-37. Satellite communications systems are being rapidly developed by the Navy for which of the following reasons?

1. To replace microwave links
2. To relieve dependence on hf communications
3. To reduce procurement and development costs
4. To replace all physically large size equipment

3-38. In an mf transmitter, the frequency generator is used during which of the following modes of operation?

1. AM
2. Fm
3. Cw
4. Fsk

3-39. The pre-ipa and ipa in an mf transmitter are which of the following types?

1. Linear, tuned
2. Linear, untuned
3. Non-linear, tuned
4. Non-linear, untuned

3-40. Of the following communications system components, which one is a device that is nonradiating, absorbs rf, and has the characteristic impedance of the antenna?

1. Helix
2. Dummy load
3. Rf tuning unit
4. Frequency synthesizer

3-41. Which of the following antenna arrays consists of quarter-wave, vertically polarized stubs?

1. Broadside
2. Parasitic
3. Top-hat
4. Triatic

IN ANSWERING QUESTIONS 3-42 THROUGH 3-44, REFER TO THE MEDIUM FREQUENCY AND BELOW RECEIVER COVERED IN CHAPTER 5.

3-42. With one exception, the receiver has the same circuitry as any high frequency receiver. What is the one exception?

1. The components are doubled up
2. The local oscillator is eliminated
3. The radio-frequency amplifier is replaced by a video amplifier
4. The audio-frequency amplifier is replaced by a traveling-wave tube

3-43. The rejection of input frequencies above 900-kilohertz is performed by which of the following circuits?

1. An attenuator
2. A video amplifier
3. A low-pass filter
4. A calibration oscillator

3-44. The demodulation of ssb, cw, and fsk signals is performed by which of the following circuits?

1. An fm detector
2. A phase splitter
3. An audio amplifier
4. A product detector

3-45. Of the following amplifiers, which one has a high gain, low noise, wide bandwidth and is operated in the microwave region?

1. A magnetic
2. An operational
3. A differential
4. A traveling-wave-tube

3-46. In a line-of-sight communications system, propagation is affected by which of the following layers of the atmosphere?

1. Ionosphere
2. Troposphere
3. Stratosphere
4. Thermosphere

3-47. Horn-driven paraboloid antennas have which of the following characteristics?

1. High gain, narrow beam width
2. Low gain, narrow beam width
3. High gain, wide beam width
4. Low gain, wide beam width

3-48. Line-of-sight systems are configured in many ways with regards to channel width and number of channels. A voice system with a channel width of 4-kilohertz has a total of how many channels available for transmission?

1. 200
2. 400
3. 600
4. 800

3-49. A one-hop transmission of a tropo-scatter system can travel what maximum distance?

1. 1200 miles
2. 1000 miles
3. 800 miles
4. 500 miles

3-50. Of the following advantages, which one is primary to the NTDS when compared with conventional data systems?

1. Speed
2. Distance
3. Security
4. Reliability

3-51. The NTDS uses which of the following data transmission links?

1. 14, 11, 4A
2. 14A, 11, 4
3. 14, 11A, 4
4. 14A, 11A, 4A

3-52. Of the following NTDS links, which one(s) is/are only used as a one-way broadcast?

1. 14
2. 4A
3. 11, 14
4. 4A, 11

3-53. Portable radio sets are used primarily for which of the following types of communications?

1. Amphibious
2. Air-to-air
3. Electronic warfare
4. Anti-submarine warfare

3-54. When designing portable and pack radios, which of the following characteristics is the prime consideration?

1. Must be solar powered
2. Must be heavy and rugged
3. Must have high output power
4. Must be light-weight and compact

IN ANSWERING QUESTIONS 3-55 THROUGH 3-58, REFER TO THE EMERGENCY LIFEBOAT TRANSMITTER COVERED IN CHAPTER 5.

3-55. Search and rescue stations are divided into groups that have distinct rescue functions. Emergency lifeboat transmissions are designed for reception by a total of how many of these groups?

1. 5
2. 2
3. 3
4. 4

3-56. What are the operating frequencies of the transmitter?

1. 8,364 kHz and 500 MHz
2. 500 MHz and 8,864 kHz
3. 500 kHz and 8,364 kHz
4. 500 kHz and 8,864 kHz

3-57. What is the primary source of power for the transmitter?

1. External ac
2. Wind generator
3. Internal battery
4. Handcrank generator

3-58. When in the automatic mode of operation, the transmitter transmits (a) while changing frequency every (b) seconds?

1. (a) The SOS distress signal
   (b) 50
2. (a) Voice messages
   (b) 20
3. (a) Voice messages
   (b) 50
4. (a) The SOS distress signal
   (b) 20

3-59. What is the maximum number of channels available when you are using the emergency portable transceiver covered in chapter 5?

1. 5
2. 2
3. 3
4. 4

3-60. A laser operates in which of the following areas of the light spectrum?

1. Red
2. Infrared
3. Ultraviolet
4. At or near visible light

3-61. The principle of the laser is much like that of, which of the following electronic components?

1. Hall generator
2. Reflex klystron
3. Traveling-wave-tube
4. Very high-Q cavity resonator

3-62. Which of the following components is a close relative of the laser?

1. Thyristor
2. Photo transistor
3. Light emitting diode
4. Photovoltaic transducer

3-63. Laser transmissions during adverse weather conditions experience which of the following problems?

1. Absorption
2. Refraction
3. Reflection
4. Diffraction

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 18—Radar Principles

NAVEDTRA 14190

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 18 of a series.

History of the course:

Sep 1998: *Original edition released. Prepared by FTMC Frank E. Sloan and FTCM Gilbert J. Coté*

Jan 2004: *Administrative update released. Reviewed by ETC(SW) Jack Weatherford. No change in technical content.*

NAVSUP Logistics Tracking Number
0504-LP-026-8430

TABLE OF CONTENTS

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

CHAPTER 1

RADAR FUNDAMENTALS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the OCC/ECC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

1. Define range, bearing, and altitude as they relate to a radar system.
2. Discuss how pulse width, peak power, and beam width affect radar performance.
3. Describe the factors that contribute to or detract from radar accuracy.
4. Using a block diagram, describe the basic function, principles of operation, and interrelationships of the basic units of a radar system.
5. Explain the various ways in which radar systems are classified, including the standard Army/Navy classification system.
6. Explain the basic operation of cw, pulse, and Doppler radar systems.

INTRODUCTION TO RADAR FUNDAMENTALS

The term RADAR is common in today's everyday language. You probably use it yourself when referring to a method of recording the speed of a moving object. The term *Radar* is an acronym made up of the words radio detection and ranging. The term is used to refer to electronic equipment that detect the presence, direction, height, and distance of objects by using reflected electromagnetic energy. Electromagnetic energy of the frequency used for radar is unaffected by darkness and also penetrates weather to some degree, depending on frequency. It permits radar systems to determine the positions of ships, planes, and land masses that are invisible to the naked eye because of distance, darkness, or weather.

The development of radar into the highly complex systems in use today represents the accumulated developments of many people and nations. The general principles of radar have been known for a long time, but many electronics discoveries were necessary before a useful radar system could be developed. World War II provided a strong incentive to develop practical radar, and early versions were in use soon after the war began. Radar technology has improved in the years since the war. We now have radar systems that are smaller, more efficient, and better than those early versions.

Modern radar systems are used for early detection of surface or air objects and provide extremely accurate information on distance, direction, height, and speed of the objects. Radar is also used to guide missiles to targets and direct the firing of gun systems. Other types of radar provide long-distance surveillance and navigation information.

BASIC RADAR CONCEPTS

The electronics principle on which radar operates is very similar to the principle of sound-wave reflection. If you shout in the direction of a sound-reflecting object (like a rocky canyon or cave), you will hear an echo. If you know the speed of sound in air, you can then estimate the distance and general direction of the object. The time required for a return echo can be roughly converted to distance if the speed of sound is known. Radar uses electromagnetic energy pulses in much the same way, as shown in figure 1-1. The radio-frequency (rf) energy is transmitted to and reflects from the reflecting object. A small portion of the energy is reflected and returns to the radar set. This returned energy is called an ECHO, just as it is in sound terminology. Radar sets use the echo to determine the direction and distance of the reflecting object.

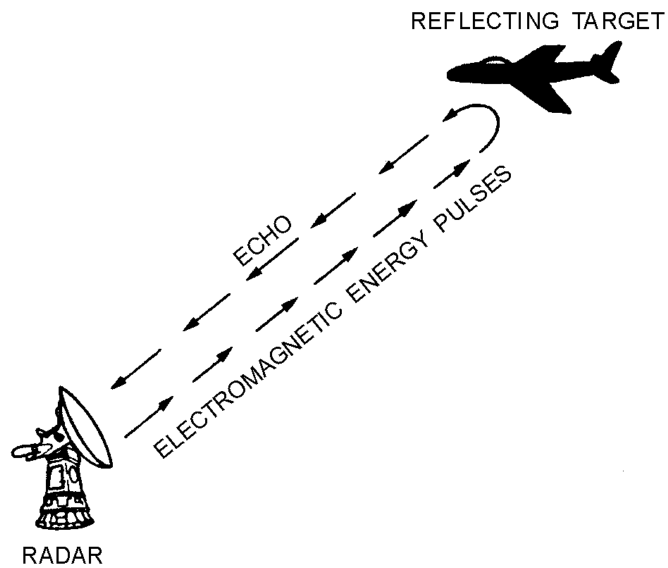


Figure 1-1.—Radar echo.

NOTE: The terms TARGET, RETURN, ECHO, CONTACT, OBJECT, and REFLECTING OBJECT are used interchangeably throughout this module to indicate a surface or airborne object that has been detected by a radar system.

Radar systems also have some characteristics in common with telescopes. Both provide only a limited field of view and require reference coordinate systems to define the positions of detected objects. If you describe the location of an object as you see it through a telescope, you will most likely refer to prominent features of the landscape. Radar requires a more precise reference system. Radar surface angular measurements are normally made in a clockwise direction from TRUE NORTH, as shown in figure 1-2, or from the heading line of a ship or aircraft. The surface of the earth is represented by an imaginary flat plane, tangent (or parallel) to the earth's surface at that location. This plane is referred to as the HORIZONTAL PLANE. All angles in the up direction are measured in a second imaginary plane that is perpendicular to the horizontal plane.

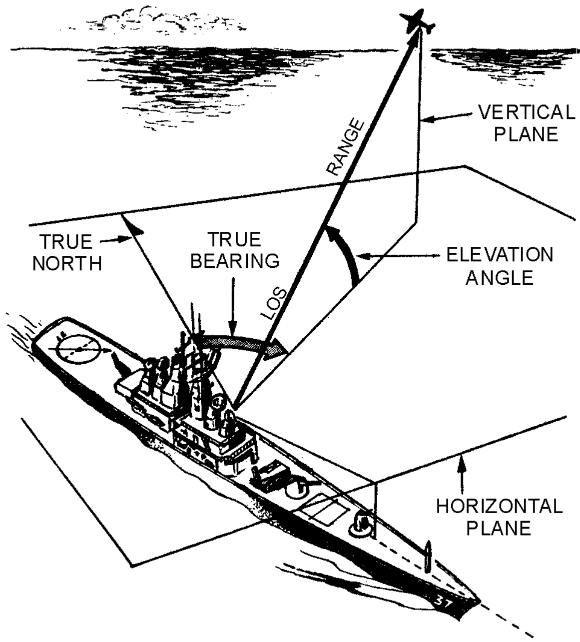


Figure 1-2.—Radar reference coordinates.

This second plane is called the VERTICAL PLANE. The radar location is the center of this coordinate system. The line from the radar set directly to the object is referred to as the LINE OF SIGHT (los). The length of this line is called RANGE. The angle between the horizontal plane and the los is the ELEVATION ANGLE. The angle measured clockwise from true north in the horizontal plane is called the TRUE BEARING or AZIMUTH angle. These three coordinates of range, bearing, and elevation describe the location of an object with respect to the antenna.

Q1. Radar surface-angular measurements are referenced to true north and measured in what plane?

Q2. The distance from a radar set to a target measured along the line of sight is identified by what term?

RANGE

Radar measurement of range, or distance, is made possible because of the properties of radiated electromagnetic energy. This energy normally travels through space in a straight line, at a constant speed, and will vary only slightly because of atmospheric and weather conditions. The effects atmosphere and weather have on this energy will be discussed later in this chapter; however, for this discussion on determining range, these effects will be temporarily ignored.

Electromagnetic energy travels through air at approximately the speed of light, which is 186,000 STATUTE MILES per second. The Navy uses NAUTICAL MILES to calculate distances; 186,000 statute miles is approximately 162,000 nautical miles. While the distance of the statute mile is approximately 5,280 feet, the distance for a nautical mile is approximately 6,080 feet.

Radar timing is usually expressed in microseconds. To relate radar timing to distances traveled by radar energy, you should know that radiated energy from a radar set travels at approximately 984 feet per microsecond. With the knowledge that a nautical mile is approximately 6,080 feet, we can figure the approximate time required for radar energy to travel one nautical mile using the following calculation:

$$\text{time for energy to travel one nautical mile}$$
$$= \frac{6{,}080 \text{ feet}}{984 \text{ feet per microsecond}}$$
$$= 6.18 \text{ microseconds (approx.)}$$

The same answer can be obtained using yards instead of feet. In the following calculation, the 6,080 foot approximation of a nautical mile is converted to 2,027 yards and energy speed is changed from 984 feet to 328 yards per microsecond:

$$\text{time for energy to travel one nautical mile} = \frac{2{,}027 \text{ yards}}{328 \text{ yards per microsecond}}$$
$$= 6.18 \text{ microseconds (approx.)}$$

A pulse-type radar set transmits a short burst of electromagnetic energy. Target range is determined by measuring elapsed time while the pulse travels to and returns from the target. Because two-way travel is involved, a total time of 12.36 (6.18 x 2) microseconds per nautical mile will elapse between the start of the pulse from the antenna and its return to the antenna from a target. This 12.36 microsecond time interval is sometimes referred to as a RADAR MILE, RADAR NAUTICAL MILE, or NAUTICAL RADAR MILE. The range in nautical miles to an object can be found by measuring the elapsed time during a *round trip* of a radar pulse and dividing this quantity by 12.36. In equation form, this is:

$$\text{range} = \frac{\text{elapsed time}}{12.36 \text{ microseconds per nautical mile}}$$

For example, if the elapsed time for an echo is 62 microseconds, then the distance is 5 miles, as shown in the following calculation:

$$\text{range} = \frac{\text{elapsed time}}{12.36 \text{ microseconds per nautical mile}}$$
$$= \frac{62 \text{ microseconds}}{12.36 \text{ microseconds per nautical mile}}$$
$$= 5 \text{ nautical miles (approx.)}$$

NOTE: Unless otherwise stated all distances will be expressed as nautical miles throughout this module.

Minimum Range

Recall from NEETS, Module 11, *Microwave Principles*, that the DUPLEXER alternately switches the antenna between the transmitter and receiver so that only one antenna need be used. This switching is necessary because the high-power pulses of the transmitter would destroy the receiver if energy were allowed to enter the receiver. As you probably already realize, timing of this switching action is critical to the operation of the radar system. What you may not realize is that the minimum range ability of the radar system is also affected by this timing. The two most important times in this action are PULSE WIDTH and RECOVERY TIME.

This timing action must be such that during the transmitted pulse (pulse width), only the transmitter can be connected to the antenna. Immediately after the pulse is transmitted, the antenna must be reconnected to the receiver.

The leading edge of the transmitted pulse causes the duplexer to align the antenna to the transmitter. This action is essentially instantaneous. At the end of the transmitted pulse, the trailing edge of the pulse causes the duplexer to line up the antenna with the receiver; however, this action is not instantaneous. A small amount of time elapses at this point that is referred to as recovery time. Therefore, the total time in which the receiver is unable to receive the reflected pulse is equal to the pulse width plus the recovery time. Note that any reflected pulses from close targets returning before the receiver is connected to the antenna will be undetected. The minimum range, in yards, at which a target can be detected is determined using the following formula (pulse width and recovery time are expressed in microseconds or fractions of microseconds):

$$\text{minimum range} = \frac{\text{pulse width} + \text{recovery time}}{2} \times 328 \text{ yards}$$

or

$$\text{minimum range} = (\text{pulse width} + \text{recovery time}) \times 164 \text{ yds}$$

For example, minimum range for a radar system with a pulse width of 25 microseconds and a recovery time of 0.1 microseconds is figured as follows:

$$\begin{aligned}\text{minimum range} &= (25 + 0.1) \times 164 \text{ yards} \\ &= 25.1 \times 164 \text{ yards} \\ &= 4,116 \text{ yards (approximate)}\end{aligned}$$

Most modern radar systems are designed with such small recovery times that this figure can often be ignored when figuring minimum range.

Maximum Range

The maximum range of a pulse radar system depends upon CARRIER FREQUENCY, PEAK POWER of the transmitted pulse, PULSE-REPETITION FREQUENCY (prf) or PULSE REPETITION RATE (prr), and RECEIVER SENSITIVITY with prf as the primary limiting factor. The peak power of the pulse determines what maximum range the pulse can travel to a target and still return a usable echo. A usable echo is the smallest signal detectable by a receiver system that can be processed and presented on an indicator.

The frequency of the rf energy in the pulse radiated by a radar is referred to as the CARRIER FREQUENCY of the radar system. The carrier frequency is often a limiting factor in the maximum range capability of a radar system because radio frequency energy above 3,000 megahertz is rapidly attenuated by the atmosphere. This decreases the usable range of radio-frequency energy. Therefore, as the carrier frequency is increased, the transmitted power must also be increased to cover the same range. Long-range coverage is more easily achieved at lower frequencies because atmospheric conditions have less effect on low-frequency energy.

Radar systems radiate each pulse at the carrier frequency during transmit time, wait for returning echoes during listening or rest time, and then radiate a second pulse, as shown in figure 1-3. The number of pulses radiated in one second is called the pulse-repetition frequency (prf), or the pulse-repetition rate (prr). The time between the beginning of one pulse and the start of the next pulse is called PULSE-REPETITION TIME (prt) and is equal to the reciprocal of prf as follows:

$$prt = \frac{1}{prf}$$

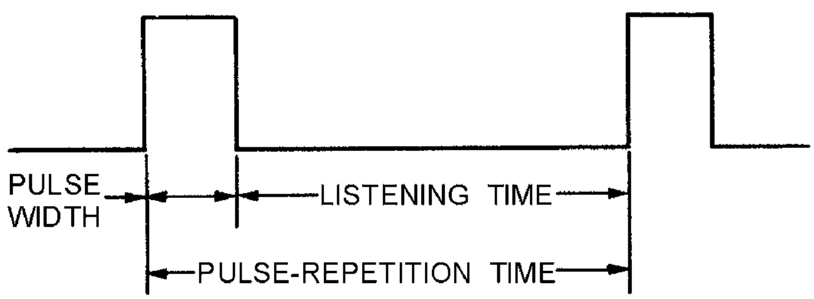


Figure 1-3.—Radar pulse relationships.

AMBIGUOUS RETURNS.—The radar timing system must be reset to zero each time a pulse is radiated. This is to ensure that the range detected is measured from time zero each time. The prt of the radar becomes important in maximum range determination because target return times that exceed the prt of the radar system appear at incorrect locations (ranges) on the radar screen. Returns that appear at these incorrect ranges are referred to as AMBIGUOUS RETURNS or SECOND-SWEEP ECHOES.

Figure 1-4 illustrates a radar system with a 1 millisecond prt. The pulses are shown at the top, and examples of two transmitted pulses hitting targets and returning are shown at the bottom. In the case of target A, the pulse travels round trip in 0.5 millisecond, which equates to a target range of 82,000 yards. Since 0.5 millisecond is less than 1 millisecond, displaying a correct range is no problem. However, target B is 196,800 yards distant from the radar system. In this case, total pulse travel time is 1.2 milliseconds and exceeds the prt limitation of 1 millisecond for this radar. While the first transmitted pulse is traveling to and returning from target B, a second pulse is transmitted and the radar system is reset to 0 again. The first pulse from target B continues its journey back to the radar system, but arrives during the timing period for the second pulse. This results in an inaccurate reading. In this case, the first return pulse from target B arrives 0.2 millisecond into the second timing period. This results in a range of 32,800 yards instead of the actual 196,800 yards. You should see from this example that pulse returns in excess of the prt of the radar system result in ambiguous ranges while pulse returns within the prt limits result in normal (unambiguous) ranges. The maximum unambiguous range for a given radar system can be determined by the following formula:

$$R_{max}\ \text{unambiguous} = \frac{162,000\ \text{miles/second}}{2} \times \text{prt}$$

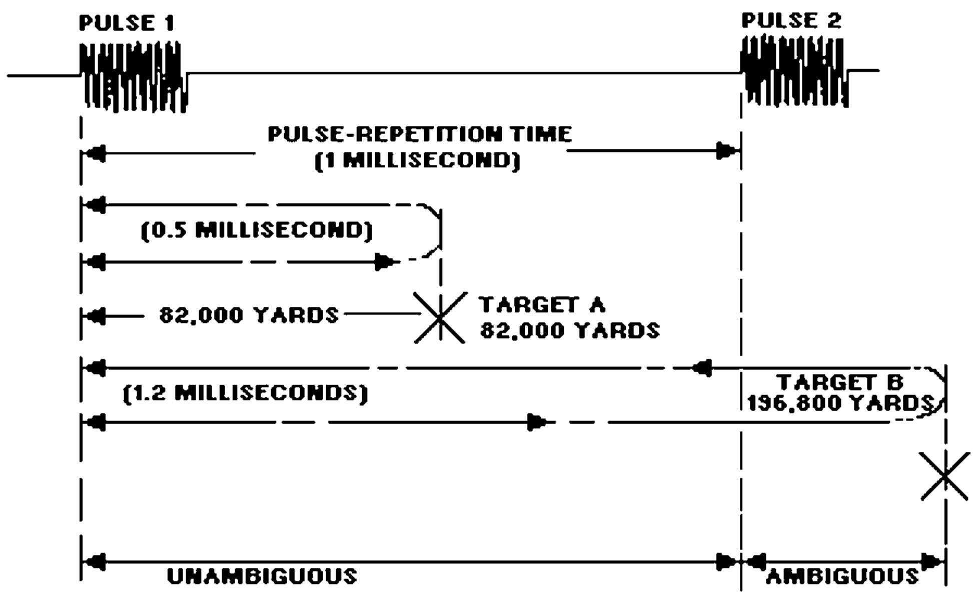


Figure 1-4.—Maximum unambiguous range.

Q3. What is the speed of electromagnetic energy traveling through air?

Q4. How much time is required for electromagnetic energy to travel 1 nautical mile and return to the source?

Q5. In addition to recovery time, what determines the minimum range of a radar set?

PULSE-REPETITION FREQUENCY AND POWER CALCULATIONS.—The energy content of a continuous-wave radar transmission may be easily figured because the transmitter operates continuously. However, pulsed radar transmitters are switched on and off to provide range timing information with each pulse. The resulting waveform for a transmitter was shown in figure 1-3. The amount of energy in this waveform is important because maximum range is directly related to transmitter output power. The more energy the radar system transmits, the greater the target detection range will be. The energy content of the pulse is equal to the PEAK (maximum) POWER LEVEL of the pulse multiplied by the pulse width. However, meters used to measure power in a radar system do so over a period of time that is longer than the pulse width. For this reason, pulse-repetition time is included in the power calculations for transmitters. Power measured over such a period of time is referred to as AVERAGE POWER. Figure 1-5 illustrates the way this average power would be shown as the *total* energy content of the pulse. The shaded area represents the total energy content of the pulse; the crosshatched area represents average power and is equal to peak power spread out over the prt. (Keep in mind, as you look at figure 1-5, that *no energy is actually present between pulses* in a pulsed radar system. The figure is drawn just to show you how average power is calculated.) Pulse-repetition time is used to help figure average power because it defines the total time from the beginning of one pulse to the beginning of the next pulse. Average power is figured as follows:

Where: P_{avg} = average power

P_{pk} = peak power

pw = pulse width

prt = pulse-repetition time

$$P_{avg} = P_{pk} \times \frac{pw}{prt}$$

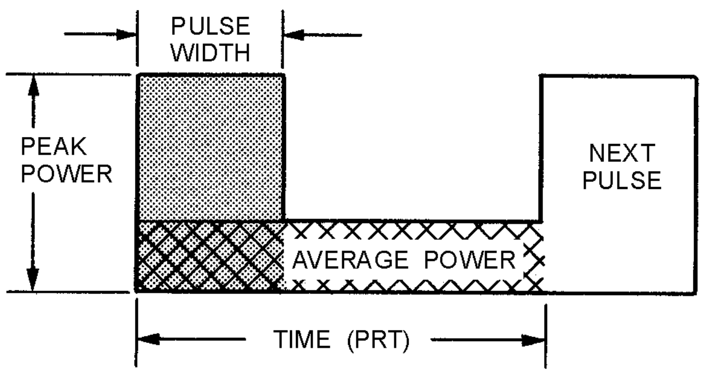


Figure 1-5.—Pulse energy content.

Because 1/prt is equal to prf, the formula may be written as follows:

$$P_{avg} = P_{pk} \times pw \times prf$$

The product of pulse width (pw) and pulse-repetition frequency (prf) in the above formula is called the DUTY CYCLE of a radar system. The duty cycle is a ratio of the *time on* to the *time off* of the transmitter, as shown in figure 1-6. The duty cycle is used to calculate both the peak power and average power of a radar system. The formula for duty cycle is shown below:

$$\text{duty cycle} = pw \times prf$$

NOTE: Pulse repetition frequency (prf) and pulse repetition rate (prr) are interchangeable terms.

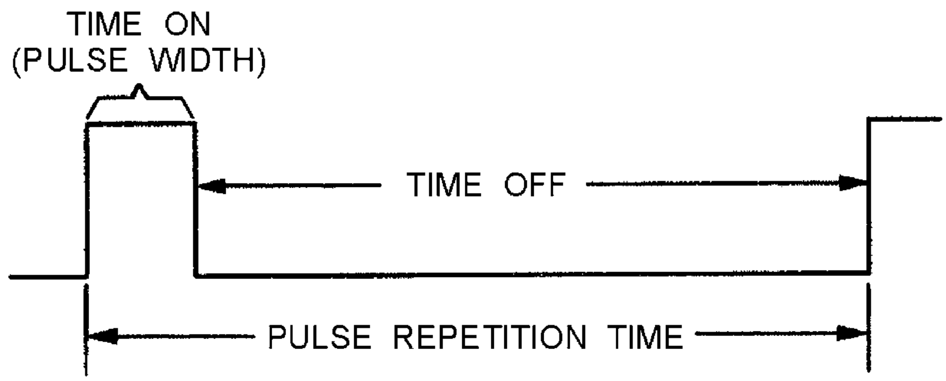


Figure 1-6.—Duty cycle.

Since the duty cycle of a radar is usually known, the most common formula for average power is expressed as:

$$P_{avg} = P_{pk} \times \text{duty cycle}$$

Transposing the above formula gives us a common formula for peak power:

$$P_{pk} = \frac{P_{avg}}{\text{duty cycle}}$$

Peak power must be calculated more often than average power. This is because, as previously mentioned, most measurement instruments measure average power directly. An example is shown below:

Where:

$$P_{avg} = 20,000 \text{ watts}$$
$$pw = 20 \text{ microseconds } (20 \times 10^{-6})$$
$$prf = 1,000 \text{ or } 10^{3} \text{ pulses per second}$$

Before figuring P*p*, you must figure duty cycle as follows:

$$\begin{aligned} \text{duty cycle} &= pw \times prf \\ &= 20 \times 10^{-6} \times 10^{3}\ 20 \times 10^{-3} \\ &= .02 \end{aligned}$$

Now that you have duty cycle, P*p* may be calculated as follows:

$$P_{pk} = \frac{P_{avg}}{\text{duty cycle}}$$

$$= \frac{20,000}{.02}$$

$$= 1,000,000 \text{ or } 10^6 \text{ watts}$$

ANTENNA HEIGHT AND SPEED.—Another factor affecting radar range is antenna height. The high-frequency energy transmitted by a radar system travels in a straight line and does not normally bend to conform to the curvature of the earth. Because of this, the height of both the antenna and the target are factors in detection range. The distance to the horizon (in nautical miles) for a radar system varies with the height of the antenna according to the following formula:

$$\text{radar horizon distance} = 1.25\sqrt{\text{antenna height in feet}}$$
$$\text{(in nautical miles)}$$

For example, assume antenna height to be 64 feet in the following calculations:

$$\begin{aligned}\text{horizon distance} &= 1.25\sqrt{\text{antenna height}}\\ &= 1.25\sqrt{64 \text{ feet}}\\ &= 1.25 \times 8 \text{ feet}\\ &= 10 \text{ nautical miles}\end{aligned}$$

A target at a range greater than the radar horizon will not be detected unless it is high enough to be above the horizon. An example of the antenna- and target-height relationship is shown in figure 1-7.

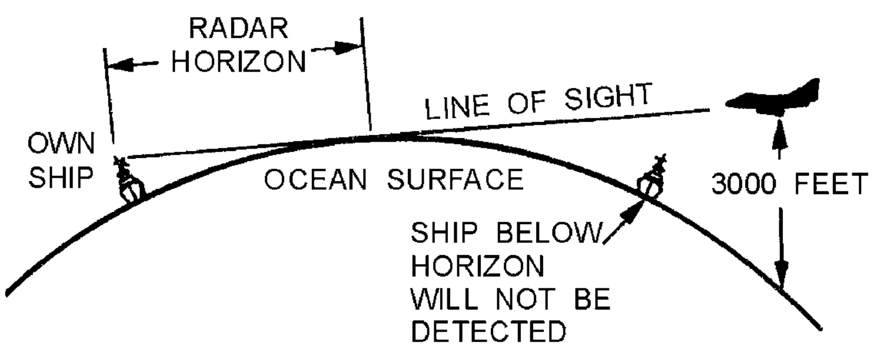


Figure 1-7.—Radar horizon.

The antenna-rotation rate also affects maximum detection range. The slower an antenna rotates, the greater the detection range of a radar system. When the antenna is rotated at 10 revolutions per minute (rpm), the beam of energy strikes each target for just one-half the time it would if the rotation were 5 rpm.

The number of strikes per antenna revolution is referred to as HITS PER SCAN. During each revolution enough pulses must be transmitted to return a usable echo.

NOTE: The more pulses transmitted to a given area (at slower antenna speeds), the greater the number of hits per scan.

As an example, if the antenna rotates at 20 rpm, it completes a revolution in 3 seconds. During this time, a transmitter with a prf of 200 pulses per second (pps) transmits 600 pulses. Since 360 degrees of azimuth must be covered, the following formula shows the number of pulses for each degree of azimuth:

$$\frac{600 \text{ pulses per revolution}}{360 \text{ degrees per revolution}} = 1.67 \text{ pulses per degree}$$

Such a low number of pulses for any given target area greatly increases the likelihood that some targets will be missed entirely; therefore, prf and antenna speed must be matched for maximum efficiency.

Q6. Atmospheric interference with the travel of electromagnetic energy increases with what rf energy characteristic?

Q7. How is prt related to prf?

Q8. What type of radar transmitter power is measured over a period of time?

Q9. What term is used to describe the product of pulse width and pulse-repetition frequency?

BEARING

The TRUE BEARING (referenced to true north) of a radar target is the angle between true north and a line pointed directly at the target. This angle is measured in the horizontal plane and in a clockwise direction from true north. The bearing angle to the radar target may also be measured in a clockwise direction from the centerline of your own ship or aircraft and is referred to as the RELATIVE BEARING. Both true and relative bearing angles are illustrated in figure 1-8.

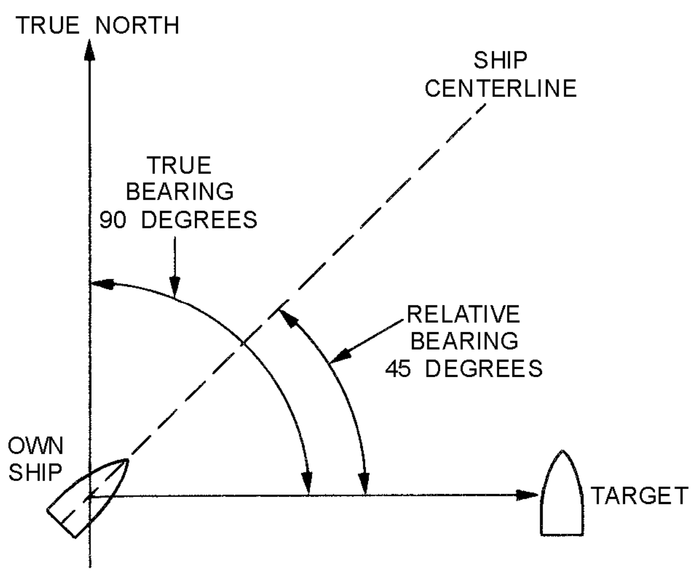


Figure 1-8.—True and relative bearings.

The antennas of most radar systems are designed to radiate energy in a one-directional lobe or beam that can be moved in bearing simply by moving the antenna. As you can see in figure 1-9, the shape of the beam is such that the echo signal strength varies in amplitude as the antenna beam moves across the target. At antenna position **A**, the echo is minimal; at position **B**, where the beam axis is pointing directly at the target, the echo strength is maximum. Thus, the bearing angle of the target can be obtained by moving the antenna to the position at which the echo is strongest. In actual practice, search radar antennas move continuously; the point of maximum echo return is determined by the detection circuitry as the beam passes the target or visually by the operator. Weapons-control and guidance radar systems are positioned to the point of maximum signal return and maintained at that position either manually or by automatic tracking circuits.

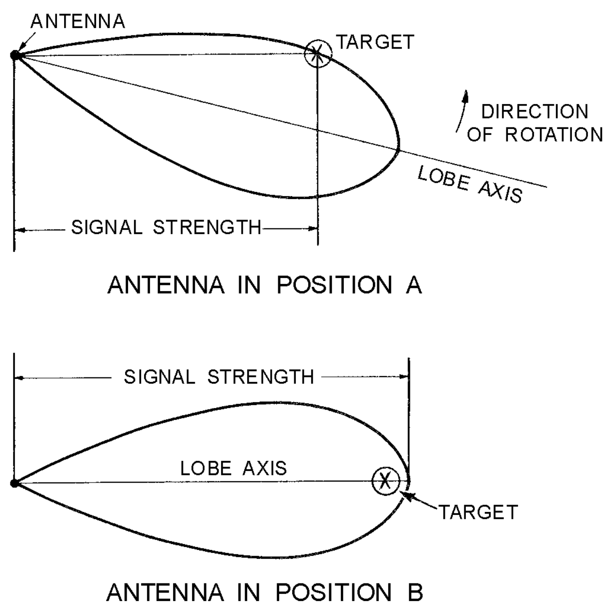


Figure 1-9.—Determination of bearing.

ALTITUDE

Many radar systems are designed to determine only the range and bearing of an object. Such radar systems are called TWO-DIMENSIONAL (2D) radars. In most cases these systems are further described as SEARCH RADAR SYSTEMS and function as early-warning devices that search a fixed volume of space. The range and bearing coordinates provide enough information to place the target in a general area with respect to the radar site and to determine distance, direction of travel, and relative speed. However, when action must be taken against an airborne target, altitude must be known as well. A search radar system that detects altitude as well as range and bearing is called a THREE-DIMENSIONAL (3D) radar.

Altitude- or height-finding search radars use a beam that is very narrow in the vertical plane. The beam is scanned in elevation, either mechanically or electronically, to pinpoint targets. Height-finding radar systems that also determine bearing must have a beam that is very narrow in both the vertical and horizontal planes. An electronic elevation-scanning pattern for a search radar set is illustrated in figure 1-10. Lines originating at the antenna indicate the number of beam positions required for complete elevation coverage. In practice the beams overlap slightly to prevent any gaps in the coverage. Each beam position corresponds to a slight change in either the frequency or phase of the radiated energy. A change in either phase or frequency of the energy causes it to leave the antenna at a different angle. Thus, the frequency or phase can be predetermined to create an orderly scanning pattern that covers the entire vertical plane. Electronic scanning permits automatic compensation for an unstable radar platform (site), such as a ship at sea. Error signals are produced by the roll and pitch of the ship and are used to correct the radar beam to ensure complete elevation coverage.

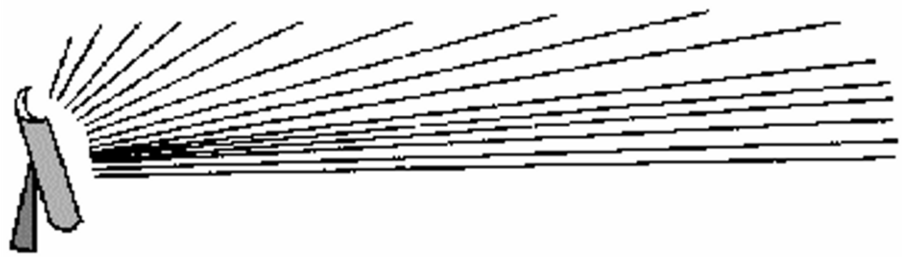


Figure 1-10.—Electronic elevation scan.

Mechanical elevation scanning is achieved by mechanically moving the antenna or radiation source. Weapons-control and tracking radar systems commonly use mechanical elevation scanning techniques. Most electronically scanned radar systems are used as air search radars. Some older air-search radar systems use a mechanical elevation scanning device; however, these are being replaced by electronically scanned radar systems.

Q10. What type of target bearing is referenced to your ship?

Q11. What type of radar detects range, bearing, and height?

Q12. What characteristic(s) of radiated energy is (are) altered to achieve electronic scanning?

TARGET RESOLUTION

The TARGET RESOLUTION of a radar is its ability to distinguish between targets that are very close together in either range or bearing. Weapons-control radar, which requires great precision, should be able to distinguish between targets that are only yards apart. Search radar is usually less precise and only distinguishes between targets that are hundreds of yards or even miles apart. Resolution is usually divided into two categories; RANGE RESOLUTION and BEARING RESOLUTION.

Range Resolution

Range resolution is the ability of a radar system to distinguish between two or more targets on the same bearing but at different ranges. The degree of range resolution depends on the width of the transmitted pulse, the types and sizes of targets, and the efficiency of the receiver and indicator. Pulse width is the primary factor in range resolution. A well-designed radar system, with all other factors at maximum efficiency, should be able to distinguish targets separated by one-half the pulse width time. Therefore, the theoretical range resolution of a radar system can be calculated from the following formula:

$$\text{range resolution (in yards)} = \frac{\text{pw (microseconds)}}{2} \times 328 \text{ yards per microsecond}$$

The above formula is often written as:

$$\text{range resolution} = \text{pw} \times 164 \text{ yards per microsecond}$$

For example, if a radar system has a pulse width of 5 microseconds, the range resolution is calculated as follows:

$$\begin{aligned}\text{range resolution} &= \text{pw} \times 164 \text{ yards per microsecond}\\ &= 5 \times 164 \text{ yards per microsecond}\\ &= 820 \text{ yards}\end{aligned}$$

In the above example, targets on the same bearing would have to be separated by more than 820 yards to show up as two targets on your indicator.

Bearing Resolution

Bearing, or azimuth, resolution is the ability of a radar system to separate objects at the same range but at different bearings. The degree of bearing resolution depends on radar beam width and the range of the targets. Range is a factor in bearing resolution because the radar beam spreads out as range increases. A RADAR BEAM is defined in width in terms of HALF-POWER POINTS. All the points off the centerline of the beam that are at one-half the power level at the center are plotted to define beam width. When the half-power points are connected to the antenna by a curve, such as that shown in figure 1-11, the resulting angular width of the curve is called the ANTENNA BEAM WIDTH. The physical size and shape of the antenna determines beam width. Beam width can vary from about 1 degree up to 60 degrees. In figure 1-11, only the target within the half-power points will reflect a useful echo. Two targets at the same range must be separated by at least one beam width to be distinguished as two objects.

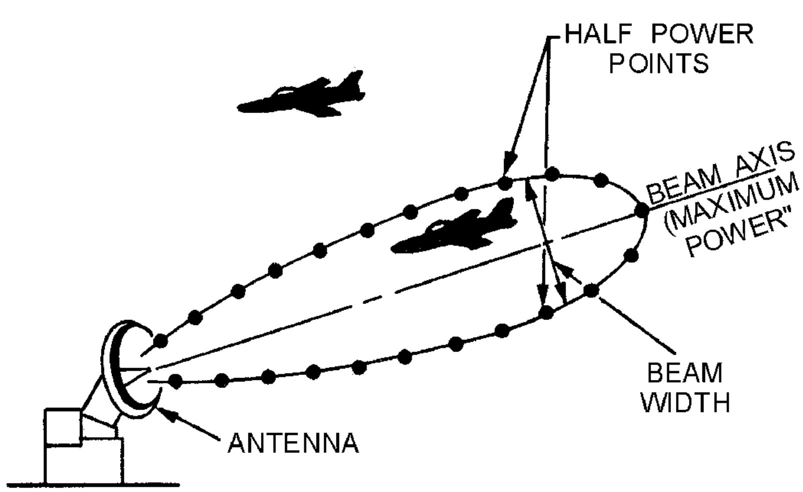


Figure 1-11.—Beam half-power points.

RADAR ACCURACY

Radar accuracy is a measure of the ability of a radar system to determine the correct range, bearing, and, in some cases, height of an object. The degree of accuracy is primarily determined by the resolution of the radar system. Some additional factors affecting accuracy are pulse shape and atmospheric conditions.

Pulse Shape

In the case of a pulse radar, the shape and width of the rf pulse influences minimum range, range accuracy, and maximum range. The ideal pulse shape is a square wave having vertical leading and trailing edges. However, equipments do not usually produce the ideal waveforms.

The factors influencing minimum range are discussed first. Since the receiver cannot receive target reflections while the transmitter is operating, you should be able to see that a narrow pulse is necessary for short ranges. A sloping trailing edge extends the width of the transmitter pulse, although it may add very little to the total power generated. Therefore, along with a narrow pulse, the trailing edge should be as near vertical as possible.

A sloping leading edge also affects minimum range as well as range accuracy since it provides no definite point from which to measure elapsed time on the indicator time base. Using a starting point at the lower edge of the pulse's leading edge would increase minimum range. Using a starting point high up on the slope would reduce the accuracy of range measurements at short ranges which are so vital for accurate solution of the fire-control problem.

Maximum range is influenced by pulse width and pulse repetition frequency (prf). Since a target can reflect only a very small part of the transmitted power, the greater the transmitted power, the greater the strength of the echo that could be received. Thus, a transmitted pulse should quickly rise to its maximum amplitude, remain at this amplitude for the duration of the desired pulse width, and decay instantaneously to zero. Figure 1-12 illustrates the effects of pulse shapes.

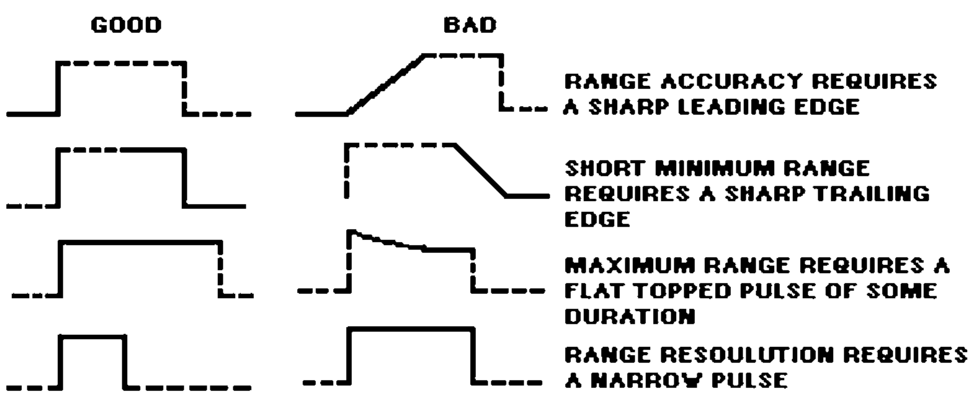


Figure 1-12.—Pulse shapes and effects.

Atmospheric Conditions

Electromagnetic wavefronts travel through empty space in straight lines at the speed of light, but the REFRACTIVE INDEX of the atmosphere affects both the travel path and the speed of the electromagnetic wavefront. The path followed by electromagnetic energy in the atmosphere, whether direct or reflected, usually is slightly curved; and the speed is affected by temperature, atmospheric pressure, and the amount of water vapor present in the atmosphere, which all affect the refractive index. As altitude increases, the combined effects of these influences, under normal atmospheric conditions, cause a small, uniform increase in signal speed. This increase in speed causes the travel path to curve slightly downward, as shown in figure 1-13. The downward curve extends the radar horizon beyond a line tangent to the earth, as illustrated in figure 1-14.

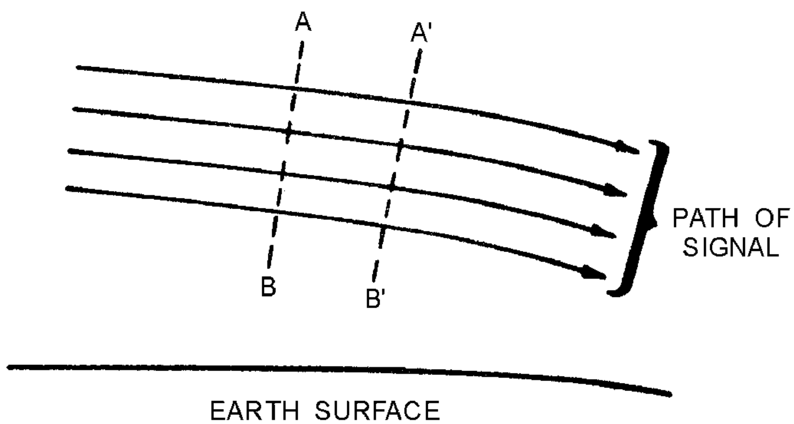


Figure 1-13.—Wavefront path.

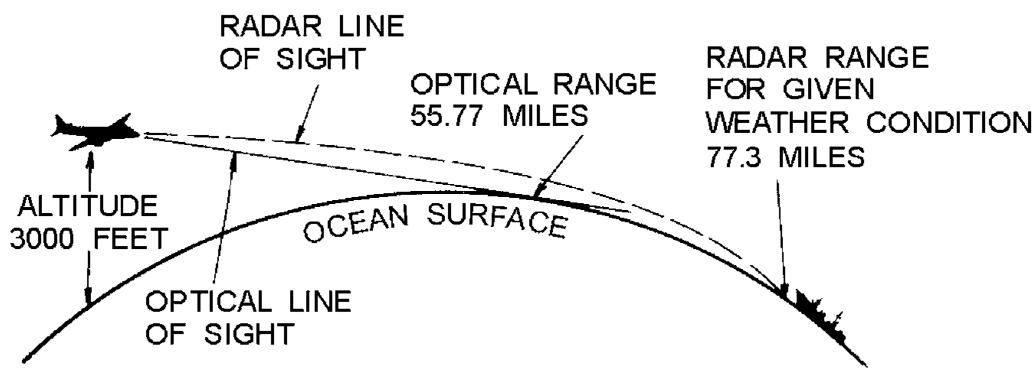


Figure 1-14.—Extension of the radar horizon.

The reason for the downward curve can be illustrated using line **AB** in figure 1-13. Line **AB** represents the surface of a wavefront with point A higher in altitude than point B. As wavefront AB moves to the point represented by A'B', the speed at A and A' is faster than the speed at B and B' since A and A' are at a greater altitude. Therefore, in a given time, the upper part of the wavefront moves farther than the lower part. The wavefront leans slightly forward as it moves. Since the direction of energy propagation is always perpendicular to the surface of a wavefront, the tilted wavefront causes the energy path to curve downward.

REFRACTION is the bending of electromagnetic waves caused by a change in the density of the medium through which the waves are passing. A visible example of electromagnetic refraction is the apparent displacement of underwater objects caused by the bending of light as it passes from the atmosphere into the water. An INDEX OF REFRACTION has been established which indicates the degree of refraction, or bending, caused by different substances. Because the density of the atmosphere changes with altitude, the index of refraction changes gradually with height.

The temperature and moisture content of the atmosphere normally decrease uniformly with an increase in altitude. However, under certain conditions the temperature may first increase with height and then begin to decrease. Such a situation is called a temperature inversion. An even more important deviation from normal may exist over the ocean. Since the atmosphere close to the surface over large bodies of water may contain more than a normal amount of moisture, the moisture content may decrease more rapidly at heights just above the sea. This effect is referred to as MOISTURE LAPSE.

Either temperature inversion or moisture lapse, alone or in combination, can cause a large change in the refraction index of the lowest few-hundred feet of the atmosphere. The result is a greater bending of the radar waves passing through the abnormal condition. The increased bending in such a situation is referred to as DUCTING and may greatly affect radar performance. The radar horizon may be extended or reduced, depending on the direction the radar waves are bent. The effect of ducting on radar waves is illustrated in figure 1-15.

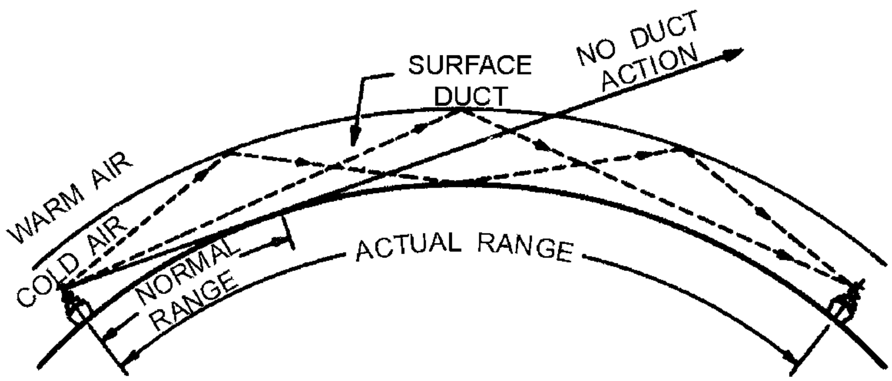


Figure 1-15.—Ducting effect on the radar wave.

Another effect of the atmosphere on radar performance is caused by particles suspended in the air. Water droplets and dust particles diffuse radar energy through absorption, reflection, and scattering so less energy strikes the target. Consequently, the return echo is smaller. The overall effect is a reduction in usable range that varies widely with weather conditions. The higher the frequency of a radar system, the more it is affected by weather conditions such as rain or clouds. In some parts of the world, dust suspended in the air can greatly decrease the normal range of high-frequency radar.

Q13. What term is used to describe the ability of a radar system to distinguish between targets that are close together?

Q14. The degree of bearing resolution for a given radar system depends on what two factors?

Q15. What happens to the speed of electromagnetic energy traveling through air as the altitude increases?

Q16. What term is used to describe a situation in which atmospheric temperature first increases with altitude and then begins to decrease?

RADAR PRINCIPLES OF OPERATION

Radar systems, like other complex electronics systems, are composed of several major subsystems and many individual circuits. This section will introduce you to the major subsystems common to most radar sets. A brief functional description of subsystem principles of operation will be provided. A much more detailed explanation of radar subsystems will be given in chapters 2 and 3. Since most radar systems in use today are some variation of the pulse radar system, the units discussed in this section will be those used in pulse radar. All other types of radar use some variation of these units, and these variations will be explained as necessary.

RADAR COMPONENTS

Pulse radar systems can be functionally divided into the six essential components shown in figure 1-16. These components are briefly described in the following paragraphs and will be explained in detail after that:

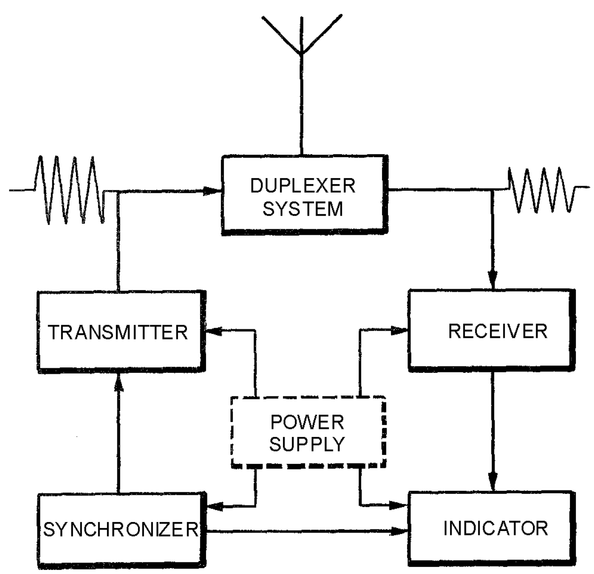


Figure 1-16.—Functional block diagram of a basic radar system.

- The SYNCHRONIZER (also referred to as the TIMER or KEYER) supplies the synchronizing signals that time the transmitted pulses, the indicator, and other associated circuits.
- The TRANSMITTER generates electromagnetic energy in the form of short, powerful pulses.
- The DUPLEXER allows the same antenna to be used for transmitting and receiving.
- The ANTENNA SYSTEM routes the electromagnetic energy from the transmitter, radiates it in a highly directional beam, receives any returning echoes, and routes those echoes to the receiver.
- The RECEIVER amplifies the weak, electromagnetic pulses returned from the reflecting object and reproduces them as video pulses that are sent to the indicator.

- The INDICATOR produces a visual indication of the echo pulses in a manner that, at a minimum, furnishes range and bearing information.

While the physical configurations of radar systems differ, any radar system can be represented by the functional block diagram in figure 1-16. An actual radar set may have several of these functional components within one physical unit, or a single one of these functions may require several physical units. However, the functional block diagram of a basic radar set may be used to analyze the operation of almost any radar set.

In the following paragraphs, a brief description of the operation of each of the major components is given.

Synchronizer (Timer)

The synchronizer ensures that all circuits connected with the radar system operate in a definite timed relationship. It also times the interval between transmitted pulses to ensure that the interval is of the proper length. Timing pulses are used to ensure synchronous circuit operation and are related to the prf. The prf can be set by any stable oscillator, such as a sine-wave oscillator, multivibrator, or a blocking oscillator. That output is then applied to pulse-shaping circuits to produce timing pulses. Associated components can be timed by the output of the synchronizer or by a timing signal from the transmitter as it is turned on.

Transmitter

The transmitter generates powerful pulses of electromagnetic energy at precise intervals. The required power is obtained by using a high-power microwave oscillator, such as a magnetron, or a microwave amplifier, such as a klystron, that is supplied by a low-power rf source. (The construction and operation of microwave components can be reviewed in NEETS, Module 11, *Microwave Principles*.) The high-power generator, whether an oscillator or amplifier, requires operating power in the form of a properly-timed, high-amplitude, rectangular pulse. This pulse is supplied by a transmitter unit called the MODULATOR. When a high-power oscillator is used, the modulator high-voltage pulse switches the oscillator on and off to supply high-power electromagnetic energy. When a microwave power amplifier is used, the modulator pulse activates the amplifier just before the arrival of an electromagnetic pulse from a preceding stage or a frequency-generation source. Normally, because of the extremely high voltage involved, the modulator pulse is supplied to the cathode of the power tube and the plate is at ground potential to shield personnel from shock hazards. The modulator pulse may be more than 100,000 volts in high-power radar transmitters. In any case, radar transmitters produce voltages, currents, and radiation hazards that are extremely dangerous to personnel. Safety precautions must always be strictly observed when working in or around a radar transmitter.

Duplexer

A duplexer is essentially an electronic switch that permits a radar system to use a single antenna to both transmit and receive. The duplexer must connect the antenna to the transmitter and disconnect the antenna from the receiver for the duration of the transmitted pulse. The receiver must be completely isolated from the transmitted pulse to avoid damage to the extremely sensitive receiver input circuitry. After the transmitter pulse has ended, the duplexer must rapidly disconnect the transmitter and connect the receiver to the antenna. As previously mentioned, the switching time is called receiver recovery time, and must be very fast if close-in targets are to be detected. Additionally, the duplexer should absorb very little power during either phase of operation. Low-loss characteristics are particularly important during the receive period of duplexer operation. This is because the received signals are of extremely low amplitude.

Antenna System

The antenna system routes the pulse from the transmitter, radiates it in a directional beam, picks up the returning echo, and passes it to the receiver with a minimum of loss. The antenna system includes the antenna, transmission lines and waveguide from the transmitter to the antenna, and the transmission line and waveguide from the antenna to the receiver. In some publications the duplexer is included as a component of the antenna system.

Receiver

The receiver accepts the weak echo signals from the antenna system, amplifies them, detects the pulse envelope, amplifies the pulses, and then routes them to the indicator. One of the primary functions of the radar receiver is to convert the frequency of the received echo signal to a lower frequency that is easier to amplify. This is because radar frequencies are very high and difficult to amplify. This lower frequency is called the INTERMEDIATE FREQUENCY (IF). The type of receiver that uses this frequency conversion technique is the SUPER HETERODYNE RECEIVER. Superheterodyne receivers used in radar systems must have good stability and extreme sensitivity. Stability is ensured by careful design and the overall sensitivity is greatly increased by the use of many IF stages.

Indicator

The indicator uses the received signals routed from the radar receiver to produce a visual indication of target information. The cathode-ray oscilloscope is an ideal instrument for the presentation of radar data. This is because it not only shows a variation of a single quantity, such as voltage, but also gives an indication of the relative values of two or more quantities. The sweep frequency of the radar indicator is determined by the pulse-repetition frequency of the radar system. Sweep duration is determined by the setting of the range-selector switch. Since the indicator is so similar to an oscilloscope, the term RADAR SCOPE is commonly used when referring to radar indicators.

Q17. What radar subsystem supplies timing signals to coordinate the operation of the complete system?

Q18. When a transmitter uses a high-power oscillator to produce the output pulse, what switches the oscillator on and off?

Q19. What radar component permits the use of a single antenna for both transmitting and receiving?

SCANNING

Radar systems are often identified by the type of SCANNING the system uses. Scanning is the systematic movement of a radar beam in a definite pattern while searching for or tracking a target. The type and method of scanning used depends on the purpose and type of radar and on the antenna size and design. In some cases, the type of scan will change with the particular system mode of operation. For example, in a particular radar system, the search mode scan may be quite different from that of the track mode scan.

Stationary-Lobe Scanning

A SINGLE STATIONARY-LOBE SCANNING SYSTEM is the simplest type of scanning. This method produces a single beam that is stationary in relation to the antenna. The antenna is then mechanically rotated continuously to obtain complete 360-degree azimuth coverage. A stationary lobe, however, cannot satisfactorily track a moving object because it does not provide enough information about the object's movement to operate automatic tracking circuits, such as those in fire-control tracking radar. A two-dimensional search radar, however, does use a single-lobe that is scanned in a 360-degree pattern because automatic tracking circuits are not normally used in 2D radars.

Single-lobe scanning is unsuitable for use as a tracking radar for several reasons. For example, let's assume that a target is somewhere on the lobe axis and the receiver is detecting signals reflected from the target. If these reflected signals begin to decrease in strength, the target likely has flown off the lobe axis. In this case, the beam must be moved to continue tracking. The beam might be moved by an operator tracking the target with an optical sight, but such tracking is slow, inaccurate, and limited by conditions of visibility. An automatic tracking system would require that the beam SCAN, or search, the target area in such a case.

Again, assume that a missile is riding (following) the axis of a single beam. The strength of the signals it receives (by means of a radar receiver in the missile) will gradually decrease as its distance from the transmitter increases. If the signal strength decreases suddenly, the missile will know, from built-in detection circuitry, that it is no longer on the axis of the lobe. But it will *not* know which way to turn to get back on the axis. A simple beam does not contain enough information for missile guidance.

Methods of Beam Scanning

The two basic methods of beam scanning are MECHANICAL and ELECTRONIC. In mechanical scanning, the beam can be moved in various ways: (1) The entire antenna can be moved in the desired pattern; (2) the energy feed source can be moved relative to a fixed reflector; or (3) the reflector can be moved relative to a fixed source. In electronic scanning, the beam is effectively moved by such means as (1) switching between a set of feeder sources, (2) varying the phasing between elements in a multielement array, or (3) comparing the amplitude and phase differences between signals received by a multielement array. A combination of mechanical and electronic scanning is also used in some antenna systems.

MECHANICAL SCANNING.—The most common type of mechanical scanning is the rotation of the antenna through 360 degrees to obtain azimuth coverage. Most search radar sets use this method. A common form of scanning for target tracking or missile beam-rider systems is CONICAL (cone-like) SCANNING. This is generally accomplished mechanically by NUTATING the rf feed point.

Nutation is difficult to describe in words but easy to demonstrate. Hold a pencil in two hands. While holding the eraser end as still as possible, swing the point in a circular motion. This motion of the pencil is referred to as nutation; the pencil point corresponds to the open, or transmitting, end of the waveguide antenna. The important fact to remember is that polarization of the beam is not changed during the scanning cycle. This means that the axis of the moving feeder does not change either horizontal or vertical orientation while the feeder is moving. You might compare the feeder movement to that of a ferris wheel; that is, the vertical orientation of each seat remains the same regardless of the position of the wheel.

Recall that a waveguide is a metal pipe, usually rectangular in cross section, used to conduct the rf energy from the transmitter to the antenna. The open end of the waveguide faces the concave side of the reflector and the rf energy it emits is bounced from the reflector surface.

A conical scan can be generated by nutation of the waveguide. In this process the axis of the waveguide itself is moved through a small conical pattern. In an actual installation of a nutating waveguide, the three-dimensional movement is fast and of small amplitude. To an observer, the waveguide appears merely to be vibrating slightly.

By movement of either the waveguide or the antenna, you can generate a conical scan pattern, as shown in figure 1-17. The axis of the radar lobe is made to sweep out a cone in space; the apex of this cone is, of course, at the radar transmitter antenna or reflector. At any given distance from the antenna, the path of the lobe axis is a circle. Within the useful range of the beam, the inner edge of the lobe always overlaps the axis of scan.

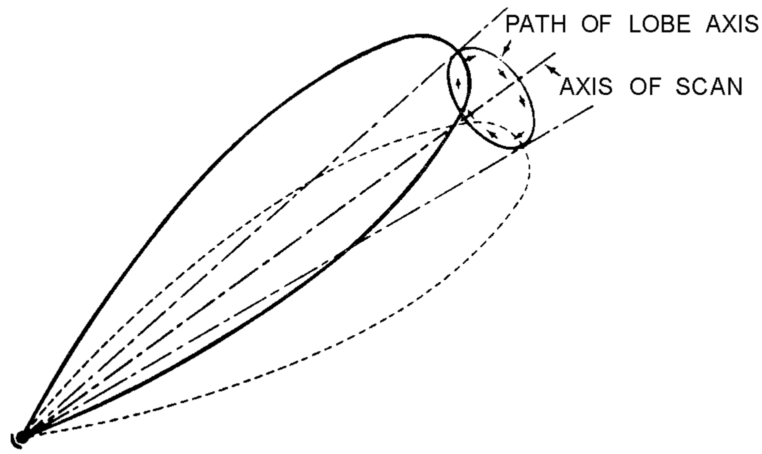


Figure 1-17.—Conical scanning.

Now assume that we use a conically scanned beam for target tracking. If the target is on the scan axis, the strength of the reflected signals remains constant (or changes gradually as the range changes). But if the target is slightly off the axis, the amplitude of the reflected signals will change at the scan rate. For example, if the target is to the left of the scan axis, as shown in figure 1-18, the reflected signals will be of maximum strength as the lobe sweeps through the left part of its cone; the signals will quickly decrease to a minimum as the lobe sweeps through the right part. Information on the instantaneous position of the beam, relative to the scan axis, and on the strength of the reflected signals is fed to a computer. Such a computer in the radar system is referred to as the angle-tracking or angle-servo circuit (also angle-error detector). If the target moves off the scan axis, the computer instantly determines the direction and amount of antenna movement required to continue tracking. The computer output is used to control servomechanisms that move the antenna. In this way, the target is tracked accurately and automatically.

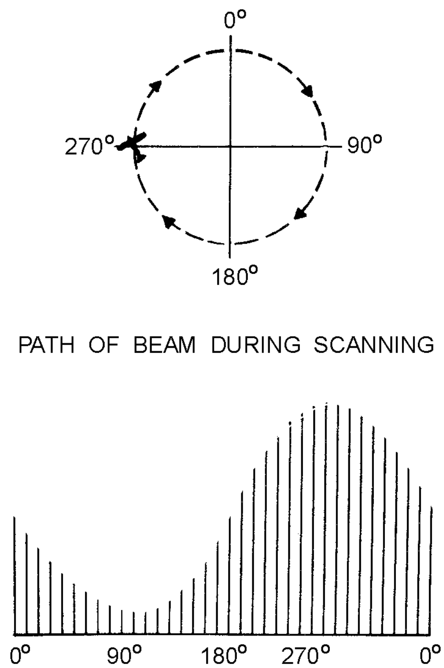


Figure 1-18.—Reflected signal strength.

Q20. What is the simplest type of scanning?

Q21. How does the operator of a single-lobe scanning system determine when the target moves off the lobe axis?

Q22. What are the two basic methods of scanning?

Q23. Rotation of an rf-feed source to produce a conical scan pattern is identified by what term?

ELECTRONIC SCANNING.—Electronic scanning can accomplish lobe motion more rapidly than, and without the inherent maintenance disadvantages of, the mechanical systems. Because electronic scanning cannot generally cover as large an area of space, it is sometimes combined with mechanical scanning in particular applications.

With MONOPULSE (SIMULTANEOUS) LOBING, all range, bearing, and elevation-angle information of a target is obtained from a single pulse. Monopulse scanning is used in fire-control tracking radars.

For target tracking, the radar discussed here produces a narrow circular beam of pulsed-rf energy at a high pulse-repetition rate. Each pulse is divided into four signals which are equal both in amplitude and phase. These four signals are radiated at the same time from each of four feedhorns that are grouped in a cluster. The resulting radiated energy is focused into a beam by a microwave lens. Energy reflected from targets is refocused by the lens back into the feedhorns. The total amount of the energy received by each horn varies, depending on the position of the target relative to the beam axis. This is illustrated in figure 1-19 for four targets at different positions with respect to the beam axis. Note that a phase inversion takes place at the microwave lens similar to the image inversion that takes place in an optical system.

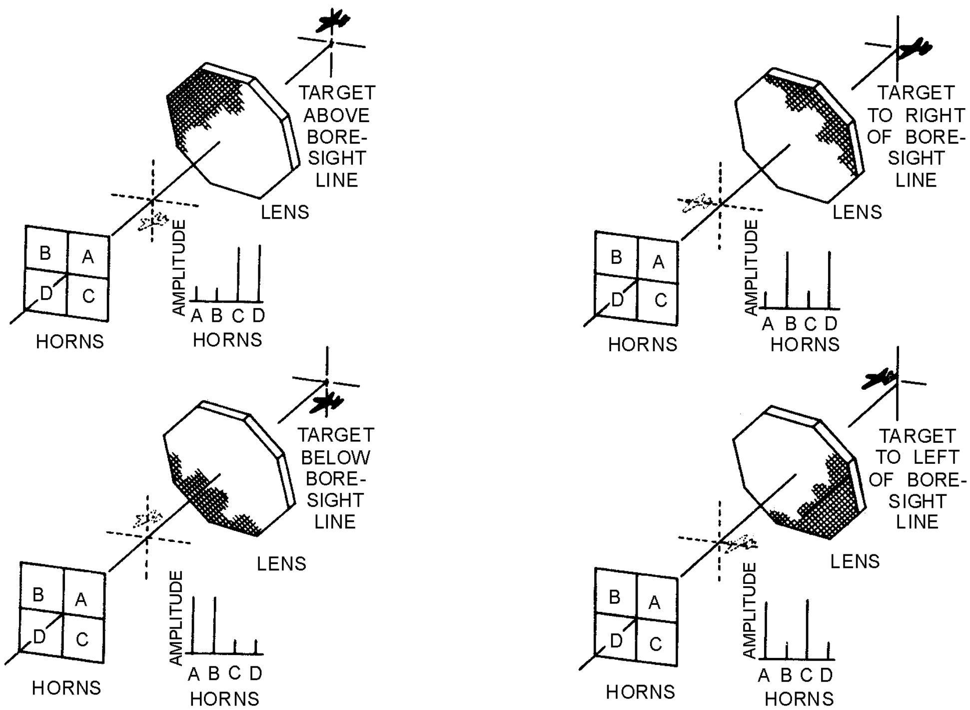


Figure 1-19.—Monopulse scanning.

The amplitude of returned signals received by each horn is continuously compared with those received in the other horns. Error signals are generated which indicate the relative position of the target with respect to the axis of the beam. Angle servo circuits receive these error signals and correct the position of the radar beam to keep the beam axis on target.

The TRAVERSE (BEARING) SIGNAL is made up of signals from horn **A** added to **C** and from horn **B** added to **D**. By waveguide design, the sum of **B** and **D** is made 180 degrees out of phase with the sum of **A** and **C**. These two are combined and the traverse signal is the difference of (**A** + **C**) − (**B** + **D**). Since the horns are positioned as shown in figure 1-19, the relative amplitudes of the horn signals give an indication of the magnitude of the traverse error. The elevation signal consists of the signals from horns **C** and **D** added 180 degrees out of phase with horns **A** and **B** [(**A** + **B**) − (**C** + **D**)]. The sum, or range, signal is composed of signals from all four feedhorns added together in phase. It provides a reference from which target direction from the center of the beam axis is measured. The range signal is also used as a phase reference for the traverse and elevation-error signals.

The traverse and elevation error signals are compared in the radar receiver with the range or reference signal. The output of the receiver may be either positive or negative pulses; the amplitudes of the pulses are proportional to the angle between the beam axis and a line drawn to the target. The polarities of the output pulses indicate whether the target is above or below, to the right or to the left of the beam axis. Of course, if the target is directly on the line of sight, the output of the receiver is zero and no angle-tracking error is produced.

An important advantage of a monopulse-tracking radar over radar using conical scan is that the instantaneous angular measurements are not subject to errors caused by target SCINTILLATION. Scintillation can occur as the target maneuvers or moves and the radar pulses bounce off different areas of the target. This causes random reflectivity and may lead to tracking errors. Monopulse tracking radar is not subject to this type of error because each pulse provides an angular measurement without regard to the rest of the pulse train; no such cross-section fluctuations can affect the measurement. An additional advantage of monopulse tracking is that no mechanical action is required.

ELECTRONIC SCANNING used in search radar systems was explained in general terms earlier in this chapter during the discussion of elevation coverage. This type of electronic scanning is often called FREQUENCY SCANNING. An in-depth explanation of frequency scanning theory can be found in the fire control technician rate training manuals.

RADAR TRANSMISSION METHODS

Radar systems are normally divided into operational categories based on energy transmission methods. Up to this point, we have mentioned only the pulse method of transmission to illustrate basic radar concepts. Although the pulse method is the most common method of transmitting radar energy, two other methods are sometimes used in special applications. These are the continuous-wave (cw) method and the frequency modulation (fm) method. All three basic transmission methods are often further subdivided to designate specific variations or combinations.

CONTINUOUS-WAVE METHOD

When radio-frequency energy transmitted from a fixed point continuously strikes an object that is either moving toward or away from the source of the energy, the *frequency* of the reflected energy is changed. This shift in frequency is known as the DOPPLER EFFECT. The difference in frequency between the transmitted and reflected energy indicates both the presence and the speed of a moving target.

Doppler Effect

A common example of the Doppler effect in action is the changing pitch of the whistle of an approaching train. The whistle appears to change pitch from a high tone, as the train approaches, to a lower tone as it moves away from the observer. As the train approaches, an apparent increase in frequency (an increase in pitch) is heard; as the train moves away, an apparent decrease in frequency (a decrease in pitch) is heard. This pitch variation is known as the Doppler effect.

Let's examine the reason for this apparent change in pitch. Assume that the transmitter emits an audio signal at a frequency of 60 hertz and that the transmitter is traveling at a velocity of 360 feet per second (fps). At the end of 1 second, the transmitter will have moved from point P to point P1 as shown in view A of figure 1-20. The total distance from point P to the observer is 1,080 feet. The velocity of sound is 1,080 feet per second; thus, a sound emitted at point P will reach the observer in 1 second. To find the wavelength of this transmitted signal, you divide the velocity of the signal (1,080 fps) by the frequency (60 hertz). The result is 18 feet, as shown below:

$$\begin{aligned} \text{wavelength} &= \frac{\text{velocity}}{\text{frequency}} \\ &= \frac{1,080 \text{ fps}}{60 \text{ Hz}} \\ &= 18 \text{ feet} \end{aligned}$$

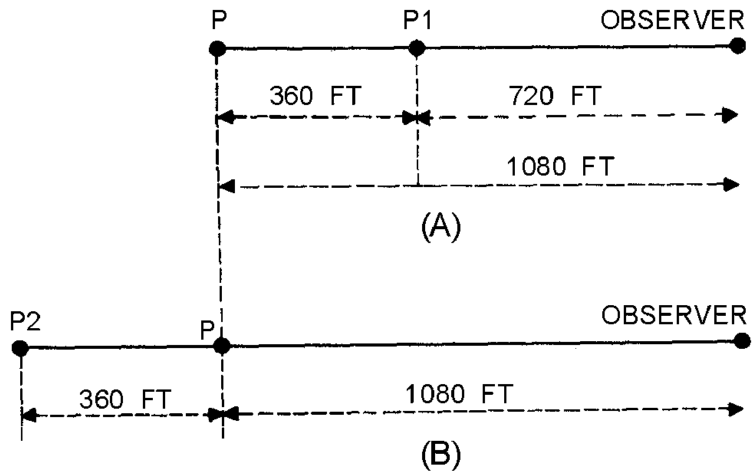


Figure 1-20.—Transmitter moving relative to an observer.

In 1 second the transmitter moves 360 feet and transmits 60 hertz. At the end of 1 second, the first cycle of the transmitted signal reaches the observer, just as the sixtieth cycle is leaving the transmitter at point P1. Under these conditions the 60 hertz emitted is located between the observer and point P1. Notice that this distance is only 720 feet (1,080 minus 360). The 60 hertz is spread over the distance from point P1 to the observer and has a wavelength of just 12 feet (720 divided by 60). To find the new frequency, use the following formula:

$$\begin{aligned} \text{frequency} &= \frac{\text{velocity}}{(\text{wavelength})} \\ &= \frac{1,080}{12} \\ &= 90 \text{ hertz} \end{aligned}$$

The original frequency, 60 hertz, has changed to an apparent frequency of 90 hertz. This new frequency only applies to the observer. Notice that the Doppler frequency variation is directly proportional to the velocity of the approaching transmitter. The faster the transmitter moves toward the observer, the greater the number of waves that will be crowded into the space between the transmitter and the observer.

Suppose the transmitter were stationary and the observer moving. When approaching the transmitter, the observer would encounter waves per unit of time. As a result, the observer would hear a higher pitch than the transmitter would actually emit.

If the transmitter were traveling away from the observer, as shown in view B of figure 1-20, the first cycle would leave the transmitter at point P and the sixtieth at point P2. The first cycle would reach the observer when the transmitter reached P2. You would then have 60 cycles stretched out over 1,080 plus 360 feet, a total of 1,440 feet. The wavelength of these 60 hertz is 1,440/60, or 24 feet. The apparent frequency is 1,080 divided by 24, or 45 hertz.

Uses of CW Doppler System

The continuous-wave, or Doppler, system is used in several ways. In one radar application, the radar set differentiates between the transmitted and reflected wave to determine the speed of the moving object.

The Doppler method is the best means of detecting fast-moving objects that do not require range resolution. As a moving object approaches the transmitter, it encounters and reflects more waves per unit of time. The amount of frequency shift produced is very small in relation to the carrier frequency. This is because the velocity of propagation of the signal is very high compared to the speed of the target. However, because the carrier frequencies used in radar are high, larger frequency shifts (in the audio-frequency range) are produced. The *amount* of shift is proportional to the *speed* of the reflecting object. One-quarter cycle shift at 10,000 megahertz will provide speed measurements accurate to a fraction of a percent.

If an object is moving, its velocity, relative to the radar, can be detected by comparing the transmitter frequency with the echo frequency (which differs because of the Doppler shift). The DIFFERENCE or BEAT FREQUENCY, sometimes called the DOPPLER FREQUENCY (fd), is related to object velocity.

The separation of the background and the radar contact is based on the Doppler frequency that is caused by the reflection of the signal from a moving object. Disadvantages of the Doppler system are that it does not determine the range of the object, nor is it able to differentiate between objects when they lie in the same direction and are traveling at the same speed. Moreover, it does not "see" stationary or slow-moving objects, which a pulse radar system can detect.

To track an object with cw Doppler, you must determine the radar range. Since the Doppler frequency is not directly related to range, another method is needed to determine object range. By using two separate transmitters that operate at two different frequencies (f_1 and f_2), you can determine range by measuring the relative phase difference between the two Doppler frequencies. In such a system, a mixer is used to combine the two transmitted frequencies and to separate the two received frequencies. This permits the use of one transmitting and receiving antenna.

Instead of using two transmitter frequencies, you can find the range by sweeping the transmitter frequency uniformly in time to cover the frequency range from f_1 to f_2. The beat, or difference, frequency between the transmitted and received signals is then a function of range. In this type of radar, the velocity as well as range is measured.

Q24. The Doppler effect causes a change in what aspect of rf energy that strikes a moving object?

Q25. The Doppler variation is directly proportional to what radar contact characteristic?

Q26. The Doppler method of object detection is best for what type objects?

Q27. The beat frequency in a swept-frequency transmitter provides what contact information?

FREQUENCY-MODULATION METHOD

In the frequency-modulation method, the transmitter radiates radio-frequency waves. The frequency of these rf waves is continually increasing and decreasing from a fixed reference frequency. At any instant, the frequency of the returned signal differs from the frequency of the radiated signal. The amount of the difference frequency is determined by the time it took the signal to travel the distance from the transmitter to the object.

An example of a frequency-modulated signal, plotted against time, is shown in figure 1-21. As shown, the 420-megahertz frequency increases linearly to 460 megahertz and then quickly drops to 420 megahertz again. When the frequency drops to 420 megahertz the frequency cycle starts over again.

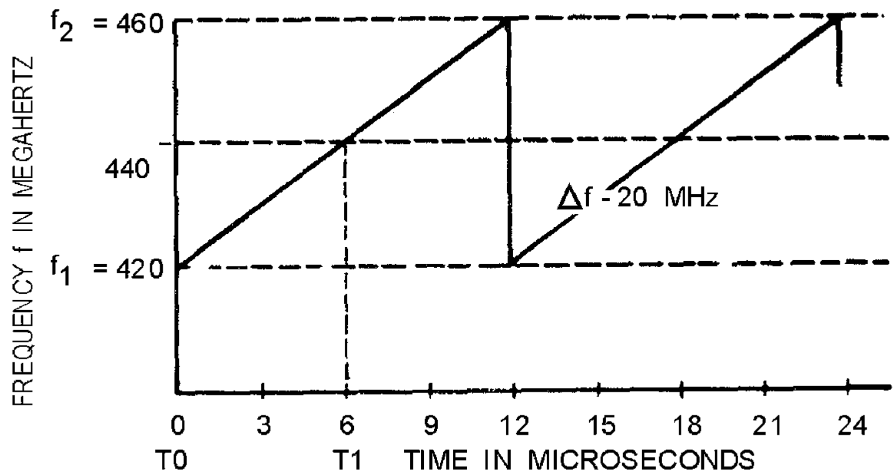


Figure 1-21.—Frequency-modulation chart.

The frequency regularly changes 40 megahertz with respect to time; therefore, its value at any time during its cycle can be used as the basis for computing the time elapsed after the start of the frequency cycle. For example, at T0 the transmitter sends a 420-megahertz signal toward an object. It strikes the object and returns to the receiver at T1, when the transmitter is sending out a new frequency of 440 megahertz. At T1, the 420-megahertz returned signal and the 440-megahertz transmitter signal are fed to the receiver simultaneously. When the two signals are mixed in the receiver, a beat frequency results. The beat frequency varies directly with the distance to the object, increasing as the distance increases. Using this information, you can calibrate a device that measures frequency to indicate range.

This system works well when the detected object is stationary. It is used in aircraft altimeters which give a continuous reading of the height above the earth of the aircraft. The system is not satisfactory for locating moving objects. This is because moving targets produce a frequency shift in the returned signal because of the Doppler effect; this affects the accuracy of the range measurement.

PULSE-MODULATION METHOD

The pulse-modulation method of energy transmission was analyzed to some extent earlier in this chapter. As the previous discussions indicated, radio-frequency energy can also be transmitted in very short bursts, called pulses. These pulses are of extremely short time duration, usually on the order of 0.1 microsecond to approximately 50 microseconds. In this method, the transmitter is turned on for a very short time and the pulse of radio-frequency energy is transmitted, as shown in view A of figure 1-22. The transmitter is then turned off, and the pulse travels outward from the transmitter at the velocity of light (view B). When the pulse strikes an object (view C), it is reflected and begins to travel back toward the radar system, still moving at the same velocity (view D). The pulse is then received by the radar system (view E). The time interval between transmission and reception is computed and converted into a visual indication of range in miles or yards. The radar cycle then starts over again by transmitting another pulse (view F). This method does not depend on the relative frequency of the returned signal or on the motion of the target; therefore, it has an important advantage over cw and fm methods.

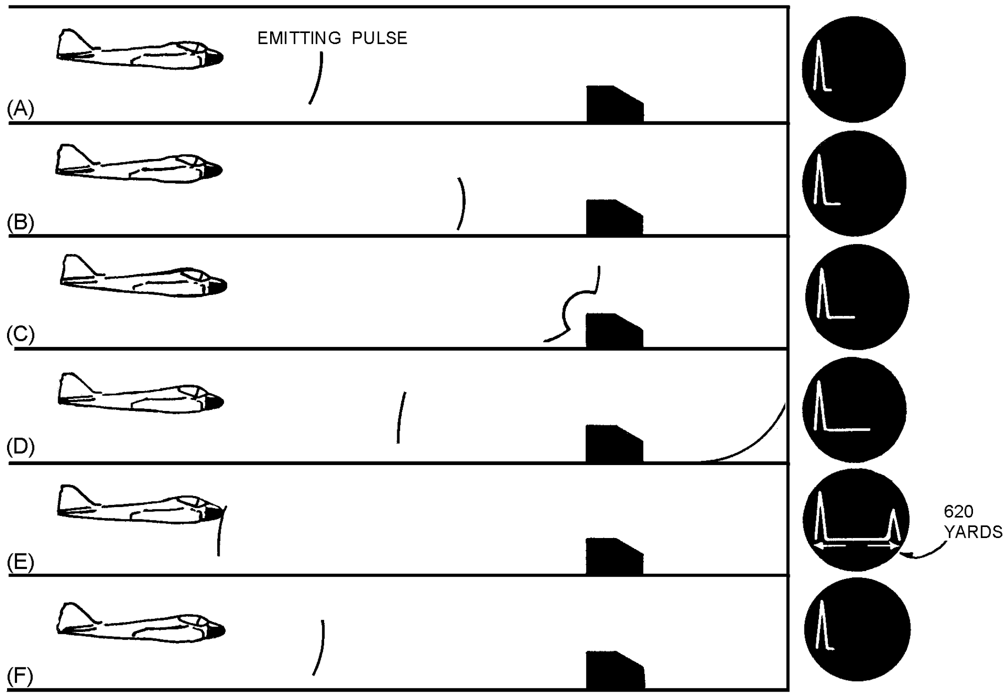


Figure 1-22.—Pulse detection.

PULSE-DOPPLER METHOD

Pulse radar systems may be modified to use the Doppler effect to detect a moving object.

A requirement for any Doppler radar is COHERENCE; that is, some definite phase relationship must exist between the transmitted frequency and the reference frequency, which is used to detect the Doppler shift of the receiver signal. Moving objects are detected by the phase difference between the target signal and background noise components. Phase detection of this type relies on coherence between the transmitter frequency and the receiver reference frequency.

In coherent detection, a stable cw reference oscillator signal, which is locked in phase with the transmitter during each transmitted pulse, is mixed with the echo signal to produce a beat or difference signal. Since the reference oscillator and the transmitter are locked in phase, the echoes are effectively compared with the transmitter in frequency and phase.

The phase relationships of the echoes from fixed objects to the transmitter is constant and the amplitude of the beat signal remains constant. A beat signal of varying amplitude indicates a moving object. This is because the phase difference between the reference oscillator signal and the echo signal changes as the range to the reflecting object changes. The constant amplitude beat signal is filtered out in the receiver. The beat signal of varying amplitude is sent to the radar indicator scope for display.

Q28. What factor determines the difference between the transmitted frequency and the received frequency in an fm transmitter?

Q29. What type of objects are most easily detected by an fm system?

Q30. What transmission method does NOT depend on relative frequency or target motion?

Q31. What transmission method uses a stable cw reference oscillator, which is locked in phase with the transmitter frequency?

RADAR CLASSIFICATION AND USE

Radar systems, like cars, come in a variety of sizes and have different performance specifications. Some radar systems are used for air-traffic control at airports and others are used for long-range surveillance and early-warning systems. A radar system is the heart of a missile guidance system. Small portable radar systems that can be maintained and operated by one person are available as well as systems that occupy several large rooms.

MILITARY CLASSIFICATION OF RADAR SYSTEMS

The large number of radar systems used by the military has forced the development of a joint-services classification system for accurate identification. The Federal Aviation Agency (FAA) also makes extensive use of radar systems for commercial aircraft in-flight and landing control, but does not use the military classification system.

Radar systems are usually classified according to specific function and installation vehicle. Some common examples are listed below:

FUNCTION	INSTALLATION VEHICLE
Search	Ground or land based
Track	Airborne
Height-finder	Shipboard

The joint-service standardized classification system further divides these broad categories for more precise identification. Table 1-1 is a listing of equipment identification indicators. Use of the table to identify a particular radar system is illustrated in figure 1-23. Note that for simplicity, only a portion of the table has been used in the illustration.

Table 1-1.—Table of Equipment Indicators

TABLE OF EQUIPMENT INDICATORS			Miscellaneous Identification
Installation (1st letter)	Type of Equipment (2d letter)	Purpose (3rd letter)	
A—Piloted aircraft	A—invisible light, heat radiation	B—Bombing	X, Y, Z—Changes in voltage, phase, or frequency
B—Underwater mobile, submarine	C—Carrier	C—Communications (receiving and transmitting	T—Training
D—Pilotless carrier	D—Radiac	D—Direction finder reconnaissance and/or surveillance	(V)—Variable grouping
F—Fixed ground	G—Telegraph or Teletype	E—Ejection and/or release	
G—General ground use	I—Interphone and public address	G—Fire control, or searchlight directing	
K—Amphibious	J—Electromechanical or Inertial wire covered	H—Recording and/or reproducing (graphic meteorological and sound)	
M—Ground, mobile	K—Telemetering	K—Computing	
P—Portable	L—Countermeasures	M—Maintenance and/or test assemblies (including tools)	
S—Water	M—Meteorological	N—Navigational aids (including altimeters, beacons, compasses, racons, depth sounding, approach and landing)	
T—Ground, transportable	N—Sound in air	Q—Special, or combination of purposes	
U—General utility	P—Radar	R—Receiving, passive detecting	
V—Ground, vehicular	Q—Sonar and underwater sound	S—Detecting and/or range and bearing, search	
W—Water surface and under water combination	R—Radio	T—Transmitting	
Z—Piloted and pilotless airborne vehicle combination	S—Special types, magnetic, etc., or combinations of types	W—Automatic flight or remote control	
	T—Telephone (wire)	X—Identification and recognition	
	V—Visual and visible light	Y—Surveillance (search detect, and multiple target tracking) and control (both fire control and air control)	
	W—Armament (peculiar to armament, not otherwise covered)		
	X—Facsimile or television	X—Facsimile or television	
	Y—Data processing		

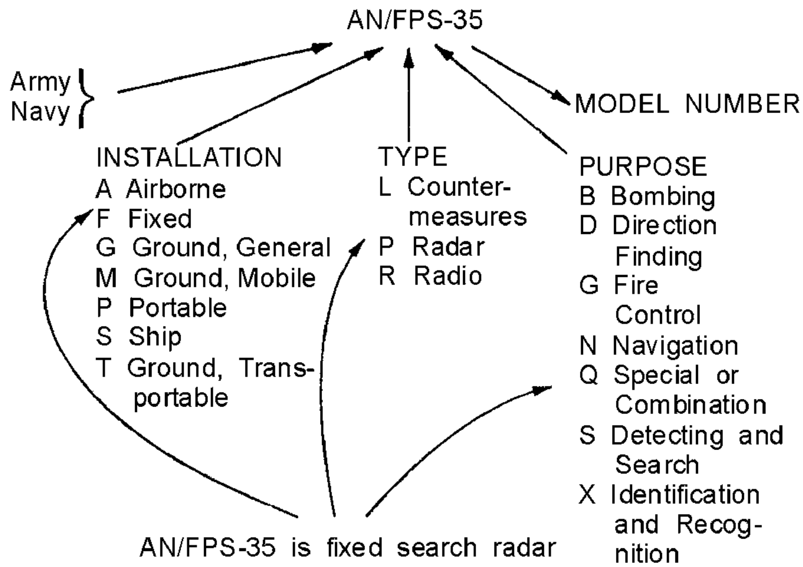


Figure 1-23.—Joint service classification system.

RADAR FUNCTIONS

No single radar system has yet been designed that can perform all of the many radar functions required by the military. Some of the newer systems combine several functions that formerly required individual radar systems, but no single system can fulfill all the requirements of modern warfare. As a result, modern warships, aircraft, and shore stations usually have several radar systems, each performing a different function.

One radar system, called SEARCH RADAR, is designed to continuously scan a volume of space to provide initial detection of all targets. Search radar is almost always used to detect and determine the position of new targets for later use by TRACK RADAR. Track radar provides continuous range, bearing, and elevation data on one or more targets. Most of the radar systems used by the military are in one of these two categories, though the individual radar systems vary in design and capability.

Some radar systems are designed for specific functions that do not precisely fit into either of the above categories. The radar speed gun is an example of radar designed specifically to measure the speed of a target. The military uses much more complex radar systems that are adapted to detect only fast-moving targets such as aircraft. Since aircraft usually move much faster than weather or surface targets, velocity-sensitive radar can eliminate unwanted clutter from the radar indicator. Radar systems that detect and process only moving targets are called MOVING-TARGET INDICATORS (mti) and are usually combined with conventional search radar.

Another form of radar widely used in military and civilian aircraft is the RADAR ALTIMETER. Just as some surface-based radars can determine the height of a target, airborne radar can determine the distance from an aircraft to the ground. Many aircraft use radar to determine height above the ground. Radar altimeters usually use frequency-modulated signals of the type discussed earlier in the chapter.

RADAR TYPES

The preceding paragraphs indicated that radar systems are divided into types based on the designed use. This section presents the general characteristics of several commonly used radar systems. Typical characteristics are discussed rather than the specific characteristics of any particular radar system.

SEARCH RADAR

Search radar, as previously mentioned, continuously scans a volume of space and provides initial detection of all targets within that space. Search radar systems are further divided into specific types, according to the type of object they are designed to detect. For example, surface-search, air-search, and height-finding radars are all types of search radar.

Surface-Search Radar

A surface-search radar system has two primary functions: (1) the detection and determination of accurate ranges and bearings of surface objects and low-flying aircraft and (2) the maintenance of a 360-degree search pattern for all objects within line-of-sight distance from the radar antenna.

The maximum range ability of surface-search radar is primarily limited by the radar horizon; therefore, higher frequencies are used to permit maximum reflection from small, reflecting areas, such as ship masthead structures and the periscopes of submarines. Narrow pulse widths are used to permit a high degree of range resolution at short ranges and to achieve greater range accuracy. High pulse-repetition rates are used to permit a maximum definition of detected objects. Medium peak power can be used to permit the detection of small objects at line-of-sight distances. Wide vertical-beam widths permit compensation for the pitch and roll of own ship and detection of low flying aircraft. Narrow horizontal-beam widths permit accurate bearing determination and good bearing resolution. For example, a common shipboard surface-search radar has the following design specifications:

- Transmitter frequency 5,450-5,825 MHz
- Pulse width .25 or 1.3 microseconds
- Pulse-repetition rate between 625 and 650 pulses per second
- Peak power between 190 and 285 kW
- Vertical beam width between 12 and 16 degrees
- Horizontal beam width 1.5 degrees

Surface-search radar is used to detect the presence of surface craft and low flying aircraft and to determine their presence. Shipboard surface-search radar provides this type of information as an input to the weapons system to assist in the engagement of hostile targets by fire-control radar. Shipboard surface-search radar is also used extensively as a navigational aid in coastal waters and in poor weather conditions. A typical surface-search radar antenna is shown in figure 1-24.

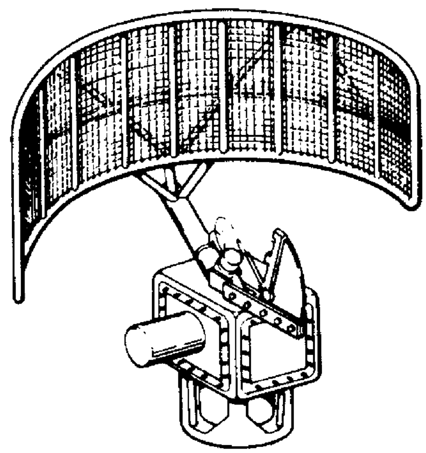


Figure 1-24.—Surface-search radar.

Q32. What type of radar provides continuous range, bearing, and elevation data on an object?

Q33. Radar altimeters use what type of transmission signal?

Q34. A surface-search radar normally scans how many degrees of azimuth?

Q35. What limits the maximum range of a surface-search radar?

Q36. What is the shape of the beam of a surface-search radar?

Air-Search Radar

Air-search radar systems initially detect and determine the position, course, and speed of air targets in a relatively large area. The maximum range of air-search radar can exceed 300 miles, and the bearing coverage is a complete 360-degree circle. Air-search radar systems are usually divided into two categories, based on the amount of position information supplied. As mentioned earlier in this chapter, radar sets that provide only range and bearing information are referred to as two-dimensional, or 2D, radars. Radar sets that supply range, bearing, and height are called three-dimensional, or 3D, radars. (3D radar will be covered in the next section.) The coverage pattern of a typical 2D radar system is illustrated in figure 1-25. A typical 2D air-search radar antenna is shown in figure 1-26.

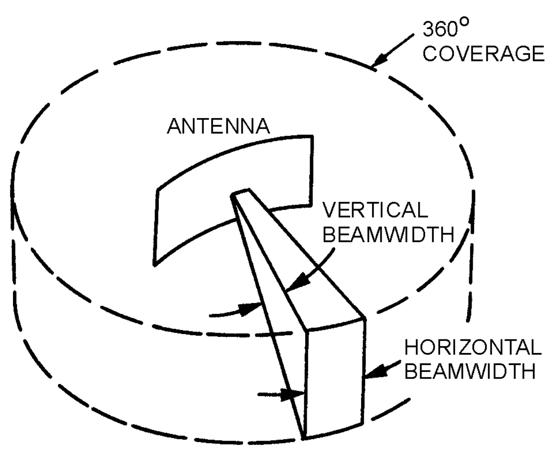


Figure 1-25.—2D radar coverage pattern.

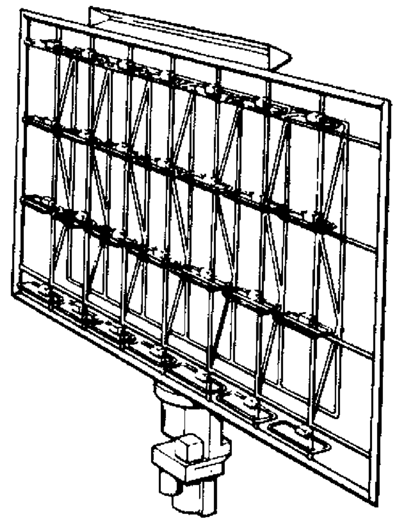


Figure 1-26.—2D air-search radar.

Relatively low transmitter frequencies are used in 2D search radars to permit long-range transmissions with minimum attenuation. Wide pulse widths and high peak power are used to aid in detecting small objects at great distances. Low pulse-repetition rates are selected to permit greater maximum range. A wide vertical-beam width is used to ensure detection of objects from the surface to relatively high altitudes and to compensate for pitch and roll of own ship. The output characteristics of specific air-search radars are classified; therefore, they will not be discussed.

Air-search radar systems are used as early-warning devices because they can detect approaching enemy aircraft or missiles at great distances. In hostile situations, early detection of the enemy is vital to a successful defense against attack. Antiaircraft defenses in the form of shipboard guns, missiles, or fighter planes must be brought to a high degree of readiness in time to repel an attack. Range and bearing information, provided by air-search radars, used to initially position a fire-control tracking radar on a target. Another function of the air-search radar system is guiding combat air patrol (CAP) aircraft to a position suitable to intercept an enemy aircraft. In the case of aircraft control, the guidance information is obtained by the radar operator and passed to the aircraft by either voice radio or a computer link to the aircraft.

Height-Finding Search Radar

The primary function of a height-finding radar (sometimes referred to as a three-coordinate or 3D radar) is that of computing accurate ranges, bearings, and altitudes of aircraft targets detected by air-search radars. Height-finding radar is also used by the ship's air controllers to direct CAP aircraft during interception of air targets. Modern 3D radar is often used as the primary air-search radar (figure 1-27). This is because of its high accuracy and because the maximum ranges are only slightly less than those available from 2D radar.

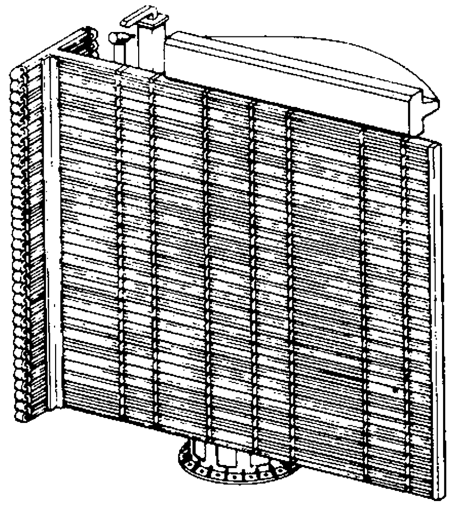


Figure 1-27.—3D air-search radar.

The range capability of 3D search radar is limited to some extent by an operating frequency that is higher than that of 2D radar. This disadvantage is partially offset by higher output power and a beam width that is narrower in both the vertical and horizontal planes.

The 3D radar system transmits several narrow beams to obtain altitude coverage and, for this reason, compensation for roll and pitch must be provided for shipboard installations to ensure accurate height information.

Applications of height-finding radars include the following:

- Obtaining range, bearing, and altitude data on enemy aircraft and missiles to assist in the control of CAP aircraft
- Detecting low-flying aircraft
- Determining range to distant land masses
- Tracking aircraft over land
- Detecting certain weather phenomena
- Tracking weather balloons
- Providing precise range, bearing, and height information for fast, accurate initial positioning of fire-control tracking radars

Q37. Air-search radar is divided into what two basic categories?

Q38. What position data are supplied by 2D search radar?

Q39. Why do 2D air-search radars use relatively low carrier frequencies and low pulse-repetition rates?

Q40. Why is the range capability of 3D radar usually less than the range of 2D radar?

TRACKING RADAR

Radar that provides continuous positional data on a target is called tracking radar. Most tracking radar systems used by the military are also fire-control radar; the two names are often used interchangeably.

Fire-control tracking radar systems usually produce a very narrow, circular beam.

Fire-control radar must be directed to the general location of the desired target because of the narrow-beam pattern. This is called the DESIGNATION phase of equipment operation. Once in the general vicinity of the target, the radar system switches to the ACQUISITION phase of operation. During acquisition, the radar system searches a small volume of space in a prearranged pattern until the target is located. When the target is located, the radar system enters the TRACK phase of operation. Using one of several possible scanning techniques, the radar system automatically follows all target motions. The radar system is said to be *locked on* to the target during the track phase. The three sequential phases of operation are often referred to as MODES and are common to the target-processing sequence of most fire-control radars.

Typical fire-control radar characteristics include a very high prf, a very narrow pulse width, and a very narrow beam width. These characteristics, while providing extreme accuracy, limit the range and make initial target detection difficult. A typical fire-control radar antenna is shown in figure 1-28. In this example the antenna used to produce a narrow beam is covered by a protective radome.

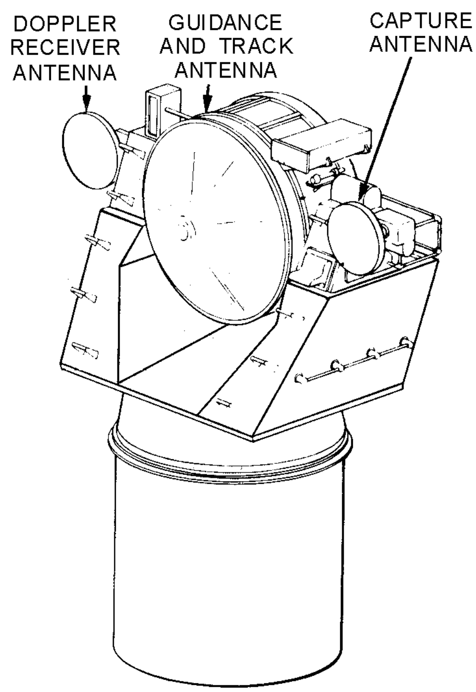


Figure 1-28.—Fire-control radar.

MISSILE-GUIDANCE RADAR

A radar system that provides information used to guide a missile to a hostile target is called GUIDANCE RADAR. Missiles use radar to intercept targets in three basic ways: (1) Beam-rider missiles follow a beam of radar energy that is kept continuously pointed at the desired target; (2) homing missiles detect and home in on radar energy reflected from the target; the reflected energy is provided by a radar transmitter either in the missile or at the launch point and is detected by a receiver in the missile; (3) passive homing missiles home in on energy that is radiated by the target. Because target position must be known at all times, a guidance radar is generally part of, or associated with, a fire-control tracking radar. In some instances, three radar beams are required to provide complete guidance for a missile. The beam-riding missile, for example, must be launched into the beam and then must ride the beam to the target. Initially, a wide beam is radiated by a capture radar to gain (capture) control of the missile. After the missile enters the capture beam, a narrow beam is radiated by a guidance radar to guide the missile to the target. During both capture and guidance operations, a tracking radar continues to track the target. Figure 1-29 illustrates the relationships of the three different radar beams.

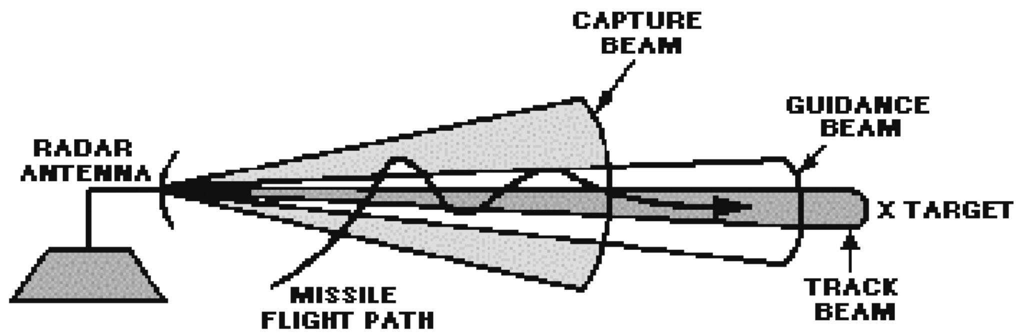


Figure 1-29.—Beam relationship of capture, guidance, and track beams.

Q41. Fire-control tracking radar most often radiates what type of beam?

Q42. Tracking radar searches a small volume of space during which phase of operation?

Q43. What width is the pulse radiated by fire-control tracking radar?

Q44. Which beam of missile-guidance radar is very wide?

CARRIER-CONTROLLED APPROACH (CCA) AND GROUND-CONTROLLED APPROACH (GCA) RADAR

CARRIER-CONTROLLED APPROACH and GROUND-CONTROLLED APPROACH radar systems are essentially shipboard and land-based versions of the same type of radar. Shipboard CCA radar systems are usually much more sophisticated systems than GCA systems. This is because of the movements of the ship and the more complicated landing problems. Both systems, however, guide aircraft to safe landing under conditions approaching zero visibility. By means of radar, aircraft are detected and observed during the final approach and landing sequence. Guidance information is supplied to the pilot in the form of verbal radio instructions, or to the automatic pilot (autopilot) in the form of pulsed control signals.

AIRBORNE RADAR

Airborne radar is designed especially to meet the strict space and weight limitations that are necessary for all airborne equipment. Even so, airborne radar sets develop the same peak power as shipboard and shore-based sets.

As with shipboard radar, airborne radar sets come in many models and types to serve many different purposes. Some of the sets are mounted in blisters (or domes) that form part of the fuselage; others are mounted in the nose of the aircraft.

In fighter aircraft, the primary mission of a radar is to aid in the search, interception, and destruction of enemy aircraft. This requires that the radar system have a tracking feature. Airborne radar also has many other purposes. The following are some of the general classifications of airborne radar: search, intercept and missile control, bombing, navigation, and airborne early warning.

SUMMARY

The following paragraphs summarize the important points of this chapter.

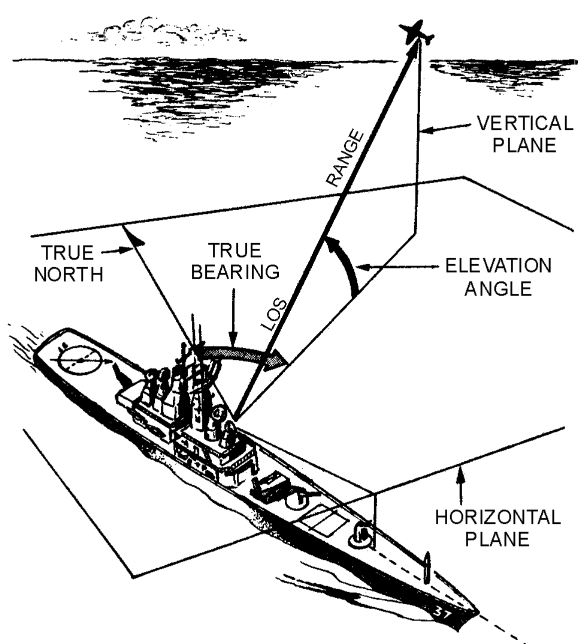
RADAR is an electronic system that uses reflected electromagnetic energy to detect the presence and position of objects invisible to the eye.

TARGET POSITION is defined in reference to true north, the horizontal plane, and the vertical plane.

TRUE BEARING is the angle between true north and the line of sight to the target, measured in a clockwise direction in the horizontal plane.

ELEVATION ANGLE is the angle between the horizontal plane and the line of sight, measured in the vertical plane.

RANGE is the distance from the radar site to the target measured along the line of sight. The concepts are illustrated in the figure.



RANGE to any target can be calculated by measuring the time required for a pulse to travel to a target and return to the radar receiver and by dividing the elapsed time by 12.36 microseconds.

$$\text{target range} = \frac{\text{elapsed time}}{\text{12.36 microseconds per nautical mile}}$$

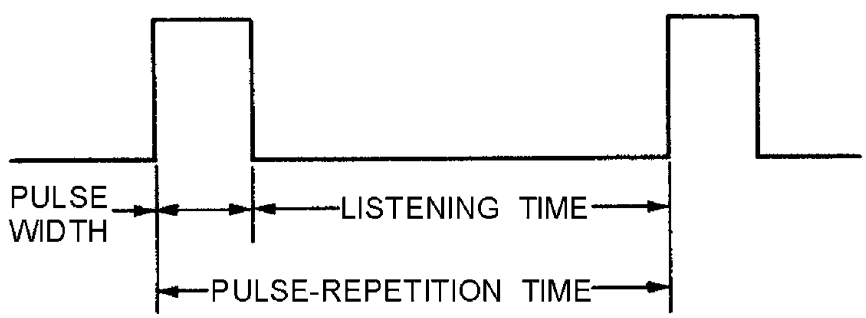
The **MINIMUM RANGE** of a radar system can be calculated from the formula:

$$\text{minimum range} = (\text{pulse width} + \text{recovery time}) \times 164 \text{ yards/microsecond}$$

The **MAXIMUM RANGE** of a pulse radar system depends on the CARRIER FREQUENCY, PEAK POWER, PULSE-REPETITION FREQUENCY, and RECEIVER SENSITIVITY.

PULSE-REPETITION TIME is the time between the beginning of one pulse and the beginning of the next pulse and is the reciprocal of prf.

$$\text{prt} = \frac{1}{\text{prf}}$$



AMBIGUOUS RETURNS are echoes from targets that exceed the prt of the radar system and result in false range readings. The maximum (unambiguous) range for a radar system can be determined by the formula:

$$R_{max} = \frac{\text{162, 000 mile/second}}{2} \times \text{prt}$$

The **PEAK POWER** of a radar system is the total energy contained in a pulse. Peak power is obtained by multiplying the maximum power level of a pulse by the pulse width.

Since most instruments are designed to measure AVERAGE POWER over a period of time, prt must be included in transmitter power measurements. The formula for average power is:

$$P_{avg} = P_{pk} \times \frac{pw}{prt}$$

or

$$P_{avg} = P_{pk} \times pw \times prf$$

The product of pw and prf is called the DUTY CYCLE of a radar system and is the ratio of transmitter time on to time off.

The formula for the peak power (using average power) of a radar system is:

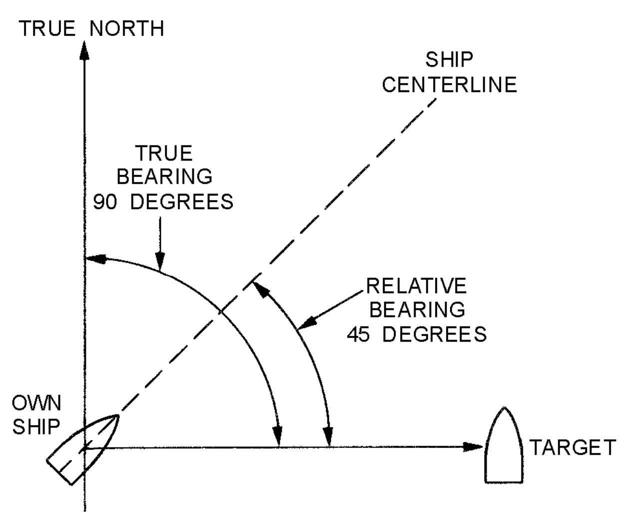
$$P_{pk} = \frac{P_{avg}}{\text{duty cycle}}$$

Antenna height and ROTATION SPEED affect radar range. Since high-frequency energy does not normally bend to follow the curvature of the earth, most radar systems cannot detect targets below the RADAR HORIZON. The distance to the horizon for a radar system can be determined by the formula:

$$\text{radar horizon distance} = 1.25\sqrt{\text{antenna height in feet}}$$
$$\text{(in nautical miles)}$$

The slower an antenna rotates, the larger the HITS PER SCAN value. The likelihood that a target will produce a usable echo is also increased.

The bearing to a target may be referenced to true north or to your own ship. Bearing referenced to true north is TRUE BEARING and bearing referenced to your ship is RELATIVE BEARING, as shown in the illustration. The bearing angle is obtained by moving the antenna to the point of maximum signal return.



Radar systems that detect only range and bearing are called TWO-DIMENSIONAL (2D) radars. Radars that detect height as well as range and bearing are called THREE-DIMENSIONAL (3D) RADARS.

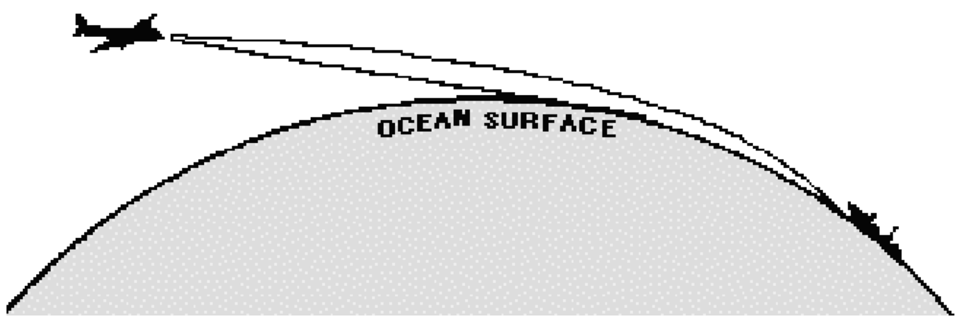
The target RESOLUTION of a radar system is its ability to distinguish between targets that are very close together.

RANGE RESOLUTION is the ability to distinguish between two or more targets on the same bearing and is primarily dependent on the pulse width of the radar system. The formula for range resolution is:

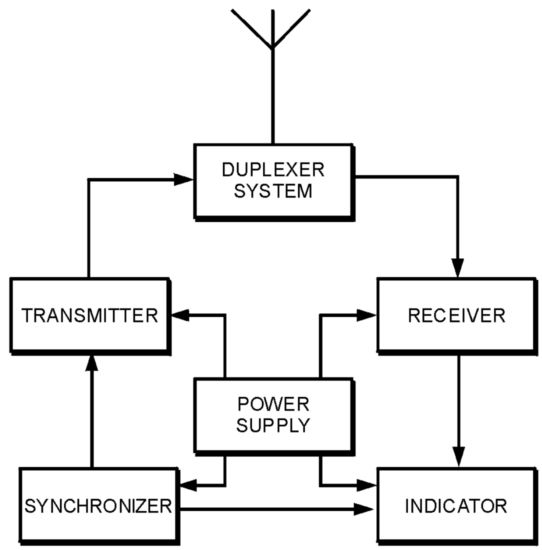
$$\text{resolution} = \text{pw} \times 164 \text{ yards per microsecond}$$

BEARING RESOLUTION is the ability of a radar to separate targets at the same range but different bearings. The degree of bearing resolution is dependent on beam width and range. The accuracy of radar is largely dependent on resolution.

ATMOSPHERIC CONDITIONS affect the speed and direction of travel of electromagnetic wavefronts traveling through the air. Under normal conditions, the wavefronts increase uniformly in speed as altitude increases which causes the travel path to curve downward. The downward curve extends the radar horizon as shown in the illustration. The density of the atmosphere, the presence of water vapor, and temperature changes also directly affect the travel of electromagnetic wavefronts.



The major components in a typical PULSE RADAR SYSTEM are shown in the illustration. The SYNCHRONIZER supplies the timing signals to coordinate the operation of the entire system. The TRANSMITTER generates electromagnetic energy in short, powerful pulses. The DUPLEXER allows the same antenna to be used to both transmit and receive. The RECEIVER detects and amplifies the return signals. The INDICATOR produces a visual indication of the range and bearing of the echo.



SCANNING is the systematic movement of a radar beam while searching for or tracking a target.

STATIONARY-LOBE SCANNING is the simplest type of scanning and is usually used in 2D search radar. Monopulse scanning, used in fire-control radars, employs four signal quantities to accurately track moving targets. The two basic methods of scanning are MECHANICAL and ELECTRONIC.

Radar systems are often divided into operational categories based on energy transmission methods—continuous wave (cw), frequency modulation (fm), and pulse modulation (pm).

The **CONTINUOUS WAVE (cw)** method transmits a constant frequency and detects moving targets by detecting the change in frequency caused by electromagnetic energy reflecting from a moving target. This change in frequency is called the DOPPLER SHIFT or DOPPLER EFFECT.

In the **FREQUENCY MODULATION (fm)** method, a signal that constantly changes in frequency around a fixed reference is used to detect stationary objects.

The **PULSE-MODULATION (pm) METHOD** uses short pulses of energy and relatively long listening times to accurately determine target range. Since this method does not depend on signal frequency or target motion, it has an advantage over cw and fm methods. It is the most common type of radar.

Radar systems are also classified by function. SEARCH RADAR continuously scans a volume of space and provides initial detection of all targets. TRACK RADAR provides continuous range, bearing, and elevation data on one or more specific targets. Most radar systems are variations of these two types.

ANSWERS TO QUESTIONS Q1. AND Q44.

A1. Horizontal plane.

A2. Range.

A3. Approximately the speed of light (162,000 nautical miles per second).

A4. 12.36 microseconds.

A5. Pulse width.

A6. Frequency.

A7.

$$\frac{1}{\mathrm{prt}} = \mathrm{prf}$$

A8. Average power.

A9. Duty cycle.

A10. Relative bearing.

A11. Three-dimensional.

A12. Frequency or phase.

A13. Target resolution.

A14. Beam width and range.

A15. Speed increases.

A16. Temperature inversion.

A17. Synchronizer.

A18. High-voltage pulse from the modulator.

A19. Duplexer.

A20. Single lobe.

A21. The reflected signals decrease in strength.

A22. Mechanical and electronic.

A23. Nutation.

A24. Frequency.

A25. Velocity.

A26. Fast-moving targets.

A27. Range.

A28. Travel time.

A29. Stationary.

A30. Pulse modulation.

A31. Pulse-Doppler.

A32. Track radar.

A33. Frequency modulated (fm).

A34. 360 degrees.

A35. Radar horizon.

A36. Wide vertically, narrow horizontally.

A37. 2D and 3D.

A38. Range and bearing.

A39. Increased maximum range.

A40. Higher operating frequency.

A41. A narrow circular beam.

A42. Acquisition.

A43. Very narrow.

A44. Capture beam.

CHAPTER 2

RADAR SUBSYSTEMS

LEARNING OBJECTIVES

Upon completion of this chapter, the student will be able to:

1. Describe, in general terms, the function of a radar synchronizer.
2. State the basic requirements and types of master synchronizers.
3. Describe the purpose, requirements, and operation of a radar modulator.
4. Describe the basic operating sequence of a keyed-oscillator transmitter.
5. Describe the basic operating sequence of a power-amplifier transmitter.
6. State the purpose of a duplexer.
7. State the operational principles of tr and atr tubes.
8. Describe the basic operating sequence of series and parallel connected duplexers.
9. List the basic design requirements of an effective radar receiver.
10. List the major sections of a typical radar receiver.
11. Using a block diagram, describe the operational characteristics of a typical radar receiver.

INTRODUCTION TO RADAR SUBSYSTEMS

Any radar system has several major subsystems that perform standard functions. A typical radar system consists of a SYNCHRONIZER (also called the TIMER or TRIGGER GENERATOR), a TRANSMITTER, a DUPLEXER, a RECEIVER, and an INDICATOR. These major subsystems were briefly described in chapter 1. This chapter will describe the operation of the synchronizer, transmitter, duplexer, and receiver of a typical pulse radar system and briefly analyze the circuits used. Chapter 3 will describe typical indicator and antenna subsystems. Because radar systems vary widely in specific design, only a general description of representative circuits is presented in this chapter.

SYNCHRONIZERS

The synchronizer is often referred to as the "heart" of the radar system because it controls and provides timing for the operation of the entire system. Other names for the synchronizer are the TIMER and the KEYER. We will use the term synchronizer in our discussion. In some complex systems the synchronizer is part of a system computer that performs many functions other than system timing.

SYNCHRONIZER FUNCTION

The specific function of the synchronizer is to produce TRIGGER PULSES that start the transmitter, indicator sweep circuits, and ranging circuits.

Timing or control is the function of the majority of circuits in radar. Circuits in a radar set accomplish control and timing functions by producing a variety of voltage waveforms, such as square waves, sawtooth waves, trapezoidal waves, rectangular waves, brief rectangular pulses, and sharp peaks. Although all of these circuits can be broadly classified as timing circuits, the specific function of any individual circuit could also be wave shaping or wave generation. The operation of many of these circuits and associated terms were described in detail in NEETS, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*.

Q1. What is the purpose of the synchronizer in a radar system?

Q2. What is the purpose of the majority of circuits in a radar system?

SYNCHRONIZER OPERATION

Radar systems may be classified as either SELF-SYNCHRONIZED or EXTERNALLY SYNCHRONIZED systems. In a self-synchronized system, the timing trigger pulses are generated in the transmitter. In an externally synchronized system, the timing trigger pulses are generated by a MASTER OSCILLATOR, which is usually external to the transmitter.

The master oscillator in an externally synchronized system may be a BLOCKING OSCILLATOR, a SINE-WAVE OSCILLATOR, or an ASTABLE (FREE-RUNNING) MULTI-VIBRATOR. When a blocking oscillator is used as a master oscillator, the timing trigger pulses are usually obtained directly from the oscillator. When a sine-wave oscillator or an astable multivibrator is used as a master oscillator, pulse-shaping circuits are required to form the necessary timing trigger pulses. In an externally synchronized radar system, the pulse repetition rate (prr) of the timing trigger pulses from the master oscillator determines the prr of the transmitter.

In a self-synchronized radar system, the prr of the timing trigger pulses is determined by the prr of the modulator or transmitter.

Associated with every radar system is an indicator, such as a cathode-ray tube, and associated circuitry. The indicator can present range, bearing, and elevation data in visual form so that a detected object may be located. Trigger pulses from the synchronizer are frequently used to produce gate (or enabling) pulses. When applied to the indicator, gate pulses perform the following functions:

1. Initiate and time the duration of the indicator sweep voltage

2. Intensify the cathode-ray tube electron beam during the sweep period so that the echo pulses may be displayed

3. Gate a range marker generator so that range marker signals may be superimposed on the indicator presentation

Figure 2-1 shows the time relationships of the various waveforms in a typical radar set. The timing trigger pulses are applied to both the transmitter and the indicator. When a trigger pulse is applied to the transmitter, a short burst of transmitter pulses (rf energy) is generated.

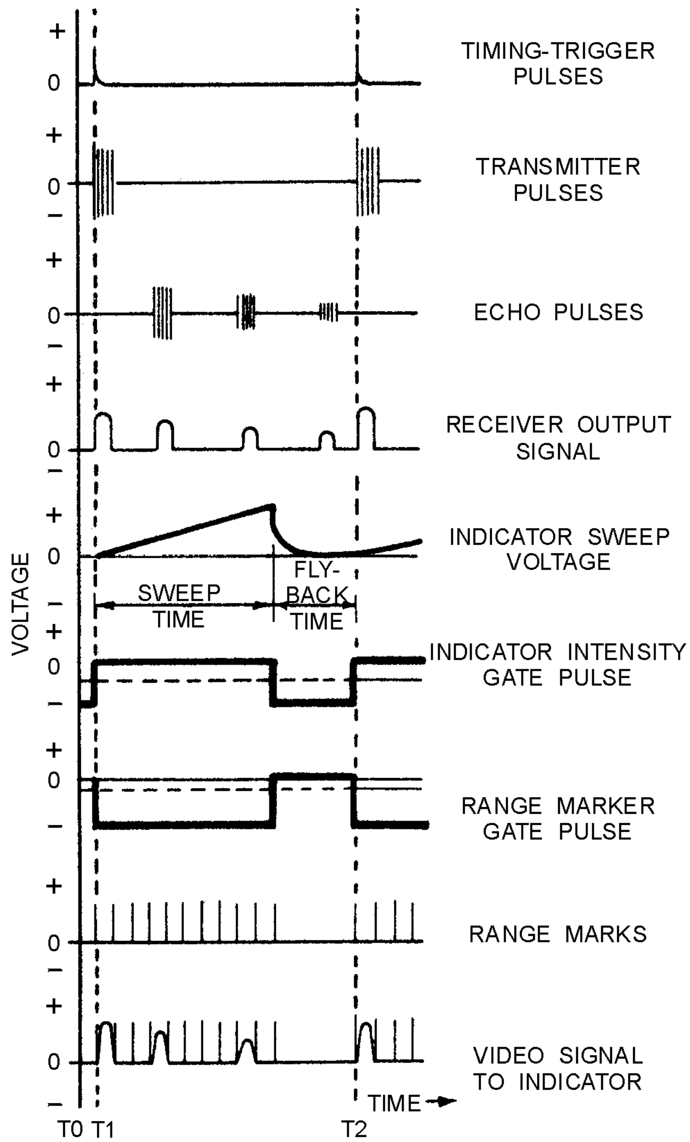


Figure 2-1.—Time relationship of waveforms.

This energy is conducted along a transmission line to the radar antenna. It is radiated by the antenna into space. When this transmitter energy strikes one or more reflecting objects in its path, some of the transmitted energy is reflected back to the antenna as echo pulses. Echo pulses from three reflecting targets at different ranges are illustrated in figure 2-1. These echoes are converted to the corresponding receiver output signals as shown in the figure. The larger initial and final pulses in the receiver output signal are caused by the energy that leaks through the duplexer when a pulse is being transmitted.

The indicator sweep voltage shown in figure 2-1 is initiated at the same time the transmitter is triggered. In other applications, it may be more desirable to delay the timing trigger pulse that is to be fed to the indicator sweep circuit. Delaying the trigger pulse will initiate the indicator sweep *after* a pulse is transmitted.

Note in figure 2-1 that the positive portion of the indicator intensity gate pulse (applied to the cathode-ray tube control grid) occurs only during the indicator sweep time. As a result, the visible cathode-ray tube trace occurs only during sweep time and is eliminated during the flyback (retrace) time. The negative portion of the range-marker gate pulse also occurs during the indicator sweep time. This negative gate pulse is applied to a range-marker generator, which produces a series of range marks.

The range marks are equally spaced and are produced only for the duration of the range-marker gate pulse. When the range marks are combined (mixed) with the receiver output signal, the resulting video signal applied to the indicator may appear as shown at the bottom of figure 2-1.

Q3. A self-synchronized radar system obtains timing trigger pulses from what source?

Q4. What type of multivibrator can be used as a radar master oscillator?

Q5. In an externally synchronized radar, what determines the prr of the transmitter?

Q6. In figure 2-1, what causes the initial and final pulses on the receiver output signal?

BASIC SYNCHRONIZER CIRCUITS

The basic synchronizer circuit should meet the following three basic requirements:

1. It must be *free running* (astable). Because the synchronizer is the heart of the radar, it must establish the zero time reference and the prf (prr).

2. It should be *stable in frequency*. For accurate ranging, the prr and its reciprocal, pulse-repetition time (prt), must not change between pulses.

3. The *frequency must be variable* to enable the radar to operate at different ranges.

Three basic synchronizer circuits can meet the above mentioned requirements. They are the SINE-WAVE OSCILLATOR, the SINGLE-SWING BLOCKING OSCILLATOR, and the MASTER-TRIGGER (ASTABLE) MULTIVIBRATOR.

Figure 2-2 shows the block diagrams and waveforms of these three synchronizers as they are used in externally synchronized radar systems. In each case, equally spaced timing trigger pulses are produced. The prr of each series of timing trigger pulses is determined by the operating frequency of the associated master oscillator.

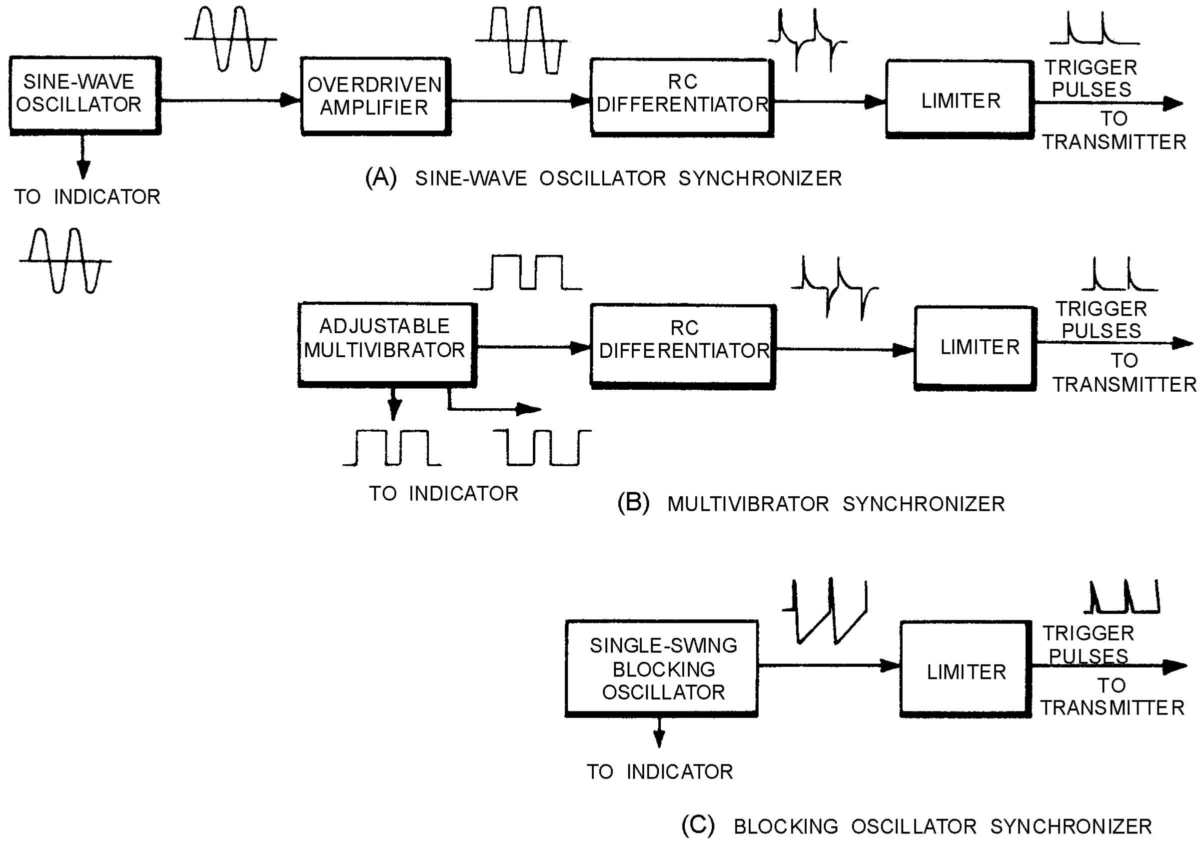


Figure 2-2.—Timers used in externally synchronized radar systems.

Sine-Wave Oscillator Synchronizer

In the sine-wave oscillator synchronizer (figure 2-2, view A), a sine-wave oscillator is used for the basic timing device (master oscillator). The oscillator output is applied to both an overdriven amplifier and the radar indicator. The sine waves applied to the overdriven amplifier are shaped into square waves. These square waves are then converted into positive and negative timing trigger pulses by means of a short-time-constant RC differentiator.

By means of a limiter, either the positive or negative trigger pulses from the RC differentiator are removed. This leaves trigger pulses of only one polarity. For example, the limiter in view A of figure 2-2 is a negative-lobe limiter; that is, the limiter removes the negative trigger pulses and passes only positive trigger pulses to the radar transmitter.

A disadvantage of a sine-wave oscillator synchronizer is the large number of pulse-shaping circuits required to produce the necessary timing trigger pulses.

Master Trigger (Astable) Multivibrator Synchronizer

In a master trigger (astable) multivibrator synchronizer (view B, figure 2-2), the master oscillator generally is an astable multivibrator. The multivibrator is either ASYMMETRICAL or SYMMETRICAL. If the multivibrator is asymmetrical, it generates rectangular waves. If the multivibrator is symmetrical, it generates square waves. In either case, the timing trigger pulses are equally spaced after a limiter removes undesired positive or negative lobes.

There are two transistors in an astable multivibrator. The two output voltages are equal in amplitude, but are 180 degrees out of phase. The output of the astable multivibrator consists of positive and negative rectangular waves. Positive rectangular waves are applied to an RC differentiator and converted into positive and negative trigger pulses. As in the sine-wave synchronizer, the negative trigger pulses are removed by means of a negative-lobe limiter, and the positive pulses are applied to the transmitter.

Both positive and negative rectangular waves from the astable multivibrator are applied to the indicator. One set of waves is used to intensify the cathode-ray tube electron beam for the duration of the sweep. The other set of waves is used to gate (turn on) the range marker generator.

Single-Swing Blocking Oscillator Synchronizer

In the single-swing, blocking-oscillator synchronizer, shown in view C of figure 2-2, a free-running, single-swing blocking oscillator is generally used as the master oscillator. The advantage of the single-swing blocking oscillator is that it generates sharp trigger pulses without additional shaping circuitry. Timing trigger pulses of only one polarity are obtained by means of a limiter.

Gating pulses for the indicator circuits are produced by applying the output of the blocking oscillator to a one-shot multivibrator or another variable time delay circuit (not shown). Crystal-controlled oscillators may be used when very stable frequency operation is required.

Q7. What basic circuits meet the requirements of an externally synchronized master oscillator?

Q8. Name a disadvantage of sine-wave oscillator synchronizers.

Q9. Which of the basic timing circuits produces sharp trigger pulses directly?

TRANSMITTERS

The TRANSMITTER produces the short duration high-power rf pulses of energy that are radiated into space by the antenna. Two main types of transmitters are now in common use. The first is the KEYED-OSCILLATOR type. In this transmitter one stage or tube, usually a magnetron, produces the rf pulse. The oscillator tube is keyed by a high-power dc pulse of energy generated by a separate unit called the MODULATOR (discussed in the following section). The second type of transmitter consists of a POWER-AMPLIFIER CHAIN. This transmitter system begins with an rf pulse of very low power. This low-level pulse is then amplified by a series (chain) of power amplifiers to the high level of power desired in a transmitter pulse. In most power-amplifier transmitters, each of the power-amplifier stages is pulse modulated in a manner similar to the oscillator in the keyed-oscillator type. Because the modulator is common to both types of transmitter systems, the operation of a typical modulator will be discussed first.

RADAR MODULATOR

The modulator controls the radar pulse width by means of a rectangular dc pulse (modulator pulse) of the required duration and amplitude. The peak power of the transmitted rf pulse depends on the amplitude of the modulator pulse.

Figure 2-3 shows the waveforms of the trigger pulse applied by the synchronizer to the modulator, the modulator pulse applied to the radar transmitter, and the transmitted rf pulse.

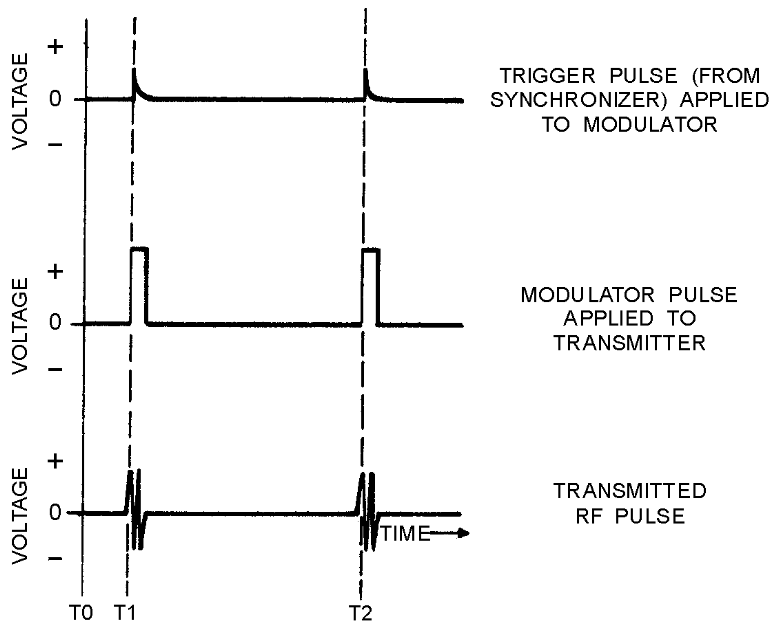


Figure 2-3.—Transmitter waveforms.

As you can see in the figure, the modulator pulse is applied to the transmitter the instant the modulator receives the trigger pulse from the synchronizer (T1, T2). The modulator pulse is flat on top and has very steep leading and trailing edges. These pulse characteristics are necessary for the proper operation of the transmitter and for the accurate determination of target range. The range timing circuits must be triggered the instant the leading edge of the transmitted rf pulse leaves the transmitter. In this way, the trigger pulse that controls the operation of the modulator also synchronizes the cathode-ray tube sweep circuits and range measuring circuits.

MAGNETRON OSCILLATORS are capable of generating rf pulses with very high peak power at frequencies ranging from 600 to 30,000 megahertz. However, if its cathode voltage changes, the magnetron oscillator shifts in frequency. To avoid such a frequency change, you must ensure that the amplitude of the modulator (dc) pulse remains constant for the duration of the transmitted rf pulse. That is, the modulator pulse must have a flat top. The range of cathode voltages over which a magnetron oscillates in the desired frequency spectrum is relatively small.

When a low voltage is applied to a magnetron, the magnetron produces a noise voltage output instead of oscillations. If this noise enters the receiver, it can completely mask the returning echoes. If a modulator pulse builds up and decays slowly, noise is produced at both the beginning and end of the pulse. Therefore, for efficient radar operation, a magnetron requires a modulator pulse that has a flat top and steep leading and trailing edges. An effective modulator pulse must perform in the following manner:

- Rise from zero to its maximum value almost instantaneously
- Remain at its maximum value for the duration of the transmitted rf pulse
- Fall from its maximum value to zero almost instantaneously

In radars that require accurate range measurement, the transmitted rf pulse must have a steep leading edge. The leading edge of the echo is used for range measurement. If the leading edge of the echo is not steep and clearly defined, accurate range measurement is not possible. The leading and trailing edges of echoes have the same shape as the leading and trailing edges of the transmitted rf pulse.

A transmitted rf pulse with a steep trailing edge is essential for the detection of objects at short ranges. If the magnetron output voltage drops gradually from its maximum value to zero, it contributes very little to the usable energy of the transmitted rf pulse. Furthermore, part of the magnetron output voltage enters the receiver and obscures nearby object echoes.

Types of Modulators

The two types of modulators are the LINE-PULSING MODULATOR and the HARD-TUBE MODULATOR. (A hard tube is a high-vacuum electron tube.) The line-pulsing modulator stores energy and forms pulses in the same circuit element. This element is usually the pulse-forming network. The hard-tube modulator forms the pulse in the driver; the pulse is then amplified and applied to the modulator. The hard tube modulator has been replaced by the line-pulsed modulator in most cases. This is because the hard-tube modulator has lower efficiency, its circuits are more complex, a higher power supply voltage is required, and it is more sensitive to voltage changes.

The line-pulsing modulator is easier to maintain because of its less complex circuitry. Also, for a given amount of power output, it is lighter and more compact. Because it is the principally used modulator in modern radar, it is the only type that will be discussed.

Figure 2-4 shows the basic sections of a radar modulator. They are as follows:

- The *power supply*.
- The *storage element* (a circuit element or network used to store energy).
- The *charging impedance* (used to control the charge time of the storage element and to prevent short-circuiting of the power supply during the modulator pulse).
- The *modulator switch* (used to discharge the energy stored by the storage element through the transmitter oscillator during the modulator pulse).

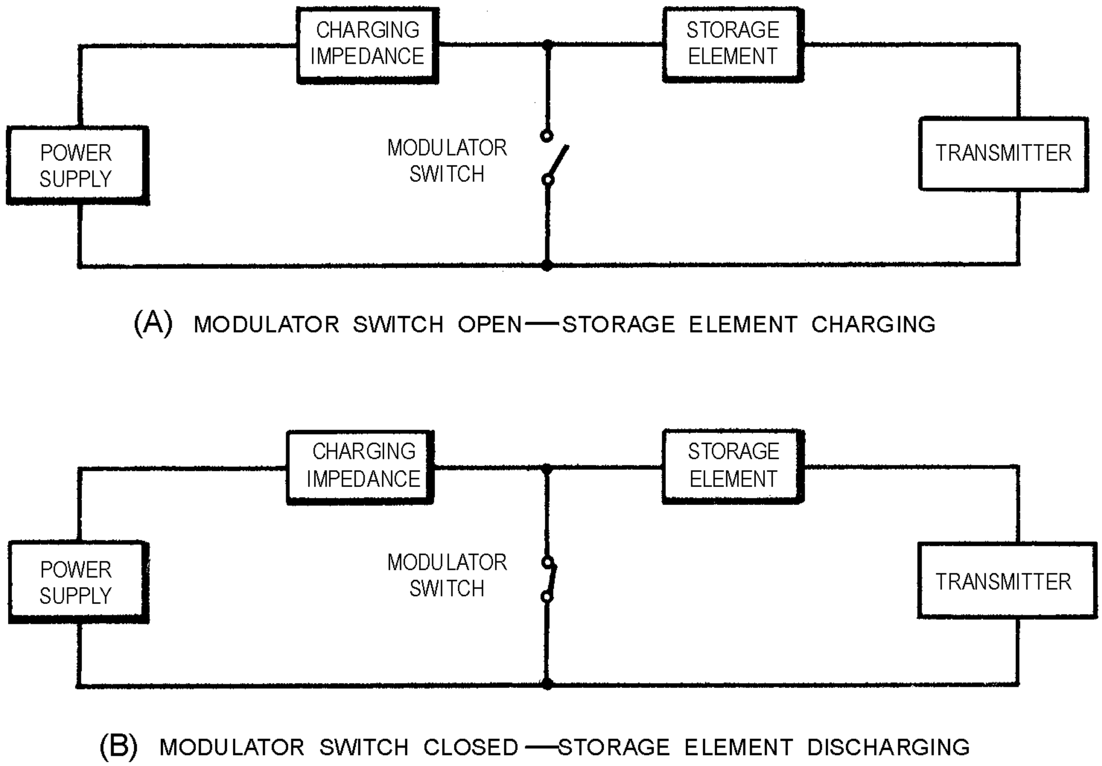


Figure 2-4.—Basic line-pulsing modulator block diagram.

View A of figure 2-4 shows the modulator switch open and the storage element charging. With the modulator switch open, the transmitter produces no power output, but the storage element stores a large amount of energy. View B shows the modulator switch closed and the storage element discharging through the transmitter. The energy stored by the storage element is released in the form of a high-power, dc modulator pulse. The transmitter converts the dc modulator pulse to an rf pulse, which is radiated into space by the radar antenna. Thus, the modulator switch is closed for the duration of a transmitted rf pulse, but open between pulses.

Many different kinds of components are used in radar modulators. The power supply generally produces a high-voltage output, either alternating or direct current. The charging impedance may be a resistor or an inductor. The storage element is generally a capacitor, an artificial transmission line, or a pulse-forming network. The modulator switch is usually a thyratron.

Modulator Storage Element

Capacitor storage elements are used only in modulators that have a dc power supply and an electron-tube modulator switch.

The capacitor storage element is charged to a high voltage by the dc power supply. It releases only a small part of its stored energy to the transmitter. The electron-tube modulator switch controls the charging and discharging of the capacitor storage element.

The artificial transmission line storage element, shown in view A of figure 2-5, consists of identical capacitors (C) and inductors (L) arranged to simulate sections of a transmission line. The artificial transmission line serves two purposes: (1) to store energy when the modulator switch is open (between transmitted rf pulses) and (2) to discharge and form a rectangular dc pulse (modulator pulse) of the required duration when the modulator switch is closed.

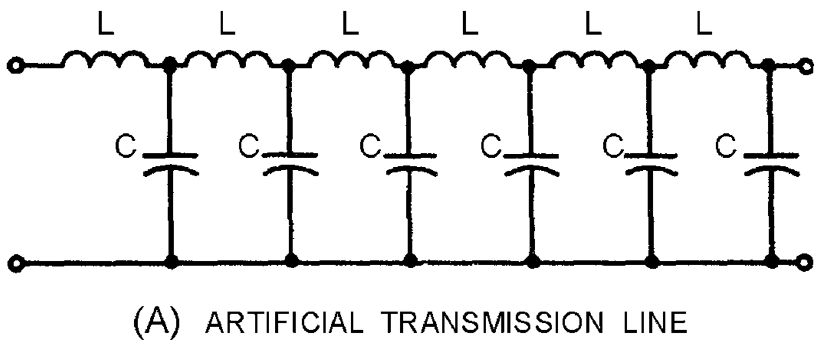


Figure 2-5A.—Modulator storage elements.

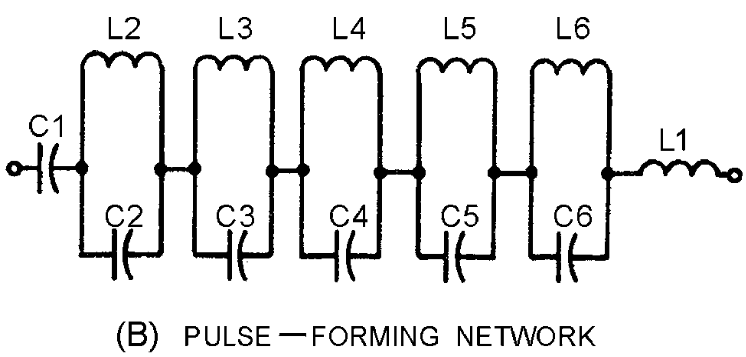


Figure 2-5B.—Modulator storage elements.

The duration of the modulator pulse depends on the values of inductance and capacitance in each LC section of the artificial transmission line in view A and the number of LC sections used. Other arrangements of capacitors and inductors (such as the pulse-forming network shown in view B) are very similar in operation to artificial transmission lines.

ARTIFICIAL TRANSMISSION LINES and PULSE-FORMING NETWORKS (pfn) are used more often than the capacitor-type storage elements.

ARTIFICIAL TRANSMISSION LINE.—Figure 2-6 shows a radar modulator that uses an artificial transmission line as its storage element. A modulator switch controls the pulse-repetition rate. When the modulator switch is open (between transmitted rf pulses), the transmission line charges.

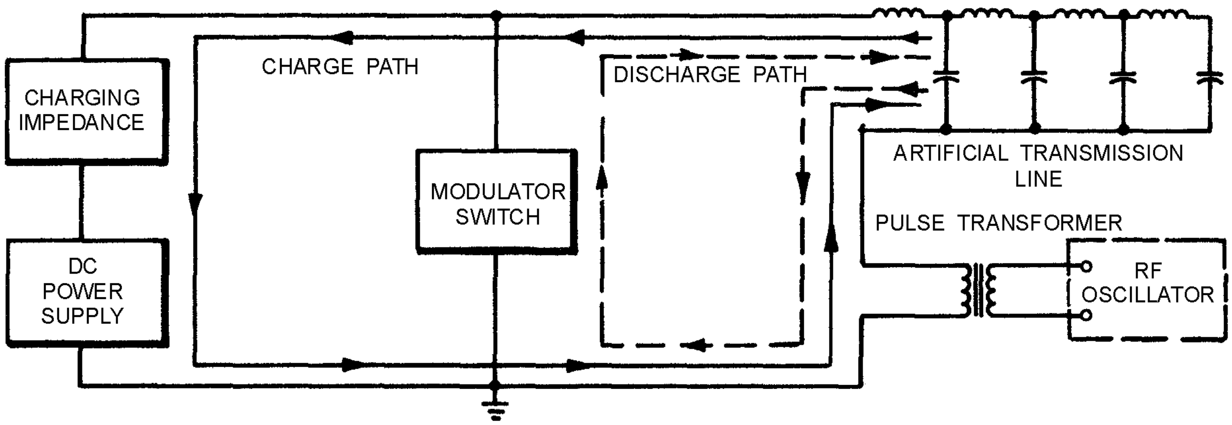


Figure 2-6.—Modulator with an artificial transmission line for the storage element.

The charge path includes the primary of the pulse transformer, the dc power supply, and the charging impedance. When the modulator switch is closed, the transmission line discharges through the series circuit. This circuit consists of the modulator switch and the primary of the pulse transformer.

The artificial transmission line is effectively an open circuit at its output end. Therefore, when the voltage wave reaches the output end of the line, it is reflected. As the reflected wave propagates from the output end back toward the input end of the line, it completely discharges each section of the line. When the reflected wave reaches the input end of the line, the line is completely discharged, and the modulator pulse ceases abruptly. If the oscillator and pulse transformer circuit impedance is properly matched to the line impedance, the voltage pulse that appears across the transformer primary equals one-half the voltage to which the line was initially charged.

The width of the pulse generated by an artificial transmission line depends on the time required for a voltage wave to travel from the input end to the output end of the line and back. Therefore, we can say the pulse width depends on the velocity of propagation along the line (determined by the inductances and capacitances of each section of the line) and the number of line sections (the length of the line).

PULSE-FORMING NETWORKS.—A pulse-forming network is similar to an artificial transmission line in that it stores energy between pulses and produces a nearly rectangular pulse. The pulse-forming network in view B of figure 2-5 consists of inductors and capacitors so arranged that they approximate the behavior of an artificial transmission line.

Each capacitor in the artificial transmission line, shown in view A, must carry the high voltage required for the modulator pulse. Because each capacitor must be insulated for this high voltage, an artificial transmission line consisting of many sections would be bulky and cumbersome.

The pulse-forming network, shown in view B of figure 2-5, can carry high voltage but does not require bulky insulation on all of its capacitors. Only series capacitor C1 must have high-voltage insulation. Because the other capacitors are in parallel with the corresponding inductors, the modulator-pulse voltage divides nearly equally among them. Thus, except for C1, the elements of the pulse-forming network are relatively small.

Pulse-forming networks are often insulated by immersing each circuit element in oil. The network is usually enclosed in a metal box on which the pulse width, characteristic impedance, and safe operating voltage of the network are marked. If one element in such a network fails, the entire network must be replaced.

Q10. What are the two basic types of transmitters?

Q11. What controls transmitter pulse width?

Q12. In addition to a flat top, what characteristics must a modulator pulse have?

Q13. What type of modulator is most commonly used in modern radar systems?

Q14. What three types of storage elements most often are used in modulators?

Q15. What characteristic is determined by the time required for a voltage wave to travel from the input end of an artificial transmission line to the output end and back again?

Modulator Switching Devices

The voltage stored in a storage-element capacitor, artificial transmission line, or pulse-forming network must be discharged through a MODULATOR SWITCHING DEVICE. The modulator switching device conducts for the duration of the modulator pulse and is an open circuit between pulses. Thus, the modulator switch must perform the following four functions:

1. Close very quickly and then reach full conduction in a small fraction of a microsecond
2. Conduct large currents (tens or hundreds of amperes) and withstand large voltages (thousands of volts)
3. Cease conducting (become an open circuit) with the same speed that it starts to conduct
4. Consume only a very small fraction of the power that passes through it

These switching and conducting requirements are met best by the THYRATRON tube. The thyratron tube is normally held below cutoff by a negative grid voltage and conducts when a positive trigger pulse is applied to its grid. Once fired, the thyratron tube continues to conduct as long as the storage element (artificial transmission line or pulse-forming network) is discharging.

During discharge of the storage element, the gas in the thyratron tube is highly ionized. While the storage element discharges, the plate-to-cathode resistance of the thyratron is practically zero. When the storage element is completely discharged, current ceases to flow through the thyratron and the gases become deionized; the negative grid bias regains control, and the thyratron is cut off (the modulator switch opens).

Most radar modulators use a high-voltage, dc power supply. Typical dc power supplies for radar modulators use a half-wave rectifier, a full-wave rectifier, or a bridge rectifier.

The modulator charging impedance, shown in figure 2-7, prevents the dc power supply from becoming short-circuited when the modulator switch closes. When the modulator switch is open, the charging impedance also controls the rate at which the storage element charges. When the charging impedance is small, the storage element charges rapidly.

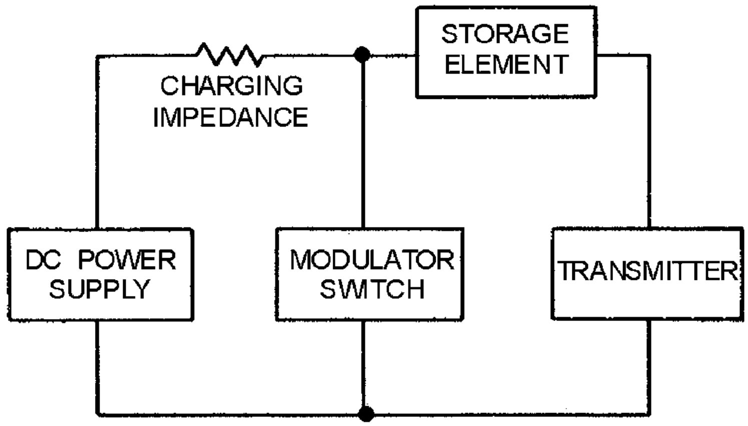


Figure 2-7.—Modulator charging impedance.

Many different kinds of charging impedance and charging circuits are used in radar modulators. The type of charging impedance and charging circuit used depends on the following five elements:

1. The type of power supply (ac or dc)
2. The type of storage element
3. The amount of modulator pulse voltage required
4. The pulse-repetition rate
5. The frequency of the available ac supply voltage

Q16. What type of tube best meets the requirements of a modulator switching element?

Q17. What modulator element controls the rate at which the storage element charges?

KEYED-OSCILLATOR TRANSMITTER

The KEYED-OSCILLATOR TRANSMITTER most often uses a MAGNETRON as the power oscillator. The following discussion is a description of a magnetron used as a keyed-oscillator radar transmitter.

Figure 2-8 shows the typical transmitter system that uses a magnetron oscillator, waveguide transmission line, and microwave antenna. The magnetron at the bottom of the figure is connected to the waveguide by a coaxial connector. High-power magnetrons, however, are usually coupled directly to the waveguide. A cutaway view of a typical waveguide-coupled magnetron is shown in figure 2-9.

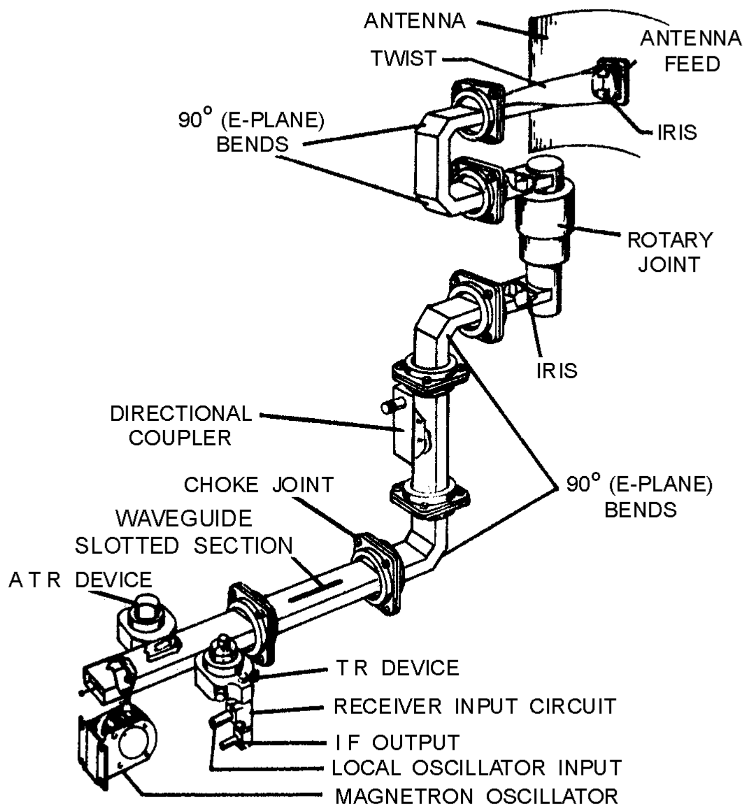


Figure 2-8.—Keyed oscillator transmitter physical layout.

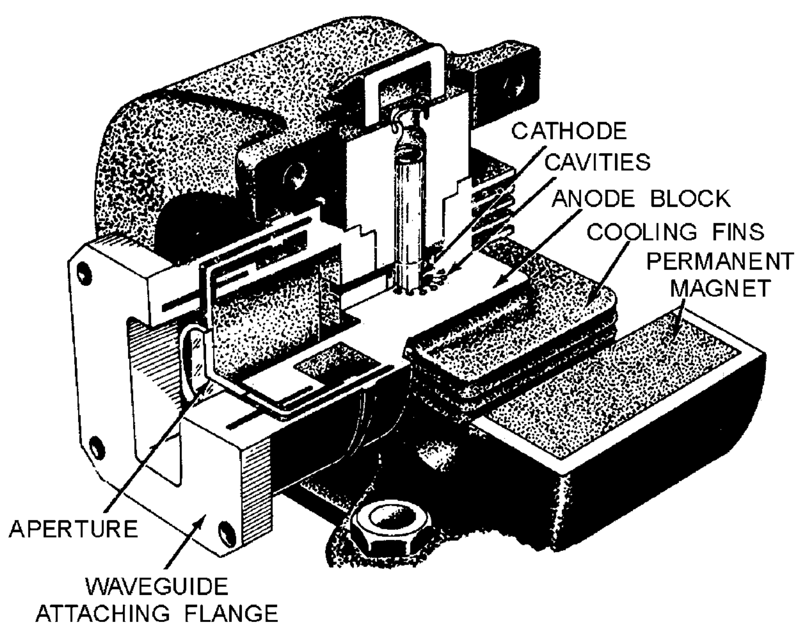


Figure 2-9.—Typical magnetron.

The magnetron is an electron tube in which a magnetic (H) field between the cathode and plate is perpendicular to an electric (E) field. Tuned circuits, in the form of cylindrical cavities in the plate, produce rf electric fields. Electrons interact with these fields in the space between the cathode and plate to produce an ac power output. Magnetrons function as self-excited microwave oscillators. These multicavity devices may be used in radar transmitters as either pulsed or cw oscillators at frequencies ranging from approximately 600 to 30,000 megahertz. (If you wish to review magnetron operation in more detail, refer to NEETS, Module 11, *Microwave Principles*.)

Let's examine the following characteristics of a magnetron used as a pulse radar transmitter oscillator stage:

- Stability
- Pulse characteristics
- The magnet
- Output coupling

Stability

In speaking of a magnetron oscillator, STABILITY usually refers to the stability of the mode of operation of the magnetron. The two main types of mode instability are MODE SKIPPING and MODE SHIFTING.

Mode skipping (or misfiring) is a condition in which the magnetron fires randomly in an undesired, interfering mode during some pulse times, but not in others. Pulse characteristics and tube noises are factors in mode skipping.

Mode shifting is a condition in which the magnetron changes from one mode to another during pulse time. This is highly undesirable and does not occur if the modulator pulse is of the proper shape, unless the cathode of the magnetron is in very poor condition.

Pulse Characteristics

PULSE CHARACTERISTICS are the make up of the high-voltage modulator pulse that is applied to the magnetron. The pulse should have a steep leading edge, a flat top, and a steep trailing edge. If the leading edge is not steep, the magnetron may begin to oscillate before the pulse reaches its maximum level. Since these low-power oscillations will occur in a different mode, the mode of the magnetron will be shifted as the pulse reaches maximum power. This mode shifting will result in an undesirable magnetron output. For the same reason (to prevent mode shifting), the top of the modulator pulse should be as flat as possible. Variations in the applied operating power will cause variations in the mode of operation. The trailing edge of the pulse should also be steep for the same reason--to prevent mode shifting.

Magnet

The purpose of the MAGNET is to produce a fairly uniform magnetic field of the desired value over the interaction space between the cathode and plate of the magnetron. The strength of the magnet is critical for proper operation. If the magnetic field strength is too high, the magnetron will not oscillate. If the magnetic field strength is too low, the plate current will be excessive and power output will be low. Frequency of operation will also be affected.

Since the strength of the magnet is critical, you should be careful when handling the magnet. Striking the magnet, especially with a ferromagnetic object, will misalign the molecular structure of the magnet and decrease the field strength.

Output Coupling

The OUTPUT COUPLING transfers the rf energy from the magnetron to the output transmission line (coaxial line or waveguide). A number of considerations impose restrictions upon the output circuit. The wavelength (frequency) and the power level of the magnetron output energy determine whether the transmission line to the antenna will be waveguide or coaxial line.

The coaxial output circuit consists of a length of coaxial line in which the center conductor is shaped into a loop and inserted into one of the magnetron cavities for magnetic coupling. The load side of the coupling line may feed either an external coaxial line or a waveguide. If the external line is coaxial, the connection may be direct or by means of choke joints. If the external line is a waveguide, the output circuit must include a satisfactory junction from the coaxial line to the waveguide. One type of junction used quite often is the PROBE COUPLER. The probe coupler acts as an antenna radiating into the waveguide.

The waveguide output may be fed directly by an opening (slot) into one of the magnetron cavities, as shown in figure 2-9. This opening must be covered by an iris window to maintain the vacuum seal.

The peak power ratings of magnetrons range from a few thousand watts (kilowatts) to several million watts (megawatts). The average power ratings are much lower, however, and vary from a few watts to several kilowatts. Additionally, many of the magnetrons used in modern radar systems are tunable in frequency. Typically, a tunable magnetron can vary the output frequency ±5 percent about the center of its frequency band. Thus the carrier frequency of radar can be changed to obtain the best operation or avoid electronic jamming on a particular frequency.

Modulator signals of many thousands of volts are applied to the magnetron cathode during operation. These high voltage levels require large glass posts to insulate the cathode and filaments from the anode block. In some high-power magnetrons, the cathode is completely enclosed in a container filled with insulating oil.

WARNING

> **All radar transmitters contain lethal voltages. Extreme care and strict observance of all posted safety precautions are essential when working on a radar transmitter.**

Q18. What is the frequency range of magnetron oscillators?

Q19. What two forms of instability are common in magnetrons?

Q20. What is the effect on magnetron operation if the magnetic field strength is too high?

Q21. What is the typical frequency range about the center frequency of a tunable magnetron?

POWER-AMPLIFIER TRANSMITTER

POWER-AMPLIFIER TRANSMITTERS are used in many recently developed radar sets. This type of transmitter was developed because of the need for more stable operation of the moving target indicator (mti). In a magnetron transmitting system, the high-power magnetron oscillator has a tendency to drift in frequency because of temperature variations, changes in the modulating pulse, and various other effects.

Frequency drift is compensated for, in part, by the use of automatic frequency control (afc) circuits designed to control the frequency of the local oscillator in the receiver system. This, however, does not completely eliminate the undesirable effects of frequency drift on mti operation.

The power-amplifier transmitter system does the same thing as the keyed-oscillator transmitter but with fewer stability problems. It generates, shapes, and amplifies pulses of rf energy for transmission.

Figure 2-10 is a block diagram of a typical power-amplifier transmitter system. In this transmitter system a multicavity klystron tube amplifies lower-powered rf pulses that have been generated and shaped in other stages. CROSSED-FIELD AMPLIFIERS (AMPLITRONS) are used in radar systems with a wide band of transmitted frequencies because they are stable over a wider frequency range. A crossed-field amplifier transmitter is discussed later in this section.

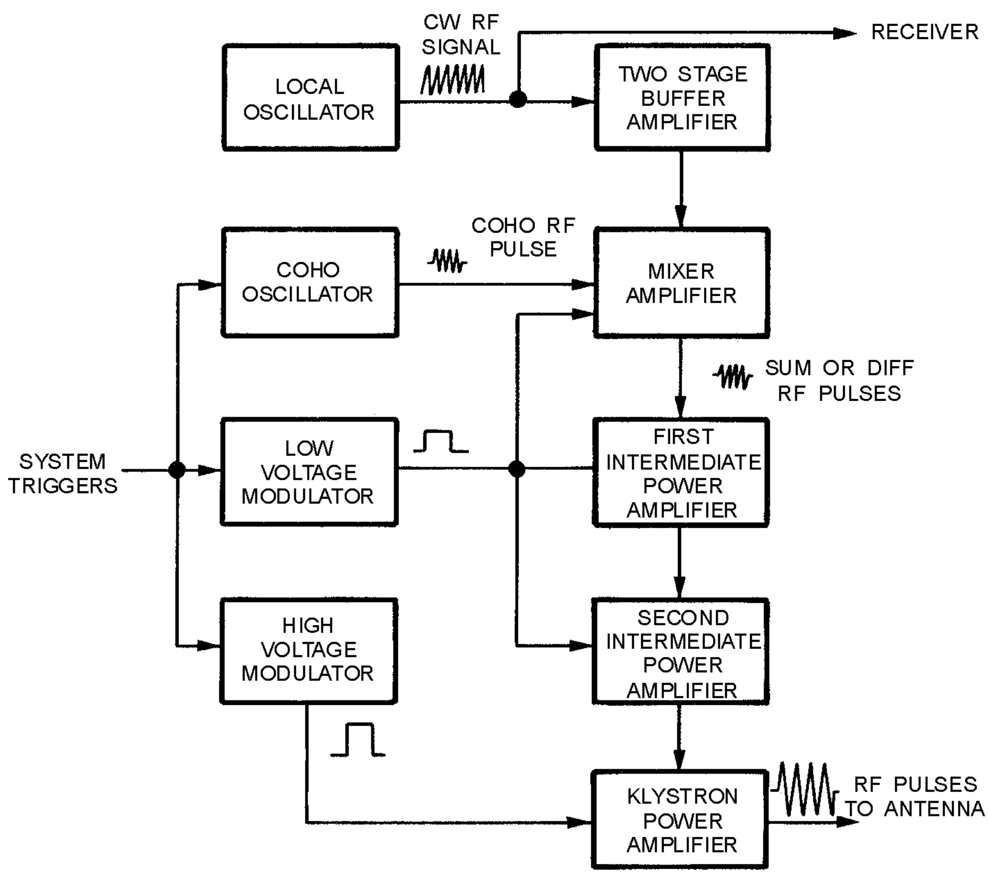


Figure 2-10.—Power amplifier transmitter block diagram.

In figure 2-10, the power-amplifier chain input signals are generated by heterodyning (mixing) two frequencies. That is, two different frequencies are fed to a mixer stage (mixer amplifier) and the resultant, either the sum or difference frequency, may be selected as the output. (The operation of mixer circuits is explained in more detail in the section on receivers.) The low-power pulse is then amplified by intermediate power amplifier stages and applied to the klystron power-amplifier. The klystron power-amplifier concentrates the rf output energy into a very narrow frequency spectrum. This concentration makes the system more sensitive to smaller targets. In addition the detection range of all targets is increased.

To examine the operation of the transmitter, we will trace the signal through the entire circuit. The local oscillator shown at the left of figure 2-10 is a very stable rf oscillator that produces two cw rf outputs. As shown, the cw output is sent to the receiver system; the cw output is also one of the two rf signals fed to the mixer amplifier by way of the two BUFFER AMPLIFIER STAGES. The buffer amplifiers raise the power level of the signal and also isolate the local oscillator.

The COHERENT OSCILLATOR (COHO) is triggered by the system trigger and produces as its output an rf pulse. This rf pulse is fed directly to the mixer amplifier.

The mixer-amplifier stage receives three signals: the coherent rf pulse, the local oscillator cw rf signal, and a dc modulating pulse from the low-voltage modulator. The coherent and local oscillator signals are mixed to produce sum and difference frequency signals. Either of these may be selected as the output. The modulator pulse serves the same purpose as in the keyed-oscillator transmitter, because it determines the pulse width and power level. The mixer stage functions only during the modulator pulse time. Thus the mixer amplifier produces an output of rf pulses in which the frequency may be either the sum or difference of the coherent and local oscillator signals.

The mixer-amplifier feeds the pulses of rf energy to an intermediate power amplifier. This amplifier stage is similar to the buffer-amplifier stage except that it is a pulsed amplifier. That is, the pulsed amplifier has operating power only during the time the modulator pulse from the low-voltage modulator is applied to the stage. The amplified output signal is fed to a second intermediate power amplifier that operates in the same manner as the first.

From the second intermediate power amplifier, the signal is fed to the KLYSTRON POWER AMPLIFIER. This stage is a multicavity power klystron. The input rf signal is used as the exciter signal for the first cavity. High-voltage modulating pulses from the high-voltage modulator are also applied to the klystron power amplifier. These high-voltage modulating pulses are stepped up across a pulse transformer before being applied to the klystron. All cavities of the klystron are tunable and are tuned for maximum output at the desired frequency.

Provisions are made in this type of transmitter to adjust the starting time of the modulating pulses applied to the coherent oscillator, mixer amplifier, intermediate power amplifiers, and klystron power-amplifier. By this means the various modulator pulses are made to occur at the same time.

This transmitter produces output rf pulses of constant power and minimum frequency modulation and ensures good performance.

Q22. What is the primary advantage of power-amplifier transmitters over keyed-oscillator transmitters?

Q23. In the power amplifier shown in figure 2-10, what two signals are mixed to produce the output signal?

Q24. What type of klystron is used as the final stage of a power-amplifier transmitter?

Figure 2-11 is a block diagram of a power-amplifier transmitter that uses a FREQUENCY SYNTHESIZER to produce the transmitted frequency rather than the heterodyning mixer. The frequency synthesizer allows the transmitter to radiate a large number of discrete frequencies over a relatively wide band. Such a system is commonly used with frequency-scan search radars that must transmit many different frequencies to achieve elevation coverage and to compensate for the roll and pitch of a ship.

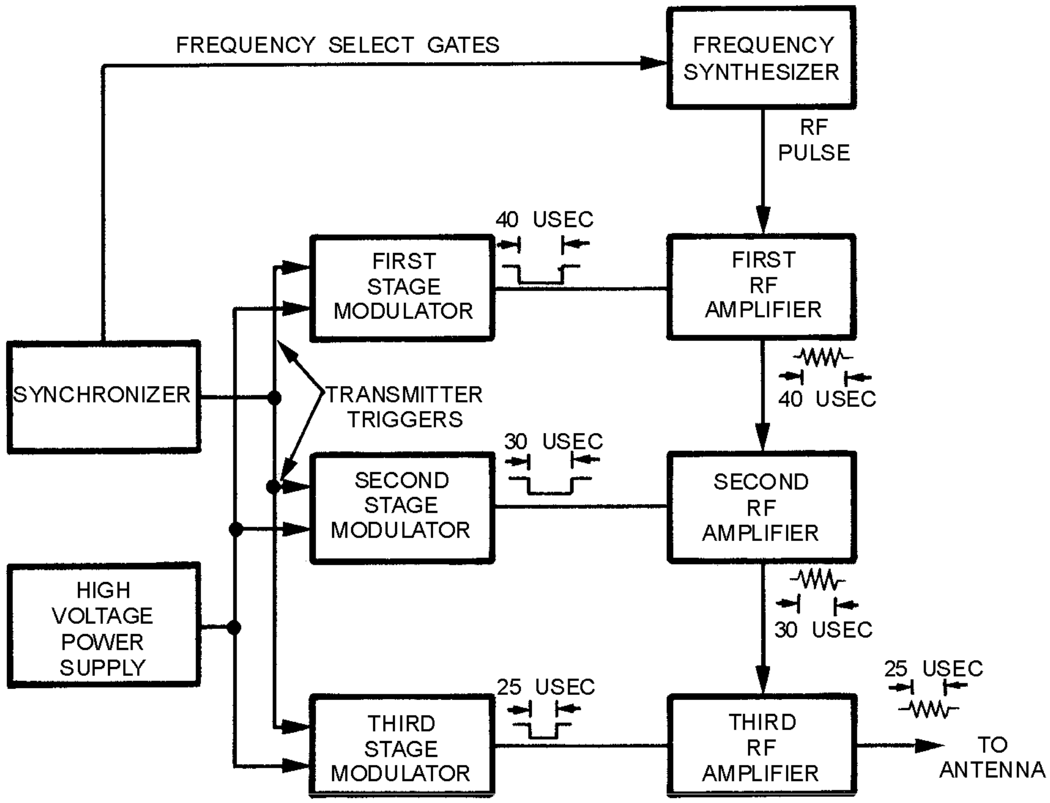


Figure 2-11.—Power amplifier transmitter using crossed-field amplifiers.

A typical frequency synthesizer consists of a bank of oscillators producing different fixed frequencies. The outputs of a relatively few fixed oscillators can be mixed in various combinations to produce a wide range of frequencies. In mti systems the selected oscillator frequencies are mixed with a coherent oscillator frequency to provide a stable reference for the mti circuits. The frequency synthesizer also produces the local oscillator signals for the receiver system. Because the transmitted pulse changes frequency on each transmission, the local oscillator signal to the receiver must also change and be included in the transmitted frequency. A system of this type is frequency-programmed by select gates from the synchronizer.

The detailed operation of frequency synthesizers is beyond the scope of this manual but may be found in the technical manuals for most frequency scan radar systems.

The first rf amplifier receives the pulses of the selected frequency from the synthesizer and a modulator pulse (from the first stage modulator) at the same time. The rf pulse is usually slightly wider than the modulator pulse which prevents the amplifier tube from pulsing when no rf energy is present. Most pulsed rf amplifiers will oscillate at an undesired frequency if pulsed without an rf input. The output of the first rf amplifier is an amplified rf pulse that is the same width as the first stage modulator pulse. The second stage modulator is designed to produce a pulse slightly narrower than the first stage modulator pulse; this also prevents the amplifier from pulsing when no rf is present. Therefore, the second stage amplifier receives a modulator pulse a short time after the first stage rf arrives at the input. As shown in figure 2-11, the same procedure is repeated in the third and final stage.

The amplifiers in this type of power-amplifier transmitter must be broad-band microwave amplifiers that amplify the input signals without frequency distortion. Typically, the first stage and the second stage are traveling-wave tubes (twt) and the final stage is a crossed-field amplifier. Recent technological advances in the field of solid-state microwave amplifiers have produced solid-state amplifiers with enough output power to be used as the first stage in some systems. Transmitters with more than three stages usually use crossed-field amplifiers in the third and any additional stages. Both traveling-wave tubes and crossed-field amplifiers have a very flat amplification response over a relatively wide frequency range.

Crossed-field amplifiers have another advantage when used as the final stages of a transmitter; that is, the design of the crossed-field amplifier allows rf energy to pass through the tube virtually unaffected when the tube is not pulsed. When no pulse is present, the tube acts as a section of waveguide. Therefore, if less than maximum output power is desired, the final and preceding cross-field amplifier stages can be shut off as needed. This feature also allows a transmitter to operate at reduced power, even when the final crossed-field amplifier is defective.

Q25. What transmitter component allows the radiation of a large number of discrete frequencies over a wide band?

Q26. What is the result of pulsing a pulsed rf amplifier when no rf is present?

DUPLEXERS

Whenever a single antenna is used for both transmitting and receiving, as in a radar system, problems arise. Switching the antenna between the transmit and receive modes presents one problem; ensuring that maximum use is made of the available energy is another. The simplest solution is to use a switch to transfer the antenna connection from the receiver to the transmitter during the transmitted pulse and back to the receiver during the return (echo) pulse. No practical mechanical switches are available that can open and close in a few microseconds. Therefore, ELECTRONIC SWITCHES must be used. Switching systems of this type are called DUPLEXERS.

BASIC DUPLEXER OPERATION

In selecting a switch for this task, you must remember that protection of the receiver input circuit is as important as are output power considerations. At frequencies where amplifiers may be used, amplifier tubes can be chosen to withstand large input powers without damage. However, the input circuit of the receiver is easily damaged by large applied signals and must be carefully protected.

An effective radar duplexing system must meet the following four requirements:

1. During the period of transmission, the switch must connect the antenna to the transmitter and disconnect it from the receiver.

2. The receiver must be thoroughly isolated from the transmitter during the transmission of the high-power pulse to avoid damage to sensitive receiver components.

3. After transmission, the switch must rapidly disconnect the transmitter and connect the receiver to the antenna. For targets close to the radar to be seen, the action of the switch must be extremely rapid.

4. The switch should absorb an absolute minimum of power both during transmission and reception.

Therefore, a radar duplexer is the microwave equivalent of a fast, low-loss, single-pole, double-throw switch. The devices developed for this purpose are similar to spark gaps in which high-current microwave discharges furnish low-impedance paths. A duplexer usually contains two switching tubes (spark gaps) connected in a microwave circuit with three terminal transmission lines, one each for the transmitter, receiver, and antenna. As shown in views A and B of figure 2-12, these circuits may be connected in parallel or in series. Both systems will be discussed in detail in this section. One tube is called the TRANSMIT-RECEIVER TUBE, or TR TUBE; the other is called the ANTITRANSMIT-RECEIVE TUBE, or ATR TUBE. The tr tube has the primary function of disconnecting the receiver, and the atr tube of disconnecting the transmitter.

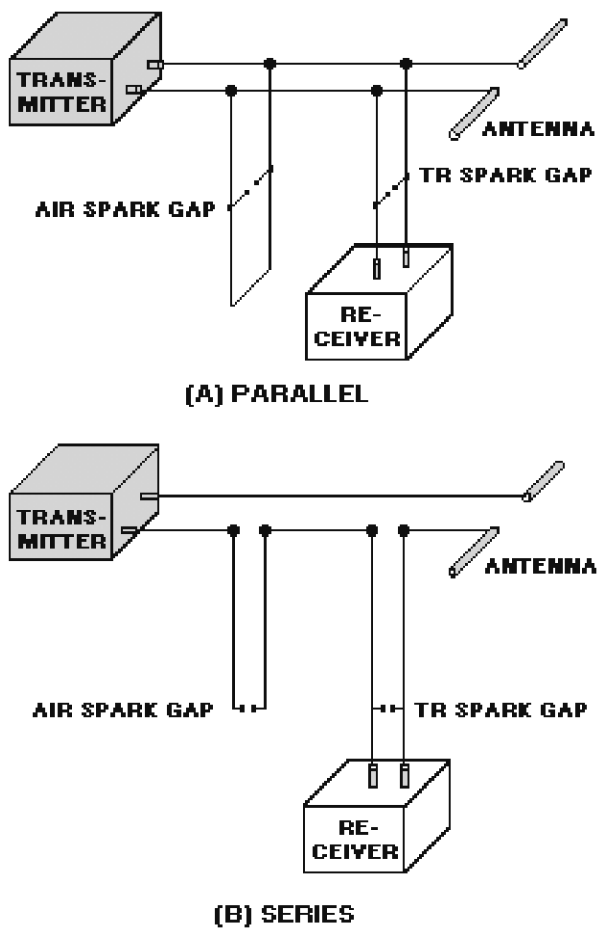


Figure 2-12.—Duplexer systems.

The overall action of the tr and atr circuits depends upon the impedance characteristics of the quarter-wavelength section of transmission line. A quarter-wavelength, or an odd multiple of the quarter-wavelength, transmission line presents opposite impedance values at the ends; one end of the line appears as a short and the other end appears as an open.

TR Tube

The type of spark gap used as a tr tube may vary. It may be one that is simply formed by two electrodes placed across the transmission line; or it may be one enclosed in an evacuated glass envelope with special features to improve operation. The requirements of the spark gap are (1) high impedance prior to the arc and (2) very low impedance during arc time. At the end of the transmitted pulse the arc should be extinguished as rapidly as possible. Extinguishing the arc stops any loss caused by the arc and permits signals from nearby targets to reach the receiver.

The simple gap formed in air has a resistance during conduction of from 30 to 50 ohms. This is usually too high for use with any but an open-wire transmission line. The time required for the air surrounding the gap to completely deionize after the pulse voltage has been removed is about 10 microseconds. During this time the gap acts as an increasing resistance across the transmission line to which it is connected. However, in a tr system using an air gap, the echo signals reaching the receiver beyond the gap will be permitted to increase to half their proper magnitude 3 microseconds after the pulse voltage has been removed. This interval is known as RECOVERY TIME.

Tr tubes are usually conventional spark gaps enclosed in partially evacuated, sealed glass envelopes, as shown in figure 2-13. The arc is formed as electrons are conducted through the ionized gas or vapor. You may lower the magnitude of voltage necessary to break down a gap by reducing the pressure of the gas that surrounds the electrodes. Optimum pressure achieves the most efficient tr operation. You can reduce the recovery time, or DEIONIZATION TIME, of the gap by introducing water vapor into the tr tube. A tr tube containing water vapor at a pressure of 1 millimeter of mercury will recover in 0.5 microseconds. It is important for a tr tube to have a short recovery time to reduce the range at which targets near the radar can be detected. If, for example, echo signals reflected from nearby objects return to the radar before the tr tube has recovered, those signals will be unable to enter the receiver.

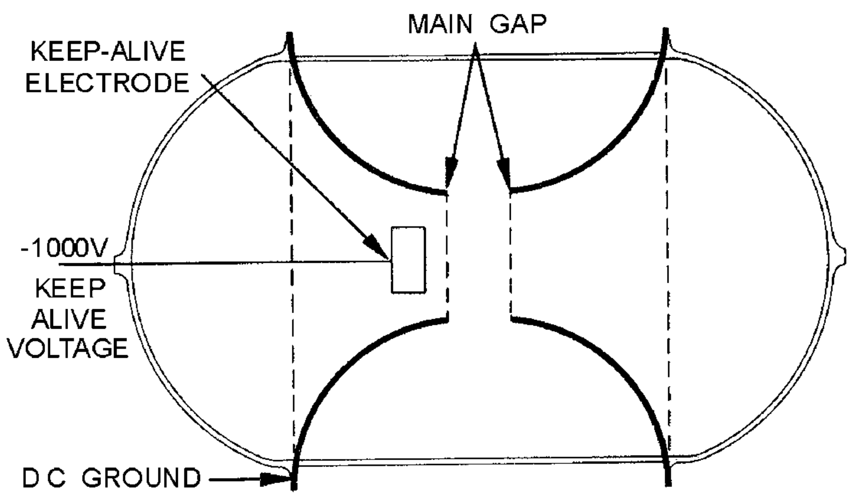


Figure 2-13.—Tr tube with keep-alive electrode.

Tr tubes used at microwave frequencies are built to fit into, and become a part of, a resonant cavity. You may increase the speed with which the gap breaks down after the transmitter fires by placing a voltage across the gap electrodes. This potential is known as KEEP-ALIVE VOLTAGE and ranges from 100 volts to 1,000 volts. A glow discharge is maintained between the electrodes. (The term GLOW DISCHARGE refers to the discharge of electricity through a gas-filled electron tube. This is distinguished by a cathode glow and voltage drop much higher than the gas-ionization voltage in the cathode vicinity.) This action provides for rapid ionization when the transmitter pulse arrives.

Failure of the tr tube is primarily caused by two factors. The first and most common cause of failure is the gradual buildup of metal particles that have been dislodged from the electrodes. Such metal bits become spattered on the inside of the glass envelope. These particles act as small, conducting areas and tend to lower the Q of the resonant cavity and dissipate power. If the tube continues in use for too long a period in this condition, the particles will form a detuning wall within the cavity and eventually prevent the tube from functioning. A second cause of failure is the absorption of gas within the enclosure by the metal electrodes. This results in a gradual reduction of pressure within the tube to a point where gap breakdown becomes very difficult. The final result is that extremely strong signals (from the transmitter) are coupled to the receiver. Because both types of failures develop gradually, the tr tube periodically must be checked carefully to determine performance level.

Q27. What type of switches are used as duplexers?

Q28. What tube in a duplexer has the primary function of disconnecting the receiver?

Q29. How may the tr tube ionization speed be increased?

ATR Tube

The atr tube is usually a simpler device than a tr tube. An atr tube might use a pure inert gas, such as argon, because recovery time generally is not a vital factor. Furthermore, a priming agent, such as keep-alive voltage, is not needed. The absence of either a chemically active gas or a keep-alive voltage results in atr tubes having longer useful lives than tr tubes.

WARNING

Tr and atr tubes may contain radioactive material. Handle with care to avoid breakage and possible contamination.

There are two basic tr-atr duplexer configurations. They are the parallel-connected and the series-connected duplexer systems. The following paragraphs describe the operation of both systems.

Parallel Connected Duplexer Operation

First, let's consider a PARALLEL-CONNECTED DUPLEXER system, as shown in figure 2-14. The tr spark gap shown in figure 2-14 is located in the receiver coupling line one-quarter wavelength from the T-junction. A half-wavelength, closed-end section of transmission line, called a STUB, is shunted across the main transmission line. An atr spark gap is located in this line one-quarter wavelength from the main transmission line and one-quarter wavelength from the closed end of the stub. As shown in the figure, antenna impedance, line impedance, and transmitter output impedance, when transmitting, are all equal. The action of the circuit during transmission is shown in figure 2-15.

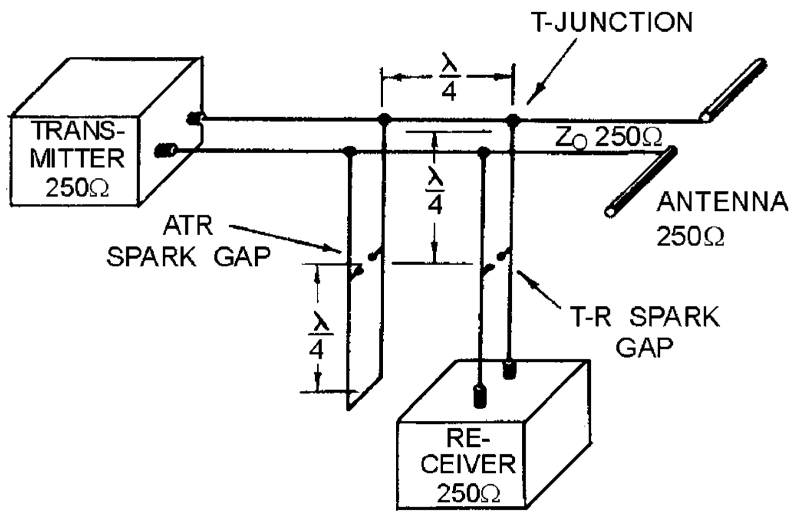


Figure 2-14.—Parallel-connected duplexer showing distance and impedance.

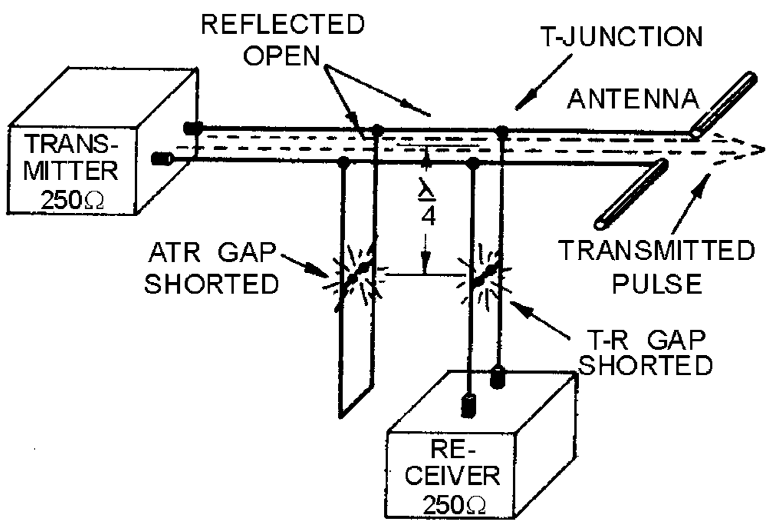


Figure 2-15.—Parallel-connected duplexer during transmission.

During the transmitting pulse, an arc appears across both spark gaps and causes the tr and atr circuits to act as shorted (closed-end) quarter-wave stubs. The circuits then reflect an open circuit to the tr and atr circuit connections to the main transmission line. None of the transmitted energy can pass through these reflected opens into the atr stub or into the receiver. Therefore, all of the transmitted energy is directed to the antenna.

During reception, as shown in figure 2-16, the amplitude of the received echo is not sufficient to cause an arc across either spark gap. Under this condition, the atr circuit now acts as a half-wave transmission line terminated in a short-circuit. This is reflected as an open circuit at the receiver T-junction, three-quarter wavelengths away. The received echo sees an open circuit in the direction of the transmitter. However, the receiver input impedance is matched to the transmission line impedance so that the entire received signal will go to the receiver with a minimum amount of loss.

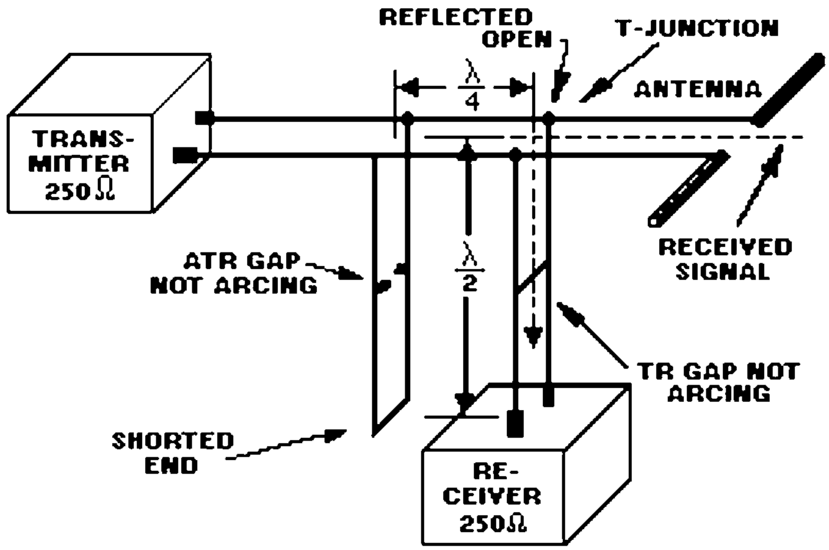


Figure 2-16.—Parallel-connected duplexer during reception.

Series-Connected Duplexer Operation

In the SERIES-CONNECTED DUPLEXER SYSTEM, shown in figure 2-17, the tr spark gap is located one-half wavelength from the receiver T-junction. The atr spark gap is located one-half wavelength from the transmission line and three-quarters wavelength from the receiver T-junction. During transmission, the tr and atr gaps fire in the series-connected duplexer system, as shown in figure 2-18. This causes a short-circuit to be reflected at the series connection to the main transmission line one-half wavelength away. The transmitted pulse "sees" a low impedance path in the direction of the antenna and does not go into the atr stub or the receiver.

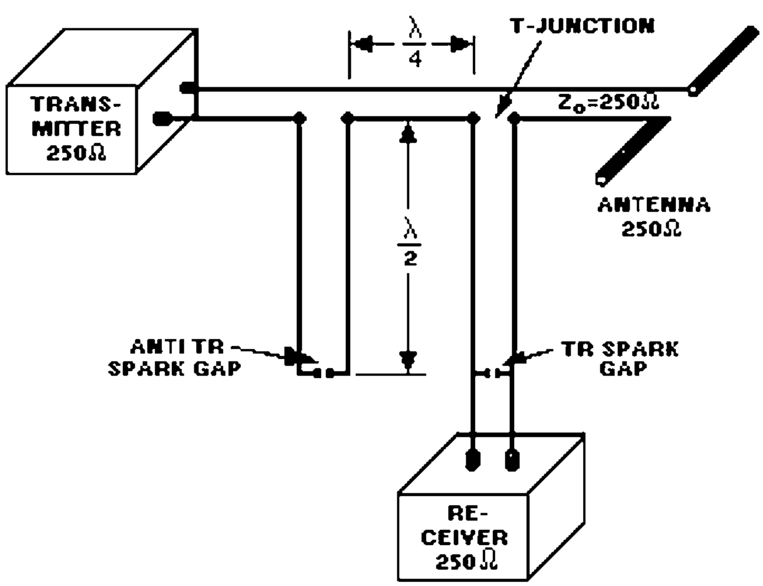


Figure 2-17.—Series-connected duplexer showing distance and impedance.

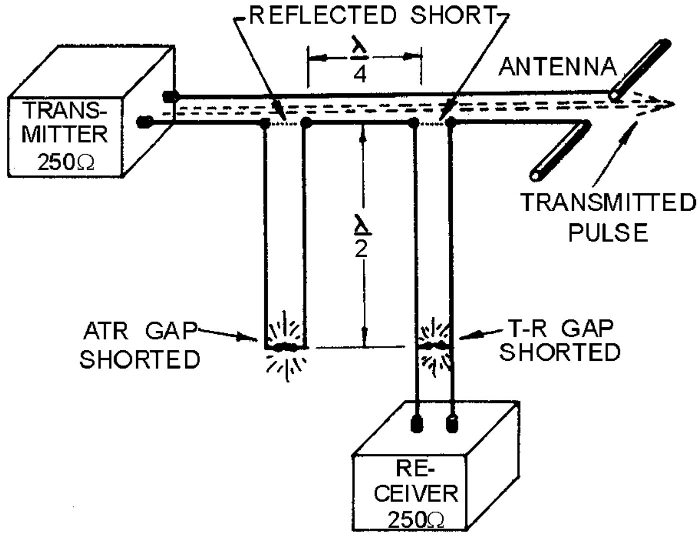


Figure 2-18.—Series-connected duplexer during transmission.

During reception, neither spark gap is fired, as shown in figure 2-19. The atr acts as a half-wave stub terminated in an open. This open is reflected as a short-circuit at the T-junction three-quarters of a wavelength away. Consequently, the received signal sees a low impedance path to the receiver, and none of the received signal is lost in the transmitting circuit.

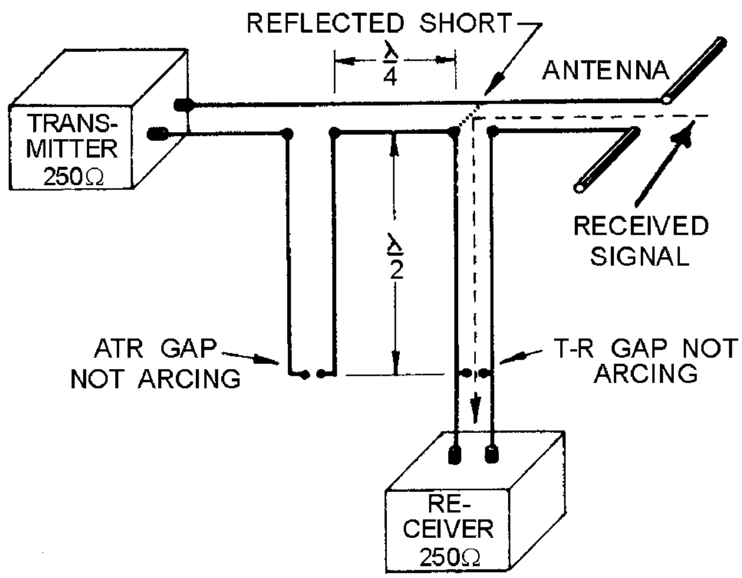


Figure 2-19.—Series-connected duplexer during reception.

DUPLEXER TYPES

Duplexers are constructed in many forms, such as RESONANT-CAVITY COAXIAL SYSTEMS, WAVEGUIDE SYSTEMS, and HYBRID RINGS. Since waveguide and hybrid-ring duplexers are most common in radar systems, those will be discussed in this section.

Waveguide Duplexer

WAVEGUIDE DUPLEXERS usually consist of tr tubes and atr tubes housed in a resonant cavity and attached to a waveguide system in some manner. Resonant-cavity tr tubes may be applied to waveguides, either directly or indirectly, to obtain switching action. The indirect method uses a coaxial line system, and then couples the coaxial line into the waveguide that feeds the antenna. If large losses are incurred by the use of a coaxial line, the resonant cavity can be coupled directly to the waveguide. Figure 2-20 shows a direct method of cavity tr switching in a waveguide system. The waveguide terminates in the antenna at one end and in a shorting plate at the other. The magnetron uses a voltage probe to excite the waveguide. The transmitted pulse travels up the guide and moves into the tr box through a slot. The cavity builds up a strong electric field across the gap, breaks it down, and detunes the cavity. This action effectively seals the opening and passes the pulse energy to the antenna.

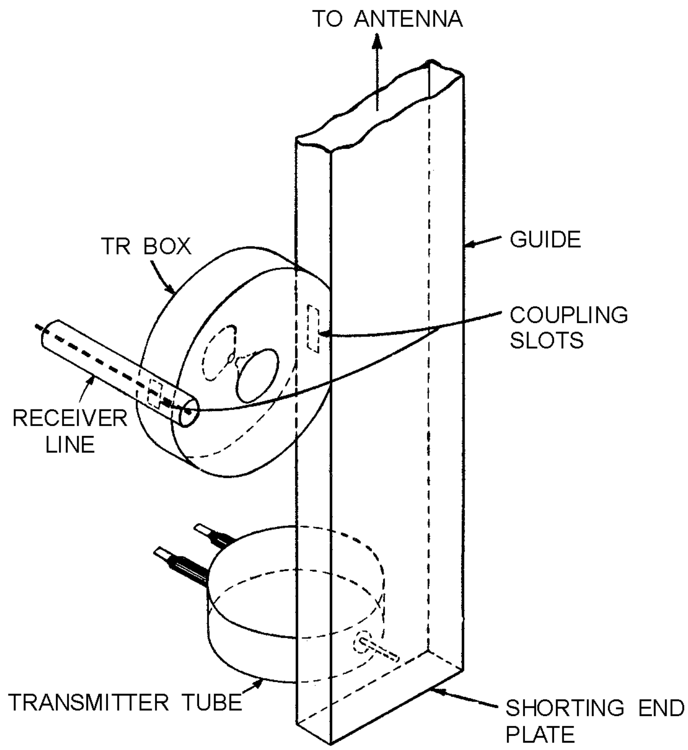


Figure 2-20.—Waveguide duplexer with cavity tr tube.

The signals received during the resting time travel down the guide to the magnetron and the shorting end plate where they are reflected. The slot coupling the waveguide to the cavity is located at a point where the standing-wave magnetic field produced by reflections in the waveguide is maximum. The maximum magnetic field, therefore, energizes the cavity. The echo signals are not strong enough to cause an arc, and the cavity field is undisturbed by the gap. Therefore, the cavity field couples rf energy into the receiver coaxial line and provides maximum energy transfer.

The cavity tr switch can also be applied to branch lines of the waveguide, as shown in figure 2-21. The magnetron is coupled to the guide by a voltage probe to produce proper excitation.

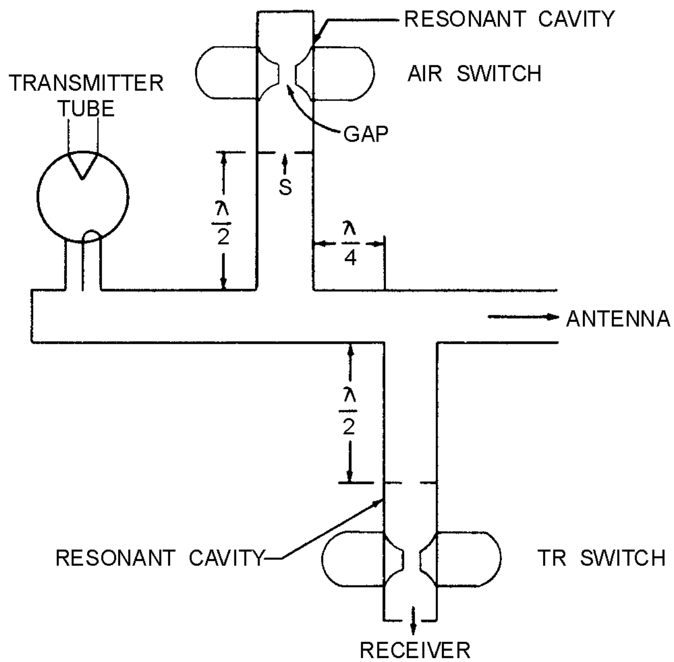


Figure 2-21.—Branched waveguide duplexer.

Maximum use of the received signals is ensured by an atr tube. The transmitted pulse travels from the magnetron to the atr branch where part of the energy is diverted into the gap. A slot (S) is placed across the waveguide one-half wavelength from the main guide, and passes the rf energy through it and into the cavity. The cavity builds up the electric field that breaks down the gap, detunes the cavity, and, as a result, effectively closes the slot. One-half wavelength away, this action effectively closes the entrance to the atr branch and limits the amount of energy entering the atr branch to a small value.

Most of the energy is, therefore, directed down the guide to the antenna. Upon reaching the receiver branch, the same effect is produced by the tr tube in the receiver line. Because the energy entering both openings is effectively limited by the gaps, maximum energy is transferred between the magnetron and the antenna.

During the resting time, the atr spark gap is not broken down by the received signals. The received signal sets up standing waves within the cavity that cause it to resonate. At resonance, the low impedance of the atr cavity is reflected as a high impedance at the entrance to the transmitter waveguide (three-quarter wavelength away). This ensures that the maximum received signal will enter the receiver branch.

Hybrid Ring Duplexer

The HYBRID RING is used as a duplexer in high-power radar systems. It is very effective in isolating the receiver during transmission. A simplified version of the hybrid-ring duplexer is shown in views A and B of figure 2-22. The operation of the duplexer, in terms of the E field distribution during transmission and reception, is illustrated in views C and D. The H lines, though present, have been omitted to simplify the explanation.

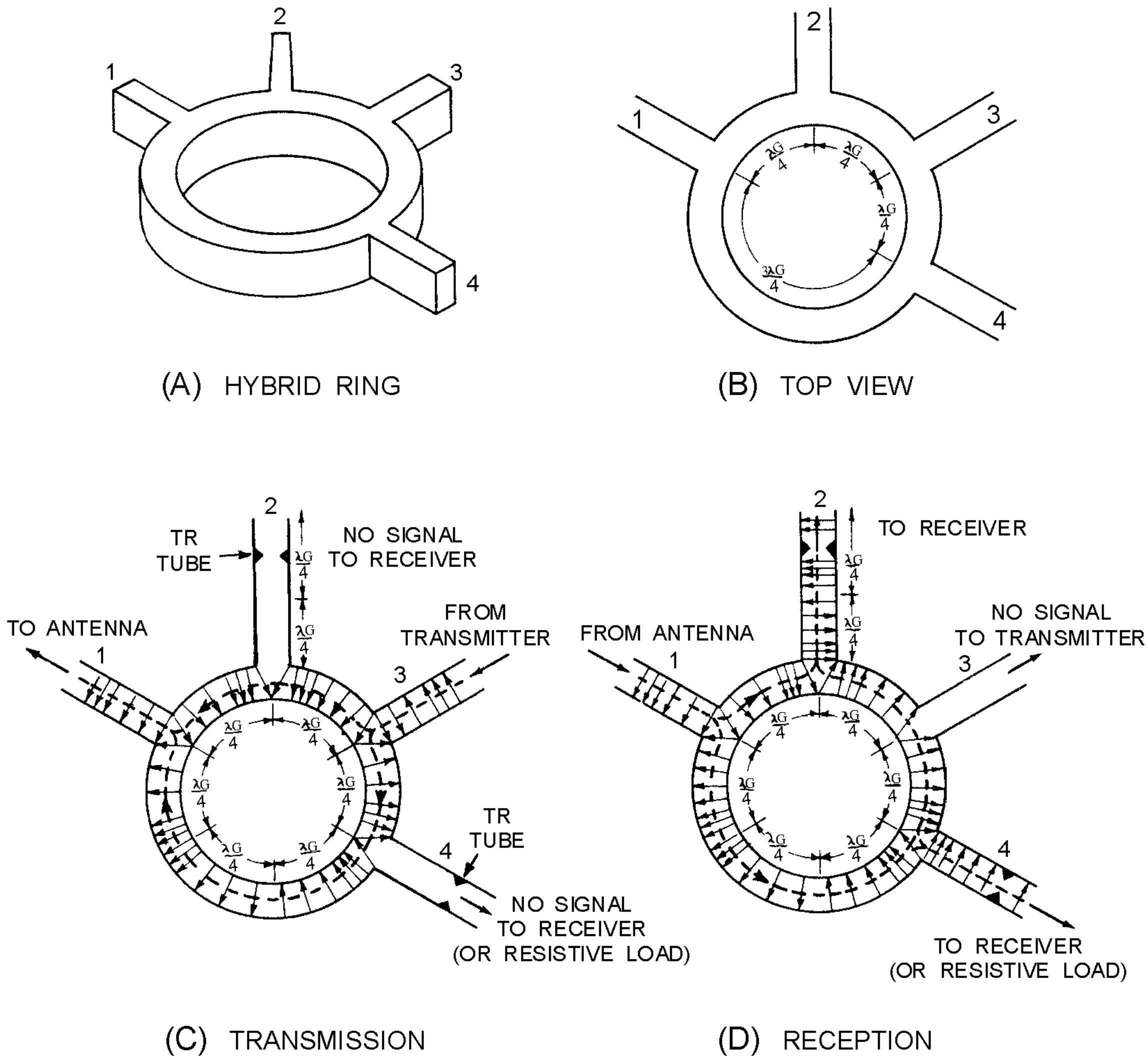


Figure 2-22.—Hybrid-ring duplexer.

During transmission the E field from the transmitter enters arm 3 and divides into two fields 180 degrees out of phase. One field moves clockwise around the ring and the other moves counterclockwise. The two fields must be 180 degrees out of phase at the entrance of an arm to propagate any energy down the arm. The field moving clockwise from arm 3 ionizes the tr tube in arm 4, and the energy is blocked from the receiver. The tr tube reflects a high impedance equivalent to an open circuit. This high impedance prevents any energy from entering the receiver - even though the two fields are out of phase at the entrance to arm 4. The field moving counterclockwise from arm 3 ionizes the tr tube in arm 2, which reflects a short circuit back to the ring junction. No energy is sent to the receiver, however, because the fields arriving at arm 2 are in phase. The clockwise and counterclockwise fields arrive at arm 1 out of phase by 180 degrees. They are then propagated through the arm to the antenna.

During reception, the relatively weak field from the antenna enters arm 1 and divides at the junction into two out-of-phase components. Neither field is sufficient to fire the tr tubes in arms 2 and 4; since the fields arrive at these arms out of phase, energy is propagated to the receiver. The energy arriving at arm 3 is in phase and will not be coupled to the transmitter. Since the operation of the arms of a hybrid ring is the same as the operation of E-type waveguide T-junctions, you may find it helpful to review NEETS, Module 11, *Microwave Principles*.

Q30. The actions of the tr and atr circuits depend on the impedance characteristics of what length of transmission line?

Q31. During which of the transmit or receive cycles are both the tr and atr tubes of a parallel-connected duplexer ionized (arcing)?

Q32. In a series-connected duplexer, what tube (tr or atr), if any, fires during the receive cycle?

Q33. To propagate energy down an arm of a hybrid ring duplexer, the two fields at the junction of the arm and the ring must have what phase relationship?

RECEIVERS

The energy that a distant object reflects back to the antenna in a radar system is a very small fraction of the original transmitted energy. The echoes return as pulses of rf energy of the same nature as those sent out by the transmitter. However, the power of a return pulse is measured in fractions of microwatts instead of in kilowatts, and the voltage arriving at the antenna is in the range of microvolts instead of kilovolts. The radar receiver collects those pulses and provides a visual display of object information.

Information about the position of the object is present visually when the reception of an echo causes the movement or appearance of a spot of light on a cathode-ray tube (crt). The crt requires a signal on the order of at least several volts for proper operation and will not respond to the high frequencies within a return pulse. Therefore, a receiver amplifier and detector must be used that are capable of producing a visible indication on the cathode-ray tube under the following conditions: (1) when the input signal to the amplifier is in the form of pulses of extremely high-frequency, (2) the amplitude of the pulses is in the microvolt range, and (3) the pulses last for only a few microseconds.

The radar receiver evolved directly from the simple radio receiver. The radar receiver operates on exactly the same principles as the radio receiver. However, the overall requirements and limitations of a radar receiver differ somewhat from those of a radio receiver because of the higher frequencies involved and the greater sensitivity desired.

In studying the radar receiver, we will first examine the overall requirements of a radar receiver. Second, we will examine a typical radar receiver that satisfies these requirements. Finally, we will discuss the individual components of the receiver.

RADAR RECEIVER REQUIREMENTS

The following characteristics determine the design requirements of an effective radar receiver:

- Noise
- Gain
- Tuning
- Distortion
- Blocking

Noise

The word NOISE is a carryover from sound-communications equipment terminology. Noise voltages in sound equipment produce actual noise in the loudspeaker output. In radar, noise voltages result in erratic, random deflection or intensity of the indicator sweep that can mask small return signals.

Were it not for noise, the maximum range at which an object would be detectable by radar could be extended almost infinitely. Objects at great range return exceedingly small echoes. However, without noise, almost any signal could be amplified to a usable level if enough stages were added to the receiver. Because of noise, the signal detection limit or sensitivity level of a receiver is reached when the signal level falls below the noise level to such an extent as to be obscured. A simple increase of amplification is of no help because both signal and noise are amplified at the same rate.

In the radar portion of the rf spectrum, external sources of noise interference are usually negligible; consequently, the sensitivity that can be achieved in a radar receiver is usually determined by the noise produced in the receiver. Not only must noise be kept down, but everything possible must be done to minimize attenuation of the video signal (echo) before it is amplified.

Gain

The GAIN of a radar receiver must be very high. This is because the strength of the signal at the antenna is at a level of microvolts and the required output to the indicator is several volts. The gain of a radar receiver is roughly in the range of 10^6 to 10^8. FEEDBACK, or REGENERATION, is one of the most serious difficulties in the design of an amplifier with such high gain. Special precautions must be taken to avoid feedback. Such precautions include careful shielding, decoupling (isolation) between voltage supplies for the different tubes, and amplification at different frequencies in separate groups of stages.

Tuning

The radar receiver requires a limited tuning range to compensate for transmitter and local oscillator frequency changes because of variations in temperature and loading. Microwave radar receivers usually use automatic frequency control (afc) for this purpose.

Distortion

If distortion occurs in the receiver, the time interval between the transmitted pulse and the received pulse changes, thereby affecting range accuracy.

Blocking

BLOCKING refers to a condition of the receiver in which the voltage pulse at the receiver input is too large. As a result, for a short time after the pulse, the receiver is insensitive or blocked to signals below a certain level. This condition results from one or more of the amplifier stages in the receiver being overdriven. After a strong pulse, the receiver may be biased to a point at which it will not amplify small signals. Recovery after blocking may be only a fraction of a microsecond, or it may take several hundred microseconds, depending upon the point in the receiver at which blocking occurs. To detect a weak echo immediately following a strong one, the receiver must have a short BLOCKING RECOVERY TIME. The blocking itself must be minimized as much as possible. If a portion of the transmitted pulse leaks into the receiver input, then the receiver may be blocked and not show small, nearby objects. In most receivers, blocking is minimized from this cause by a duplexer. The duplexer protects the receiver by isolating it during the transmitted pulse.

RECEIVER BLOCK DIAGRAM

The SUPERHETERODYNE receiver is almost always used in microwave radar systems. A typical superheterodyne radar receiver is shown in figure 2-23. A receiver of this type meets all the requirements listed above. Signals from the antenna enter the receiver via the duplexer. A low-noise rf amplifier is usually the first stage of modern radar receivers. Some receivers, however, send the antenna signal directly to the mixer, as shown by the dashed path. The low-noise amplifiers used in modern systems are usually solid-state devices, such as tunnel-diode, parametric, or microwave transistor amplifiers.

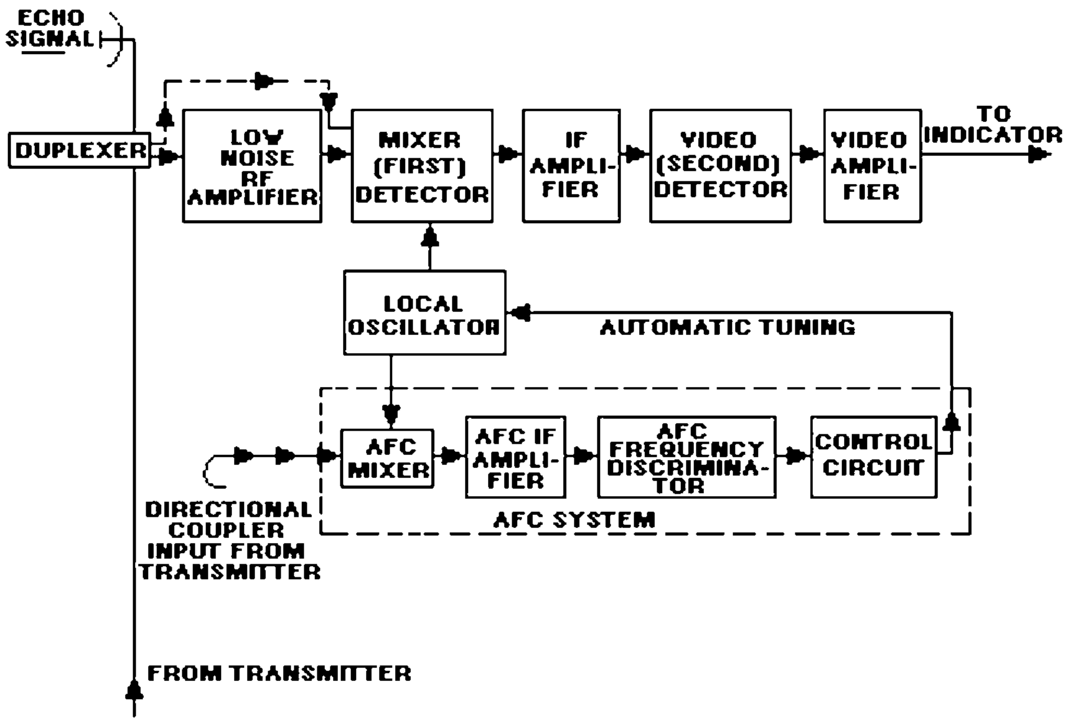


Figure 2-23.—Typical superheterodyne radar receiver.

The MIXER stage is often called the FIRST DETECTOR. The function of this stage is to convert the received rf energy to a lower, intermediate frequency (IF) that is easier to amplify and manipulate electronically. The intermediate frequency is usually 30 or 60 megahertz. It is obtained by heterodyning the received signal with a local-oscillator signal in the mixer stage. The mixer stage converts the received signal to the lower IF signal without distorting the data on the received signal.

After conversion to the intermediate frequency, the signal is amplified in several IF AMPLIFIER stages. Most of the gain of the receiver is developed in the IF amplifier stages. The overall bandwidth of the receiver is often determined by the bandwidth of the IF stages.

The output of the IF amplifiers is applied to the SECOND DETECTOR. It is then rectified and passed through one or more stages of amplification in the video amplifier(s). The output stage of the receiver is normally an emitter follower. The low-impedance output of the emitter follower matches the impedance of the cable. The video pulses are coupled through the cable to the indicator for video display on the crt.

As in all superheterodyne receivers, controlling the frequency of the local oscillator keeps the receiver tuned. Since this tuning is critical, some form of automatic frequency control (afc) is essential to avoid constant manual tuning. Automatic frequency control circuits mix an attenuated portion of the transmitted signal with the local oscillator signal to form an IF signal. This signal is applied to a frequency-sensitive discriminator that produces an output voltage proportional in amplitude and polarity to any change in IF frequency. If the IF signal is at the discriminator center frequency, no discriminator output occurs. The center frequency of the discriminator is essentially a reference frequency for the IF signal. The output of the DISCRIMINATOR provides a control voltage to maintain the local oscillator at the correct frequency.

Different receiving systems may vary in the type of coupling between stages, the type of mixer, the detector, the local oscillator, and the number of stages of amplification at the different frequencies. However, the receiver is always designed to have as little noise as possible. It is also designed to have sufficient gain so that noise, rather than lack of gain, limits the smallest visible signal.

RECEIVER COMPONENTS

This section will analyze in more detail the operation of the receiver circuits mentioned above. The circuits discussed are usually found in some form in all radar superheterodyne receivers.

Low-Noise Amplifier

LOW-NOISE AMPLIFIERS, sometimes called PREAMPS, are found in most modern radar receivers. As previously mentioned, these amplifiers are usually solid-state microwave amplifiers. The most common types are tunnel diode and parametric amplifiers. These amplifiers are discussed in detail in NEETS, Module 11, *Microwave Principles*. Some older systems may still use a traveling-wave tube (twt) as a low-noise first stage amplifier. However, the solid-state amplifiers produce lower noise levels and more gain.

Local Oscillator

Most radar receivers use a 30 or 60 megahertz intermediate frequency. The IF is produced by mixing a local oscillator signal with the incoming signal. The local oscillator is, therefore, essential to efficient operation and must be both tunable and very stable. For example, if the local oscillator frequency is 3,000 megahertz, a frequency change of 0.1 percent will produce a frequency shift of 3 megahertz. This is equal to the bandwidth of most receivers and would greatly decrease receiver gain.

The power output requirement for most local oscillators is small (20 to 50 milliwatts) because most receivers use crystal mixers that require very little power.

The local oscillator output frequency must be tunable over a range of several megahertz in the 4,000-megahertz region. The local oscillator must compensate for any changes in the transmitted frequency and maintain a constant 30 or 60 megahertz difference between the oscillator and the transmitter frequency. A local oscillator that can be tuned by varying the applied voltage is most desirable.

The REFLEX KLYSTRON is often used as a local oscillator because it meets all the requirements mentioned above. The reflex klystron is a very stable microwave oscillator that can be tuned by changing the repeller voltage.

Most radar systems use an automatic frequency control (afc) circuit to control the output of the local oscillator. A block diagram of a typical afc circuit is included in figure 2-23. Note that the afc circuits form a closed loop. This circuit is, in fact, often called the afc loop.

A sample of the transmitter energy is fed through the afc mixer and an IF amplifier to a discriminator. The output of the discriminator is a dc error voltage that indicates the degree of mistuning between the transmitter and the local oscillator. In this particular example let's assume that the IF is 30 megahertz. If the output of the mixer is correct, the discriminator will have no output. If the mixer output is above 30 megahertz, the output of the discriminator will be positive dc pulses; if the mixer output is below 30 megahertz, the discriminator output will be negative dc pulses. In either case, this output is fed through an amplifier to the control circuit. The control circuit adjusts the operating frequency of the local oscillator so that no mistuning exists and the IF is 30 megahertz. In this example the local oscillator is a reflex klystron and the control circuit provides he repeller plate voltage for the klystron; thus, the klystron directly controls the local oscillator frequency. In this manner the local oscillator is maintained exactly 30 megahertz below the transmitter frequency.

Q34. What is the greatest limiting factor in a receiver's detectable range?

Q35. What type of receiver is most often used in radar systems?

Q36. What IF frequencies are normally used in radar receivers?

Q37. Which component of the receiver produces the signal that is mixed with the received signal to produce the IF signal?

Mixer

Many older radar receivers do not use a low-noise amplifier as the receiver front end; they simply send the echo signal directly to a crystal mixer stage. A crystal is used rather than an electron-tube diode because, at microwave frequencies, the tube would generate excessive noise. Electron tubes are also limited by the effects of transit time at microwave frequencies. The crystal most commonly used is the point-contact crystal diode; however, recent developments in the field of solid-state microwave devices may soon replace the point-contact diode with devices that produce even less noise. The Schottky-barrier diode is an example of a relatively recent development that produces less noise than the point-contact crystal.

The simplest type of radar mixer is the SINGLE ENDED or UNBALANCED CRYSTAL MIXER, shown in figure 2-24. The mixer illustrated uses a tuned section of coaxial transmission line one-half wavelength long. This section matches the crystal to the signal echo and the local oscillator inputs. Local oscillator injection is accomplished by means of a probe. In the coaxial assembly, the signal is injected by means of a slot. This slot would normally be inserted in the duplexer waveguide assembly and be properly oriented to provide coupling of the returned signal. In this application, the unwanted signals at the output of the mixer (carrier frequency, the local oscillator frequency, and sum of these two signals) are effectively eliminated by a resonant circuit tuned to the intermediate, or difference frequency. One advantage of the unbalanced crystal mixer is its simplicity. It has one major disadvantage; its inability to cancel local oscillator noise. Difficulty in detecting weak signals will exist if noise is allowed to pass through the mixer along with the signal.

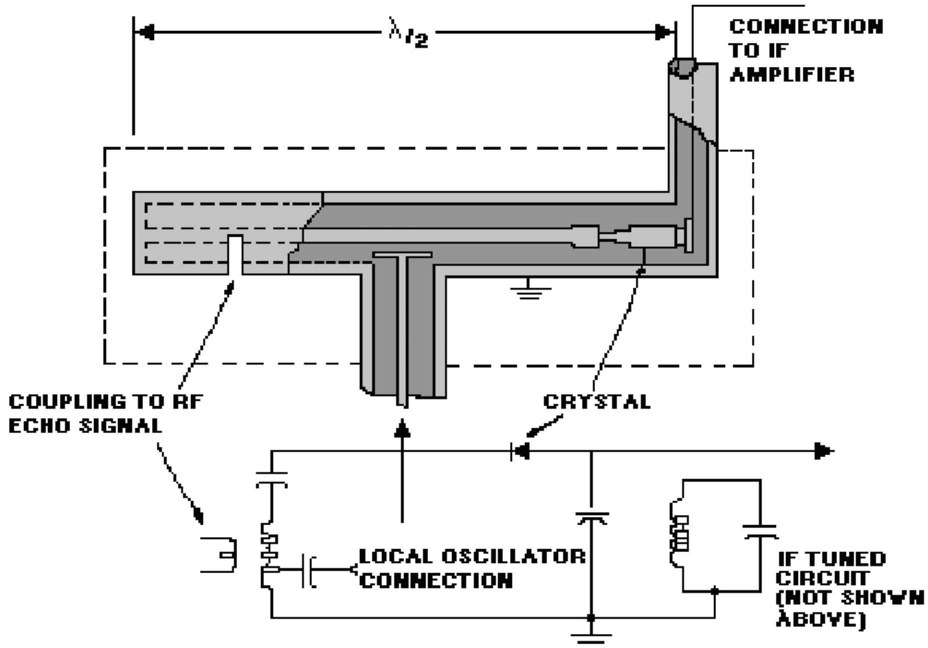


Figure 2-24.—Single-ended crystal mixer.

One type of mixer which cancels local oscillator noise is the BALANCED, OR HYBRID, MIXER (sometimes called the MAGIC T). Figure 2-25 shows this type of mixer. In hybrid mixers, crystals are inserted directly into the waveguide. The crystals are located one-quarter wavelength from their respective short-circuited waveguide ends (a point of maximum voltage along a tuned line). The crystals are also connected to a balanced transformer, the secondary of which is tuned to the desired IF. The local oscillator signal is introduced into the waveguide local oscillator arm and distributes itself as shown in view A of figure 2-26. Observe that the local oscillator signal is in phase across the crystals. In view B the echo signal is introduced into the echo signal arm of the waveguide and is out of phase across the crystals. The resulting fields are shown in view C.

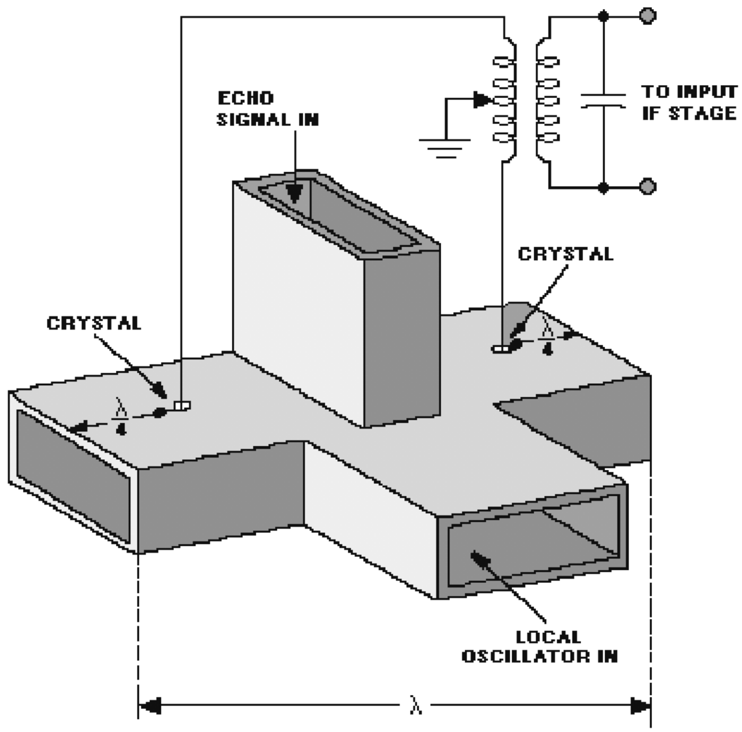


Figure 2-25.—Balanced (hybrid) crystal mixer.

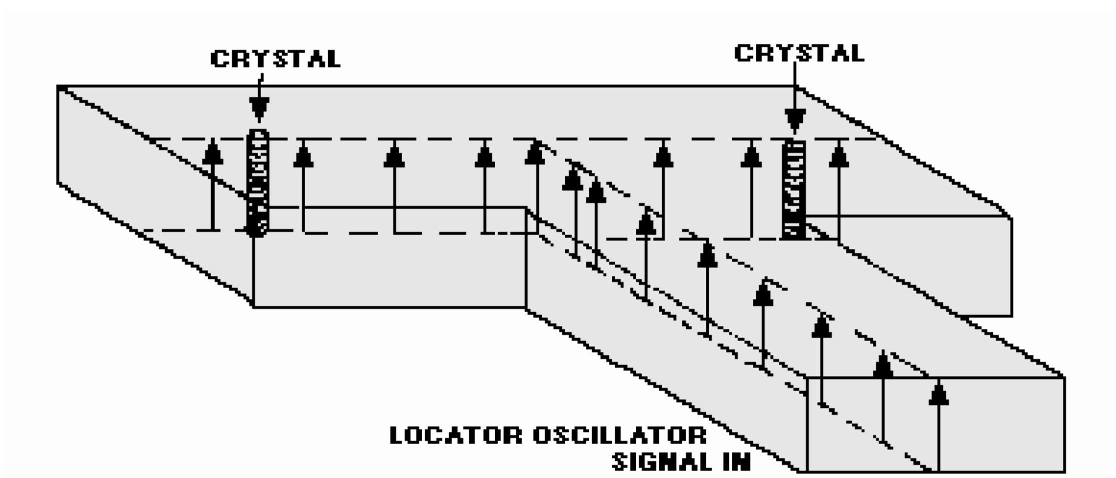


Figure 2-26A.—Balanced mixer fields. WAVEGUIDE AND LOCAL OSCILLATOR ARM.

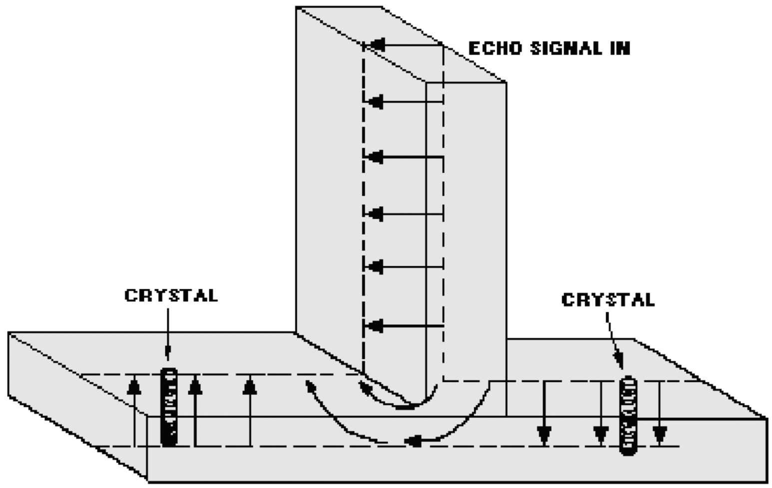


Figure 2-26B.—Balanced mixer fields. WAVEGUIDE AND ECHO SIGNAL ARM.

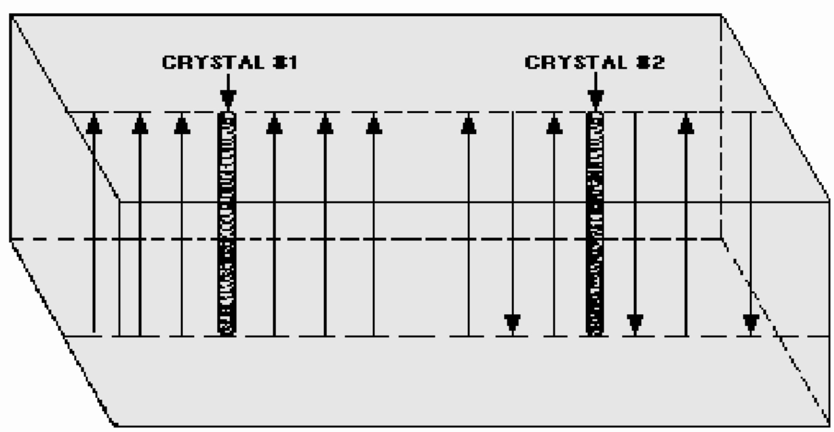


Figure 2-26C.—Balanced mixer fields. WAVEGUIDE.

A difference in phase exists between echo signals applied across the two crystals. The signal applied to the crystals from the local oscillator is in phase. Therefore, at some point both signals applied to crystal #1 will be in phase, and the signals applied to crystal #2 will be out of phase. This means that an IF signal of one polarity will be produced across crystal #1 and an IF signal of the opposite polarity will be produced across crystal #2. When these two signals are applied to the balanced output transformer (figure 2-25), they will add. Outputs of the same polarity will cancel across the balanced transformer.

This action eliminates the noise of the local oscillator. Noise components introduced from the local oscillator are in phase across the crystals and are, therefore, cancelled in the balanced transformer. The rf characteristics of the crystals must be nearly equal, or the noise of the local oscillator will not completely cancel. Note that only the noise produced by the local oscillator is canceled. Noise arriving with the echo signal is not affected.

IF Amplifier Stage

The IF AMPLIFIER SECTION of a radar receiver determines the gain, signal-to-noise ratio, and effective bandwidth of the receiver. The typical IF amplifier (commonly called an IF strip) usually contains from three to ten amplifier stages. The IF amplifier has the capability to vary both the bandpass and the gain of a receiver. Normally, the bandpass is as narrow as possible without affecting the actual signal energy. When a selection of pulse widths is available, such as short and long pulses, the bandpass must be able to match the bandwidth of the two different signals. Gain must be variable to provide a constant voltage output for input signals of different amplitudes. Figure 2-27 is a block diagram of an IF amplifier that meets these requirements.

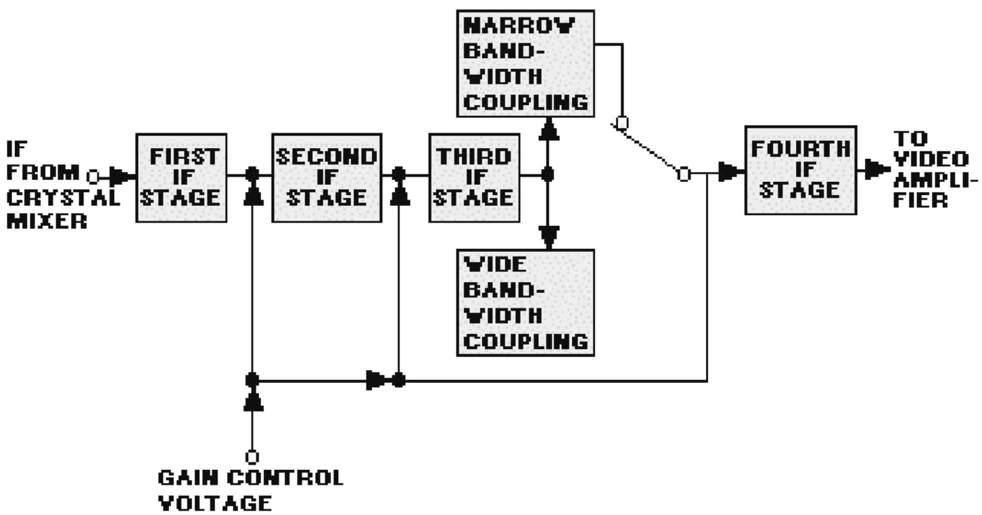


Figure 2-27.—IF amplifier block diagram.

The most critical stage of the IF section is the input (first stage). The quality of this stage determines the noise figure of the receiver and the performance of the entire receiving system with respect to detection of small objects at long ranges. Gain and bandwidth are not the only considerations in the design of the first IF stage. A consideration perhaps of more importance is noise generation. Noise generation in this stage must be low. Noise generated in the input IF stage will be amplified by succeeding stages and may exceed the echo signal in strength.

Detectors

The detector in a microwave receiver serves to convert the IF pulses into video pulses. After amplification, these are applied to the indicator. The simplest form of detector, and the one most commonly used in microwave receivers, is the DIODE DETECTOR.

A diode detector circuit is shown in view A of figure 2-28. The secondary of T1 and C1 form a tuned circuit that is resonant at the intermediate frequency. Should an echo pulse of sufficient amplitude be received, the voltage (e_i) developed across the tuned circuit is an IF pulse. Its shape is indicated by the dashed line in view B. Positive excursions of e_i cause no current to flow through the diode. However, negative excursions result in a flow of diode current and a subsequent negative voltage (e_o) to be developed across R1 and C2. Between peak negative voltage excursions of the e_i wave, capacitor C2 discharges through R1. Thus, the e_o waveform is a negative video pulse with sloping edges and superimposed IF ripple, as indicated by the solid line in view B. A negative polarity of the output pulse is ordinarily preferred, but a positive pulse may be obtained by reversing the connections of the diode. In view A, inductance L1, in combination with wiring capacitance and C2, forms a low-pass filter. This filter attenuates the IF components in the e_o waveform but results in a minimum loss of video high-frequency components.

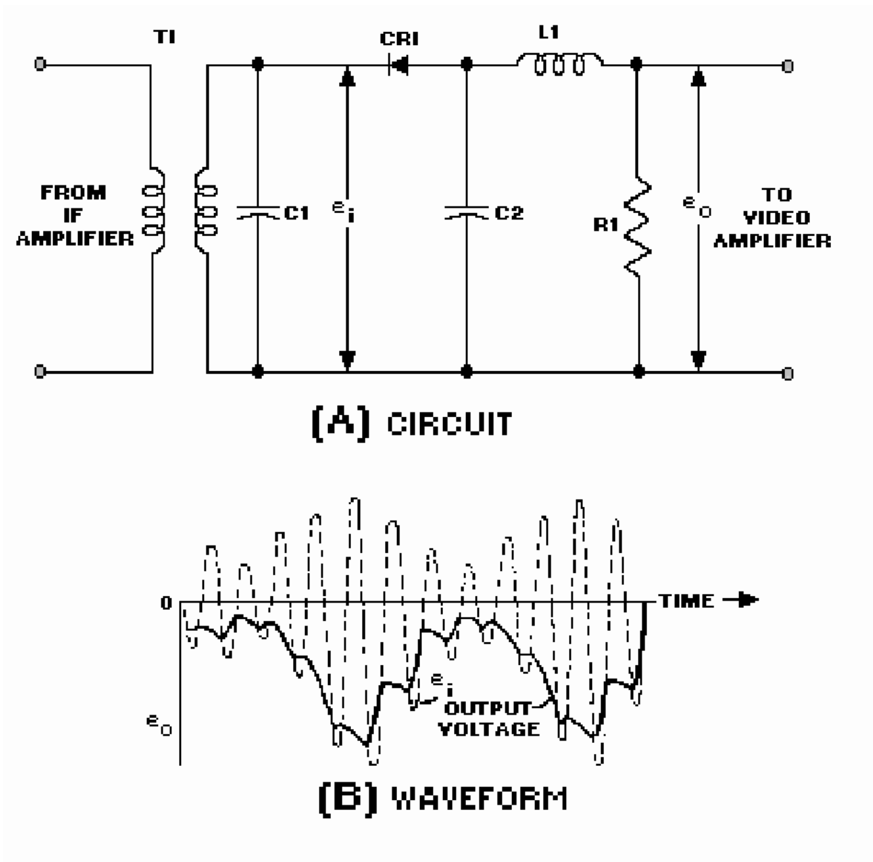


Figure 2-28.—Diode detector.

Video Amplifiers

The video amplifier receives pulses from the detector and amplifies these pulses for application to the indicating device. A video amplifier is fundamentally an RC coupled amplifier that uses high-gain transistors or pentodes. However, a video amplifier must be capable of a relatively wide frequency response. Stray and interelectrode capacitances reduce the high-frequency response of an amplifier, and the reactance of the coupling capacitor diminishes the low-frequency response. These problems are overcome by the use of FREQUENCY COMPENSATION NETWORKS in the video amplifier. The types of frequency compensation networks that may be used in a video amplifier stage are discussed in detail in NEETS, Module 8, *Introduction to Amplifiers*.

Q38. What receiver circuit actually produces the IF frequency?

Q39. The IF amplifiers are connected in what amplifier configuration?

Q40. Which receiver component converts the IF pulses to video pulses?

RECEIVER SPECIAL CIRCUITS

The performance efficiency of radar receivers is often greatly decreased by interference from one or more of several possible sources. Weather and sea return are the most common of these interference sources, especially for radar systems that operate above 3,000 megahertz. Unfavorable weather conditions can completely mask all radar returns and render the system useless. Electromagnetic interference from external sources, such as the deliberate interference by an enemy, called jamming or electronic counter measures (ECM), can also render a radar system useless. Many special circuits have been designed to help the radar receiver counteract the effects of external interference. These circuits are called VIDEO ENHANCEMENT FEATURES, ANTIJAMMING CIRCUITS, or ELECTRONIC COUNTER-COUNTERMEASURES (ECCM) CIRCUITS. This section will discuss, in general terms, some of the more common video enhancement features associated with radar receivers.

Automatic Gain Control (AGC)

Most radar receivers use some means to control the overall gain. This usually involves the gain of one or more IF amplifier stages. Manual gain control by the operator is the simplest method. Usually, some more complex form of automatic gain control (agc) or instantaneous automatic gain control (iagc) is used during normal operation. Gain control is necessary to adjust the receiver sensitivity for the best reception of signals of widely varying amplitudes. Agc and iagc circuits are designed with, a shut-off feature so that receiver gain may be adjusted manually. In this way, manual gain control can be used to adjust for best reception of a particular signal.

The simplest type of agc adjusts the IF amplifier bias (and gain) according to the average level of the received signal. Agc is not used as frequently as other types of gain control because of the widely varying amplitudes of radar return signals.

With agc, gain is controlled by the largest received signals. When several radar signals are being received simultaneously, the weakest signal may be of greatest interest. Iagc is used more frequently because it adjusts receiver gain for each signal.

The iagc circuit is essentially a wide-band, dc amplifier. It instantaneously controls the gain of the IF amplifier as the radar return signal changes in amplitude. The effect of iagc is to allow full amplification of weak signals and to decrease the amplification of strong signals. The range of iagc is limited, however, by the number of IF stages in which gain is controlled. When only one IF stage is controlled, the range of iagc is limited to approximately 20 dB. When more than one IF stage is controlled, iagc range can be increased to approximately 40 dB.

Sensitivity Time Control (STC)

In radar receivers, the wide variation in return signal amplitudes make adjustment of the gain difficult. The adjustment of receiver gain for best visibility of nearby target return signals is not the best adjustment for distant target return signals. Circuits used to adjust amplifier gain with time, during a single pulse-repetition period, are called stc circuits.

Sensitivity time-control circuits apply a bias voltage that varies with time to the IF amplifiers to control receiver gain. Figure 2-29 shows a typical stc waveform in relation to the transmitted pulse. When the transmitter fires, the stc circuit decreases the receiver gain to zero to prevent the amplification of any leakage energy from the transmitted pulse. At the end of the transmitted pulse, the stc voltage begins to rise, gradually increasing the receiver gain to maximum. The stc voltage effect on receiver gain is usually limited to approximately 50 miles. This is because close-in targets are most likely to saturate the receiver; beyond 50 miles, stc has no affect and the receiver operates normally.

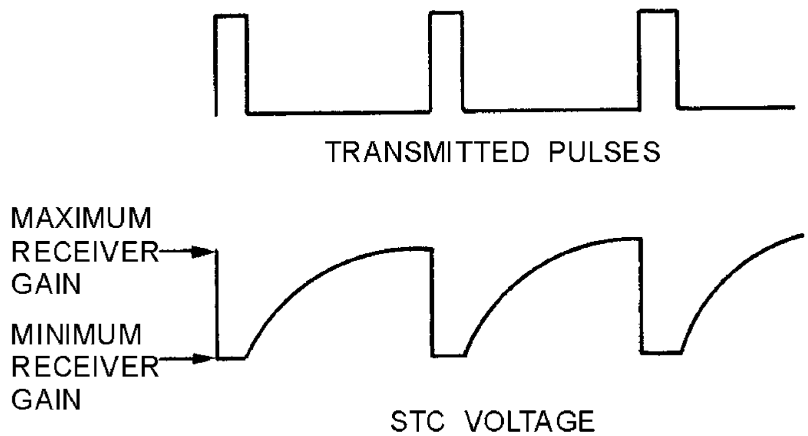


Figure 2-29.—Stc voltage waveform.

The combination of stc and iagc circuits results in better overall performance than with either type of gain control alone. Stc decreases the amplitude of nearby target return signals, while iagc decreases the amplitude of larger-than-average return signals. Thus, normal changes of signal amplitudes are adequately compensated for by the combination of iagc and stc.

Antijamming Circuits

Among the many circuits used to overcome the effects of jamming, two important ones are GATED AGC CIRCUITS and FAST-TIME-CONSTANT CIRCUITS. A gated agc circuit permits signals that occur only in a very short time interval to develop the agc. If large-amplitude pulses from a jamming transmitter arrive at the radar receiver at any time other than during the gating period, the agc does not respond to these jamming pulses.

Without gated agc, a large jamming signal would cause the automatic gain control to follow the interfering signal. This would decrease the target return signal amplitude to an unusable value. Gated agc produces an output signal for only short time periods; therefore, the agc output voltage must be averaged over several cycles to keep the automatic gain control from becoming unstable.

Gated agc does not respond to signals that arrive at times other than during the time of a target return signal. However, it cannot prevent interference that occurs during the gating period. Neither can gating the agc prevent the receiver from overloading because of jamming signal amplitudes far in excess of the target return signal. This is because the desired target is gated to set the receiver gain for a signal of that particular amplitude. As an aid in preventing radar receiver circuits from overloading during the reception of jamming signals, fast-time-constant coupling circuits are used. These circuits connect the video detector output to the video amplifier input circuit.

A fast-time-constant (ftc) circuit is a differentiator circuit located at the input of the first video amplifier. When a large block of video is applied to the ftc circuit, only the leading edge will pass. This is because of the short time constant of the differentiator. A small target will produce the same length of signal on the indicator as a large target because only the leading edge is displayed. The ftc circuit has no effect on receiver gain; and, although it does not eliminate jamming signals, ftc greatly reduces the effect of jamming.

Q41. Which of the two types of automatic gain control, agc or iagc, is most effective in radar use for the Navy?

Q42. Immediately after the transmitter fires, stc reduces the receiver gain to what level?

Q43. How does ftc affect receiver gain, if at all?

SPECIAL RECEIVERS

The basic receiver of a radar system often does not meet all the requirements of the radar system, nor does it always function very well in unfavorable environments. Several special receivers have been developed to enhance target detection in unfavorable environments or to meet the requirements of special transmission or scanning methods. A radar system with a moving target indicator (mti) system or a monopulse scanning system requires a special type of receiver. Other types of special receivers, such as the logarithmic receiver, have been developed to enhance reception during unfavorable conditions. These receivers will be discussed in general terms in this section.

Moving Target Indicator (mti) System

The MOVING TARGET INDICATOR (mti) system effectively cancels CLUTTER (caused by fixed unwanted echoes) and displays only moving target signals. Clutter is the appearance on a radar indicator of confusing, unwanted echoes which interfere with the clear display of desired echoes. Clutter is the result of echoes from land, water, weather, and so forth. The unwanted echoes can consist of GROUND CLUTTER (echoes from surrounding land masses), SEA CLUTTER (echoes from the irregular surface of the sea), or echoes from the clouds and rain. The problem is to find the desired echo in the midst of the clutter. To do this, the mti system must be able to distinguish between fixed and moving targets and then must eliminate only the fixed targets. This is accomplished by phase detection and pulse-to-pulse comparison.

Target echo signals from stationary objects have the same phase relationship from one receiving period to the next. Moving objects produce echo signals that have a different phase relationship from one receiving period to the next. This principle allows the mti system to discriminate between fixed and moving targets.

Signals received from each transmitted pulse are delayed for a period of time exactly equal to the pulse-repetition time. The delayed signals are then combined with the signals received from the next transmitted pulse. This is accomplished in such a manner that the amplitudes subtract from each other as shown in figure 2-30, views Aand B. Since the fixed targets have approximately the same amplitude on each successive pulse, they will be eliminated. The moving target signals, however, are of different amplitudes on each successive pulse and, therefore, do not cancel. The resulting signal is then amplified and presented on the indicators.

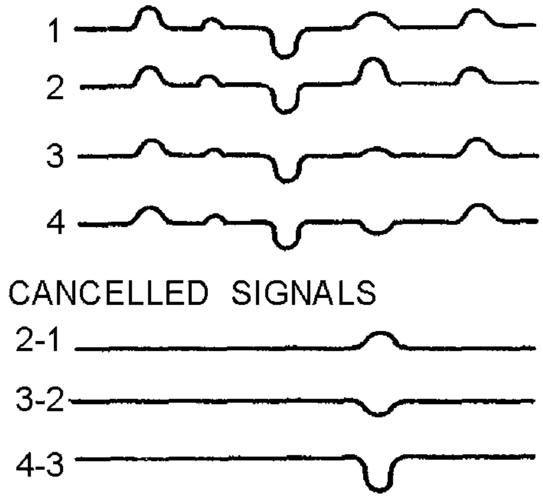


Figure 2-30A.—Fixed target cancellation.

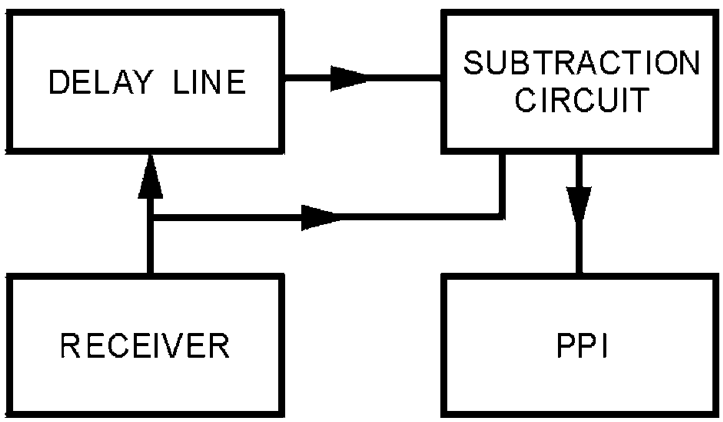


Figure 2-30B.—Fixed target cancellation.

In figure 2-31, 30-megahertz signals from the signal mixer are applied to the 30-megahertz amplifier. The signals are then amplified, limited, and fed to the phase detector. Another 30-megahertz signal, obtained from the coherent oscillator (coho) mixer, is applied as a lock pulse to the coho. The coho lock pulse is originated by the transmitted pulse. It is used to synchronize the coho to a fixed phase relationship with the transmitted frequency at each transmitted pulse. The 30-megahertz, cw reference signal output of the coho is applied, together with the 30-megahertz echo signal, to the phase detector.

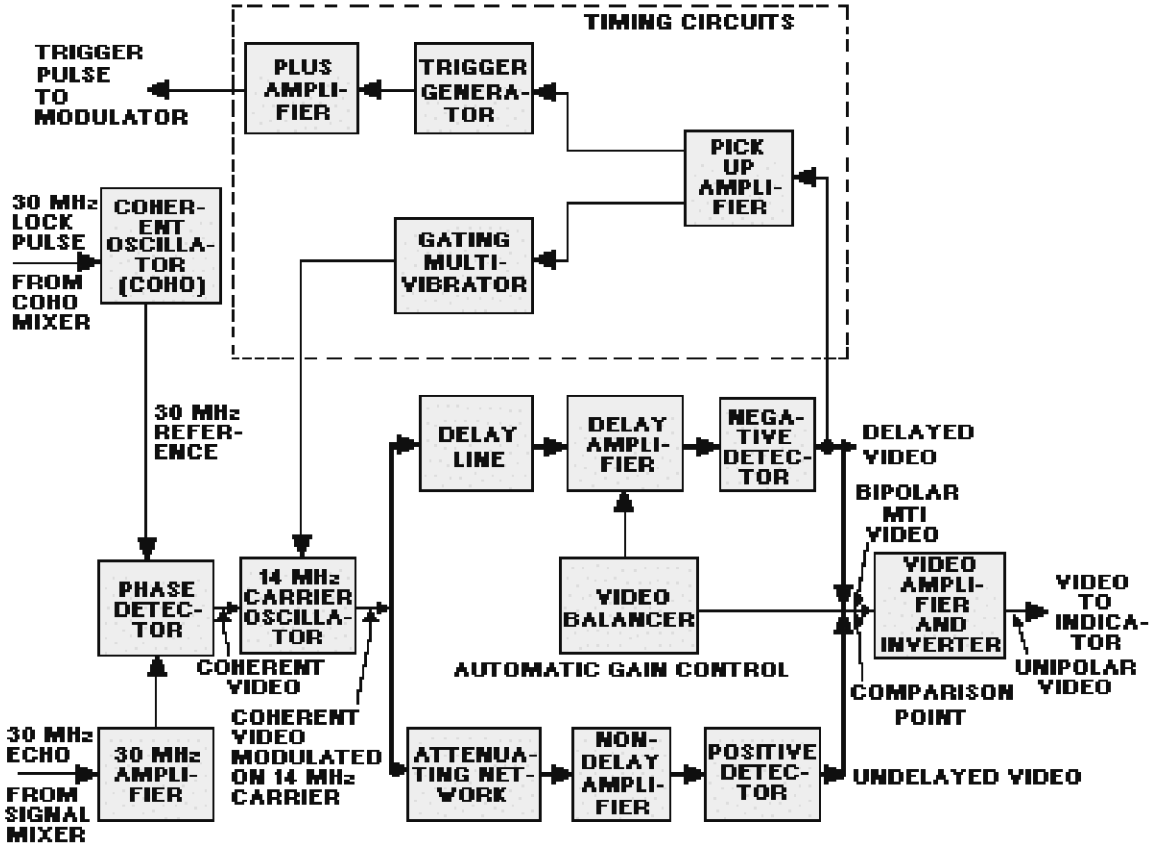


Figure 2-31.—Mti block diagram.

The phase detector produces a video signal. The amplitude of the video signal is determined by the phase difference between the coho reference signal and the IF echo signals. This phase difference is the same as that between the actual transmitted pulse and its echo. The resultant video signal may be either positive or negative. This video output, called coherent video, is applied to the 14-megahertz cw carrier oscillator.

The 14-megahertz cw carrier frequency is amplitude modulated by the phase-detected coherent video. The modulated signal is amplified and applied to two channels. One channel delays the 14-megahertz signal for a period equal to the time between transmitted pulses. The signal is then amplified and detected. The delay required (the period between transmitted pulses) is obtained by using a mercury delay line or a fused-quartz delay line, which operates ultrasonically at 14 megahertz.

The signal to the other channel is amplified and detected with no delay introduced. This channel includes an attenuating network that introduces the same amount of attenuation as does the delay line in the delayed video channel. The resulting nondelayed video signal is combined in opposite polarity with the delayed signal. The amplitude difference, if any, at the comparison point between the two video signals is amplified; because the signal is bipolar, it is made unipolar. The resultant video signal, which represents only moving targets, is sent to the indicator system for display.

An analysis of the mti system operation just described shows that signals from fixed targets produce in the phase detector recurring video signals of the same amplitude and polarity. (Fixed targets have an unchanging phase relationship to their respective transmitted pulses.) Thus, when one video pulse is combined with the preceding pulse of opposite polarity, the video signals cancel and are not passed on to the indicator system.

Signals from *moving targets*, however, will have a varying phase relationship with the transmitted pulse. As a result, the signals from successive receiving periods produce signals of different amplitudes in the phase detector. When such signals are combined, the difference in signal amplitude provides a video signal that is sent to the indicator system for display.

The timing circuits, shown in figure 2-31, are used to accurately control the transmitter pulse-repetition frequency to ensure that the pulse-repetition time remains constant from pulse to pulse. This is necessary, of course, for the pulses arriving at the comparison point to coincide in time and achieve cancellation of fixed targets.

As shown in figure 2-31, a feedback loop is used from the output of the delay channel, through the pickoff amplifier, to the trigger generator and gating multivibrator circuits. The leading edge of the square wave produced by the detected carrier wave in the delayed video channel is differentiated at the pickoff amplifier. It is used to activate the trigger generator and gating multivibrator. The trigger generator sends an amplified trigger pulse to the modulator, causing the radar set to transmit.

The gating multivibrator is also triggered by the negative spike from the differentiated square wave. This stage applies a 2,000-microsecond negative gate to the 14-megahertz oscillator. The oscillator operates for 2,400 microseconds and is then cut off. Because the delay line time is 2,500 microseconds, the 14-megahertz oscillations stop before the initial waves reach the end of the delay line. This wave train, when detected and differentiated, turns the gating multivibrator on, producing another 2,400-microsecond wave train. The 100 microseconds of the delay line is necessary to ensure that the mechanical waves within the line have time to damp out before the next pulse-repetition time. In this manner the pulse-repetition time of the radar set is controlled by the delay of the mercury, or quartz delay line. Because this delay line is also common to the video pulses going to the comparison point, the delayed and the undelayed video pulses will arrive at exactly the same time.

Q44. What type of target has a fixed phase relationship from one receiving period to the next?

Q45. What signal is used to synchronize the coherent oscillator to a fixed phase relationship with the transmitted pulse?

Q46. What is the phase relationship between the delayed and undelayed video?

Logarithmic Receiver

The LOGARITHMIC RECEIVER uses a linear logarithmic amplifier, commonly called a LIN-LOG AMPLIFIER, instead of a normal IF amplifier. The lin-log amplifier is a nonsaturating amplifier that does not ordinarily use any special gain-control circuits. The output voltage of the lin-log amplifier is a linear function of the input voltage for low-amplitude signals. It is a logarithmic function for high-amplitude signals. In other words, the range of linear amplification does not end at a definite saturation point, as is the case in normal IF amplifiers. The comparison of the response curves for normal IF and lin-log amplifiers is shown in figure 2-32. The curves show that a continued increase in the input to the lin-log amplifier causes a continued increase in the output, but at a reduced rate. Therefore, a large signal does not saturate the lin-log amplifier; rather, it merely reduces the amplification of a simultaneously applied small signal. A small echo signal can often be detected by the lin-log receiver when a normal receiver would be saturated.

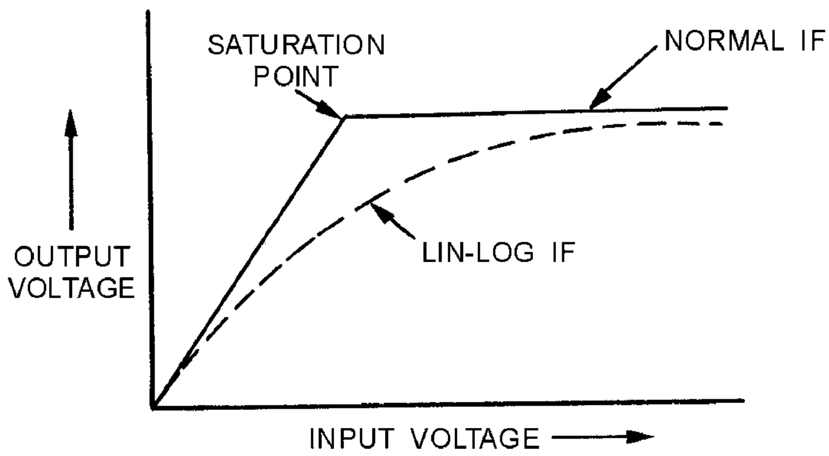


Figure 2-32.—Lin-Log amplifier versus normal IF amplifier.

A typical circuit for obtaining a lin-log response is shown in figure 2-33. If detectors 2 and 3 were not present, the output voltage would be limited by the saturation point of the final IF stage, as it is in a normal IF section. However, when the final stage of the lin-log is saturated, larger signals cause an increase in the output of the next to last stage. This increase is detected by detector 2 and summed with the output of detector 1. This sum produces an increase in the output even though the final stage is saturated. Detector 3 causes the output to continue to increase after the second stage saturates. The overall gain becomes less and less as each stage saturates, but some degree of amplification is still available. The proper choice of IF stage gains and saturation points produces an approximately logarithmic response curve.

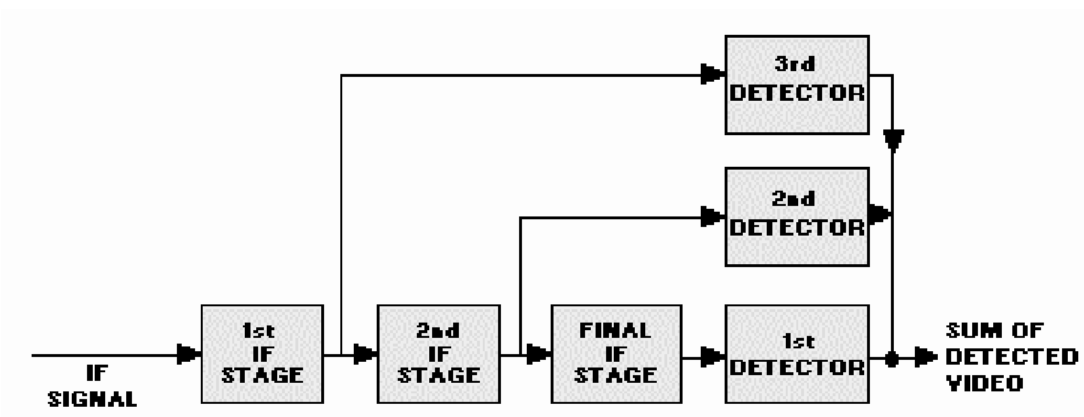


Figure 2-33.—Lin-Log receiver block diagram.

Figure 2-34, shows the response curves of the three IF stages in the lin-log amplifier shown in figure 2-33. The responses of the individual stages produce a segmented overall response curve for the receiver.

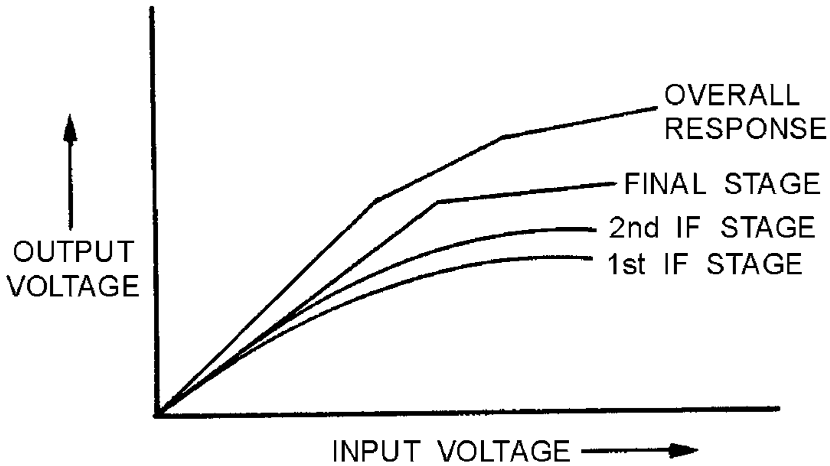


Figure 2-34.—Lin-Log amplifier stage response curves.

Monopulse Receiver

The most common of the automatic tracking radars is the MONOPULSE RADAR. The monopulse radar obtains the three target position coordinates of range, bearing, and elevation from a single pulse. The receiver for a monopulse radar must have three separate channels to process range, bearing, and elevation information. The block diagram of a simplified monopulse receiver is shown in figure 2-35.

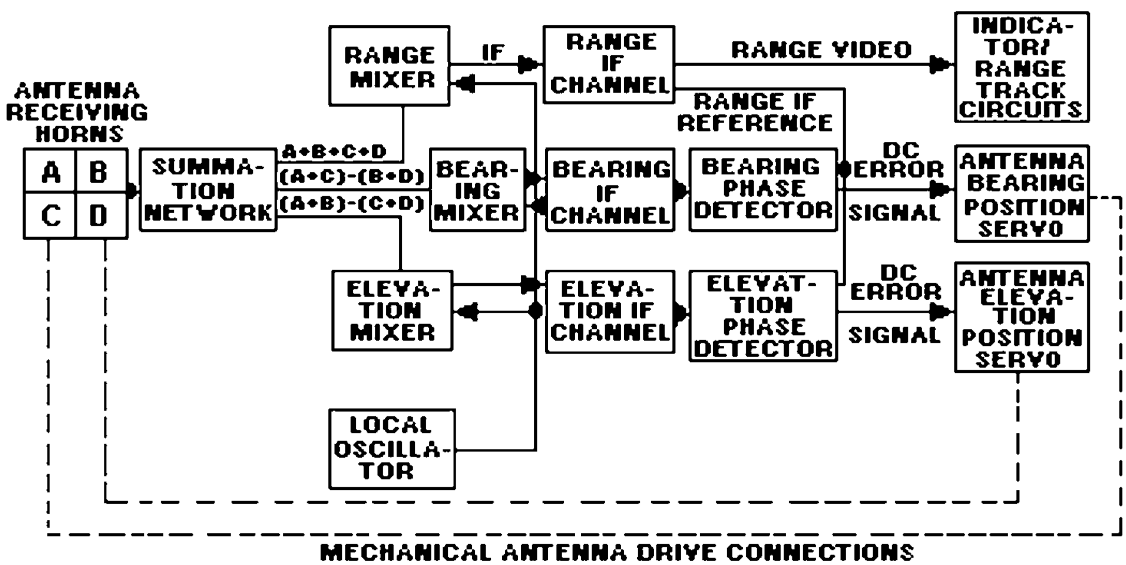


Figure 2-35.—Monopulse receiver block diagram.

As in a conventional receiver, each channel of the monopulse receiver converts the return echo to an IF frequency by mixing the returned signal with a common local oscillator signal. The sum of the energy from all four return signals is mixed with the local oscillator signal to produce range IF information. Bearing information is obtained by subtracting the energy from horns B and D from the energy from horns A and C:

$$(A + C) - (B + D)$$

and mixing the difference with the local oscillator signal. The result is a bearing IF signal. Elevation information is obtained in the same way, except the energy from horns C and D is subtracted from the energy from horns A and B:

$$(A + B) - (C + D)$$

If the target is on the elevation and bearing axis, the summations will both be zero; therefore, neither the bearing nor elevation channels will receive an input signal. If either of the bearing or elevation signals is off the axis, an input to the IF channel is produced. This input is subsequently converted to an IF signal in the appropriate channel.

The major difference between the monopulse receiver and the conventional receiver is the requirement for a dc error voltage output from the bearing and elevation channels. The range channel of a monopulse receiver is sent to a conventional ranging circuit for presentation are on an indicator or for use by a range-tracking circuit. However, since most monopulse radars are automatic tracking radars, the outputs of the bearing and elevation channels must be converted to dc error signals for use by automatic bearing and elevation tracking systems. The dc error voltages are applied to the antenna bearing and elevation positioning servos. These servos reposition the antenna until the errors are nulled.

The phase detectors compare the phase of the bearing and elevation IF with a reference IF from the range channels. This *comparison* produces the dc error pulses needed to drive the antenna servos. The signals from both the bearing and elevation channels are the result of a summation process. They can be either positive (in-phase) or negative (180-degrees out of phase) when compared to the reference IF signal. For example, if the output of horns A and C is smaller than the output of horns B and D, a negative or 180-degree-out-of-phase signal is produced by the bearing channel (A + C) - (B + D). If output A + C is greater than output B + D, a positive or in-phase signal is produced by the bearing channel. The phase of the bearing and elevation output signals determines the direction in which the antenna moves; the magnitude of the signal determines the amount of movement. Since two signals must be present at the phase detector to produce an output, an error signal occurs only when a return echo is not on the antenna beam axis.

This technique produces an error signal when the target moves off the radiated beam axis in either bearing or elevation. The error signal causes the antenna to move in the proper direction and for the proper duration to cancel the error signal. This method of automatic tracking is commonly used by weapons-control tracking radar systems.

Q47. When a large signal and a small signal are applied to a lin-log amplifier at the same time, what is the effect on the small signal?

Q48. What happens to the overall gain of a lin-log amplifier as each stage saturates?

Q49. A monopulse receiver has how many separate channels?

Q50. If a target is on the bearing axis of the radiated beam, what is the input to the bearing IF channel?

Q51. What characteristic of the bearing and elevation output signals determines the direction of antenna movement?

SUMMARY

The following paragraphs summarize the important points of this chapter.

The **SYNCHRONIZER** is essential to any radar because it controls and times the operation of the entire system. Radar systems may be self-synchronized by triggers from the transmitter or externally synchronized from a master oscillator.

Most modern systems are synchronized by a **MASTER OSCILLATOR**, which may be a SINEWAVE OSCILLATOR, an ASTABLE MULTIVIBRATOR, or a BLOCKING OSCILLATOR.

Each of these oscillators fulfills the basic requirements of a synchronizer, which must be:

- free running
- stable in frequency
- frequency variable (in steps)

The **TRANSMITTER** produces the short-duration, high-power, rf pulses of energy that are radiated into space by the antenna.

The **MODULATOR** controls the radar pulse, width and amplitude.

KEYED-OSCILLATOR TRANSMITTERS produce a high-power output pulse by keying a high-power oscillator, such as a MAGNETRON.

POWER-AMPLIFIER TRANSMITTERS amplify a low-level pulse to the desired power level using a series of microwave amplifiers such as TRAVELING-WAVE TUBES or KLYSTRONS.

The **DUPLEXER** is a device that allows the same antenna to both transmit and receive. Most duplexers use the impedance characteristics of transmission lines and waveguides in conjunction with TR and ATR tubes to route the energy to the correct place. One of the most important functions of the duplexer is isolation of the receiver during transmission.

The **RECEIVER** detects the very small target return echo and amplifies it to a usable level for display on the indicator.

A typical **SUPERHETERODYNE RECEIVER** consists of a low-noise amplifier, a mixer, a local oscillator, an IF amplifier, a detector, and a video amplifier.

Some special purpose receivers are the MOVING TARGET INDICATOR and MONOPULSE RECEIVERS.

ANSWERS TO QUESTIONS Q1. THROUGH Q51.

A1. Controls system operation and timing.

A2. Timing and control.

A3. Transmitter.

A4. Free-running.

A5. The master oscillator.

A6. Leakage from the duplexer.

A7. Sine-wave oscillator, single-swing blocking oscillator, and master-trigger (astable) multivibrator.

A8. It requires additional shaping circuits.

A9. Blocking oscillators.

A10. Keyed oscillator and power-amplifier chain.

A11. The modulator.

A12. Steep leading and trailing edges.

A13. Line-pulsed.

A14. Capacitor, artificial transmission line, or pulse-forming network.

A15. Pulse width.

A16. Thyratron.

A17. The charging impedance.

A18. 600-30,000 megahertz.

A19. Mode skipping and mode shifting.

A20. The magnetron will not oscillate.

A21. ±5 percent.

A22. Frequency stability.

A23. Local oscillator and coherent oscillator.

A24. Multicavity klystron.

A25. Frequency synthesizer.

A26. Oscillations at an undesired frequency.

A27. Electronic.

A28. Tr tube.

A29. Apply keep-alive voltage.

A30. Quarter-wavelength section.

A31. Transmit.

A32. Neither fires.

A33. 180 degrees out of phase.

A34. Noise.

A35. Superheterodyne.

A36. Thirty or sixty megahertz.

A37. Local oscillator.

A38. Mixer.

A39. Cascade.

A40. Detector.

A41. IAGC.

A42. Zero.

A43. FTC has no effect on receiver gain.

A44. Stationary.

A45. Coho lock pulse.

A46. Opposite.

A47. Amplification is reduced.

A48. Decreases.

A49. Three.

A50. Zero.

A51. Phase.

CHAPTER 3

RADAR INDICATORS AND ANTENNAS

LEARNING OBJECTIVES

Upon completion of this chapter, the student will be able to:

1. Describe the purpose of the A scope, the range-height indicator (rhi), and the plan position indicator (ppi).
2. State the relationship between range and sweep speed and length on a radar indicator.
3. Explain the purpose of timing triggers, video, and antenna position inputs to a radar indicator.
4. List the major units of a ppi and describe their functions.
5. Describe the basic operation of sweep deflection and sweep rotation in a ppi.
6. List and describe the operation of the three range measurement circuits.
7. Describe antenna directivity and power gain characteristics.
8. Describe the focusing action of a basic parabolic antenna.
9. Describe the basic radiation patterns of the most common parabolic reflectors.
10. Describe the basic characteristics of horn radiators.

INTRODUCTION

Radar systems require an antenna to both transmit and receive radar energy and an indicator system to display the video information generated. This chapter will briefly describe some commonly used indicators and antenna systems. Antenna systems are described in more detail in NEETS, Module 10, *Introduction to Wave Generation, Transmission Lines, and Antennas*, and Module 11, *Microwave Principles*.

RADAR INDICATORS

The information available from a radar receiver may contain as many as several million separate data bits per second. From these and other data, such as the orientation of the antenna, the indicator should present to the observer a continuous, easily understandable, graphic picture of the relative position of radar targets. It should provide size, shape, and insofar as possible, indications of the type of targets. A cathode-ray tube (crt) fulfills these requirements to an astonishing degree. The cathode-ray tube's principal shortcoming is that it cannot present a true three-dimensional picture.

The fundamental geometrical quantities involved in radar displays are the RANGE, AZIMUTH ANGLE (or BEARING), and ELEVATION ANGLE. These displays relate the position of a radar target to the origin at the antenna. Most radar displays include one or two of these quantities as coordinates of the crt face.

The actual range of a target from the radar, whether on the ground, in the water, or in the air is known as SLANT RANGE. The majority of displays use as one coordinate the value of slant range, its horizontal projection (GROUND RANGE), or its vertical projection (ALTITUDE). Since slant range is involved in every radar situation, it inevitably appears in at least one display on every set. Slant range is the coordinate that is duplicated most often when more than one type of display is used. This is partly because displays presenting range have the highest signal-to-noise discrimination and partly for geometrical reasons.

Range is displayed by means of a linear time-base sweep, starting from a given point or line at a definite time in each pulse cycle. Thus, distances along this range sweep actually measure slant range. The sweep speed determines the scale factor, which relates the distance on the tube to actual range. The sweep length is the total distance represented. Distances are expressed in miles (statute or nautical) or yards. The origin of the range sweep may be on or off the tube face.

The angle at which the antenna is pointing, either in azimuth or elevation, may provide two-dimensional information in the display; that is, range and azimuth, or range and elevation.

A radar indicator, sometimes called a radar repeater, acts as the master timing device in *analyzing* the return of the video in a radar system. It also provides that capability to various other locations physically remote from the radar system. Each indicator should have the ability to select the outputs from any desired radar system aboard the ship. This is usually accomplished by the use of a RADAR DISTRIBUTION SWITCHBOARD. The switchboard contains a switching arrangement that has inputs from each radar system aboard ship and provides outputs to each repeater. The radar desired is selected by means of a selector switch on the repeater. For the repeater to present correct target position data, the indicator must have the following three inputs from the selected radar:

1. *Trigger timing pulses*. These pulses ensure that the sweep on the repeater starts from its point of origin each time the radar transmits. As discussed earlier, the repeater displays all targets at their actual range from the ship based on the time lapse between the instant of transmission and the instant the target's echo is received.

2. *The returning echo*. The echo, in rf form, is detected (converted to a video signal) by the radar receiver and applied to the repeater.

3. *Antenna information*. The angular sweep position of a plan position indicator (ppi) repeater must be synchronized to the angular position of the radar antenna to display target bearing (azimuth) information.

The three most common types of displays, called scopes, are the A-scope, the RANGE-HEIGHT INDICATOR (RHI) SCOPE, and the PLAN POSITION INDICATOR (PPI) SCOPE. The primary function of these displays will be discussed in this section. However, detailed descriptions will be limited to the ppi scope, which is the most common display.

THE A SCOPE

The A-scope display, shown in figure 3-1, presents only the range to the target and the relative strength of the echo. Such a display is normally used in weapons control radar systems. The bearing and elevation angles are presented as dial or digital readouts that correspond to the actual physical position of the antenna.

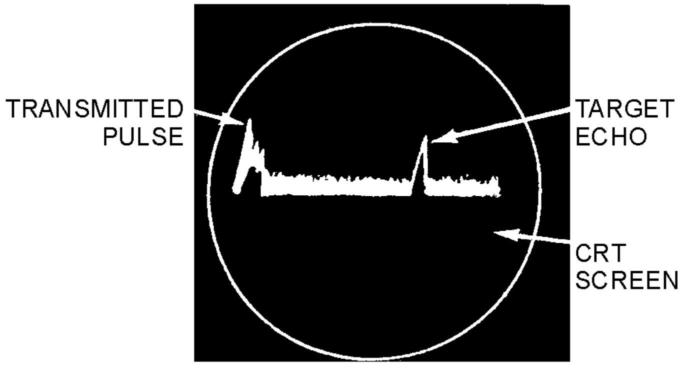


Figure 3-1.—A scope.

The A-scope normally uses an electrostatic-deflection crt. The sweep is produced by applying a sawtooth voltage to the horizontal deflection plates. The electrical length (time duration) of the sawtooth voltage determines the total amount of range displayed on the crt face.

The ranges of individual targets on an A-scope are usually determined by using a movable range gate or step that is superimposed on the sweep. Ranging circuits will be discussed in more detail later in this section.

RANGE-HEIGHT INDICATOR (RHI)

The range-height indicator (rhi) scope, shown in figure 3-2, is used with height-finding search radars to obtain altitude information. The rhi is a two-dimensional presentation indicating target range and altitude. The sweep originates in the lower left side of the scope. It moves across the scope, to the right, at an angle that is the same as the angle of transmission of the height-finding radar. The line of sight to the horizon is indicated by the bottom horizontal line. The area directly overhead is straight up the left side of the scope. Target echoes are displayed on the scope as vertical PIPS or BLIPS (spots of increased intensity that indicate a target location). The operator determines altitude by adjusting a movable height line to the point where it bisects the center of the blip. Target height is then read directly from an altitude dial or digital readout. Vertical range markers are also provided to estimate target range.

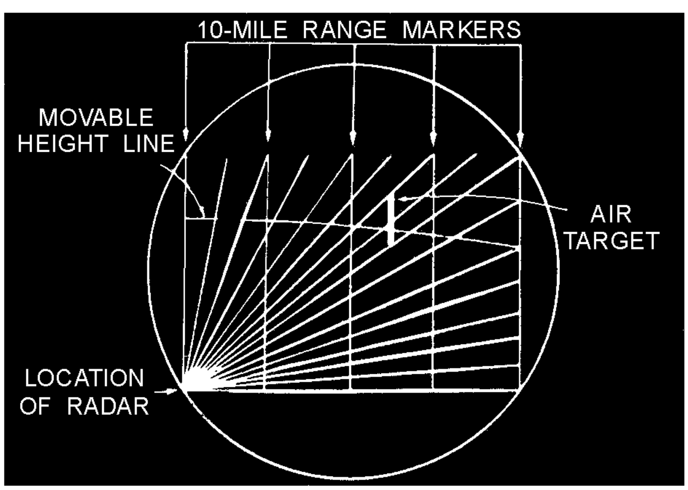


Figure 3-2.—RHI scope.

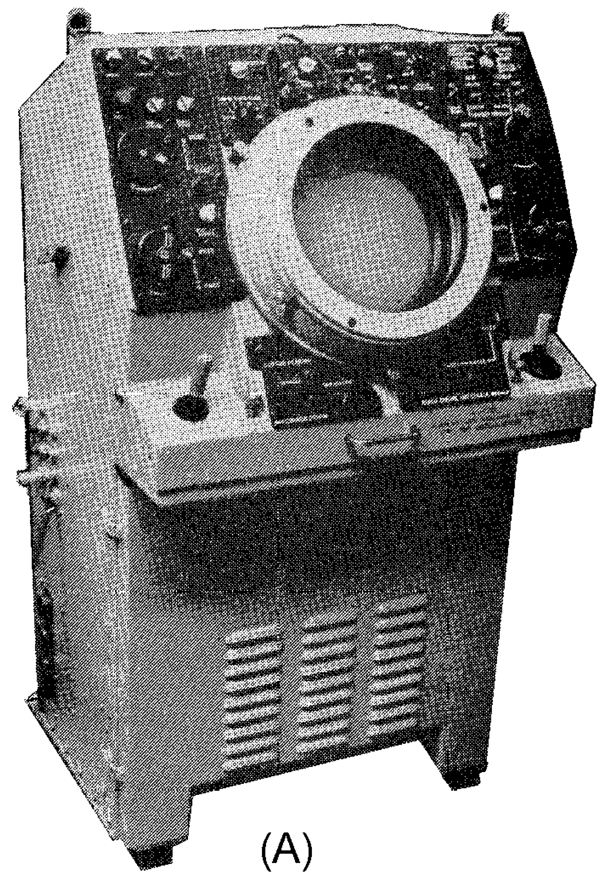
Q1. What are the three fundamental quantities involved in radar displays?

Q2. What are the required radar inputs for proper indicator operation?

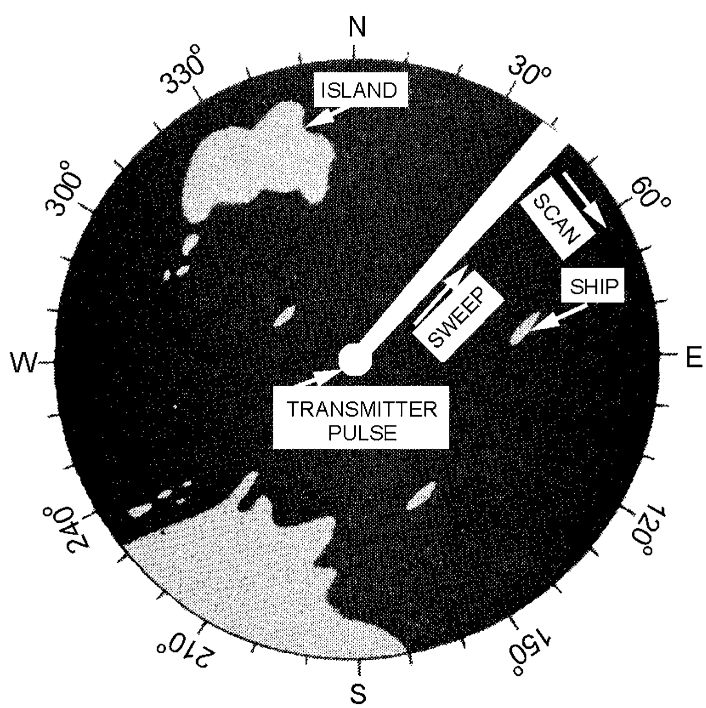
Q3. What coordinates are displayed on an rhi scope?

PLAN POSITION INDICATOR (PPI).

The ppi scope shown in figure 3-3, is by far the most used radar display. It is a polar coordinate display of the area surrounding the radar platform. Own ship is represented as the origin of the sweep, which is normally located in the center of the scope, but may be offset from the center on some sets. The ppi uses a radial sweep pivoting about the center of the presentation. This results in a map-like picture of the area covered by the radar beam. A long-persistence screen is used so that the display remains visible until the sweep passes again.



(A)



(B)

Figure 3-3.—PPI scope.

Bearing to the target is indicated by the target's angular position in relation to an imaginary line extending vertically from the sweep origin to the top of the scope. The top of the scope is either true north (when the indicator is operated in the true bearing mode) or ship's heading (when the indicator is operated in the relative bearing mode).

PPI Block Diagram

The basic block diagram, figure 3-4, illustrates the major units of a plan position indicator. Synchronization of events is particularly important in the presentation system. At the instant a radar transmitter fires (or at some predetermined time thereafter), circuits which control the presentation on the indicator must be activated. These events must be performed to a high degree of accuracy to ensure accurate range determination. The synchronization of these events is provided by the gate circuit.

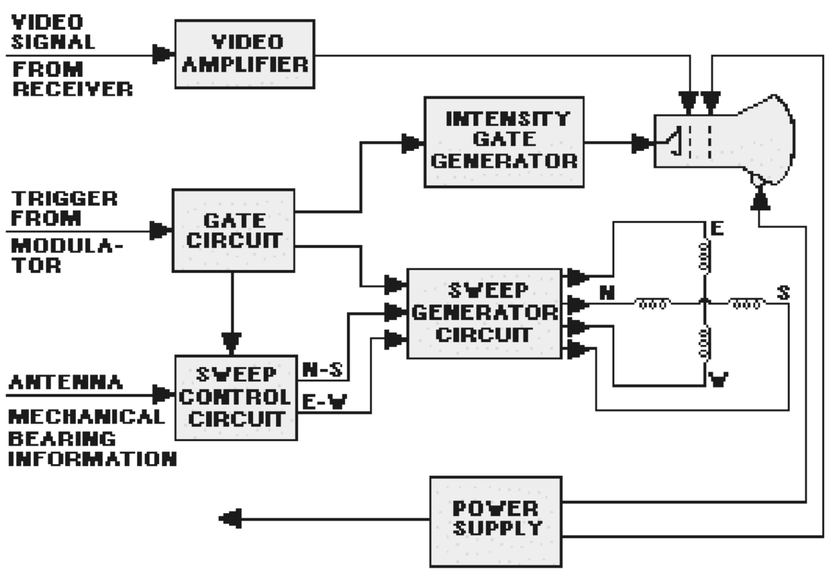


Figure 3-4.—Basic ppi block diagram.

GATE CIRCUIT.—The gate circuit develops pulses which synchronize the indicator with the transmitter. The gate circuit itself is synchronized by trigger pulses from the synchronizer. It then provides timing for the intensity gate generator, sweep generator circuit, and the sweep control circuit.

SWEEP CONTROL CIRCUIT.—The sweep control circuit converts mechanical bearing information from the antenna into voltages which control sweep circuit azimuth.

SWEEP GENERATOR CIRCUIT.—The sweep generator circuit produces currents which deflect an electron beam across the crt. Varying voltages from the sweep control circuit are applied to deflection coils. Gate voltages determine sweep rate, and therefore, the effective distance (range) covered by each sweep. Sweep potentials consist of separate north-south and east-west voltages; the amplitudes of these voltages determine sweep azimuth. The sweep generator is synchronized by an input from the gate circuit.

INTENSITY GATE GENERATOR.—The intensity gate generator provides a gate which unblanks the crt during sweep periods. The intensity of the trace appearing on the crt is determined by the dc level of this gate. This circuit is also synchronized by the gate circuit.

VIDEO AMPLIFIER.—The video amplifier circuit amplifies the video signal from the receiver and applies it to the crt intensity-modulating element (control grid).

POWER SUPPLY.—The power supply produces all voltages needed to operate the indicator. It also includes protective devices and metering circuits.

Although not shown in the basic block diagram, many indicators contain circuits which aid in range and bearing determination. These circuits are also synchronized by the gate circuit.

Sweep Deflection

In modern indicator systems, electromagnetic deflection of the crt electron beam is preferred to electrostatic deflection. Reasons for this choice are (1) increased control of the beam, (2) improved deflection sensitivity, (3) better beam position accuracy, and (4) simpler construction of the crt.

The primary difference between electromagnetic and electrostatic cathode-ray tubes lies in the method of controlling deflection and focusing of the electron beam. Both types employ electron guns and use electrostatic fields to accelerate and control the flow of electrons. The physical construction of a crt employing electromagnetic deflection is similar to an electrostatic type. The construction of a crt employing electromagnetic deflection is shown in figure 3-5.

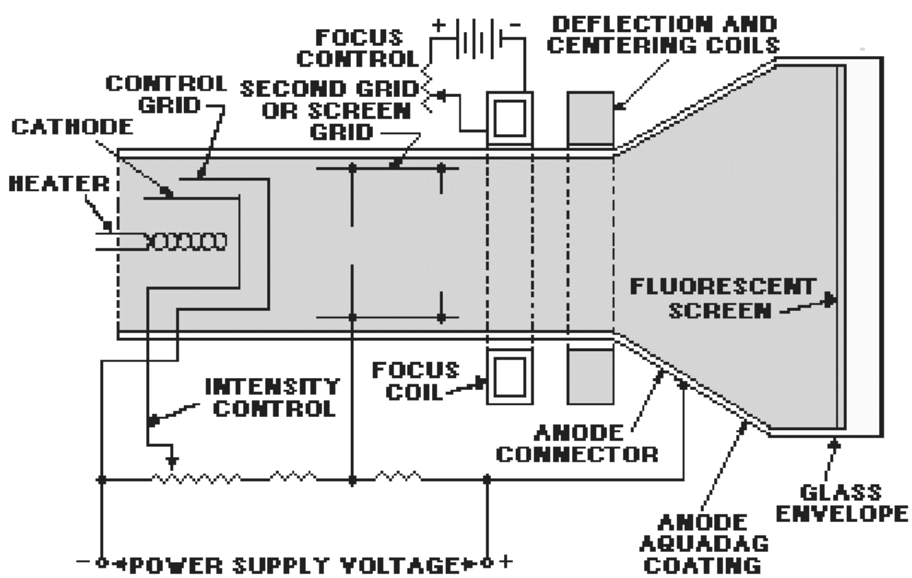


Figure 3-5.—Electromagnetic crt construction.

The electron gun in figure 3-5 is made up of a heater, cathode, control grid, second or screen grid, focus coil, and anode (composed of a special coating). Focusing the electron beam on the face of the screen is accomplished by the focus coil. A direct current through the windings sets up a strong magnetic field at the center of the coil. Electrons move precisely along the axis of the tube and pass through the focusing field with no deflection. This is because they move parallel to the magnetic field at all times.

Any electron which enters the focusing field at an angle to the axis of the tube has a force exerted on it that is perpendicular to its direction of motion. A second force on this electron is perpendicular to the magnetic lines and is, therefore, constantly changing in direction. These forces cause the electron to move in a helical or corkscrew path shown in figure 3-6. With the proper velocity of the electron and strength of the magnetic field, the electron will be caused to move at an angle which allows it to converge with other electrons at some point on the crt screen. Focusing is accomplished by adjusting the current flow through the focusing coils.

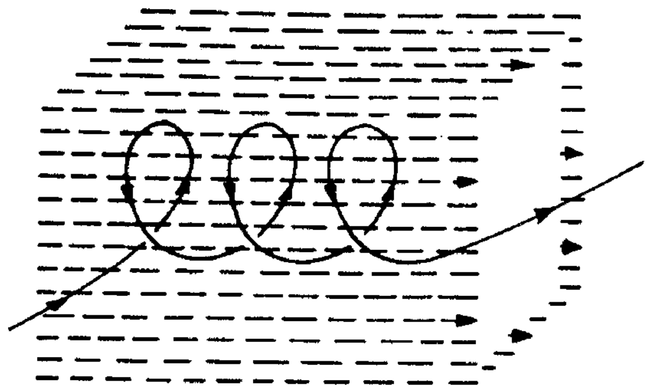


Figure 3-6.—Helical motion of electron through a uniform magnetic field.

The focused electron beam is deflected by a magnetic field that is generated by current flow through a set of deflection coils, as shown in figure 3-5. These coils are mounted around the outside surface of the neck of the crt. Normally, four deflection coils (**N**, **S**, **E**, and **W**) are used, as shown in figure 3-4. Two coils in series are positioned in a manner that causes the magnetic field produced to be in a vertical plane. The other two coils, also connected in series, are positioned so that their magnetic field is in a horizontal plane. The coils (**N-S**) which produce a horizontal field are called the VERTICAL DEFLECTION COILS and the coils (**E-W**) which produce a vertical field are called the HORIZONTAL DEFLECTION COILS. This may be more clearly understood if you recall that an electron beam will be deflected at right angles to a deflecting field. The deflection coils are illustrated in view A of figure 3-7. View B shows the **N-S** windings in schematic form.

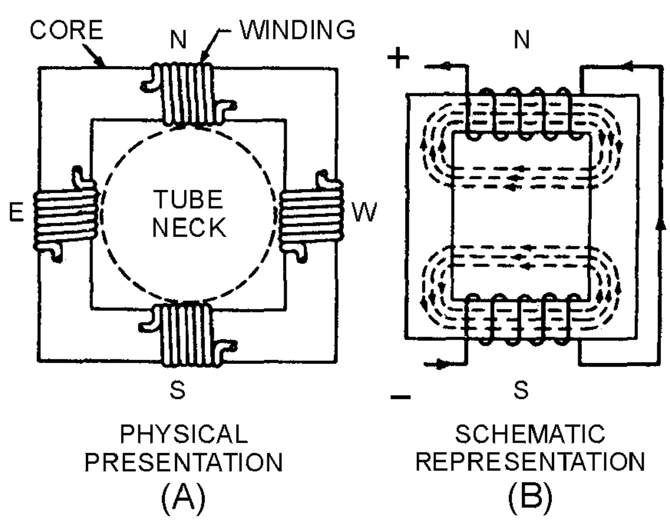


Figure 3-7.—Deflection yoke.

Electron deflection in the electromagnetic crt is proportional to the strength of the magnetic fields. Magnetic field strength depends on current in the coils. The sweep circuits associated with electromagnetically deflected cathode-ray tubes must provide currents, rather than voltage, to produce the desired beam deflection.

A sawtooth current is required to produce a linear trace. A deflection coil may be considered equivalent to the circuit shown in view A of figure 3-8. Because of the inductance of the coil, a trapezoidal voltage must be applied across the coil to produce a sawtooth of current through it. This is illustrated in view B. (Refer to NEETS, Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits*, for a review of wave shaping.)

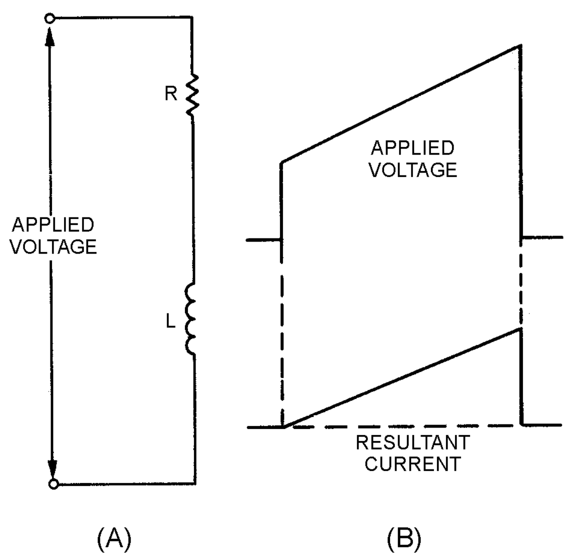


Figure 3-8.—Deflection coil equivalent circuit and waveform.

Sweep Rotation

Azimuth indication of the ppi requires that the range trace rotate about the center of the screen. A very simple means of achieving sweep rotation is to cause the deflection coil to rotate about the neck of the crt in synchronization with the antenna motion. This method, however, has the disadvantages of inaccuracy and maintenance complications inherent to any mechanical gear-train assembly.

Most modem ppi systems employ fixed deflection coils and use special circuits to electronically rotate the magnetic field. Figure 3-9 illustrates a method of electronically producing a rotating sweep. In view A, a range sweep current, i, is applied to the vertical deflection coils only. The resulting magnetic field, represented by ! , lies along the axis of these coils. The resulting range trace, shown by the short straight line, is vertical because the electron beam is deflected perpendicular to the magnetic field. In view B, range sweep currents are applied to both sets of coils, and the resultant magnetic field takes a position between the axes of the two sets of coils. Because of this shift of the magnetic field, the range trace is rotated 45 degrees clockwise from its previous position. In view C, the sweep current is applied to the horizontal deflection coils only, and the range trace lies 90 degrees clockwise from its original position. Further rotation is obtained if the polarities of the deflection coil currents are varied in proper sequence, as illustrated in views D and E.

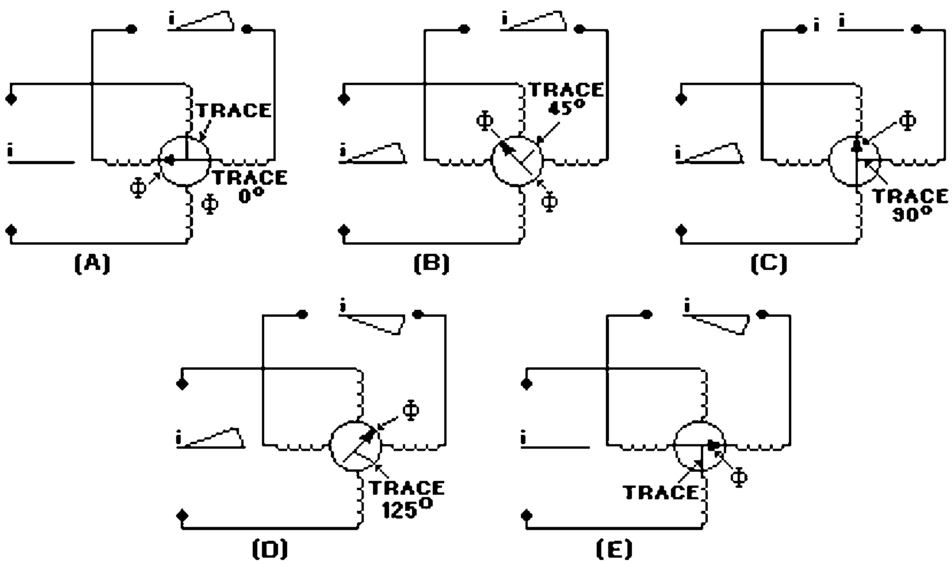


Figure 3-9.—Trace rotation.

To synchronize sweep rotation with antenna rotation, you must convert antenna azimuth (bearing) information into electrical signals. These signals, usually provided by synchros, control the amplitudes and polarities of the sawtooth sweep currents applied to the deflection coils.

Figure 3-10 illustrates the waveforms of current required to produce a rotating range sweep. The amplitudes of the sawtooth sweep currents are varied sinusoidally (like a sine wave), corresponding to the rotation of the antenna. Notice that there is a 90 degree phase difference between the amplitude variations of the horizontal and vertical waveforms.

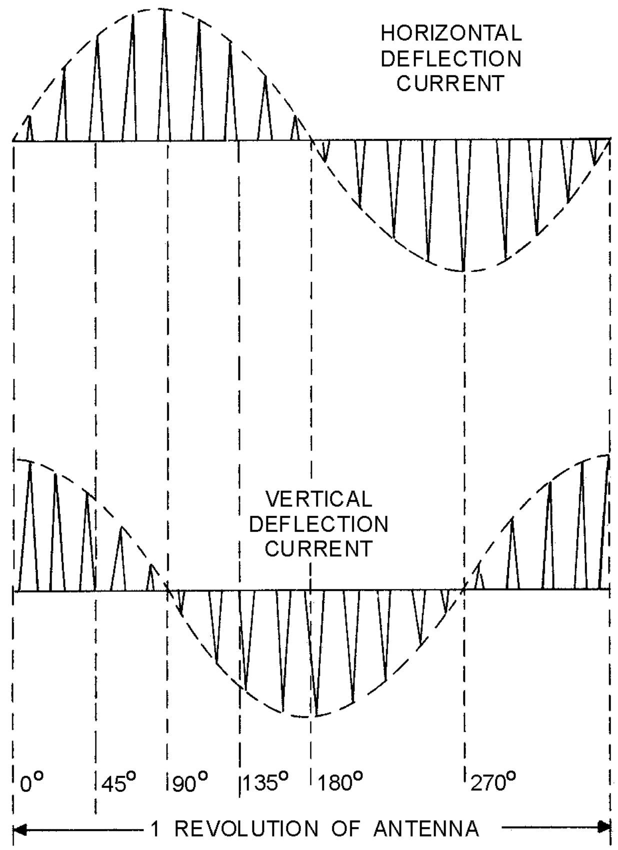


Figure 3-10.—Deflection coil currents.

CRT Screen Persistence

A ppi requires a crt in which the screen is coated with a long-persistence phosphor. This is necessary because each target reflects energy for only a short period of time during each rotation of the antenna. Therefore, the target indication on the face of the crt must be able to continue to glow during the portion of antenna rotation when the target is not reflecting energy.

Q4. What coordinates are presented on a ppi scope?

Q5. What type of deflection is preferred for a crt electron beam?

Q6. Which of the two types of deflection coils (fixed or rotating) is used most often?

RANGING CIRCUITS

The accuracy of target-range data provided by a radar varies with the use of the radar. For example, a weapons systems radar operating in a search mode is required to be accurate within a small percentage of its maximum range. However, an intercept radar, operating in a tracking mode, must supply range data that is even more accurate; it must be within a few yards of the actual range.

In some applications of radar, the indicator sweep is calibrated by a transparent overlay with an engraved range scale. This overlay enables the operator to estimate the range of targets. In other applications, electronic range marks are supplied to the indicator. They usually appear as vertical pulses on A-scopes and as concentric circles on ppi scopes. The distance between range marks is generally determined by the type of equipment and its mode of operation.

In a weapons systems radar that requires extremely accurate target-range data, a movable range marker may be used. The range marker is obtained from a range-marker generator and may be a movable range gate or range step. When a ppi scope is used, a range circle of adjustable diameter may be used to measure range accurately. In some cases, movement of the range marker is accomplished by adjusting a calibrated control from which range readings are obtained.

The following discussion describes the operation of three types of range-marker generators: the RANGE-GATE GENERATOR, the RANGE-MARKER GENERATOR, and the RANGE-STEP GENERATOR. The range-gate generator, used in conjunction with a blocking oscillator, generates a movable range gate. The range-marker generator and the range-step generator, used in conjunction with an astable multivibrator, generate fixed range marks and a movable range step, respectively.

Range-Gate Generator

Figure 3-11 shows a simplified block diagram of a typical range-gate generator. The pulse-repetition frequency is controlled by a master oscillator, or multivibrator, in which the output is coupled to a trigger thyratron (both in the synchronizer). The output of the trigger thyratron is used to trigger the radar modulator and the scope sweep circuits, thus starting the transmitter pulse and the range sweep at the same instant, referred to as time T0.

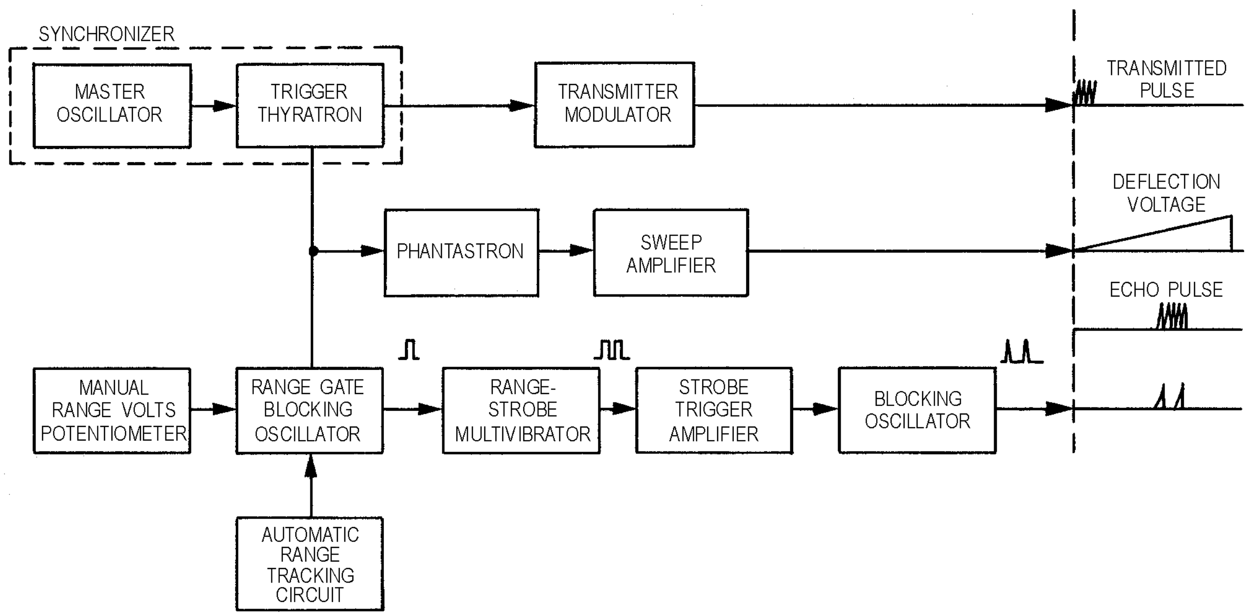


Figure 3-11.—Range-gate generator.

The PHANTASTRON in the sweep circuits is a variable timing circuit that supplies a sweep sawtooth to the sweep amplifier. The width of the gate and sawtooth is dependent upon the range selected by the radar operator.

The range-gate circuit receives its input pulse from the trigger thyratron and generates a delayed range-gate pulse. The delay of this pulse from time **T0** is dependent on either the range of the target when the radar is tracking, or the manual positioning of the range-volts potentiometer when the radar is not tracking (in the search mode). The range-gate triggers the range-strobe multivibrator, from which the output is amplified and sent to the blocking oscillator (which sharpens the pulses), as shown in figure 3-11. This range gate is used to select the target to be tracked. When in the track mode, the range gate brightens the trace or brackets the blip (depending on the system) to indicate what target is being tracked. Range-gate generators are used most often in weapons-control track radar A-scope presentations, but they can also be used with ppi presentations. When used with a ppi presentation, the range gate must be movable in both range and bearing.

The range-gate generator can easily be modified to produce a range strobe instead of a range gate. A range strobe is simply a single brightened spot that is movable both in range and bearing. In operation, the range strobe or range gate control also controls a dial or digital readout to provide a range readout to the operator.

Range-Marker Generator

Several types of range-marker generators are in common use. Figure 3-12 shows a simplified version of a circuit that produces both range markers and the basic system timing triggers. The master oscillator in this case is a blocking oscillator that operates at a frequency of 80.86 kilohertz. By dividing 80.86 kilohertz into 1 (t = 1/frequency), we find the time required for one cycle of operation is 12.36 microseconds. Thus the blocking oscillator produces pulses 1 radar mile apart. These are fed to the 5:1 divider circuit. Five of the 1-mile marks are required to produce an output from the divider circuit. These five-mile marks are sent to the indicator for display and to the 10:1 divider circuit. In the latter case, ten of the five-mile marks are required to produce an output from the 10:1 divider. Thus the output triggers are 50 miles apart. These basic timing triggers are for a radar with a range of fifty miles. The period between triggers could be extended through the use of additional dividers for use with longer range systems.

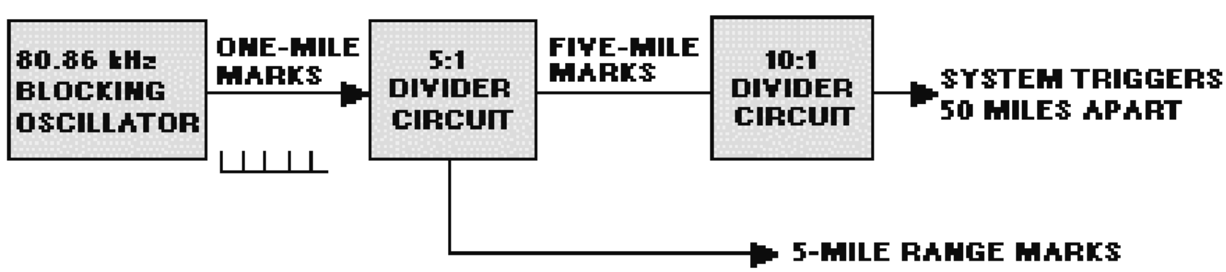


Figure 3-12.—Range-marker generator.

Another version of a range-mark generator is shown in figure 3-13. This circuit provides range marks at 1,000-, 2,000-, or 3,000-yard intervals. Generation of the marks begins with the ringing oscillator, which is started by a delayed master trigger from the synchronizer. A ringing oscillator produces a sinusoidal output of a fixed duration and frequency when triggered. The output is synchronized to the input trigger. In this circuit, the trigger causes the oscillator to produce a 162-kilohertz signal that lasts for 4 1/2 cycles. The emitter follower isolates the ringing oscillator from the countdown multivibrator and clips the oscillator output signals. This action allows only the positive half of each sine wave to reach the multivibrator. The positive triggers from the ringing oscillator are at 1,000-yard intervals. This input signal results in an output from the countdown multivibrator of 1,000-, 2,000-, or 3,000-yard range marks, depending on the position of the RANGE MARK SELECT SWITCH.

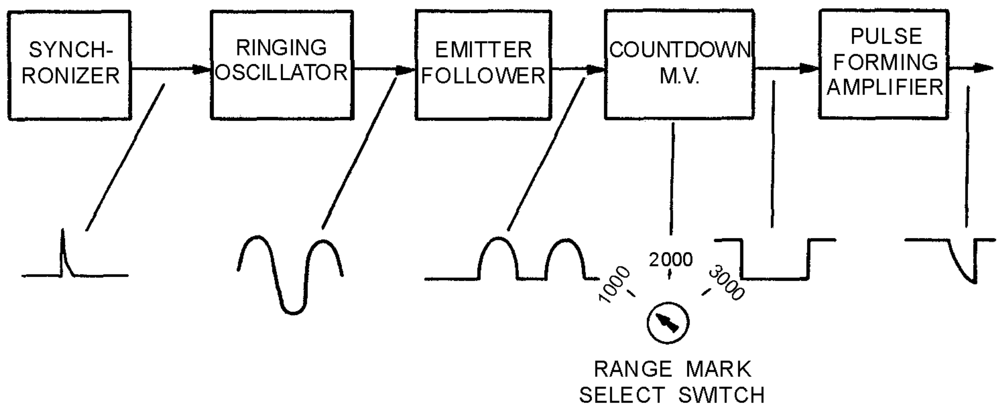


Figure 3-13.—Range-marker generator.

Range-Step Generator

The range step is often used to determine target range on an A-scope presentation. The appearance of a range step on an A-scope is illustrated in figure 3-14.

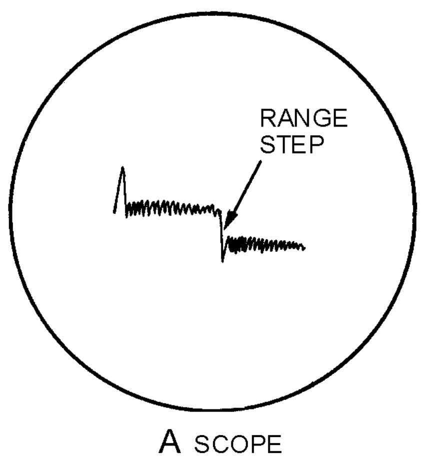
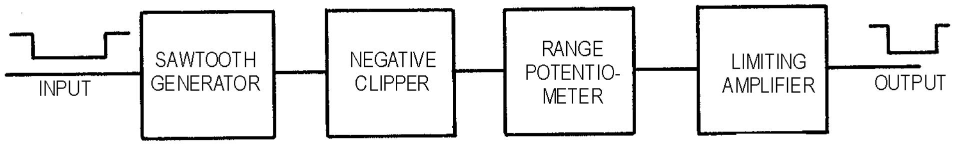
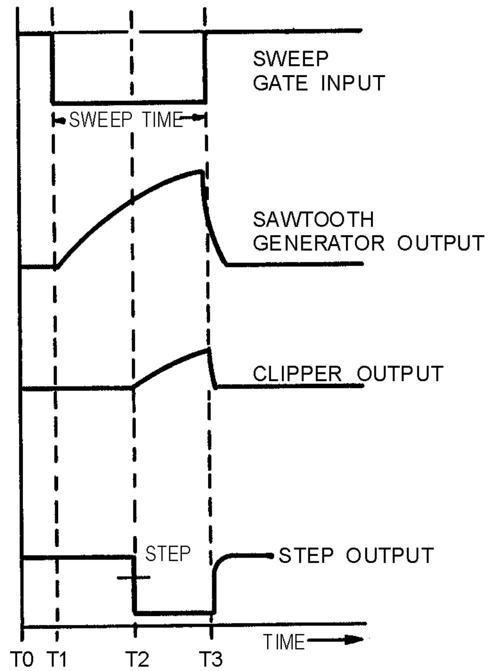


Figure 3-14.—Range-step presentation.

View A of figure 3-15 is a block diagram of a simple range-step generator consisting of a sawtooth generator, a negative clipper, a range potentiometer, and a limiting amplifier. The position of the range step along the indicator's time base is controlled by the range potentiometer. When the range step coincides with the leading edge of a target's echo pulse, the range can be read directly from a calibrated readout associated with the potentiometer.



(A)



(B)

Figure 3-15.—Range-step generation.

View B shows the time relationships of the voltage waveforms produced by the range-step generator. During the sweep gate, the sawtooth generator produces a sawtooth voltage that is sent to the clipper. The point at which the sawtooth is clipped is controlled by the range potentiometer. The clipped sawtooth is shaped in the limiting amplifier to produce the output voltage waveform. The portion of the output waveform from T1 to T3 is applied to the vertical-deflection plates of the indicator crt to produce the display shown in figure 3-14.

Q7. What type of ranging circuit is most often used with a radar that requires extremely accurate range data?

Q8. The range sweep in a range-gate generator is started at the same time as what other pulse?

Q9. Range-marker generators produce pulses based on what radar constant?

Q10. What radar scope uses a range step for range measurement?

RADAR ANTENNAS

In this section, we will briefly review the requirements of radar antennas. Antenna characteristics are discussed in detail in NEETS, Module 10, *Introduction to Wave-Generation, Transmission Lines, and Antenna* and in Module 11, *Microwave Principles*. A review of these modules would be helpful at this point to prepare you for the following radar antenna discussion.

Antennas fall into two general classes, OMNIDIRECTIONAL and DIRECTIONAL. Omnidirectional antennas radiate rf energy in all directions simultaneously. They are seldom used with modern radars, but are commonly used in radio equipment, in iff (identification friend or foe) equipment, and in countermeasures receivers for the detection of enemy radar signals. Directional antennas radiate rf energy in patterns of LOBES or BEAMS that extend outward from the antenna in one direction for a given antenna position. The radiation pattern also contains minor lobes, but these lobes are weak and normally have little effect on the main radiation pattern. The main lobe may vary in angular width from one or two degrees in some radars to 15 to 20 degrees in other radars. The width depends on the system's purpose and the degree of accuracy required.

Directional antennas have two important characteristics, DIRECTIVITY and POWER GAIN. The directivity of an antenna refers to the degree of sharpness of its beam. If the beam is narrow in either the horizontal or vertical plane, the antenna is said to have high directivity in that plane. Conversely, if the beam is broad in either plane, the directivity of the antenna in that plane is low. Thus, if an antenna has a narrow horizontal beam and a wide vertical beam, the horizontal directivity is high and the vertical directivity is low.

When the directivity of an antenna is increased, that is, when the beam is narrowed, less power is required to cover the same range because the power is concentrated. Thus, the other characteristic of an antenna, power gain, is introduced. This characteristic is directly related to directivity.

Power gain of an antenna is the ratio of its radiated power to that of a reference (basic) dipole. Both antennas must have been excited or fed in the same manner and each must have radiated from the same position. A single point of measurement for the power-gain ratio must lie within the radiation field of each antenna. An antenna with high directivity has a high power gain, and vice versa. The power gain of a single dipole with no reflector is unity. An array of several dipoles in the same position as the single dipole and fed from the same line would have a power gain of more than one; the exact figure would depend on the directivity of the array.

The measurement of the bearing of a target, as detected by the radar, is usually given as an angular position. The angle may be measured either from true north (true bearing), or with respect to the bow of a ship or nose of an aircraft containing the radar set (relative bearing). The angle at which the echo signal returns is measured by using the directional characteristics of the radar antenna system. Radar antennas consist of radiating elements, reflectors, and directors to produce a narrow, unidirectional beam of rf energy. A pattern produced in this manner permits the beaming of maximum energy in a desired direction. The transmitting pattern of an antenna system is also its receiving pattern. An antenna can therefore be used to transmit energy, receive energy, or both. The simplest form of antenna for measuring azimuth (bearing) is a rotating antenna that produces a single-lobe pattern.

The remaining coordinate necessary to locate a target in space may be expressed either as elevation angle or as altitude. If one is known, the other can be calculated from basic trigonometric functions. A method of determining the angle of elevation or the altitude is shown in figure 3-16. The slant range is obtained from the radar scope as the distance to the target. The angle of elevation is the angle between the axis of the radar beam and the earth's surface. The altitude in feet is equal to the slant range in feet multiplied by the sine of the angle of elevation. For example if the slant range in figure 3-16 is 2,000 feet and the angle of elevation is 45 degrees, the altitude is 1,414.2 feet (2,000 × .7071). In some radar equipments that use antennas that may be moved in elevation, altitude determination is automatically computed.

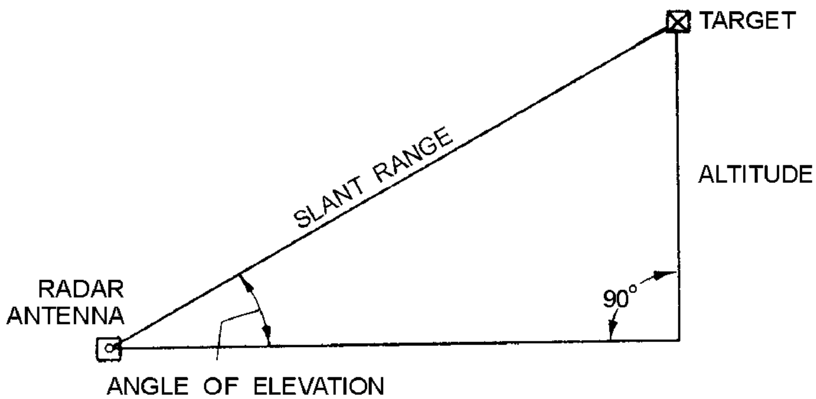


Figure 3-16.—Radar determination of altitude.

PARABOLIC REFLECTORS

A SPHERICAL WAVEFRONT spreads out as it travels and produces a pattern that is neither too sharp nor too directive. On the other hand, a PLANE wavefront does not spread out because all of the wavefront moves forward in the same direction. For a sharply defined radar beam, the need exists to change the spherical wavefront from the antenna into a plane wavefront. A parabolic reflector is one means of accomplishing this.

Radio waves behave similarly to light waves. Microwaves travel in straight lines as do light rays. They may be focused and/or reflected just as light rays can. In figure 3-17, a point-radiation source is placed at the focal point F. The field leaves this antenna with a spherical wavefront. As each part of the wavefront reaches the reflecting surface, it is shifted 180 degrees in phase and sent outward at angles that cause all parts of the field to travel in parallel paths. Because of the shape of a parabolic surface, all paths from F to the reflector and back to line XY are the same length. Therefore, all parts of the field arrive at line XY the same time after reflection.

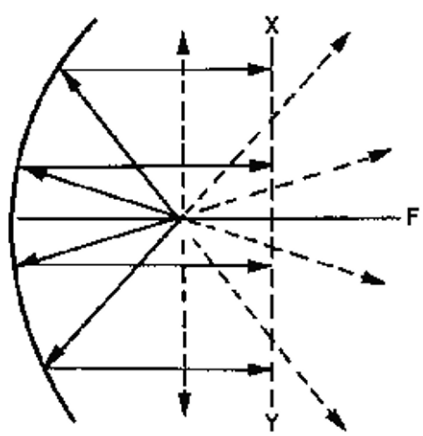


Figure 3-17.—Parabolic reflector radiation.

If a dipole is used as the source of radiation, there will be radiation from the antenna into space (dotted lines in figure 3-17) as well as toward the reflector. Energy that is not directed toward the paraboloid has a wide-beam characteristic that would destroy the narrow pattern from the parabolic reflector. This occurrence is prevented by the use of a hemispherical shield (not shown) that directs most radiation toward the parabolic surface. By this means, direct radiation is eliminated, the beam is made sharper, and power is concentrated in the beam. Without the shield, some of the radiated field would leave the radiator directly. Since it would not be reflected, it would not become a part of the main beam and thus could serve no useful purpose. The same end can be accomplished through the use of a PARASITIC array, which directs the radiated field back to the reflector, or through the use of a feed horn pointed at the paraboloid.

The radiation pattern of a parabola contains a major lobe, which is directed along the axis of revolution, and several minor lobes, as shown in figure 3-18. Very narrow beams are possible with this type of reflector. View A of figure 3-19 illustrates the parabolic reflector.

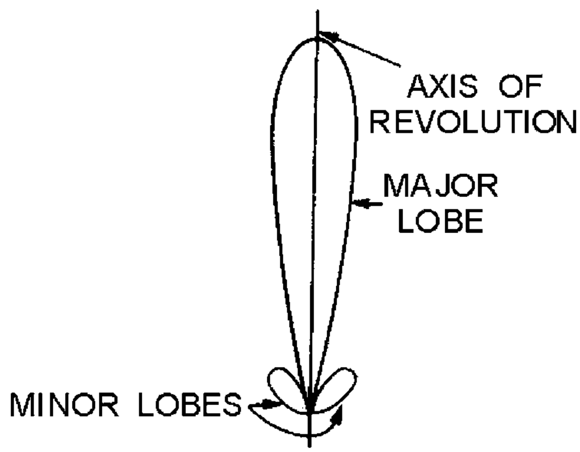


Figure 3-18.—Parabolic radiation pattern.

Truncated Paraboloid

View B of figure 3-19 shows a horizontally truncated paraboloid. Since the reflector is parabolic in the horizontal plane, the energy is focused into a narrow horizontal beam. With the reflector truncated, or cut, so that it is shortened vertically, the beam spreads out vertically instead of being focused. Since the beam is wide vertically, it will detect aircraft at different altitudes without changing the tilt of the antenna. It also works well for surface search radars to overcome the pitch and roll of the ship.

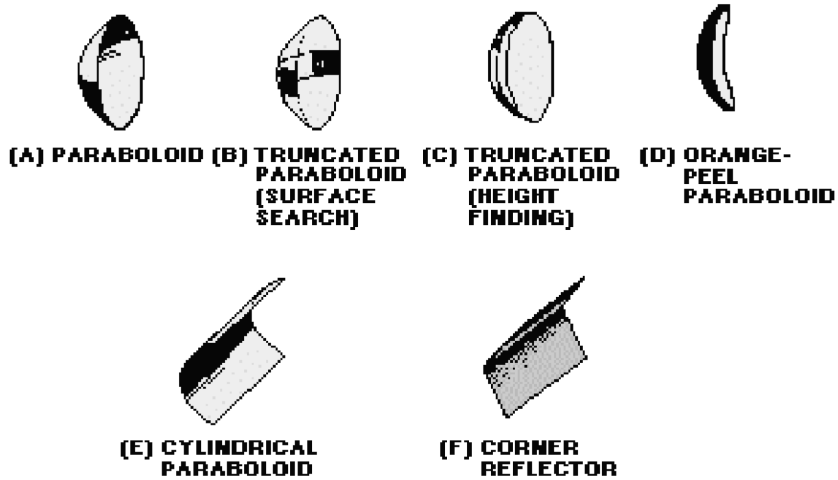


Figure 3-19.—Reflector shapes.

The truncated paraboloid reflector may be used in height-finding systems if the reflector is rotated 90 degrees, as shown in view C. Because the reflector is now parabolic in the vertical plane, the energy is focused into a narrow beam vertically. With the reflector truncated, or cut, so that it is shortened horizontally, the beam spreads out horizontally instead of being focused. Such a fan-shaped beam is used to determine elevation very accurately.

Orange-Peel Paraboloid

A section of a complete circular paraboloid, often called an ORANGE-PEEL REFLECTOR because of its shape, is shown in view D of figure 3-19. Since the reflector is narrow in the horizontal plane and wide in the vertical, it produces a beam that is wide in the horizontal plane and narrow in the vertical. In shape, the beam resembles a huge beaver tail. This type of antenna system is generally used in height-finding equipment.

Cylindrical Paraboloid

When a beam of radiated energy noticeably wider in one cross-sectional dimension than in the other is desired, a cylindrical paraboloidal section approximating a rectangle can be used. View E of figure 3-19 illustrates this antenna. A parabolic cross section is in one dimension only; therefore, the reflector is directive in one plane only. The cylindrical paraboloid reflector is either fed by a linear array of dipoles, a slit in the side of a waveguide, or by a thin waveguide radiator. Rather than a single focal point, this type of reflector has a series of focal points forming a straight line. Placing the radiator, or radiators, along this focal line produces a directed beam of energy. As the width of the parabolic section is changed, different beam shapes are obtained. This type of antenna system is used in search and in ground control approach (gca) systems.

Q11. Which of the two general classes of antennas is most often used with radar?

Q12. The power gain of an antenna is directly related to what other antenna property?

Q13. A parabolic reflector changes a spherical wavefront to what type of wavefront?

CORNER REFLECTOR

The corner-reflector antenna consists of two flat conducting sheets that meet at an angle to form a corner, as shown in view F of figure 3-19. This reflector is normally driven by a half-wave radiator located on a line which bisects the angle formed by the sheet reflectors.

BROADSIDE ARRAY

The desired beam widths are provided for some vhf radars by a broadside array, such as the one shown in figure 3-20. The broadside array consists of two or more half-wave dipole elements and a flat reflector. The elements are placed one-half wavelength apart and parallel to each other. Because they are excited in phase, most of the radiation is perpendicular or broadside to the plane of the elements. The flat reflector is located approximately one-eighth wavelength behind the dipole elements and makes possible the unidirectional characteristics of the antenna system.

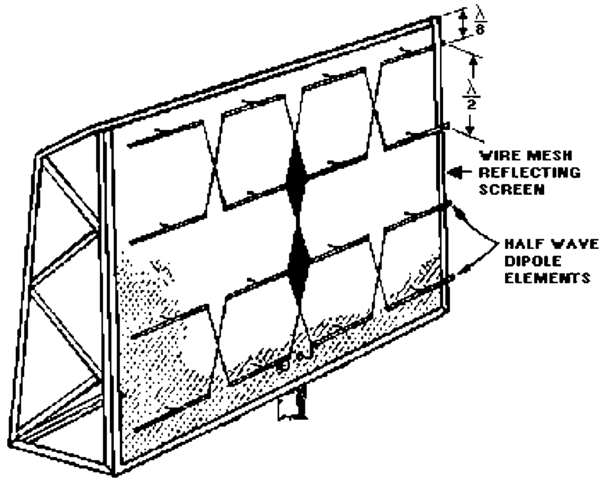


Figure 3-20.—Broadside array.

HORN RADIATORS

Horn radiators, like parabolic reflectors, may be used to obtain directive radiation at microwave frequencies. Because they do not involve resonant elements, horns have the advantage of being usable over a wide frequency band.

The operation of a horn as an electromagnetic directing device is analogous to that of acoustic horns. However, the throat of an acoustic horn usually has dimensions much smaller than the sound wavelengths for which it is used, while the throat of the electromagnetic horn has dimensions that are comparable to the wavelength being used.

Horn radiators are readily adaptable for use with waveguides because they serve both as an impedance-matching device and as a directional radiator. Horn radiators may be fed by coaxial or other types of lines.

Horns are constructed in a variety of shapes as illustrated in figure 3-21. The shape of the horn, along with the dimensions of the length and mouth, largely determines the field-pattern shape. The ratio of the horn length to mouth opening size determines the beam angle and thus the directivity. In general, the larger the opening of the horn, the more directive is the resulting field pattern.

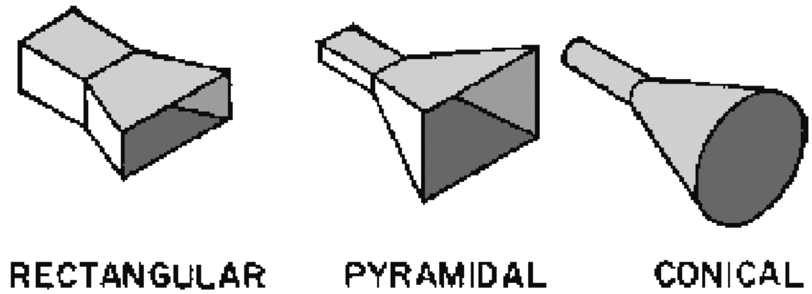


Figure 3-21.—Horn radiators.

FEEDHORNS

A waveguide horn, called a FEEDHORN, may be used to feed energy into a parabolic dish. The directivity of this feedhorn is added to that of the parabolic dish. The resulting pattern is a very narrow and concentrated beam. In most radars, the feedhorn is covered with a window of polystyrene fiberglass to prevent moisture and dirt from entering the open end of the waveguide.

One problem associated with feedhorns is the SHADOW introduced by the feedhorn if it is in the path of the beam. (The shadow is a dead spot directly in front of the feedhorn.) To solve this problem the feedhorn can be offset from center. This location change takes the feedhorn out of the path of the rf beam and eliminates the shadow. An offset feedhorn is shown in figure 3-22.

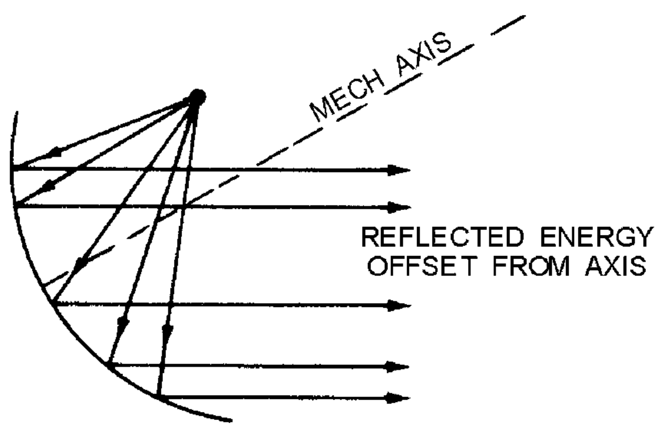


Figure 3-22.—Offset feedhorn.

AIRBORNE RADAR ANTENNAS

Airborne radar equipment is used for several specific purposes. Some of these are bombing, navigation, and search. Radar antennas for this equipment are invariably housed inside nonconducting radomes, not only for protection but also to preserve aerodynamic design. Some of these radomes are carried outside the fuselage, while others are flush with the skin of the fuselage. In the latter case, the radar antenna itself is carried inside the fuselage, and a section of the metallic skin is replaced by the nonconducting radome. The radar antenna and its radome must operate under a wide variety of temperature, humidity, and pressure conditions. As a result, mechanical construction and design must minimize any possibility of failure. Transmission lines are usually hermetically sealed to prevent moisture accumulation inside them. Such accumulation would introduce losses. Because the low air pressures encountered at high elevations are very conducive to arcing, pressurization of equipment is widely used (the pressure is maintained by a small air pump). In some airborne radar equipments, practically all of the equipment is sealed in an airtight housing, along with the antenna and transmission line. The antenna radome forms a portion of the housing.

Airborne radar antennas are constructed to withstand large amounts of vibration and shock; the radar antennas are rigidly attached to the airframe. The weight of the radar antenna, including the rotating mechanism required for scanning, is kept to a minimum. In addition, the shape of the radome is constructed so as not to impair the operation of the aircraft.

The airborne radar antenna must have an unobstructed view for most useful operation. Frequently, the antenna must be able to scan the ground directly under the aircraft and out toward the horizon. To meet this requirement, the antenna must be mounted below the fuselage. If scanning toward the rear is not required, the antenna is mounted behind and below the nose of the aircraft. If only forward scanning is needed, the antenna is mounted in the nose. When an external site is required, a location at the wing tip is common. A fire-control radar antenna is frequently located near the turret guns or in a special nacelle, where it can scan toward the rear or sides of the aircraft.

Q14. How many major lobes are produced by a paraboloid reflector?

Q15. What type of radiator normally drives a corner reflector?

Q16. The broadside array consists of a flat reflector and what other elements?

Q17. Horn radiators serve what purpose other than being directional radiators?

SUMMARY

The following is a brief summary of the important points of this chapter.

A radar **INDICATOR** presents the information (video) from the radar receiver in a usable manner. The display usually consists of one or more of the coordinates of range, bearing, and altitude.

The **CATHODE-RAY TUBE (crt)** is the best available device for displaying the two-dimensional relationship produced by radar coordinates. The most commonly used crt displays are the A-SCOPE, the RHI, and the PPI. The A-scope presents range information only. The rhi displays range and height information. The ppi is the most widely used radar display indicator and presents range and bearing.

The range of a radar contact is determined by special **RANGING CIRCUITS**. The following three basic types of ranging circuits are used.

RANGE-GATE GENERATORS produce a movable gate that measures range based on elapsed time and can be used on A-scope and ppi displays.

RANGE-MARKER GENERATORS produce fixed interval range marks that can be used to estimate the range to a detected target. Range marks appear as an intensified series of vertical dots on an rhi and as concentric circles on a ppi.

The **RANGE-STEP GENERATOR** produces a movable step that is displayed on an A-scope presentation.

RADAR ANTENNAS are usually directional antennas that radiate energy in a one directional lobe or beam. The two most important characteristics of directional antennas are directivity and power gain. Radar antennas often use parabolic reflectors in several different variations to focus the radiated energy into a desired beam pattern. Other types of antennas used with radar systems are the corner reflector, the broadside array, and horn radiators.

ANSWERS TO QUESTIONS Q1. THROUGH Q17.

A1. Range, bearing, and elevation.

A2. Triggers, video, and antenna information.

A3. Range and elevation.

A4. Range and bearing.

A5. Electromagnetic.

A6. Fixed.

A7. Range gate or range step.

A8. Transmitter.

A9. The radar mile (12.36 microseconds).

A10. The A scope.

A11. Directional.

A12. Directivity.

A13. Plane.

A14. One.

A15. Half-wave.

A16. Two or more half-wave dipoles.

A17. Waveguide impedance matching devices.

CHAPTER 4

RADAR SYSTEM MAINTENANCE

LEARNING OBJECTIVES

Upon completion of this chapter, the student will be able to:

1. Interpret the transmitter frequency spectrum in terms of frequency distribution, power output, receiver response, and an acceptable spectrum curve.
2. Describe the methods for measuring the average and peak power outputs of a radar transmitter.
3. Describe the methods of measuring receiver sensitivity.
4. Define receiver bandwidth in terms of the receiver response curve and state the most common methods of measuring tr tube recovery time.
5. List the support systems associated with a typical shipboard radar system and describe the basic function of each.
6. State the general rules for the prevention of personnel exposure to rf radiation and X-ray emissions.

INTRODUCTION TO RADAR MAINTENANCE

The effectiveness of your radar system depends largely upon the care and attention you give it. An improperly adjusted transmitter, for example, can reduce the accuracy of a perfectly aligned receiver; the entire system then becomes essentially useless. Maintenance, therefore, must encompass the entire system for best operation.

Because of the complexity of most radar systems, trying to detail step-by-step procedures for specific maintenance actions in this chapter is impractical. However, the basic procedures for some maintenance actions that are common to most radar systems will be discussed. Also, an overview of support systems for radars will be presented. This will include electrical power, dry-air systems, and liquid cooling systems. Finally, safety precautions inherent to radars are listed.

TRANSMITTER PERFORMANCE CHECKS

The transmitter of a radar is designed to operate within a limited band of frequencies at an optimum power level. Operation at frequencies or power levels outside the assigned band greatly decreases the efficiency of the transmitter and may cause interference with other radars. Therefore, transmitter performance must be monitored closely for both frequency and output power.

TRANSMITTER FREQUENCY

Whether of the fixed-frequency or tunable type, the radar transmitter frequency should be checked periodically. If the transmitter is of the fixed-frequency type and found to be operating outside its normal operating band, the problem is probably a defective part. The defective component must be replaced. If the transmitter is tunable, the transmitter must again be tuned to the assigned frequency.

Each time a radar transmitter generates an rf pulse, it produces electromagnetic energy. You should recall from your study of NEETS, Module 12, *Modulation Principles*, that the square wave used to modulate the transmitter carrier wave has (1) the fundamental square-wave frequency and (2) an infinite number of odd harmonics of the fundamental square wave frequency. When this square wave is used to modulate the transmitter carrier frequency, both the fundamental and odd harmonic frequencies of the square wave heterodyne with the transmitter carrier frequency. The heterodyning process produces in each transmitted rf pulse the following frequencies:

1. The fundamental carrier frequency
2. The sum and difference frequencies between the carrier and fundamental square-wave frequencies
3. The sum and difference frequencies between the odd harmonics of the square wave and the carrier frequencies

For a complete discussion of this process, you should review module 12.

Actually, the radar energy is distributed more or less symmetrically over a band of frequencies. This frequency distribution of energy is known as the FREQUENCY SPECTRUM. An analysis of frequency spectrum characteristics may be made with a SPECTRUM ANALYZER. The spectrum analyzer presents a graphic display of energy versus frequency. An extensive explanation of spectrum analyzer use can be found in the Electronics Installation and Maintenance Book (EIMB), *Test Methods and Practices*, NAVSEA 0967-LP-000-0130.

Spectrum Analysis

When properly performed and interpreted, a spectrum analysis will reveal misadjustments and troubles that would otherwise be difficult to locate. Therefore, you should be able to perform a spectrum analysis and understand the results.

You may be wondering why we are so interested in the frequency spectrum of an rf pulse. To better understand why, look at the spectrum of a transmitter as compared to the response curve of a receiver in figure 4-1. The receiver's response curve has a broader bandwidth than the transmitted spectrum, which ensures complete coverage. But the receiver responds best to frequencies in the middle of the bandwidth. This causes the receiver response to taper off from both sides of the center frequency until the response passes through the half-power points, as shown on the curve. Usually the receiver response beyond these points is too low to be useful and is not considered. Notice that the spectrum of the transmitter is centered inside the response curve of the receiver, thus yielding maximum efficiency.

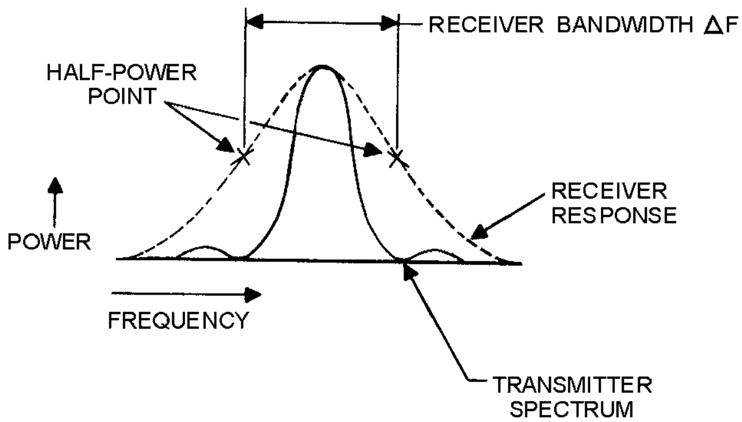


Figure 4-1.—Transmitter spectrum compared with receiver response.

Any frequency, when modulated by another frequency, will produce a base frequency with sideband frequencies (sum and difference). In other words, the output of a pulsed radar will contain more than one frequency. The output frequency spectrum of the pulsed radar transmitter does not consist of just a single frequency that is turned on and off at the pulse-repetition frequency (prf). Consider the spectrum as a base frequency (carrier) that is modulated by short rectangular pulses occurring at the prf of the radar. Two distinct modulating components are present: One component consists of the prf and its associated harmonics; the other component consists of the fundamental and odd-harmonic frequencies that make up the rectangular modulating pulse.

The distribution of power over the radar frequency spectrum depends on the amount of modulation. A pulsed radar spectrum is illustrated in figure 4-2. The vertical lines represent the modulation frequencies produced by the prf and its associated harmonics; the lobes represent the modulation frequencies produced by the fundamental pulse frequency and its associated harmonics. The amplitude of the main lobe falls to zero on each side of the carrier. The side lobes are produced by the odd harmonics of the fundamental pulse frequency. The zero points are produced by the even harmonics of the fundamental pulse frequency. In an ideal spectrum each frequency above the carrier has its counterpart in another frequency below the carrier. These frequencies are equally spaced and have equal power. Therefore, the pattern is symmetrical about the carrier. The main lobe, of course, contains the major portion of the transmitted rf energy.

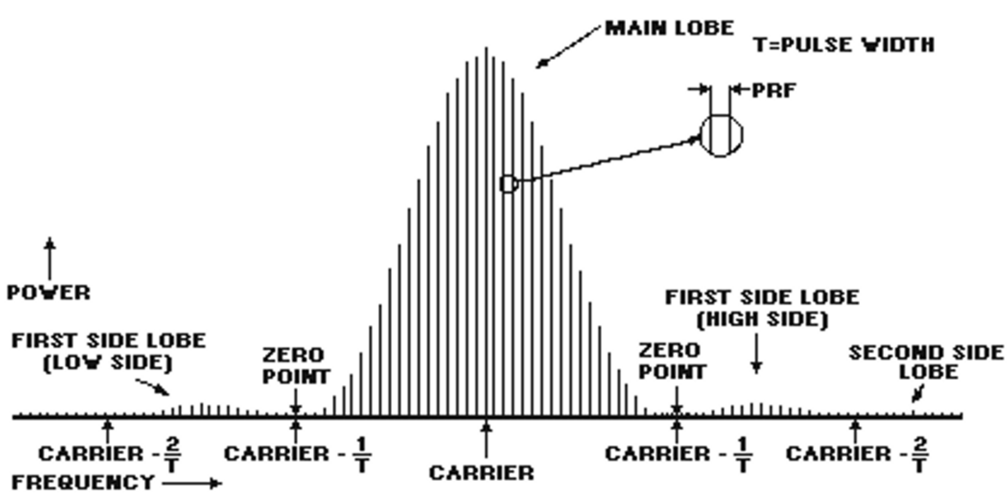


Figure 4-2.—Spectrum of a pulse-modulated carrier.

A radar transmitter in good condition should produce a spectrum curve similar to the curves shown in view A or B in figure 4-3. Good curves are those in which the two halves are symmetrical and contain deep, well-defined minimum points (minima) on both sides of the main peak.

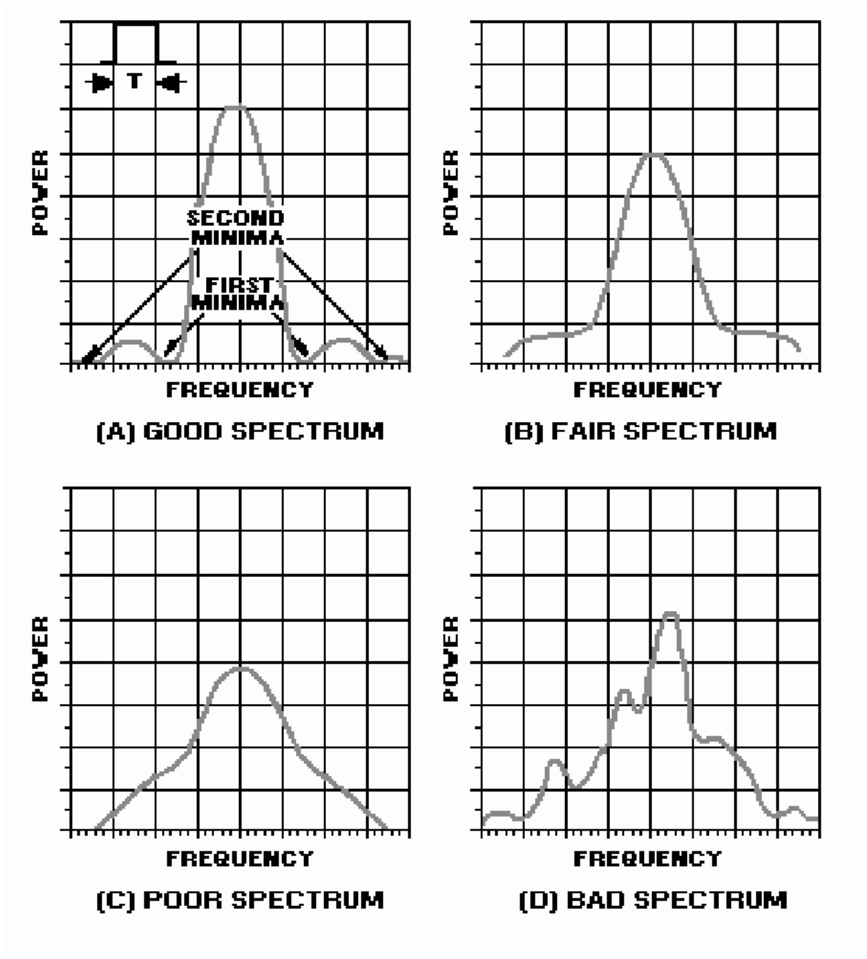


Figure 4-3.—Comparison of radar spectra.

A curve without well-defined minima, as in the curve shown in view C, indicates that the transmitter output is being frequency modulated during the pulse. This condition may occur when a pulse without sufficiently steep sides or a flat peak is applied to the transmitter. It may also occur when a transmitter tube is unstable or is operated without proper voltage, current, or magnetic field.

An extremely irregular spectrum, as in the curve in view D, is an indication of severe frequency modulation. This condition usually causes trouble with the receiver automatic frequency control (afc) as well as a general loss of signal strength. You can often improve a faulty spectrum by adjusting the transmission line stubs or by replacing the transmitter tube. When the spectrum has two large peaks that are quite far apart, it indicates that the transmitter tube is DOUBLE MODING (shifting from one frequency to another). This could be caused by standing waves in the transmission line or a faulty transmitter tube. Standing waves may be caused by a faulty line connection, a bad antenna rotating joint, or obstructions in the line. (Standing waves are described in NEETS, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*.)

In the case of a good or fair spectrum curve with sharply defined minimum points on both sides of the main lobe, the distance between these two points is proportional to the duration of the transmitted pulse.

The device most commonly used to check the frequency spectrum of a radar transmitter is the spectrum analyzer.

Frequency-Measuring Devices

Devices used to determine the basic carrier frequency of a radar transmitter are the ELECTRONIC FREQUENCY COUNTER, the WAVEMETER, and the ECHO BOX. One or more of these devices may be included in a special RADAR TEST SET designed for a specific system or type of radar. Radar test sets quite often consist of several types of test equipment. This combination of test equipments enables both transmitter and receiver performance checks to be carried out with one test instrument. Electronic frequency counters, frequency meters, and wavemeters are discussed in NEETS, Module 16, *Introduction to Test Equipment*. The echo box is discussed in the next section. The specific equipments and procedures required to measure the frequency of any radar system are found in the associated system technical manuals and related PMS documents.

Q1. The spectrum of a radar transmitter describes what characteristic of the output pulse?

Q2. Where should the transmitter spectrum be located with respect to the receiver response curve?

Q3. The ideal radar spectrum has what relationship to the carrier frequency?

Q4. The display screen of a spectrum analyzer presents a graphic plot of what two signal characteristics?

The Echo Box

The ECHO BOX is an important test instrument for indicating the overall radar system performance. The echo-box test results reflect the combined relative effectiveness of the *transmitter* as a transmitter of energy and the receiver as a *receiver* of energy.

The echo box, or RESONANCE CHAMBER, basically consists of a resonant cavity, as shown in view A of figure 4-4. You adjust the resonant frequency of the cavity by varying the size of the cavity (the larger the cavity the lower the frequency). A calibrated tuning mechanism controls the position of a plunger and, therefore, the size of the cavity. The tuning mechanism is adjusted for maximum meter deflection, which indicates that the echo box is tuned to the precise transmitted frequency. The tuning mechanism also indicates on a dial (figure 4-5, view A) both the coarse transmitted frequency and a numerical reading. This reading permits the technician to determine the transmitted frequency with greater accuracy by referring to a calibration curve on a chart (figure 4-5, view B).

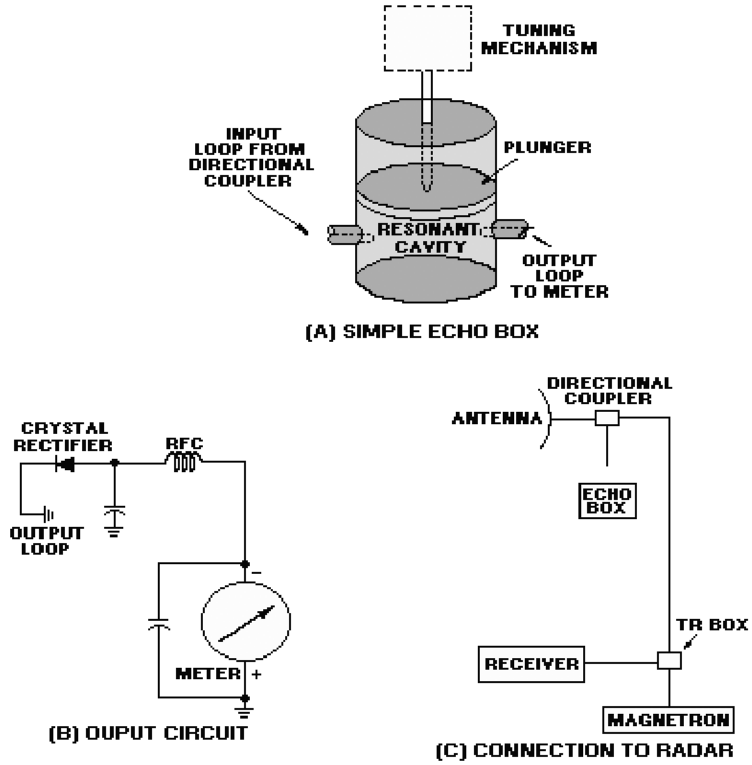


Figure 4-4.—Echo box.

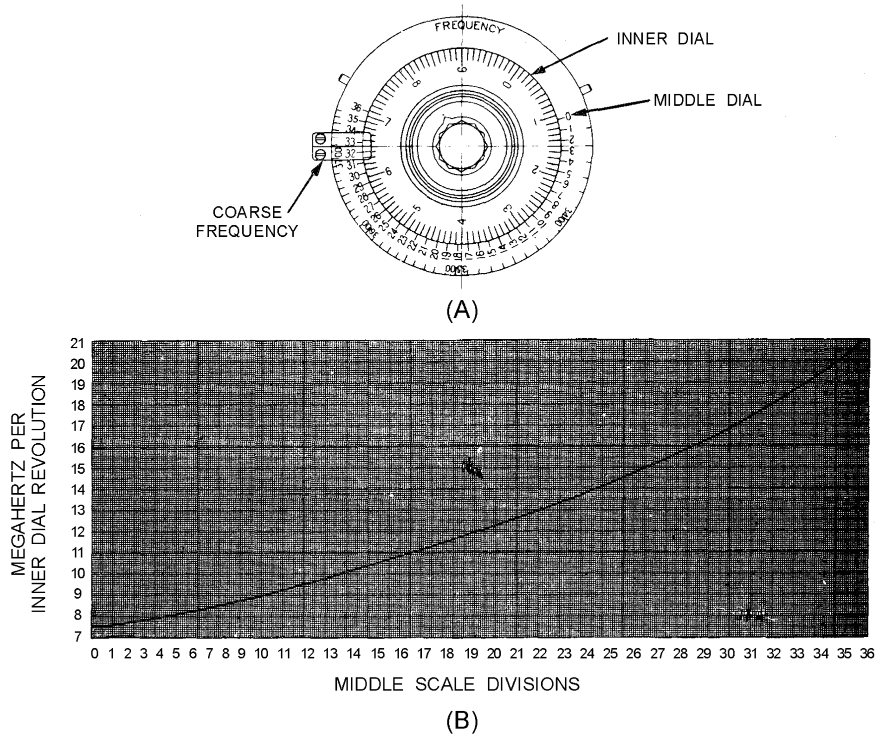


Figure 4-5.—Reading the echo box dial.

Energy is coupled into the cavity from the radar by means of an rf cable connected to the input loop. Energy is coupled out of the cavity to the rectifier and meter by means of the output loop. You can vary the amount of coupling between the echo box and the crystal rectifier by changing the position of the output loop. A schematic diagram of the output circuit is shown in figure 4-4, view B. The energy picked up by the loop is rectified, filtered, and applied to the meter. The method of connecting the echo box in a radar system is shown in figure 4-4, view C.

RING TIME MEASUREMENTS

Some of the energy generated by the radar transmitter is picked up by the echo box by means of the directional coupler. This energy causes oscillations (known as RINGING) within the echo box that persist for some time after the end of the radar pulse, much in the fashion of an echo that persists in a large room after a loud noise. As this echo dies down, a part of it is fed back into the radar receiving system, again by means of the directional coupler. The ringing causes a saturating signal to appear on the radar indicator (figure 4-6). The longer this ringing extends, the better the performance of the radar.

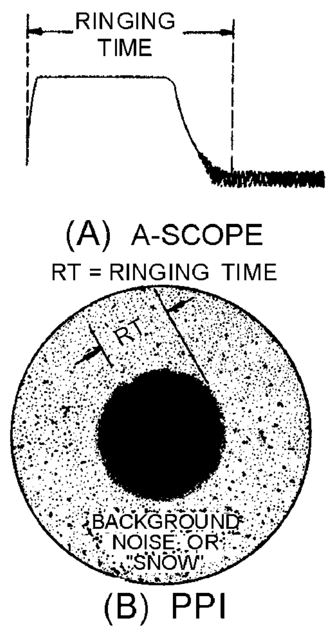


Figure 4-6.—Ring time saturation of A-scope and ppi.

The length of time the echo box *should* ring under the particular conditions of the test is called the EXPECTED RING TIME. You may determine whether or not the radar is performing well by comparing the expected ring time with the ring time observed.

The ring time to be expected on a good radar depends on the particular type of radar being tested; on the way the echo box is installed - that is, whether a directional coupler or a pickup dipole is used; on the length and type of cable used; on the individual ringing ability of the particular echo box in use; on the frequency of the radar; and on the temperature of the echo box at the time of the test. Corrections are made for all of these factors according to the procedure given in the technical manual for the echo box being used.

You may use an echo box without correction to detect a *change* in the performance of a radar. You simply log and compare the ring time from day to day. You should recognize that these readings do not permit the comparison of a particular radar with a standard of performance; however, you can use the readings to tell whether or not its performance is deteriorating.

Because ring time measurements are the most valuable single feature of the echo box, they must be measured properly. Ring time measurements are made on the A-scope or on the ppi.

In measuring the ring time, you should make sure the echo-box ringing (not some fixed-target echo or block of echoes) is being monitored. You can determine this condition by adjusting the radar gain control and noting if the ring time varies on the scope. The echo box ringing will change in duration; fixed target echoes, however, will not change duration.

To obtain the best results, you should repeat every ring time measurement at least four times; then average the readings. You should take special care to ensure that all readings are accurate. If two or more technicians use the same echo box, they should practice together until their ring time measurements agree.

TRANSMITTER POWER MEASUREMENT

Because high peak power and radio frequencies are produced by radar transmitters, special procedures are used to measure output power. High peak power is needed in some radar transmitters to produce strong echos at long ranges. Low average power is also desirable because it enables transmitter components to be compact, more reliable, and to remain cooler during operation. Because of these considerations, the lowest possible duty cycle (pw x prf) must be used for best operation. The relationships of peak power, average power, and duty cycle were described in chapter 1. Peak power in a radar is primarily a design consideration. It depends on the interrelationships between average power, pulse width, and pulse-repetition time.

You take power measurements from a radar transmitter by sampling the output power. In one sampling method, you use a pickup horn in front of the antenna. Air losses and weather conditions make the horn placement extremely critical and also affect the accuracy of the sample. A more accurate and convenient method can be used. In this method, you sample the output power through a directional sampling coupler located at the point in the transmitter where a power reading is desired. Power-amplifier transmitters usually have sampling couplers after each stage of amplification.

Some radar sets have built-in power-measuring equipment; others require the use of general purpose test equipment or a special test set. In any case, the measuring instruments are most often referenced to 1 milliwatt; readings are taken in dBm (a discussion of the decibel measurement system was presented in NEETS, Module 11, *Microwave Principles*).

When taking power measurements, you must allow for power losses. You must add the directional coupler attenuation factor and the loss in the connecting cable to the power meter reading. The sum is the total power reading. For example, the directional coupler has an attenuation factor of 20 dB, the connecting cable has a loss rating of 8 dB, and the reading obtained on the power meter is 21 dBm. Therefore, the transmitter has an output power that is 49 dBm (21 + 20 + 8). Power readings in dBm obtained by the above procedure are normally converted to watts to provide useful information. Although the conversion can be accomplished mathematically, the procedure is relatively complex and is seldom necessary. Most radar systems have a conversion chart, such as the one shown in figure 4-7, attached to the transmitter or the test equipment. As you can see on the chart, 49 dBm is easily converted to 80 watts average power.

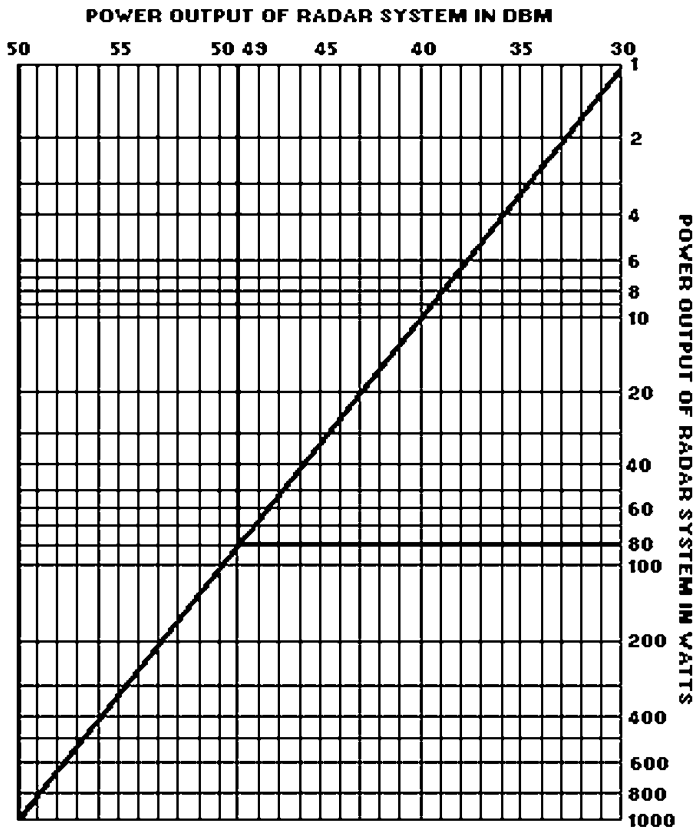


Figure 4-7.—Conversion of power in dBm to watts (average).

You can convert average power to peak power by dividing average power by the duty cycle of the radar. If the radar in the above example has a duty cycle of 0.001, then the peak power can be calculated with the following formula:

$$\text{peak power } (P_{pk}) = \frac{\text{average power } (P_{avg})}{\text{duty cycle}}$$

$$= \frac{80 \text{ watts}}{0.001}$$

$$= 80,000 \text{ watts or } 80 \text{ kilowatts}$$

Many radar systems have charts available to convert average power to peak power.

Q5. The peak power of a radar depends on the interrelationship of what other factors?

Q6. Transmitter power readings are most often referenced to what power level?

RECEIVER PERFORMANCE CHECKS

The performance of a radar receiver is determined by several factors, most of which are established in the design engineering of the equipment. In the paragraphs that follow, factors concerned with maintenance are considered. Important factors are (1) receiver sensitivity, which includes noise figure determination and minimum discernible signal (mds) measurement; (2) tr recovery time; and (3) receiver bandwidth.

Many radar systems contain circuits that serve special functions. Three of these special circuits are instantaneous automatic gain control (iagc), sensitivity time control (stc), and fast time constant (ftc). These circuits may be found in combination or alone, depending on the purpose of the radar. When the test methods and procedures about to be described are used, these special functions should not be used. If an automatic frequency control (afc) circuit is included in the radar, it may be permitted to operate during receiver tests. A good way you can check afc circuit operation is to complete the tests specified for manual tuning and then switch to afc. If the afc circuit operation is normal, test indications should not differ.

RECEIVER SENSITIVITY

Insufficient detection range in a radar system can be caused by decreased sensitivity in the radar receiver. This condition results mainly from the great number of adjustments and components associated with the receiver. A decrease of receiver sensitivity has the same effect on range performance as does a decrease of transmitter power. For example, a 6 dB loss of receiver sensitivity shortens the effective range of a radar just as much as a 6 dB loss in transmitter power. Such a drop in transmitter power is evident and is easy to detect. On the other hand, a 6 dB loss in receiver sensitivity, which can easily result from a slight misadjustment in the receiver, is difficult to detect unless accurate measurements are made.

Figure 4-8 shows a comparison of radar system performance versus maximum range. The system performance loss in dB includes both transmitter and receiver losses. You should note that with a loss of 5 dB in both receiver and transmitter (a total of 10 dB), only 55 percent of the maximum range of the system is realized.

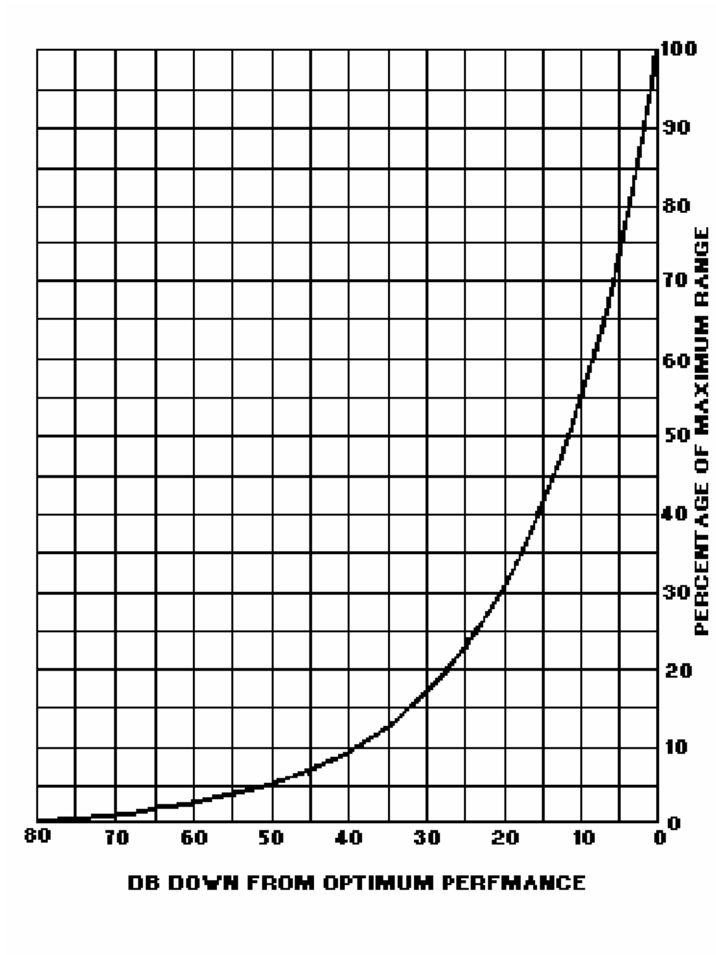


Figure 4-8.—System performance versus maximum range.

The sensitivity of the radar receiver is a measure of its ability to pick up weak signals. The greater the sensitivity of the receiver, the better the receiver picks up weak signals. You can determine receiver sensitivity by measuring the power level of the MINIMUM DISCERNIBLE SIGNAL (mds). Mds is defined as the weakest signal that produces a visible receiver output (on a scope). Its value is determined by the receiver output noise level (noise tends to obscure weak signals). Because mds measurement depends on the receiver noise level, measuring either mds or noise level (called NOISE FIGURE) will indicate receiver sensitivity.

Many radar systems have built-in receiver sensitivity test circuits. These test circuits indicate the sensitivity of the receiver to the technician or operator.

To measure the mds, you must measure the power of a test pulse in which the level is just sufficient to produce a visible receiver output. If a radar receiver has the mds level specified in the maintenance manual, then the noise figure should also be correct. Therefore, measurement of the mds is a satisfactory substitute for a noise-figure determination and is less complicated.

Because receiver sensitivity readings are taken periodically for comparison purposes, the identical pulse length must be used for each measurement. Maintenance instructions for the radar set usually specify the correct pulse length to be used in receiver sensitivity tests. In most cases, it is the same as the transmitter pulse length.

Before any measurements of receiver sensitivity can be made, the receiver must be accurately tuned to the transmitter frequency. If the receiver frequency differs from the transmitter frequency, the most likely cause is an improperly adjusted or malfunctioning local oscillator or transmitter frequency drift. Such problems can be caused by heat or aging components. Local oscillator tuning procedures differ widely according to the type of radar system; therefore, you should follow the tuning procedures in the system maintenance manuals.

Two basic methods are used to measure radar receiver sensitivity. One is the PULSE METHOD, in which a pulse of measured amplitude and width is coupled to the receiver. In the second method, you use an fm generator to vary the signal generator output frequency across the receiver bandwidth. This latter method ensures the test signal is within the bandpass of the receiver.

The sensitivity of the receiver is equal to the sum of the reading on the signal generator and the attenuations of the connecting cable and directional coupler. Receiver sensitivity is expressed as a negative dBm; for example, -90 dBm expresses the sensitivity of a receiver that can detect a signal 90 dB less than the 1-milliwatt reference level. A typical receiver sensitivity reading on a modern radar should be in the vicinity of -105 dBm.

RECEIVER BANDWIDTH TEST

Receiver bandwidth is defined as the frequency spread between the half-power points on the receiver response curve. Receiver bandwidth is specified for each radar, but wide variations are often tolerated. If either the bandwidth or the shape of the receiver response curve is not within tolerances, a detailed check of circuit components may be necessary. A considerable change in the value of circuit components is required to alter the response. You should check receiver response after any extensive repair to an IF amplifier.

Figure 4-9 shows a typical response curve of a radar receiver. The half-power points are shown as 3 dB below maximum response. Since the curve is plotted in terms of voltage, these points are also represented by the 70.7 percent voltage points as shown in the figure.

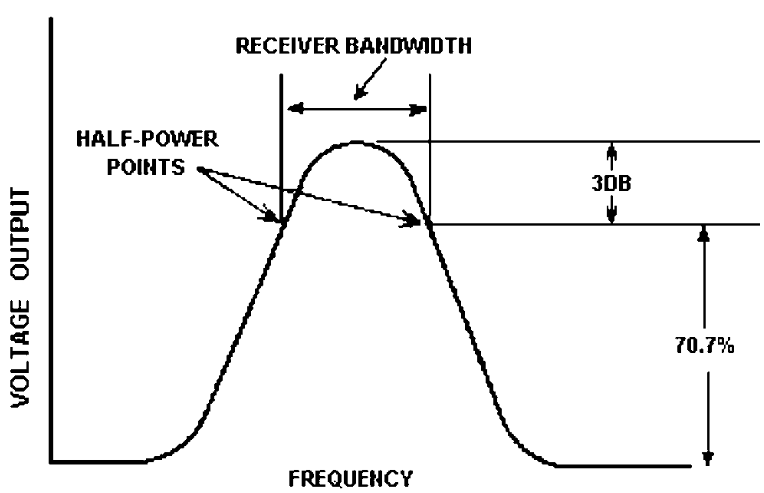


Figure 4-9.—Typical receiver response curve.

TR RECOVERY TIME

The time required for tr recovery is determined by the time taken by the tr switch (tube) to deionize after each transmitter pulse. It is usually defined as the time required for the receiver to return to within 6 dB of normal sensitivity after the end of the transmitter pulse. However, some manufacturers use the time required for the sensitivity to return to within 3 dB of normal sensitivity. Tr recovery time is a factor that limits the minimum range of a radar because the radar receiver is unable to receive until the tr switch is deionized. In various radars, the recovery time may differ from less than 1 microsecond to about 20 microseconds.

The primary function of the tr switch is to protect the sensitive crystal detectors from the powerful transmitter pulse. Even the best tr switches allow some power to leak through; but when the switch is functioning properly, leakage power is so small that it does not damage the crystal. However, the useful life of a tr tube is limited because the amount of leakage to the receiver increases with use.

To ensure efficient performance, some technicians make a policy of replacing the tr tube after a certain number of hours of use. A better practice is to measure the tr recovery time at frequent intervals and make a graph or chart. A graph or chart will immediately disclose any change in performance. Figure 4-10 shows how the recovery time is correlated with leakage power. Note that the end of the useful life of the tr tube is indicated by an increase in recovery time.

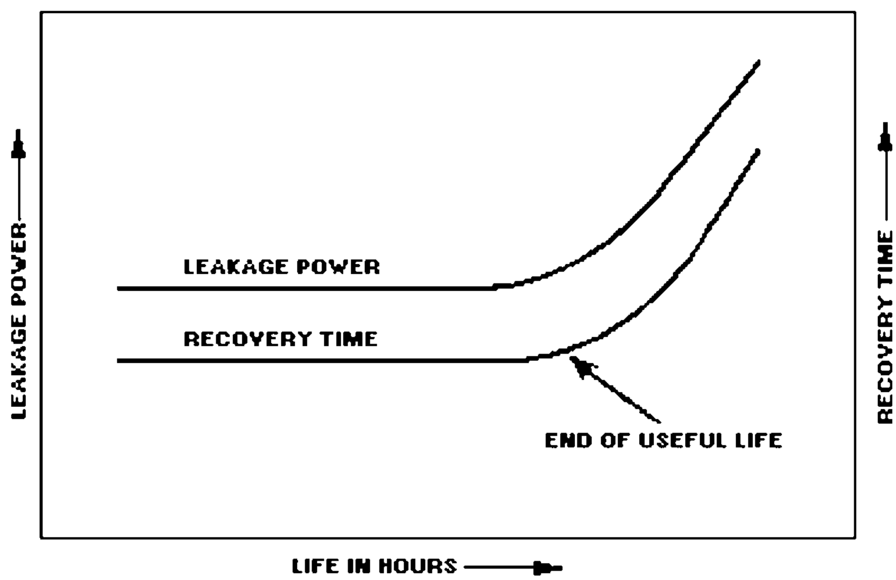


Figure 4-10.—Tr recovery time versus leakage power.

This method of checking the condition of a tr tube is reliable because recovery time increases before leakage power becomes excessive. In practice, a tr tube is replaced when any sharp increase in recovery time becomes apparent.

Ambient temperature also has an effect on recovery time. The colder a tr tube, the greater its recovery time. When tests are conducted under widely varying temperature conditions, this effect must be considered.

One method you can use in testing a tr tube is to measure the KEEP-ALIVE current. This current keeps the tr tube partially ionized, which makes the firing more instantaneous and thus helps protect the receiver crystals. The keep-alive current is normally about 100 microamperes but falls off as the end of the tr tube life approaches. You can also measure the keep-alive voltage between the plate of the tr tube and ground when the voltage source is known to have the correct output. You then record this voltage for use as a reference for future checks. However, these checks are not as reliable as recovery time testing.

Specific procedures for measuring tr leakage and recovery time can be found in the equipment technical manuals.

Q7. A loss of receiver sensitivity has the same effect on range performance as what other loss?

Q8. You determine receiver sensitivity by measuring the power level of what signal?

Q9. When measuring receiver sensitivity, what quantities must you add to the dBm reading obtained on the signal generator or test set?

STANDING WAVE MEASUREMENTS

(You may want to refer to NEETS, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas* for a review of standing waves before going further.) Measurements of standing waves can indicate the approximate operating frequency, the presence of defective transmission-line sections, and the condition of the antenna. Standing waves present on transmission lines and waveguides indicate an impedance mismatch between a transmitter or receiver and its antenna. When this condition occurs, the transfer of energy between these units becomes inefficient. Reflection of energy at the load end of a transmission line results in a wave that travels toward the generator end. This reflected wave varies continuously in phase in much the same way that the incident wave varies in phase. At certain points, a half wavelength apart, the two waves are exactly in phase; the resultant voltage is at maximum. At points a quarter wavelength from the maximums, the two waves are in opposition and voltage nodes (null points) are produced. The ratio of maximum-to-minimum voltage at such points is called the VOLTAGE STANDING WAVE RATIO (vswr). The ratio of maximum-to-minimum current along a transmission line is the same as the vswr. A high vswr (1.5 to 1 or higher) indicates that the characteristic impedance of a transmission line differs greatly from the terminating impedance; a low vswr (1 to 1 is best) indicates a good impedance match between the transmission line characteristic impedance and the terminating impedance.

For radar applications, a low vswr is desired for the following reasons: (1) Reflections in the transmission line cause improper transmitter operation and can result in faulty pulsing (this effect is most pronounced when the line is long, as compared with a wavelength of the transmitted energy); (2) arc-over may occur at the maximum voltage points; and (3) hot spots can occur in the transmission line and cause mechanical breakdown. Since transmission lines for radar equipment are normally coaxial cables or waveguides, slotted lines or directional couplers must be used for standing-wave measurements.

Q10. Receiver bandwidth is defined as those frequencies spread between what two points of the receiver response curve?

Q11. The end of the usefulness of a tr tube is indicated by an increase in what quantity?

SUPPORT SYSTEMS

When you think of radar equipment with its complex electronic circuitry and other sophisticated equipment, you may forget that the entire radar relies on other systems. These other systems are referred to as SUPPORT SYSTEMS and are not normally thought of as part of the radar. These support systems include ELECTRICAL POWER, DRY-AIR, and LIQUID-COOLING SYSTEMS. Without these support systems, radars could not function. Therefore, you must be aware of these support systems and understand their relationship to your radar equipment.

ELECTRICAL POWER

Let us now look at a typical ship's power distribution system. The power system on your ship or aircraft is probably similar in many ways. We will briefly discuss an overall power distribution system and the areas that are closely related to radar equipment.

Power Distribution System

Most ac power distribution systems in naval vessels are 440-volt, 60-hertz, 3-phase, 3-wire, ungrounded systems. The ac power distribution system consists of the power source, equipment to distribute the power, and the equipment which uses the power. A partial distribution chart is shown in figure 4-11.

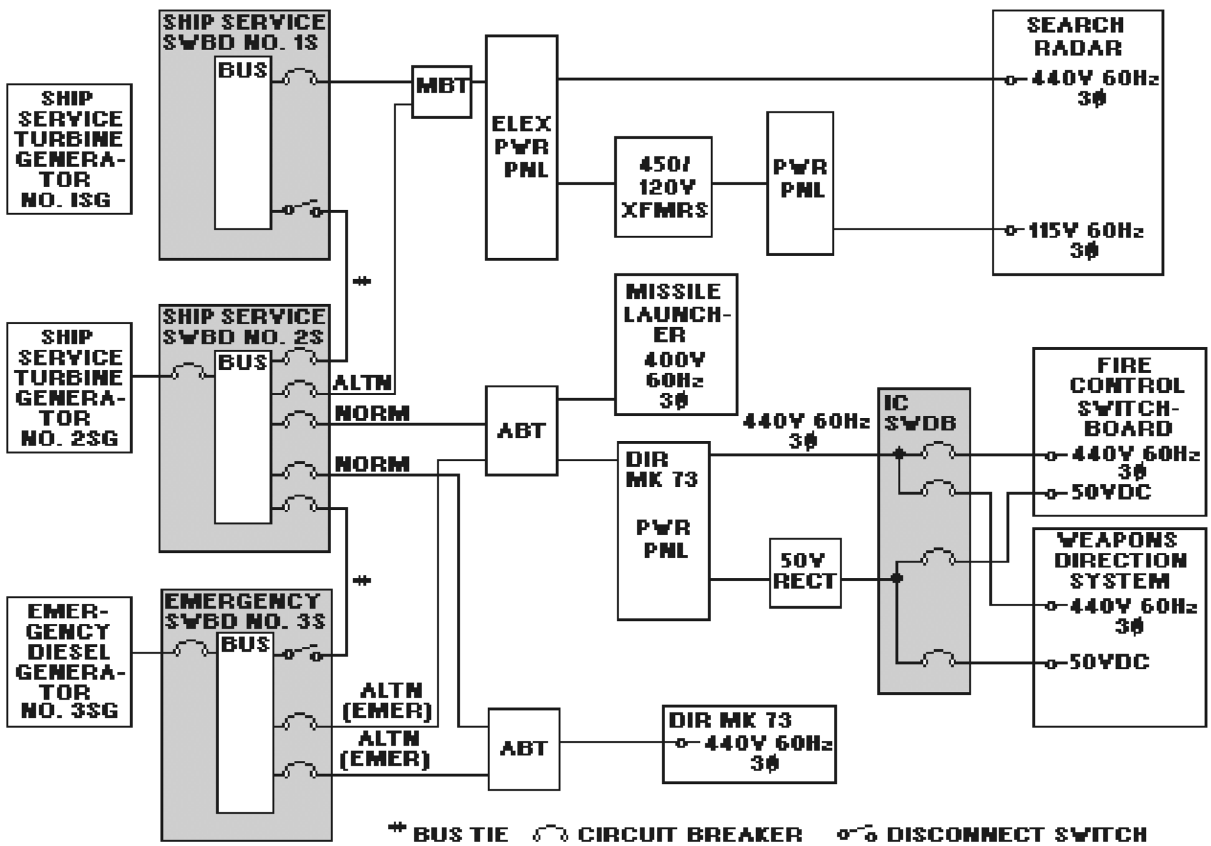


Figure 4-11.—60 Hz distribution.

The power source can be the ship service turbine generator or the emergency diesel generator. Power is normally distributed through the ship service distribution switchboards and power panels. Some large ships also use load centers (not shown) that function as remote switchboards.

Power is used by any equipment that requires electrical power for its operation (lights, motors, director power drives, radar equipment, weapon direction equipment, computers, etc.). The maintenance of the ship service generators, the emergency generators, and distribution switchboards is the responsibility of the ship's engineers (machinist's mates, electrician's mates, enginemen, etc.).

Emergency Power

If power from the ship service distribution system is interrupted, the emergency power distribution system is activated. The emergency system supplies an immediate and automatic source of electrical power to selected loads that are vital to the safety and defense of the ship. This system includes one or more emergency diesel generators and switchboards. The emergency generator is started automatically when a sensor detects the loss of normal power.

Bus Transfer Equipment

Bus transfer equipment is installed on switchboards, at load centers, on power panels, and on loads that are fed by both normal and alternate and/or emergency feeders (figure 4-11). Either the normal or alternate source of the ship's service power can be selected. Emergency power from the emergency distribution system can be used if an emergency feeder is also provided.

Automatic bus transfer (ABT) equipment is used to provide power to vital loads, while nonvital loads can be fed through manual bus transfer (MBT) equipment. For example, the interior communications (IC) switchboard is fed through an ABT in which the alternate input is from the emergency switchboard. A search radar might be fed through an MBT.

Miscellaneous Power

Many other supply voltages are used in radar systems and subsystems. They are usually used as reference voltages for specific functions. When you are missing a power input to your equipment, work backwards from the load to the source. Usually, the power panels and bus transfer units that feed the equipment are located nearby, possibly in the same space or in a passageway.

Keep in mind that technicians have corrected many suspected casualties merely by restoring a minor power input or signal reference, sometimes after hours of troubleshooting.

Q12. Most shipboard distribution systems use ac power that has what number of phases?

Q13. How is emergency power applied when normal power is lost?

Q14. What device is used to switch power from the normal source to an alternate source for nonvital users?

Q15. What procedure should you use when a power input to your equipment is missing?

DRY-AIR SYSTEMS

Some radars depend on inputs of dry air for proper operation. Radar dry air is normally supplied by the ship's central dry-air system. This system produces high-pressure (hp) air and low-pressure (lp) dry air for distribution to user equipment, such as a search or a fire control radar.

Electronics Dry-Air Branch

The electronics dry-air branch is fed from the vital service lp air main through the Type II (desiccant) or Type III (combination refrigerant and desiccant) dehydrators, as shown in figure 4-12. The purpose of the electronics dry-air branch is to provide several electronic equipments with air that is dry enough for proper operation. Microwave components, such as waveguides, cavities, and power amplifiers, require dry air to prevent arcing and internal corrosion. The electronics dry-air branch must satisfy the dry-air requirements of the electronic user equipment. Dry air of less than the required specifications will degrade equipment performance. It may also incur major repairs, overhaul, or replacement of expensive electronic components.

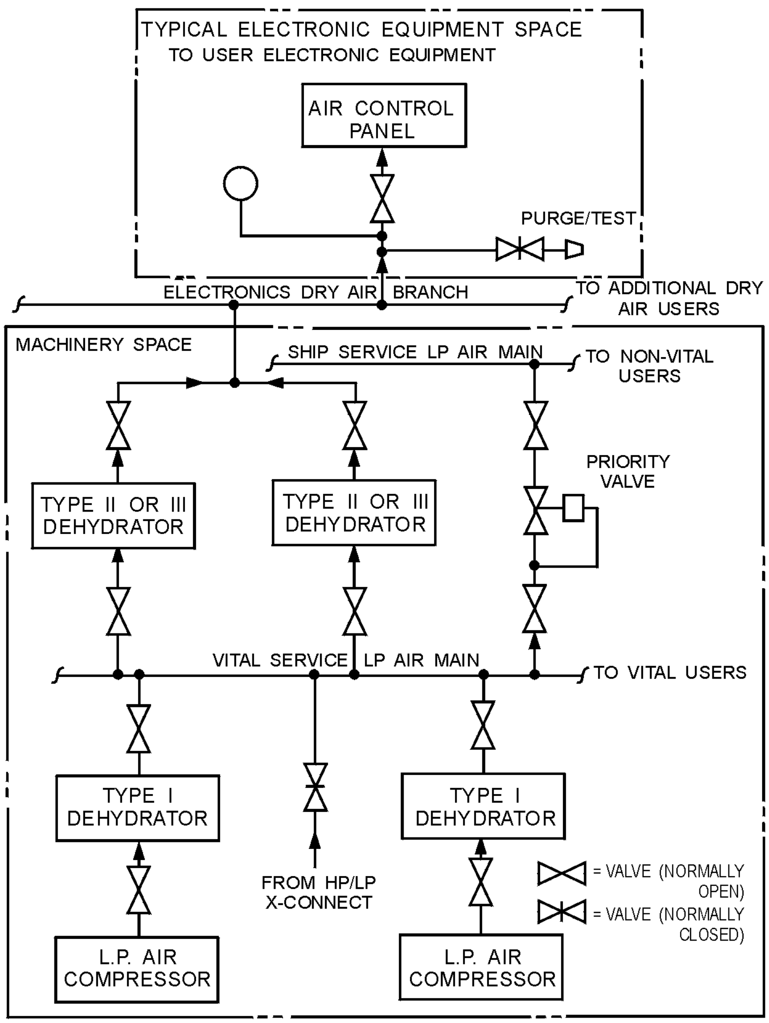


Figure 4-12.—Typical lp air system layout.

Air Control Panel

The dry-air distribution system (figure 4-12) delivers dry air to each air control panel of the user equipment. The air control panels are used to control and regulate the dry-air pressure to that required by the electronic user equipment.

The air control panel (figure 4-13) provides a means of monitoring the dry-air supply to the user equipment. The type of control panel used varies, depending on the outlet pressure and flow rate required.

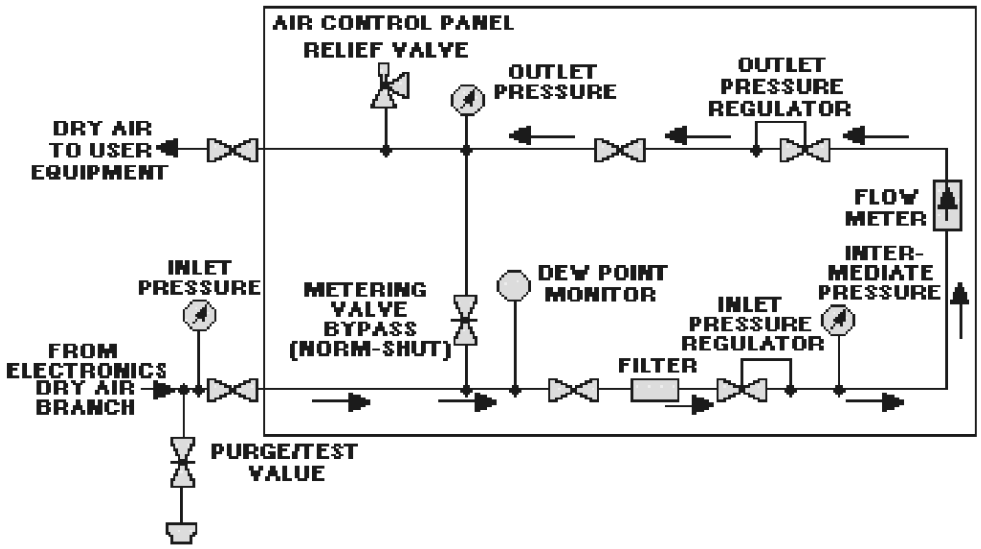


Figure 4-13.—Air control panel flow diagram.

The dew point (related to moisture content) and the flow of the lp dry air can be monitored at the air control panel. Also, the dry-air pressure can be monitored at the input to the control panel, at the input to the flowmeter (in which accuracy is calibrated at a certain pressure), and at the output of the control panel. A filter is installed to trap particles that affect proper pressure regulation. A metering valve bypass and a pressure relief valve are provided in case of malfunctions. The metering valve bypass permits manual control of air pressure to the user equipment.

Electronic Equipment Dehydrators.

Dehydrators or compressor-dehydrators are supplied as part of various radars. Many of them were provided prior to installation of properly configured central dry-air systems. These dehydrators are intended for emergency use in the event of the failure of the central dry-air system. In a typical configuration (figure 4-14), the outlet air from the local dehydrator is connected between the air control panel outlet and the user equipment or radar by a three-way valve.

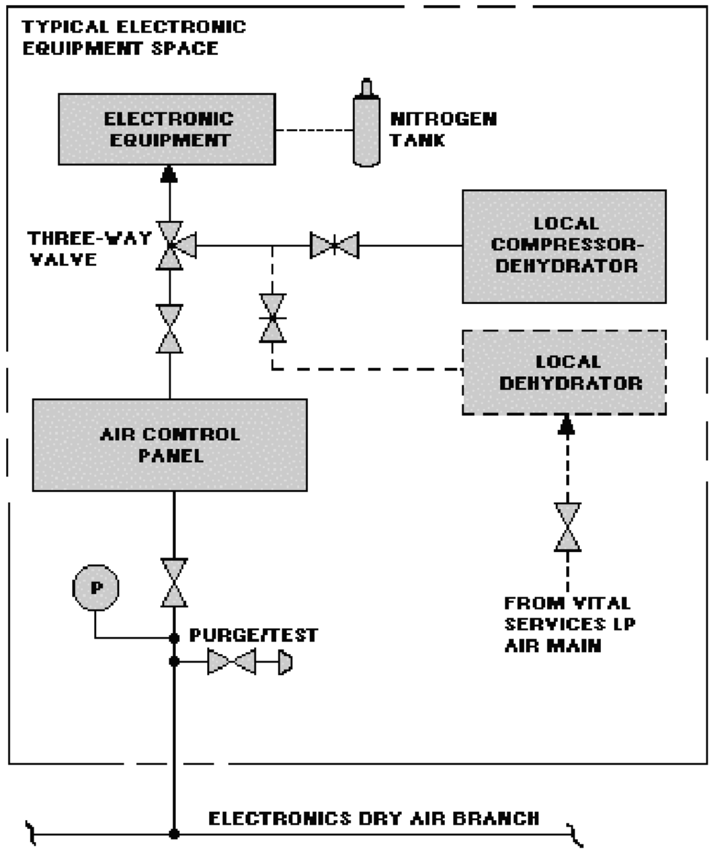


Figure 4-14.—Typical local dehydrator interface.

Local dehydrators depend on the ship's lp air for an inlet supply, while the local compressor-dehydrators can operate independently of the ship's air supply. Some units of electronic equipment that have local dehydrator units are pressure interlocked within the dehydrator unit. When the outlet air pressure is below a set value, the interlock prevents the equipment from going to a full OPERATE condition. When the central dry-air system is used, the pressure interlock is bypassed.

Some radars provide a tank of nitrogen as an emergency source that can be connected in place of dry air. Special safety precautions must be taken when you handle compressed gases because of the possibility of explosion. Nitrogen does not support life; when released in a confined space, it can cause asphyxiation.

Q16. What is the normal source of dry air for a radar system?

Q17. What is the major difference between the electronics dry-air branch and the vital service lp air main?

Q18. What is the air control panel designed to control?

COOLING SYSTEMS

Radar equipment, particularly the high-power transmitters, generate large amounts of heat. This heat must be dissipated to prevent damage to the equipment and to prevent erratic circuit operation. Most radar equipment rooms have high-capacity air-conditioning systems to control the ambient room temperature;

however, equipment cabinets must have additional cooling to control the internal temperature. In the case of transmitters (and other high-voltage circuits), individual components may require cooling.

Cabinets that generate relatively small amounts of heat may only require a system of fans or blowers to maintain constant air circulation. In some cases the air is circulated through a liquid-cooled heat exchanger located inside the cabinet.

Most low-power amplifier tubes are air cooled; most high-power tubes, such as klystrons, crossed-field amplifiers, and magnetrons, are liquid cooled.

The main source of power and heat in a power amplifier package is the high-voltage power supply. Part of the power produced by the power amplifier is transmitted in the form of rf energy; the remainder of the power eventually converts to heat, and cooling is required to dissipate the heat.

Radars that use blowers for cooling will usually have an airflow sensing switch. If the blower fails, the switch will open and remove power from appropriate power supplies. Radars employing liquid cooling normally distribute the liquid into a large number of separate paths, because the flow requirements are quite dissimilar. Each of the various paths will have a low flow interlock. If one of the liquid cooling paths becomes restricted, the low flow interlock switch will open and remove power from the radar.

Liquid cooling systems also include pressure gauges and switches, temperature gauges, and overtemperature switches. Many systems have pressure or flow regulators. Some systems include audio and/or visual alarms that energize before damage actually occurs. In some cases this allows the problem to be corrected without turning off the equipment.

Figure 4-15 illustrates a typical transmitter cooling system showing the many protective devices.

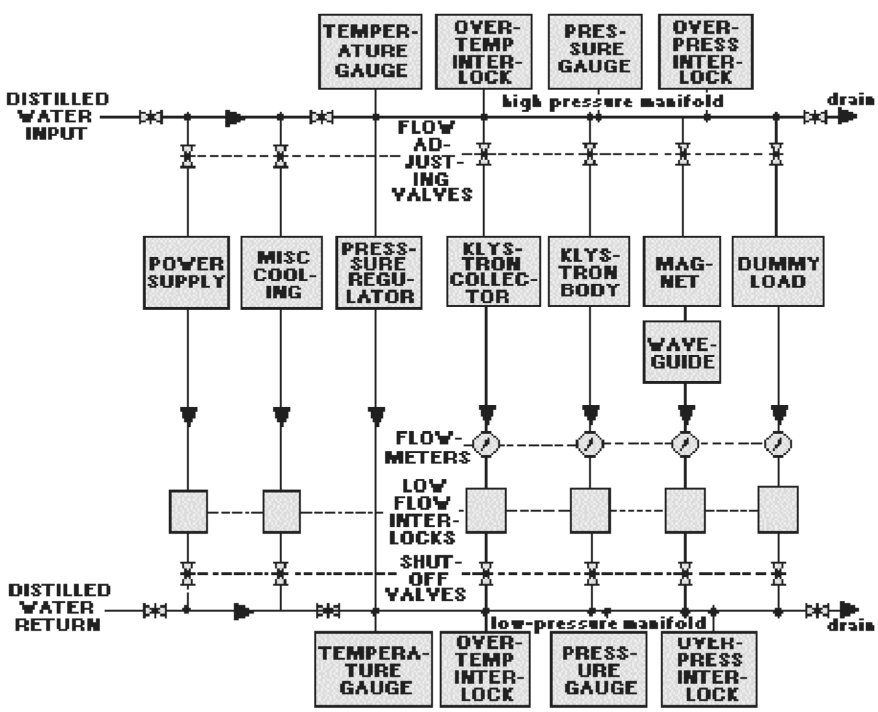


Figure 4-15.—Typical transmitter cooling system.

Distilled water is one of the best mediums for cooling high-power components, and, in many cases, the only medium that may be used.

For a distilled-water-cooling system to operate satisfactorily, the temperature, quantity, purity, flow, and pressure of the water must be controlled. This control is provided by various valves, regulators, sensors, meters, and instruments that measure the necessary characteristics and provide the required regulation.

Liquid-cooling systems consist of a sea water or a chilled (fresh) water section that cools the distilled water circulating through the electronic equipment. The main components of cooling systems are piping, valves, regulators, heat exchangers, strainers, circulating pumps, expansion tanks, gages, and demineralizers. Other specialized components are sometimes necessary to monitor cooling water to the electronic equipment.

A typical liquid-cooling system is composed of a PRIMARY LOOP and a SECONDARY LOOP (figure 4-16). The primary loop provides the initial source of cooling water and the secondary loop transfers the heat load from the electronic equipment to the primary loop. The source of cooling water for the primary loop is either sea water from a sea water supply or chilled water from the ship's air-conditioning plant. The cooling water used in the secondary loop is distilled water. Ultrapure systems are maintained by a demineralizer and use double-distilled water obtained through the Navy Supply System.

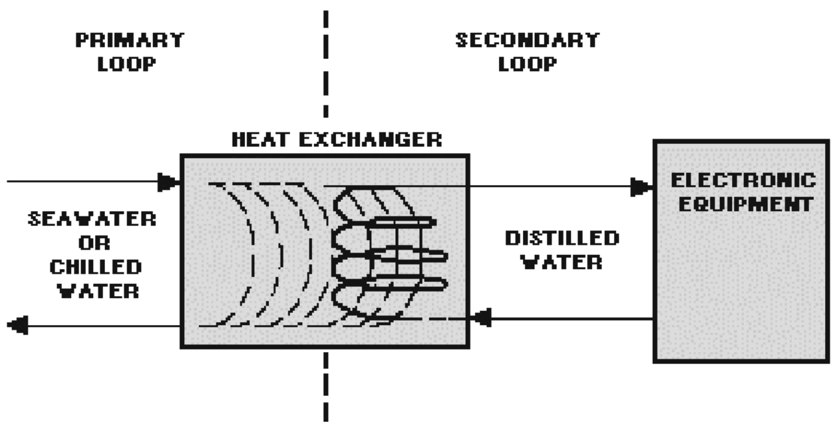


Figure 4-16.—Liquid cooling system block diagram.

Additional information about liquid cooling systems can be found in *Basic Liquid Cooling Systems for Shipboard Electronic Equipment Technician's Handbook*, NAVSEA 0948-LP-122-8010.

Q19. What type of cooling is used to control ambient room temperature?

Q20. A typical liquid-cooling system is composed of what loops?

Q21. What loop of a cooling system is often supplied by sea water?

SAFETY

Many safety and health hazards are involved with operating and maintaining high-power radars. These hazards result from high levels of rf radiation, X-ray emissions, the necessity of working aloft, and the generation of extremely high voltages.

Navy professionals are very safety conscious and, as a result, the number of accidents that occur on the job is small. Most of the safety precautions applicable to radar are published in radar technical manuals. Many of the safety regulations included in technical manuals are the result of actual experiences. Therefore, you should give them careful thought and strict observance.

RF RADIATION HAZARDS

Radar peak power may reach a million watts or more. Rf radiation hazards exist in the vicinity of radar transmitting antennas. These hazards are present not only in front of an antenna but also to the sides and sometimes even behind it because of spillover and reflection. At some frequencies, exposure to excessive levels of radiation will not produce a sufficient sensation of pain or discomfort to warn you of injury. If you suspect any injury, see your ship's doctor or corpsman. Be sure to acquaint yourself with the actual radiation hazard zones of the radars on your ship.

Personnel should observe the following precautions to ensure that persons are not exposed to harmful rf radiation:

- Visual inspection of feedhorns, open ends of waveguides, and any other opening that emits rf energy should not be made unless the equipment is properly secured and tagged for that purpose.
- Operating and maintenance personnel should observe all rf radiation hazard signs posted in the operating area.
- All personnel should observe rf radiation hazard (radhaz) warning signs (figure 4-17) that point out the existence of rf radiation hazards in a specific location or area. (You may encounter other types of rf radiation hazard signs, depending on the situation.)
- Ensure that radiation hazard warning signs are available and posted.
- Ensure that those radar antennas that normally rotate are rotated continuously while radiating or are trained to a known safe bearing.
- Ensure that those antennas that do not normally rotate are pointed away from inhabited areas (ships, piers, and the like) while radiating.
- Dummy loads should be employed where applicable in transmitting equipment during testing or checkout.

Figure 4-17.—Sample of one type of radhaz sign.

X-RAY EMISSIONS

X rays may be produced by the high-voltage electronic equipment in radars. X rays can penetrate human tissue and cause damage of a temporary or permanent nature. Unless the dosage is extremely high, no ill effects will be noticeable for days, weeks, or even years after the exposure.

The sources of these X rays are usually confined to magnetrons, klystrons, and cathode-ray tubes. Personnel should not linger near any of these types of equipments when the equipment covers have been removed. Klystrons, magnetrons, rectifiers, or other tubes that employ an excitation of 15,000 volts or more may emit X rays out to a few feet; thus, unshielded personnel standing or working close to the tubes will be endangered.

When performing maintenance on X-ray emitting devices, you should take the following precautions:

- Observe all warning signs (figure 4-18) on the equipment and all written precautions in the equipment technical manuals.

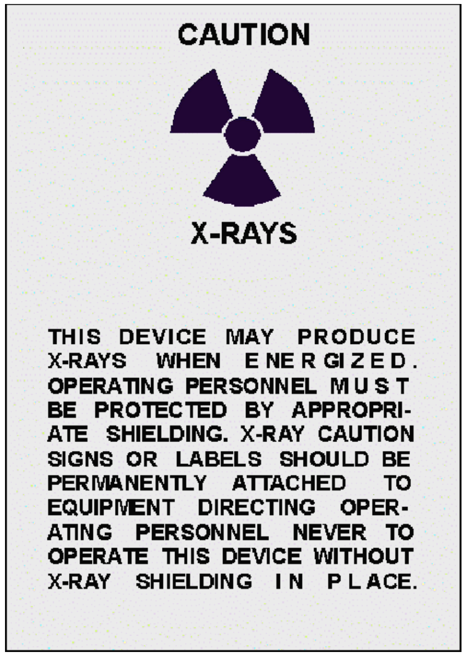


Figure 4-18.—X-ray caution label.

- Unless called for in the technical manual, do not bypass interlocks to permit the servicing of operating equipment with the X-ray shield removed.

- Be sure to replace all protective X-ray shielding when servicing is complete.

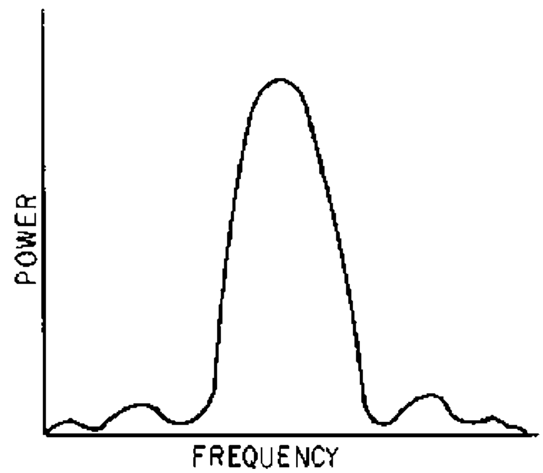
SUMMARY

This chapter has presented information on radar maintenance procedures. The information that follows summarizes the important points of this chapter.

Transmitter **PERFORMANCE CHECKS** are essential for you to maintain an efficient radar system. The transmitter output must be monitored closely for both frequency and power.

Transmitter energy is distributed symmetrically over a band of frequencies known as the **SPECTRUM**.

A **SPECTRUM CURVE** for a transmitter in good condition is shown in the illustration.



The **SPECTRUM ANALYZER** and the ECHO BOX are two instruments used to check transmitter performance.

One of the more important measurements that can be performed with the echo box is RING TIME. Ring time gives a relative indication of both transmitter output power and receiver sensitivity.

Transmitter **OUTPUT POWER MEASUREMENTS** are a good indication of overall transmitter operation. POWER MEASUREMENTS are usually of average power read in dBm. The average power dBm reading must be converted to watts and the peak power calculated. The formula for peak power is:

$$\text{peak power } (P_{pk}) = \frac{\text{average power } (P_{avg})}{\text{duty cycle}}$$

RECEIVER PERFORMANCE CHECKS determine receiver sensitivity, tr recovery time, and receiver bandwidth.

You usually measure receiver sensitivity by measuring the MINIMUM DISCERNIBLE SIGNAL (mds) using the pulse method.

TR RECOVERY time is the time required for the tr tube to DEIONIZE after each transmitted pulse. You should keep a graph of tr recovery time to determine when the tr tube should be replaced. If not replaced in a timely manner, a weak tr tube will allow damage to the radar receiver.

Few radars can function without SUPPORT SYSTEMS. These support systems include ELECTRICAL POWER, DRY-AIR, and LIQUID-COOLING SYSTEMS.

The radar technician should learn the source and distribution routes for NORMAL and EMERGENCY POWER for the radar.

The **DRY AIR** needed for electronic equipment can be supplied by the ship's electronics dry-air system through an air control panel or from local dehydrators.

Radar transmitters generate large amounts of heat. Most of this heat is dissipated by a combination of AIR CONDITIONING, CABINET AIR BLOWERS, and a DISTILLED-WATER COOLING system.

Personnel working on radars should always be aware of the hazards of RF RADIATION and X-RAY EMISSION.

All posted SAFETY PRECAUTIONS should be strictly observed.

ANSWERS TO QUESTIONS Q1. THROUGH Q21.

A1. Frequency distribution.

A2. In the center.

A3. Symmetrical above and below the carrier frequency.

A4. Power and frequency.

A5. Average power, pulse width, and prt.

A6. 1 milliwatt.

A7. Transmitter power loss.

A8. Minimum discernible signal (mds).

A9. Attenuations of the directional coupler and the connecting cable.

A10. Half-power points.

A11. Recovery time.

A12. Three.

A13. Automatically.

A14. Manual bus transfer (MBT) unit.

A15. Work backwards from the load to the source.

A16. Ship's central dry-air system.

A17. Degree of dehydration.

A18. Pressure.

A19. Air conditioning.

A20. Primary and secondary.

A21. The primary loop.

APPENDIX I

GLOSSARY

A-SCOPE—A radar display on which slant range is shown as the distance along a horizontal trace.

ACQUISITION—Operational phase of a fire- control or track radar during which the radar system searches a small volume of space in a prearranged pattern.

AIR-CONTROL PANEL—Panel that monitors the dry-air input at each user equipment.

ALTITUDE—Vertical distance of an aircraft or object above a given reference, such as ground or sea level.

AMBIGUOUS RETURNS—Echoes that exceed the prt of a radar and appear at incorrect ranges.

AMPLITRON—See Crossed-Field Amplifier.

ANTENNA BEAM WIDTH—Width of a radar beam measured between half-power points

ANTENNA SYSTEM—Routes rf energy from the transmitter, radiates the energy into space, receives echoes, and routes the echoes to the receiver.

ANTIJAMMING CIRCUIT—Electronic circuit used to minimize the effects of enemy countermeasures, thereby permitting radar echoes to be visible on the indicator.

ANTITRANSMIT-RECEIVE TUBE (atr)—Tube that isolates the transmitter from the antenna and receiver. Used in conjunction with tr tube.

ARTIFICIAL TRANSMISSION LINE—An LC network that is designed to simulate characteristics of a transmission line.

ASYMMETRICAL MULTIVIBRATOR—Multivibrator that generates rectangular waves.

AUTOMATIC GAIN CONTROL—Circuit used to vary radar receiver gain for best reception of signals that have widely varying amplitudes.

AVERAGE POWER—Output power of a transmitter as measured from the start of one pulse to the start of the next pulse.

AZIMUTH—Angular measurement in the horizontal plane in a clockwise direction.

BALANCED MIXER—Waveguide arrangement that resembles a T and uses crystals for coupling the output to a balanced transformer.

BEAM—See Lobe.

BEARING RESOLUTION—Ability of a radar to distinguish between targets that are close together in bearing.

BEAT FREQUENCIES—Difference and sum frequencies which result from combining two different frequencies.

BLIP—See Pip.

BLOCKING—A condition in an amplifier, caused by overdriving one or more stages, in which the amplifier is insensitive to small signals immediately after reception of a large signal.

BROADSIDE ARRAY—An antenna array in which the direction of maximum radiation is perpendicular to the plane of the array.

BUFFER AMPLIFIER STAGE—Amplifier stage that isolates one circuit from another.

CARRIER FREQUENCY—The frequency of an unmodulated transmitter output.

CARRIER-CONTROLLED APPROACH—Shipboard radar system used to guide aircraft to safe landings in poor visibility conditions.

CLUTTER—Confusing, unwanted echoes that interfere with the observation of desired signals on a radar indicator.

COHERENCE—A definite phase relationship between two energy waves, such as transmitted frequency and reference frequency.

COHERENT OSCILLATOR—In cw radar this oscillator supplies phase references to provide coherent video from target returns.

CONICAL SCANNING—Scanning in which the movement of the beam describes a cone, the axis of which coincides with that of the reflector.

CONTACT—In radar, an object that reflects rf energy; target.

CORNER REFLECTOR—Two flat reflectors that meet at an angle and are normally fed by a half-wave radiator.

CROSSED-FIELD AMPLIFIER—High-power electron tube that converts dc to microwave power by a combination of crossed electric and magnetic fields.

CYLINDRICAL PARABOLIC REFLECTOR—A parabolically shaped reflector that resembles part of a cylinder.

DEFLECTION COILS—In a cathode-ray tube, coils used to bend an electron beam a desired amount.

DEIONIZATION TIME—In a spark gap, the time required for ionized gas to return to its neutral state after the spark is removed.

DESIGNATION—Operational phase of a fire- control or track radar during which the radar is directed to the general direction of a desired target.

DIFFERENCE FREQUENCY—See Beat Frequency.

DIODE DETECTOR—A demodulator that uses one or more diodes to provide a rectified output with an average value that is proportional to the original modulation.

DIRECTIONAL ANTENNA—An antenna that radiates most effectively in only one direction.

DIRECTIVITY—Ability of an antenna to radiate or receive more energy in some directions than in others. The degree of sharpness of the antenna beam.

DISCRIMINATOR—A circuit in which amplitude variations are derived in response to phase or frequency variations.

DISTILLED WATER—Water that has been purified through a process of evaporation and condensation.

DOPPLER EFFECT—In radar, the change in frequency of a received signal caused by the relative motion between the radar and the target.

DOPPLER FREQUENCY—Difference between transmitted and reflected frequencies; caused by the Doppler effect.

DOUBLE-MODING—In a transmitter output tube, the abrupt and random change from one frequency to another.

DRY-AIR SYSTEM—Provides dehumidified air for electronic equipment that is moisture critical.

DUCTING—Trapping of an rf wave between two layers of the earth's atmosphere or between an atmospheric layer and the earth.

DUPLEXER—A radar device that switches the antenna from the transmitter to the receiver and vice versa.

DUTY CYCLE—In a transmitter, ratio of time on to time off.

ECHO—The rf signal reflected back from a radar target.

ECHO BOX—A resonant cavity device that is used to check the overall performance of a radar system. It receives a portion of the transmitted pulse and retransmits it back to the receiver as a slowly decaying transient.

ELECTRICAL POWER SYSTEM—Provides the necessary input power.

ELECTRONIC COUNTER-COUNTERMEASURES (ECCM) CIRCUITS—See Antijamming Circuits.

ELECTRONIC EQUIPMENT DEHYDRATOR—Provides alternate dry-air input in the event of failure of the central dry-air system. May include a compressor.

ELECTRONIC FREQUENCY COUNTER—An instrument that counts the number of cycles (pulses) occurring during a precise time interval.

ELECTRONIC SCANNING—Scanning in which the axis of the beam is moved, relative to the antenna axis, in a desired pattern.

ELECTRONICS DRY-AIR BRANCH—A common line for providing dry air to various electronic equipment, such as search radar, fire-control radar, and repeaters.

ELEVATION ANGLE—The angle between the horizontal plane and the line of sight.

EMERGENCY POWER—Temporary source of limited electrical power used upon the loss of the normal power source.

EXTERNALLY SYNCHRONIZED RADAR—Radar system in which timing pulses are generated by a master oscillator external to the transmitter.

FAST-TIME-CONSTANT CIRCUIT—Differentiator circuit in the first video amplifier that allows only the leading edges of target returns, no matter how small or large, to be used.

FEEDBACK—The return of a portion of the output of a circuit to its input.

FEEDHORN—A horn radiator used to feed a reflector.

FIRST DETECTOR—See Mixer.

FREQUENCY COMPENSATION NETWORK—Circuit modification used to improve or broaden the linearity of its frequency response.

FREQUENCY SCANNING—Varying the output frequency to achieve electronic scanning.

FREQUENCY SPECTRUM—In a radar, the entire range of frequencies contained in an rf pulse or signal.

FREQUENCY SYNTHESIZER—A bank of oscillators in which the outputs can be mixed in various combinations to produce a wide range of frequencies.

GAIN—Any increase in the strength of a signal.

GATED AGC—Circuit that permits automatic gain control to function only during short time intervals.

GLOW DISCHARGE—Discharge of electricity through a gas in an electron tube.

GROUND CLUTTER—Unwanted echoes from surrounding land masses that appear on a radar indicator.

GROUND RANGE—The distance on the surface of the earth between a radar and its target. Equal to slant range only if both radar and target are at the same altitude.

GROUND-CONTROLLED APPROACH—Radar system used to guide aircraft to safe landings in poor visibility conditions.

GUIDANCE RADAR—System which provides information that is used to guide a missile to a target.

HALF-POWER POINT—A point on a waveform or radar beam that corresponds to half the power of the maximum power point.

HARD-TUBE MODULATOR—A high-vacuum electron tube modulator that uses a driver for pulse forming.

HEIGHT-FINDING RADAR—Radar that provides target altitude, range, and bearing data.

HITS PER SCAN—The number of times an rf beam strikes a target per antenna revolution.

HORIZONTAL PLANE—Imaginary plane that is tangent (or parallel) to the earth's surface at a given location.

HORN ANTENNA—See Horn Radiator.

HORN RADIATOR—A tubular or rectangular microwave antenna that is tapered and is widest at the open end.

HYBRID RING—A circular waveguide arrangement with four branches. When properly terminated, energy is transferred from any one branch into any two of the remaining three branches.

HYBRID MIXER—See Balanced Mixer.

IF AMPLIFIER—Usually a narrow-bandwidth IF amplifier that is tuned to one of the output frequencies produced by the mixer.

INDEX OF REFRACTION—The degree of bending of an rf wave when passing from one medium to another.

INDICATOR—Equipment that provides a visual presentation of target position information.

INSTANTANEOUS AUTOMATIC GAIN CONTROL (IAGC)—Circuit that can vary the gain of the radar receiver with each input pulse to maintain the output peak amplitude nearly constant.

INTERMEDIATE FREQUENCY (IF)—A lower frequency to which an rf echo is converted for ease of amplification.

KEEP-ALIVE CURRENT—See Keep-Alive Voltage.

KEEP-ALIVE VOLTAGE—Dc voltage applied to a tr gap electrode to produce a glow discharge that allows the tube to ionize faster when the transmitter fires.

KEYED-OSCILLATOR TRANSMITTER—A transmitter in which one stage is used to produce the rf pulse.

KEYER—See Synchronizer.

KLYSTRON POWER AMPLIFIER—Multicavity microwave electron tube that uses velocity modulation.

LIN-LOG AMPLIFIER—Amplifier in which the response is linear for weak signals and logarithmic for large signals.

LINE OF SIGHT—Straight line from a radar antenna to a target.

LINE-PULSING MODULATOR—Circuit that stores energy and forms pulses in the same circuit element, usually the pulse-forming network (pfn).

LIQUID-COOLING SYSTEM—Source of cooling for high-heat producing equipments, such as microwave components, radar repeaters, and transmitters.

LOBE—An area of greater signal strength in the transmission pattern of an antenna.

LOGARITHMIC RECEIVER—Receiver that uses a linear logarithmic amplifier (lin-log) instead of a normal linear amplifier.

LOW-NOISE AMPLIFIER—See Preamplifier.

MAGIC T—See Balanced Mixer.

MAGNETRON OSCILLATOR—Electron tube that provides a high power output. Theory of operation is based on interaction of electrons with the crossed electric and magnetic fields in a resonant cavity.

MASTER OSCILLATOR—In a transmitter, the oscillator that establishes the carrier frequency of the output.

MECHANICAL SCANNING—The reflector, its feed source, or the entire antenna is moved in a desired pattern.

MINIMUM DISCERNIBLE SIGNAL (MDS)—The weakest signal that produces a usable signal at the output of a receiver. The weaker the signal, the more sensitive the receiver.

MIXER—In radar, a circuit that combines the received rf signal with a local-oscillator signal to effectively convert the received signal to a lower IF frequency signal.

MODE SHIFTING—In a magnetron, shifting from one mode to another during a pulse.

MODE SKIPPING—Rather than firing on each successive pulse as desired, the magnetron fires randomly.

MODES—Operational phases (of a radar).

MODULATOR SWITCHING DEVICE—Controls the on (discharge) and off (charge) time of the modulator.

MODULATOR—Produces a high-voltage pulse that turns the transmitter on and off.

MONOPULSE (SIMULTANEOUS) LOBING—Radar receiving method using two or more (usually four) partially overlapping lobes. Sum and difference channels locate the target with respect to the axis of the antenna.

MONOPULSE RADAR—A radar that gets the range, bearing, and elevation position data of a target from a single pulse.

MONOPULSE RECEIVER—See Monopulse Lobing.

MOISTURE LAPSE—Abnormal variation of moisture content at different altitudes because of high moisture located just above large bodies of water.

MOVING TARGET INDICATOR—A device that limits the display of radar information to moving targets.

NAUTICAL MILE—The length of a minute of arc of a great circle of the earth (6,076 ft.)

NAUTICAL RADAR MILE—See Radar Mile.

NOISE—In radar, erratic or random deflection or intensity of the indicator sweep that tends to mask small echo signals.

NOISE FIGURE—The ratio of output noise to input noise in a receiver.

NUTATING—Moving an antenna feed point in a conical pattern so that the polarization of the beam does not change.

OMNIDIRECTIONAL ANTENNA—An antenna that radiates equally in all directions (nondirectional).

ORANGE-PEEL PARABOLOID—A section of a complete circular paraboloid that is narrow in the horizontal plane and wide in the vertical plane.

PARABOLIC REFLECTOR—An antenna reflector in the shape of a parabola. It converts spherical wavefronts from the radiating element into plane wavefronts.

PARALLEL-CONNECTED DUPLEXER—Configuration in which the tr spark gap is connected across the two legs of the transmission line one-quarter wavelength from the Tjunction.

PARASITIC ARRAY—An antenna array containing one or more elements not connected to the transmission line.

PEAK POWER—Maximum power of the rf pulse from a radar transmitter.

PERSISTANCE—The length of time a phosphor dot glows on a crt before disappearing.

PHANTASTRON—A variable-length sawtooth generator used to produce a sweep on an A-scope.

PIP (BLIP)—On a crt display, a spot of light or a base-line irregularity representing the radar echo.

PLANE WAVEFRONTS—Waves of energy that are flat, parallel planes and perpendicular to the direction of propagation.

PLANNED-POSITION INDICATOR—A radar display in which range is indicated by the distance of a bright spot or pip from the center of the screen and the bearing is indicated by the radial angle of the spot.

POWER GAIN—In an antenna, the ratio of its radiated power to that of a reference.

POWER-AMPLIFIER (CHAIN) TRANSMITTER—Transmitter that uses a series of power amplifiers to create a high level of power.

PREAMPLIFIER (PREAMP)—An amplifier that raises the output of a low-level source for further processing without appreciable degradation of the signal-to-noise ratio.

PRIMARY LOOP—In a cooling system, the primary source of cooling for the distilled water.

PROBE COUPLER—A resonant conductor placed in a waveguide or cavity to insert or extract energy.

PULSE WIDTH—Duration of time between the leading and trailing edges of a pulse.

PULSE-FORMING NETWORK (PFN)—An LC network that alternately stores and releases energy in an approximately rectangular wave.

PULSE-REPETITION RATE (PRR)—Average number of pulses per unit of time; pulse rate.

PULSE-REPETITION FREQUENCY (PRF)—The rate at which pulses are transmitted, given in hertz or pulses per second; reciprocal of pulse-repetition time.

PULSE-REPETITION TIME (PRT)—Interval between the start of one pulse and the start of the next pulse; reciprocal of pulse-repetition frequency.

RADAR—An acronym for RAdio Detecting And Ranging.

RADAR ALTIMETER—Airborne radar that measures the distance of the aircraft above the ground.

RADAR BEAM—The space in front of a radar antenna where a target can be effectively detected or tracked.

RADAR DISTRIBUTION SWITCHBOARD—An electrical switching panel used to connect inputs from any of several radars to repeaters.

RADAR MILE—Time interval (12.36 microseconds) for rf energy to travel out from a radar to a target and back to the radar; radar nautical mile.

RADAR TEST SET—Combination of several test circuits and equipment used to test various characteristics of a radar.

RANGE—The length of a straight line between a radar set and a target.

RANGE-HEIGHT INDICATOR—A radar display on which slant range is shown along the X axis and height along the Y axis.

RANGE-GATE—A movable gate used to select radar echoes from a very short-range interval.

RANGE MARKER—A movable vertical pulse on an A-scope or ring on a ppi scope used to measure the range of an echo or to calibrate the range scale.

RANGE RESOLUTION—Ability of a radar to distinguish between targets that are close together.

RANGE STEP—On an A-scope sweep, vertical displacement used to measure the range of an echo.

RECEIVER—In radar, a unit that converts rf echoes to video and/or audio signals.

RECEIVER SENSITIVITY—The degree to which a receiver can usefully detect a weak signal; the lower limit of useful signal input to the receiver.

RECOVERY TIME—In a radar, the time interval between the end of the transmitted pulse and the time when echo signals are no longer attenuated by the tr gap.

REFLECTING OBJECT—In radar, an air or surface contact that provides an echo.

REFLEX KLYSTRON—A microwave oscillator that is tuned by changing the repeller voltage.

REFRACTION—The bending of rf waves as the waves pass through mediums of different density.

REFRACTIVE INDEX—In a wave-transmission medium, the ratio between the phase velocity in free space and in the medium.

REGENERATION—See Feedback.

RELATIVE BEARING—Bearing of target measured in a clockwise direction from "dead ahead" of a ship or plane.

RESONANCE CHAMBER—See Echo Box.

RETURN—The rf signal reflected back from a radar target; echo.

RF RADIATION HAZARD—Health hazard caused by exposure to electromagnetic radiation or high-energy particles (ions). Abbreviated RADHAZ.

RING TIME—In radar, the time during which the output of an echo box remains above a specified level.

RINGING—Rf oscillations caused by shock excitation of a resonant circuit (cavity).

SCANNING—Systematic movement of a radar beam to cover a definite pattern or area in space.

SEA CLUTTER—Unwanted echoes from the irregular surface of the sea that appear on a radar indicator.

SEARCH RADAR SYSTEM—Early-warning device that searches a fixed volume of space.

SECOND DETECTOR DEMODULATOR—The part of the receiver that separates the audio or video component from the modulated intermediate frequency.

SECOND-SWEEP ECHOES—See Ambiguous Returns.

SECONDARY LOOP—In a cooling system, the loop that transfers the heat from the heat source (electronic equipment) to the primary loop; usually distilled water.

SELF-SYNCHRONIZED RADAR—A type of radar in which the timing pulses are generated within the transmitter.

SENSITIVITY TIME CONTROL (STC)—A circuit that varies the gain of a receiver as a function of time.

SERIES-CONNECTED DUPLEXER—Configuration in which the tr spark gap is connected in series in one leg of the transmission line one-half wavelength away from the T- junction.

SHADOW—A dead spot (minimum radiation) caused by the physical obstruction of transmitted waves by a feed horn.

SINGLE-ENDED MIXER—See Unbalanced Crystal Mixer.

SINGLE, STATIONARY-LOBE SCANNING SYSTEM—Antenna (with a single, stationary beam) that is rotated to obtain 360-degree coverage.

SLANT RANGE—See Range.

SPECTRUM ANALYZER—A test instrument that provides a visual display of the frequency distribution of a transmitter output.

SPHERICAL WAVEFRONTS—Waves of energy that spread out in concentric circles.

STABILITY—In a magnetron, the ability to maintain normal operating characteristics.

STATUTE MILE—I5,280 ft.

STUB—A short section of transmission line connected in parallel with the main transmission line.

SCINTILLATION—Apparent change in target reflectivity. Motion of the target causes radar pulses to bounce off different parts of the target, such as fuselage and wingtip.

SUPERHETERODYNE RECEIVER—A type of receiver that uses a mixer to convert the rf echo to an IF signal for amplification.

SUPPORT SYSTEM—For a radar, a system that provides an auxiliary input, such as dry air, electrical power, or liquid cooling.

SYMMETRICAL MULTIVIBRATOR—Circuit that generates square waves.

SYNCHRONIZER—Circuit that supplies timing signals to other radar components.

TARGET—In radar, a specific object of radar search or detection.

TARGET RESOLUTION—The ability of a radar to distinguish between two or more targets that are close to each other.

THREE-DIMENSIONAL RADAR (3D)—Measures the range, bearing, and altitude of a target.

THYRATRON—Gas tube used as a modulator switching device.

TIMER—See Synchronizer.

TR RECOVERY TIME—Time required for a fired tr or atr tube to deionize to a normal level of conductance.

TRACK—Operational phase of a fire-control or track radar during which the radar beam is kept on the target.

TRACK RADAR—Radar that provides continuous range, bearing, and elevation data by keeping the rf beam on the target.

TRANSMIT-RECEIVE TUBE (TR)—Gas-filled rf switch that is used as a duplexer.

TRANSMITTER—Equipment that generates, amplifies, and modulates electromagnetic energy.

TRANSMITTER FREQUENCY (CARRIER FREQUENCY)—The frequency of the unmodulated output of a transmitter.

TRAVERSE (BEARING) SIGNAL—In a monopulse radar system, the combination of individual lobe signals that represents target offset direction and amplitude from the antenna axis.

TRIGGER GENERATOR—See Synchronizer.

TRIGGER PULSES—In radar, pulses that are used to initiate specific events.

TRUE BEARING—Angle between a target and true north measured clockwise in the horizontal plane.

TRUE NORTH—Geographic north.

TRUNCATED PARABOLOID—A paraboloid reflector that has been cut away at the top and bottom to increase beam width in the vertical plane.

TWO-DIMENSIONAL RADAR (2D)—Measures the range and bearing to a target.

UNBALANCED CRYSTAL MIXER—Circuit consisting of a section of coaxial transmission line one-half wavelength long that is tuned to the difference (intermediate) frequency between the local oscillator and rf echo signals.

VERTICAL PLANE—Imaginary plane that is perpendicular to the horizontal plane.

VIDEO ENHANCEMENT FEATURES—See Antijamming Circuits.

VOLTAGE STANDING WAVE RATIO (VSWR)—In a waveguide, the ratio of the electric field at a maximum point to that of an adjacent minimum point.

WAVEGUIDE DUPLEXER—Consists of tr and atr tubes housed in a resonant cavity attached to a waveguide system.

WAVEMETER—An instrument for measuring the wavelength of an rf wave.

X-RAY EMISSION—Penetrating radiation similar to light, but with shorter wavelength, that can penetrate human tissue.

APPENDIX II

REFERENCES USED TO DEVELOP THIS NRTC

NOTE: Although the following references were current when this NRTC was published, their continued currency cannot be assured. When consulting these references, keep in mind that they may have been revised to reflect new technology or revised methods, practices, or procedures; therefore, you need to ensure that you are studying the latest references.

CHAPTER ONE

Shipboard Electronics Material Officer, NAVEDTRA 10478-A, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1982.

Basic Electronics, Vol. 1, NAVPERS 10087-C, Bureau of Naval Personnel, Washington, DC.

Aviation Fire Control Technician 3 & 2, NAVEDTRA 10387-B, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1977.

Fire Control Technician M 3, NAVEDTRA 10224, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1981.

Fire Control Technician M 3 & 2, NAVEDTRA 10209-A, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1974.

CHAPTER TWO

Aviation Fire Control Technician 3 & 2, NAVEDTRA 10387-B, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1977.

Basic Electronics, Vol. 2, NAVEDTRA 10087-CI, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1979.

Fire Control Technician M 1 & C, NAVPERS 10210, Bureau of Naval Personnel, Washington, D.C.

Fire Control Technician M 3 & 2, NAVEDTRA 10209-A, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1974.

CHAPTER THREE

Shipboard Electronics Material Officer, NAVEDTRA 10478-A, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1982.

Basic Electronics, Vol. 2, NAVEDTRA 10087-C1, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1979.

Aviation Fire Control Technician 3 & 2, NAVEDTRA 10387-B, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1977.

CHAPTER FOUR

Fire Control Technician G 3 & 2, NAVEDTRA 10207-B, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1976.

Electronics Technician 3 & 2, Vol. 2, NAVEDTRA 10195-A, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1979.

Aviation Fire Control Technician 3 & 2, NAVEDTRA 10387-B, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1977.

Fire Control Technician M 3, NAVEDTRA 10224, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1981.

Basic Electronics, Vol 2, NAVEDTRA 10087-C 1, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1979.

MODULE 18

INDEX

S

T

V

W

X

ASSIGNMENT 1

Textbook assignment: Chapter 1, “Radar Fundamentals,” pages 1-1 through 1-45.

1-1. Radar uses what form of energy to detect ships, planes, and land masses?

1. Sound waves
2. Visible light
3. Infrared radiation
4. Electromagnetic energy

1-2. What radar measurement of an object is referenced to true north?

1. Height
2. Surface angle
3. Vertical angle
4. One-way distance

1-3. The elevation angle to the target is the angle between which of the following references?

1. Ship's heading and line of sight
2. Vertical plane and line of sight
3. Horizontal plane and line of sight
4. Vertical plane and horizontal plane

1-4. Electromagnetic energy travels through air at approximately what speed?

1. 984 feet per microsecond
2. 186,000 statute miles per second
3. 162,000 nautical miles per second
4. Each of the above is correct

1-5. For an object that is detected 15 miles from a radar set, what is the approximate time required for the rf energy to travel to and return from the object?

1. 7 microseconds
2. 185 microseconds
3. 271 microseconds
4. 927 microseconds

1-6. The minimum range of a radar depends on the length of time of the transmitter pulse (pulse width) and recovery time. During this period, the radar can NOT receive energy. A radar set with a pulse width of 5 microseconds and a recovery time of 0. 2 microseconds has a minimum range of approximately

1. 614 yards
2. 787 yards
3. 852 yards
4. 4,100 yards

1-7. Of the following radar characteristics, which has NO effect on maximum range capability?

1. Recovery time
2. Carrier frequency
3. Receiver sensitivity
4. Pulse-repetition frequency

1-8. Which of the following characteristics of a radar system determines the degree to which the radiated energy is affected by atmospheric conditions?

1. Pulse peak power
2. Carrier frequency
3. Receiver sensitivity
4. Pulse-repetition frequency

1-9. Pulse-repetition time is described as the

1. reciprocal of pulse width
2. period of time from the beginning to the end of a transmitter pulse
3. period of time required for the pulse to travel to the target and return
4. period of time from the beginning of one transmitter pulse to the beginning of the next

1-10. Which of the following terms describes radar returns that exceed the prt of the radar?

1. Clutter
2. Ambiguous
3. Reciprocal
4. Acquisition

1-11. What type of transmitter power is measured over a period of time?

1. Peak
2. Return
3. Average
4. Reciprocal

1-12. Which of the following formulas can be used to compute the duty cycle of a radar set?

1. $\frac{\text{prt}}{\text{pw}}$

2. $\frac{\text{pw}}{\text{prt}}$

3. $\frac{\text{peak power}}{\text{average power}}$

4. $\text{average power} \times \text{peak power}$

IN ANSWERING QUESTIONS 1-13 AND 1-14, ASSUME THAT A RADAR SET HAS THE FOLLOWING CHARACTERISTICS:
AVERAGE POWER = 700 WATTS
PULSE WIDTH = 3. 5 MICROSECONDS
PRF = 400 HERTZ

1-13. What is the peak power (in kilowatts) for this radar set?

1. 500
2. 1,500
3. 2,000
4. 2,500

1-14. What is the pulse-repetition time (in microseconds) for this radar set?

1. 1,500
2. 2,000
3. 2,500
4. 4,000

1-15. For a radar set with a pulse width of 25 microseconds and a prf of 600 pulses per second, what is the duty cycle?

1. 0.015
2. 0.024
3. 0.15
4. 0.24

1-16. What is the maximum radar horizon distance for a radar set with an antenna 100 feet above the surface of the earth?

1. 8 miles
2. 12.5 miles
3. 80 miles
4. 125 miles

1-17. A radar set with a prf of 1,000 pps and an antenna rotation rate of 15 rpm produces what maximum number of pulses per degree?

1. 2.7
2. 11.1
3. 14.4
4. 36

1-18. The relative bearing of a radar echo is measured with respect to which of the following reference points?

1. True north
2. Magnetic north
3. Centerline of your own ship
4. The position of the antenna

1-19. Altitude- or height-finding radar systems require a beam with which of the following characteristics?

1. Narrow in the vertical plane
2. Narrow in the horizontal plane
3. Broad in the horizontal plane
4. Broad in the vertical plane

1-20. Target resolution is the ability of a radar to distinguish between targets that are at nearly the same range and/or bearing. Range resolution is dependent on which of the following factors?

1. Peak power and beam width
2. Pulse width and beam width
3. Pulse width and target size
4. Peak power and target size

1-21. What is the approximate range resolution (in yards) of a radar set with a pulse width of 15 microseconds and a peak power of 2 kilowatts?

1. 10
2. 24
3. 1,093
4. 2,460

1-22. The width of a radar beam and range to a detected object are the determining factors in which of the following radar system characteristics?

1. Range resolution
2. Bearing resolution
3. Beam half-power points
4. Altitude detection accuracy

1-23. Which of the following factors has/have the greatest effect on the accuracy of a pulse radar?

1. Resolution
2. Average power
3. Atmospheric conditions
4. Each of the above

1-24. To detect nearby objects, the output pulse of a radar transmitter should possess which of the following characteristic shapes?

1. Narrow and square
2. Narrow with a sloping trailing edge
3. Wide and trapezoidal
4. Wide and square

1-25. Refractions and speed changes are known to occur in electromagnetic wavefronts. These phenomena are caused by which of the following conditions?

1. Temperature
2. Vapor content
3. Atmospheric pressure
4. All of the above

1-26. Temperature inversions and/or moisture lapses in the atmosphere may extend or reduce the range of a radar by creating

1. ducts
2. ionic layers
3. surface waves
4. reflective layers

1-27. In a pulse radar system, which of the following components should you expect to find?

1. Synchronizer and transmitter
2. Duplexer and antenna system
3. Receiver and indicator
4. All of the above

1-28. In a pulse radar system, what component controls timing throughout the system?

1. Power supply
2. Synchronizer
3. Indicator
4. Receiver

1-29. What component of the transmitter supplies the trigger pulse, which acts as a switch to turn the klystron on and off?

1. Magnetron
2. Modulator
3. Indicator
4. Power supply

1-30. Which of the following radar components allows the use of one antenna for both transmitting and receiving radar energy?

1. Synchronizer
2. Modulator
3. Duplexer
4. Mixer

1-31. To make amplification of received rf energy easier, the returning rf signal is converted to a lower, intermediate frequency. What component of the radar system performs this function?

1. A duo-diode duplexer
2. A three-stage synchronizer
3. A cross-field amplifier
4. A superheterodyne receiver

1-32. The sweep frequency of a radar indicator is determined by what parameter of the radar system?

1. Duty cycle
2. Pulse width
3. Carrier frequency
4. Pulse-repetition frequency

1-33. The type and method of scanning used by a radar system depend upon which of the following radar system design considerations?

1. Antenna size
2. Type of radar
3. Purpose of the radar
4. All of the above

1-34. In a single stationary-lobe scanning system, complete azimuth coverage is achieved by which of the following methods?

1. Multiple overlapping beams
2. An omnidirectional antenna
3. A continuously rotating antenna
4. Very wide, flat beams

1-35. Which of the following methods can be used to achieve radar-beam scanning?

1. Mechanical
2. Electronic
3. Combined mechanical and electronic
4. Each of the above

1-36. In a conical-scan antenna, nutation of the radar beam is usually accomplished by which of the following methods?

1. By moving the reflector
2. By moving the feed point
3. By varying the signal phase at the feed point
4. By moving both the feed point and the reflector

1-37. At any given distance from the antenna, the radar beam axis of a conical-scan antenna follows what pattern?

1. A circle around the scan axis
2. An ellipse in the vertical plane
3. An ellipse in the horizontal plane
4. Two circles covering the scan axis in figure eights

1-38. In a monopulse scanning radar, the relative position of a target with respect to the radar-beam axis is determined by comparing which of the following signal components?

1. The phases of the radiated rf energy
2. The phases of the returning rf energy
3. The amplitudes of the returning rf energy in each horn
4. The amplitudes of the returning rf energy in each successive pulse

1-39. In a monopulse scanning system, which of the following feedhorn signal combinations makes up the bearing signal?

1. (A + D) - (A + C)
2. (A + B) - (C + D)
3. (A + C) - (B + D)
4. A + B + C + D

1-40. In a monopulse scanning system, which of the following combinations make up the range signal?

1. (A + D) - (A + C)
2. (A + B) - (C + D)
3. (A + C) - (B + D)
4. A + B + C + D

1-41. Monopulse receivers use what signal as a phase reference?

1. Range
2. Traverse
3. Elevation
4. Angle-tracking

1-42. In the cw radar transmission method, the Doppler effect provides which of the following target information?

1. Speed of the target
2. Presence of the target
3. Both 1 and 2 above
4. Relative bearing of the target

1-43. The frequency of the returned signal increases when a target is approaching and decreases when a target is moving away in which of the following types of radar systems?

1. Search
2. Doppler
3. Pulse-modulation
4. Frequency-modulation

1-44. Continuous-wave radar that uses the Doppler effect is best used in detecting which of the following types of targets?

1. Stationary
2. Past-moving
3. Targets with a high degree of range resolution
4. Targets with a high degree of bearing resolution

1-45. Range information can be obtained in a Doppler radar by which of the following methods?

1. Sweeping the transmitter frequency
2. Using two separate transmitters
3. Both 1 and 2 above
4. Using two separate antennas

1-46. Frequency-modulated radars transmit a wave that continuously changes in frequency about a center frequency. Using frequency modulation, the range to a target is determined by using which of the following methods?

1. By comparing the frequency of the transmitted signal with the returned frequency from the target
2. By comparing the magnitude of transmitted pulses with the magnitude of returned pulses
3. By comparing the velocity of the received energy with the velocity of the transmitted energy
4. By measuring the Doppler shift that occurs in the returning signal

1-47. Which of the following statements describes the advantage of using pulse modulation (pm) rather than continuous-wave (cw) in a radar system?

1. Pm may be used to detect moving targets; cw is effective only against stationary targets
2. Pm may be used to determine relative velocity much more accurately than cw
3. Pm does not require frequency stabilization of the carrier wave; cw does
4. Pm does not depend on target motion; cw does

1-48. In Doppler radar, some definite relationship must exist between the transmitted frequency and the reference frequency. For what purpose is this relationship used?

1. To detect the heterodyning signal
2. To detect the continuous-wave signal
3. To detect the Doppler shift of the received signal
4. To detect the frequency shift of the transmitted frequency

1-49. Which of the following JETDS classifications identifies a shipboard fire control radar set?

1. AN/SPS-39
2. AN/SPG-55
3. AN/APG-12
4. AN/MRC-20

1-50. For most military applications, which of the following radar systems is/are used?

1. Track radar only
2. Search radar only
3. Both track and search radars are used
4. Moving-target indicators

1-51. Detection of surface and low-altitude air targets is the primary purpose of which of the following types of radar?

1. Surface search
2. Height finding
3. Fire control
4. Air search

1-52. Which of the following are typical characteristics of surface-search radars?

1. High pulse-repetition rates
2. Narrow pulse widths
3. High frequencies
4. All of the above

1-53. Long-range aircraft detection is provided by which of the following types of radar?

1. Track
2. Air search
3. Fire control
4. Surface search

1-54. Which of the following are characteristics of a typical air-search radar?

1. Low frequency
2. Narrow pulse width
3. High pulse-repetition rate
4. All of the above

1-55. Which of the following types of radar provides accurate range, bearing, and altitude of aircraft?

1. Air search
2. Guidance
3. Height-finding
4. Surface search

1-56. Which of the following radars would most likely be used to direct CAP aircraft during an intercept?

1. Track radar
2. Fire-control radar
3. Surface-search radar
4. Three-coordinate radar

1-57. The range capability of a 3D radar is limited by which of the following characteristics?

1. Low prf
2. Long prt
3. Low output power
4. High operating frequency

1-58. Fire control radars must be directed to the general location of a desired target. This is because of which of the following characteristics?

1. Low output power level
2. Low degree of accuracy
3. Narrow beam pattern
4. Poor resolution

1-59. When a fire-control radar antenna approaches the general direction of a target, the radar enters which of the following modes of operation?

1. Acquisition
2. Designation
3. Lock-on
4. Track

1-60. Which of the following characteristics is/are typical of a fire-control radar?

1. Very high prf
2. Very narrow pw
3. Very narrow beam
4. All of the above

1-61. Complete control of a beam-rider missile requires what minimum number of radar beams?

1. 1
2. 2
3. 3
4. 4

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Radar Subsystems," pages 2-1 through 2-51.

2-1. Which of the following units of a radar system determines timing for all units of the radar?

1. Automatic tracker
2. Synchronizer
3. Transmitter
4. Receiver

2-2. Which of the following classifications describes a radar system that uses a master oscillator to produce timing pulses?

1. Externally synchronized
2. Self-synchronized
3. Unsynchronized
4. Free-running

2-3. Which of the following oscillators may be used as the master oscillator in a radar system?

1. A stable multivibrator
2. Sine-wave oscillator
3. Blocking oscillator
4. Each of the above

2-4. Which of the following oscillators, used in a synchronizer to provide timing trigger pulses, does NOT require a pulse-shaping circuit in its output?

1. Phase-shift oscillator
2. Square-wave oscillator
3. Sine-wave oscillator
4. Blocking oscillator

2-5. A radar system in which timing triggers are determined by the pulse-repetition rate of the modulator uses what type of synchronization?

1. Self
2. External
3. Free-running
4. Stable blocked

2-6. Which of the following radar indicator functions is/are controlled by gate pulses from the synchronizer?

1. Sweep duration
2. Sweep initiation
3. Range-mark generator gating
4. All of the above

2-7. Indicator sweep voltage in a radar is normally initiated at which of the following times?

1. At the same time as the transmitter trigger
2. Before the transmitter trigger
3. After the transmitter trigger
4. Each of the above

2-8. The flyback retrace of a radar system indicator is prevented from appearing on the cathode-ray tube by removing the

1. negative-intensity gate pulse applied to the control grid
2. positive-intensity gate pulse applied to the control grid
3. negative-intensity gate pulse applied to the cathode
4. positive-intensity gate pulse applied to the cathode

2-9. Which of the following characteristics is NOT a requirement of a basic radar system timing circuit?

1. It must be free-funning
2. Frequency must be variable
3. It must develop random frequencies
4. It should provide a stable frequency

2-10. Which of the following circuits converts square waves to positive and negative triggers?

1. A negative limiter
2. A positive limiter
3. A long-time-constant RC differentiator
4. A short-time-constant RC differentiator

2-11. Which of the following circuits is used to remove either the negative or positive triggers from the output of a sine-wave oscillator?

1. A clipper
2. A limiter
3. An LC network
4. A differentiator

2-12. When the master oscillator in a multivibrator timer is asymmetrical, the output of the master oscillator is in the form of

1. rectangular waves
2. negative pulses
3. square waves
4. sine waves

2-13. The positive and negative output pulses of the astable multivibrator are sent to the indicator for which of the following purposes?

1. To intensify the crt beam
2. To gate the range-marker generator
3. Both 1 and 2 above
4. To gate the transmitter output

2-14. Which of the following oscillators generates sharp trigger pulses without additional circuitry?

1. Sine-wave oscillator
2. Wien-bridge oscillator
3. One-shot multivibrator
4. Single-swing blocking oscillator

2-15. A radar transmitter is triggered directly by high voltage pulses from what unit?

1. The timer
2. The antenna
3. The indicator
4. The modulator

2-16. The peak power of a transmitted rf pulse depends on which of the following factors?

1. Width of the modulator pulse
2. Amplitude of the modulator pulse
3. Prf of trigger pulses from the timer
4. Delay time between trigger and modulator pulse outputs

2-17. The transmitter range timing circuit must be triggered at the instant the transmitted pulse leaves the transmitter. For which of the following reasons is this timing so important?

1. To ensure accurate range
2. The ensure long range targets are detected
3. To ensure that the target is "painted" on the crt by each transmitted pulse
4. To keep the magnetron oscillating at a fixed frequency

2-18. For proper operation of the magnetron, the modulator pulse must have which of the following characteristics?

1. A flat top
2. A steep leading edge
3. A steep trailing edge
4. All of the above

2-19. In order that nearby targets may be detected, which of the following characteristics must the transmitted pulse have?

1. A steep leading edge
2. A steep trailing edge
3. A sloping leading edge
4. A sloping trailing edge

2-20. Compared to the hard-tube modulator, the line-pulsing modulator has which of the following advantages?

1. It is more complex
2. It is more efficient
3. It is more sensitive to voltage changes
4. It requires a higher power-supply voltage

2-21. The modulator of a radar basically consists of a power supply, a switch, a storage element, and a/an

1. IF strip
2. oscillator
3. transmitter
4. charging impedance

2-22. Which of the following devices can be used as the storage element in a radar modulator?

1. A capacitor
2. A pulse-forming network
3. An artificial transmission line
4. Each of the above

2-23. A signal introduced at the input end of an artificial transmission line moves through the circuit to the output end and is reflected back to the input. The output end of an artificial transmission line appears to the input signal as what type of circuit?

1. Open
2. Short
3. Inductive reactance
4. Capacitive reactance

2-24. The discharge pulse from the artificial transmission line causes a voltage of what magnitude to appear across the primary of the pulse transformer?

1. Twice the original charge voltage
2. One-half the original charge voltage
3. The same as the original charge voltage
4. One-fourth the original charge voltage

2-25. What parameter, if any, of the output pulse from an artificial transmission line is affected by the inductance and capacitance of each section of the line?

1. Width
2. Amplitude
3. Frequency
4. None of the above

2-26. A pulse-forming network exhibits electrical behavior similar to which of the following devices?

1. A resistance-capacitance network
2. An artificial transmission line
3. A capacitor
4. An inductor

2-27. The requirements of a modulator switch are to (1) reach full conduction quickly, (2) consume low power, (3) start and stop conduction suddenly, and (4) conduct high currents and handle high voltages. Which of the following tubes meets these requirements?

1. A tetrode
2. A thyratron
3. A magnetron
4. A beam-powered amplifier

2-28. What modulator circuit characteristic determines the charging rate of the storage element?

1. Resistance
2. Capacitance
3. Charging impedance
4. Pulse-repetition frequency

2-29. Which of the following types of instability are common to magnetron oscillators?

1. Mode skipping
2. Mode shifting
3. Both 1 and 2 above
4. Magnetic fluctuation

2-30. Which of the following magnetron characteristics can be caused by low magnetic field strength?

1. Low power output
2. Excessive plate current
3. Incorrect operating frequency
4. All of the above

2-31. Which of the following maximum tuning ranges is typical for a tunable magnetron?

1. ±10 percent around the center frequency
2. ±5 percent around the center frequency
3. 1 to 5 percent above the center frequency
4. 1 to 10 percent below the center frequency

2-32. Compared to the keyed-oscillator transmitter, power-amplifier transmitters are used more often with mti radar systems because power-amplifier transmitters provide

1. better stability
2. lower output power
3. higher output power
4. greater frequency range

2-33. Which of the following tubes should be used as the power amplifier in a radar transmitter?

1. The magnetron
2. The thyratron
3. The reflex klystron
4. The multicavity klystron

2-34. Which of the following components determines the pulse width of a power-amplifier transmitter?

1. The modulator
2. The mixer-amplifier
3. The local oscillator
4. The power-amplifier tube

2-35. The intermediate stages of a power-amplifier transmitter have operating power only during which of the following times?

1. When the coherent rf pulse is applied
2. When the local oscillator signal is applied
3. During the time the modulator pulse is applied
4. Immediately after the coherent rf pulse is removed

2-36. Using a frequency synthesizer instead of a heterodyning mixer as the frequency generating source for a power-amplifier transmitter is an advantage because the frequency synthesizer

1. is more stable
2. is simpler to construct
3. produces a single frequency
4. produces discrete frequencies over a wide band

2-37. Unwanted oscillations in an rf amplifier transmitter are prevented because of which of the following pulse relationships?

1. The rf pulse is wider than the modulator pulse
2. The rf pulse is narrower than the modulator pulse
3. The rf pulse frequency is equal to the local oscillator frequency
4. The rf pulse frequency is less than the local oscillator frequency

2-38. A power-amplifier transmitter that transmits a broad band of frequencies typically uses which of the following tubes as the final stage?

1. A crossed-field amplifier
2. A multicavity klystron
3. A magnetron
4. A twt

2-39. What is the primary function of the radar duplexing system?

1. To prevent the formation of standing waves in the waveguide system
2. To permit the use of one antenna for transmission and reception
3. To increase the effective range of the radar
4. To increase antenna directivity

2-40. A defective duplexer in a radar will most likely cause damage to which of the following components?

1. The receiver
2. The waveguide
3. The magnetron
4. The local oscillator

2-41. Why is it desirable that the duplexer quickly connect the receiver to the antenna after the transmitted pulse?

1. So that line-match will remain balanced
2. So that the transmitter power dissipated will be minimum
3. So that echoes from nearby targets will be received
4. So that echoes from nearby targets will not prolong ionization

2-42. The action of tr-atr circuits depends upon the impedance characteristics of which of the following lengths of transmission line segments?

1. 1 wave length
2. 1/2 wave length
3. 1/4 wave length
4. 1/8 wave length

2-43. Which of the following requirements is/are essential for proper tr spark gap operation?

1. High impedance during arc time
2. Low impedance during arc time
3. High impedance prior to arc time
4. Both 2 and 3 above

2-44. What is the purpose of introducing water vapor into a tr tube?

1. It increases recovery time
2. It prevents early ionization
3. It decreases deionization time
4. It increases the gap breakdown potential

2-45. Keep-alive voltage is applied to the tr tube for which of the following reasons?

1. To maintain ionization within the tube after the breakdown voltage is removed
2. To ensure that the tube will rapidly return to the deionized state
3. To maintain a glow discharge within the tube so that firing will occur rapidly
4. To prevent breakdown within the tube prior to pulse transmission so that firing will not be premature

2-46. Atr tubes generally have a longer duty life than tr tubes because atr tubes do NOT use

1. radioactive materials and chemically active gas
2. chemically active gas and keep-alive voltages
3. keep-alive voltages and radioactive materials
4. keep-alive voltages and pure inert gas

2-47. In a series-connected duplexer, what spark gap, if any, fires during reception?

1. The tr only
2. The atr only
3. Both the tr and atr
4. None of the above

2-48. Indirectly coupled waveguide duplexers are normally connected to the main waveguide by which of the following devices?

1. A two-wire line
2. A coaxial cable
3. A short quarter-wave stub
4. A short section of waveguide

2-49. The direct-coupled waveguide duplexer is connected to the waveguide at the location of

1. minimum current flow
2. minimum magnetic field intensity
3. maximum magnetic field intensity
4. maximum electric field intensity

2-50. In a hybrid-ring duplexer, the fields at the entrance of an arm must have what phase relationship to propagate energy down the arm?

1. 0 degrees
2. 90 degrees
3. 180 degrees
4. 270 degrees

2-51. Which of the following are requirements of a microwave receiver?

1. Amplify extremely high-frequency pulses
2. Detect and amplify pulses in the microvolt range
3. Detect pulses with a duration of a few microseconds
4. All of the above

2-52. The maximum range at which a radar receiver can detect an object is limited by which of the following factors?

1. Noise
2. Signal distortion
3. Receiver bandwidth
4. Transmitter frequency

2-53. An effective radar receiver should have a gain factor that is in which of the following ranges?

1. 10^1 to 10^2
2. 10^3 to 10^4
3. 10^6 to 10^8
4. 10^9 to 10^{10}

2-54. An overdriven amplifier stage in a receiver may cause which of the following conditions?

1. Blocking
2. Inaccurate ranges
3. Increased sensitivity
4. Large signal distortion

2-55. The intermediate frequency is produced in what stage of a microwave receiver?

1. The mixer
2. The IF amplifier
3. The second detector
4. The local oscillator

2-56. What section of a radar receiver usually determines the overall bandwidth?

1. The mixer
2. The IF amplifier
3. The video amplifier
4. The local oscillator

2-57. What component in a receiver afc circuit produces an output voltage proportional in amplitude and polarity to any change in the intermediate frequency?

1. The mixer
2. The IF amplifier
3. The discriminator
4. The local oscillator

2-58. An efficient local oscillator must have which of the following characteristics?

1. Tunable frequency
2. Stable output frequency
3. Operation in the 4,000 megahertz range
4. All of the above

2-59. Which of the following devices would be used as a local oscillator in a radar receiver?

1. A magnetron
2. A crystal diode
3. A reflex klystron
4. A parametric amplifier

2-60. Which of the following advantages is gained by using a crystal mixer in a microwave receiver?

1. Less noise
2. Reduced saturation
3. Increased overall gain
4. Improved oscillator stability

2-61. The resonant circuit at the output of an unbalanced crystal mixer serves which of the following purposes?

1. It amplifies the IF signal
2. It produces the IF signal
3. It eliminates unwanted signals
4. It amplifies the local oscillator signal

2-62. The balanced transformer connected to the crystals of a balanced mixer has a secondary that is tuned to what frequency?

1. The desired IF
2. The local oscillator frequency
3. The afc discriminator frequency
4. The transmitter carrier frequency

2-63. The IF amplifier stage of a radar receiver determines which of the following receiver characteristics?

1. The gain
2. The effective bandwidth
3. The signal-to-noise ratio
4. All of the above

2-64. The detector in a radar receiver converts the IF pulses into what form?

1. Video pulses
2. Square waves
3. Dc voltage levels
4. Continuous-wave signals

2-65. Agc automatically adjusts the gain of the receiver by controlling which of the following quantities?

1. Detector bias
2. IF amplifier bias
3. Mixer output signal level
4. Video amplifier output signal level

2-66. A radar receiver uses iagc for which of the following purposes?

1. To reduce the amplitude of echoes from distant targets
2. To prevent full amplification of strong signals
3. To permit full amplification of weak signals
4. Both 2 and 3 above

2-67. In a radar receiver, which of the following purposes is served by using stc?

1. Prevents full amplification of echoes from nearby targets
2. Permits full amplification of echoes from distant targets
3. Both 1 and 2 above
4. Prevents target echoes within a selected range from being received

2-68. In the input of the first video amplifier of a radar receiver, the differentiator circuit performs which of the following functions?

1. Ftc
2. Gagc
3. Afc
4. Iagc

2-69. The primary function of the mti system is to display which of the following types of targets?

1. Moving targets only
2. Motionless targets only
3. Moving and motionless targets during each sweep
4. Moving and motionless targets during alternate sweeps

2-70. Delaying the received signals in the mti system permits them to be combined with the next set of received signals so that only desired signals are displayed. The signals displayed are formed by which of the following methods?

1. Division
2. Addition
3. Subtraction
4. Multiplication

2-71. In the mti system, what is the purpose of the coho lock pulse?

1. To synchronize the coho and transmitted frequency phase relationship
2. To control the transmitter pulse-repetition frequency
3. To synchronize the phase of the timing circuits with the phase detector
4. To control the polarity of the coherent video

2-72. The amplitude of coherent video is determined by the phase difference between which of the following signals?

1. Coho reference and transmitted pulse
2. Coho reference and coho lock pulse
3. Coho reference and IF echo
4. IF echo and received echo

2-73. The purpose of the mti system timing circuits is to

1. synchronize the coho and transmitted frequency phase relationship
2. control the transmitter pulse-repetition frequency
3. synchronize the phase of the video balancer with the phase detector
4. select the polarity of the coherent video

2-74. The lin-log amplifier provides (a) a linear output voltage for what amplitude of input signal and (b) a logarithmic output voltage for what amplitude of input signal?

1. (a) Low (b) low
2. (a) Low (b) high
3. (a) High (b) high
4. (a) High (b) low

2-75. What channel in a monopulse receiver is used as the reference channel?

1. IF
2. Range
3. Bearing
4. Elevation

ASSIGNMENT 3

Textbook assignment: Chapter 3, “Radar Indicators and Antennas,” pages 3-1 through 3-22. Chapter 4, “Radar Maintenance,” pages 4-1 through 4-23.

3-1. Which of the following geometrical quantities is/are used as coordinates for radar displays?

1. Range
2. Bearing
3. Elevation
4. All of the above

3-2. Of the following target parameters, which would most likely be displayed on all radar sets?

1. Altitude
2. Slant range
3. Ground range
4. All of the above

3-3. Which of the following quantities determines total distance represented on a crt display.

1. Crt size
2. Sweep speed
3. Sweep length
4. Echo spacing

3-4. To correctly represent the location of a target, a radar repeater must receive which of the following quantities?

1. Video
2. Timing pulses
3. Antenna information
4. All of the above

3-5. The A-scope crt normally uses what type of sweep deflection?

1. Mechanical
2. Electrostatic
3. Electromagnetic
4. Electromechanical

3-6. The rhi scope provides the operator with information concerning the target's

1. range only
2. altitude only
3. range and altitude
4. range and bearing

3-7. Own ship position is at the center of the scope in which of the following radar displays?

1. A-scope
2. B-scope
3. Rhi scope
4. Ppi scope

3-8. Synchronization of events in ppi circuitry is of special importance for which of the following reasons?

1. To ensure that bearing readings are accurate
2. To ensure that range readings are accurate
3. To ensure the deflection coils do not overheat
4. To ensure the power supply is activated at the instant the transmitter fires

3-9. Pulses used to synchronize the ppi with the transmitter are developed in which of the following circuits?

1. Gate
2. Sweep control
3. Sweep generator
4. Intensity gate generator

3-10. Which of the following circuits produces currents that deflect the electron beam across the crt?

1. Gate
2. Sweep control
3. Sweep generator
4. Intensity gate generator

3-11. Electromagnetic deflection is preferred over electrostatic deflection in ppi scopes because it provides which of the following advantages?

1. Better control of the beam
2. Better beam position accuracy
3. Better deflection sensitivity
4. All of the above

3-12. Focusing is accomplished in an electromagnetic crt by varying the

1. potential on the deflection plates
2. potential between the first anode and the cathode
3. current through the coil around the neck of the tube
4. do bias current in the deflection coils

3-13. Because the electromagnetic crt uses magnetic deflection, the sweep circuits must provide the deflection coils with which of the following signals?

1. Linear trace current
2. Trapezoidal voltage
3. Sinusoidal voltage
4. Direct current

3-14. In a ppi scope that uses electromagnetic deflection, the amplitudes and polarities of the sawtooth currents are determined by which of the following inputs?

1. Target position and speed
2. Antenna rotation speed
3. Antenna position
4. Both 2 and 3 above

3-15. On the screen of a ppi scope, range markers appear as

1. vertical pulses
2. radial grid lines
3. concentric circles
4. horizontal grid lines

3-16. The range sweep in a range-gate generator is started at the same time the transmitter fires. A pulse from which of the following circuits causes this timing?

1. Receiver
2. Indicator
3. Transmitter
4. Synchronizer

3-17. When used with a ppi presentation, a range gate must have which of the following characteristics?

1. Movable in range
2. Movable in bearing
3. Both 1 and 2 above
4. Fixed in range and bearing

3-18. Range-markers are produced on the basis of which of the following timing constants?

1. Radar mile
2. Transmitter prf
3. Receiver bandwidth
4. Transmitter pulse width

3-19. To read range directly, the range step is placed in what position relative to an echo pulse?

1. The range step is centered on the echo pulse
2. The range step coincides with the leading edge of the echo pulse
3. The range step coincides with the trailing edge of the echo pulse
4. The range step covers the entire echo pulse

3-20. Compared to omnidirectional antennas, directional antennas provide which of the following advantages?

1. Power gain and selectivity
2. Power gain and directivity
3. Sensitivity and selectivity
4. Sensitivity and directivity

3-21. If the vertical beam width of a radar antenna is decreased, what will be the effect on (a) power gain and (b) vertical directivity?

1. (a) Decrease (b) decrease
2. (a) Decrease (b) increase
3. (a) Increase (b) increase
4. (a) Increase (b) decrease

3-22. An array of twelve dipoles are set in the same position as a reference dipole and are fed with the same line. The power gain will be

1. less than unity
2. one-twelfth the reference
3. twelve times the reference
4. dependent on the directivity of the array

3-23. If the slant range and altitude of a target are known, which of the following coordinates can be computed using trigonometric functions?

1. Elevation angle
2. True-bearing angle
3. Relative-bearing angle
4. All of the above

3-24. To convert diverging waves into parallel waves, where must the radiating element be placed in relation to a parabolic reflector?

1. At the focal point of the reflector
2. 1/4 wavelength from the reflector
3. 1/2 wavelength from the reflector
4. At the focal point of the hemispherical shield

3-25. What is the function of the hemispherical shield of the parabolic reflector?

1. To polarize all reflected waves in the vertical plane
2. To polarize all reflected waves in the horizontal plane
3. To convert the spherical waves radiated by the dipole into vertical lines of rf energy
4. To reflect rf energy radiated forward of the dipole back to the parabolic reflector

3-26. Which of the following types of parabolic reflectors has a focal line rather than a single focal point?

1. Truncated
2. Rotational
3. Orange-peel
4. Cylindrical

3-27. A broadside array causes maximum energy to be radiated perpendicular to the plane of the dipole for which of the following reasons?

1. Because dipoles are excited in phase with each other
2. Because dipoles are parallel to each other
3. Because dipoles are 1/2 wavelength apart
4. Because dipoles are 1/8 wavelength away from the reflector

3-28. The directivity of a horn radiator is dependent on which of the following physical dimensions of the horn?

1. The shape
2. The length
3. The mouth size
4. All of the above

3-29. Feedhorn shadows can be eliminated by taking which of the following actions?

1. By making the horn smaller
2. By making the reflector smaller
3. By offsetting the horn from the center of the reflector
4. By putting the horn behind the reflector

3-30. Airborne radars have unique physical design requirements. Which of the following functions is performed by the radome?

1. It serves as the antenna
2. It provides aerodynamic shape
3. It protects the antenna from low air pressure
4. All of the above

3-31. If a fixed-frequency radar transmitter is found to be off its normal operating band, which of the following corrective actions should be taken?

1. Retune the transmitter
2. Change the assigned frequency
3. Replace the defective part causing the error
4. Check the frequency again, an error has been made

3-32. If a transmitter carrier wave is modulated by a square wave, what maximum number of different frequencies will be produced?

1. An infinite number
2. 2
3. 3
4. 4

3-33. Which of the following statements describes a radar transmitter frequency spectrum?

1. The distribution of energy over a band of frequencies
2. The distribution of energy over time
3. The prf multiplied by the duty cycle
4. The pulse width versus peak power

3-34. What total number of modulating components are present in the output spectrum of a pulse radar transmitter?

1. 1
2. 2
3. 3
4. 4

3-35. An ideal radar frequency spectrum would be best described in which of the following ways?

1. It is symmetrical
2. It has a wide lobe
3. It has narrow side lobes
4. It has no minimum points

3-36. In a good spectrum curve the distance between the two minima is proportional to which of the following transmitter parameters?

1. Prt
2. Prf
3. Peak power
4. Pulse width

3-37. The echo box is a good instrument for measuring overall radar system performance because it indicates the combined effectiveness of which of the following components?

1. Antenna and duplexer
2. Transmitter and antenna
3. Transmitter and receiver
4. Receiver and synchronizer

3-38. Oscillations in an echo box are known as ringing. Which of the following conditions cause this ringing?

1. A weak transmitter
2. A saturated receiver
3. A normally operating receiver
4. A normally operating transmitter

3-39. What constitutes the single most useful measurement you can make with the echo box?

1. Ring time
2. Duty cycle
3. Power distribution
4. Frequency distribution

3-40. Desirable transmitter output power characteristics include what relative levels of (a) peak power and (b) average power?

1. (a) Low (b) low
2. (a) Low (b) high
3. (a) High (b) high
4. (a) High (b) low

3-41. Most transmitter power readings are referenced to which of the following quantities?

1. 1 microwatt
2. 1 milliwatt
3. 1 watt
4. 1 kilowatt

3-42. Which of the following factors determines the overall performance of a radar receiver?

1. Bandwidth
2. Sensitivity
3. Recovery time of the tr
4. All of the above

3-43. Of the following receiver special circuits, which one is used during sensitivity tests?

1. Afc
2. Agc
3. Ftc
4. Iagc

3-44. The ability of a receiver to detect weak signals can be determined by which of the following measurements?

1. Noise figure
2. Minimum discernable signal
3. Both 1 and 2 above
4. Bandwidth

3-45. When several mds checks are to be taken over a period of time, the length of the test pulse used in the tests should

1. be the same on each check
2. be different on each check
3. vary with the transmitter pulse length on each check
4. vary with the noise figure on each check

3-46. When expressing the sensitivity of a radar receiver, which of the following quantities is used?

1. The signal generator reading
2. The combined attenuation value of the connecting cable and directional coupler
3. The sum of both 1 and 2 above
4. The attenuation value of the signal generator

3-47. Tr recovery time places limits on which of the following quantities?

1. Minimum range
2. Maximum range
3. Receiver bandwidth
4. Receiver sensitivity

3-48. Which of the following methods is/are used to determine the effectiveness of a tr tube?

1. Measure the keep-alive current
2. Measure the keep-alive voltage
3. Graph the correlation between recovery time and leakage power
4. All of the above

3-49. The presence of standing waves on a transmission line indicates which of the following conditions?

1. Excessive output power
2. An impedance mismatch
3. Excessive pulse width
4. Excessive prf

3-50. Of the following conditions, which would be a likely indication of a high vswr?

1. Insufficient reflection
2. Cold spots in the transmission line
3. Arc-over at the maximum points
4. All of the above

3-51. Most primary shipboard ac distribution systems provide which of the following types of electrical power?

1. 60 Hz, 3 phase, ungrounded
2. 60 Hz, 1 phase, ungrounded
3. 400 Hz, 1 phase, ungrounded
4. 400 Hz, 3 phase, ungrounded

3-52. If your equipment is missing a certain voltage input, which of the following actions should you take first?

1. Call an electrician
2. Energize the emergency generator
3. Check the input to the switchboard
4. Check the power panel that feeds your equipment

3-53. What is the normal source of dry air for a shipboard radar system?

1. Compressed-air bottles
2. Central dry-air system
3. Dehumidifying ovens
4. Local dehydrators

3-54. The air control panel is designed to regulate

1. flow
2. purity
3. pressure
4. dew point

3-55. Which of the following units may be available as an emergency back-up to the central dry-air system?

1. Nitrogen tank
2. Local dehydrator
3. Local compressor-dehydrator
4. Each of the above

3-56. Which of the following methods is used to cool radar system components?

1. Air blowers
2. Liquid-cooling loops
3. Air-conditioning systems
4. Each of the above

3-57. A radar cooling system has a low-flow alarm in the sea-water loop. What is the primary purpose of this alarm?

1. It allows correction before damage occurs
2. It increases the flow in the distilled-water loop
3. It removes power from the system
4. It increases the sea-water pressure

3-58. Which of the following characteristics of cooling water for electronic equipment must be carefully controlled?

1. Purity
2. Pressure
3. Quantity
4. All of the above

3-59. For which of the following reasons do personnel sometimes develop a false sense of security concerning exposure to radiation?

1. Rf radiation does not always produce pain
2. Rf radiation is visible only at night
3. Only search radars are hazardous
4. Rf hazards occur only at night

3-60. Injury from X-rays would most likely result from which of the following actions?

1. Standing near unshielded high-voltage components
2. Working alone on low-voltage power supplies
3. Bypassing interlocks on shielded equipment
4. Working too close to a crt with a potential of 1,500 volts

NONRESIDENT
TRAINING
COURSE

Navy Electricity and Electronics Training Series

Module 19—The Technician's Handbook

NAVEDTRA 14191

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 19 of a series.

History of the course:

Sep 1998: *Original edition released. Prepared by TMCM Jack L. Formyduval.*
Jan 2004: *Administrative update released. Reviewed and revised by ETC(SW) Jack Weatherford. Minor revision to technical content.*

TABLE OF CONTENTS

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

CHAPTER 1

THE TECHNICIAN'S HANDBOOK

Here, in one compact module, you will be able to find essential information and reference material. Whether you want to know safety precautions, first aid, or any number of helpful pieces of information, you will find it indexed.

We have included electrical and electronic formulas, data tables, and general maintenance hints. In most cases you will find references to other more detailed sources included.

We solicit your suggestions, maintenance hints, and constructive criticism. You will receive credit in future editions of this handbook if your input is used.

MISHAP PREVENTION AND AFTERCARE

Most of us working with electricity take risks. Usually we get our jobs done without any harmful results. Mishaps or injuries usually result from not understanding a risk or danger.

The first part of this handbook is designed to help you eliminate or minimize mishaps. It also provides you with a good review of what to do in case of a mishap.

SAFETY OBSERVATIONS FOR THE ELECTRICAL AND ELECTRONICS TECHNICIAN

Working safely is the most important thing you can do. Because of their importance, several precautions are included as the first subject in this handbook. Of course there are more precautions, but these are some you should think about. The keyword here is *think*. Think safety.

- Never work alone.
- Never receive an intentional shock.
- Only work on, operate, or adjust equipment if you are authorized.
- Don't work on energized equipment unless absolutely necessary.
- Keep loose tools, metal parts, and liquids from above electrical equipment. Never use steel wool or emery cloth on electric and electronic circuits.
- Never attempt to repair energized circuits except in an emergency.
- Never measure voltage in excess of 300 volts while holding the meter wire or probe.
- Use only one hand when operating circuit breakers or switches.
- Use proper tag-out procedures for regular and preventive maintenance.

- Be cautious when working in voids or unvented spaces.
- Beware the dangers of working aloft. Never attempt to stop a rotating antenna manually.
- Keep protective closures, fuse panels, and circuit breaker boxes closed unless you are actually working on them.
- Never bypass an interlock unless you are authorized to do so by the commanding officer, and then properly tag the bypass.
- Use extreme caution when handling cathode-ray tubes. They implode violently if broken. The anode contact may have a residual electrical charge. Make sure you discharge the anode before handling.

You can find additional and more detailed information on safety in the Electronics Installation and Maintenance Book (EIMB), *General*, NAVSEA SE000-00-EIM-100, paragraph 3-4. Another excellent reference source is the Naval Electronics Systems Command's *Electronic Safety Handbook*, E0410-AA-HBK-010/00K ELEXSAFE.

FIRST AID

First aid is the emergency care you give to sick or injured persons. It consists only of providing temporary assistance or treatment until medical help is available. In addition to knowing what to do for a victim, you should also know what not to do.

This section should be used to reinforce the knowledge you already have about first aid. First aid is included in detail for the purposes of review, study, and ready reference.

First aid study classes are usually available through your medical department or the American Red Cross.

Your knowledge of first aid measures and their proper application may mean the difference between life and death, between rapid recovery and long hospitalization, or between temporary disability and permanent injury.

The objectives of first aid are to save life and prevent further injury. First aid is not a substitute, however, for proper medical treatment. Keep in mind the objectives of first aid. Everyone in the Navy must know when and how to apply first aid measures and must be prepared to give assistance to persons injured in battle, collision, fire, or accidents.

In administering first aid, you have three primary tasks:

- Maintain breathing
- Stop bleeding
- Prevent or reduce shock

The first step, of course, is to determine the extent of the victim's injuries. When you treat a victim, first consideration usually must be given to the most serious injury. In general, the order of treatment is to restore breathing, stop bleeding, and treat for shock.

Work quickly, but do not rush around frantically. Do not waste time looking for ready-made materials; do the best you can with whatever is at hand. Send for medical help as soon as possible.

Although each case involving injury or sickness presents its own special problems, the following general rules apply to practically all situations. Become familiar with these basic rules before you go on to first aid treatment for specific types of injuries.

1. Keep the victim lying down, head level with the body, until you have found out what kind of injury has occurred and how serious it is. If the victim shows one of the following difficulties, however, follow the rule given for that specific problem:
   a. Vomiting or bleeding about the mouth and semi-consciousness. If the victim is in danger of sucking in blood, vomited matter, or water, place the victim on his/her side or back with the head turned to one side and lower than the feet.
   b. Shortness of breath. If the victim has a chest injury or breathing difficulties, place the victim in a sitting or semi-sitting position.
   c. Shock. If the victim is in shock, place the person on his or her back with the head slightly lower than the feet.
2. Move the victim no more than is absolutely necessary. To determine the extent of the victim's injuries, carefully rip or cut the clothing along the seams. If done improperly, the removal of the victim's clothing could cause great harm, especially if fracture injuries are involved. When the clothing is removed, ensure that the victim does not become chilled. Shoes may also be cut off to avoid causing pain or increasing an injury.
3. The victim need not see the actual injury(ies). You can supply reassurance and make the victim more comfortable by ensuring him or her that the injuries incurred are understood and medical attention will be given as soon as possible.
4. Do not touch open wounds or burns with fingers or other objects, except when sterile compresses or bandages are not available and it is absolutely necessary to stop severe bleeding.
5. Do not try to give an unconscious person any solid or liquid substance by mouth. The person may vomit and get some of the material into the lungs when he or she breathes, causing choking and possibly death.
6. If a bone is broken, or you suspect that one is broken, do not move the victim until you have immobilized the injured part. This may prove life saving in cases of severe bone fractures or spinal cord injuries, because the jagged bone may sever nerves and blood vessels, damage tissues, and increase shock. Of course, threat of fire, necessity to abandon ship, or other similar situations may require that the victim be moved. But the principle that further damage could be done by moving the victim should always be kept in mind and considered against other factors.
7. When transporting an injured person, always see that the litter is carried feet forward no matter what the injuries are. This will enable the rear bearer to observe the victim for any respiratory obstruction or stoppage of breathing.
8. Keep the injured person comfortably warm — warm enough to maintain normal body temperature.

Very serious and mutilating injuries may require heroic first aid measures on your part. However, the greater the number of injuries, the more judgment and self-control you must exhibit to prevent yourself and well-intentioned bystanders from trying to do too much.

Electric Shock

Electric shock may cause anything from mild surprise to death. The effects of the shock are usually unknown. It is often hard to determine how an electrical shock victim has been affected.

SYMPTOMS OF ELECTRIC SHOCK.—When you find someone who has received a severe electric shock, the person's skin is usually very white or pale blue. In the case of victims with dark skin, it may be necessary to rely primarily on the color of the mucous membranes on the inside of the mouth or under the eye lid or under the nail bed. A person in or going into electric shock has a bluish color to these membranes instead of a healthy pink. The victim's pulse is very weak or absent. The person is unconscious, and usually the skin is burned. A stiffness of the body may happen in a few minutes. This is caused by the muscles reacting to shock. You should not consider this condition as rigor mortis. You should make sure the victim is no longer touching the live circuit and then start artificial respiration. People have recovered after body stiffness has set in.

RESCUE OF VICTIMS.—The rescue of a shock victim depends on your immediate administration of first aid.

WARNING

Do not attempt to administer first aid or come in physical contact with an electric shock victim before the power is shut off or, if the power cannot be shut off immediately, before the victim has been removed from the live conductor.

When attempting to administer first aid to an electric shock victim, proceed as follows:

Shut off the power.

If the power cannot be deactivated, remove the victim immediately, observing the following precautions:

—Protect yourself with dry insulating material. Use a dry board, a belt, dry clothing, or other available nonconductive material to free the victim (by pulling, pushing, or rolling) from the power-carrying object. DO NOT TOUCH the victim.

Immediately after you remove the victim from contact with the live circuit, administer artificial respiration/ventilation or cardiopulmonary resuscitation as necessary.

ANYONE WHO RECEIVES A SIGNIFICANT SHOCK SHOULD BE TAKEN TO SICK BAY OR A MEDICAL FACILITY AND OBSERVED FOR SEVERAL HOURS.

Artificial Ventilation

A person who has stopped breathing is not necessarily dead, but is in immediate critical danger. Life depends on oxygen that is breathed into the lungs and then carried by the blood to every body cell. Since body cells cannot store oxygen, and since the blood can hold only a limited amount (and only for a short time), death will surely result from continued lack of breathing.

The heart may continue to beat and the blood may still be circulated to the body cells for some time after breathing has stopped. Since the blood will, for a short time, contain a small supply of oxygen, the body cells will not die immediately. Thus, for a few minutes, there is some chance that the person's life may be saved. A person who has stopped breathing but who is still alive is said to be in a state of respiratory failure. The first aid treatment for respiratory failure is called artificial ventilation.

The purpose of artificial ventilation is to provide a method of air exchange until natural breathing is established. Artificial ventilation should be given only when natural breathing has stopped; it must NOT be given to any person who is still breathing. Do not assume that breathing has stopped merely because a person is unconscious or because a person has been rescued from the water, from poisonous gas, or from contact with an electric wire. Remember, **DO NOT GIVE ARTIFICIAL VENTILATION TO A PERSON WHO IS BREATHING NATURALLY**. If the victim does not begin spontaneous breathing after you use the head or jaw tilt techniques (discussed later) to open the airway, artificial ventilation must be attempted immediately. If ventilation is inadequate, one of the "thrust" methods of clearing the airway must be performed, followed by another attempt of artificial ventilation.

MOUTH-TO-MOUTH.—To perform this method of ventilation, clear the victim's mouth of obstructions (false teeth and foreign matter), place one hand under the victim's neck and the heel of the other hand on the forehead, and, using the thumb and index finger, pinch the nostrils shut. Tilt the head back to open the airway. Take a deep breath, cover the victim's mouth with your own, and blow into the victim's mouth. Then remove your mouth from the victim's to allow the victim to exhale. Observe the victim's chest for movement. If the victim has not started to breathe normally, start artificial ventilation with four quick ventilation in succession, allowing the lungs to only partially inflate. If the victim still does not respond, then you must fully inflate the victim's lungs at the rate of **12 TO 15 VENTILATIONS PER MINUTE, or ONE BREATH EVERY 5 SECONDS**.

MOUTH-TO-NOSE.—This type ventilation is effective when the victim has extensive facial or dental injuries or is very young, as it permits an effective air seal.

To administer this method, place the heel of one hand on the victim's forehead and use the other hand to lift the jaw. After sealing the victim's lips, take a deep breath, place your lips over the victim's nose, and blow. Observe the chest for movement and place your ear next to the victim's nose to listen for, or feel, air exchange. Again, you must continue your efforts at the rate of 12 to 15 ventilation per minute, or one breath every 5 seconds, until the victim can breathe without assistance.

NOTE: Sometimes during artificial ventilation, air enters the stomach instead of the lungs. This condition is called GASTRIC DISTENTION. It can be relieved by moderate pressure exerted with a flat hand between the navel and rib cage. Before applying pressure, turn the victim's head to the side to prevent choking on stomach contents that are often brought up during the process.

BACK PRESSURE ARM LIFT.—This method is an alternate technique used when other methods are not possible. Place the victim on the stomach, face to one side, neck hypo-extended, with hands under the head. Quickly clear the mouth of any foreign matter. Kneel at the victim's head and place your hands on the victim's back so that the heels of the hands lie just below a line between the armpits, with thumbs touching and fingers extending downward and outward. Rock forward, keeping your arms straight, and exert pressure almost directly downward on the victim's back, forcing air out of the lungs. Then rock backward, releasing the pressure and grasping the arms just above the elbows. Continue to rock backward, pulling the arms upward and inward (toward the head) until resistance and tension in the victim's shoulders are noted. This expands the chest, causing active intake of air (inspiration). Rock forward and release the victim's arms. This causes passive exiting of air (expiration). Repeat the cycle of *press, release, lift,* and *release* 10 to 12 times a minute until the victim can breathe naturally.

Cardiac Arrest and Cardiopulmonary Resuscitation

Cardiac arrest is the complete stoppage of heart function. If the victim is to live, action must be taken immediately to restore heart function.

In this situation, the immediate administration of cardiopulmonary resuscitation (CPR) by a rescuer using correct procedures greatly increases the chances of a victim's survival. To be effective, CPR must be started within 4 minutes of the onset of cardiac arrest. CPR consists of external heart compression and artificial ventilation. This compression is performed on the outside of the chest, and the lungs are ventilated either by mouth-to-mouth or mouth-to-nose techniques. The victim should be lying on a firm surface.

CAUTION

A rescuer who has not been properly trained should not attempt CPR. Everyone who works around electricity should be trained. (To learn CPR, consult a hospital corpsman.) Improperly done, CPR can cause serious damage. Therefore, it is never practiced on a healthy individual for training purposes; a training aid is used instead.

ONE RESCUER TECHNIQUE.—If a cardiac arrest is not witnessed, the rescuer must not assume that an arrest has occurred solely because the victim is lying on the floor and appears to be unconscious. First, try to arouse the victim. You can try shaking the victim's shoulders gently to obtain a response. Next, quickly check vital signs; if there is no response, apply artificial ventilation. Establish an open airway and ventilate the victim four times. Check the carotid (neck) pulse as shown in figure 1-1. If no pulse is felt and there are no visible signs of breathing, start CPR immediately.

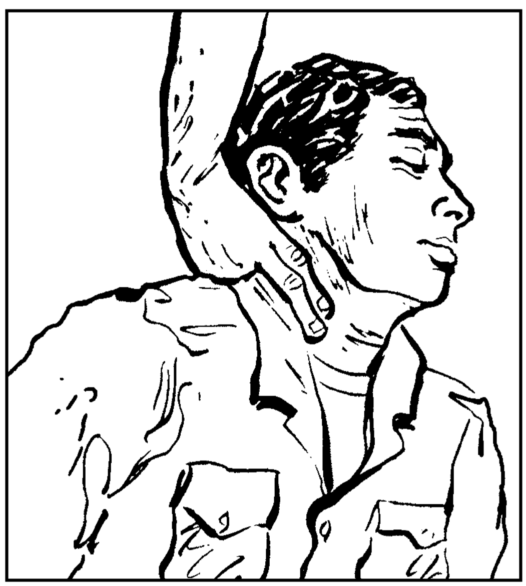


Figure 1-1.—Feeling for the carotid pulse

To start external cardiac compression, place the victim on the back, establish an open airway, and kneel at right angles to the victim's body. Then locate the victim's sternum (breastbone). You have a choice of two methods of doing this. One method is to bare the chest and locate the sternum by drawing an imaginary line from one nipple to the other to identify the proper area of the sternum, which is darkened in figure 1-2. The other method is to locate the lower tip of the sternum with the index and middle fingers, placing the heels of your hands above your fingers in the darkened area.

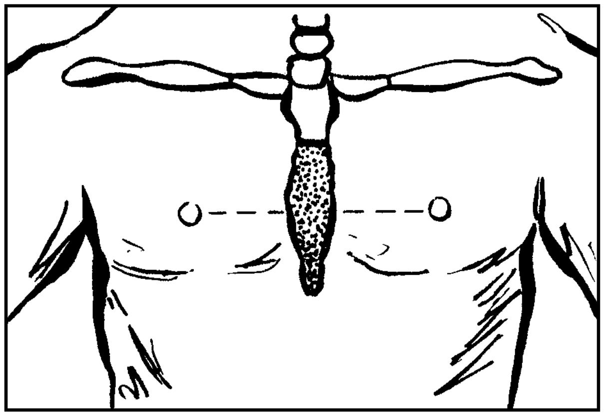


Figure 1-2.—Locating the sternum.

There is a small piece of cartilage at the lower end of the sternum (figure 1-2). A fracture of this area can damage the liver, causing hemorrhage (heavy bleeding) and death. When you place the heels of your hands on the victim's chest, make sure they are above the tip of the sternum.

Place the heel of one hand directly on the sternum and the heel of the other on top of the first. Figure 1-3, view A, shows this technique. Interlock your fingers and *keep them off the victim's chest!*

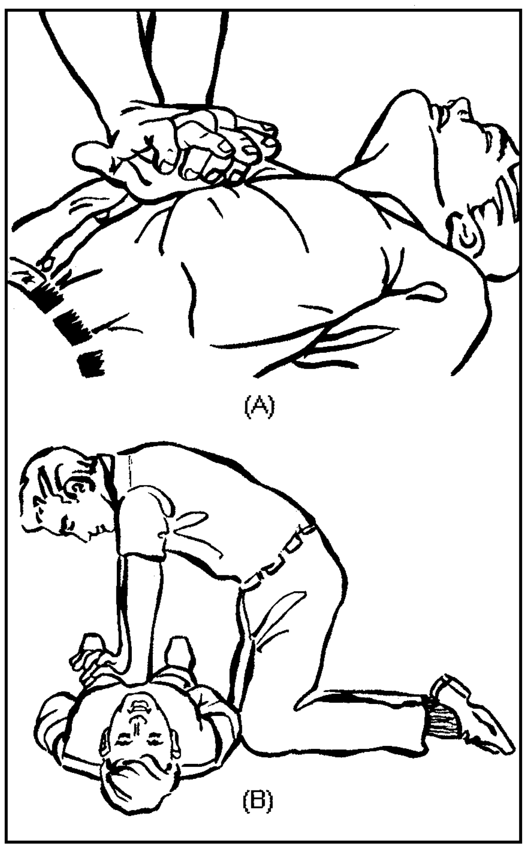


Figure 1-3.—Position for cardiac compression.

Lean or rock forward with elbows locked, and apply vertical pressure to depress the sternum (adult) 1 1/2 to 2 inches. View B depicts this. Then release the pressure, keeping hands in place. Administer 60 to 80 compressions per minute.

You will feel less fatigue if you use the proper technique, and a more effective compression will result.

Ineffective compression occurs when the elbows are not locked, the rescuer is not directly over the sternum, or the hands are improperly placed on the sternum.

When one rescuer performs CPR, as shown in figure 1-4, the ratio of compressions to ventilations is 15 to 2. After 15 compressions, you must give the victim 2 ventilations. This ratio must continue for four full cycles. Then check for pulse and breathing. If there are still no signs of recovery, continue CPR until the victim can breathe unassisted or you are relieved by medical personnel.

Figure 1-4.—One rescuer CPR technique.

Before reviewing the next technique, let's go over the steps to take in an unwitnessed cardiac arrest involving one rescuer.

1. Determine whether the victim is conscious.
2. Check the vital signs.
3. Ventilate four times (you may have to remove an airway obstruction at this time!).
4. Again check the vital signs; if none:
    a. Begin compression-ventilation rate of 15 to 2 for four complete cycles.
    b. Check pulse, breathing, and pupils. If no change
    c. Continue compression-ventilation rate of 15 to 2 until victim is responsive or you are relieved by medical personnel.

TWO RESCUER TECHNIQUE.—If two people trained in CPR are on the scene, one must perform compressions while the other performs artificial ventilation. The ratio for two-person CPR is 5 compressions to 1 ventilation. One rescuer is positioned at the chest area and the other beside the victim's head. The rescuers should be on opposite sides of the victim.

To avoid confusion, one rescuer must be designated the leader. The leader must make the preliminary checks of the victim's vital signs and perform the initial four ventilations. The second rescuer will perform the compressions.

When CPR is started, the compressions should be given in a constant, methodical rhythm. The rescuer giving the compressions counts them out loud. As the fifth compression is released, the other rescuer ventilates the victim. The compressions should be continued while ventilation is being given.

Hemorrhage

Blood is circulated throughout the body by means of three different kinds of blood vessels: arteries, veins, and capillaries. Arteries are large vessels that carry the blood away from the heart; veins are large vessels that carry the blood back to the heart; and capillaries form a connecting network of smaller vessels between the arteries and the veins.

Hemorrhage (escape of blood) occurs whenever there is a break in the wall of one or more blood vessels. In most small cuts, only capillaries are injured. Deeper wounds result in injury to veins or arteries. Bleeding which is severe enough to endanger life seldom occurs except when arteries or veins are cut.

The average adult body contains about 5 quarts (4.75 liters) of blood. One pint of blood can usually be lost without harmful effect—in fact, this is the amount usually given by blood donors. However, the loss of 2 pints (.95 liter) will usually cause shock; shock becomes greater and greater as the amount of blood loss increases (shock will be discussed later in this chapter). If half the blood in the body is lost, death almost always results.

Capillary blood is usually brick red in color. If capillaries are cut, the blood oozes out slowly. Blood from veins is dark red. If a vein is cut, the blood escapes in a steady, even flow. If an artery near the surface is cut, the blood will gush out in spurts that are synchronized with the heartbeats; but if the cut artery is deeply buried, the bleeding will appear to be a steady stream. Arterial blood is usually bright red in color.

In actual practice, you might find it difficult to decide whether bleeding is from a vein or an artery; but the distinction is not usually important. The important thing to know is that bleeding must be controlled as quickly as possible.

METHODS OF CONTROLLING BLEEDING.—The only way to stop serious bleeding is by the application of pressure. In practically all cases, bleeding can be stopped if **PRESSURE** is applied **DIRECTLY TO THE WOUND**. If direct pressure does not stop the bleeding, pressure should be applied at the appropriate pressure point. In those very rare cases where bleeding is so severe that it cannot be controlled by either of these methods, pressure can be applied by means of a tight, constricting band called a tourniquet.

PROCEDURES.—The actual procedures you should use to stop bleeding are detailed in the following paragraphs:

Direct Pressure.—In almost every case, bleeding can be stopped by the application of pressure directly on the wound. Figure 1-5 is an example of direct pressure. Place a dressing (sterile or clean, if possible) over the wound and firmly fasten it in position with a bandage. If bleeding does not stop, firmly secure another dressing over the first, or apply direct pressure with your hand to the dressing.

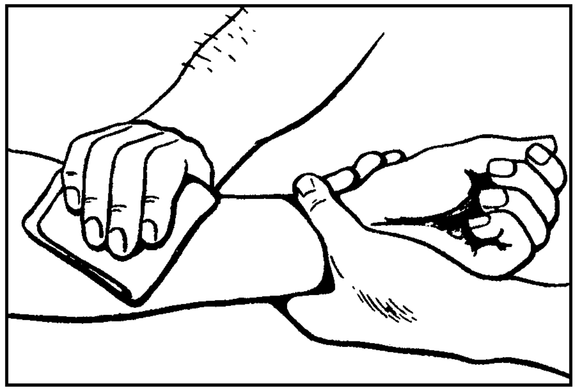


Figure 1-5.—Direct pressure.

In cases of severe hemorrhage, do not worry too much about the dangers of infection. The basic problem is to stop the flow of blood. If no material is available, simply apply pressure with your bare hand. Remember, **DIRECT PRESSURE** is the first method to use when you are trying to control hemorrhage.

Pressure Points.—Bleeding from a cut artery or vein may often be controlled by pressure applied to the appropriate pressure point. A pressure point is a place where the main artery to the injured part lies near the skin surface and over a bone. Pressure at such a point is applied with the fingers (digital pressure) or with the hand; no first aid materials are required. The object of the pressure is to compress the artery against the bone, thus shutting off the flow of blood from the heart to the wound.

There are 11 principal points on each side of the body where hand or finger pressure can be used to stop hemorrhage. These points are shown in figure 1-6.

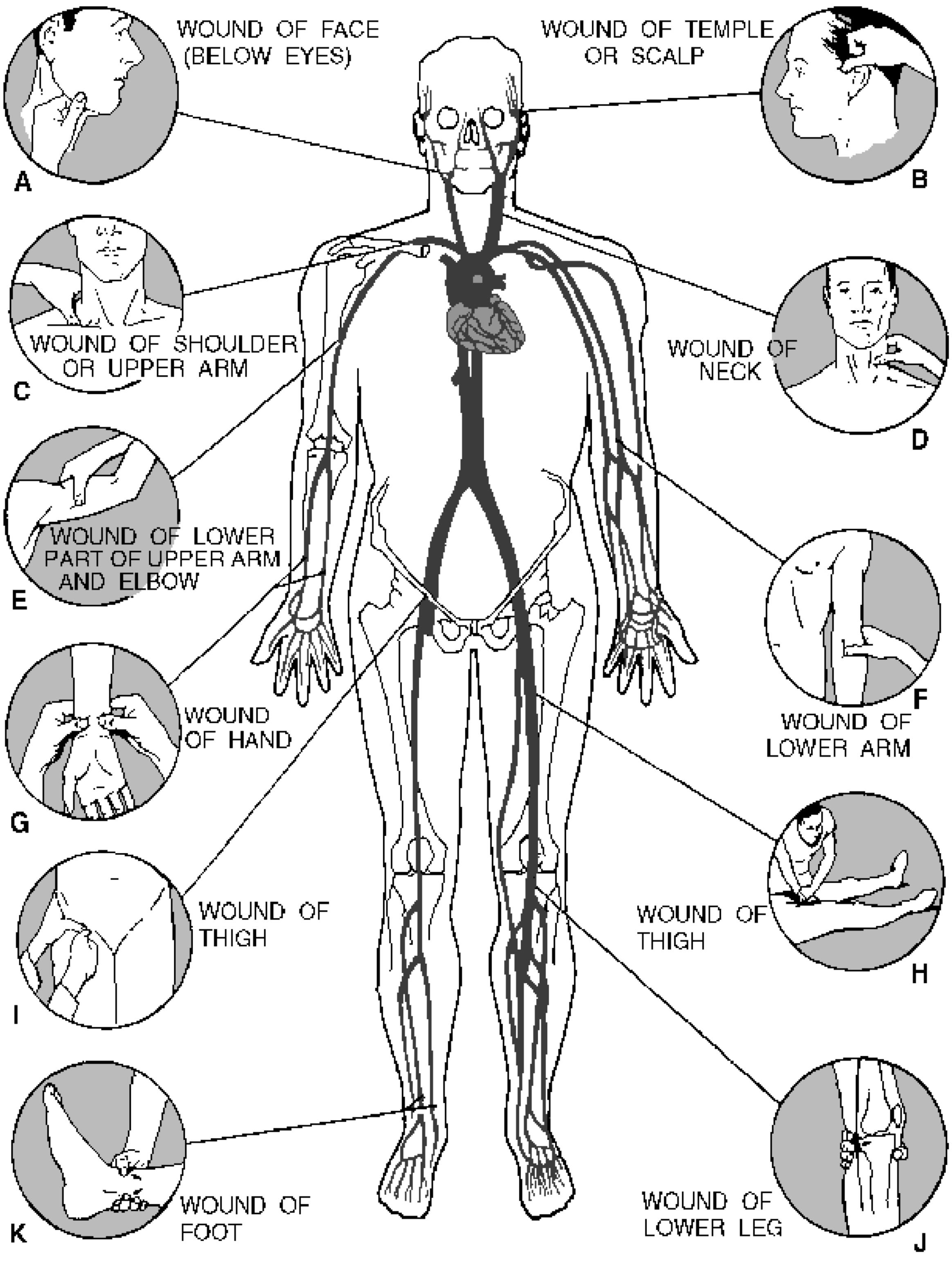


Figure 1-6.—Pressure points for control of bleeding.

You should memorize these pressure points so that you will know immediately which point to use for controlling hemorrhage from a particular part of the body. The correct pressure point is the one that is

(1) NEAREST THE WOUND and (2) BETWEEN THE WOUND AND THE MAIN PART OF THE BODY.

Applying finger pressure is very tiring, and it can seldom be maintained for more than 15 minutes. Pressure points are recommended for use while direct pressure is being applied to a serious wound. While pressure is being applied at the appropriate pressure point, an assistant can bandage the wound (or wounds). If available, a battle dressing should be used. Figure 1-7 shows the battle dressing and its use. After opening the dressing, be careful not to contaminate it. Place the compress portion over the wound, then bind it tightly in place with the attached straps. If bleeding continues to be severe even after direct pressure and pressure points have been used, you may have to apply a tourniquet.

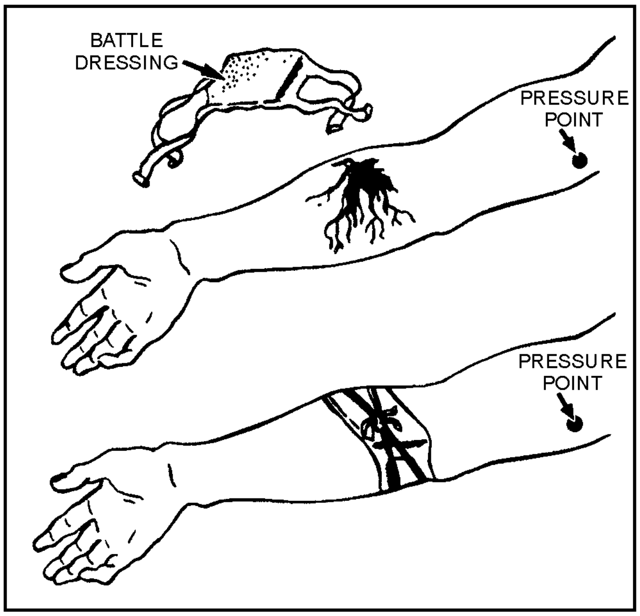


Figure 1-7.—Battle dressing.

Use of the Tourniquet.—A tourniquet is a constricting band that is used to cut off the supply of blood to an injured limb. It cannot be used to control bleeding from the head, neck, or trunk, since its use in these locations would result in greater injury or death. A tourniquet should be used *only* if the control of hemorrhage by other means proves to be impossible.

Basically, a tourniquet consists of a pad, a band, and a device for tightening the band so that the blood vessels will be compressed. There are several different kinds of ready-made tourniquets. A variety of materials can be used to improvise tourniquets. Any round, smooth pressure object may be used for the pad — a compress, a roller bandage, a stone, a rifle shell — and any long, flat material may be used as the band. However, the band must be flat: belts, stockings, flat strips of rubber, or neckerchiefs can be used; but rope, wire, string, or very narrow pieces of cloth should not be used because they will cut into the flesh. A short stick may be used to twist the band, thus tightening the tourniquet.

A tourniquet must always be applied *above* the wound — that is, toward the body — and it must be applied as close to the wound as practicable.

The best object to be used for the pad is either a bandage, a compress, or some similar pressure object. The pad goes under the band. It must be placed directly over the artery, or it will actually decrease the pressure on the artery and thus allow greater flow of blood. If a tourniquet placed over a pressure object does not stop the bleeding, the pressure object is probably in the wrong place. If this occurs, shift the object around until the tourniquet, when tightened, will control the bleeding. If no suitable pressure object is available, use the tourniquet without it.

To apply an emergency tourniquet made from something like a neckerchief, wrap the material once around the limb and tie an overhand knot; place a short stick on the overhand knot and tie a square knot over it. Then twist the stick rapidly to tighten the tourniquet. The stick may be tied in place with another strip of material. Figure 1-8 shows how to apply a tourniquet.

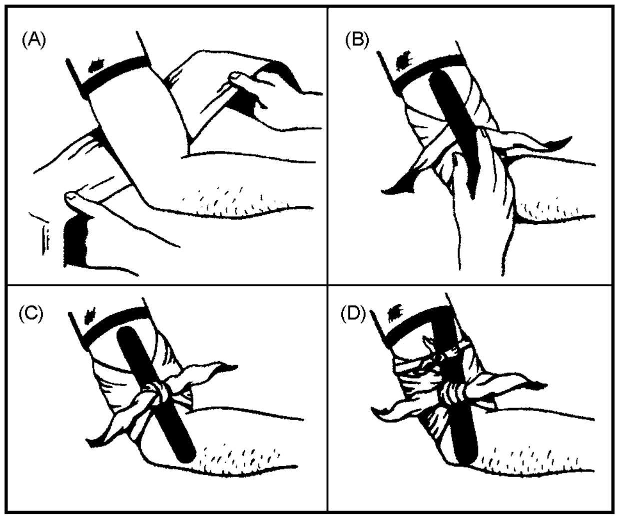


Figure 1-8.—Applying a tourniquet.

To be effective, a tourniquet must be tight enough to stop the blood flowing to the limb. If the pressure from the tourniquet is less than the arterial pressure, arterial bleeding will continue. Also, insufficient tourniquet pressure may actually increase the amount of bleeding from the veins. So be sure to draw the tourniquet tight enough to stop the bleeding. However, do not make it any tighter than necessary.

After you have brought the bleeding under control with the tourniquet, apply a sterile compress or dressing to the wound and fasten it in position with a bandage.

CAUTION

NEVER apply a tourniquet unless the hemorrhage is so severe that it cannot be controlled in any other way. By the time the tourniquet is required, the victim will have lost a considerable amount of blood; therefore, once a tourniquet has been applied, it should be released only by medical personnel.

Here are the points to remember when you use a tourniquet:

- Do not use a tourniquet unless you cannot control the bleeding by any other means.
- Do not use a tourniquet for bleeding from the head, face, neck, or trunk. Use it only on the limbs.
- Always apply a tourniquet *above the wound* and as close to the wound as possible.

- Be sure you draw the tourniquet tight enough to stop the bleeding, but do not make it any tighter than necessary.

- Do not loosen a tourniquet after it has been applied except in extreme emergency.

- Do not cover a tourniquet with a dressing. If you must cover the injured person in some way, make sure that all other people concerned with the case know about the tourniquet. Using crayon, magic marker, or blood, mark a large T on the victim's forehead or on a medical tag attached to the wrist.

Shock

If you've ever hit your finger with a hammer and felt — in addition to the pain — weak, dizzy, and nauseous, then you have experienced a mild form of shock. In such an instance, the symptoms appear immediately after the injury; but they may not show up for several hours.

Shock is a condition in which blood circulation is seriously disturbed. Crushed or fractured bones, burns, prolonged bleeding, and asphyxia all cause shock. It may be slight or it may be severe enough to cause death. Because all injuries will result in some form of shock, you must learn its symptoms and know how to treat the victim.

HOW TO RECOGNIZE SHOCK.—A person who is going into shock may show quite a few signs or symptoms, some of which are indicated in figure 1-9 and are discussed below. Remember, however, that signs of shock do not always appear at the time of the injury; indeed, in many serious cases they may not appear until hours later.

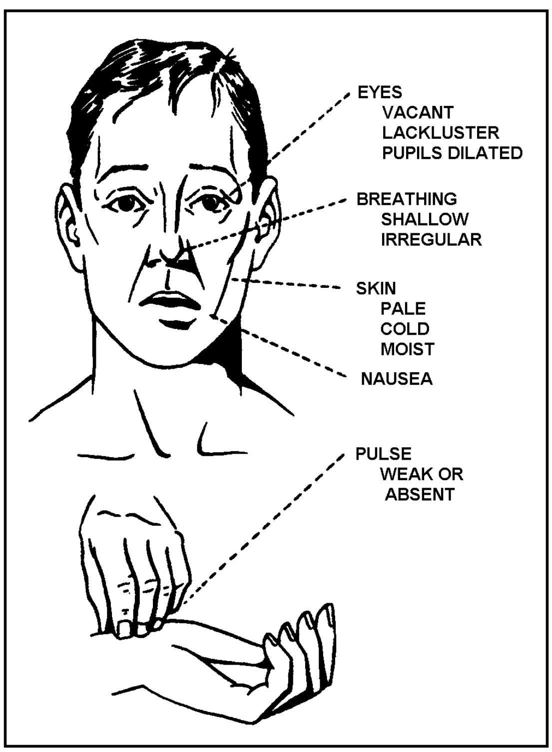


Figure 1-9.—Symptoms of shock.

The symptoms of a person suffering from shock are caused, directly or indirectly, by the disturbance of the circulation of the blood. The pulse is weak and rapid. Breathing is likely to be shallow, rapid, and irregular, because the poor circulation of the blood affects the breathing center in the brain. The temperature near the surface of the body is lowered because of the poor blood flow; so the face, arms, and legs feel cold to the touch. Sweating is likely to be very noticeable. A person in shock is usually very pale, but in some cases the skin may have a bluish or reddish color. As mentioned previously when we were discussing electric shock, in the care of victims with dark skin, it may be necessary to rely primarily on the color of the mucous membranes on the inside of the mouth or under the eyelid or under the nail bed. A person in or going into shock has a bluish color to these membranes instead of a healthy pink. The pupils of the eyes are usually dilated (enlarged).

A conscious person in shock may complain of thirst and have a feeling of weakness, faintness, or dizziness. The victim may feel nauseous, restless, frightened, and/or anxious. As shock deepens, these signs gradually disappear and the victim becomes less and less responsive to what is going on. Even pain may not arouse the shock victim. Finally, the victim may become unconscious.

You will not likely see all these symptoms of shock in any one case. Some of them appear only in late stages of shock when the disturbance of the blood flow has become so great that the person's life is in serious danger. Sometimes the signs of shock may be disguised by other signs of injury. You must know what symptoms indicate the presence of shock, but do not ever wait for symptoms to develop before beginning the treatment for shock. Remember, **EVERY SERIOUSLY INJURED PERSON IS LIKELY TO DEVELOP SERIOUS SHOCK**.

PREVENTION AND TREATMENT OF SHOCK.—You should begin treatment for shock as soon as possible. Prompt treatment may prevent the occurrence of shock or, if it has already developed, prevent its reaching a critical point. Keep the victim lying down and warm. If conscious, the victim should be encouraged and assured that expert medical help will arrive soon.

KEEP AN INJURED PERSON WARM ENOUGH FOR COMFORT, BUT DO NOT LET THE VICTIM BECOME OVERHEATED.

The best position to use for the prevention or the treatment of shock is one which encourages the flow of blood to the brain. If the injured person can be placed on his/her back on a bed, a cot, or a stretcher, you can raise the lower end of the support about 12 inches so that the feet will be higher than the head. The circumstances of the accident may prevent the use of a bed, a cot, or a stretcher. In such cases, you might still be able to raise the feet and legs enough to help the blood flow to the brain. Sometimes you can take advantage of a natural slope of ground and place the casualty so that the head is lower than the feet.

In every case, of course, you will have to consider what type of injury is present before you can decide on the best position. For example, a person with a chest wound may have so much trouble breathing that you will have to raise the head slightly. If the face is flushed rather than pale, or if you have any reason to suspect a head injury, do not raise the feet; instead, you should keep the head level with or slightly higher than the feet. If the person has broken bones, you will have to judge what position would be best both for the fractures and for shock. A fractured spine must be immobilized before the victim is moved at all, if further injuries are to be avoided. If you have any doubts about the correct position to use, have the victim lie flat on his/her back. **THE BASIC POSITION FOR TREATING SHOCK IS ONE IN WHICH THE HEAD IS LOWER THAN THE FEET**. Do the best you can under the particular circumstances to get the injured person into this position. In any case, never let a seriously injured person sit, stand, or walk around.

Liquids should be administered sparingly, and not at all if medical attention will be available within a short time. If necessary, small amounts of warm water, tea, or coffee may be given to a victim who is conscious. Persons having serious burns are an exception. Burn victims require large amounts of fluids. Water, tea, fruit juices, and sugar water may be given freely to a victim who is conscious, able to swallow, and has no internal injuries. Slightly salted water is also beneficial. Alcohol must never be given to a person in shock.

An injured person may or may not be in pain. The amount of pain felt depends in part on the person's physical condition and the type of injury. Extreme pain, if not relieved, can increase the degree of shock. Make the victim as comfortable as possible. Fractures should be immobilized and supported. Immobilization greatly reduces, and sometimes eliminates, pain. Normally, you should not administer drugs, but aspirin may be given for mild pain.

Heat is important in the treatment of shock to the extent that the injured person's body heat must be conserved. Exposure to cold, with resulting loss of body heat, can cause shock to develop or to become worse. You will have to judge the amount of covering to use by considering the weather and the general circumstances of the accident. Often a light covering will be enough to keep the casualty comfortable. Wet clothing should be removed and dry covering provided, even on a hot day. Use blankets or any dry material to conserve body heat. Artificial means of warming (hot water bottles, heated bricks, heated sand) should not ordinarily be used. Artificial heat may cause loss of body fluids (by sweating) and it brings the blood closer to the surface, thus defeating the body's own efforts to supply blood to the vital organs and to the brain. Also, the warming agent may burn the victim.

Burns

The seriousness of a burn depends on two factors: the extent of the burned area and the depth of the burn. Shock can be expected from burns involving 15 percent or more of the body. Burns involving 20 percent endanger life. Without adequate treatment, burns of over 30 percent are usually fatal. The depth of the injury determines whether it is a first, second, or third degree burn.

First degree burns are mildest. Symptoms are slight pain, redness, tenderness, and increased temperature of the affected area.

Second degree burns are more serious. The inner skin may be damaged, resulting in blistering, severe pain, some dehydration, and possible shock.

Third degree burns are worst of all. The skin is destroyed, and possibly also the tissue and muscle beneath it. The skin may be charred, or it may be white and lifeless (from scalds). After the initial injury, pain may be less severe because of destroyed nerve ends. The person may have chills. Some form of shock will result.

Probably the most important aspect is the extent of the burned area. A first degree burn covering a large area could be more serious than a small third degree burn. A sunburn, for example, ranging from mild to serious, is easily obtained, particularly if you are not accustomed to the exposure. If you were to fall asleep while sunbathing, second degree burns, or even third degree burns of a possibly fatal nature, could result.

The most effective immediate treatment of burns and of pain is to immerse the burned area in cold water or to apply cold compresses if immersion is impracticable. Cold water not only minimizes pain, but also reduces the burning effect in the deeper layers of the skin. Gently pat dry the area with lint-free cloth or gauze. Aspirin is also effective for the relief of pain. Continue treatment until no pain is felt when the burned area is exposed to the air.

Burn victims require large amounts of water, which should be slightly salted. Because of the nature of the injury, most burns are sterile. The best treatment for uninfected burns, therefore, is merely to protect the area by covering it with the cleanest (preferably sterile) dressing available. Never apply ointments to a burn nor use petrolatum gauze.

Do not attempt to break blisters or to remove shreds of tissue or adhered particles of charred clothing. Never apply a greasy substance (butter, lard, or petroleum jelly), antiseptic preparations, or ointments. These may cause further complications and interfere with later treatment by medical personnel.

Heatstroke

Sunstroke is more accurately called heatstroke since a person does not have to be exposed to the sun for this condition to develop. It is a less common but far more serious condition than heat exhaustion, since it carries a 20 percent fatality rate. The main feature of heatstroke is the extremely high body temperature, 105º F (41º C) or higher, that accompanies it. In heatstroke, the victim has a breakdown of the sweating mechanism and is unable to eliminate excessive body heat built up while exercising. If the body temperature rises too high, the brain, kidneys, and liver may be permanently damaged.

Sometimes the victim may have preliminary symptoms, such as headache, nausea, dizziness, or weakness. Breathing will be deep and rapid at first, later shallow and almost absent. Usually the victim will be flushed, very dry, and very hot. The pupils will be constricted (pinpoint) and the pulse fast and strong. Figure 1-10 compares these symptoms with those of heat exhaustion.

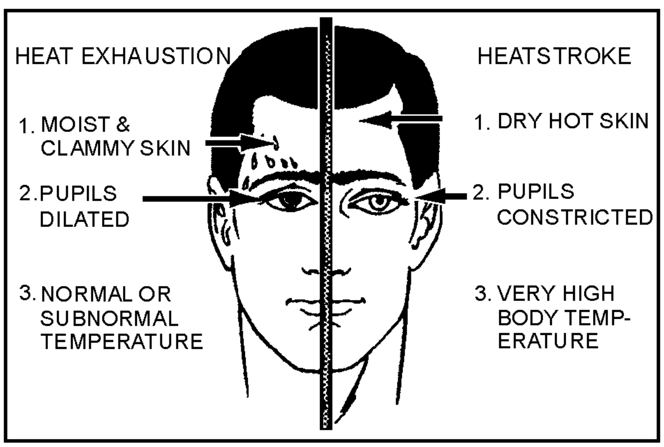


Figure 1-10.—Symptoms of heatstroke and heat exhaustion.

When you provide first aid for heatstroke, remember that this is a true life-and-death emergency. The longer the victim remains overheated, the higher the chances of irreversible body damage or even death occurring. First aid treatment for heatstroke is designed to reduce body heat.

Reduce body heat immediately by dousing the body with cold water, or applying wet, cold towels to the whole body. Move the victim to the coolest possible place and remove as much clothing as possible. Maintain an open airway. Place the victim on his/her back, with the head and shoulders slightly raised. If cold packs are available, place them under the arms, around the neck, at the ankles, and in the groin. Expose the victim to a fan or air-conditioner since drafts will promote cooling. Immersing the victim in a cold water bath is also effective. Give the victim (if conscious) cool water to drink. Do not give any hot drinks or stimulants. Get the victim to a medical facility as soon as possible. Cooling measures must be continued while the victim is being transported.

Heat Exhaustion

Heat exhaustion (heat prostration or heat collapse) is the most common condition caused by working or exercising in hot spaces. Heat exhaustion produces a serious disruption of blood flow to the brain, heart, and lungs. This causes the victim to experience weakness, dizziness, headache, loss of appetite, and nausea.

Signs and symptoms of heat exhaustion are similar to those of shock: the victim will appear ashen gray; the skin will be cold, moist, and clammy; and the pupils of the eyes may be dilated (enlarged). The vital (blood pressure, temperature, pulse, and respiration) signs usually are normal; however, the victim may have a weak pulse together with rapid and shallow breathing. Body temperature may be below normal.

You should treat heat exhaustion victims as if they were in shock. Loosen the clothing, apply cool wet cloths, move the victim to either a cool or an air-conditioned area, and fan the victim. Do not allow the person to become chilled. If the victim is conscious, administer a solution of 1 teaspoon of salt dissolved in a quart of cool water. If the victim vomits, do not give any more fluids. Transport the victim to a medical facility as soon as possible.

HELPFUL INFORMATION

The second part of this handbook has been compiled to provide the technician with a collection of helpful information. Included are many commonly used formulas, data tables, and general maintenance hints used in, with, and around electricity.

BASIC ELECTRICAL FORMULAS

Basic electrical formulas are included to aid you in solving electrical problems. These formulas are for capacitance, current, inductance, power, reactance, impedance, resistance, voltage, and transformers. Additional formulas can be found in the appropriate NEETS module.

Capacitance

The property of an electrical device to store energy is **CAPACITANCE**. This energy is stored in a way to oppose a change in voltage. A **CAPACITOR** is used to store this electrical energy. The **FARAD** is the basic unit of measurement of capacitance.

Formulas for capacitance:

$$C = \frac{Q}{E}$$

C = capacitance in farads

Q = coulombs (a unit of charge equal to 6.28×10^{18} electrons)

E = volts

$$C = 0.2249 \frac{(kA)}{d}$$

A = area of one plate, in square inches

C = capacitance in picofarads

d = distance between the plates in inches

k = dielectric constant of the insulating material

0.2249 is a constant resulting from conversion from metric to British units

Common insulating materials for capacitors and their dielectric constant are:

MATERIAL	CONSTANT
Vacuum	1.0000
Air	1.0006
Paraffin paper	3.5
Glass	5 to 10
Mica	3 to 6
Rubber	2.5 to 35
Wood	2.5 to 8
Glycerine (15º C)	56
Petroleum	2
Pure Water	81

The time to charge a capacitor to 63.2 percent of applied voltage or discharge it to 36.8 percent of its initial voltage is known as the TIME CONSTANT (t) of the circuit. Figure 1-11 shows an RC time constant chart. One time constant (t) in seconds equals R × C, with R in ohms and C in farads.

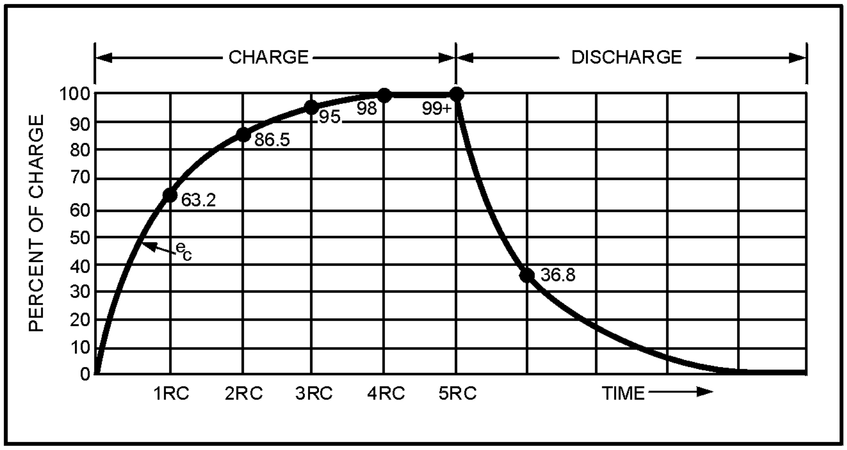


Figure 1-11.—RC time constants.

Figure 1-12 is a universal time constant chart for RC and LR circuits. One time constant (t) in seconds equals R × C, with R in ohms and C in farads, or L/R with L in henries and R in ohms.

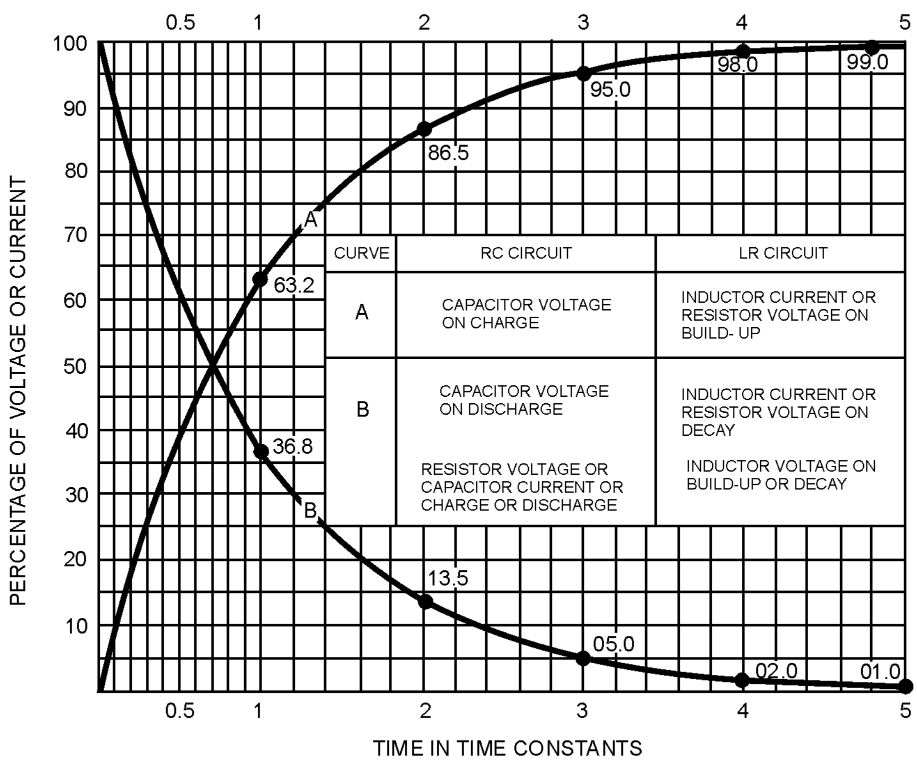


Figure 1-12.—Universal time constant chart for RC and LR circuits.

Adding capacitors in series:

If only two capacitors are used:

$$C_T = \frac{C_1 \times C_2}{C_1 + C_2}$$

If more than two capacitors are used:

$$C_T = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \ldots + \frac{1}{C_n}}$$

Adding capacitors in parallel:

$$C_T = C_1 + C_2 + C_3 + \ldots + C_n$$

CAUTION

Capacitors retain an electrical charge. Be sure to discharge all capacitors and circuits containing capacitors before working on them.

A more detailed description of capacitors and capacitance can be found in NEETS, Module 2, *Introduction to Alternating Current and Transformers.*

Current

Electrons (negative charges) move through a conductor when an electric field is applied. Electron current is defined as the directed flow of electrons from negative to positive.

Current is measured in AMPERES (AMP). One amp of current flows when one coulomb (6.28×10^{18} electrons) passes a point in one second.

The Ohm's law formulas for current are:

$$I = \frac{E}{R}$$

I = current in amps

E = voltage in volts

R = resistance in ohms

$$I = \sqrt{\frac{P}{R}}$$

P = power in watts

$$I = \frac{P}{E}$$

Ac current formulas are:

Average current

$$I_{avg} = 0.636 \times I_{max}$$

Effective current

$$I_{eff} = 0.707 \times I_{max}$$

Maximum current

$$I_{max} = 1.414 \times I_{eff}$$

Ohm's law

$$I_{eff} = \frac{E_{eff}}{R}$$

$$I_{avg} = \frac{E_{avg}}{R}$$

$$I_{max} = \frac{E_{max}}{R}$$

Ohm's law for reactive circuits:

$$I = \frac{E}{X_L} \text{ or } I = \frac{E}{X_C}$$

Ohm's law for circuits containing resistance and reactance:

$$I = \frac{E}{Z}$$

Current across the primary (I_p) of a transformer:

$$I_p = \frac{E_S I_S}{E_p}$$

Current across the secondary (I_s):

$$I_S = \frac{E_p I_p}{E_S}$$

NOTE

Human reaction to electrical shock is determined by the amount of current flowing through the body. A 100-milliampere shock for 1 second is usually fatal!

More information about current can be found in NEETS, Module 1, *Introduction to Matter, Energy, and Direct Current.*

Inductance

Inductance is the characteristic of an electrical conductor that opposes a change in electrical current. The symbol for inductance is L and the basic unit of measurement is the *HENRY(H)*.

An inductor has an inductance of 1 henry if an electromotive force (emf) of 1 volt is induced in the inductor when the current through the inductor is changing at the rate of one ampere per second.

Mathematically:

$$E_{ind} = L\frac{\Delta I}{\Delta t}$$

E_{ind} = induced voltage
L = inductance in henrys
ΔI = change in current in amperes
Δt = change in time in seconds

Mutual inductance:

$$M = K\sqrt{L_1 L_2}$$

M = mutual inductance in henrys
K = coefficient of coupling
$L_1 L_2$ = inductance of coils in henrys

Series inductors without magnetic coupling:

$$L_T = L_1 + L_2 + L_3 \ldots + L_n$$

L_T = total inductance in henrys
L_1, L_2, L_3 = inductance of each inductor
L_n = any number of additional inductors (L_4, L_5, and so forth) that could be used

Series inductors with magnetic coupling:

$$L_T = L_1 + L_2 \pm 2_M$$

M = mutual inductance between two inductors; plus sign is used when the magnetic fields of two inductors aid each other and minus when they oppose.

Parallel inductors without coupling:

$$\frac{1}{L_T} = \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} \ldots + \frac{1}{L_n}$$

Provided the coefficient of coupling between inductors is zero.

Resistive/inductive circuit:

The time required for the current in an inductor to increase to 63.2 percent of its final value or decrease to 36.8 percent is known as the time constant.

Mathematically expressed:

$$t = \frac{L}{R}$$

$$t = \text{seconds}$$

$$L = \text{henrys}$$

$$R = \text{ohms}$$

Figure 1-13 shows an L/R time constant chart.

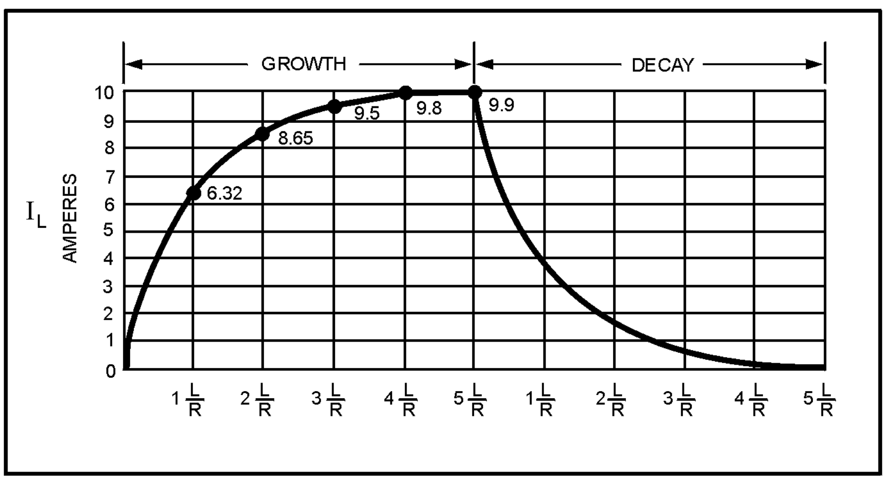


Figure 1-13.—L/R time constant.

You may want to refer back to figure 1-12, which shows the L/R relationship using the universal time constant chart.

Inductive reactance:

$$X_L = 2\pi f L$$
$$X_L = \text{reactance in ohms}$$
$$\pi = 3.1416$$
$$f = \text{frequency in hertz}$$
$$L = \text{inductance in henrys}$$

Ohm's law for reactive circuit:

$$I = \frac{E}{X_L}$$

You can find more information about inductance in NEETS, Module 2, *Introduction to Alternating Current and Transformers*.

Power

Electrical **POWER** pertains to the rate at which work is being done. Work is done whenever a force causes motion. The instantaneous rate at which work is done is called the electric power rate and is measured in **WATTS**.

Formulas for power in dc circuits are:

$$P = I E$$
$$P = I^2 R$$
$$p = \frac{E^2}{R}$$

For ac circuits:

True power:

$$P = (I_R)^2 R$$

True power is measured in watts.

I_R = resistive current in amperes.

R = resistance in ohms.

Reactive power:

$$Q = (I_X)^2 X$$

Reactive power is measured in volt-amperes-reactive.

I_X = reactive current in amps.

X = total reactance in ohms.

Another way to calculate reactive power is:

Reaction power:

$$Q = (I_L)^2 X_L - (I_C)^2 X_C$$

or

$$Q = (I_C)^2 X_C - (I_L)^2 X_L$$

Subtract the smaller from the larger:

I_C = capacitive current in amperes

X_C = capacitive reactance in ohms

I_L = inductive current in amperes

X_L = inductive reactance in ohms

Apparent power:

$$s = (I_Z)^2 Z$$

Apparent power is measured in volt - amperes (VA).

I_Z = impedance current in amperes

Z = impedance in ohms

or

$$S = (\text{true power})^2 + (\text{reactive power})^2$$

Power factor ($\cos\Theta$) :

$$\cos\Theta = \frac{\text{true power}}{\text{apparent power}}$$

or

$$\cos\Theta = \frac{(I_R)^2 R}{(I_Z)^2 Z}$$

or in a series circuit :

$$\cos\Theta = \frac{R}{Z}$$

$\cos\Theta$ is represented as a decimal or percentage.

You can find more detailed information about power in NEETS, Module 2, *Introduction to Alternating Current and Transformers*.

Reactance

REACTANCE is the result of the difference between the values of X_C (capacitive reactance) and X_L (inductive reactance). Reactance is represented by the letter X and its basic measurement is in ohms.

The formula for reactance is:

$$X = X_L - X_c \text{ or } X = X_c - X_L$$

If you want more detailed information on reactance, look in NEETS, Module 2, *Introduction to Alternating Current and Transformers*.

Impedance

IMPEDANCE is the combined opposition of current flow by reactance and resistance and is represented by the symbol Z.

Formulas for finding impedance:

$$Z = \sqrt{R^2 + (X_L - X_c)^2}$$

or

$$Z = \sqrt{R^2 + (X_c - X_L)^2}$$

or

$$Z = \sqrt{R^2 + X^2}$$

Resistance

RESISTANCE is the opposition to current flow. It is measured in ohms and is represented by the letter R.

Formulas for resistance:

$$R = \frac{E}{I}$$

$$R = \frac{E^2}{P}$$

$$R = \frac{P}{I^2}$$

Voltage

VOLTAGE exists when a charge exists between two bodies. When a one coulomb charge exists, one unit of electrical potential energy is created. This is called a difference of potential, an electromotive force, or a voltage. It is measured in volts and represented by the letter E.

Formulas for voltage:

$$E = IR$$

$$E = \frac{P}{I}$$

$$E = \sqrt{PR}$$

Formulas for ac circuits:

(Average voltage)

$$(E_{avg} = 0.636 \times E_{max})$$

(Effective voltage)

$$E_{eff} = 0.707 \times E_{max}$$

(Maximum voltage)

$$E_{max} = 1.414 \times E_{eff}$$

Voltage across the primary of a transformer:

$$E_p = \frac{E_s N_p}{N_s}$$

Voltage across the secondary of a transformer:

$$E_S = \frac{E_p N_S}{N_p}$$

Transformers

A **TRANSFORMER** is a device that transfers electrical energy from one circuit to another by electromagnetic induction (transformer action). Voltage induced into the secondary from the primary is determined by the turns ratio.

Turns ratio formula:

$$\frac{E_S}{E_p} = \frac{N_S}{N_p}$$

or

$$E_p N_S = E_S N_p$$

transposing:

$$E_S = \frac{E_p N_S}{N_p}$$

Where:

E_S = voltage induced in the secondary

E_p = voltage applied to the primary

N_S = number of turns in the secondary

N_p = ampere – turns in the secondary winding

Turns and current ratios:

$$I_p N_p = I_S N_S$$

$I_p N_p$ = ampere – turns in the primary winding

$I_S N_S$ = ampere – turns in the secondary winding

By dividing both sides of the equation by $I_p N_s$, you obtain:

$$\frac{N_p}{N_S} = \frac{I_S}{I_p}$$

Since:

$$\frac{E_S}{E_p} = \frac{N_S}{N_p}$$

Then:

$$\frac{E_p}{E_S} = \frac{N_p}{N_S}$$

And:

$$\frac{E_p}{E_S} = \frac{I_S}{I_p}$$

Where:

E_p = voltage applied to the primary in volts

E_S = voltage across the secondary in volts

I_p = current in the primary in amperes

I_S = current in the secondary in amperes

Transformer power:

$$P_S = P_p - P_L$$

P_S = power delivered to the load by the secondary

P_p = power delivered to the primary by the source

P_L = power losses in the transformer

Transformer efficiency:

$$\text{Efficiency in } (\%) = \frac{P_o}{P_i} \times 100$$

Where:

P_o = total output power delivered to the load

P_i = total input power

Impedance matching transformers:

$$\frac{N_p}{N_S} = \sqrt{\frac{Z_p}{Z_S}}$$

WARNING

Transformers are often used to STEP-UP voltage. You may find a low voltage across the primary and a much higher voltage across the secondary. Use extreme caution, especially when working around television and other crt high voltage transformers. They often step voltages up to, or in excess of, 30,000 volts.

BASIC ELECTRONICS FORMULAS

Basic electronics formulas are included to aid you in solving any electronics problem that you may encounter. These formulas are for antennas, resonance, transistors, vacuum tubes, wavelength, and radar. Additional formulas may be found in the appropriate NEETS module.

Antennas

An antenna is a conductor or a group of conductors used either for radiating electromagnetic energy into space or collecting it from space or both.

Antenna gain remains the same for the antenna whether it is transmitting or receiving. Antenna gain (G) can be described as the effectiveness of a directional antenna in a particular direction, compared to a standard or reference antenna. Some antenna formulas are shown below:

Gain formula:

$$G = KD$$

K = radiation efficiency factor ($K \leq 1$)

D = directivity

Effective aperture:

$$A_e = \frac{W}{P}$$

W = power delivered to a matched load

P = power density

Also:

$$A_e = \frac{\lambda 2G}{4\pi}$$

λ = wavelength (covered in depth later in this section)

G = gain

Directivity of an antenna:

$$D = \frac{U_m}{U_o}$$

U_m = maximum radiation intensity

U_o = average radiation intensity

Also:

$$D = \frac{U_m}{W/4\pi} = \frac{4\pi U_m}{W}$$

W = total power radiated

Field strength:

$$E = \frac{5870P}{D}$$

E = field intensity in millivolts

P = transmitter power in watts

D = distance in miles

Antenna length (L):

(Half-wave up to 30 megahertz)

$$L(\text{feet}) = \frac{492 \times 0.95}{f(MH_z)} = \frac{468}{f(MH_z)}$$

(Half-wave above 30 megahertz)

$$L(\text{feet}) = \frac{492 \times .94}{f(MH_z)} = \frac{462}{f(MH_z)}$$

$$L(\text{inches}) = \frac{5540}{f(MH_z)}$$

WARNING

Rf voltages may be induced in ungrounded metal objects such as wire guys, wire cables (hawsers), handrails, or ladders. You could receive a shock or rf burn if you come in contact with these objects. Obtain proper permission prior to going topside or "working aloft."

Rf burns are usually deep, penetrating, and third degree. They must heal from the inside out. If you are burned, seek medical attention. A person in an RF field will usually have a body temperature rise. The eyes and reproductive organs are especially susceptible to RF energy. Read and heed all warning signs!

You can find more information about antennas in NEETS, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*.

Resonance

RESONANCE is a condition that exists in a circuit when inductance, capacitance, and the applied frequency are such that inductive reactance and capacitive reactance cancel each other.

Formula for resonant frequency (f_r):

$$fr = \frac{1}{2\pi\sqrt{LC}}$$

Where:

fr = resonant frequency in hertz

L = inductance in henrys

C = capacitance in farads

π (pi) = 3.1416

NOTE: The formula for resonance is the same for series or parallel circuits when $X_L = X_C$.

Transistors

Semiconductor devices that have three or more elements are called **TRANSISTORS**. The term is derived from **TRANS**fer and res**ISTOR**. This term describes the operation of the transistor — the transfer of an input signal current from a low-resistance circuit to a high-resistance circuit.

Some transistor formulas are shown below.

Transistor total current:

$$I_E = I_B + I_C$$

Where:

I_E = emitter current

I_B = base current

I_C = collector current

Common emitter gain:

$$\text{Beta } (\beta) \text{ or } h_{fe} = \frac{\Delta I_C}{\Delta I_B}$$

Where:

hfe = h = hybrid

f = forward current ratio

e = common emitter configuration

Δ = delta (indicates a change)

Common base gain:

$$\text{Alpha } (\alpha) = \frac{\Delta I_c}{\Delta I_E}$$

or

When β is known:

$$\alpha = \frac{\beta}{\beta + 1}$$

Alpha is always less than 1 for a common base configuration.

Common collector gain:

$$\text{Gamma or } \gamma = \frac{I_E}{I_B}$$

When beta is known then:

$$\gamma = \beta + 1$$

TRANSISTOR RUGGEDNESS.—Transistors are generally more rugged mechanically than electron tubes. They are susceptible to damage by electrical overloads, heat, humidity, and radiation. Unless you are careful, damage can occur during maintenance.

DAMAGE PREVENTION.—To prevent damage and avoid electrical shock, use the following precautions when working on transistorized equipment:

- Check your test equipment and soldering irons for leakage current from the power source. If leakage current exists, use an isolation transformer to eliminate the current.

- Connect a ground between the test equipment and circuit under test.
- Do not exceed the maximum allowable voltages for circuit components and transistors.
- Ohmmeter ranges that require more than one milliampere should not be used for testing transistors.
- Battery eliminators should not be used to furnish power for transistorized equipment. They have poor voltage regulation.
- The heat applied to a transistor, when soldered connections are required, should be kept to a minimum by the use of low-wattage soldering irons and heat shunts or heat sinks.
- When replacing transistors, never pry them from the printed circuit board.
- Check all circuits for defects before replacing transistors.
- Remove power from the equipment prior to replacing a transistor.
- Use extreme care when using test probes on a transistorized circuit. It is easy to short across adjacent terminals with conventional probes. Try insulating the probe tips and leaving a very short section of the point exposed.

You can find more about transistors in NEETS, Module 7, *Introduction to Solid-State Devices and Power Supplies*.

Vacuum Tubes

The characteristics of a vacuum tube are measured by two factors: **AMPLIFICATION FACTOR**, mu (μ), and **TRANSCONDUCTANCE** (gm).

Formula for amplification factor:

$$\mu = \frac{\Delta E_p}{\Delta E_g}$$

$$E_p = \text{change in plate voltage}$$

$$E_g = \text{change in grid voltage}$$

Formula for transconductance (gm) is:

$$gm = \frac{I_p}{E_g}$$

$$I_p = \text{change in plate current}$$

$$E_g = \text{change in grid voltage}$$

Information on vacuum tubes can be found in NEETS, Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies*.

CAUTION

Because vacuum tubes become hot and most are made of glass, use caution while removing or replacing them. Use vacuum tube pullers or at least wear some type of hand protection.

Wavelength

Wavelength is the distance in space occupied by one cycle of a radio wave at any given instant. If a radio wave could be frozen in time and measured, the distance from the leading edge of one cycle to the leading edge of the next cycle would be the wavelength. Wavelength varies from a few hundredths of an inch at the high frequencies to many miles at extremely low frequencies. Common practice is to express wavelength in meters. The Greek letter lambda (!) is used to signify wavelength. Formulas for wavelength, period, and velocity are shown below.

Wavelength formula:

$$\lambda = \frac{v}{f}$$

Lambda (λ) = wavelength in feet

v = velocity of propagation in feet per second

f = frequency in Hz

Frequency formula:

$$f = \frac{1}{T}$$

T = time of one wave period (cycle) in seconds

f = frequency in Hz

Period formula:

$$T = \frac{1}{f}$$

Velocity formula:

$$v = \lambda f$$

You can find more information on wavelength, frequency, period, and velocity in NEETS, Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas*.

Radar

Some helpful radar information/formulas are shown below.

PW — Pulse Width—The width of the transmitted RF pulse from the radar (the term can also be used regarding other non-RF information).

$$PW = DC \times PRT$$

PRT — Pulse Repetition Time—The duration of the time between radar transmitter pulses (leading edge of pulse to leading edge of the next pulse).

$$PRT = \frac{1}{PRF}$$

PRR or PRF — PRF Pulse Repetition Rate or Pulse Repetition Frequency—the number of pulses that occur during one second.

$$PRF = \frac{1}{PRT}$$

DC — Duty Cycle—The amount of actual transmitter pulse time (PW) divided by the amount of transmitter total time (PRT). For radar applications, the duty cycle will always be less than one.

$$DC = \frac{PW}{PRT} \text{ or } \frac{P_{avg}}{P_{pk}}$$

P_{pk} — Power Peak (normally referenced in kilowatts)—The actual power of the transmitted RF pulse (PW).

$$P_{pk} = \frac{P_{avg}}{DC}$$

P_{avg} — Average Power (normally referenced in watts)—The transmitted power relative to one PRT.

$$P_{avg} = DC \times P_{pk}$$

$$\text{Nautical radar mile} = \frac{\text{elapsed time}}{12.36 \text{ microseconds}}$$

Minimum radar range = (pulse width + *recovery time) × 164 yards

*In most modern radar systems, recovery time is negligible and does not need to be considered when figuring minimum radar range.

$$\text{Radar horizon distance (nautical miles)} =$$
$$1.25\sqrt{\text{antenna height (ft)}}$$

Radar range resolution (in yards) = pulse width × 164 yards per microsecond.

When you use the term decibel (dB), you are referring to a logarithmic comparison between two signals, usually the output and the input. In power measurement applications, however, a reference of 0 dBm equalling 1 milliwatt is usually used. The term dBm is used to represent power levels above, below, or at 1 milliwatt.

The following formulas are used for figuring dB:

As a power ratio:

$$dB = 10 \log_{10} \frac{P_2}{P_1}$$

As a voltage ratio:

$$dB = 20 \log_{10} \frac{E_2}{E_1}$$

As a current ratio:

$$dB = 20 \log_{10} \frac{I_2}{I_1}$$

In circuits where impedances may vary:

$$\text{Voltage dB} = 20 \log_{10} \frac{E_2 R_1}{E_1 R_2}$$

$$\text{Current dB} = 20 \log_{10} \frac{I_2 R_2}{I_1 R_1}$$

Some basic information to remember:

Power

A gain of 1 dB power is equal to 1. 25 × that power

A gain of 3 dB power is equal to 2. 00 × that power

A gain of 10 dB power is equal to 10. 0 × that power

A gain of 1 dB power is equal to 0. 80 × that power

A loss of 3 dB power is equal to 0. 50 × that power

A loss of 10 dB power is equal to 0. 10 × that power

Voltage

A gain of 1 dB is equal to 1. 118 × that voltage

A gain of 6 dB is equal to 2. 000 × that voltage

A gain of 20 dB is equal to 10. 00 × that voltage

A loss of 1 dB is equal to 0. 894 × that voltage

A loss of 6 dB is equal to 0. 500 × that voltage

A loss of 20 dB is equal to 0. 100 × that voltage

POWER CONVERSION

For ease of power conversion, this listing provides rough, basic data:

Log or dB	Gain or mw
11	12.5
10	10
9	8
8	6.25
7	5
6	4
5	3.125
4	2.5
3	2
2	1.6
1	1.25
0	1
-1	0.8
-2	0.625
-3	0.5
-4	0.4
-5	0.312
-6	0.25
-7	0.2
-8	0.16
-9	0.125
-10	0.1
-11	0.08

The figures in the above listing are not precise, but are accurate for most applications. For figures in between the above numbers, logarithm interpolation must be done. Table 1-34 provides a seven-place table of common logarithms.

The following are examples of power conversion:

Example 1:

Convert 56 dBm to watts

5 6 dBm (m = referenced to 1 mw)

6 dBm = 4mw

5 = mantissa or exponent = 10^5

$= 4\text{mW} \times 10^5$

$= 4\ \text{W} \times 10^2$

= 400 watts

Example 2:

Convert -52 dBm to μw (microwatts)

-5 2 dBm

$-2\ \text{dBm} = .625\ \text{mw}$

$-5 = \text{mantissa or exponent} = 10^{-5}$

$= .625\ \text{mw}\ 10^{-5}$

$= .625\ \mu\text{w} \times 10^{-2}$

$= .00625\ \mu\text{w}$

Example 3:

Convert 80 kw to dBm

$$= \frac{8 \times 10^{7}\ mw}{79\ dbm} \quad 8\ mw = 9\ dB$$

Convert 5 µw to dBm

$$= \frac{5 \times 10^{-3}\ mw}{-37\ dBm} \quad 5\ mw = 7\ dBm$$

DATA TABLES

Data tables are provided for reference. Information in the tables are the usually accepted standards. Various military standards have also been used to provide these tables.

Capacitor Identification

Two methods of capacitor identification are used. The first is the typographical method, and the second is the color code method. Typographically marked identification will be discussed first. It is the type marking where a number is printed on the capacitor. You should note that on different physical styles of capacitors, the printed number may be in either microfarads or picofarads. Two styles of capacitors have been selected from Military Standard 198E (MIL-STD-198E) to show how the part number stamped on the capacitor is broken down.

Table 1-1 shows the CB style of capacitor part number breakdown.

Table 1-1.—CB Style Capacitor Part Number Breakdown

CB50 R B 050 K

STYLE - The two-letter symbol "CB" identifies button-style, mica-dielectric, fixed capacitors; the two-digit number identifies the shape, dimensions, and operating temperature range.

TERMINAL ASSEMBLY - The single-letter symbol identifies the terminal assembly as follows:

Symbol	Terminal assembly
P - - - - - - - - - - - -	Single L
R - - - - - - - - - - - -	Double L

CHARACTERISTIC - The single-letter symbol identifies the relative stability of the capacitor with temperature change as follows:

Symbol	Temperature coefficient	Capacitance drift
	(ppm/°C)	
B - - -	Not specified	Not specified
D - - -	-100 to +100	0.5% or 0.5 pF, whichever is greater
E - - -	-20 to +100	±(0.1% +0.1 pF)
F - - -	-50 to +100	±(0.1% +0.1 pF)

CAPACITANCE - The three-digit number identifies the nominal capacitance value, expressed in picofarads (pF); the first two digits represent significant figures and the last digit specifies the number of zeros to follow.

CAPACITANCE TOLERANCE - The single-letter symbol identifies the capacitance tolerance as follows:

Symbol	Capacitance tolerance
	Percent (±)
F - - -	1
G - - -	2
J - - -	5
K - - -	10

Table 1-2 shows how the part number on the CMR style of capacitor is broken down

Table 1-2.—CMR Style Capacitor Part Number Breakdown

CMR05 C 1R0 D O C M

STYLE - The three-letter symbol "CMR" identifies established reliability, mica dielectric, fixed capacitors; the two-digit number identifies the shape and dimensions of the capacitor.

CHARACTERISTIC - The single-letter symbol indicates the relative stability of the capacitor with temperature change as follows:

Symbol	Temperature coefficient (ppm/°C)	Capacitance drift
C - - -	-200 to +200	±(0.5% + 0.1 pF)
E - - -	-20 to +100	±(0.1% + 0.1 pF)
F - - -	0 to +70	±(0.05% + 0.1 pF)

CAPACITANCE - The three-digit number identifies the nominal capacitance value, expressed in picofarads (pF). Where the nominal capacitance is 10 pF or greater, the first two digits represent significant figures and the last digit specifies the number of zeros to follow (for example: 10 pF = 100; 150 pF = 151; 1500 = 152). Where the nominal capacitance is less than 10 pF (a whole number), the letter "R" shall represent the decimal point, and the other digits are significant (for example: 1 pF = 1R0).

CAPACITANCE TOLERANCE - The single-letter symbol identifies the capacitance tolerance as follows:

Symbol	Capacitance tolerance (±)
D - - - - - - - - - - - -	0.5 pF
F - - - - - - - - - - - -	1 percent
G - - - - - - - - - - - -	2 percent
J - - - - - - - - - - - -	5 percent

OPERATING TEMPERATURE RANGE - A single-letter symbol identifies the operating temperature range as follows:

Symbol	Operating temperature range
O - - - - - - - - - - -	-55° to +125°C
P - - - - - - - - - - -	-55° to +150°C

RATED VOLTAGE - The single-letter symbol identifies the rated voltage as follows:

Symbol	Rated voltage (Volts, dc)
Y - - - - - - - - - - -	50
A - - - - - - - - - - -	100
C - - - - - - - - - - -	300
D - - - - - - - - - - -	500

FAILURE RATE LEVEL - The single-letter symbol identifies the failure rate level as follows:

Symbol	Failure rate level (% 1,000 h)
M - - - - - - - - - -	1.0
P - - - - - - - - - -	0.1
R - - - - - - - - - -	0.01
S - - - - - - - - - -	0.001

Table 1-3 is a partial cross reference list of the CYR10 (MIL-C-23269/1) style of capacitor. As an example, if you need a 3.3 pF, 500 VDC capacitor in the CYR10 style, with a failure rate (FR) of 1 percent per 1,000 hours, the part number would be M23269/01-3009.

Table 1-3.—CYR10 Style Capacitor Cross-Reference

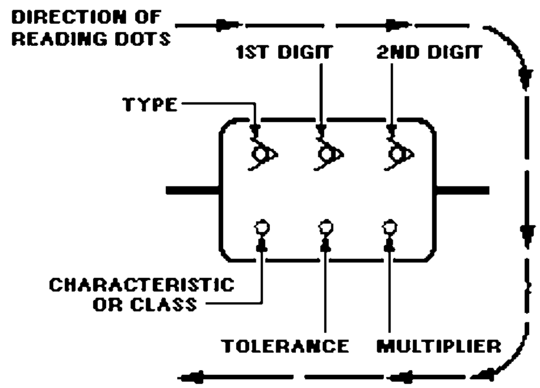
STYLE CYR10 (MIL-C-23269/1)

OPERATING TEMPERATURE RANGE -55° TO +125°C -- TEMPERATURE COEFFICIENT 140 ± 25 PPM/°C -- CAPACITANCE DRIFT 0.1% OR 0.1 pF, WHICHEVER IS GREATER

Capacitance value	DC rated voltage	Capacitance tolerance	Dash number M23269/01- FR level in %/1,000 hours			
			M (1.0)	P (0.1)	R (0.01)	S (0.001)
pF	volts, dc					
0.5	500	±0.25 pF	3001	4001	5001	6001
1.0		±0.25 pF	3002	4002	5002	6002
1.5		±0.25 pF	3003	4003	5003	6003
2.2		±0.25 pF	3004	4004	5004	6004
2.7		±0.25 pF	3006	4006	5006	6006
3.3		±0.25 pF	3009	4009	5009	6009
3.6		±0.25 pF	3010	4010	5010	6010
3.9		±0.25 pF	3012	4012	5012	6012
4.7		±0.25 pF	3015	4015	5015	6015
5.6		±0.25 pF	3017	4017	5017	6017
5.6		±5%	3018	4018	5018	6018
6.8		±0.25 pF	3021	4021	5021	6021
6.8		±5%	3022	4022	5022	6022
8.2		±0.25 pF	3025	4025	5025	6025
8.2		±5%	3026	4026	5026	6026

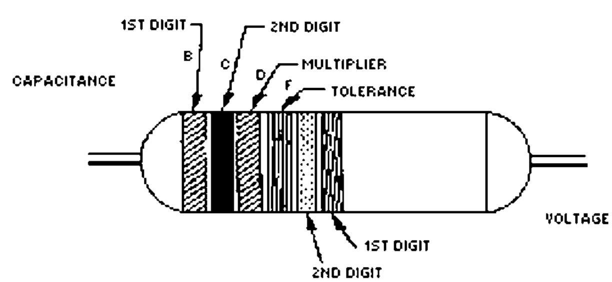
Table 1-3 is not a complete list of the CYR10 style capacitor. The CYR10 is not the only style capacitor listed this way. You should refer to Military Standard 198E (MIL-STD-198E), *Capacitors, Selection and Use of,* for more information.

The color coding identification method is becoming obsolete. This method is included for the technician who is required to work on some older model equipment.



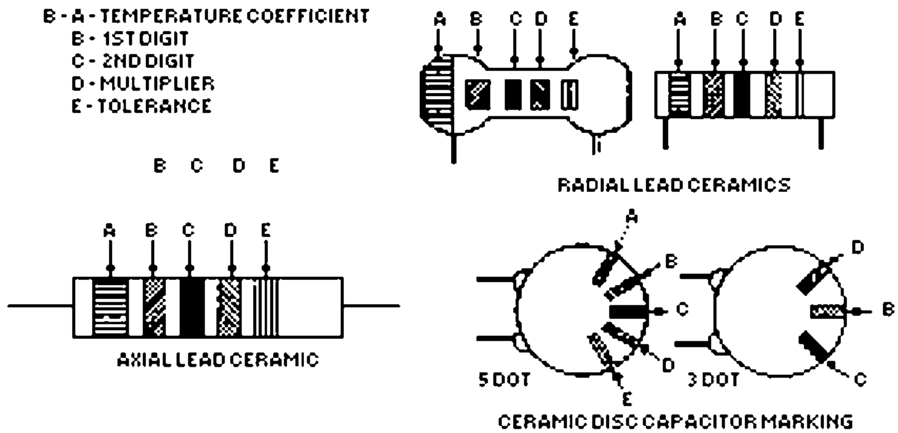
TYPE	COLOR	1ST DIGIT	2ND DIGIT	MULTIPLIER	TOLERANCE (PERCENT)	CHARACTERISTIC OR CLASS
JAN.	BLACK	0	0	1.0	±20	APPLIES TO
MICA	BROWN	1	1	10		
	RED	2	2	100	±2	TEMPERATURE
	ORANGE	3	3	1,000		
	YELLOW	4	4	10,000		COEFFICIENT
	GREEN	5	5	100,000		
	BLUE	6	6	1,000,000		OR METHODS
	VIOLET	7	7	10,000,000		
	GRAY	8	8	100,000,000		OF TESTING
EIA.	WHITE	9	9	1,000,000,000		
	GOLD			0.1	±5	
MOLDED	SILVER			0.01	± 10	
PAPER	BODY				±20	

Figure 1-14.—Six-dot color code for mica and molded paper capacitors.



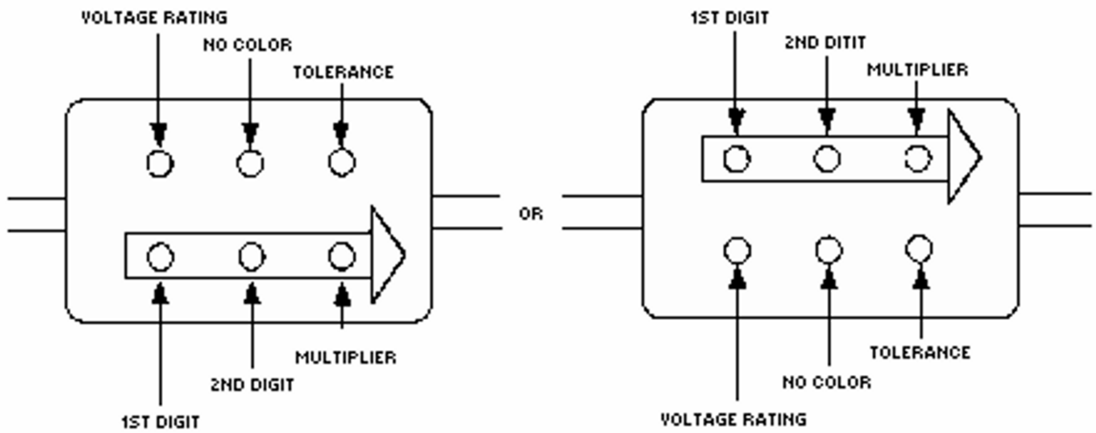
COLOR	CAPACITANCE			TOLERANCE (PERCENT)	VOLTAGE RATING	
	1ST DIGIT	2ND DIGIT	MULTIPLIER		1ST DIGIT	2ND DIGIT
BLACK	0	0	1	±20	0	0
BROWN	1	1	10		1	1
RED	2	2	100		2	2
ORANGE	3	3	1,000	±30	3	3
YELLOW	4	4	10,000	±40	4	4
GREEN	5	5	100,000	±5	5	5
BLUE	6	6	1,000,000		6	6
VIOLET	7	7			7	7
GRAY	8	8			8	8
WHITE	9	9		±10	9	9

Figure 1-15.—Six-band color code for tubular paper dielectric capacitors.



COLOR	1ST DIGIT	2ND DIGIT	MULTIPLIER	TOLERANCE MORE THAN 10 pF (IN PERCENT)	TOLERANCE LESS THAN 10 pF (IN pF)	TEMPERATURE COEFFICIENT
BLACK	0	0	1.0	±20	± 2.0	0
BROWN	1	1	10	± 1		-30
RED	2	2	100	± 2		-80
ORANGE	3	3	1,000			-150
YELLOW	4	4	10,000			-220
GREEN	5	5		± 5	± 0.5	-330
BLUE	6	6				-470
VIOLET	7	7				-750
GRAY	8	8	.01		± 0.25	+ 30
WHITE	9	9	.1	± 10	± 1.0	+120 TO -750 (EIA) +500 TO -330 (JAN)
SILVER						+100 (JAN)
GOLD						BYPASS OR COUPLING (EIA)

Figure 1-16.—Ceramic capacitor color code.



COLOR	1ST DIGIT	2ND DIGIT	MULTIPLIER	TOLERANCE (PERCENT)	VOLTAGE RATING
BLACK	0	0	1.0		
BROWN	1	1	10	±1	100
RED	2	2	100	±2	200
ORANGE	3	3	1,000	±3	300
YELLOW	4	4	10,000	±4	400
GREEN	5	5	100,000	±5	500
BLUE	6	6	1,000,000	±6	600
VIOLET	7	7	10,000,000	±7	700
GRAY	8	8	100,000,000	±8	800
WHITE	9	9	1,000,000,000	±9	900
GOLD			.1		1000
SILVER			.01	±10	2000
BODY				±20	*

* WHERE NO COLOR IS INDICATED, THE VOLTAGE RATING MAY BE AS LOW AS 300 VOLTS.

Figure 1-17.—Mica capacitor color code.

Table 1-4 shows some principal capacitor applications by type and military specification.

Table 1-4.—Principal Applications of Capacitors

MILITARY SPECIFICATION	APPLICATION												
	Establishing Reliability	Capacitor Type	Blocking	Buffing	By-passing	Coupling	Filtering	Tuning	Temperature compensating	Trimming	Motor starting	Timing	Noise suppression
MIL-C-5		Mica	x	x	x	x	x	x				x	
MIL-C-20	x	Ceramic			x	x	x	x	x				
MIL-C-62		Aluminum			x		x						
MIL-C-81		Ceramic Trimmer		x		x		x					
MIL-C-10950		Mica			x	x		x		x			
MIL-C-11015		Ceramic	x		x	x							
MIL-C-14409		Piston Trimmer						x		x			
MIL-C-19978	x	Plastic	x	x	x	x	x						
MIL-C-23183		Vacuum	x		x	x	x	x					
MIL-C-23269	x	Glass	x		x	x		x					
MIL-C-39001	x	Mica	x	x	x	x	x	x				x	x
MIL-C-39003	x	Solid Tantalum	x		x	x	x				x		
MIL-C-39006	x	Wet Tantalum	x		x	x	x						
MIL-C-39014	x	Ceramic			x	x	x						
MIL-C-39018	x	Aluminum	x		x	x	x						
MIL-C-39022	x	Met. Plastic	x		x	x	x						
MIL-C-55365	x	Solid Tantalum, Chip			x	x	x						
MIL-C-55514	x	Plastic	x		x								
MIL-C-55681	x	Ceramic, Chip			x	x	x				x		
MIL-C-83421	x	Met. Plastic	x	x	x	x	x						

Table 1-5 is a capacitor style to military specification cross referencing. This cross reference guide is included for general information only; some styles are not preferred standards and, therefore, are not included in this standard.

Table 1-5.—Style to Military Specification Cross-Reference

STYLE	SPECIFICATION	DESCRIPTION	CLASS	STATUS	REPLACEMENT
CA	MIL-C 12889	Paper, By-Pass	Non-ER	I	19978
CB	10950	Mica, Button, Feed-Thru	Non-ER	A	
CC	20	Ceramic, Encap., Temp. Comp.	Non-ER	PI	CCR
CCR	20	Ceramic, Encap., Temp. Comp.	ER	A	
CDR	55681	Ceramic, Chip	ER	A	
CE	62	Aluminum Electrolytic	Non-ER	PI	39018
CFR	55514	Plastic, Non-Herm. Sealed	ER	A	
CG	23183	Vacuum or Gas, Variable	Non-ER	A	
CH	18312	Metallized Paper, or Plastic	Non-ER	I	39022
CHR	39022	Metallized Plastic, Herm. Sealed	ER	A	
CJ	3871	Aluminum, Motor Start	Non-ER	C	EIA RS-463 39014
CK	11015	Ceramic, Encapsulated	Non-ER	PI	
CKR	39014	Ceramic, Encapsulated	ER	A	
CKS	123	Ceramic, Encapsulated and Chip	Hi-Rel	A	
CL	3965	Tantalum, Foil and Wet Slug	Non-ER	I	39006
CLR	39006	Tantalum, Foil and Wet Slug	ER	A	
CM	5	Mica, Molded, Silvered, and RF	Non-ER	PI	39001

Table 1-5.—Style to Military Specification Cross-Reference. —Continued

STYLE	SPECIFICATION	DESCRIPTION	CLASS	STATUS	REPLACEMENT
CMR	39001	Mica, Silvered	ER	A	
CMS	87164	Mica, Silvered	Hi-Rel	A	
CN	91	Paper, Non-Metal Cases	Non-ER	C	55514
CP	25	Paper, Herm. Sealed	Non-ER	I	19978
CPV	14157	Paper or Plastic, Herm. Sealed	Non-ER	C	19978
CQ	19978	Paper or Plastic, Herm. Sealed	Non-ER	I	CQR
CQR	19978	Paper or Plastic, Herm. Sealed	ER	A	
CRH	83421	Metallized Plastic, Herm. Sealed	ER	A	
CRL	83500	Tantalum, Wet Slug	Non-ER	A	
CS	26655	Tantalum, Solid, Herm. Sealed	Non-ER	C	39003
CSR	39003	Tantalum, Solid, Herm. Sealed	ER	A	
CSS	39003	Tantalum, Solid, Herm. Sealed	Hi-Rel	A	
CT	92	Air, Variable	Non-ER	A	
CTM	27287	Plastic, Non-Metal Case	Non-ER	I	55514
CU	39018	Aluminum Electrolytic	Non-ER	PI	CUR
CUR	39018	Aluminum Electroyltic	ER	A	
CV	81	Ceramic, Variable	Non-ER	A	
CWR	55365	Tantalum, Solid, Chip	ER	A	
CX	49137	Tantalum, Solid, Non-Herm. Sealed	Non-ER	A	
CY	11272	Glass	Non-ER	I	23269
CYR	23269	Glass	Non-ER	A	
CZ	11693	Metallized Paper or Plastic F.T.	Non-ER	I	CZR
CZR	11693	Metallized Paper or Plastic F.T.	ER	A	
PC	14409	Piston Trimmer	Non-ER	A	

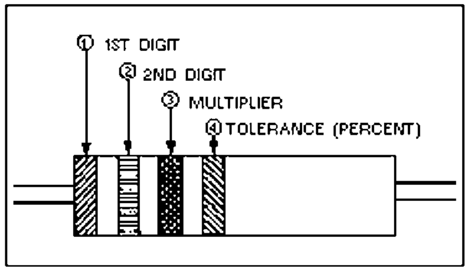
A = Active for design
C = Canceled
I = Inactive for design
PI = Partially Inactive for design

ER = Extended reliability
NON-ER = Not extended reliability
HI-REL = High reliability

Military Standard 198E (MIL-STD-198E) contains information concerning capacitors and should be helpful in selecting any replacement.

Resistor Identification

This section contains information that will aid you in identifying the specifications indicated on resistors.



COLOR	1ST DIGIT	2ND DIGIT	MULTIPLIER	TOLERANCE (PERCENT)
BLACK	0	0	1	
BROWN	1	1	10	
RED	2	2	100	
ORANGE	3	3	1,000	
YELLOW	4	4	10,000	
GREEN	5	5	100,000	
BLUE	6	6	1,000,000	
VIOLET	7	7	10,000,000	
GRAY	8	8	100,000,000	
WHITE	9	9	1,000,000,000	
GOLD			.1	5
SILVER			.01	10
NO COLOR				20

RESISTORS FOR MILITARY USE MAY HAVE A FIFTH HAND TO INDICATE RELIABILITY IN TERMS OF FAILURE RATE, AS FOLLOWS:

NO COLOR: NO TEST MADE
BROWN : 1.0 PERCENT PER 1000 HOURS
RED : 0.1 PERCENT PER 1000 HOURS
ORANGE : 0.01 PERCENT PER 1000 HOURS
YELLOW : 0.001 PERCENT PER 1000 HOURS

Figure 1-18.—Resistor color code.

Table 1-6 is a resistor selection chart. As an example, let's suppose you need a 1-watt, composition resistor. Look under the "type" heading to find composition. Then look under the "power and max voltage ratings" headings to find 1W/500V. The style you select should be RCR32. This is the first part of the part number. To find the last part of the desired part number, the ohmic value, refer to table 1-7, the resistor type designation part number breakdown.

Tables 1-6 and 1-7 are excerpts from Military Standard 199C(MIL-STD-199C) and are included as examples of the information contained in MIL-STD-199C. If another type of resistor is needed, the complete part number breakdown can be located in MIL-STD-199C.

Table 1-6.—Resistor Selection Chart

Military specification	Type	Styles available in standard	Power and max voltage ratings	Resistance tolerance (+ percent)	Ohmic range
MIL-R-26	Wirewound (Power Type)	RW29	11W	5, 10	.1 to 5.6 K
		RW31	14W		.1 to 6.8 K
		RW33	26W		.1 to 18 K
		RW35	55W		.1 to 43 K
		RW37	113W		.1 to 91 K
		RW38	159V		.1 to.15 M
		RW47	210W		.1 to.18 M
		RW56	14W		.1 to 9.1 K
MIL-R-22684	Film (Insulated)	RL42. . .TX	2W/500V	2, 5	10 to 1.5 M
MIL-R-18546	Wirewound (Power Type, Chassis Mounted)	RE77	75W	1	.05 to 29.4 K
		RE80	120W	1	.1 to 35.7 K
MIL-R-39008	Composition (Insulated), Established Reliability	RCR05	.125W/150V	5, 10	2.7 to 22 M
		RCR07	.25W/250V		2.7 to 22 M
		RCR20	.5W/350V		1.0 to 22 M
		RCR32	1W/500V		1.0 to 22 M
		RCR42	2W/500V		10 to 22 M
MIL-R-55182	Film, Established Reliability	RNR50	.05W/200V .1W/200V	.1, .5, 1	10 to.796 M
		RNR55	.1W/200V .125/200V		10 to 2.0 M
		RNR60	.125W/250V .25W/300V		1.0 to 4.02 M
		RNR65	.25W/300V .5W/350V		1.0 to 8.06 M
		RNR70	.5W/350V .75W/500V		1.0 to 15 M
		RNR75	1W/750V 2W/750V		24.9 to 2 M
		RNC90 .	.3W/300V 6W/300V	.05, .01, .005	4.99 to 100K

1/ M = megohms; K = kilohms.

Table 1-6.—Resistor Selection Chart—Continued

Military specification	Type	Styles available in standard	Power and max voltage ratings	Resistance tolerance (+ percent)	Ohmic range
MIL-R-39005	Wirewound (Accurate), Established Reliability	RBR52 RBR53 RBR54 RBR55 RBR56 RBR57 RBR71 RRBR75	.5W/600V .33W/300V .25W/300V .15W/200V .125W/150V .75W/600V .125W/150V .125W/150V	.01, .05, .1, 1	.1 to.806 M .1 to.499 M .1 to.255 M .1 to.150 M .1 to.1 M .1 to 1.37 M .1 to.1 M .1 to 71.5 K
MIL-R-39007	Wirewound (Power Type), Established Reliability	RWR78 RWR80 RWR81 RWRB2 RWR84 RWR89	10W 2W 1W 1.5W 7W 3W	.1, .5, 1	.1 to 39.2 K .1 to 1.21 K .1 to.464 K .1 to.931 K .1 to 12.4 K .1 to 3.57 K
MIL-R-39017	Film (Insulated), Established Reliability	RLR05 RLR07 RLR20 RLR32	.125W/200V .25W/250V .5W/350V 1W/500V	1, 2	4.7 to.3 M 10 to 2.49 M 4.3 to 3.01 M 10 to 1.0 M
MIL-R-39009	Wirewound (Power Type, Chassis Mounted), Established Reliability	RER40 RER45 RER50 RER55 RER60 RER65 RER70 RER75	5W 10W 20W 30W 5W 10W 20W 30W	1	1 to 1.65 K 1 to 2.80 K 1 to 6.04 K 1 to 19.6 K .1 to 3.32 K .1 to 5.62 K .1 to 12.1K .1 to 39.2 K
MIL-R-55342	Film, Chip, Established Reliability	RM0502 RM0505 RM0705 RM1005 RM1505 RM2208	.02W/40V .15W/40V .10W/40V .15W/40V .10W/50V .225W/50V	1, 5, 10	5.6 to .1 M 5.6 to .47 M 5.6 to .1 M 5.6 to .47 M 5.6 to .1 M 5.6 to 15 M

1/ M = megohms; K = kilohms.

Table 1-7.—Resistor Type Designation Part Number Breakdown

RCR07 G 470 J M

STYLE: The three-letter symbol "RCR" identifies established reliability, insulated, composition, fixed resistors; the two-digit number identifies the size and power rating.

CHARACTERISTIC: The single-letter symbol "G" identifies the resistance-temperature characteristic as shown in table 301-I.

RESISTANCE: The three-digit number identifies the nominal resistance value, expressed in ohms; the first two digits represent significant figures and the last digit specifies the number of zeros to follow. Where fractional values of an ohm are required, the letter "R" represents the decimal point. (See 3.3 and the following example.)

EXAMPLE:

2R7	2.7 ohms
100	10 ohms
105	1 megohm
226	22 megohm

RESISTANCE TOLERANCE: The single-letter symbol identifies the resistance tolerance as follows:

J	± 5 percent
K	± 10 percent

FAILURE RATE LEVEL: The single-letter symbol identifies the failure rate level as follows:

M	1.0 percent/1,000 hr
P	0.1 percent/1,000 hr
R	0.01 percent/1,000 hr
S	0.001 percent/1,000 hr

Transformer Lead Identification

This area contains color coding identification as it relates to transformers.

Figure 1-19 shows the color codes for power transformers, IF transformers, and interstage-audio transformers.

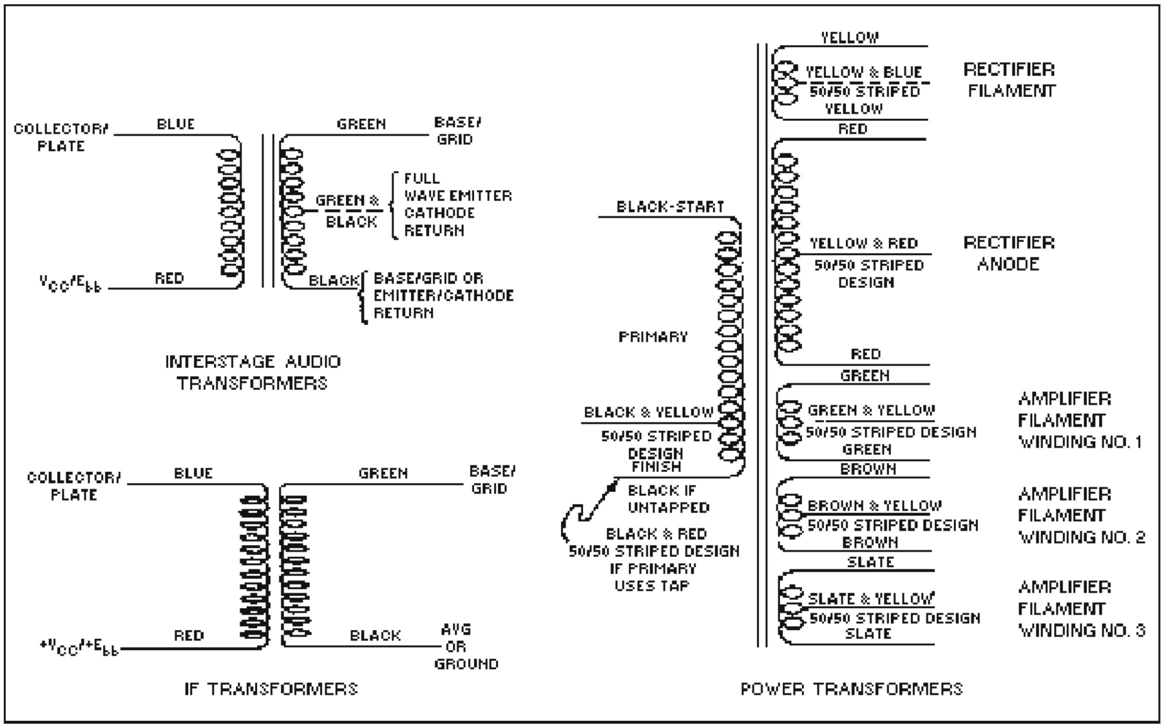


Figure 1-19.—Color code for transformers.

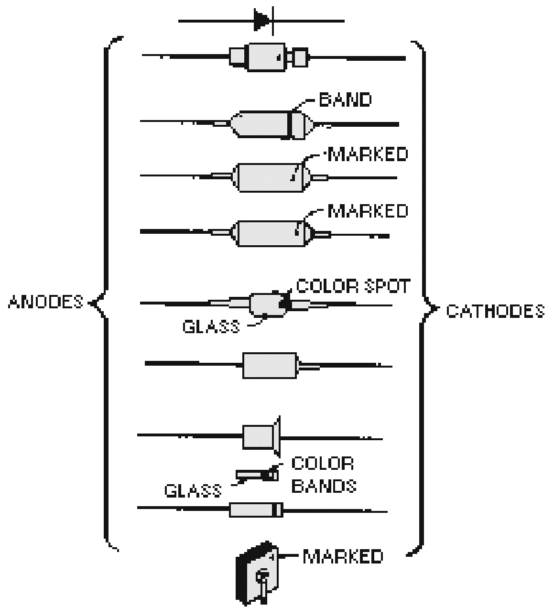
Chassis Wiring Identification

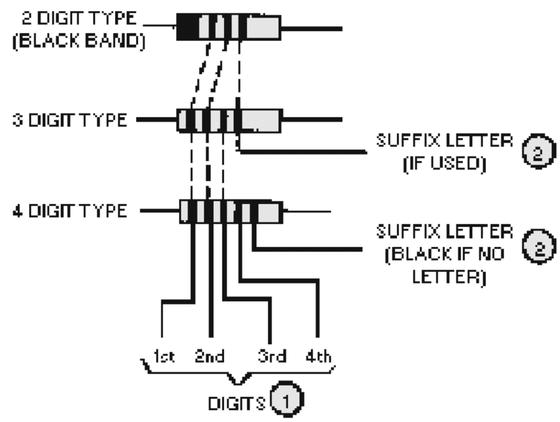
The standard colors used in chassis wiring for the purpose of equipment circuit identification follow:

CIRCUIT	COLOR
GROUNDS, GROUNDED ELEMENTS, AND RETURNS	BLACK
BEATERS OR FILAMENTS, OFF GROUND	BROWN
POWER SUPPLY $+V_{cc}/+E_{bb}$	RED
SCREEN GRIDS	ORANGE
EMITTERS/CATHODES	YELLOW
BASES/CONTROL GRIDS	GREEN
COLLECTORS/PLATES	BLUE
POWER SUPPLY, $-V_{cc}$ /$-E_{bb}$	VIOLET (PURPLE)
AC POWER LINES	GRAY
MISCELLANEOUS, ABOVE OR BELOW GROUND RETURNS, AUTOMATIC VOLUME CONTROL (AVC)	WHITE

Semiconductor Case Outlines, Color Coding, Lead Identification, and Pin Placement

Case outlines, color coding, lead identification, and pin placements of common semiconductor devices are used frequently by the technician and are included in figures 1-20, 1-21, and 1-22.





COLOR	1 DIGIT	2 DIODE SUFFIX LETTER
BLACK	0	-
BROWN	1	A
RED	2	B
ORANGE	3	C
YELLOW	4	D
GREEN	5	E
BLUE	6	F
VIOLET	7	G
GRAY	8	H
WHITE	9	J
SILVER	-	-
GOLD	-	-
NONE	-	-

Figure 1-20.—Semiconductor diode markings and color code system.

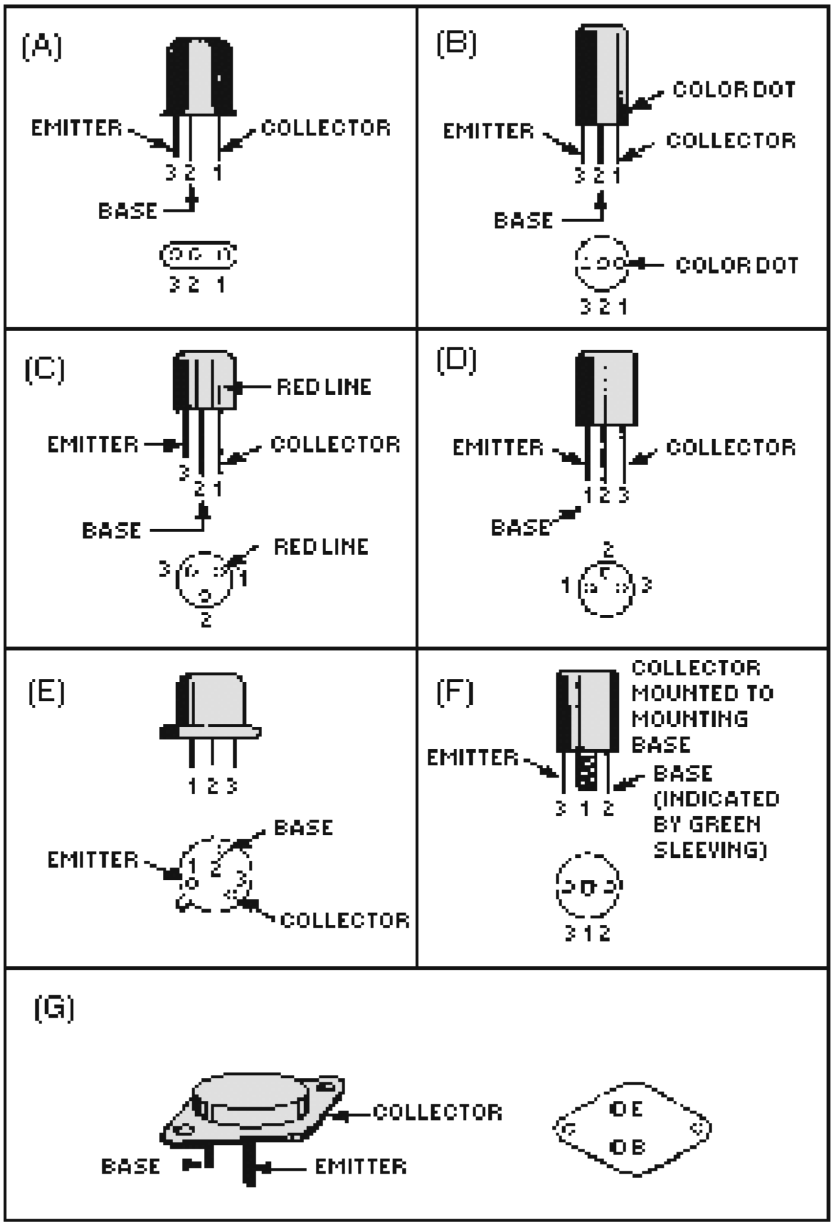


Figure 1-21.—Transistor lead identification and case outline.

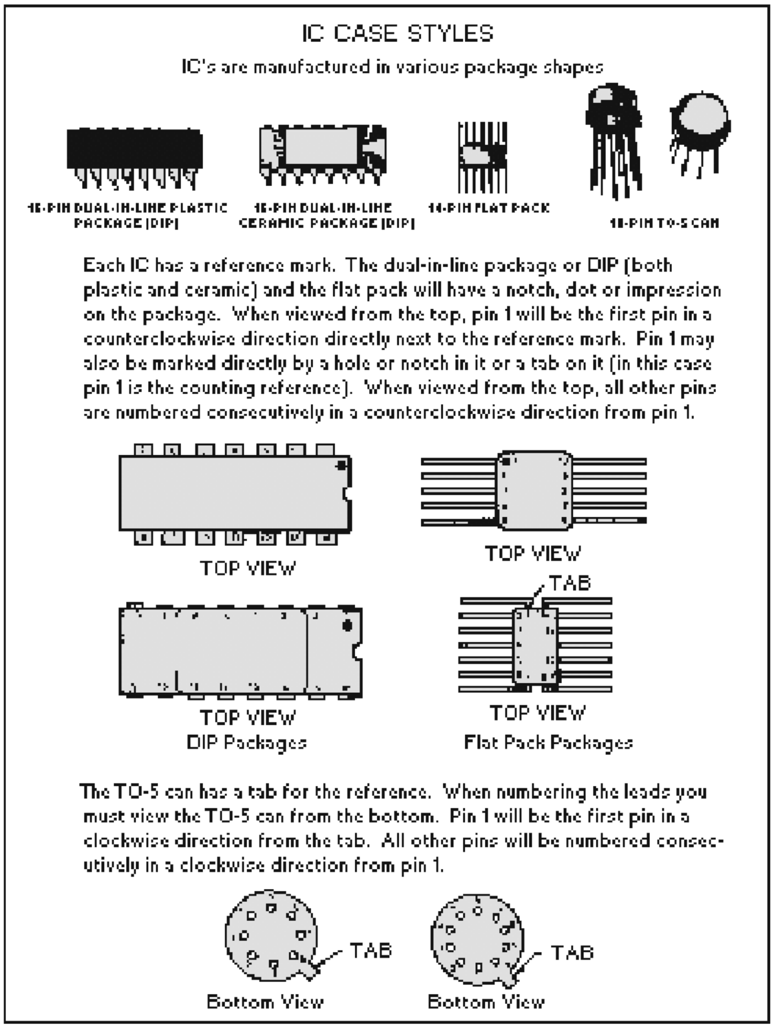


Figure 1-22.—IC identification and pin placement.

Batteries

The two fundamental types of batteries are the **PRIMARY CELL** and **SECONDARY CELL**. Primary cells are those commonly used in flashlights and some portable, hand-held test equipment. Common sizes and part numbers are:

SIZE	PART NUMBER
AA	BA58
C	BA42
D	BA30

Secondary cell batteries are the type used in automobiles. They are rechargeable.

Safety precautions concerning charging, handling, and storage of batteries can be found in the Electronics Installation and Maintenance Book (EIMB), *General*, NAVSEA SE000-00-EIM-100. Stock numbers and part numbers can be found in NAVSUP Publication 4400, the *Afloat Shopping Guide*. The federal supply classification (FSC) number for batteries is 6135. You can also find more information on batteries in NEETS, Module 1, *Introduction to Matter, Energy, and Direct Current*.

Cables

Tables 1-8 and 1-9 contain type, construction, and application data on shipboard cable. These tables contain current, discontinued, and some recently obsolete types of cables and cords.

Table 1-8.—Types and Construction/Description of Shipboard Cable

CVSF	400-Hz aircraft servicing: three synthetic rubber insulated conductors and one uninsulated conductor, overall polychloroprene jacket.
DLT	Divers lifeline and telephone: four rubber insulated conductors cabled around an insulated steel core, reinforced polychloroprene jacket overall.
DSS	Double conductor, shielded: rubber insulation, overall braided shield, polychloroprene or chlorosulfonated polyethylene jacket.
DSWS	Double conductor, shielded: rubber insulation, overall braided shield, polychloroprene jacket.
FSS	Four conductors, shielded: rubber insulated, overall braided shield, polychloroprene or chlorosulfonated polyethylene jacket.
JAS	Jet aircraft servicing: four rubber insulated conductors, two conductors Navy size 250, two conductors Navy size 6, reinforced polychloroprene jacket.
MCSF-4	Multiple conductor, acoustic minesweeping, power: two American Wire Gauge (AWG) 6 and two AWG 1 conductors, rubber insulation, reinforced polychloroprene jacket.
MSP	Multiple conductor: fifty-nine conductors, sixteen AWG 22 having fluorocarbon insulation and a braided copper shield, eighteen AWG 20 having polyvinyl chloride insulation and a braided copper shield (nine singles, one triad and three pairs, each shielded), twenty-five Navy size 3 having polyvinyl insulation (eight pairs and three triads, each shielded), polychloroprene jacket.
MSPW	Multiple conductor: fifty-nine conductors; sixteen AWG 22 having fluorocarbon insulation and a braided copper shield, eighteen AWG 20 having polyvinyl chloride insulation and a braided copper shield (nine singles, one triad and three pairs, each shielded), twenty-five Navy size 3 having polyvinyl insulation (eight pairs and three triads, each shielded), polychloroprene jacket, watertight.
MWF	Multiple conductor: rubber or cross-linked polyethylene insulation, arctic type neoprene jacket.
S2S	Two conductors, shielded: cross-linked polyethylene insulations, braided shield, rubber insulation over shield, outer-braided shield; reinforced rubber, insulated, arctic type polychloroprene jacket.
THOF	Three conductors, heat and oil resistant, flexible: synthetic rubber insulation standard thermoplastic jacket on THOF-42, and polychloroprene jacket on THOF-400 and THOF-500.
TRF	Single conductor, flexible: rubber insulation, polychloroprene jacket.
TPUM-6	Telephone, portable, multiple conductor: copper-clad steel conductors, polypropylene insulation, six pairs cabled, polyurethane jacket applied in two layers.
TRXF	Single conductor: polychloroprene jacket.
TSP	Twisted pairs: polyvinyl chloride insulated, special thermoplastic jacket, watertight, unarmored.
TSPA	Twisted pairs: polyvinyl chloride insulated, special thermoplastic jacket, watertight, armored.

Table 1-8.—Types and Construction/Description of Shipboard Cable—Continued

TSS	Three conductors, special purpose, shielded: rubber insulation, overall braided shield, polychloroprene or chlorosulfonated polyethylene jacket.
1SWF	Singles, shielded: polyethylene insulation, braided shield on each conductor, arctic type polychloroprene jacket.
2SWF	Pairs, shielded, watertight, flexible: polyethylene insulation, braided shield over each pair, arctic type polychloroprene jacket.
5SS	Five conductors, shielded, sonar: rubber insulation, braided shield on one conductor only, and a braided shield over the assembled five conductors, polychloroprene jacket overall.
7SS	Seven conductors, shieled: rubber insulation, overall braided shield, polychloroprene or chlorosulfonated polyethylene jacket.

Table 1-9.—Shipboard Cable Application Data

Application	Cable type 2 Non-flexing service	 Repeated flexing service
Outboard submersible: For hydrophones, transducers, outboard dial telephones, retractable antennae and similar equipment. Types MWF, 1SWF, and 2SWF are for hydrophones, transducers, and telephone lines in the weather. Types 1PR-A20E, 1PR-16, 7PR-16, 3PR-16, 1Q-16, ITR-16, and 7SPR-16S are only for submarine outboard use.	MSPW TSPA 1PR-A20E 1PR-16 7PR-16 2SPR-16 3PR-16 1Q-16 ITR-16 7SPR-16S	MSP, TSP, 5S5, S2S, DSS, FSS, TSS, MWF, DSWS, MCSF, 1SWF, 2SWF, TPUM
Welding electrode circuit		TRF TRXF
Shore-to-ship power		THOF-400 THOF-500
Diver's line and telephone		DLT
400-Hz aircraft servicing		CVSF-4
DC aircraft servicing		JAS-250
1/ The order of listing of cables for general application data has no significant meaning for their usage.		
2/ Many cables are manufactured in variations of armored, unarmored, and unarmored with overall shielding.		

Table 1-10 provides data on the allowable temperature ratings and current-carrying capacities (in amperes) of some single copper conductors in free air at a maximum ambient temperature of 86º F (30º C). With temperatures greater than 86º F, the current-carrying capacity would be less.

Table 1-10.—Current-Carrying Capacities (in Amperes) of Some Single Copper Conductors at 30º C

Size	Moisture Resistant Rubber or Thermoplastic	Varnished Cambric or Heat Resistant Thermoplastic	Silicone Rubber or Fluorinated Ethylene Propylene (FEP)	Polytetra-Fluoroethylene (Teflon)
0000	300	385	510	850
000	260	330	430	725
00	225	285	370	605
0	195	245	325	545
1	165	210	280	450
2	140	180	240	390
3	120	155	210	335
4	105	135	180	285
6	80	100	135	210
8	55	70	100	115
10	40	55	75	110
12	25	40	55	80
14	20	30	45	60

More information about electrical cable used aboard ship can be found in NEETS, Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading; Cable Comparison Guide*, NAVSEA 0981-052-8090; and *Design Data Book*, NAVSEA 0902-LP-006-0000, Section DDS-304-1. Cable supply information can be found in NAVSUP Publication 4400, *Afloat Shopping Guide*, under federal supply classification (FSC) 6145. Hook-up or chassis wire is covered in Military Specification 76B (MIL-W-76B). Table 1-11 shows the current-carrying capacity or AMPACITY of equipment hook-up wire.

Table 1-11.—Current-Carrying Capacity of Equipment Hook-up Wire—Continued

Wire Size		Copper Conductor (100°C) Nominal Resistance	Maximum Current in Amperes			
			Copper Wire		Aluminum Wire	
AWG	Circular Mils	(Ohms/1000 ft)	Wiring in Free Air	Wiring Confined	Wiring in Free Air	Wiring Confined
32	63.2	188.0	0.53	0.32		
30	100.5	116.0	0.86	0.52		
28	159.8	72.0	1.4	0.83		
26	254.1	45.2	2.2	1.3		
24	404.0	28.4	3.5	2.1		
22	642.4	22.0	7.0	5.0		
20	1022	13.7	11.0	7.5		
18	1624	6.50	16	10		
16	2583	5.15	22	13		
14	4107	3.20	32	17		
12	6530	2.02	41	23		
10	10 380	1.31	55	33		
8	16 510	0.734	73	46	60	38
6	26 250	0.459	101	60	83	50
4	41 740	0.290	135	80	108	66
2	66 370	0.185	181	100	152	82
1	93 690	0.151	211	125	174	105
0	105 000	0.117	245	150	202	123
00	133 100	0.092	283	175	235	145
000	167 800	0.074	328	200	266	162
0000	211 600	0.059	380	225	303	190

Table 1-12 lists the preferred general purpose rf cable selected by the armed services as the most satisfactory types to be used in electronics equipment.

Table 1-12.—Preferred General Purpose RF Cable

Jan Type	Overall Diameter (ins)	Impedance (OHMS)	Operating Voltage (Volts RMS)	Remarks
RG-11A/U	0.412	75.0	5,000	Medium size, flexible video cable
RG-12A/U	0.475	75.0	5,000	Same as RE-11A/U, armored
RG-34B/U	0.630	75.0	6,500	Large size, high power, low attenuation, flexible
RG-35B/U	0.945	75.0	10,000	Large size, high power, low attenuation, video and communications, armored
RG-55B/U	0.206	53.0	1,900	Small size, double braid
RG-58C/U	0.195	50.0	1,900	Small size, flexible
RG-59B/U	0.242	75.0	2,300	Small size, video
RG-84A/U	1.000	75.0	10,000	Same as RG-35B/U, except with lead sheath vice armor for underground installation
RG-85A/U	1.565	75.0	10,000	Same as RG-84A/U except with special armor
RG-164/U	0.870	75.0	10,000	Same as RG-35B/U except no armor
RG-212/U	0.332	50.0	3,000	Wave, formerly RG-5B/U
RG-213/U	0.405	50.0	5,000	Medium size, flexible, formerly RG-8A/U
RG-214/U	0.425	50.0	5,000	Medium size, double braid, flexible, formerly RG-9B/U
RG-215/U	0.475	50.0	15,000	Same as RG-214/U but armored. Formerly RG-10A/U
RG-216/U	0.425	75.0	5,000	Medium size, flexible video and communication, formerly RG-13A/U
RG-217/U	0.545	50.0	7,000	Medium size, power transmission line, formerly RG-14A/U
RG-218/U	0.870	50.0	11,000	Large size, low attenuation, high power transmission line, formerly RG-17A/U
RG-219/U	0.945	50.0	11,000	Same as RG-218/U, but armored, formerly RG-18A/U
RG-220/U	1.120	50.0	14,000	Very large, low attenuation, high power transmission line, formerly RG-19A/U
RG-221/U	1.195	50.0	14,000	Same as RG-220/U but armored, formerly RG-20A/U
RG-223/U	0.216	50.0	1,900	Small size, double braid, formerly RG-55A/U
RG-224/U	0.615	50.0	7,000	Same as RG-217/U but armored.

Specifications on special types of rf cable can be found in the *Military Standardization Handbook* 216 (MIL-HDBK-216).

Connectors

General purpose connectors and rf connectors are covered in this section. We'll cover the general purpose connectors first.

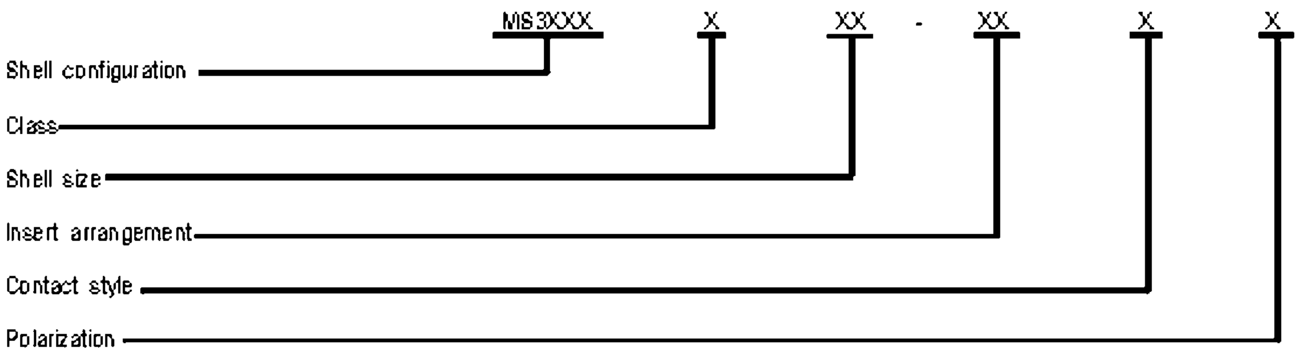
General purpose connectors were formerly designated with the prefix "**AN**." You may find older connectors with this prefix. The superseding connector has the same part number except the "AN" has been replaced by "**MS**." Table 1-13 shows the method used to break down a connector for identification. This breakdown is for MIL-C-5015 connectors. Identification breakdown for other MIL-C connectors can be found in *Naval Shore Electronics Criteria, Installation Standards and Practices*, NAVELEX 0280-LP-900-8000.

Table 1-13.—MS Connector Identification

MIL-STD-1353B
30 August 1983

MIL-C-5015 PART NUMBER SYSTEM

Part No. example: MS3400D22-22S



Shell configuration:	Front release crimp	Rear release crimp
Wall mounting receptacle	MS3400	MS3450
Box mounting receptacle	MS3402	MS3452
Jam nut mounting receptacle	MS3404	MS3454
Straight plug	MS3406	MS3456
Self-locking plug	-	MS3459
Cable connecting receptacle	MS3401	MS3451

Class:

D - Environment resisting - High impact shock MS3400
L - Environment resisting - Fluid resisitant (electroless N) For "space" applications only.
W - Environment resisting - Fluid resistant (Cad OD over suitable underplate)
KS - Firewall, self-locking, stainless steel
KT - Firewall, self-locking, cadmium plated ferrous alloy

(L, W, KS: MS3450)

Shell size: Shell size in 16th of an inch.

Insert arrangements: See MIL-C-STD-1651.

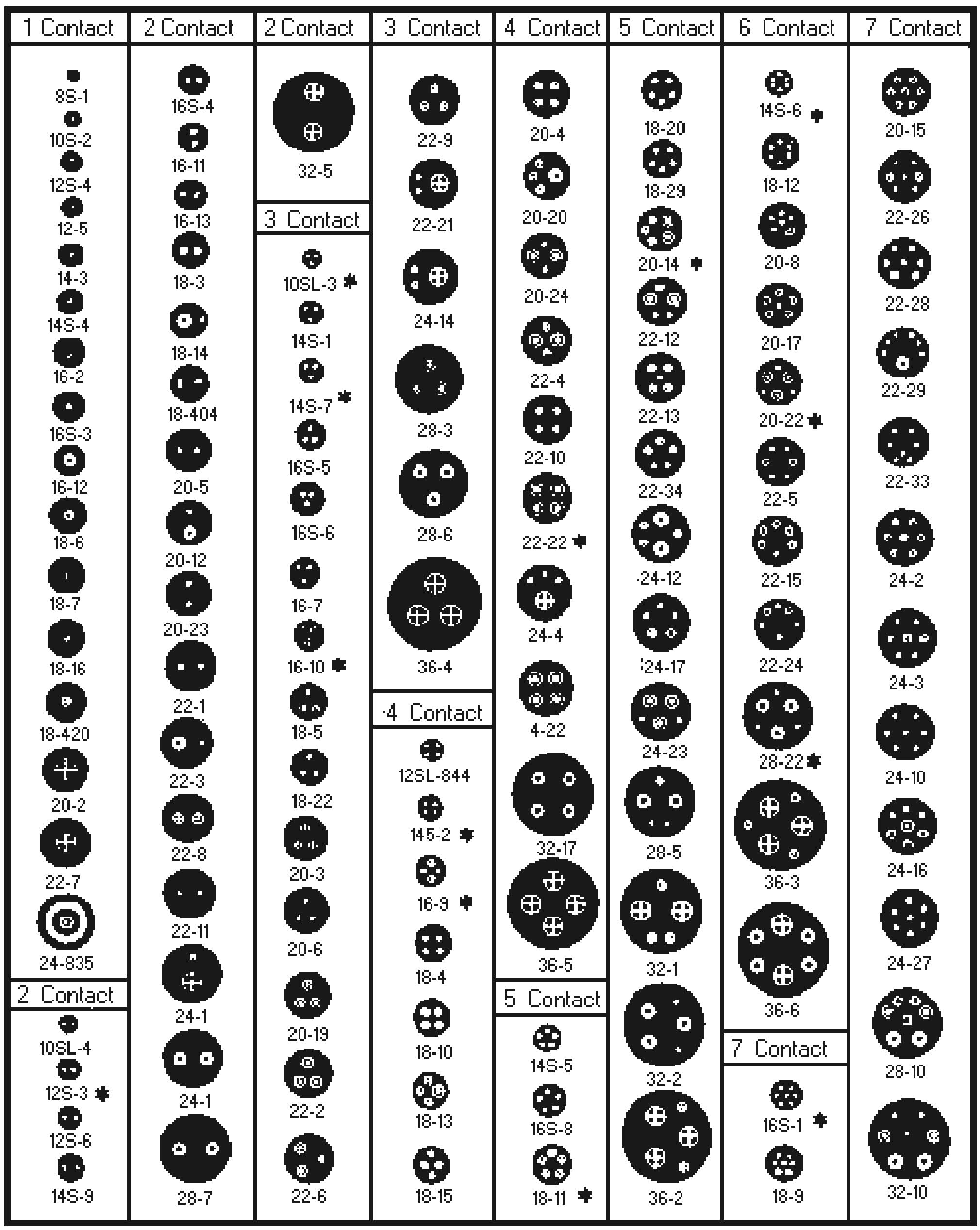
Contact style:

P - Pin contact - MIL-C-39029/29 and /44.
S - Socket contact - MIL-C-39029/30 and /45.

Polarization: Normal polarization is considered preferred; however, alternate polarizations, when required by a system, do not require nonstandard part approval.

Supply information on connectors can be found in NAVSUP Publication 4400, *Afloat Shopping Guide*, under federal supply classification (FSC) 5935.

Insert arrangements for MS type connectors, MIL-C-5015, are shown in figure 1-23. Alternate positions of connector inserts are shown in figure 1-24.



✱ INDICATES AIR STANDARDS COORDINATING COMMITTEE PREFERRED.
NOTE: FACE VIEW OF PIN INSERTS. ALL INSERTS SHOWN IN NORMAL POSITION.

Figure 1-23.—Insert arrangements type connectors, MIL-C-5015.

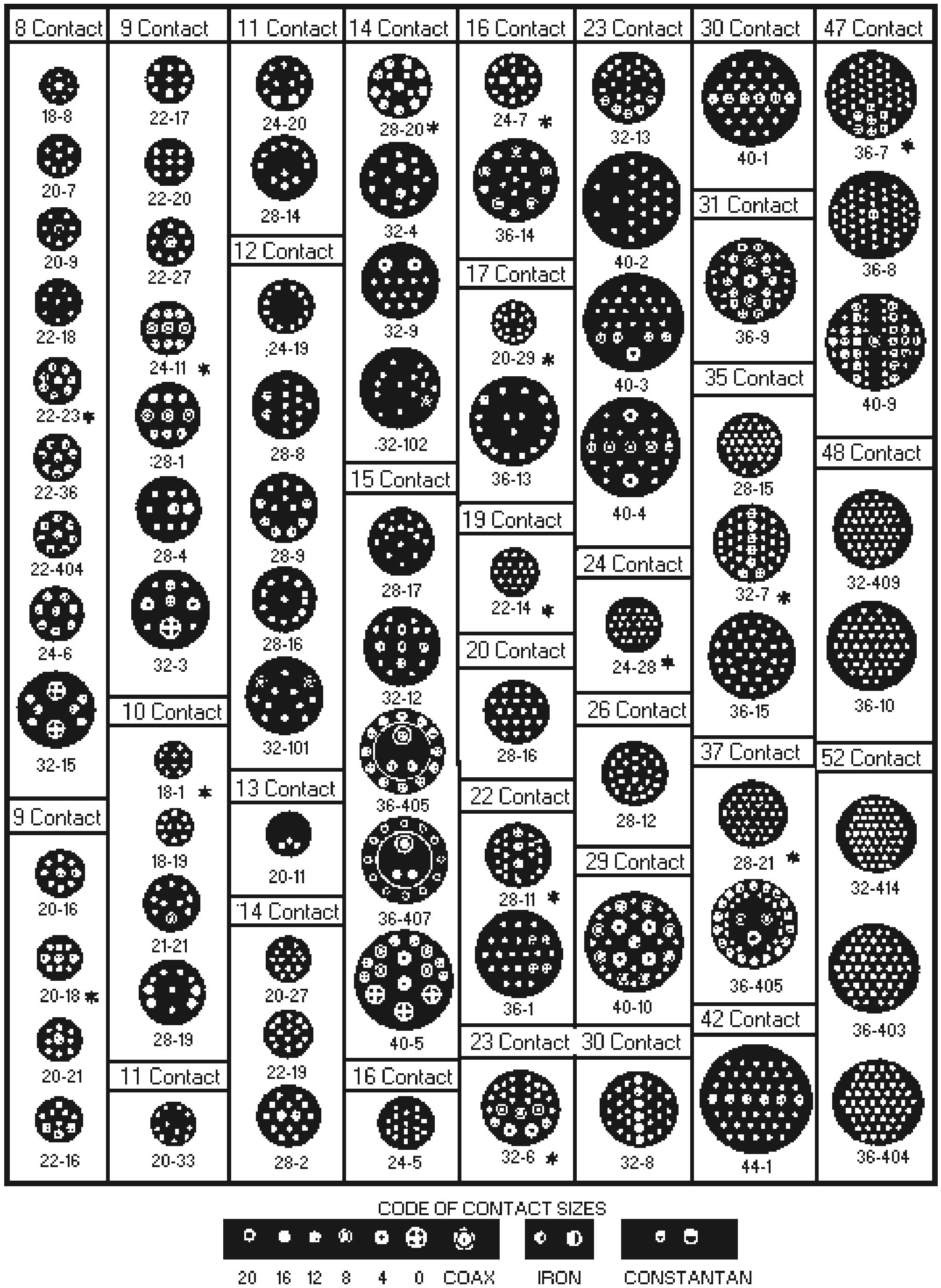


Figure 1-23.—Insert arrangements type connectors, MIL-C-5015. —Continued.

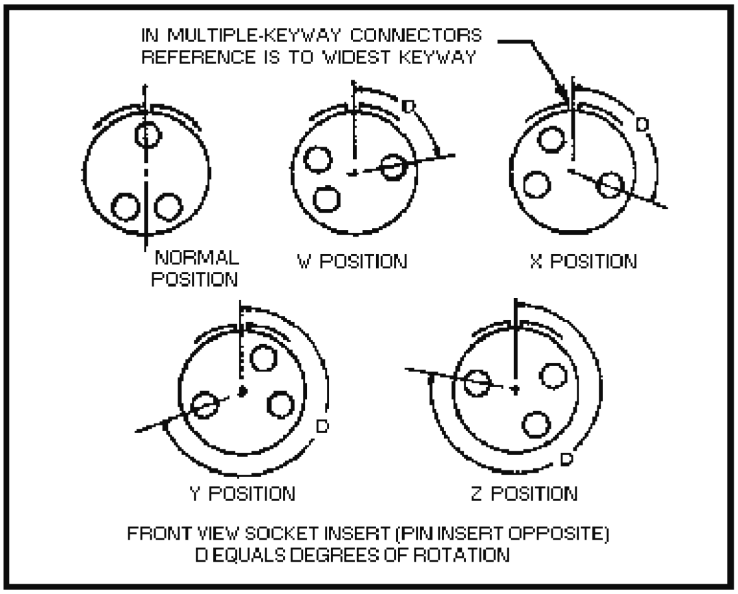


Figure 1-24.—Alternate positions of connector inserts.

MIL-C-5015, MIL-C-26482, and MIL-C-26500 connectors are designated with an MS number such as MS3101. MIL-C-81511, MIL-C-83723, and other later specifications identify the connector by the specification number, a slash, and the connector number. For example: MIL-C-81511/3.

Rf connectors and coaxial cable assemblies are used to carry radio frequency (rf) power from one point to another with a known rate of loss. Rf connectors are available as plugs, jacks, panel jacks, and receptacles. Plugs and jacks are attached to the ends of coaxial cables; panel jacks and receptacles are mounted to panels and chassis.

Baby **N** connector (**BNC**) series connectors are small, lightweight, and feature a quick connect/disconnect, bayonet-lock coupling. The connectors use small rf cables such as **RG-58/U** or **RG-59/U** and operate up to peak voltages of 500 volts. Manufacture is under Military Specification C-3608A (MIL-C-3608A). Figure 1-25 shows three typical **BNC** connectors. Figure 1-26 shows you how to attach **BNC** connectors to coaxial cable. Table 1-14 indicates which **BNC** connector to use with what coaxial cable type.

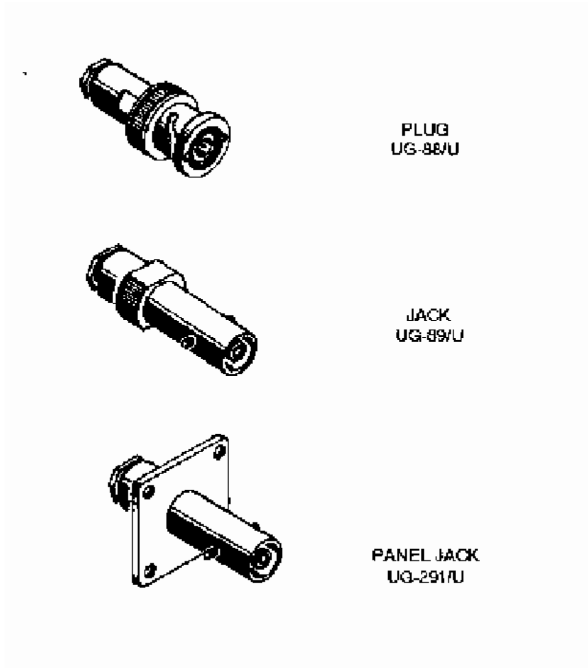


Figure 1-25.—Typical BNC connectors.

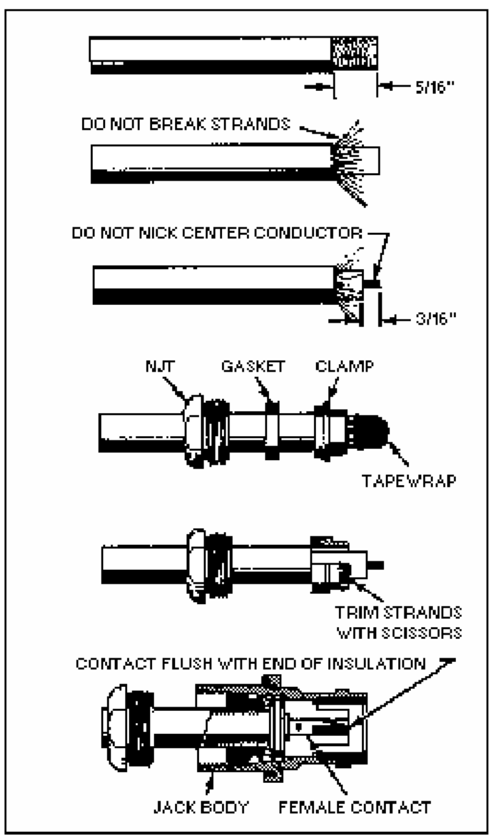


Figure 1-26.—Attaching BNC connectors to coaxial cable.

Table 1-14.—BNC Series Connectors with Associated Cables

Plug	Jack	Panel Jack	For Use With Cable Types
Improved Version:			
UG-88E/U	UG-89C/U	UG-291C/U	RG-55/U, 58/U and 223/U
UG-260D/U	UG-261C/U	UG-262C/U	RG-59/U, 62/U and 71/U
Captivated Contact Version (Amphenol):			
31-301	31-302	31-300	RG-55/U, 58/U, 141/U and 142/U
31-304	31-305	31-303	Rg-59/U, 62/U, 71/U and 140/U

HN series connectors have a 50-ohm impedance and threaded coupling connectors designed for high-voltage applications. These connectors are weatherproof. The frequency range is 0-4 gigahertz. The **HN** series is used with medium size coaxial cable such as **RG-8/U, RG-9/U, RG-87/U, RG-213/U, RG-214/U,** and **RG-225/U**. Figure 1-27 shows three typical **HN** connectors. Figure 1-28 shows how **HN** connectors are attached to coaxial cable. Table 1-15 indicates which HN connector to use with what coaxial cable type.

N series connectors are low-voltage, 50-ohm, threaded coupling connectors designed for use with small and medium size rf cable. They have a 1,000 volts peak rating and are weatherproof. There is a group of **N** series connectors that are 70 ohms and are numbered **UG-98A/U** and **UG-96A/U**. These 70-ohm **N** connectors are designed for cables such as **RG-61C**, **RG-11/U**, and **RG-13/U**. The 70-ohm connectors will not mate with 50-ohm connectors of this series. Figure 1-29 shows three typical **N** series connectors. Figure 1-30 illustrates the method used to attach **N** connectors to coaxial cable. Table 1-16 shows which **N** connector to use with what rf cable type.

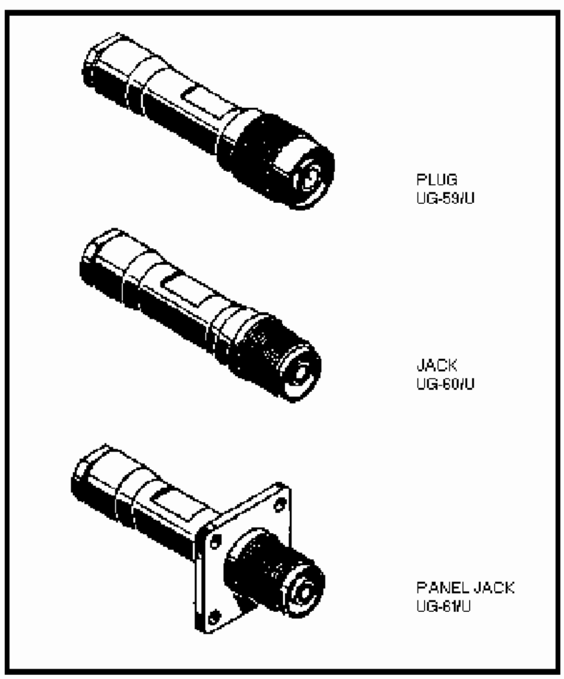


Figure 1-27.—Typical HN connectors.

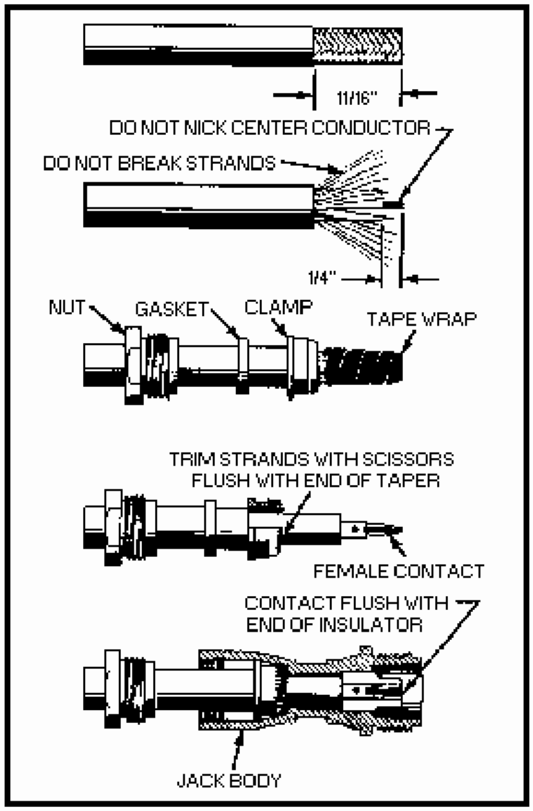


Figure 1-28.—Attaching HN connectors to coaxial cable.

Table 1-15.—HN Series Connectors with Associated Cables

Plug	Jack	Panel Jack	For Use With Cable Types
*UG-59E/U **UG-1213/U	*UG-60E/U **UG-1214/U	*UG-61E/U **UG-1215/U	RG8/U, 9/U, 213/U and 214/U
Improved Version; **Captivated Contact Version			

C series connectors are weatherproof, quick-connect/disconnect, bayonet-locking type connectors. They are used with medium size cables, such as **RG-5/U, RG-8/U**, and **RG-9/U**. They operate up to a peak of 1,000 volts and at frequencies up to 10 gigahertz. Their impedance is 50 ohms. Figure 1-31 shows three typical C series connectors. Figure 1-32 shows you how to attach the C series connectors to a coaxial cable.

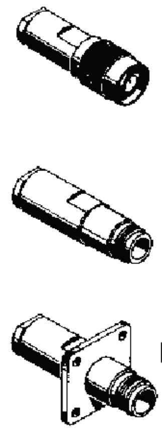


Figure 1-29.—Typical N connectors.

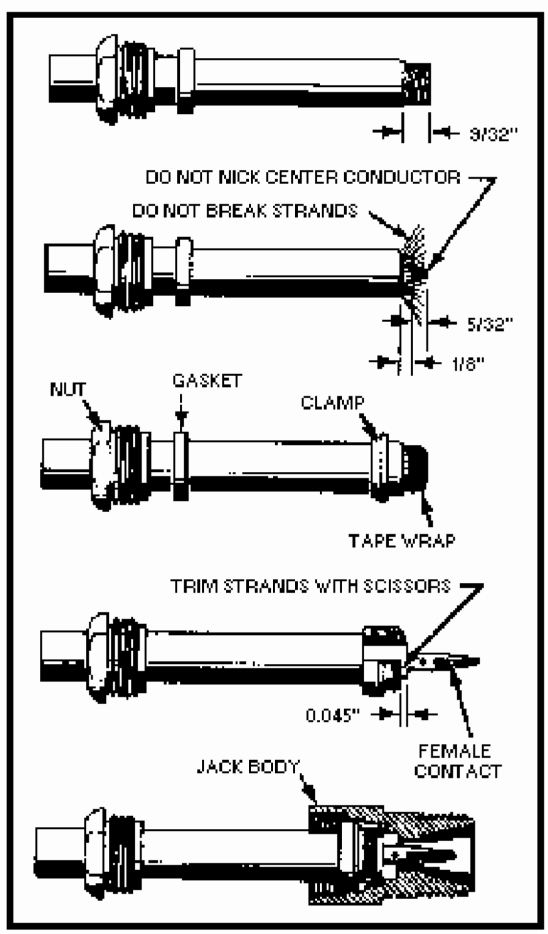


Figure 1-30.—Attaching N connectors to coaxial cable.

Table 1-16.—N Series Connectors with Associated Cables

Plug	Jack	Panel Jack	For Use With Cable Types
Improved Version:			
UG-18D/U	UG-20D/U	UG-19D/U	RG-5/U, 6/U, 21/U and 212/U
UG-21E/U	UG-23E/U	UG-23E/U	RG-8/U, 9/U, 213/U and 214/U
UG-594A/U		UG-160D/U	
UG-536B/U			RG-55/U and 58/U
Captivated Contact Version:			
UG-1185/U	UG-1186/U	UG-1187/U	RG-8/U,9/U,213/ U AND 214/U

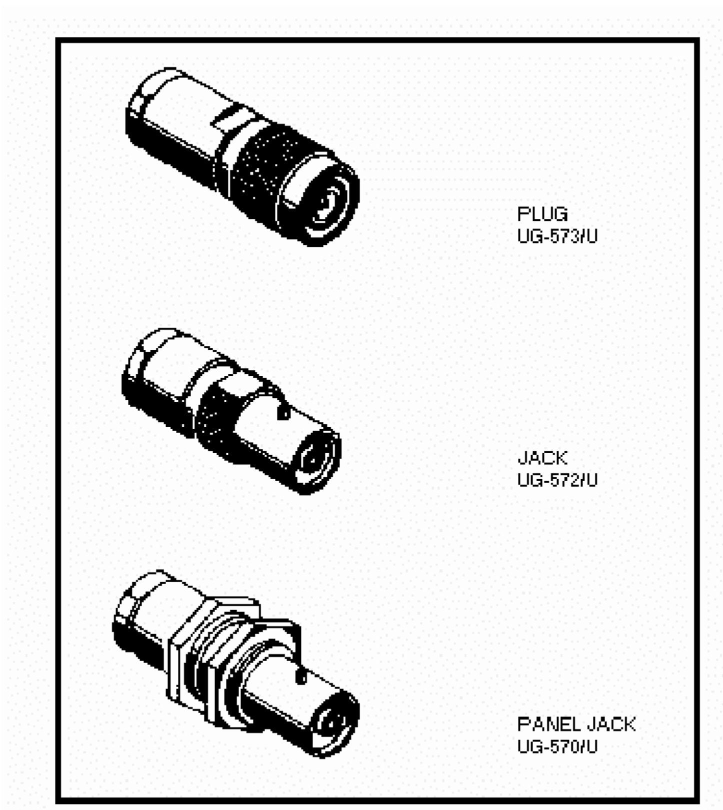


Figure 1-31.—Typical C connectors.

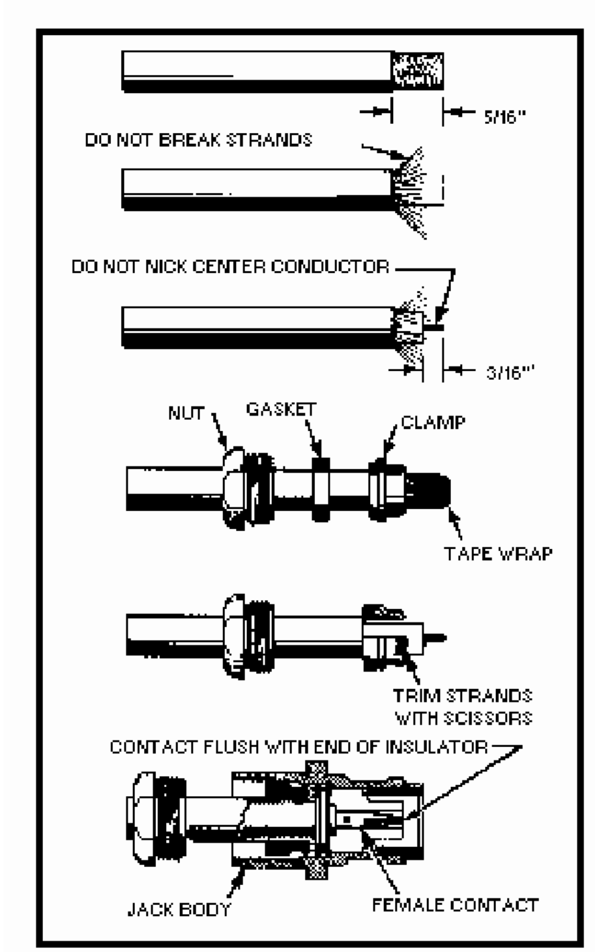


Figure 1-32.—Attaching C connectors to coaxial cable.

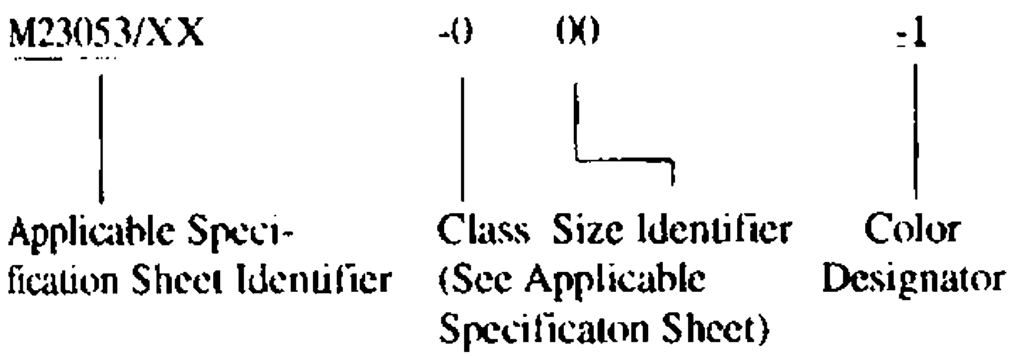
Table 1-17.—C Series Connectors with Associated Cables

Plug	Jack	Panel Jack	For Use With Cable Types
UG-573B/U	UG-572A/U	UG-570A/U	RG-8/U, 9/U, 213/U and 214/U
UG-701B/U		UG-571A/U	
UG-626B/U	UG-633A/U	UG-629A/U	RG-5/U, 6/U, and 212/U
		UG-630A/U	
UG-707A/U			
			RG-14/U and 217/U

You can find more information on rf connectors in *Military Handbook* 216 (MIL-HDBK-216).

Insulation, Heat Shrinkable Sleeving (Shrink, Tubing), and Cable Straps

Heat shrinkable sleeving is intended for use as a snug-fitting electrical insulator. It is used to insulate wire bundles, splices, bus bars, connectors, terminals, metal, or fibrous tubing. It is also used as extra insulation over hotspot areas and as a cable blast shield in rocket launchings. Heat shrinkable sleeving is found under Military Specification I-23053C (MIL-1-23053C). Part numbers under this military specification are coded as follows:



Color code designations are:

DESIGNATOR	COLOR	DESIGNATOR	COLOR
0	Black	7	Violet (Purple)
1	Brown	8	Gray (Slate)
2	Red	9	White
3	Orange	C	Clear
4	Yellow	P	Pink
5	Green	T	Tan
6	Blue		

The particular uses for heat shrinkable sleeving depend on the specific properties described by the individual specification sheet. Intended uses are indicated below:

Military
Part Number

M23053/1:	Intended for use on heavy duty cables or harness systems such as ground support.
/2 and /3:	Used for light-duty harnessing or wire bundling.
/4:	Used for one-step potting, encapsulation, or moisture sealing and corrosion protection of electrical components or terminations.
/5:	Used for light-duty harness jackets, wire color coding, marking, or identification.
/6:	Used for wire identification, marking, or strain relief.
/7:	Used for light-duty wire identification and component covering.
/8:	Used for wire or termination strain relief at elevated temperatures.
/9:	Canceled.
/10:	Used for high-or low-temperature applications or where resistance to melting in high-blast flame is required.
/11:	Used where strain relief is necessary at high temperatures.
/12:	Used at high temperatures where resistance to flame is important to protect high-temperature cable, components, and terminations.
/13:	Used in elevated-temperature applications or where exposure to elevated-temperature solvents is expected.
/14:	Used as component and electronic lead strain relief where low expansion ratios are satisfactory. Operates over a fairly wide temperature range.
/15:	Used for repair of heavy duty cables and splice covers.
/16	Used on heavy duty cables or harness systems that are subjected to high levels of physical abuse and exposure to fuels and oils coupled with high-and low- temperature extremes.

Tables 1-18 and 1-19 provide part numbers for two types of shrinkable tubing, **M23053/5** and **M23053/4**. These were chosen because of their wide range of sizes **(M23053/5)** and their abilities to provide potting, encapsulating, or moisture sealing of electrical components **(M23053/4)**.

Table 1-18.—Shrinkable Tubing Part Numbers

Military Part Number	As supplied	After unrestricted shrinkage
	I.D. min.	I. D. max.
Class 1	Inches	Inches
M23053/5-101-*	.046	.023
M23053/5-102-*	.063	.031
M23053/5-103-*	.093	.046
M23053/5-104-*	.125	.062
M23053/5-105-*	.187	.093
M23053/5-106-*	.250	.125
M23053/5-107-*	.375	.187
M23053/5-108-*	.500	.250
M23053/5-109-*	.750	.375
M23053/5-110-*	1.000	.500
M23053/5-111-*	1.500	.750
M23053/5-112-*	2.000	1.000
M23053/5-113-*	3.000	1.500
M23053/5-114-*	4.000	2.000

* Asterisk in the part number shall be replaced by color code designations.

Table 1-19.—Shrinkable Tubing Part Numbers

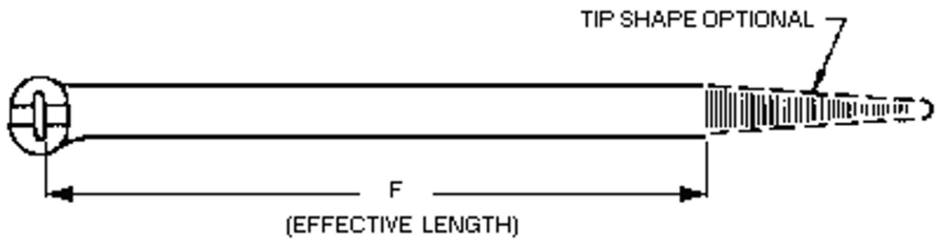
Military Part Number	As supplied		After unrestricted shrinkage	
	I.D. min.		I.D. max.	
Class 1		Inches		Inches
M23053/4-101-*	.125		.023	
M23053/4-102-*	.187		.060	
M23053/4-103-*	.250		.080	
M23053/4-104-*	.375		.135	
M23053/4-105-*	.500		.195	
M23053/4-106-*	.750		.313	
M23053/4-107-*	1.000		.400	
M23053/4-108-*	.300		.050	
Class 2				
M23053/4-201-*	.238		.125	
M23053/4-202-*	.355		.187	
M23053/4-203-*	.475		.250	
M23053/4-204-*	.712		.375	
M23053/4-205-*	.950		.500	
M23053/4-206-*	1.425		.750	
Class 1 = Semi-rigid, flame retardant Class 2 = Flexible, flame retardant * Asterisk in the part number shall be replaced by color code designations.				

When ordering, you should replace the asterisk (*) in the part number with the color code designation.

For example: The part number for **M23053/4**, one-half inch, class 1, black shrinkable tubing would be **M23053/4-105-0**.

Table 1-20 provides data on adjustable nylon cable straps. These straps are adjustable only in one direction. They are not designed to be loosened.

Table 1-20.—Adjustable Nylon Cable Strap Data



Part Number	Bundle diameter Min	Bundle diameter Max	Tensile strength (pounds) (ref)	MS installation tool	MIN F LENGTH
MS3367-1	1/16	1-3/4	50	MS90387-1	6.000 (152.4)
MS3367-2	1/16	4	50	MS90387-1	13.00 (330.20)
MS3367-3	3/16	3-1/2	120	MS90387-2	11.88 (301.75)
MS3367-4	1/16	5/8	18	MS90387-1	2.50 (63.50)
MS3367-5	1/16	1-1/4	30	MS90387-1	4.45 (113.03)
MS3367-6	3/16	8	120	MS90387-2	26.00 (660.40)
MS3367-7	1/16	3	50	MS90387-1	9.950 (252.73)

Color dash no.	Color	Color dash no.	Color
-0(a)	Black	-5	Green
-1	Brown	-6	Blue
-2	Red	-7	Purple
-3	Orange	-8	Gray
-4	Yellow	-9(a)	Natural

(a) ONLY STRAPS IN NATURAL AND BLACK ARE STOCKED BY THE GOVERNMENT.

Fuses and Circuit Breakers

New type military fuse designations can be identified by using table 1-21. Old style military fuse designations can be identified by using table 1-22.

Table 1-21.—New Style Military Fuse Identification

Example: Fuse Part Number = F 02 A 250V 1A S

F -Military designation which stands for fuse

02 -Style - indicates dimensions and construction

A - Characteristic (Time Delay)

A = Standard
B = Delay
C = Fast

250V - Voltage rating. Numerical value of maximum voltage followed by the letter "V".

1A - Current rating - Nominal current in amperes followed by the letter "A".

S - Platings

s = silver
No identification for other platings

Table 1-22.—Old Style Military Fuse Identification

Example: Fuse Part Number F 02 G 1R00 A

F Stands for fuse

02 Style, construction, and dimensions.

G Voltage Rating - Letter value of maximum voltage taken from table.

Voltage Code Letter	Voltage Volts or less
A	32
B	52
C	90
D	125
G	250
H	500
J	1,000
K	2,500
N	5,000
P	10,000

1R00 Current Rating - Nominal current in current in amperes. R = decimal point. Example given is 1 amp

A Characteristic - Time Delay rating

A = Standard
B = Delay
C = Fast

Commercial fuse identification and a fuse cross-reference can be found in NEETS, Module 3, *Introduction to Circuit Protection, Control, and Measurement*; and in Military Standard 1360A (MIL-STD-1360A). These will assist you in selecting or identifying fuses.

Circuit breakers are too numerous to cover in this text. They are used in houses, vehicles, ships, and airplanes. Military Standard 1498 (MIL-STD-1498) contains information to help you select or identify circuit breakers.

Classification of Rf Emissions

The system of designating rf emissions is arranged according to modulation type, mode, and supplementary characteristics. For example: **A3B** indicates amplitude modulation, telephony, two independent sidebands, and a suppressed carrier. Table 1-23 will assist you in breaking down the emission classification code.

Table 1-23.—Emission Types

Emission	Type
Modulation Types	
Amplitude	A
Frequency	F
Pulse	P
Modulation (Transmission Mode)	
None	0
Telegraphy (keyed r-f carrier)	1
Telegraphy (tone)	2
Telephony	3
Facsimile	4
Television	5
Four Channel Diplex Telegraphy	6
Multichannel Voice Frequency Telegraphy Complex Forms	7
Supplemental Characteristics	
Double Sideband	none
Single Sideband	
-reduced carrier	A
-full carrier	H
-suppressed carrier	J
Two Independent Sidebands	
-suppressed carrier	B
Vestigial Sideband Pulse	C
-amplitude modulated	D
-width modulated	E
-phase modulated	F
-code modulated	G

*Capital or lower case letter

**Commercial practice is to reduce carrier 20 dB, to provide sufficient carrier for receiver afc lock-in, where afs receivers are used.

Note: a number preceding the emission designation indicates the bandwidth in kilohertz.

Conversion and Equivalent Tables

Table 1-24 provides the multiplying factors necessary to convert from one unit of measure to another and vice versa.

Table 1-24.—Conversion Chart

To Convert	To	Multiply By	Conversely, Multiply By
Acres	Square feet	4.356 x 10^{4}	2.296 " 10^{-5}
Acres	Square meters	4047	2.471 " 10^{-4}
Ampere-hour	Coulombs	3600	2.778 " 10^{-4}
Amperes	Microamperes	1,000,000	0.000,001
Amperes	Milliamperes	1,000	0.001
Amperes per sq cm	Amperes per sq in	6.452	0.1550
Amperes-turns	Gilberts	1.257	0.7958
Amperes-turns per cm	Amperes-turns per inch	2.540	0.3937
Amperes-turns per cm	Oersteds	1.257	0.7958
Ampere-turns per in	Oersteds	0.495	2.02
Ampere-turns per meter	Oersteds	.01257	79.58
Ampere-turns per weber	Gilberts per maxwell	1.257 " 10^{-8}	7.958 " 10^{7}
Atmospheres	MM of mercury at 0º C	760	1.316 " 10^{-8}
Atmospheres	Feet of water at 4º C	33.90	2.950 " 10^{-2}
Atmospheres	Inches of mercury at 0º C	29.92	3.342 " 10^{-2}
Atmospheres	Kilograms per sq meter	1.033 " 10^{4}	9.678 " 10^{-5}
Atmospheres	Pounds per sq inch	14.70	6.804 " 10^{-2}
BTU	Foot-pounds	778.3	1.285 " 10^{-3}
BTU	Joules	1054.8	9.480 " 10^{-4}
BTU	Kilogram-calories	0.2520	3.969
BTU per hour	Horsepower-hours	3.929 " 10^{-4}	2545
Bushels	Cubic feet	1.2445	0.8036
Celsius (Centigrade) deg	Fahrenheit deg	(º C " 9/5) + 32	(º F - 32) " 5/9
Circular Mils	Sq centimeters	5.067 " 10^{-6}	1.973 " 10^{5}
Circular mils	Square mils	0.7854	1.273
Cubic feet	Cords	7.8125 " 10^{-3}	128
Cubic feet	Gallons (liquid US)	7.481	0.1337
Cubic feet	Liters	28.32	3.531 " 10^{-2}
Cubic inches	Cubic centimeters	16.39	6.102 " 10^{-2}
Cubic inches	Cubic feet	5.787 " 10^{-4}	1728
Cubic inches	Cubic meters	1.639 " 10^{-5}	6.102 " 10^{-4}
Cubic inches	Gallons (liquid US)	4.329 " 10^{-3}	231
Cubic meters	Cubic feet	35.31	2.832 " 10^{-2}
Cubic meters	Cubic yards	1.381	0.7646
Degrees (angle)	Radians	1.745 " 10^{-2}	57.30
Dynes	Pounds	2.248 " 10^{-6}	4.448 " 10^{5}
Ergs	Foot-pounds	7.367 " 10^{-8}	1.356 " 10^{7}
Fards	Microtarads	1,000,000	0.000,001
Farads	Picofarads	1,000,000,000,000	0.000,000,000,001
Fathoms	Feet	6	0.16666

Table 1-24.—Conversion Chart—Continued

To Convert	To	Multiply By	Conversely, Multiply By
Feet	Centimeters	30.48	3.281 " 10^{-2}
Feet	Varas	0.3594	2.782
Feet of water at 4º C	Inches of mercury at 0º C	0.8826	1.133
Feet of water at 4º C	Kilograms per sq meter	304.8	3.281 " 10^{-3}
Feet of water at 4º C	Pounds per sq foot	62.43	1.602 " 10^{-2}
Foot-pounds	Horsepower-hours	5.050 " 10^{-7}	1.98 " 10^{-6}
Foot-pounds	Kilogram-meters	0.1383	7.233
Foot-pounds	Kilowatt-hours	3.766 " 10^{-7}	2.655 " 10^{6}
Gallons	Cubic meters	3.785 " 10^{-8}	264.2
Gallons (liquid US)	Gallons (liquid Br Imp)	0.8327	1.201
Gausses	Lines per sq inch	6.452	0.1550
Gilberts per cm	Oersteds	1	1
Grains (for humidity calculations)	Pounds (avoirdupois)	1.429 " 10^{-4}	7000
Grams	Dynes	980.7	1.020 " 10^{-3}
Grams	Grains	15.43	6.481 " 10^{-2}
Grams	Ounces (avoirdupois)	3.527 " 10^{-2}	28.35
Grams	Poundals	7.093 " 10^{-2}	14.10
Grams per cm	Pounds per inch	5.600 " 10^{-3}	178.6
Grams per cu cm	Pounds per cu in	3.613 " 10^{-3}	27.68
Grams per sq cm	Pounds per sq ft	2.0481	0.4883
Hectares	Acres	2.471	0.4047
Henrys	Microhenrys	1,000,000	0.000,001
Henrys	Millihenrys	1000	0.001
Henrys per meter	Gausses per Oersted	7.958 " 10^{5}	1.257 " 10^{-6}
Horsepower (boiler)	BTU per hour	3.347 " 10^{4}	2.986 " 10^{-5}
Horsepower (metric) (542.5 ft-lb per sec)	BTU per minute	41.83	2.390 " 10^{-2}
Horsepower (metric) (542.5 ft-lb per sec)	Ft-lb per minute	3.255 " 10^{4}	3.072 " 10^{-5}
Horsepower (metric) (542.5 ft-lb per sec)	Kilogram-calories per min	10.54	9.485 " 10^{-2}
Horsepower (550 ft-lb per sec)	BTU per minute	42.41	2.357 " 10^{-2}
Horsepower (550 ft-lb per sec)	Ft-lb per minute	3.3 " 10^{4}	3.030 " 10^{-5}
Horsepower (550 ft-lb per sec)	Kilowatts	0.745	1.342
Horsepower (550 ft-lb per sec)	Watts	746	1.342 " 10^{-3}
Horsepower (metric) (542.5 ft-lb per sec)	Horsepower (550 ft-lb per sec)	0.9863	1.014
Horsepower (550 ft-lb per sec)	Kilogram-calories per min	10.69	9.355 " 10^{-2}

Table 1-24.—Conversion Chart—Continued

To Convert	To	Multiply By	Conversely, Multiply By
Inches	Centimeters	2.540	3.3937
Inches	Feet	8.33 " 10^{-2}	12
Inches	Miles	1.578 " 10^{-5}	6.336 " 10^{4}
Inches	Mil	1000	0.001
Inches	Yards	2.778 " 10^{-2}	36
Inches of mercury at 0º C	Pounds per sq inch	0.4912	2.036
Inches of water at 4º C	Kilograms per sq meter	25.40	3.937 " 10^{-2}
Inches of water at 4º C	Ounces per sq inch	0.5782	1.729
Inches of water at 4º C	Pounds per sq foot	5.202	0.1922
Inches of water at 4º C	Inches of mercury	7.355 " 10^{-2}	13.60
Joules	Foot-pounds	0.7376	1.356
Joules	Ergs	10 7	10^{-3}
Kilogram-calories	Kilogram-meters	426.9	2.343 " 10^{3}
Kilogram-calories	Kilojoules	4.186	0.2389
Kilograms	Tons, long (avdp 2240 lb)	9.842 " 10^{-4}	1016
Kilograms	Tons, short (avdp 2000 lb)	1.102 " 10^{-8}	907.2
Kilograms	Pounds (avoirdupois)	2.205	0.4536
Kilograms per sq meter	Pounds per sq foot	0.2048	4.882
Kilometers	Feet	3281	3.048 " 10^{-4}
Kilovolts	Volts	1000	0.001
Kilowatt-hours	BTU	3413	2.930 " 10^{-4}
Kilowatt-hours	Foot-pounds	2.655 " 10^{6}	3.766 " 10^{-7}
Kilowatt-hours	Joules	3.6 " 10^{7}	2.778 " 10^{-7}
Kilowatt-hours	Kilogram-calories	860	1.163 " 10^{-3}
Kilowatt-hours	Kilogram-meters	3.671 " 10^{5}	2.724 " 10^{-7}
Kilowatt-hours	Pounds carbon oxidized	0.235	4.26
Kilowatt-hours	Pounds water evaporated from and at 212º F	3.53	0.283
Kilowatt-hours	Pounds water raised from 32º to 212º F	24.52	4.078 " 10^{-2}
Kilowatts	Watts	1000	0.001
Leagues	Miles	2.635	0.3795
Lines per inch 2	Gausses	0.1550	6.452
Liters	Bushels (dry, US)	2.838 " 10^{-2}	35.24
Liters	Cubic Centimeters	1000	0.001
Liters	Cubic meters	0.001	1000
Liters	Cubic inches	61.02	1.639 " 10^{-2}
Liters	Gallons (liq US)	0.2642	3.785
Liters	Pints (liq US)	2.113	0.4732
Log N	Log 10N	0.4343	2.303
Lumens per sq ft	Foot-candles	1	1
Lux	Foot-candles	0.0929	10.764
Maxwells	Lines	1	1
Maxwells	Webers	10^{-2}	10^{8}

Table 1-24.—Conversion Chart—Continued

To Convert	To	Multiply By	Conversely, Multiply By
Maxwells per cm2	Gausses	1	1
Meters	Yards	1.094	0.9144
Meters	Varas	1.179	0.848
Meters per min	*Knots (naut mi per hour)	30.866	30.866
Meters per min	Feet per minute	3.281	0.3048
Meters per min	Kilometers per hour	0.06	16.67
Mhos	Micromhos	1,000,000	0.000,001
Microhms per cm cube	Microhms per inch cubs	0.3937	2.540
Miles (nautical	Feet	6076.103	1.645 " 10^{-4}
Miles (nautical)	Kilometers	1.852	0.5396
Miles (statute)	Kilometers	1.609	0.6214
Miles (statute)	*Miles (nautical)	0.8688	1.151
Miles (statute)	Fast	5280	1.894 " 10^{-4}
Miles par hour	Kilometers per mi	2.682 " 10^{-2}	37.28
Miles per hour	Feet per minute	88	1.136 " 10^{-2}
Miles par hour	*Knots (naut mi per hr)	0.8688	1.151
Miles per hour	Kilometers per hour	1.609	0.6214
Nepers	Decibels	8.686	0.1151
Picofarads	Microfarads	0.000,001	1,000,000
Pounds of water (dist)	Cubic feet	1.603 " 10^{-2}	62.38
Pounds of water (dist)	Gallons	0.1198	8.347
Pounds per cu ft	Kilograms per cu meter	16.02	6.243 " 10^{-2}
Pounds per cu inch	Pounds per cu foot	1728	5.787 " 10^{-4}
Pounds per sq ft	Pounds per sq inch	6.944 " 10^{-3}	144
Pounds per sq in	Kilograms per sq meter	703.1	1.442 " 10^{-3}
Poundals	Dynes	1.383 " 10^{4}	7.233 " 10^{-5}
Poundals	Pounds (avoirdupois)	3.108 x 10^{-2}	32.17
Radians	Mils	10^{3}	10^{-3}
Radians	Minutes	3.438 x 10^{3}	2.909 " 10^{-4}
Radians	Seconds	2.06265 " 10^{5}	4.848 " 10^{-6}
Slugs	Pounds	32.174	3.108 " 10^{-2}
Sq inches	Circular mils	1.273 " 10^{6}	7.854 " 10^{-7}
Sq inches	Sq centimeters	6.452	0.1550
Sq feet	Sq meters	9.290 " 10^{-2}	10.76
Sq miles	Sq yards	3.098 " 10^{-6}	3.228 " 10^{-7}
Sq miles	Acres	640	1.562 " 10^{-3}
Sq miles	Sq kilometers	2.590	0.3861
Sq millimeters	Circular mils	1973	5.067 " 10^{-4}
Tons, short (avdp 2000 lb)	Tonnes (1000 kg)	0.9072	1.102
Tons, long (avdp 2240 lb)	Tonnes (1000 kg)	1.016	0.9842
Tons, long (avdp 2240 lb)	Tons, short (avdp 2000 lb)	1.120	0.8929
Tons, (U.S. Shipping)	Cubic feet	40	0.025
Watts	BTU per minute	5.689 " 10^{-2}	17.58
Watts	Ergs per second	107	10^{-7}

Table 1-24.—Conversion Chart—Continued

To Convert	To	Multiply By	Conversely, Multiply By
Watts	Ft-lb per minute	44.26	2.260 " 10^{-2}
Watts	Horsepower (550 ft-lb per sec)	1.341 " 10^{-3}	745.7
Watts	Horsepower (metric) (542.5 ft-lb per sec)	1.360 " 10^{-3}	735.5
Watts	Kilogram-calories per min	1.433 " 10^{-2}	69.77
Webers	Volt-seconds	.1	1
Webers per ampere-turn	Maxwell per gilbert	7.958 " 10^{-7}	1.257 " 10^{-8}
Webers per Meter^2	Gausses	10^4	10^{-4}

Electrical, Electronic/Logic, and Fiber Optic Symbols.

Figures 1-33, 1-34, and 1-35 contain symbols used in electrical, electronic/logic, and fiber optic circuits.

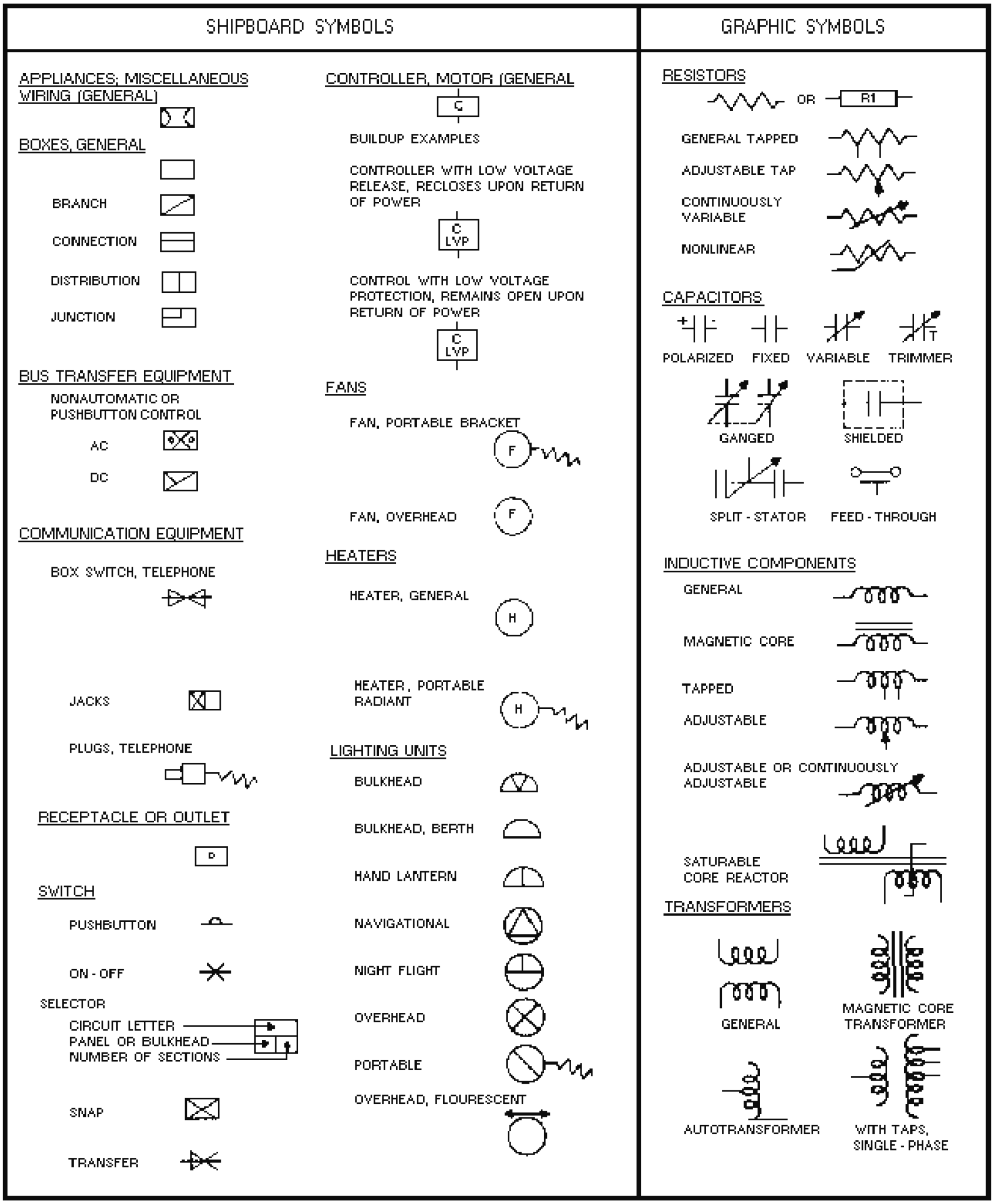


Figure 1-33.—Electrical symbols.

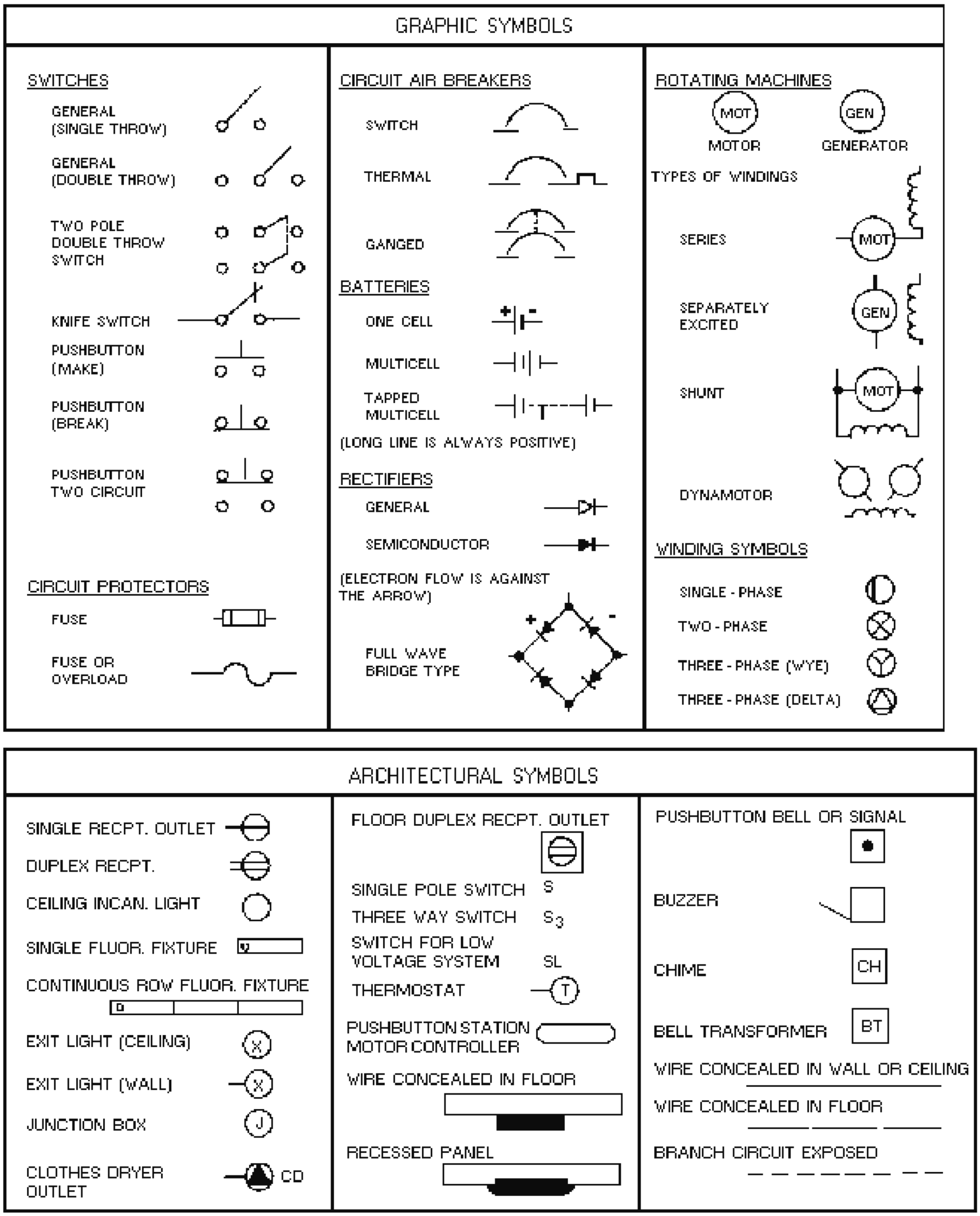
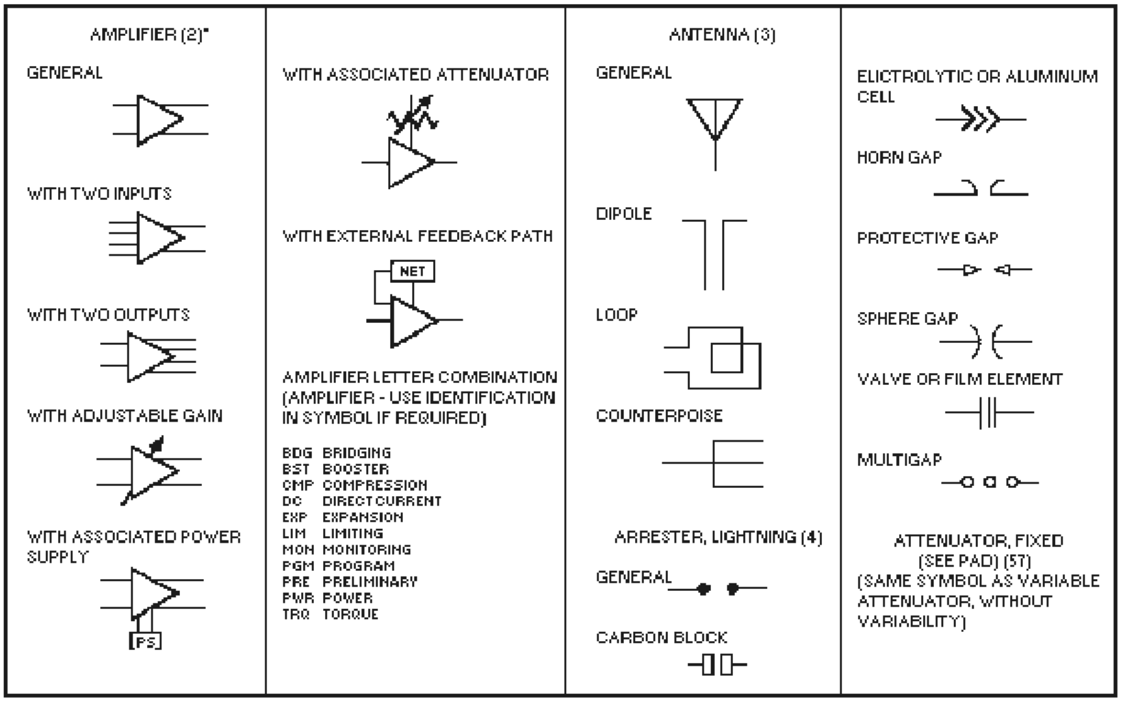


Figure 1-33.—Electrical symbols.—Continued.



* NUMBER IN PARENTHESES INDICATES LOCATION OF SYMBOL IN MIL-STD PUBLICATION

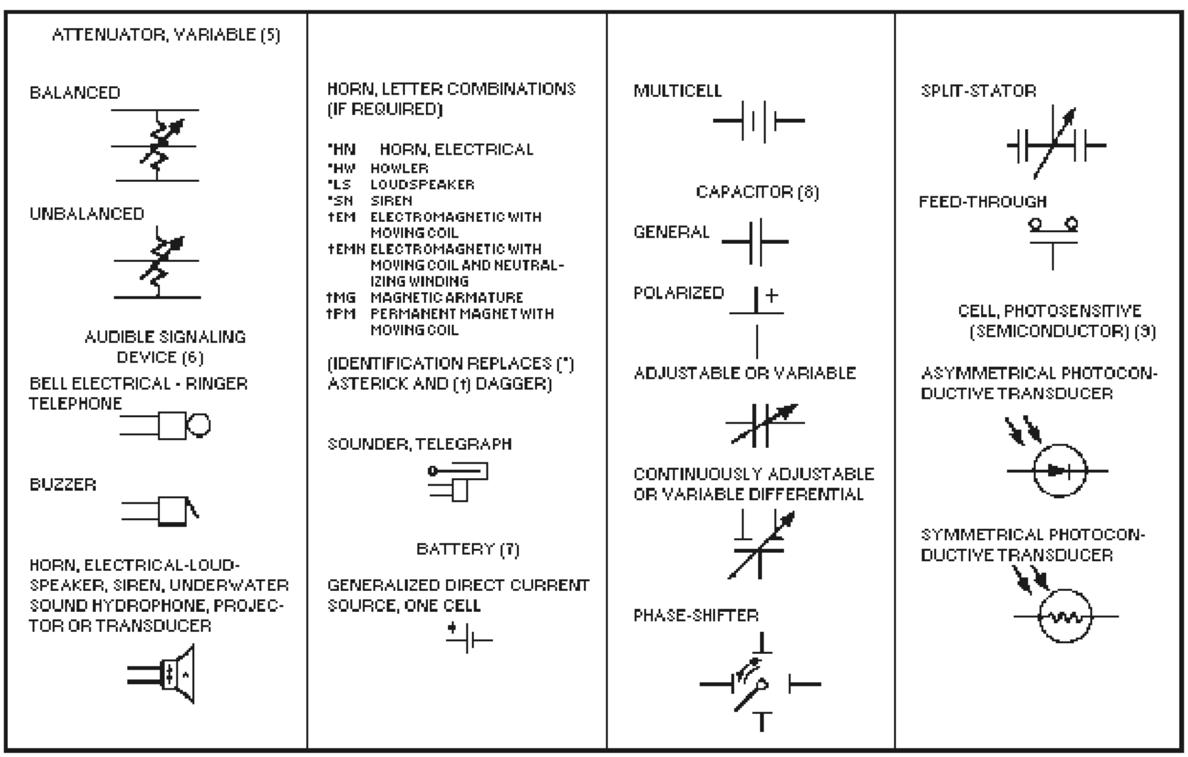


Figure 1-34.—Electronic/logic symbols.

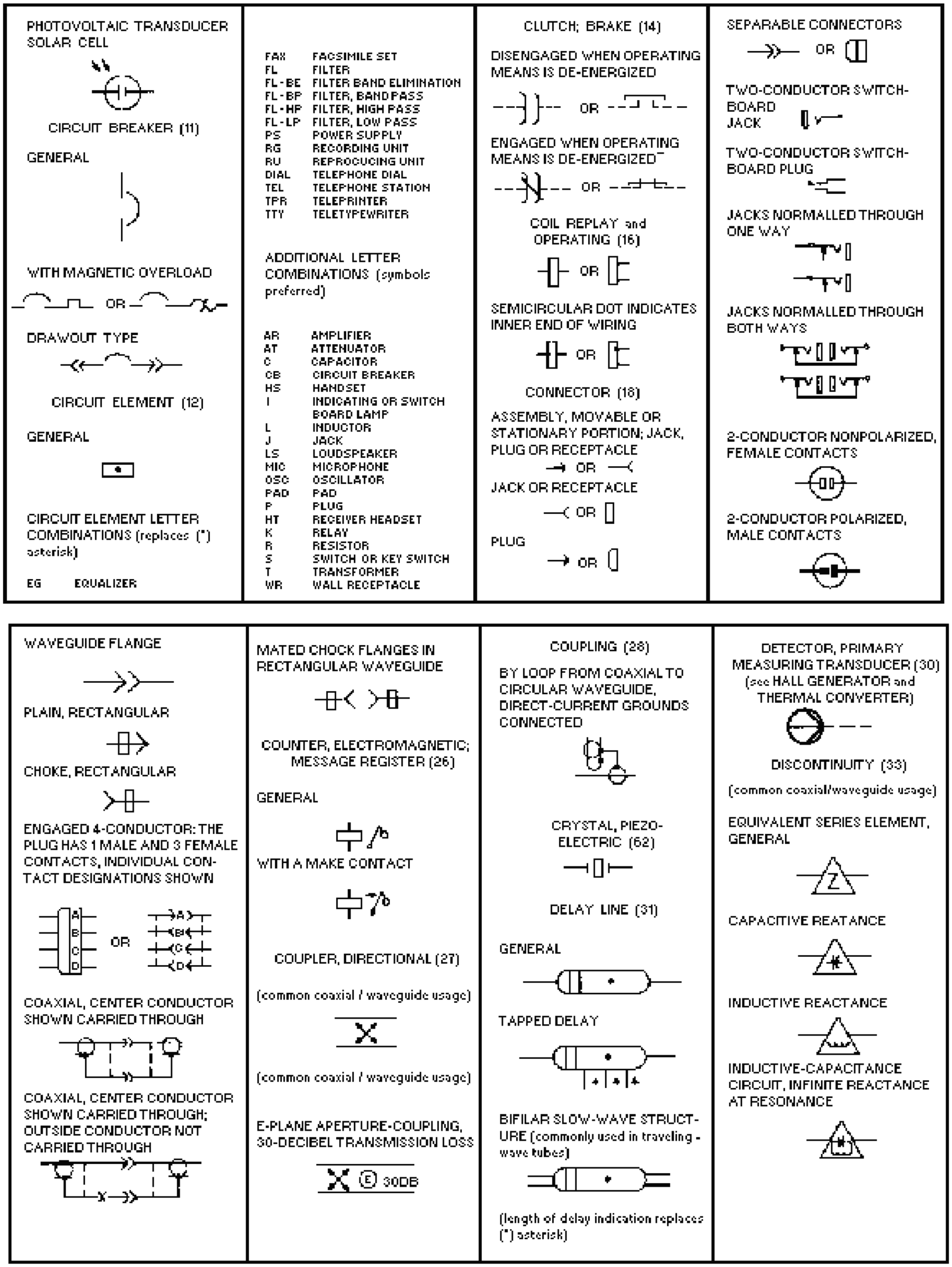
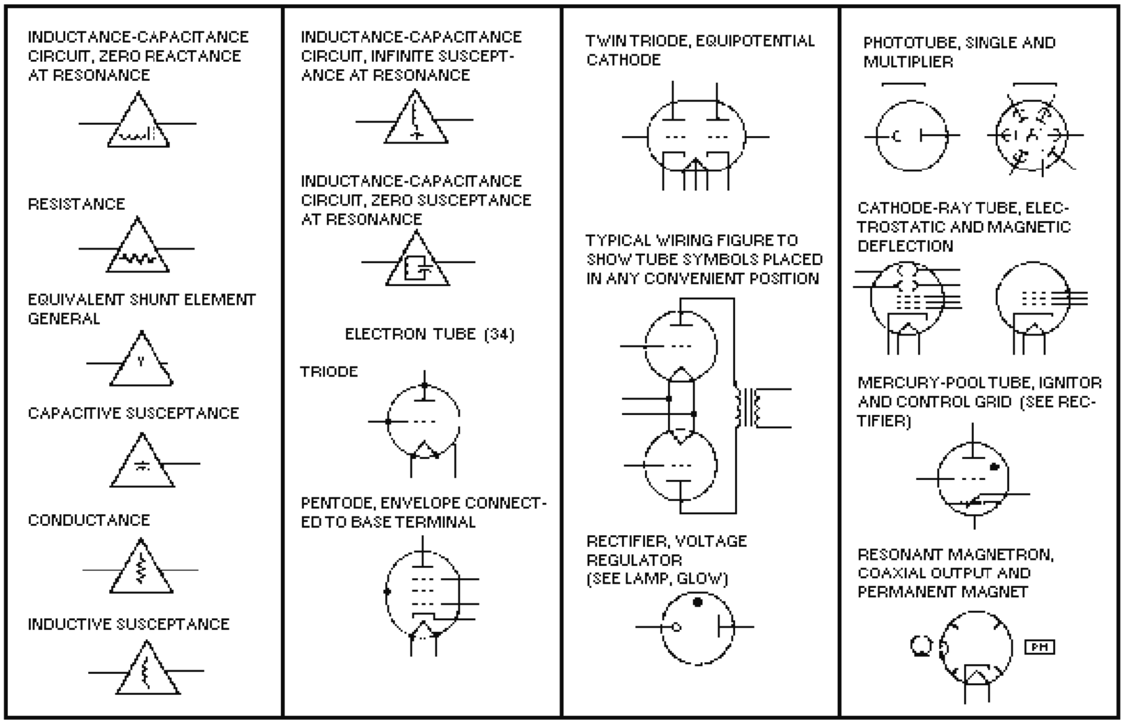


Figure 1-34.—Electronic/logic symbols.—Continued.



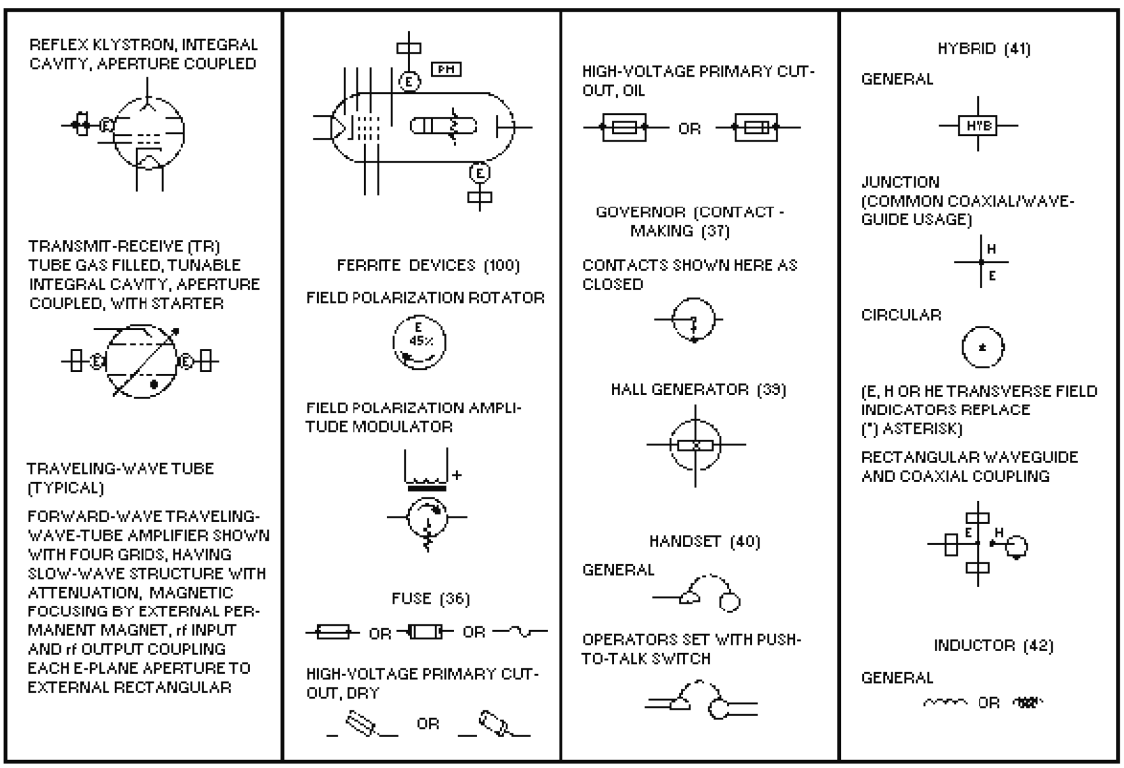
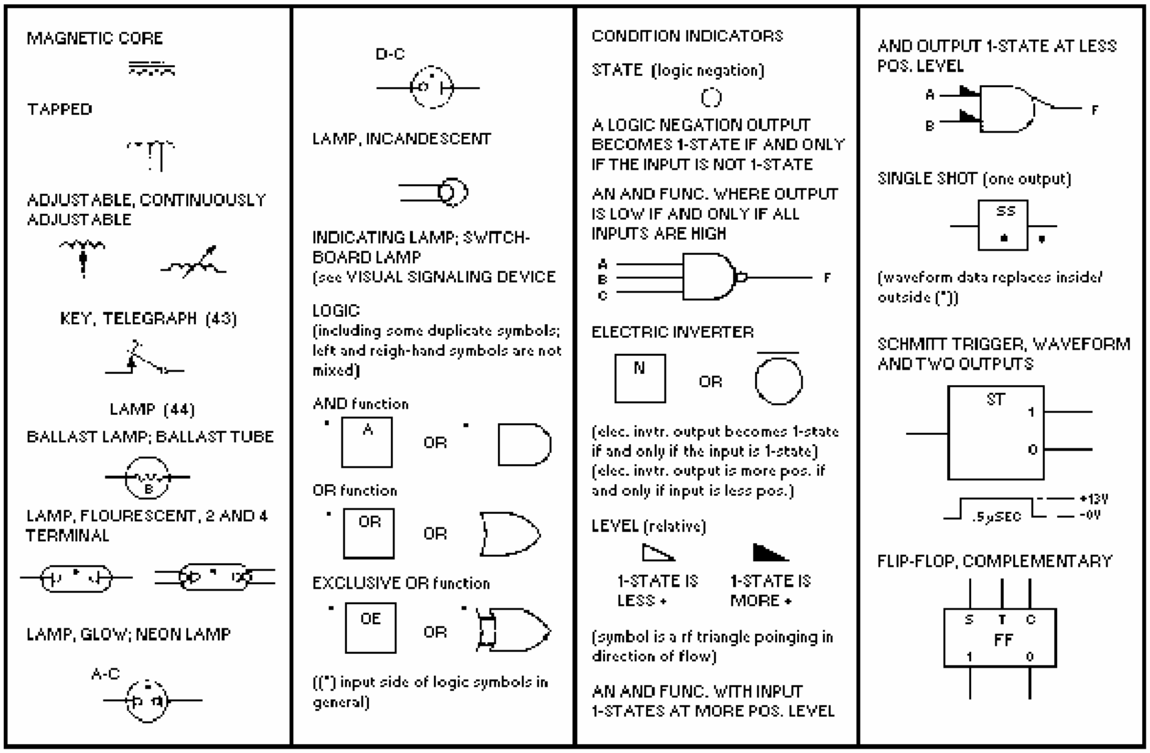


Figure 1-34.—Electronic/logic symbols.—Continued.



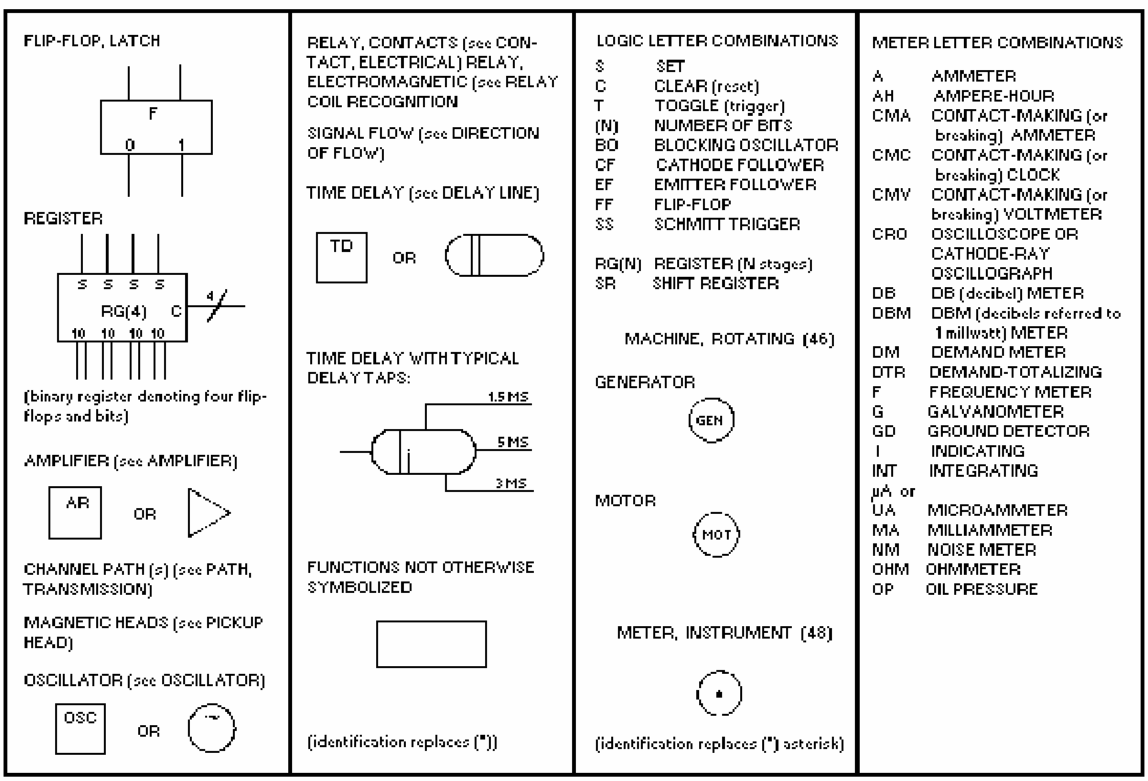
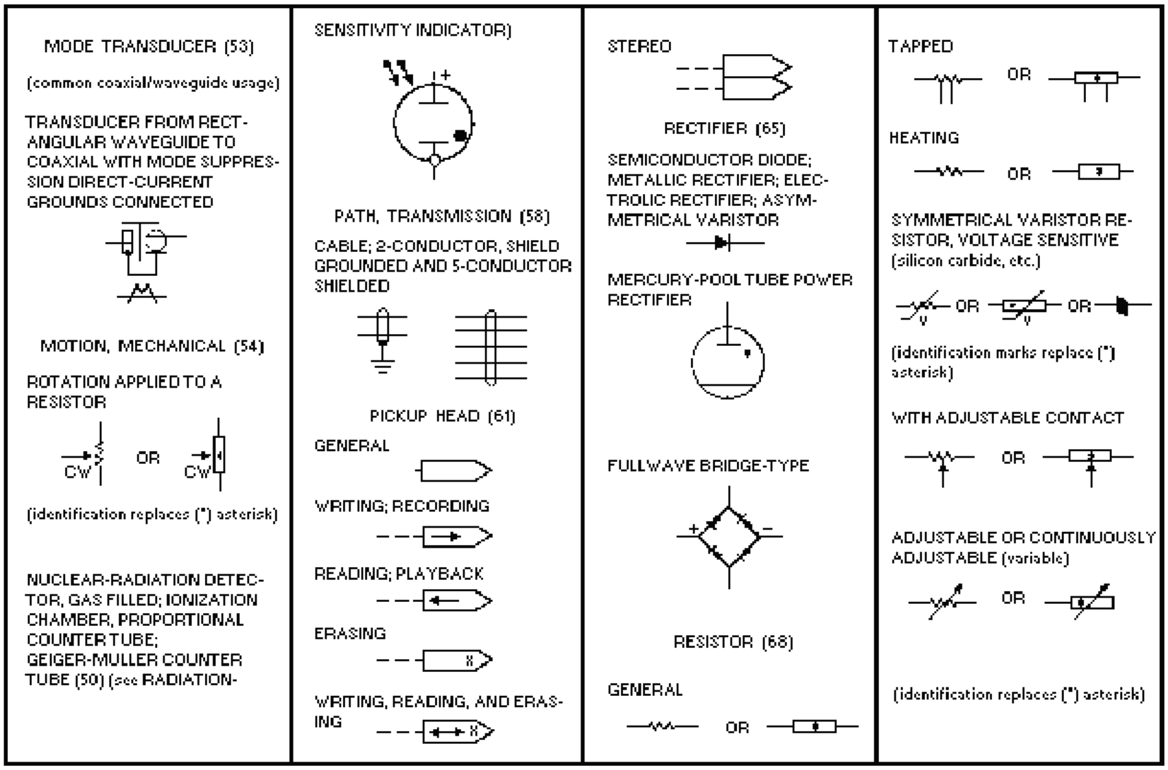


Figure 1-34.—Electronic/logic symbols.—Continued.



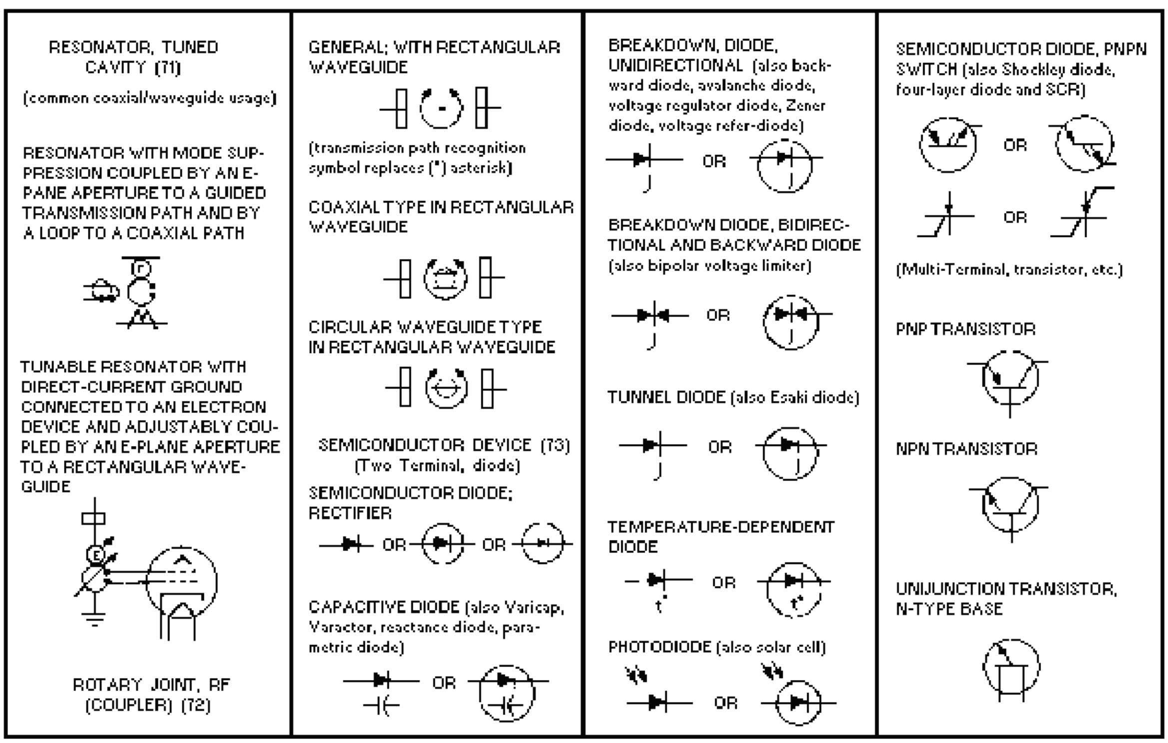
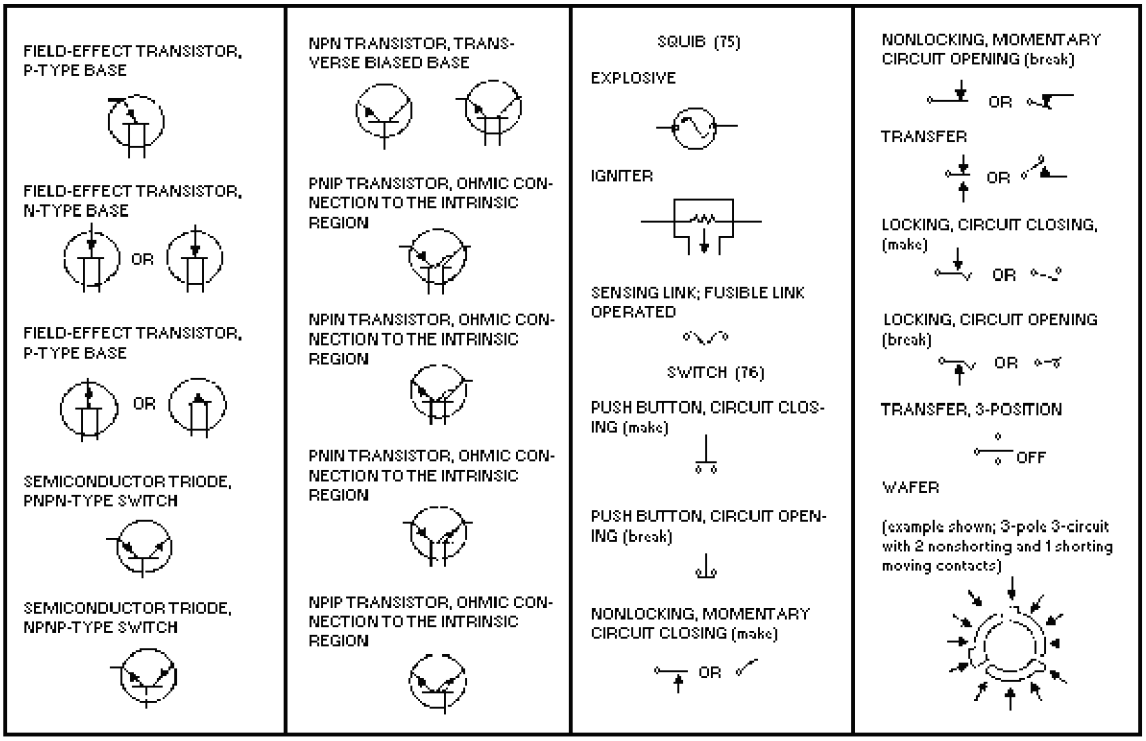


Figure 1-34.—Electronic/logic symbols.—Continued.



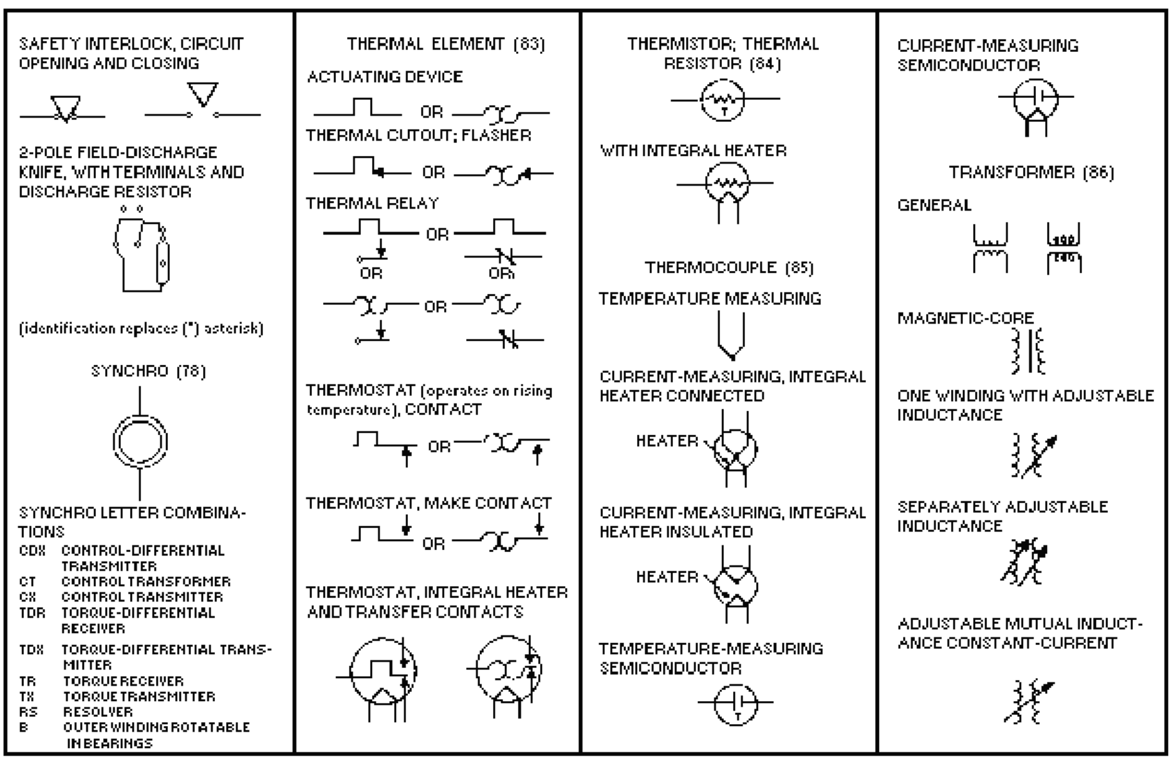


Figure 1-34.—Electronic/logic symbols.—Continued.

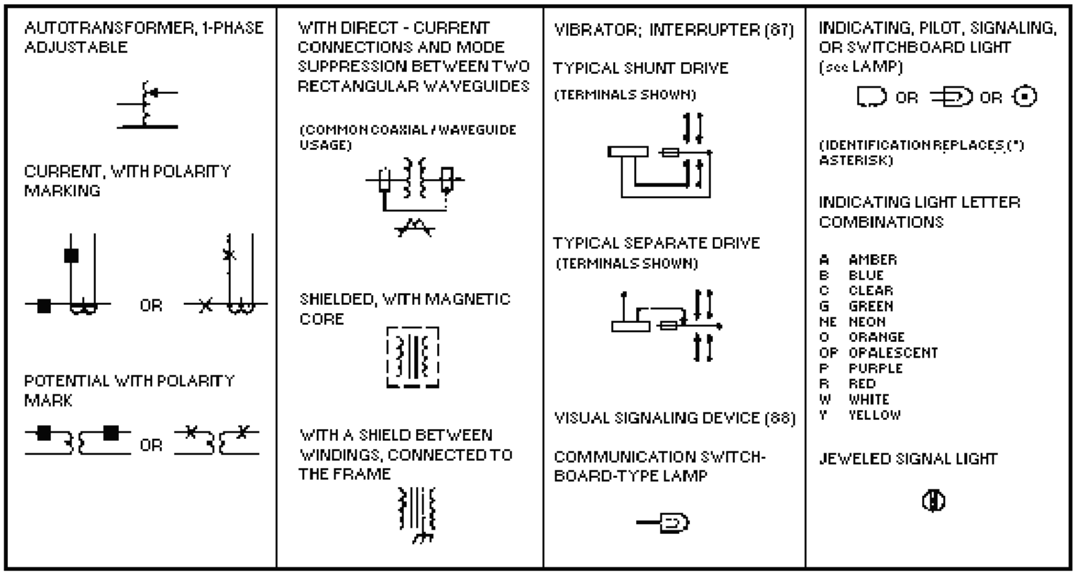


Figure 1-34.—Electronic/logic symbols.—Continued.

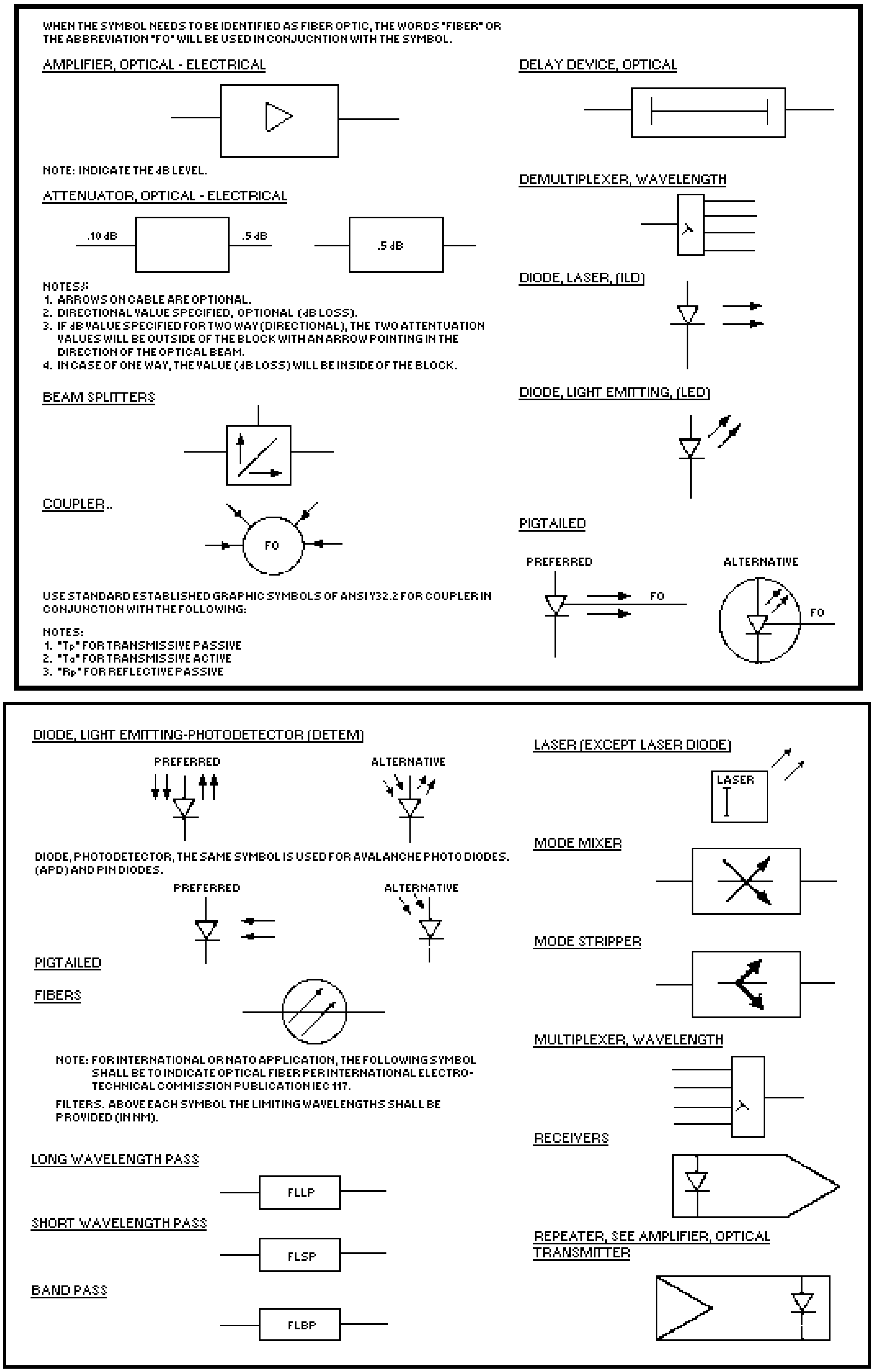


Figure 1-35.—Fiber optic symbols.

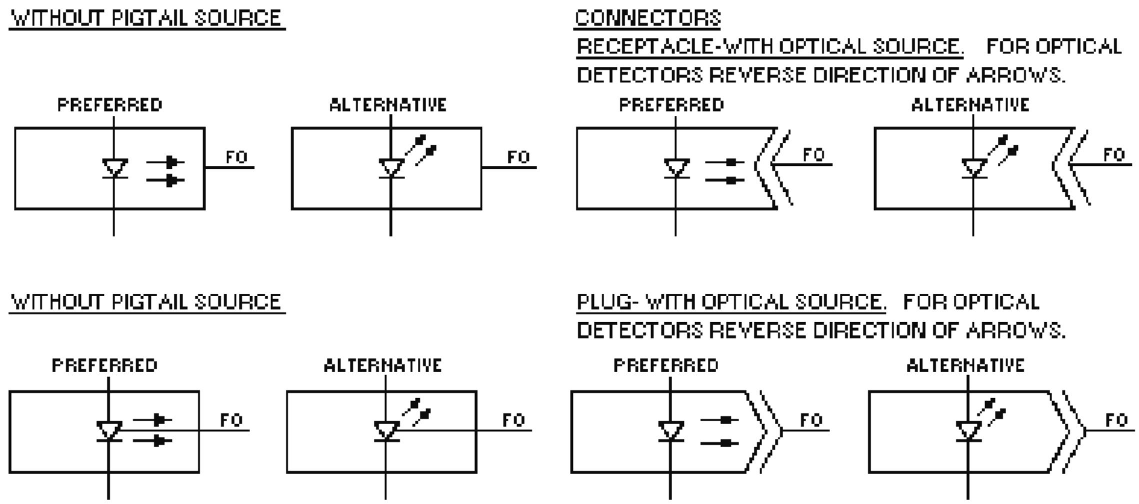


Figure 1-35.—Fiber optic symbols.—Continued.

Frequency Spectrum Designation

The complete spectrum of communications frequencies is broken down into ranges or bands. The United States practice is to designate a two- or three-letter abbreviation for the name. The practice of the International Telecommunications Union (ITU) is to designate a number. Table 1-25 shows the bands and their designators. Table 1-26 indicates the frequency spectrum broken down as to usage.

Table 1-25.—Frequency Spectrum

FREQUENCY	DESIGNATOR	IUT DESIGNATOR
BELOW 300 Hz	ELF (EXTREMELY LOW FREQUENCY)	--
300 Hz - 3 kHz	ILF (INFRA LOW FREQUENCY) SOMETIMES VF (VOICE FREQUENCY)	--
3 kHz - 30 kHz	VLF (VERY LOW FREQUENCY)	4
30 kHz - 300 kHz	LF (LOW FREQUENCY)	5
300 kHz - 3 MHz	MF (MEDIUM FREQUENCY)	6
3 MHz - 30 MHz	HF (HIGH FREQUENCY)	7
30 MHz - 300 MHz	VHF (VERY HIGH FREQUENCY)	8
300 MHz - 3 GHz	UHF (ULTRAHIGH FREQUENCY)	9
3 GHz - 30 GHz	SHF (SUPERHIGH FREQUENCY)	10
30 GHz - 300 GHz	EHF (EXTREMELY HIGH FREQUENCY)	11
300 GHz – 3 THz	THF (TREMENDOUSLY HIGH FREQUENCY)	12

Table 1-26.—Frequency Spectrum Usage

DIVISION	RANGE	USAGE
ELF	Long	Communication, Navigation, Experimental
ILF	Long	Communication, Navigation, Experimental
VLF	Long	Communication, Navigation
LF	Long Medium	Communication, Broadcasting, Navigation
MF	Medium	Communication, Broadcasting, Navigation
HF	Long	Communication, Broadcasting
VHF		Communication Television, Radar
(Lower) (Upper)	Medium Beyond horizon	
UHF	Beyond line of sight	Communication
	Line of sight	Radar
SHF	Line of sight	Radar, Doppler
EHF	Line of sight	Short-range radar
THF	Line of sight	Experimental

Television Channel Assignments

Table 1-27 lists the VHF and UHF television channel frequencies. The video carrier is 1.25 MHz above the lower channel limit. The sound carrier is .25 MHz below the upper channel limit. For example: Channel 10 sound carrier is 197 MHz, and the video carrier is 193.25 MHz.

Table 1-27.—Television Channel Frequencies

Channel	Band Limits (MHz)	Channel	Band Limits (MHz)
2	54-60	43	644-650
3	60-66	44	650-656
4	66-72	45	656-662
5	76-82	46	662-668
6	82-88	47	668-674
7	174-180	48	674-680
8	180-186	49	680-686
9	186-192	50	686-692
10	192-198	51	692-698
11	198-204	52	698-704
12	204-210	53	704-710
13	210-216	54	710-716
14	470-476	55	716-722
15	476-482	56	722-728
16	482-488	57	728-734
17	488-494	58	734-740
18	494-500	59	740-746
19	500-506	60	746-752
20	506-512	61	752-758
21	512-518	62	758-764
22	518-524	63	764-770
23	524-530	64	770-776
24	530-536	65	776-782
25	536-542	66	782-788
26	542-548	67	788-794
27	548-554	68	794-800
28	554-560	69	800-806
29	560-566	70	806-812
30	566-572	71	812-818
31	572-578	72	818-824
32	578-584	73	824-830
33	584-590	74	830-836
34	590-596	75	836-842
35	596-602	76	842-848
36	603-608	77	848-854
37	608-614	78	854-860
38	614-620	79	860-866
39	620-626	80	866-872
40	626-632	81	872-878
41	632-638	82	878-884
42	638-644	83	884-890

Joint Electronic Type Designation System (JETDS)

This system, formerly known as the Joint Army-Navy (AN) nomenclature system, was designed so that a common designation could be used for all the services' equipment. Figure 1-36 shows you how to identify equipment in the JETDS (AN) System.

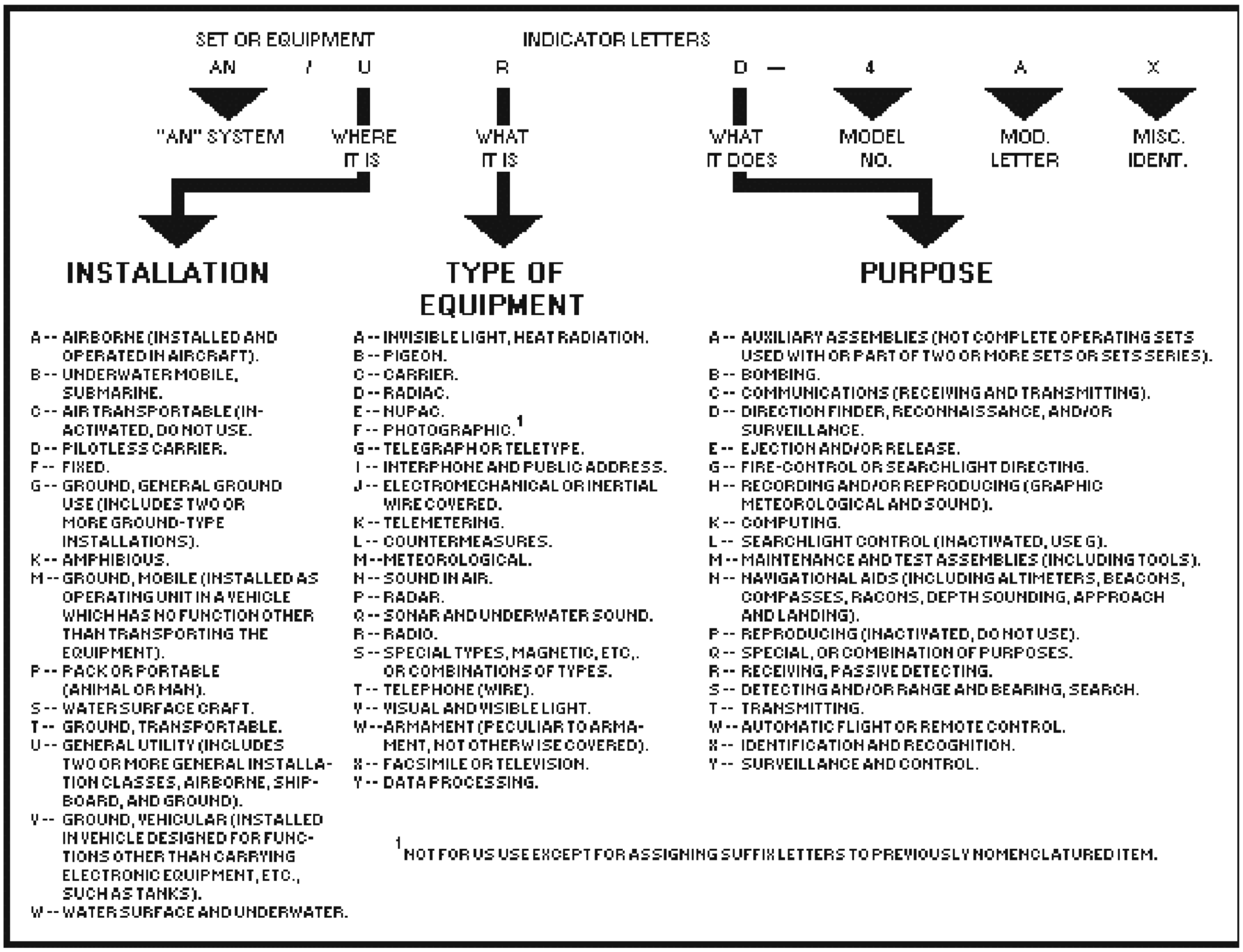


Figure 1-36.—Joint Electronics Type Designation System (AN).

Microcircuit Part Numbers.

The military designator for microcircuits is M38510. Table 1-28 shows by example how the military part number M38510/00104BCB is broken down.

Table 1-28.—Microcircuit Part Number Breakdown

1	2	3	4	5	6
M38510/	001	04	B	C	B

1. Military Specification Designator
2. Detail Specification (Group of devices of similar function)
3. Device Type (Specific part type in group)
4. Device Class:
   Class A - Manual space program
   Class B - Avionics, space satellites
   Class C - Prototype, noncritical ground systems
5. Case outline:
   A--1/4" " 1/4" Flat Pack 14 Pin
   B--1/4" " 1/8" Flat Pack 14 Pin
   C--Dip 14 Pin
   D--1/4" " 3/8" Flat Pack 14 Pin
   E--Dip 16 Pin
   F--1/4" " 3/8" Flat Pack 16 Pin
   G--Can to 99 8 Pin
   H--1/4" " 1/4" Flat Pack 10 Pin
   I--Can to 100 10 Pin
   J--Dip 24 Pin
   K--3/8" " 5/8" Flat Pack 24 Pin
   L--3/8" " 1/2" Flat Pack 24 Pin
   X--To 5
   Y--To 3
   Z--1/4" " 3/8" Flat Pack 24 Pin
6. Lead finish:
   A--Hot Solder
   B--Tin Plate
   C--Gold Plate

Table 1-29 is a microcircuit-part-number-to-circuit-type crossover list. By using this table and table 1-28, we find our example part number M38510/00104BCB is a 5400 microcircuit type in class B, with 14 pin DIP, and tin plate leads.

Table 1-29.—Microcircuit Part Numbers to Circuit Type Crossover List

M38510/Ckt Type	M38510/Ckt Type	M38510/Ckt Type	M38510/Ckt Type	M38510/Ckt Type
00101/5430	00102/5420	00103/5410	00104/5400	00105/5404
00106/5412	00107/5401	00108/5405	00109/5403	00201/5472
00202/5473	00203/54107	00204/5476	00205/5474	00206/5470
00207/5479	00301/5440	00302/5437	00303/5438	00401/5402
00402/5423	00403/5425	00404/5427	00501/5450	00502/5451
00503/5453	00504/5454	00601/5482	00602/5483	00603/9304
00604/5480	00701/3121	00701/5486	00801/5406	00802/5416
00803/5407	00804/5417	00805/5426	00901/5495	00902/5496
00903/54164	00904/54165	00905/54194	00906/54195	00909/54198
00910/54166	01001/5442	01002/5443	01003/5444	01004/5445
01005/54145	01006/5446	01007/5447	01008/5448	01009/5449
01101/54181	01101/7181	01101/9341	01101/54182	01102/9342
01201/54121	01202/54122	01203/54123	01204/9601	01205/9602
01301/5492	01302/5493	01303/54160	01304/54163	01305/54162
01306/54161	01307/5490	01308/54192	01309/54193	01310/54196
01311/54197	01312/54177	01401/54150	01402/9312	01403/54153
01404/9309	01405/54157	01405/9322	01406/54151	01501/5475
01502/5477	01503/54116	01503/9308	01504/9314	01601/5408
01602/5409	01701/54174	01702/54175	01703/54173	01801/54170
01901/54180	01902/556	02001/54L30	02002/54L20	02003/54L10
02004/54L00	02005/54L04	02006/54L01	02006/54L03	02101/54L71
02102/54L72	02103/54L73	02104/54L78	02105/54L74	02201/54H72
02202/54H73	02203/54H74	02204/54H76	02205/54H101	02206/54H103
02301/54H30	02302/54H20	02303/54H10	02304/54H00	02305/54H04
02306/54H01	02307/54H22	02401/54H40	02501/54L90	02502/54L93
02503/54L193	02504/93L10	02505/93L16	02601/54L86	02603/7644
02701/54L02	02801/54L95	02802/54L164	02803/93L28	02804/93L00
02805/76L70	02901/54L42	02902/54L43	02903/54L44	02904/54L46
02905/54L47	02906/76L42A	02907/93L01	03001/15930	03001/930
03002/15935	03002/19535	03002/935	03002/940	03003/15936
03003/936	03004/15946	03004/946	03005/15962	03005/962
03101/15932	03102/15944	03103/15957	03104/15958	03105/15933
03201/15951	03301/15945	03302/15948	03303/15950	03304/9094
03501/MH0026	03604/54LS96	04001/54H50	04002/54H51	04003/54H53
04004/54H54	04005/54H55	04101/54L51	04102/54L54	04103/54L55
04201/54L121	04202/54L122	04301/93L18	04401/93L24	04501/93L14
04502/93L08	04601/93L09	04602/93L12	04603/93L22	05001/4011A
05002/4012A	05003/4023A	05101/4013A	05102/4027A	05103/4043A
05201/4000A	05202/4001A	05203/4002A	05204/4025A	05301/4007A
05302/4019A	05303/4030A	05401/4008A	05501/4009A	05502/4010A
05503/4049A	05504/4050A	05505/4041A	05601/4017A	05602/4018A
05603/4020A	05604/4022A	05605/4024A	05701/4006A	05702/4014A
05703/4015A	05704/4021A	05705/4031A	05706/4034A	05801/4016A
05802/4066A	05901/4028A	06001/10501	06002/10502	06003/10505
06004/10506	06005/10507	06006/10509	06101/10531	06102/10531
06103/10576	06104/10535	06201/10504	06202/10597	06301/10524
06302/10525	07001/54S00	07002/54S03	07003/54S04	07004/54S05
07005/54S10	07006/54S20	07007/54S22	07008/54S30	07009/54S133
07010/54S134	07101/54S74	07102/54S112	07103/54S113	07104/54S114
07105/54S174	07106/54S175	07201/54S40	07301/54S02	07401/54S51
07402/54S64	07403/54S65	07501/54S86	07502/54S135	07601/54S194
07602/54S195	07701/54S138	07702/54S139	07703/54S280	07801/54S181
07802/54S182	07901/54S151	07902/54S153	07903/54S157	07904/54S158
07905/54S251	07906/54S257	07907/54S258	07908/54S253	08001/54S11
08002/54S15	08003/54S08	08004/54S09	08101/54S140	08201/54S85

Table 1-29.—Microcircuit Part Numbers to Circuit Type Crossover List—Continued

M38510/Ckt Type	M38510/Ckt Type	M38510/Ckt Type	M38510/Ckt Type	M38510/Ckt Type
10101/UA741	10101/52741	10101/741	10102/52747	10102/7A747
10102/747	10103/LM101A	10103/52101A	10104/LM108A	10104/52108A
10105/LH2101A	10106/LH2108A	10107/LM118	10108/1558	10201/LM723
10201/UA723	10201/52723	10201/723	10202/LM104	10203/LM105
10301/UA710	10301/52710	10301/710	10302/UA711	10302/52711
10302/711	10303/LM106	10303/52106	10304/LM111	10304/52111
10305/LH2111	10305/LM2111	10401/55107	10402/55108	10403/55114
10403/9614	10404/55115	10404/9615	10405/55113	10406/7831
10407/7832	10501/UA733	10501/52733	10601/LM102	10602/LM110
10602/52110	10603/LH2110	10603/LM2110	10701/LM109	10701/52109
10702/LM141H-05	10703/LM141H-12	10704/LM141H-15	10705/LM141H-24	10706/LM140K-05
10707/LM140K-12	10708/LM140K-15	10709/LM140K-24	10801/3018A	10802/3045
10901/SE555	10901/555	10902/SE556	11001/LM148	11002/LM149
11003/4141	11003/4156	11004/4136	11005/LM124	11101/DG181A
11102/DG182A	11103/DG184A	11104/DG185A	11105/DG187A	11106/DG188A
11107/DG190A	11108/DG191A	11201/LM139	11202/LM193	11301/DAC-08
11302/DAC-08A	11401/LF155	11402/LF156	11403/LF157	11404/LF155A
11405/LF156A	11406/LF157A	11501/LM120H-05	11502/79M05	11502/LM120H-12
11502/79M12	11503/LM120H-15	11503/79M15	11504/LM120H-24	11504/79M24
11505/LM120K-05	11505/7905	11506/LM120K-12	11506/7912	11507/LM120K-15
11507/7915	11508/LM120K-24	11508/7924	11901/061	11902/062
11903/064	11904/LF151	11904/071	11904/771	11905/LF153
11905/072	11905/772	11906/LF147	11906/074	11906/774
15001/5485	15002/9324	15101/5413	15102/5414	15102/7414
15103/54132	15201/54154	15201/9311	15202/54155	15203/54156
15204/8250	15205/8251	15206/8252	15206/9301	15301/54125
15302/54126	15401/54120	15501/MC3101	15501/54H08	15502/MC3106
15502/54H11	15503/MC3111	15503/54H21	15601/54147	15602/54148
15603/9318	15701/9338	15801/9321	15802/9317	15901/9300
15902/9328	16001/9334	16101/5432	16201/5428	16301/54365
16302/54366	16303/54367	16304/54368	20101/MCM5303	20101/54186
20102/MCM5304	20201/IM5603A	20201/IM5603	20201/54S387	20202/IM5623
20301/AM27S10	20301/5300-1	20301/7610	20301/82S126	20301/93417
20302/AM27S11	20302/5301-1	20302/7611	20302/82S129	20302/93427
20401/IM5604	20401/5305-1	20401/7620	20401/82S130	20401/93436
20402/IM5624	20402/5306-1	20402/7621	20402/82S131	20402/93446
20501/HHX7620-8	20502/HMX7621-8	20601/HMX7640-8	20601/5352-1	20601/7642
20601/82S136	20601/93452	20602/HMX7641-8	20602/5353-1	20602/7643
20602/82S137	20602/93453	20603/7644	20701/5330	20701/7602
20701/82S23	20702/5331	20702/7603	20702/82S123	20801/5340-1
20801/7640	20801/82S140	20801/93438	20802/5341-1	20802/7641
20802/82S141	20802/93448	20803/82S115	20804/5348-1	20805/5349-1
20901/7684	20901/82S184	20902/7685	20902/82S185	20903/5380-1
20903/7680	20903/82S180	20903/93450	20904/5381-1	20904/7681
20904/82S181	20904/93451	20905/82S2708	20905/93461	20906/93460
21001/53S1680	21001/76160	21001/82S190	21001/93510	21002/53S1681
21002/76161	21002/82S191	21002/93511	22001/2708	23001/93410
23002/93411	23003/93421	23004/93L420	23101/82S10	23102/82S11
23102/93425	23103/93L415	23104/93L425	23403/54LS244	23501/TMS4060
23502/TMS4050	23503/TMS4060	23504/TMS4050	23505/MM5280	23506/MM5280
23601/MCM6605	23602/MCM6604A	23602/MKB4096	23603/MCM6605	23604/MCM6604A
23604/MKB4096	23701/AM9130CFC	23702/AM9130AFC	23703/AM9130CDM	23703/AM9130CFM
23704/AM9130ADM	23704/AM9130AFM	23705/AM91L30CF	23706/AM91L30AF	23707/AM91L30CDM
23707/AM91L30CFM	23708/AM91L30ADM	23708/AM91L30AFM	23709/AM9140CFC	23710/AM9140AFC
23711/AM9140CDM	23711/AM9140CFM	23712/AM9140ADM	23712/AM9140AFM	23713/AM91L40CFC
23714/AM91L40AFC	23715/AM91L40CDM	23715/AM91L40CFM	23716/AM91L40ADM	23716/AM91L40AFM

Table 1-29.—Microcircuit Part Numbers to Circuit Type Crossover List—Continued

M38510/Ckt Type	M38510/Ckt Type	M38510/Ckt Type	M38510/Ckt Type	M38510/Ckt Type
23901/54C929	23901/6508	23902/54C930	23902/6518	24001/2117
24002/2117	24002/4116	30001/54LS00	30002/54LS03	30003/54LS04
30004/54LS05	30005/54LS10	30006/54LS12	30007/54LS20	30008/54LS22
30009/54LS30	30101/54LS73	30102/54LS74	30103/54LS112	30104/54LS113
30105/54LS114	30106/54LS174	30107/54LS175	30108/54LS107	30109/54LS109
30110/54LS76	30201/54LS40	30202/54LS37	30203/54LS38	30204/54LS28
30301/54LS02	30302/54LS27	30303/54LS266	30401/54LS51	30401/9LS51
30402/54LS54	30402/9LS54	30501/54LS32	30502/54LS86	30601/54LS194
30602/54LS195	30603/54LS95	30605/54LS164	30606/54LS295	30607/54LS395
30608/54LS165	30609/54LS166	30701/54LS138	30702/54LS139	30703/54LS42
30704/54LS47	30801/54LS181	30901/54LS151	30902/54LS153	30903/54LS157
30904/54LS158	30905/54LS251	30906/54LS257	30907/54LS258	30908/54LS253
30909/54LS298	31001/54LS11	31002/54LS15	31003/54LS21	31004/54LS08
31005/54LS09	31101/54LS85	31201/54LS83A	31202/54LS283	31301/54LS13
31302/54LS14	31303/54LS132	31401/54LS123	31402/54LS221	31403/54LS122
31501/54LS90	31502/54LS93	31503/54LS160	31504/54LS161	31505/54LS168
31506/54LS169	31507/54LS192	31508/54LS193	31509/54LS191	31510/54LS92
31511/54LS162	31512/54LS163	31513/54LS190	31601/54LS75	31602/54LS279
31603/54LS259	31604/54LS375	31701/54LS124	31702/54LS324	31801/54LS261
31901/54LS670	32001/54LS196	32002/54LS197	32003/54LS290	32004/54LS293
32101/93415	32102/54LS26	32201/54LS365	32202/54LS366	32203/54LS367
32204/54LS368	32301/54LS125	32302/54LS126	32401/54LS240	32402/54LS241
32501/54LS273	32502/54LS373	32503/54LS374	32504/54LS377	32601/54LS155
32602/54LS156	32701/54LS390	32702/54LS393	32703/54LS490	32801/54LS242
32802/54LS243	32803/54LS245	32901/54LS280	33106/25LS174	33107/25LS175
36001/54LS148	36002/54LS348	40001/6800	42001/8080A	42101/54S412
42101/8212	42201/8224	42301/8228	44001/2901A	44101/2905
44102/2906	44103/2907	44104/2915	44105/2916	44106/2917
44201/2918	46001/9900A	47001/1802		

You can find more information on microcircuits by referring to Military Specification 38510 (MIL-M-38510), Military Standard 1562D (MIL-STD-1562D), and NEETS, Module 14, *Introduction to Microelectronics*.

Shipboard Announcing System

Table 1-30 is a breakdown list of the shipboard announcing system matched to the circuit designator.

Table 1-30.—Shipboard Announcing System

CIRCUIT	SYSTEM
*1MC	General
*2MC	Propulsion plant
*3MC	Aviators'
4MC	Damage Control
*5MC	Flight Deck
*6MC	Intership
7MC	Submarine Control
8MC	Troop administration and control
*9MC	Underwater troop communication
*10MC	Dock Control (obsolete)
*11-16MC	Turret (obsolescent)
*17MC	Double Purpose Battery (obsolescent)
18MC	Bridge
19MC	Aviation Control
*20MC	Combat Information (obsolescent)
21MC	Captain's Command
22MC	Electronic Control
23MC	Electrical control
24MC	Flag Command
25MC	Ward Room (obsolescent)
26MC	Machinery Control
27MC	Sonar and Radar Control
*28MC	Squadron (obsolescent)
*29MC	Sonar Control and Information
30MC	Special Weapons
31MC	Escape trunk
32MC	Weapons control
33MC	Gunnery Control (obsolescent)
34MC	Lifeboat (obsolescent)
35MC	Launcher Captains'
36MC	Cable Control (obsolete)
37MC	Special Navigation (obsolete)
38MC	Electrical (obsolete)
39MC	Cargo Handling
40MC	Flag Administrative
41MC	Missile Control and Announce (obsolete)
42MC	CIC Coordinating
43MC	Unassigned
44MC	Instrumentation Space
45MC	Research operations
*46MC	Aviation Ordnance and Missile Handling
47MC	Torpedo Control
48MC	Stores conveyor (obsolescent)
49MC	Unassigned
50MC	Integrated operational intelligence center
51MC	Aircraft Maintenance and handling control
52MC	Unassigned

Table 1-30.—Shipboard Announcing System—Continued

CIRCUIT	SYSTEM
53MC	Ship Administrative
54MC	Repair officer's control
55MC	Sonar Service
56MC	Unassigned
57MC	Unassigned
58MC	Hanger Deck Damage Control
59MC	SAMID Alert

*- Central amplifier systems

Shipboard Alarm and Warning Systems

Table 1-31 is a breakdown list of the shipboard alarm and warning systems matched to a circuit designator.

Table 1-31.—Shipboard Alarm and Warning System

CIRCUIT	SYSTEM
BZ	Brig cell door alarm and lock operating
BW	Catapult Bridle Arresterman safety Ind.
CX	Bacteriological Lab. & Pharmacy Comb. Refer Failure
DL	Secure communications space door position alarm
DW	Wrong direction alarm
EA	Reactor compartment or fireroom emergency alarm
1EC	Lubricating oil low pressure alarm-propulsion machinery
2EC	Lubricating oil low pressure alarm-auxiliary machinery
1ED	Generator high temperature alarm
2ED	Oxygen-nitrogen generator plant low temperature alarm
EF	Generator bearing high temperature alarm
EG	Propeller pitch control, hydraulic oil system low pressure alarm
EH	Gas turbine exhaust high temperature alarm
EJ	Feed pressure alarm
1EK	Pneumatic control air pressure alarm
3EK	Catapult steam cutoff and alarm
EL	Radar cooling lines temperature and flow alarm
EP	Gas turbine lubricating oil high temperature alarm
1EQ	Desuperheater high temperature alarm
2EQ	Catapult steam trough high temperature alarm
3ES	Reactor fill alarm
ET	Boiler temperature alarm
EV	Toxic vapor detector alarm
1EW	Propulsion engines circulating water high temperature
2EW	Auxiliary machinery circulating water high temperature
EZ	Condenser vacuum alarm
F	High temperature alarm
4F	Combustion gas and smoke detector
9F	High temperature alarm system-ASROC launcher
11F	FBM storage area temperature and humidity alarm
12F	Gyro ovens temperature and power failure alarm
FD	Flooding alarm
FH	Sprinkling alarm
FR	Carbon dioxide release alarm
FS	Flight Deck Readylight Signal system
FZ	Security alarm (CLASSIFIED)
4FZ	Torpedeo alarm (CLASSIFIED)
HF	Air flow indicator and alarm
LB	Steering Emergency Signal system
LS	Submersible steering gear alarm
MG	Gas turbine overspeed alarm
NE	Nuclear facilities air particle detector alarm
NH	Navigation Horn Operating System

Table 1-31.—Shipboard Alarm and Warning System—Continued

CIRCUIT	SYSTEM
QA	Air lock warning
QD	Air filter and flame arrester pressure differential alarm, or gasoline compartment exhaust blower alarm
QX	Oxygen-nitrogen plant ventilation exhaust alarm
RA	Turret emergency alarm
RD	Safety observer warning
RW	Rocket and torpedo warning
4SN	Scavenging air blower high temperature alarm
SP	Shaft position alarm
TD	Liquid level alarm
1TD	Boiler water level alarm
2TD	Deaerating feed tank water level alarm
5TD	Reactor compartment bilge tank alarm
6TD	Primary shield tank, expansion tank level alarm
7TD	Reactor plant fresh water cooling expansion tank level alarm
8TD	Reactor secondary shield tank level alarm
9TD	Lubricating oil sump tank liquid level alarm
11TD	Induction air sump alarm
12TD	Diesel oil sea water compensating system tank liquid level alarm
14TD	Auxiliary fresh water tank low level alarm
16TD	Pure water storage tank low level alarm
17TD	Reserve feed tank alarm
18TD	Effluent tanks and contaminated laundry tank high level alarm
19TD	Sea water expansion tank low level alarm
20TD	Gasoline drain tank high level alarm
21TD	Moisture separator drain cooler high level alarm
24TD	Reactor plant on board discharge tank level alarm
25TD	Crossover drains high level alarm
29TD	Sonar dome fill tank low level alarm
30TD	JP-5 fuel drain tank high level alarm
TW	Train Warning system
W	Whistle Operating System

Sound-Powered Telephone Circuits

Table 1-32 is a breakdown list of the sound-powered telephone circuits matched to circuit designators.

Table 1-32.—Sound-Powered Telephone Circuits

CIRCUIT	PRIMARY CIRCUITS TITLE
JA	Captain's battle circuit
JC	Weapons control circuit
10JC	Missile battery control circuit
JD	Target detectors circuit
JF	Flag officer's circuit
1JG	Aircraft control circuit
2JG	Aircraft information circuit
2JG1	Aircraft strike coordination circuit
2JG2	Aircraft strike requirement and reporting circuit
2JG3	Aircraft information circuit CATTC direct line
3JG	Aircraft service circuit
4JG1	Aviation fuel and vehicular control circuit
4JG2	Aviation fueling circuit forward
4JG3	Aviation fueling circuit aft
5JG1	Aviation ordnance circuit
5JG2	Aviation missile circuit
6JG	Arresting gear and barricade control circuit
9JG	Aircraft handling circuit
10JG	Airborne aircraft information circuit
11JG	Optical landing system control circuit
JH	Switchboard cross connecting circuit
JL	Lookouts circuit
JK	Double purpose fuse circuit
JM	Mine control circuit
JN	Illumination control circuit
JO	Switchboard operators circuit
2JP	Dual purpose battery control circuit
4JP	Heavy machine gun control circuit
5JP	Light machine gun control circuit
6JP	Torpedo control circuit
8JP	ASW weapon control circuit
9JP	Rocket battery control circuit
10JP	Guided missile launcher control circuit
3JV	Engineer's circuit (boiler)
4JV	Engineer's circuit (fuel and stability)
5JV	Engineer's circuit (electrical)
6JV	Ballast control circuit
11JV	Waste control circuit
JW	Ship control bearing circuit
JX	Radio and signals circuit
2JZ	Damage and stability control
3JZ	Main deck repair circuit

Table 1-32.—Sound-Powered Telephone Circuits—Continued

CIRCUIT	PRIMARY CIRCUITS TITLE
4JZ	Forward repair circuit
5JZ	After repair circuit
6JZ	Midships repair circuit
7JZ	Engineer's repair circuit
8JZ	Flight deck repair circuit
9JZ	Magazine sprinkling and ordnance repair circuit forward
10JZ	Magazine sprinkling and ordnance repair circuit aft
11JZ	Gallery deck and island repair circuit

Table 1-32.—Sound-Powered Telephone Circuits

CIRCUIT	PRIMARY CIRCUITS TITLE
	Auxiliary Circuits
XJA	Auxiliary captain's battle circuit
X1JG	Auxiliary aircraft control circuit
X1JV	Auxiliary maneuvering and docking circuit
XJX	Auxiliary radio and signals circuit
X2JZ	Auxiliary damage and stability control circuit
	Supplementary Circuits
X1J	Ship administration circuit
X2J	Leadsman and anchor control circuit
X3J	Engineer watch officer's circuit
X4J	Degaussing control circuit
X5J	Machinery room control circuit
X6J1	Electronic service circuit
X6J7	ECM service circuit
X6J11-14	NTDS service circuits
X7J	Radio-sonde information circuit
10JP1	Starboard launcher circuit
10JP2	Port launcher circuit
11JP	FBM checkout and control circuit
JQ	Double purpose sight setters circuit
JR	Debarkation control circuit
JS	Plotters' transfer switchboard circuit
1JS	CIC information circuit
2JS	NTDS coordinating circuit No. 1
3JS	NTDS coordinating circuit No. 2
20JS1	Evaluated radar information circuit
20JS2	Evaluator's circuit
20JS3	Radar control officer's circuit

Table 1-32.—Sound-Powered Telephone Circuits—Continued

CIRCUIT	PRIMARY CIRCUITS TITLE
	SUPPLEMENTARY CIRCUITS (CONTINUED)
20JS4	Weapons liaison officer's circuit
21JS	Surface search radar circuit
22JS	Long range air search radar circuit
23JS	Medium range air search radar circuit
24JS	Range height finder radar circuit
25JS	AEW radar circuit
26JS	Radar information circuit
31JS	Track analyzer No. 1 air radar information check
32JS	Track analyzer No. 2 air radar information check
33JS	Track analyzer No. 3 air radar information check
34JS	Track analyzer No. 4 air radar information check
35JS	Raid air radar information circuit
36JS	Combat air patrol air radar information circuit
61JS	Sonar information circuit
80JS	ECM plotters' circuit
81JS	ECM information circuit
82JS	Supplementary radio circuit
JT	Target designation control circuit
1JV	Maneuvering and docking circuit
2JV	Engineers' circuit (engines)
X8J	Replenishment-at-sea circuit
X9J	Radar trainer circuit
X10J	Cargo transfer control circuit
X10J1	Cargo transfer circuit-Lower decks
X10J10	Cargo transfer circuit-Upper decks
X11J	Captain's and admiral's cruising circuit
X12J	Capstan control circuits
X13J	Aircraft crane control circuits
X14J	Missile handling and nuclear trunk crane circuit
X15J	SINS information circuit
X16J	Aircraft elevator circuit
X17J	5-inch ammunition hoist circuit
X18J	Machine gun ammunition hoist circuits
X19J	Missile component elevator circuit
X20J	Weapons elevator circuits
X21J	Catapult circuit
X22J	Catapult steam control circuit
X23J	Stores conveyor circuit
X24J	Cargo elevator circuit
X25J	Sonar service circuit
X26J	Jet engine test circuit
X28J	Dumbwaiter circuit

Table 1-32.—Sound-Powered Telephone Circuits—Continued

CIRCUIT	PRIMARY CIRCUITS TITLE
	SUPPLEMENTARY CIRCUITS (CONTINUED)
X29J	Timing and recording circuit
X34J	Alignment cart service circuit
X40J	Casualty communication circuit
X41J	Special weapons shop service circuit
X42J	Missile assembly and handling circuit
X43J	Weapons system service circuit
X44J	ASROC service circuit
X45J	Special weapons security circuit
X50J	Fog foam circuit
X61J	Nuclear support facilities operations and handling circuit

Screw, Drill, and Tap Data

Table 1-33 contains machine screw information, such as threads per inch, drill, and tap information.

Table 1-33.—Screw, Drill, and Tap Data

MACHINE SCREW		THREADS PER INCH		CLEARANCE DRILL		TAP DRILL	
NO.	DIA.	COARSE	FINE	NO.	DIA.	NO.	DIA.
0	0.060		80	52	0.063	56	0.046
1	0.073	64	72	47	0.078	53	0.059
2	0.086	56	64	42	0.093	50	0.079
		48				47	0.079
3	0.099			37	0.104		
			56			45	0.082
		40				43	0.089
4	0.112			31	0.120		
			48			42	0.093
		40				38	0.101
5	0.125			29	0.136		
			44			37	0.104
		32				36	0.107
6	0.138			27	0.144		
			40			33	0.113
		32				29	0.136
8	0.164			18	0.169		
			36			29	0.136
		24				25	0.149
10	0.190			9	.196		
			32			21	0.159
		24				16	0.177
12	0.216				.228		
			28			14	0.182
		20				7	.201
1/4	0.250				17/64		
			28			3	.213

*Size for use in hand-topping brass or soft steel; for copper, aluminum, bakelite, or similar material use one size larger.

Logarithms, Common

Table 1-34 is a seven-place table of logarithms.

Table 1-34.—Common Logarithms

N	0	1	2	3	4	5	6	7	8	9
10	0000000	0043214	0086002	0128372	0170333	0211893	0253059	0293838	0334238	0374265
11	0413927	0453230	0492180	0530784	0569049	0606978	0644580	0681859	0718820	0755470
12	0791812	0327854	0863598	0899051	0934217	0969100	1003705	1038037	1072100	1105897
13	1139434	1172713	1205739	1238516	1271048	1303338	1335389	1367206	1398791	1430148
14	1461280	1492191	1522883	1553360	1583625	1613680	1643529	1673173	1702617	1731863
15	1760913	1789769	1818436	1846914	1875207	1903317	1931246	1958997	1986571	2013971
16	2041200	2068259	2095150	2121876	2148438	2174839	2201081	2227165	2253093	2278867
17	2304489	2329961	2355284	2380461	2405492	2430380	2455127	2479733	2504200	2528530
18	2552725	2576786	2600714	2624511	2648178	2671717	2695129	2718416	2741578	2764618
19	2787536	2810334	2833012	2855573	2878017	2900346	2922561	2944662	2966652	2988531
20	3010300	3031961	3053514	3074960	3096302	3117539	3138672	3159703	3180633	3201463
21	3222193	3242825	3263359	3283796	3304138	3324385	3324538	3364597	3384565	3404441
22	3424227	3443923	3463530	3483049	3502480	3521825	3541084	3560259	3579348	3598355
23	3617278	3636120	3654880	3673559	3692159	3710679	3729120	3747483	3765770	3783979
24	3802112	3820170	3838154	3856063	3873898	3891661	3909351	3926970	3944517	3961993
25	3979400	3996737	4014005	4031205	4048337	4065402	4082400	4099331	4116197	4132998
26	4149733	4166405	4183013	4199557	4216039	4232459	4248816	4265113	4281348	4297523
27	4313638	4329693	4345689	4361626	4377506	4393327	4409091	4424798	4440448	4456042
28	4471580	4487063	4502491	4517864	4533183	4548449	4563660	4578819	4593925	4608978
29	4623980	4638930	4653829	4668676	4683473	4698220	4712917	4727564	4742163	4756712
30	4771213	4785665	4800069	4814426	4828736	4842998	4857214	4871384	4885507	4899585
31	4913617	4927604	4941546	4955443	4969296	4983106	4996871	5010593	5024271	5037907
32	5051500	5065050	5078559	5092025	5105450	5118834	5132176	5145478	5158738	5171959
33	5185139	5198280	5211381	5224442	5237465	5250448	5263393	5276299	5289167	5301997
34	5314789	5327544	5340261	5352941	5365584	5378191	5390761	5403295	5415792	5428254
35	5440680	5453071	5465427	5477747	5490033	5502284	5514500	5526682	5538830	5505944
36	5563025	5575072	5587086	5599066	5611014	5622929	5634811	5646661	5658478	5670264
37	5682017	5693739	5705429	5717088	5728716	5740313	5751878	5763414	5774918	5786392
38	5797836	5809250	5820634	5831988	5843312	5854607	5865873	5877110	5888317	5899496
39	5910646	5921768	5932861	5943926	5954962	5965971	5976952	5987905	5998831	6009729
40	6020600	6031444	6042261	6053050	6063814	6074550	6085260	605944	6106602	6117233
41	6127839	6138418	6148972	6159501	6170003	6180481	6190933	6201361	6211763	6222140
42	6232493	6242821	6253125	6263404	6273659	6283889	6294096	6304279	6314438	6324573
43	6334685	6344773	6354837	6364879	6374897	6384893	6394865	6404814	6414741	6424645
44	6434527	6444386	6454223	6464037	6473830	6483600	6493349	6503075	6512780	6522463
45	6532125	6541765	6551384	6560982	6570559	6580114	6589648	6599162	6608655	6618127
46	6627578	6637099	6646420	6655810	6665180	6674530	6683859	6693169	6702459	6711728
47	6720979	6730209	6739420	6748611	6757783	6766936	6776070	6785184	6794279	6803355
48	6812412	6821451	6830470	6839471	6848454	6857417	6866363	6875290	6884198	6893089
49	6901961	6910815	6919651	6928469	6937269	6946052	6954817	6963564	6972293	6981005
50	6989700	6998377	7007037	7015680	7024305	7032914	7041505	7050080	7058637	7067178
51	7075702	7084209	7092700	7101174	7109631	7118072	7126497	7134905	7143298	7151674
52	7160033	7168377	7176705	7185017	7193313	7201593	7209857	7218106	7226339	7234557
53	7242759	7250945	7259116	7267272	7275413	7283538	7291648	7299743	7307823	7315888
54	7323938	7331973	7339993	7347998	7355989	7363965	7371926	7379873	7387806	7395723

Table 1-34.—Common Logarithms—Continued

N	0	1	2	3	4	5	6	7	8	9
55	7403627	7411516	7419391	7427251	7435098	7442930	7450748	7458552	7466342	7474118
56	7481980	7489629	7497363	7505084	7512791	7520484	7528164	7535831	7543483	7551123
57	7338749	7566361	7573960	7581546	7589119	7596678	7604225	7611758	7619278	7626786
58	7634280	7641761	7649230	7656686	7664128	7671559	7678976	7686381	7693773	7701153
59	7708520	7715875	7723217	7730547	7737864	7745170	7752463	7759743	7767012	7774268
60	7781513	7788745	7795965	7803173	7810369	7817554	7824726	7831887	7839036	7846173
61	7853298	7860412	7867514	7874605	7881684	7888751	7895807	7901852	7909885	7916906
62	7923917	7930916	7937904	7944880	7951846	7958800	7965743	7972675	7979596	7986506
63	7993405	8000294	8007171	8014037	8020893	8027737	8034571	8041394	8048207	8055009
64	8061800	8068580	8075350	8082110	8088859	8095597	8102325	8109043	8115750	8122447
65	8129134	8135810	8142476	8149132	8155777	8162413	8169038	8175654	8182259	8188854
66	8195439	8202015	8208580	8215135	8221681	8228216	8234742	8241258	8247765	8254261
67	8260748	8267225	8273693	8280151	8286599	8293038	8299467	8305887	8312297	8318698
68	8325089	8331471	8337844	8344207	8350561	8256906	8363241	8369567	8375884	8382192
69	8388491	8394780	8401061	8407332	8413505	8419848	8426092	8432328	8438554	8444772
70	8450980	8457180	8463371	8469553	8475727	8481891	8488047	8494194	8500333	8506462
71	8512583	8518696	8524800	8530895	8536982	8543060	8549130	8555192	8561244	8567289
72	8573325	8579353	8585372	8591383	8597386	8603380	8609366	8615344	8621314	8627275
73	8633229	8639174	8645111	8651040	8656961	8662873	8668778	8674675	8680564	8686444
74	8692317	8698182	8704039	8709888	8715729	8721563	8727388	8733206	8739016	8744818
75	8750613	8756399	8762178	8768950	8773713	8779470	8785218	8790959	8796692	8802418
76	8808136	8813847	8819550	8825245	8830934	8836614	8842288	8847954	8853612	8859263
77	8864907	8870544	8876173	8881795	8887410	8893017	8898617	8904210	8909796	8915375
78	8920946	8926510	8932068	8937618	8943161	8948697	8954225	8959747	8965262	8970770
79	8976271	8981765	8987252	8992732	8998205	9003671	9009131	9014583	9020029	9025468
80	9030900	9036325	9041744	9047155	9052560	9057959	9063350	9068735	9074114	9079485
81	9084850	9090209	9095560	9100905	9106244	9111576	9116902	9122221	9127533	9132839
82	9138139	9143432	9148718	9153998	9159272	9164539	9169800	9175055	9180303	9185545
83	9190781	9196010	9201233	9206450	9211661	9216865	9222063	9227255	9232440	9237620
84	9242793	9247960	9253121	9258276	9263424	9268567	9273704	9278834	9283959	9289077
85	9294189	9299296	9304396	9309490	9314579	9319661	9324738	9329808	9334873	9339932
86	9344985	9350032	9355073	9360108	9365137	9370161	9375179	9380191	9385197	9390198
87	9395193	9400182	9405165	9410142	9415114	9420081	9425041	9429996	9434945	9439889
88	9444827	9449759	9454686	9459607	9464523	9469433	9474337	9479236	9484130	9489018
89	9493900	9498777	9503649	9508515	9513375	9518230	9523080	9527924	9532763	9537597
90	9542425	9547248	9552065	9556878	9561684	9566486	9571282	9576073	9580858	9585639
91	9590414	9595184	9599948	9604708	9609462	9614211	9618955	9623693	9628427	9633155
92	9637878	9642596	9647309	9652017	9656720	9661417	9666110	9670797	9675480	9680157
93	9684829	9689497	9694159	9698816	9703469	9708116	9712758	9717396	9722028	9726656
94	9731279	9735896	9740509	9745117	9749720	9754318	9758911	9763500	9768083	9772662
95	9777236	9781805	9786369	9790929	9795484	9800034	9804579	9809119	9813655	9818186
96	9822712	9827234	9831751	9836263	9840770	9845273	9849771	9854265	9858754	9863238
97	9867717	9872192	9876663	9881128	9885590	9890046	9894498	9898946	9903389	9907827
98	9912261	9916690	9921115	9925535	9929951	9934362	9938769	9943172	9947569	9951963
99	9956352	9960737	9965117	9969492	9973864	9978231	9982593	9986952	9991305	9995655

Trigonometric Functions

Table 1-35 is a table of trigonometric functions.

Table 1-35.! Trigonometric Functions

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
0.0	.00000	1.0000	.00000		90.0	4.0	.06976	0.9976	.06993	14.301	86.0
.1	.00175	1.0000	.00175	573.0	.9	.1	.07150	.9974	.07168	13.951	.9
.2	.00349	1.0000	.00349	286.5	.8	.2	.07324	.9973	.07344	13.617	.8
.3	.00524	1.0000	.00524	191.0	.7	.3	.07498	.9972	.17519	13.300	.7
.4	.00698	1.0000	.00698	143.24	.6	.4	.07672	.9971	.07695	12.996	.6
.5	.00873	1.0000	.00873	114.59	.5	.5	.07846	.9969	.07870	12.706	.5
.6	.01047	0.9999	.10147	95.49	.4	.6	.08020	.9968	.08046	12.429	.4
.7	.01222	.9999	.01222	81.85	.3	.7	.08194	.9966	.08221	12.163	.3
.8	.01396	.9999	.01396	71.62	.2	.8	.08368	.9965	.08397	11.909	.2
.9	.01571	.9999	.01571	63.66	.1	.9	.08542	.9963	.08573	11.664	.1
1.0	.01745	0.9998	.01746	57.20	89.0	5.0	.08716	0.9962	.08749	11.430	85.0
.1	.01920	.9998	.01920	52.08	.9	.1	.08889	.9960	.089215	11.205	.9
.2	.02094	.9998	.02095	47.74	.8	.2	.09063	.9959	.09101	10.988	.8
.3	.02269	.9997	.02269	44.07	.7	.3	.09237	.9957	.09277	10.780	.7
.4	.02443	.9997	.02444	40.92	.6	.4	.09411	.9956	.09453	10.579	.6
.5	.02618	.9997	.02619	38.19	.5	.5	.09585	.9954	.09629	10.385	.5
.6	.02792	.9996	.02793	35.80	.4	.6	.09758	.9952	.09805	10.199	.4
.7	.02967	.9996	.02968	33.69	.4	.7	.09932	.9951	.09981	10.019	.3
.8	.03141	.9995	.03143	31.82	.2	.8	.10106	.9949	.10158	9.845	.2
.9	.03316	.9995	.03317	30.14	.1	.9	.10279	.9947	.10334	9.677	.1
2.0	.03490	0.9994	.03492	28.64	88.0	6.0	.10453	0.9945	.10510	9.514	84.0
.1	.03664	.9993	.03667	27.27	.9	.1	.10626	.9943	.10687	9.357	.9
.2	.03839	.9993	.03842	26.03	.8	.2	.10800	.9942	.10863	9.205	.8
.3	.04013	.9992	.04016	24.90	.7	.3	.10973	.9940	.11040	9.058	.7
.4	.04188	.9991	.04191	23.86	.6	.4	.11147	.9938	.11217	8.915	.6
.5	.04362	.9990	.04366	22.90	.5	.5	.11320	.9936	.11394	8.777	.5
.6	.04536	.9990	.04541	22.02	.4	.6	.11494	.9934	.11570	8.643	.4
.7	.04711	.9989	.04716	21.20	.3	.7	.11667	.9932	.11747	8.513	.3
.8	.04885	.9988	.04891	20.45	.2	.8	.11840	.9930	.11924	8.386	.2
.9	.05059	.9987	.05066	19.74	.1	.9	.12014	.9928	.12101	8.264	.1
3.0	.05234	0.9986	.05241	19.081	87.0	7.0	.12187	0.9925	.12278	8.144	83.0
.1	.05408	.9985	.05416	18.464	.9	.1	.12360	.9923	.12456	8.028	.9
.2	.05582	.9984	.05591	17.886	.8	.2	.12533	.9921	.12633	7.916	.8
.3	.05756	.9983	.05766	17.343	.7	.3	.12706	.9919	.12810	7.806	.7
.4	.05931	.9982	.05941	16.832	.6	.4	.12880	.9917	.12988	7.700	.6
.5	.06105	.9981	.06116	16.350	.5	.5	.13053	.9914	.13165	7.596	.5
.6	.06279	.9980	.06291	15.895	.4	.6	.13226	.9912	.13343	7.495	.4
.7	.06453	.9979	.06467	15.464	.3	.7	.13399	.9910	.13521	7.396	.3
.8	.06627	.9978	.06642	15.056	.2	.8	.13572	.9907	.13698	7.300	.2
.9	.06802	.9977	.06817	14.669	.1	.9	.13744	.9905	.13876	7.207	.1
	cos	sin	cot	tan	deg		cos	sin	cot	tan	

Table 1-35.! Trigonometric Functions! Continued

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
8.0	.13917	0.9903	.14054	7.115	82.0	12.0	0.2079	0.9781	0.2126	4.705	78.0
.1	.14090	.9900	.14232	7.026	.9	.1	.2096	.9778	.2144	4.665	.9
.2	.14263	.9898	.14410	6.940	.8	.2	.2133	.9774	.2162	4.625	.8
.3	.14436	.9895	.14588	6.855	.7	.3	.2130	.9770	.2180	4.586	.7
.4	.14608	.9893	.14767	6.772	.6	.4	.2147	.9767	.2199	4.548	.6
.5	.14781	.9890	.14945	6.691	.5	.5	.2164	.9763	.2217	4.511	.5
.6	.14954	.9888	.15124	6.612	.4	.6	.2181	.9759	.2235	4.474	.4
.7	.15126	.9885	.15302	6.535	.3	.7	.2198	.9755	.2254	4.437	.3
.8	.15299	.9882	.15481	6.460	.2	.8	.2215	.9751	.2272	4.402	.2
.9	.15471	.9880	.15660	6.386	.1	.9	.2233	.9748	.2290	4.366	.1
9.0	.15643	0.9877	.15836	6.314	81.0	13.0	0.2250	0.9744	0.2309	4.331	77.0
.1	.15816	.9874	.16017	6.243	.9	.1	.2267	.9740	.2327	4.297	.9
.2	.15988	.9871	.16196	6.174	.8	.2	.2284	.9736	.2345	4.264	.8
.3	.16160	.9869	.16376	6.107	.7	.3	.2300	.9732	.2364	4.230	.7
.4	.16333	.9866	.16555	6.041	.6	.4	.2317	.9728	.2382	4.198	.6
.5	.16505	.9863	.16734	5.976	.5	.5	.2334	.9724	.2401	4.165	.5
.6	.16677	.9860	.16914	5.912	.4	.6	.2351	.9720	.2419	4.134	.4
.7	.16849	.9857	.17093	5.850	.3	.7	.2368	.9715	.2438	4.102	.3
.8	.17021	.9854	.17273	5.789	.2	.8	.2385	.9711	.2456	4.071	.2
.9	.17193	.9851	.17453	5.730	.1	.9	.2402	.9707	.2475	4.041	.1
10.0	.1736	0.9848	.1763	5.671	80.0	14.0	0.2419	0.9703	0.2493	4.011	76.0
.1	.1754	.9845	.1781	5.614	.9	.1	.2436	.9699	.2512	3.981	.9
.2	.1771	.9842	.1799	5.558	.8	.2	.2453	.9694	.2530	3.952	.8
.3	.1788	.9839	.1817	5.503	.7	.3	.2470	.9680	.2549	3.923	.7
.4	.1805	.9836	.1835	5.449	.6	.4	.2487	.9686	.2568	3.895	.6
.5	.1822	.9833	.1853	5.396	.5	.5	.2504	.9681	.2586	3.867	.5
.6	.1840	.9829	.1871	5.343	.4	.6	.2521	.9677	.2605	3.839	.4
.7	.1857	.9826	.1890	5.292	.3	.7	.2538	.9673	.2623	3.812	.3
.8	.1874	.9823	.1908	5.242	.2	.8	.2554	.9668	.2642	3.785	.2
.9	.1891	.9820	.1926	5.193	.1	.9	.2571	.9664	.2661	3.758	.1
11.0	.1908	0.9816	.1944	5.145	79.0	15.0	0.2588	0.9659	0.2679	3.732	75.0
.1	.1925	.9813	.1962	5.097	.9	.1	.2605	.9655	.2698	3.706	.9
.2	.1942	.9810	.1980	5.050	.8	.2	.2622	.9650	.2717	3.681	.8
.3	.1959	.9806	.1998	5.005	.7	.3	.2639	.9646	.2736	3.655	.7
.4	.1977	.9803	.2016	4.959	.6	.4	.2656	.9641	.2754	3.630	.6
.5	.1994	.9799	.2035	4.915	.5	.5	.2672	.9636	.2773	3.606	.5
.6	.2011	.9796	.2053	4.872	.4	.6	.2689	.9632	.2792	3.582	.4
.7	.2028	.9792	.2071	4.829	.3	.7	.2706	.9627	.2811	3.558	.3
.8	.2045	.9789	.2089	4.787	.2	.8	.2723	.9622	.2830	3.534	.2
.9	.2062	.9785	.2107	4.745	.1	.9	.2740	.9617	.2849	3.511	.2
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

Table 1-35.! Trigonometric Functions! Continued

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
16.0	0.2756	0.9613	0.2867	3.487	74.0	20.0	0.3420	0.9397	0.3640	2.747	70.0
.1	.2773	.9608	.2886	3.465	.9	.1	.3437	.9391	.3659	2.733	.9
.2	.2790	.9603	.2905	3.442	.8	.2	.3453	.9385	.3679	2.718	.8
.3	.2807	.9598	.2924	3.420	.7	.3	.3469	.9379	.3699	2.703	.7
.4	.2823	.9593	.2943	3.398	.6	.4	.3486	.9373	.3719	2.689	.6
.5	.2840	.9588	.2962	3.376	.5	.5	.3502	.9367	.3739	2.675	.5
.6	.2857	.9583	.2981	3.354	.4	.6	.3518	.9361	.3759	2.660	.4
.7	.2874	.9578	.3000	3.333	.3	.7	.3535	.9354	.3779	2.646	.3
.8	.2890	.9573	.3019	3.312	.2	.8	.3551	.9348	.3799	2.633	.2
.9	.2907	.9568	.3038	3.291	.1	.9	.3567	.9342	.3819	2.619	.1
17.0	0.2924	0.9563	0.3067	3.271	73.0	21.0	0.3584	0.9336	0.3839	2.605	69.0
.1	.2940	.9558	.3076	3.271	.9	.1	.3600	.9330	.3859	2.592	.9
.2	.2957	.9553	.3096	3.230	.8	.2	.3616	.9323	.3879	2.578	.8
.3	.2974	.9548	.3115	3.211	.7	.3	.3633	.9317	.3899	2.565	.7
.4	.2990	.9542	.3134	3.191	.6	.4	.3649	.9311	.3919	2.552	.6
.5	.3007	.9537	.3153	3.172	.5	.5	.3665	.9304	.3939	2.539	.5
.6	.3024	.9532	.3172	3.152	.4	.6	.3681	.9298	.3959	2.526	.4
.7	.3040	.9527	.3191	3.133	.3	.7	.3697	.9291	.3979	2.513	.3
.8	.3057	.9521	.3211	3.115	.2	.8	.3714	.9285	.4000	2.500	.2
.9	.3074	.9516	.3230	3.096	.1	.9	.3730	.9278	.4020	2.488	.1
18.0	0.3090	0.9511	0.3249	3.078	72.0	22.0	0.3746	0.9272	0.4040	2.475	68.0
.1	.3107	.9505	.3269	3.060	.9	.1	.3762	.9265	.4061	2.463	.9
.2	.3123	.9500	.3288	3.042	.8	.2	.3778	.9259	.4081	2.450	.8
.3	.3140	.9494	.3307	3.024	.7	.3	.3795	.9252	.4101	2.438	.7
.4	.3156	.9489	.3327	3.006	.6	.4	.3811	.9245	.4122	2.426	.6
.5	.3173	.9483	.3346	2.989	.5	.5	.3727	.9239	.4142	2.414	.5
.6	.3190	.9478	.3365	2.971	.4	.6	.3843	.9232	.4163	2.402	.4
.7	.3206	.9472	.3385	2.954	.3	.7	.3859	.9225	.4183	2.391	.3
.8	.3223	.9466	.3404	2.937	.2	.8	.3875	.9219	.4204	2.379	.2
.9	.3239	.9461	.3424	2.921	.1	.9	.3891	.9212	.4224	2.367	.1
19.0	0.3256	0.9455	0.3443	2.904	71.0	23.0	0.3907	0.9205	0.4245	2.356	67.0
.1	.3272	.9449	.3463	2.888	.9	.1	.3923	.9198	.4265	2.344	.9
.2	.3289	.9444	.3482	2.872	.8	.2	.3939	.9191	.4286	2.333	.8
.3	.3305	.9438	.3502	2.856	.7	.3	.3955	.9184	.4307	2.322	.7
.4	.3322	.9432	.3522	2.840	.6	.4	.3971	.9178	.4327	2.311	.6
.5	.3338	.9426	.3541	2.824	.5	.5	.3987	.9171	.4348	2.300	.5
.6	.3355	.9421	.3561	2.808	.4	.6	.4003	.9164	.4369	2.289	.4
.7	.3371	.9415	.3581	2.793	.3	.7	.4019	.9157	.4390	2.278	.3
.8	.3387	.9409	.3600	2.778	.2	.8	.4035	.9150	.4411	2.267	.2
.9	.3403	.9403	.3620	2.762	.1	.9	.4051	.9143	.4431	2.257	.1
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

Table 1-35.! Trigonometric Functions! Continued

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
24.0	0.4067	0.9135	0.4452	2.246	66.0	28.0	0.4695	0.8829	0.5317	1.881	62.0
.1	.4083	.9128	.4473	2.236	.9	.1	.4710	.8821	.5340	1.873	.9
.2	.4099	.9121	.4494	2.225	.8	.2	.4726	.8813	.5362	1.865	.8
.3	.4115	.9114	.4515	2.215	.7	.3	.4741	.8805	.5384	1.857	.7
.4	.4131	.9107	.4536	2.204	.6	.4	.4756	.8796	.5407	1.849	.6
.5	.4147	.9100	.4557	2.194	.5	.5	.4772	.8788	.5430	1.842	.5
.6	.4163	.9092	.4578	2.184	.4	.6	.4787	.8780	.5452	1.834	.4
.7	.4179	.9085	.4599	2.174	.3	.7	.4802	.8771	.5475	1.827	.3
.8	.4195	.9078	.4621	2.164	.2	.8	.4818	.8763	.5498	1.819	.2
.9	.4210	.9070	.4642	2.154	.1	.9	.4833	.8755	.5520	1.811	.1
25.0	0.4226	0.9063	0.4663	2.145	65.0	29.0	0.4848	0.8746	0.5543	1.804	61.0
.1	.4242	.9056	.4684	2.135	.9	.1	.4863	.8738	.5566	1.797	.9
.2	.4258	.9048	.4706	2.125	.8	.2	.4879	.8729	.5589	1.789	.8
.3	.4274	.9041	.4727	2.116	.7	.3	.4894	.8721	.5612	1.782	.7
.4	.4289	.9033	.4748	2.106	.6	.4	.4909	.8712	.5635	1.775	.6
.5	.4305	.9028	.4770	2.097	.5	.5	.4924	.8704	.5658	1.767	.5
.6	.4321	.9018	.4791	2.087	.4	.6	.4939	.8695	.5681	1.760	.4
.7	.4337	.9011	.4813	2.078	.3	.7	.4955	.8686	.5704	1.753	.3
.8	.4352	.9003	.4834	2.069	.2	.8	.4970	.8678	.5726	1.746	.2
.9	.4368	.8996	.4856	2.059	.1	.9	.4985	.8669	.5750	1.739	.1
26.0	0.4384	0.8988	0.4877	2.050	64.0	30.0	0.5000	0.8660	0.5774	1.7321	60.0
.1	.4399	.8980	.4899	2.041	.9	.1	.5015	.8652	.5797	1.7251	.9
.2	.4415	.8973	.4921	2.032	.8	.2	.5030	.8643	.5820	1.7162	.8
.3	.4431	.8965	.4942	2.023	.7	.3	.5045	.8634	.5844	1.7113	.7
.4	.4446	.8957	.4964	2.014	.6	.4	.5040	.8625	.5867	1.7045	.6
.5	.4462	.8949	.4986	2.006	.5	.5	.5075	.8616	.5890	1.6977	.5
.6	.4478	.8942	.5008	1.997	.4	.6	.5090	.8607	.5914	1.6909	.4
.7	.4493	.8934	.5029	1.988	.3	.7	.5105	.8599	.5938	1.6842	.3
.8	.4509	.8926	.5051	1.980	.2	.8	.5120	.8590	.5961	1.6715	.2
.9	.4524	.8918	.5073	1.971	.1	.9	.5135	.8581	.5985	1.6709	.1
27.0	0.4540	0.8910	0.5095	1.963	63.0	31.0	0.5150	0.8572	0.6009	1.6643	59.0
.1	.4555	.8902	.5117	1.954	.9	.1	.5165	.8643	.6032	1.6577	.9
.2	.4571	.8894	.5139	1.946	.8	.2	.5180	.8554	.6056	1.6512	.8
.3	.4586	.8886	.5161	1.937	.7	.3	.5195	.8545	.6080	1.6447	.7
.4	.4602	.8878	.5184	1.929	.6	.4	.5210	.8536	.6104	1.6383	.6
.5	.4617	.8870	.5206	1.921	.5	.5	.5225	.8526	.6128	1.6319	.5
.6	.4633	.8862	.5228	1.913	.4	.6	.5240	.8517	.6152	1.6255	.4
.7	.4648	.8854	.5250	1.905	.3	.7	.5255	.8508	.6176	1.6191	.3
.8	.4664	.8846	.5272	1.897	.2	.8	.5270	.8499	.6200	1.6128	.2
.9	.4679	.8838	.5295	1.889	.1	.9	.5284	.8490	.6224	1.6066	.1
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

Table 1-35.! Trigonometric Functions! Continued

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
32.0	0.5299	0.8480	0.6249	1.6003	58.0	36.0	.05878	0.8090	0.7265	1.3764	54.0
.1	.5314	.8471	.6273	1.5941	.9	.1	.5892	.8080	.7292	1.3713	.9
.2	.5329	.8462	.6297	1.5880	.8	.2	.5906	.8070	.7319	1.3663	.8
.3	.5344	.8453	.6322	1.5818	.7	.3	.5920	.8059	.7346	1.3613	.7
.4	.5358	.8443	.6346	1.5757	.6	.4	.5934	.8049	.7373	1.3564	.6
.5	.5373	.8434	.6371	1.5697	.5	.5	.5948	.8039	.7400	1.3514	.5
.6	.5388	.8425	.6395	1.5637	.4	.6	.5962	.8028	.7427	1.3465	.4
.7	.5402	.8415	.6420	1.5577	.3	.7	.5976	.8018	.7454	1.3416	.3
.8	.5417	.8406	.6445	1.5517	.2	.8	.5990	.8007	.7481	1.3367	.2
.9	.5432	.8396	.6469	1.5458	.1	.9	.6004	.7997	.7508	1.3319	.1
33.0	0.5446	0.8387	0.6494	1.5399	57.0	37.0	0.6018	0.7986	0.7536	1.3270	53.0
.1	.5461	.8377	.6519	1.5340	.9	.1	.6032	.7976	.7563	1.3222	.9
.2	.5476	.8368	.6544	1.5282	.8	.2	.6046	.7965	.7590	1.3175	.8
.3	.5490	.8358	.6569	1.5224	.7	.3	.6060	.7955	.7518	1.3127	.7
.4	.5505	.8348	.6594	1.5166	.6	.4	.6074	.7944	.7646	1.3079	.6
.5	.5519	.8339	.6619	1.5108	.5	.5	.6088	.7934	.7673	1.3032	.5
.6	.5534	.8329	.6644	1.5051	.4	.6	.6101	.7923	.7701	1.2985	.4
.7	.5548	.8320	.6669	1.4994	.3	.7	.6115	.7912	.7729	1.2938	.3
.8	.5563	.8310	.6694	1.4938	.2	.8	.6129	.7902	.7757	1.2892	.2
.9	.5577	.8300	.6720	1.4882	.1	.9	.6143	.7891	.7785	1.2846	.1
34.0	0.5592	0.8290	0.6745	1.4826	56.0	38.0	0.6157	0.7880	0.7813	1.2799	52.0
.1	.5606	.8281	.6771	1.4770	.9	.1	.6170	.7869	.7841	1.2753	.9
.2	.5621	.8271	.6796	1.4715	.8	.2	.6184	.7859	.7869	1.2708	.8
.3	.5635	.8261	.6822	1.4659	.7	.3	.6198	.7848	.7898	1.2662	.7
.4	.5650	.8251	.6847	1.4605	.6	.4	.6211	.7837	.7926	1.2617	.6
.5	.5664	.8241	.6873	1.4550	.5	.5	.6225	.7826	.7954	1.2572	.5
.6	.5678	.8231	.6899	1.4496	.4	.6	.6239	.7815	.7983	1.2527	.4
.7	.5693	.8221	.6924	1.4442	.3	.7	.6252	.7804	.8012	1.2482	.3
.8	.5707	.8211	.6950	1.4388	.2	.8	.6266	.7793	.8040	1.2437	.2
.9	.5721	.8202	.6970	1.4335	.1	.9	.6280	.7782	.8069	1.2393	.1
35.0	0.5736	0.8192	0.7002	1.4281	55.0	39.0	0.6293	0.7771	0.8098	1.2349	51.0
.1	.5750	.8181	.7028	1.4229	.9	.1	.6307	.7760	.8127	1.2305	.9
.2	.5764	.8171	.7054	1.4176	.8	.2	.6320	.7749	.8156	1.2261	.8
.3	.5779	.8161	.7080	1.4124	.7	.3	.6334	.7738	.8185	1.2218	.7
.4	.5793	.8151	.7107	1.4071	.6	.4	.6347	.7727	.8214	1.2174	.6
.5	.5807	.8141	.7133	1.4019	.5	.5	.6361	.7716	.8243	1.2131	.5
.6	.5821	.8131	.7159	1.3968	.4	.6	.6474	.7705	.8273	1.2088	.4
.7	.5835	.8121	.7186	1.3916	.3	.7	.6388	.7694	.8302	1.2045	.3
.8	.5850	.8111	.7212	1.3865	.2	.8	.6401	.7683	.8332	1.2002	.2
.9	.5864	.8100	.7239	1.3814	.1	.9	.6414	.7672	.8361	1.1960	.1
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

Table 1-35.! Trigonometric Functions! Continued

deg	sin	cos	tan	cot		deg	sin	cos	tan	cot	
40.0	0.6428	0.7660	0.8291	1.1918	50.0	43.0	0.6820	0.7314	0.9325	1.0724	47.0
.1	.6441	.7649	.8421	1.1875	.9	.1	.6833	.7302	.9358	1.0686	.9
.2	.6455	.7638	.8451	1.1833	.8	.2	.6845	.7290	.9391	1.0649	.8
.3	.6468	.7627	.8481	1.1792	.7	.3	.6858	.7278	.9424	1.0612	.7
.4	.6481	.7615	.8511	1.1750	.6	.4	.6871	.7266	.9457	1.0575	.6
.5	.6494	.7604	.8541	1.1708	.5	.5	.6884	.7254	.9490	1.0538	.5
.6	.6508	.7593	.8571	1.1667	.4	.6	.6896	.7242	.9523	1.0501	.4
.7	.6521	.7581	.8601	1.1626	.3	.7	.6909	.7230	.9556	1.0464	.3
.8	.6534	.7570	.8632	1.1585	.2	.8	.6921	.7218	.9590	1.0428	.2
.9	.6547	.7559	.8662	1.1544	.1	.9	.6934	.7206	.9623	1.0392	.1
41.0	0.6561	0.7547	0.8693	1.1504	49.0	44.0	0.6947	0.7193	0.9657	1.0355	46.0
.1	.6574	.7536	.8724	1.1463	.9	.1	.6959	.7181	.9691	1.0319	.9
.2	.6587	.7524	.8754	1.1423	.8	.2	.6972	.7169	.9725	1.0283	.8
.3	.6600	.7513	.8785	1.1383	.7	.3	.6984	.7157	.9759	1.0247	.7
.4	.6613	.7501	.8816	1.1343	.6	.4	.6997	.7145	.9793	1.0212	.6
.5	.6626	.7490	.8847	1.1303	.5	.5	.7009	.7133	.9827	1.0176	.5
.6	.6639	.7478	.8878	1.1263	.4	.6	.7022	.7120	.9861	1.0141	.4
.7	.6652	.7466	.8910	1.1224	.3	.7	.7034	.7108	.9896	1.0105	.3
.8	.6665	.7455	.8941	1.1184	.2	.8	.6794	.7337	.9260	1.0799	.2
.9	.6678	.7443	.8972	1.1145	.1	.9	.6807	.7325	.9293	1.0761	.1
42.0	0.6691	0.7431	0.9004	1.1106	48.0						
.1	.6704	.7420	.9036	1.1067	.9						
.2	.6717	.7408	.9067	1.1028	.8						
.3	.6730	.7396	.9099	1.0990	.7						
.4	.6743	.7385	.9131	1.0951	.6						
.5	.6756	.7373	.9163	1.0913	.5						
.6	.6769	.7361	.9195	1.0875	.4						
.7	.6782	.7349	.9228	1.0837	.3						
.8	.6794	.7337	.9260	1.0799	.2						
.9	.6807	.7325	.9293	1.0761	.1						
	cos	sin	cot	tan	deg		cos	sin	cot	tan	deg

Julian Date Calendar

Table 1-36 is a Julian date calendar. In leap years you should add one day after 28 February. Leap years occur every four years. The last leap year was 1992; therefore, the next few leap years will be 1996, 2000, and 2004.

Table 1-36.—Julian Date Calendar

Day	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Day
1	001	032	060	091	121	152	182	213	244	274	305	335	1
2	002	033	061	092	122	153	183	214	245	275	306	336	2
3	003	034	062	093	123	154	184	215	246	276	307	337	3
4	004	035	063	094	124	155	185	216	247	277	308	338	4
5	005	036	064	095	125	156	186	217	248	278	309	339	5
6	006	037	065	096	126	157	187	218	249	279	310	340	6
7	007	038	066	097	127	158	188	219	250	280	311	341	7
8	008	039	067	098	128	159	189	220	251	281	312	342	8
9	009	040	068	099	129	160	190	221	252	282	313	343	9
10	010	041	069	100	130	161	191	222	253	283	314	344	10
11	011	042	070	101	131	162	192	223	254	284	315	345	11
12	012	043	071	102	132	163	193	224	255	285	316	346	12
13	013	044	072	103	133	164	194	225	256	286	317	347	13
14	014	045	073	104	134	165	195	226	257	287	318	348	14
15	015	046	074	105	135	166	196	227	258	288	319	349	15
16	016	047	075	106	136	167	197	228	259	289	320	350	16
17	017	048	076	107	137	168	198	229	260	290	321	351	17
18	018	049	077	108	138	169	199	230	261	291	322	352	18
19	019	050	078	109	139	170	200	231	262	292	323	353	19
20	020	051	079	110	140	171	201	232	263	293	324	354	20
21	021	052	080	111	141	172	202	233	264	294	325	355	21
22	022	053	081	112	142	173	203	234	265	295	326	356	22
23	023	054	082	113	143	174	204	235	266	296	327	357	23
24	024	055	083	114	144	175	205	236	267	297	328	358	24
25	025	056	084	115	145	176	206	237	268	298	329	359	25
26	026	057	085	116	146	177	207	238	269	299	330	360	26
27	027	058	086	117	147	178	208	239	270	300	331	361	27
28	028	059	087	118	148	179	209	240	271	301	332	362	28
29	029	*	088	119	149	180	210	241	272	302	333	363	29
30	030		089	120	150	181	211	242	273	303	334	364	30
31	031		090		151		212	243		304		365	31

* In leap year, after February 28, add 1 to the tabulated number.

Windchill Factor

The windchill factor is a computation of the still-air temperature that would have the same cooling effect on exposed human skin as a given combination of temperature and wind speed. You should use table 1-37 as a computation chart to figure windchill factor.

Table 1-37.—Windchill Factors

ESTIMATED WIND SPEED (IN MPH)	ACTUAL THERMOMETER READING (°F)											
	50	40	30	20	10	0	-10	-20	-30	-40	-50	-60
	EQUIVALENT TEMPERATURE (°F)											
CALM	50	40	30	20	10	0	-10	-20	-30	-40	-50	-60
05	48	37	27	16	6	-5	-15	-26	-36	-47	-57	-68
10	40	28	16	4	-9	-24	-33	-46	-58	-70	-83	-95
15	36	22	9	-5	-18	-32	-45	-58	-72	-85	-99	-112
20	32	18	4	-10	-25	-39	-53	-67	-82	-96	-110	-124
25	30	16	0	-15	-29	-44	-59	-74	-88	-104	-118	-133
30	28	13	-2	-18	-33	-48	-63	-79	-94	-109	-125	-140
35	27	11	-4	-20	-35	-51	-67	-82	-98	-113	-129	-145
40	26	10	-6	-21	-37	-53	-69	-85	-100	-116	-132	-148
Wind speeds greater than 40 mph have little added effect.	LITTLE DANGER (for properly clothed person) Maximum danger of false sense of securitly.				INCREASING DANGER Danger from freezing of exposed flesh			GREAT DANGER				

Effects of Heat and Humidity

Humidity combines with heat to create a more uncomfortable apparent temperature. By using table 1-38 you can figure the apparent temperature caused by various combinations of air temperature and humidity. Remember, in heat waves the apparent temperatures may run 15 to 30 degrees higher in more humid areas.

Table 1-38.—Effects of Heat and Humidity

	Air Temperature° 70	75	80	85	90	95	100	105	110	115	120
Relative Humidity	Apparent Temperature°										
0%	64	69	73	78	83	87	91	95	99	103	107
10%	65	70	75	80	85	90	95	100	105	111	116
20%	66	72	77	82	87	93	99	105	112	120	130
30%	67	73	78	84	90	96	104	113	123	135	148
40%	68	74	79	86	93	101	110	123	137	151	
50%	69	75	81	88	96	107	120	135	150		
60%	70	76	82	90	100	114	132	149			
70%	70	77	85	93	106	124	144				
80%	71	78	86	97	113	136					
90%	71	79	88	102	122						
100%	72	80	91	108							°Degrees Fahrenheit

When apparent temperatures are between 90 and 105 degrees, heat cramps, heat exhaustion, and heatstroke are possible after prolonged exposure and physical activity. These become likely when apparent temperatures are between 105 and 130 degrees. Over 130 degrees, heatstroke is imminent. You should note that heatstroke can be fatal if medical care is delayed.

GENERAL MAINTENANCE

The general maintenance section provides you with information on cleaning solvents, lubricants, corrosion control, use of the oscilloscope, troubleshooting, classes of overhaul, and types of equipment modifications. It also provides information such as material identification, names of organizations that provide outside technical assistance, and publications and documents that will assist you in day-to-day maintenance.

Corrosion Control (Cleaning and Lubricating)

A corrosive atmosphere can damage unprotected electric and electronic equipment. You should be aware of the harmful effects of moisture and, in particular, salt spray and salt-impregnated air. To prevent corrosion, you should maintain an effective cleaning and lubricating schedule. Standard preventive maintenance (PMS) procedures provide only minimum protection. Any schedule should include dusting and cleaning, lubrication of moving parts, and the use of approved solvents or wetting agents to remove any dust, dirt, oil film, salt, or other contaminant.

Table 1-39 is a list of standard Navy lubricants and solvents and their uses, as specified in Military Standard 454M (MIL-STD-454M).

Table 1-39.—Standard Navy Lubricants and Solvents

SPECIFICATION NUMBER AND TITLE	UNIT OF ISSUE	GENERAL USE
W-P-236 Petrolatum, Technical	1 lb. can 5 lb. can	For use as a light grade of lubricating grease but not recommended for use as a lubricant in heavily loaded or hot running bearings. It may be used as a constituent in certain types of rust preventive compounds.
P-D-680 Dry Cleaning Solvent	5 gal. pail	For general cleaning of air filters, electronic equipment, and other general purpose cleanup.
MIL-G-23827 Grease, Aircraft and Instrument	1 oz. tube 4 oz. tube 8 oz. Tube 1 lb. can 5 lb. can 35 lb. pail	In ball, roller, needle bearings, gears and sliding and rolling surfaces of such equipment as instruments, cameras, electronic gear and aircraft control systems. Particularly suitable for equipment which must operate at both very low and very high temperatures for short periods. Does not contain extreme pressure or special antiwear additives. It is destructive to paint, natural rubber, and neoprene.
MIL-G-81322 Grease, Aircraft	5 lb. can 35 lb. pail	For lubrication and protection against corrosion of plain ball and roller bearings, and preservation of threads on ammunitions.
MIL-L-17331 Lubricating Oil Steam Turbine	5 gal. 55 gal.	In main turbines and gears, auxiliary turbine installation, certain hydraulic equipment general mechanical lubrication, and air compressors.
MIL-L-2105 Lubrication Oil Gear	5 gal. 1 gal. 55 gal.	For lubrication of automotive gear units, heavy duty industrial-type enclosed gear units, steering gears, and fluid-lubricated universal joints of automotive equipment.
MIL-L-6085 Lubricating Oil Instrument	1 1/2 oz. btl. 4 oz. can 1 qt.	For aircraft instruments, electronic equipment, or where a low evaporation oil is required for both high and low temperature application, and where oxidation and corrosion resistance are desirable. Destructive to paint, neoprene and rubber.
MIL-L-6086 Lubricating Oil Gear	1 gal. can 1 pt. can 1 gal. can 5 gal. drum	For use under extremely low temperature, mild extreme pressure-type oil with load carrying additive. General use in aircraft use in aircraft gear mechanisms, exclusive of engines.
MIL-L-17331 & MIL-L-17672 Lubricating Oil General Purpose	1 pt. 5 gal. 55 gal.	For all applications which require other than special lubricants, and which are subject to normal variation between ambient and operating temperature. Use in lieu of MIL-L-6085 when oil will be in contact with neoprene.

Table 1-40.—Old and New Specification Solvents

OLD MILITARY SPECIFICATION	NEW MILITARY SPECIFICATION	REFERENCE
14-P-1	VV-P-236	See Table 1-39
14-L-3	MIL-G-18709	See Table 1-39
14-G-10	MIL-G-16908	See Table 1-39
14-L-11	VV-G-632	See Table 1-39
14-O-12	VV-I-530	See Table 1-39
14-O-13	MIL-L-9000	See Table 1-39
14-O-15	MIL-L-17331	See Table 1-39
14-O-20	MIL-L-6085	See Table 1-39
AN-O-6a	MIL-L-7870	See Table 1-39
KS 7470	MIL-L-17672	See Table 1-39
MIL-S-16067	P-D-680	See Table 1-39
VV-0-401	VV-I-530	See Table 1-39
P-S-661	P-D-680	See Table 1-39
MIL-G-3545	MIL-G-81322	See Table 1-39
MIL-G-3278	MIL-G-23827	See Table 1-39

Table 1-41.—Manufacturer's Designations

MANUFACTURER DESIGNATION	MILITARY SPECIFICATION	UNIT OF ISSUE
Lubri-Plate No. 105	None	2 oz.
Lubri-Plate No. 110	None	1 lb.
Molykote "G"	None	1 lb.
Molykote M-77	None	1 lb.
Stoddard Solvent	P-D-680	
140-F	P-D-680	
MOS, Lube-Power	MIL-M-7866	10 oz.
GE 10C	VV-I-530	
GE SS4005	MIL-S-8660	1 oz.
Dow-Corning DC-4	MIL-S-8660	1 oz.
McLube MOS_2-210 (formerly MOS_2-200)	None	As Requested
McLube MOS_2 - 1118	None	As Requested
Thermotex 000	None	1 lb.

Table 1-42.—Lubricants Used in Electronics Equipment But Not Listed In MIL-STD-454M

SPECIFICATION NUMBER AND TITLE	UNIT OF ISSUE	GENERAL USE
51-F-23 Hydraulic Fluid	5 gal. 55 gal.	Used in connection with the hydraulic transmission of power. For use with Synthetic Seal.
ASTM D-3699 Kerosene	5 gal.	General uses such as a cleaner for machinery or tools.
MIL-L-7870 Lubrication Oil General Purpose	4 oz. 1 qt. 1 gal.	Specially designed for use where an oil of low evaporation, possessing rust-protective properties, is desired.
VV-G-632 General Purpose Grease	35 lb. 100 lb.	Automotive chassis, suitable for lubrication of machinery equipped with pressure grease fitting.
MIL-G-81322.Grease Aircraft	1 lb. 8 oz.	Used in antifriction bearings operating at high speeds and high temperatures.
MIL-C-11090 Cleaning Compound	5 gal. 55 gal.	Used as a solvent for cleaning grease and oils.
MIL-L-17672 Lubrication Oil General Purpose	1 gal. 55 gal.	Used in steam turbines, hydraulic systems, water generators and hydraulic turbine governors.
VV-L-751 Lubrication oil	35 lb.	Cold weather. Warm weather. Hot weather. Used for lubricating chain, wire rope, exposed gears.

Table 1-43 contains a list of common cleaning and preservation materials that were compiled from Naval Air Technical Manual 16-1-540, *Avionic Cleaning and Corrosion Prevention Control*.

Table 1-43.—Cleaning Materials

Non-abrasive cleaning & polishing pad
Isopropyl alcohol TT-I-735
General purpose lubricating oil VV-L-800
Instrument grease MIL-G-8137
General purpose grease MIL-G-81322
Zip-lock plastic bags
Distilled water
Paint brush
Toothbrush
Pipe cleaners
Q tips
Face shield
Goggles
Rubber gloves
Magnifying glass
Vacuum cleaner
Hot air gun
Inspection mirror
Rubber bucket
Plastic spray bottle

Remember to use the proper safety precautions applicable to toxic, volatile solvents and flammable lubricants. You can refer to Naval Ships' Technical Manual (NSTM), Chapter 670, *Stowage, Handling, and Disposal of Hazardous General Use Consumables*, NAVSEA S9086-WK-STM-000.

Using the Oscilloscope

An oscilloscope can be used for more than just studying the shape of a waveform. By looking at lissajous patterns and using an *octopus*, you can compare the phase and frequency relationship of two signals and check electronic components in a circuit.

LISSAJOUS PATTERNS.—The simplest lissajous patterns are produced by two sine waves of the same frequency and amplitude being applied to the horizontal and vertical deflection voltage inputs of an oscilloscope. Figure 1-37 shows patterns for several common phase relationships. These can be used to estimate the approximate phase angle of the two signals being studied.

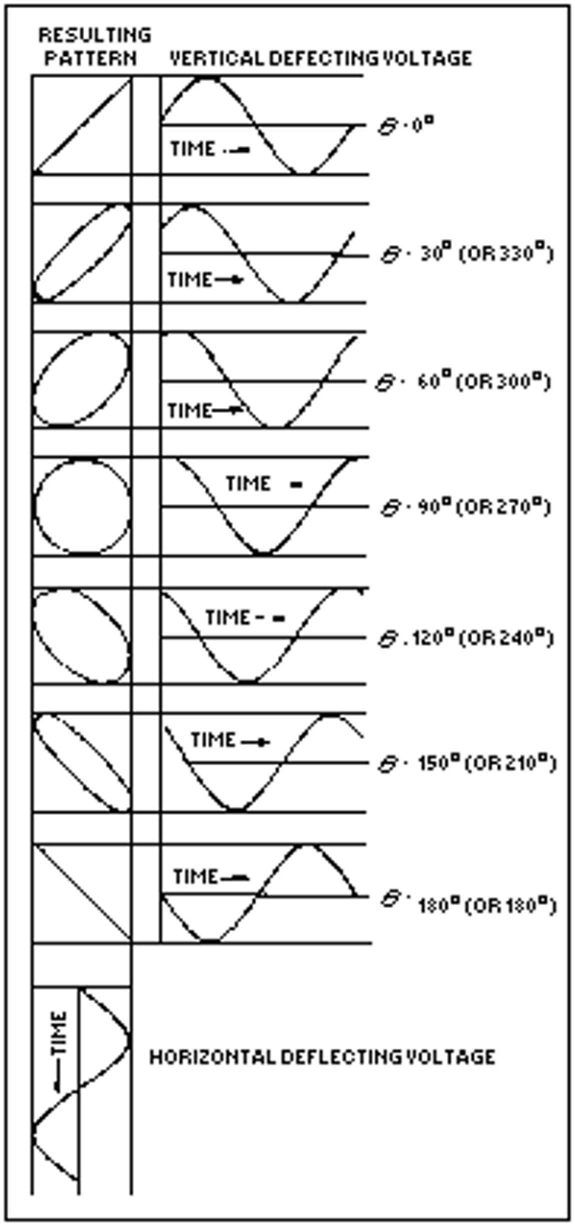


Figure 1-37.—Lissajous patterns, showing the effects of phase relationships.

Figure 1-38 will aid you in computing a phase angle if a more precise calculation is needed. We will use the graticule on the oscilloscope, a ratio formula, and sine (sin) table to compute the angle.

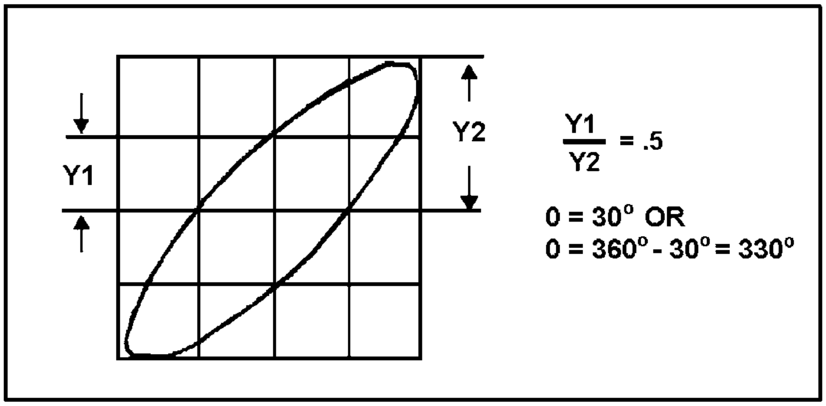


Figure 1-38.—Computation of phase angle.

To find the angle, we should first divide Y1 by Y2. We can then take that number, look it up in the sine portion of table 1-35, and read the angle.

For example, let's let each graticule in figure 1-36 represent 1 centimeter. Then,

$$Y1 = 1$$

$$Y2 = 2$$

$$\frac{Y1}{Y2} = .5$$

If we look for .5 in the sine column of table 1-35, we find that .5 is the value for the sine of 30 degrees.

The frequency ratio between two sine waves can also be determined from lissajous patterns. Figure 1-39 shows various frequency ratios between signals. Figure 1-40 and 1-41, views A, show how phase relationship can affect these patterns. If tangent lines are drawn across the top and down the side of the pattern, the ratio of points (free ends and loops) that touch these lines equals the frequency ratio. Figure 1-41 is an example of this method. Refer back to figure 1-39 and notice the relationship of loops and open ends in each example. You can find more detailed information on lissajous patterns in the Electronics Installation and Maintenance Book (EIMB), *Test Methods and Practices*, NAVSEA 0976-LP-000-0130.

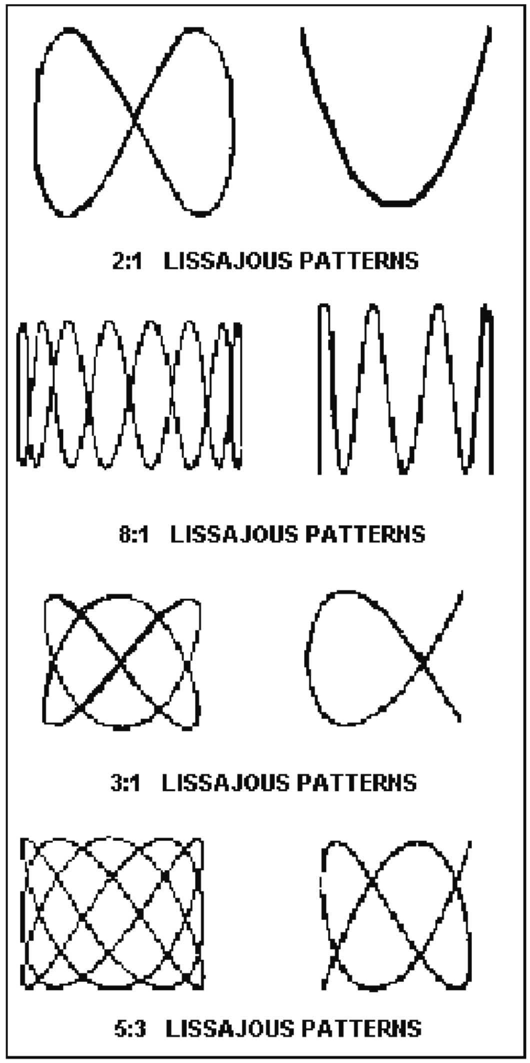


Figure 1-39.—Lissajous patterns of different frequency ratios.

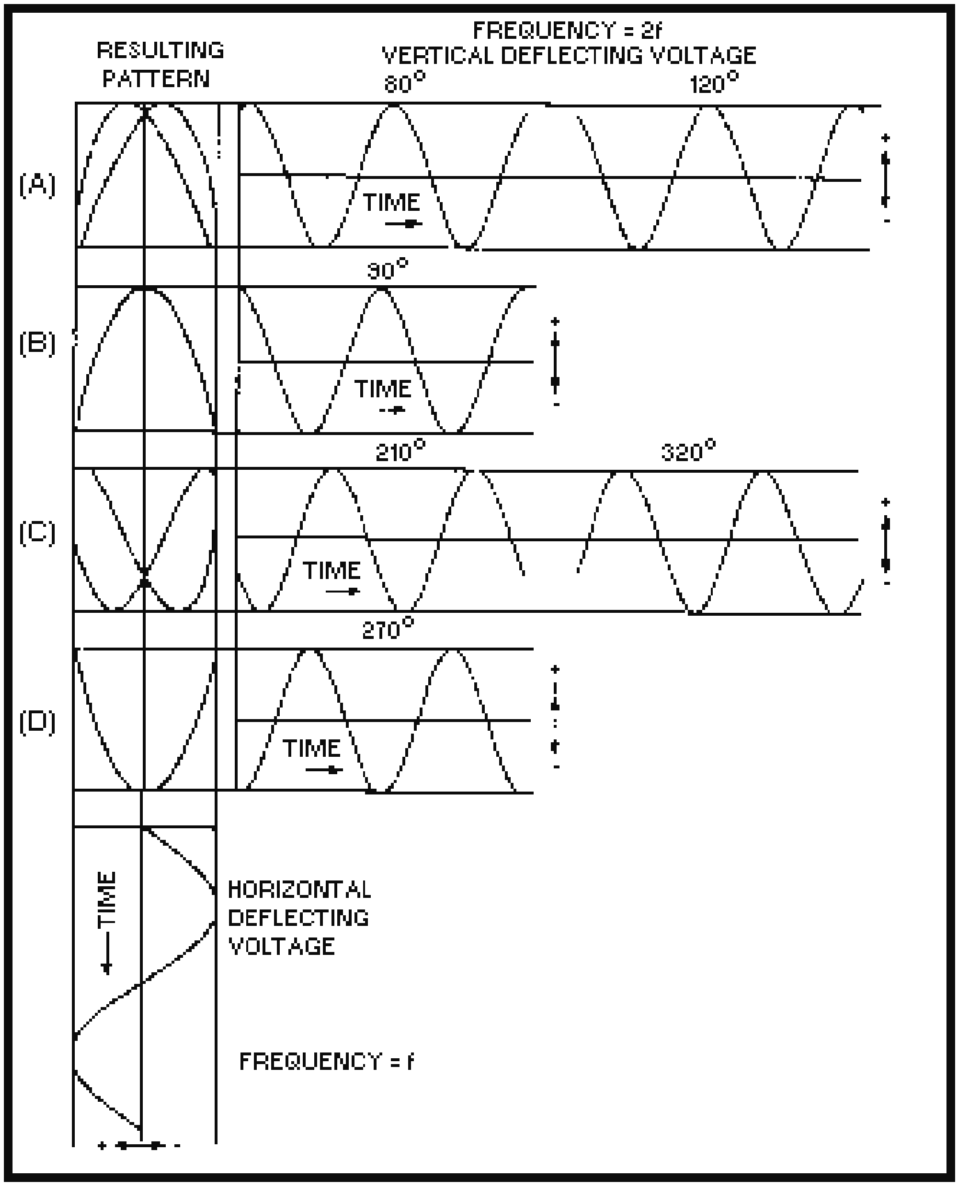


Figure 1-40.—Lissajous patterns for various phase relationships.

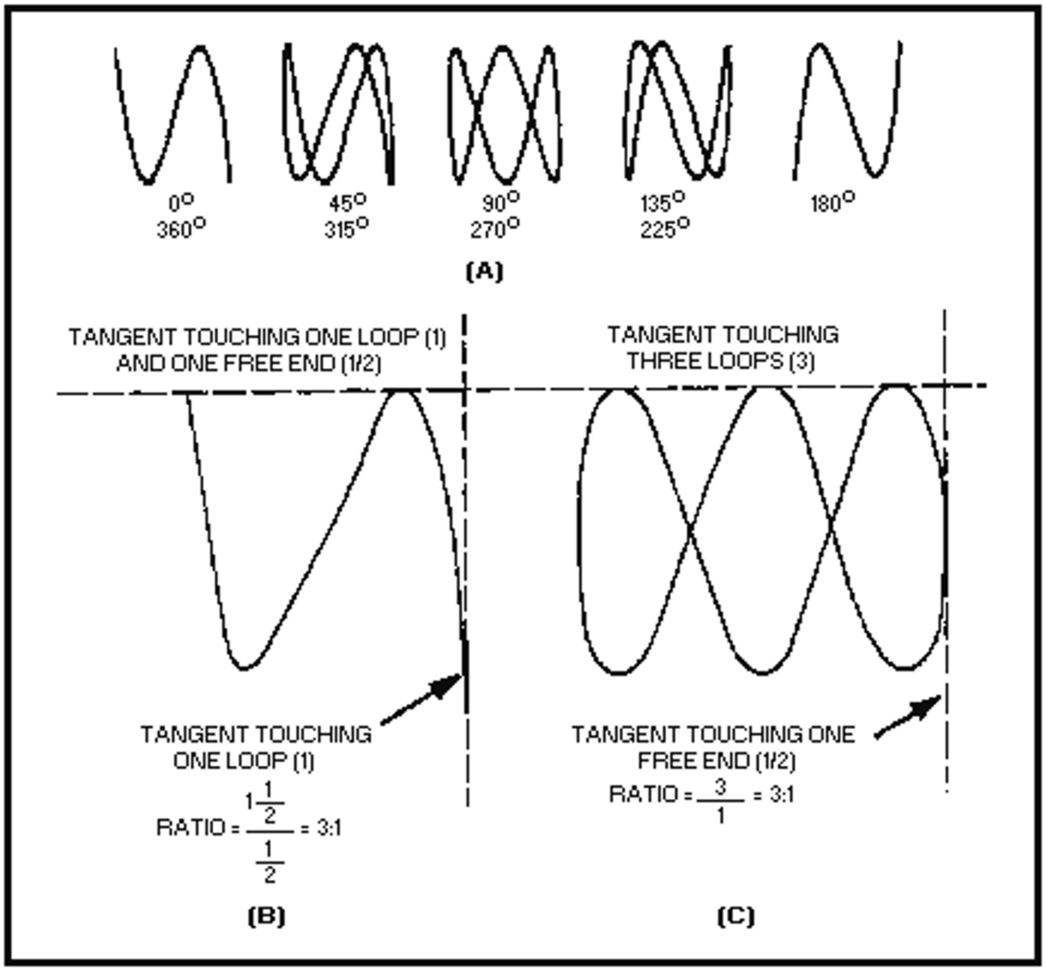


Figure 1-41—3:1 Lissajous patterns and calculation of frequency ratio.

THE OCTOPUS.—The octopus is a small, homemade test set used with an oscilloscope to check electronic components *in circuit*. It can be made easily and cheaply using parts from the supply system. Figure 1-42 is a schematic of an octopus that uses either a 6.3-volt filament transformer or an audio oscillator for input power. The benefits of *in circuit* troubleshooting with an octopus are (1) reduced maintenance time, (2) less chance of damage from soldering-iron heat, and (3) a visual display of the component's condition.

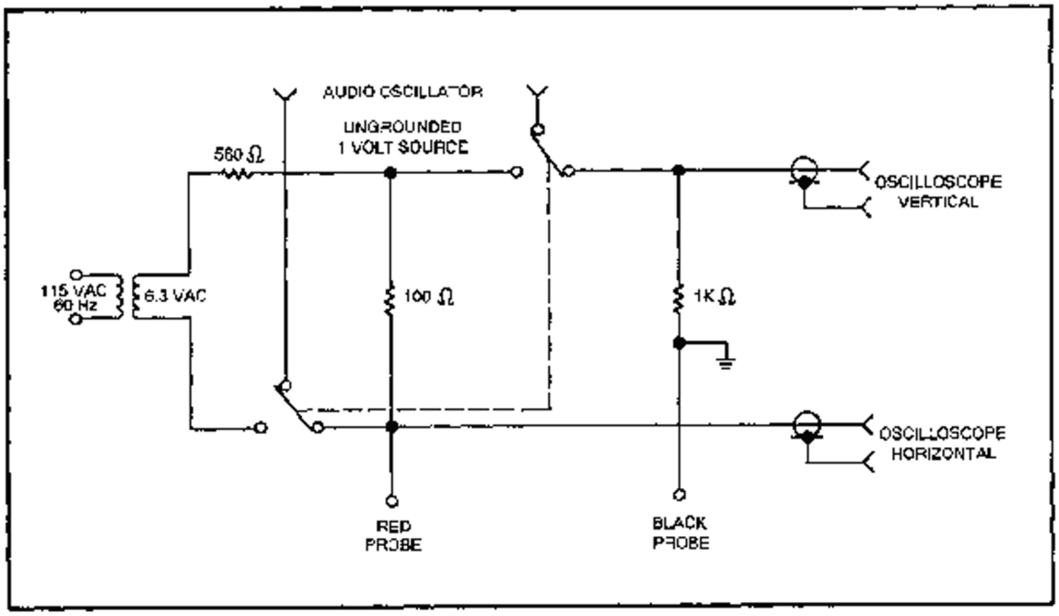


Figure 1-42.—Octopus schematic diagram (typical).

The octopus tests all components for shorts, high resistance, and opens; it checks front-to-back ratios on junction components (transistors and diodes); and it analyzes ICs and reactive components (capacitors and inductors). Figure 1-43 shows some typical oscilloscope displays obtained when the octopus is used. Figures 1-44, 1-45, and 1-46 depict transistor, potentiometer, and combination component displays, respectively. Detailed operating procedures can be found in topic 6 of NEETS, Module 16, *Introduction to Test Equipment*, and in the Electronic Installation and Maintenance Book (EIMB), *Test Methods and Practices*, NAVSEA SE000-00-EIM-130.

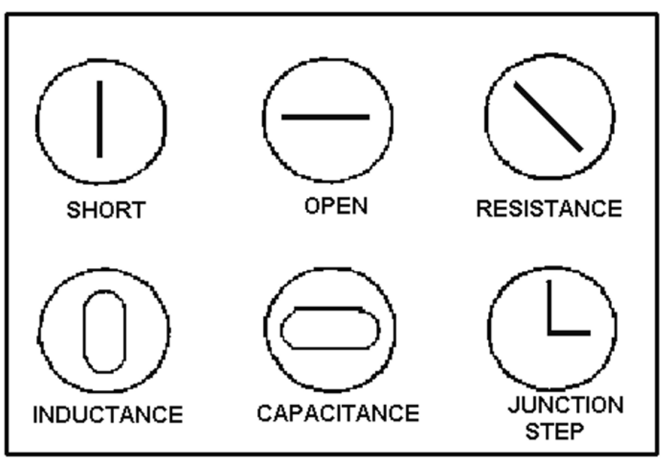


Figure 1-43.—Typical oscilloscope displays for an octopus.

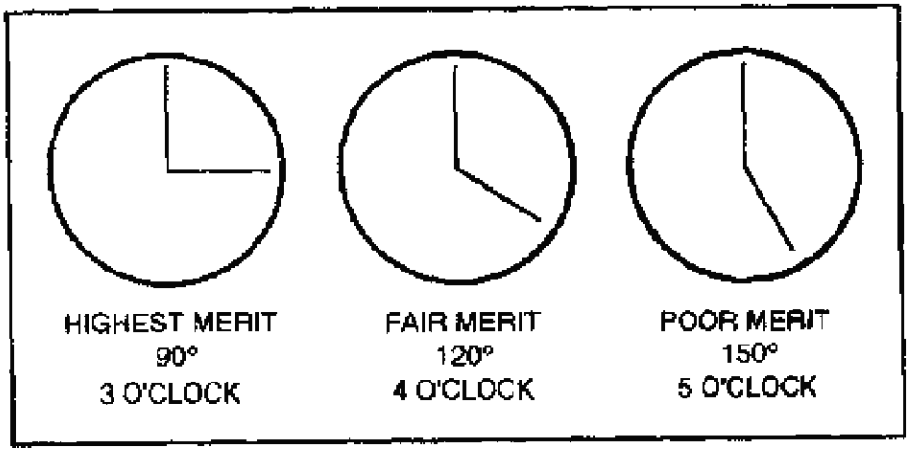


Figure 1-44.—Transistor check, single junction.

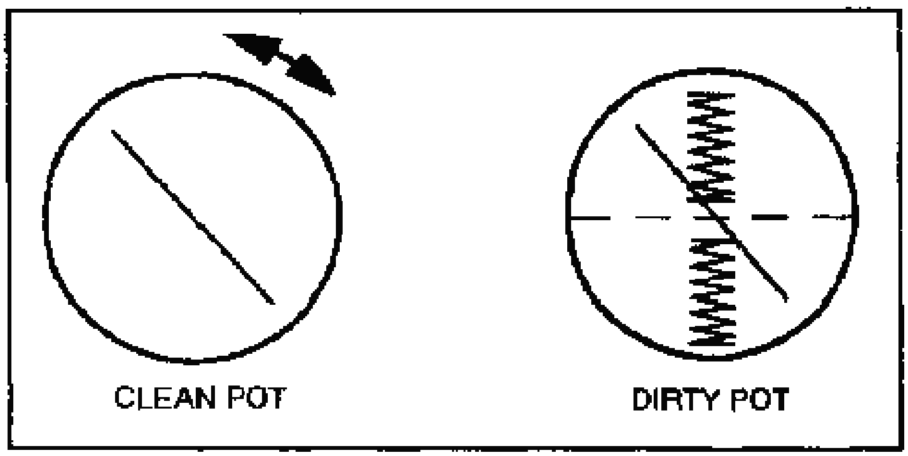


Figure 1-45.—Potentiometer noise check.

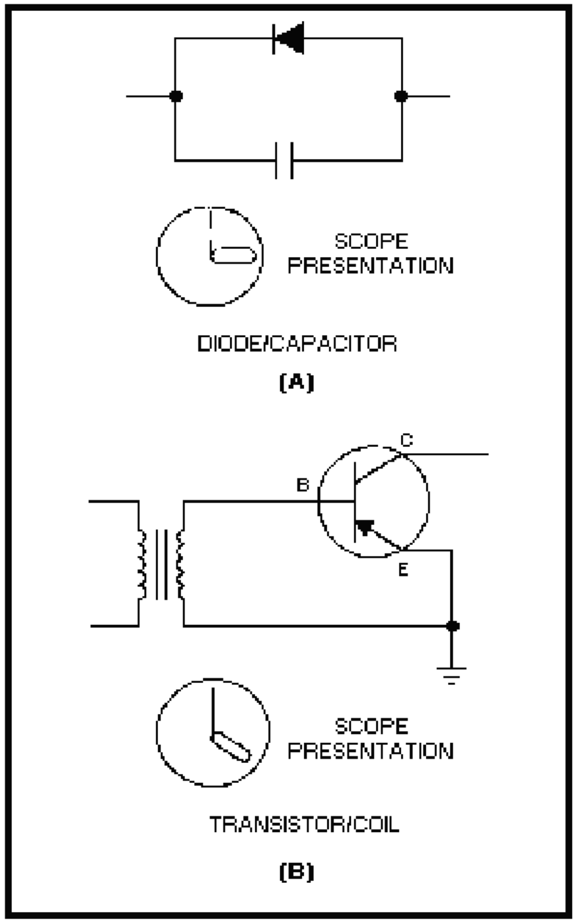


Figure 1-46.—Combination displays.

Six-Step Troubleshooting Procedure

You may have the job of maintaining or helping to maintain some electrical or electronic unit, subsystem, or system. Some of these jobs may be complex, but even a complex job can be broken down into simple steps. Basically, any repair of electric or electronic equipment should be done in the following order:

1. Symptom recognition. This is the action of recognizing some disorder or malfunction in electronic equipment.

2. Symptom elaboration. Obtaining a more detailed description of the trouble symptom is the purpose of this step.

3. Listing probable faulty functions. This step is applicable to equipment that contains more than one functional area or unit. From the information you have gathered, where could the trouble logically be located?

4. Localizing the faulty function. In this step you determine which of the functional units of the multiunit equipment is actually at fault.

5. Localizing trouble to the circuit. You will do extensive testing in this step to isolate the trouble to a specific circuit.

6. Failure analysis. This step is multipart. Here you determine which part is faulty, repair/replace the part, determine what caused the failure, return the equipment to its proper operating status, and record the necessary information in a recordkeeping book for other maintenance personnel in the future. While not a part of this step, the technician should reorder any parts used in repair of the faulty equipment.

Sometimes you may run into difficulty in finding (or troubleshooting) the problem. Some hints that may help in your efforts are:

- Observe the equipment's operation for any and all faults
- Check for any defective components with your eyes and nose
- Analyze the cause of the failure for a possible underlying problem

Classes of Overhaul Work

There are five classes of equipment overhaul (A, B, C, D, and E). The class defines the type and scope of work to be done on each equipment by the overhauling activity. (Do not confuse equipment overhaul with the term regular overhaul.)

CLASS A OVERHAUL.—A class A overhaul includes overhaul, repair, and/or modification; for example a modification could be an Ordnance Alteration (ORDALT), Special Program Alteration (SPALT), Ship Alteration (SHIPALT), or a field change that will sustain or improve the performance of a system or component to meet its most-recent design and technical specifications. The end product should be like new in appearance and operation.

CLASS B OVERHAUL.—A class B overhaul includes overhaul and repair that will restore the performance of a system or component to its original design and technical specifications. Modifications or alterations are not done unless specified by the customer.

CLASS C OVERHAUL.—A class C overhaul includes only repair work on a system or component specified by a work request or work required to correct malfunctions specified by the customer.

CLASS D OVERHAUL.—A class D overhaul includes work related to the open, inspect and report type of work request. It is intended to be diagnostic in nature and may require various tests. It is normally associated with preoverhaul test and inspection (POT& I).

CLASS E OVERHAUL.—A class E overhaul includes work required to incorporate all alterations and/or modifications specified for a system or component.

Alterations and Modifications to Equipment

Alterations and modifications to shipboard systems and equipment may take several forms. Some of these are Ship Alterations (SHIPALTS), Ordnance Alterations (ORDALTS), Special Program Alterations (SPALTS), and Air Alterations (AIRALTS). These alterations (with the exception of electronic equipment field changes) are categorized as follows:

- A military alteration that changes or improves the operational or military characteristics of a ship.

- A technical alteration that generally concerns personnel safety and equipment effectiveness.

- An alteration-equivalent-to-repair (AER) could be one of three types. One involves substitution, without change in design, of approved, different material, available from standard stock. The second involves replacement of worn or damaged parts, assemblies, or equipment with those of later and more efficient design that have been approved by the responsible systems command. The last type is used for strengthening of parts that need repair or replacement to improve the reliability of the parts, provided no other change in design is involved.

ALTERATION RESPONSIBILITIES.—Ship alterations (SHIPALTS) involve material under the technical control of the Naval Sea Systems Command (NAVSEA). Alterations which affect shipboard systems and equipment under the technical control of other systems commands; for example, air alterations (AIRALTS), ordnance alterations (ORDALTS), and special program alterations (SPALTS), are not SHIPALTS. However, they may require concurrent SHIPALTS if changes affect shipboard system interface.

ELECTRONIC EQUIPMENT FIELD CHANGES.—Field changes are identified by type and class. The type depends on the material included in the change kit or furnished by the installing activity. The class refers to the funding and the installation responsibility.

Details concerning various types of and approval authority for alterations can be found in the Electronics Installation and Maintenance Book (EIMB), *General*, NAVSEA SE000-00-EIM-100.

Material Identification

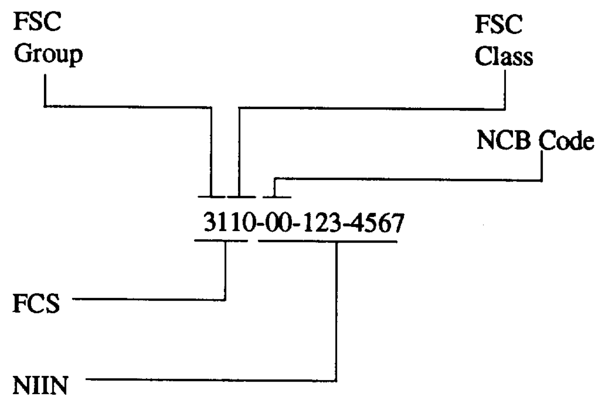
At some time in your work, you will probably have to replace a defective part or component. If you are familiar with national stock numbers (NSNs), Navy item control numbers (NICNs), part numbers (PNs), and the Coordinated Shipboard Allowance List (COSAL), getting the replacement should be a simple chore.

NATIONAL STOCK NUMBERS (NSNs).—An NSN is a 13-digit stock number used to identify an item of material in the Federal Catalog System of the Department of Defense. It consists of a four-digit federal supply classification (FSC) and a nine-digit national item identification number (NIIN). The first two digits of the FSC denote the group or major division of materials and the last two digits denote the class of subdivision of material within a group.

Examples of groups are:

GROUP	TITLE
31	Bearings
48	Valves
59	Electrical and Electronic Systems Components
79	Cleaning supplies

A complete listing of groups is provided in NAVSUP P-485, *Afloat Supply Procedures*. The NIIN consists of a two-digit national codification bureau (NCB) code and seven digits which, in conjunction with the NCB code, uniquely identify each NSN item in the Federal Supply System. For example:



Two NCB codes are assigned for the United States, 00 and 01. Code 00 identifies all FSNs (11-digit federal stock numbers used prior to NSNs) assigned prior to 31 March 1975. Code 01 identifies the numbers assigned after that time. The NCBs must be included and be correct, or the material may be rejected or you may receive the wrong material.

NAVY ITEM CONTROL NUMBERS (NICNs).—Material not included in the Federal Catalog System, but stocked or monitored in the Navy Supply System, are listed by 13-character Navy item control numbers (NICNs). These NICNs are readily identified by a two-position alpha code which signifies the type of NICN. This code and a seven-position alphanumeric uniquely identify each NICN item in the Navy Supply System. NICN codes that are currently used and examples of NICNs follow:

NICN CODE	APPLICATION	EXAMPLE
LE	Poseidon items common to Trident	1220-LE-F00-4016
LF	Stock numbers for forms	1018-LF-504-2201
LK	Aircraft change kit numbers	1234-LK-UA1-2345
LP	Stock numbers for publications	0530-LP-485-0000
LS	Special programs alteration kit numbers	1234-LS-123-4567
LX	ASO local control numbers	1560-LX-NPI-2345
LL	Local control numbers (Temporary)	4820-LL-000-1234
	Local control numbers (Permanent)	7520-LL-CAO-0001

Note: The permanent local control number can always be recognized by a C in the seventh position.

Parts can be ordered by using a NIIN or NICN.

PART NUMBERS (PNs).—A part number (also known as reference number) may be used to identify a material item or to assist you in finding the current NSN. Part numbers include old NSNs, FSNs, electron tube type numbers, and electronic equipment circuit symbol numbers. Two other important sources for reference numbers are manufacturers' part numbers and Navy drawing and piece numbers. They can be easily converted to NSNs by using the *Master Cross-Reference List* (MCRL). Table 1-44 shows excerpts from the MCRL. If the part or reference numbers do not cross to an NSN, the P/N can be used to order the replacement. A good source of part numbers is the parts list in the equipment technical manual.

Table 1-44.—Excerpts from Master Cross-Reference List (MCRL)

FEDERAL SUPPLY CODE FOR MANUFACTURERS

J17 REF NO	FSCM	NSN	ISC	RNVC	RNCC	SADC	DA	ITEM NAME
CBTV545A	80009	5625-00-714-3952*	3	2	5			OSCILLOSCOPE
		5625-00-225-0248						
CBTYPE102	7 7075	8030-00-734-9314	5	2	3		N	COMPOUND, CALKIN
CBTZ2200D	23040	2530-00-062-0345*	2	2	5			CUP, HYDRAULIC B
		2530-00-278-2267						
CBTZ6256A	23040	3020-00-499-9631	C	2	3		N	SPROCKET, DRIVE
CBT12	85 597	3460-00-586-6650	5	2	3			CHIP BREAKER
CBT13-023	801 03	5910-00-500-9119	3	2	3			CAPACITOR, FIXED

PART - 1

(REFERENCE NUMBER - TO - NEW)

NSN	ISC	REF NO	FSCM	RNVC	RNCC	SADC	DA	ITEM NAME
6905-00-714-3979	3	5905-00-503-5984*						
6625-00-714-3990	5	10656-307	94756	2	3		N	PULLEY, FLAT WEB
6625-00-714-3992*	3	CBTV545A	80009	2	5			OSCILLOSCOPE
		PC834900041-5	33597	2	3			OSCILLOSCOPE
		1011931	55232	2	5			OSCILLOSCOPE
		1548298	10001	2	3			OSCILLOSCOPE
		1842288	30003	2	5			OSCILLOSCOPE
		545A	80009	2	3			OSCILLOSCOPE
		9975856	18876	7	5			OSCILLOSCOPE
		6625-00 225-0248						
6625-00-714-3993	6	6587	47496	2	3			METER, ELECTRICA

PART - 2

(NSN - TO - REFERENCE NUMBER)

COORDINATED SHIPBOARD ALLOWANCE LIST (COSAL).—The COSAL can help you to identify repair or replacement parts. Part IIB of the COSAL is a cross-reference (microfiche only) from circuit symbol number to PN/NIIN/NICN. These parts are normally carried onboard ship for ready issue.

Outside Assistance

Your command may from time to time request assistance from another activity. This outside assistance is usually for the purpose of training, technical assistance on unusual design, planning, installation, or solving maintenance problems. Many of these activities exist throughout the fleet and shore establishments. Their capabilities and areas of responsibility differ just as equipment and systems differ. Several of the more widely known activities are included below.

MOBILE TECHNICAL UNITS (MOTU).—MOTUs provide on-the-job training and technical assistance for shipboard NAVSEA-SYSCOM/NAVELEXSYSCOM systems and equipment. They are staffed by senior military personnel and Contractor Engineering and Technical Services (CETS) representatives. MOTU's and NAUSEA combined about 10 years ago to form FTSCLANT/PAC.

MOTUs are located in the following areas:

FTSCLANT/PAC
Pearl Harbor, Hawaii
Norfolk, Va.
Groton, Conn.
San Diego, Calif.
Naples, Italy
Yokosuka, Japan
Mayport, Fla.
NSB Kings Bay, Ga.
Seattle, Wash.

NAVAL SEA SYSTEMS COMMAND (NAVSEA).— NAVSEA provides technical assistance through the use of direct fleet support technicians (TECHREP). These technicians are not to be used primarily as repairmen. The objective of their services is to promote fleet readiness and maintenance self-sufficiency. NAVSEATECHREP are located at Naval Sea Support Centers (NAVSEACENs) in Portsmouth, Va., and San Diego, Calif. They are also located at Fleet Support Offices (FSOs) in Mayport, Fla., Charleston, S.C., and New London, Conn. Selected equipment may have services provided by NAVSES, Philadelphia, or NAVSHIPWPNSYSENGSTA (NSWSES), Port Hueneme, Calif.

NAVAL ELECTRONIC SYSTEMS COMMAND (NAVELEX).— **NAVELEX** equipment is supported by the Fleet Liaison Program. Training and technical assistance is provided by civilian technicians at six NAVELEX field activities. Five are Naval Electronic Systems Engineering Centers (NESECs) and one is a Naval Electronic Systems Engineering Activity (NESEA). Fleet Liaison Offices are located at NESEC Washington, D.C., NESEC Charleston, S.C., NESEC Portsmouth, Va., NESEC San Diego, Calif., NESEC Vallejo, Calif., and NESEA St. Inigoes, Md.

NAVAL AIR SYSTEMS COMMAND (NAVAIR).— **NAVAIR** established the Navy Engineering and Technical Services (NETS) program to provide a source of technical and training assistance expertise. The program is comprised of military and civilian personnel. These people are qualified to provide advice, instruction, and training to support the installation, operation, and maintenance of Navy weapons, weapon-systems, and equipment. NETS technicians are assigned and administered by the Pacific Missile Test Center, Point Mugu, Calif., and the Naval Aviation Engineering Service Unit (NAESU), Philadelphia, Pa.

The Pacific Missile Test Center provides engineering and technical services on air-launched missile systems, air-launched guided weapons, Navy target systems, conventional ordnance, and associated support equipment. All aircraft equipment and systems not specified above are the responsibility of the Naval Aviation Engineering Service Unit (NAESU).

NAVAL AVIATION ENGINEERING SERVICE UNIT (NAESU).— **NAESU** provides field engineering assistance and instruction in installation, repair, and operation of all types of aviation systems and equipment to naval aviation fleet and shore activities throughout the world. This is accomplished by detachments (NAESU DETs) at the following locations:

Atlanta, Ga.	Miramar, Calif.
Atsugi, Japan	Misawa, Japan
Barbers Point, Hawaii	Moffett Field, Calif.
Beaufort, S.C.	Naples, Italy
Bermuda	New Orleans, La.
Brunswick, Maine	New River, N.C.
Cecil Field, Fla.	Norfolk, Va
Cherry Point, N.C.	Oceana, Va.
China Lake, Calif.	Okinawa, Japan
Corpus Christi, Tex.	Patuxent River, Md.
Cubi Point, Philippines	Pensacola, Fla.
Dallas, Tex.	Point Mugu, Calif.
Detroit, Mich.	Rota, Spain
El Toro, Calif.	San Diego, Calif.
Glenview, Ill.	Sigonella, Sicily
Agana, Guam	South Weymouth, Mass.
Iwakuni, Japan	Washington, D.C.
Jacksonville, Fla.	Whidbey Island, Wash.
Kaneohe Bay, Hawaii	Willow Grove, Pa.
Key West, Fla.	Yuma, Ariz.
Lemoore, Calif	Memphis, Tenn

CARRIER AND FIELD SERVICE UNITS (CAFSUs).— **CAFSUs** furnish technical guidance and assistance to shipyards, ship repair facilities, and shore and fleet personnel concerning the installation, operation, maintenance, and testing of shipboard NAVAIR equipment. This equipment includes catapults, arresting gear, visual landing aids, flight deck lighting, pilot landing aid television (PLAT) systems, Fresnel-lens optical landing systems (FLOLS), and integrated launch and recovery television surveillance (ILARTS) systems.

CAFSUs are under the administrative control of the Naval Air Engineering Center (NAEC), Lakehurst, N.J., and are located at the following activities:

NAS Norfolk, Va.	SRF Subic Bay, Philippines
NAVSTA Mayport, Fla.	NAEC Philadelphia, Pa.
NAF Naples, Italy	SRF Yokosuka, Japan
NAS North Island, Calif.	NAS Alemeda, Calif.

INTERMEDIATE MAINTENANCE ACTIVITIES (IMAs).—Afloat **IMAs** (tenders and repair ships) and shore IMAs (SIMAs) provide maintenance support for repairs beyond the capabilities of ship's force. IMAs also have facilities for test equipment calibration and emergency parts manufacture. Electrical, electronic, and ordnance repair divisions provide repairs on various equipment including gyrocompasses, navigational equipment, film projectors, internal communications, sonar, radar, IFF, radio receivers and transmitters, test antennas, guns and small arms, torpedoes, fire control, and missile systems. Table 1-45 lists the various IMAs and their locations.

Table 1-45.—Intermediate Maintenance Activities

A. DESTROYER TENDER (AD)			
SIERRA	AD	18	CHARLESTON
YOSEMITE	AD	19	MAYPORT
SAMUEL GOMPERS	AD	37	SAN DIEGO
PUGET SOUND	AD	38	GAETA
YELLOWSTONE	AD	41	NORFOLK
ACADIA	AD	42	SAN DIEGO
CAPE COD	AD	43	SAN DIEGO
SHENANDOAH	AD	44	NORFOLK
B. REPAIR SHIP (AR)			
VULCAN	AR	5	NORFOLK
JASON	AR	8	PEARL HARBOR
C. SUBMARINE TENDER (AS)			
FULTON	AS	11	QUINCY
ORION	AS	18	LA MADDALENA
PROTEUS	AS	19	GUAM
HUNLEY	AS	31	HOLY LOCH
HOLLAND	AS	32	CHARLESTON
SIMON LAKE	AS	33	KINGS BAY
CANOPUS	AS	34	CHARLESTON
L Y SOEAR	AS	36	NORFOLK
DIXON	AS	37	SAN DIEGO
EMORY S LAND	AS	39	NORFOLK
FRANK CABLE	AS	40	CHARLESTON
MCKEE	AS	41	SAN DIEGO
D. SHORE INTERMEDIATE MAINTENANCE ACTIVITY			
SIMA CHARLESTON			CHARLESTON
SIMA GUANTANAMO BAY			GUANTANAMO
SIMA LITTLE CREEK			LITTLE CREEK
SIMA (NRMF) NEWPORT			NEWPORT R.I.
SIMA MAYPORT			MAYPORT
SIMA NORFOLK			NORFOLK
SIMA PORTSMOUTH VA			PORTSMOUTH VA
SIMA (NRMF) PHILADELPHIA			NB PHILA
SIMA SAN DIEGO			SAN DIEGO
SIMA PEARL HARBOR			PEARL HARBOR
SIMA SAN FRANCISCO			ALAMEDA
SIMA LONG BEACH			LONG BEACH

Publications and Documents

Various publications, some of which are discussed below, are available for guidance in maintenance work or for reference and study. In general, these publications are available from the Naval Publications and Forms Center through the supply system.

NAVAL SHIPS' TECHNICAL MANUAL (NSTM).—The Naval Ships' Technical Manual (NSTM) is a prime reference for information on NAVSEA equipment. Chapter 400, *Electronics*, is most useful as it provides major policies and instructions pertaining to electronics work and material under NAVSEA and NAVELEX responsibility. Other chapters of interest to electrical and electronics technicians are:

300	Electrical Plant General
302	Electric Motors and Controllers
310	Electric Power Generators and Conversion Equipment
320	Electric Power Distribution Systems
330	Lighting
430	Interior Communication Installations
434	Motion Picture Equipment
491	Electrical Measuring and Test Instruments
510	Ventilating, Heating, Cooling, and Air-Conditioning Systems for Surface Ships
532	Liquid-Cooling Systems for Electronic Equipment
9006	Submarine Antennas and Masts
634	Deck Coverings

ELECTRONICS INSTALLATION AND MAINTENANCE BOOK (EIMB).—The Electronics Installation and Maintenance Book series supplements instructions and data supplied in equipment technical manuals. The EIMB is intended to reduce time-consuming research on electronic equipment and circuit theory. These handbooks fall into two categories: general information and equipment-oriented handbooks. The latter includes general test procedures, adjustments, and general servicing information. All handbooks of the series are listed below.

TITLE	NUMBERS
GENERAL	
General	SE000-00-EIM-100
Installation Standards	0967-LP-000-0110
Electronic Circuits	0967-000-0120
Test Methods & Practices	0967-LP-000-0130
Reference Data	0967-000-0140
EMI Reduction	0967-000-0150
General Maintenance	SE000-00-EIM-160
EQUIPMENT-ORIENTED	
Communications	SE000-00-EIM-010
Radar	SE000-00-EIM-020
Sonar	SE000-00-EIM-030
Test Equipment	SE000-00-EIM-040
Radiac	0967-000-0050
Countermeasures	SE000-00-EIM-07

ENGINEERING INFORMATION BULLETIN (EIB).—The *Engineering Information Bulletin* is published biweekly and distributed to all naval ships and electronics installation and maintenance activities. It is authoritative and is a means of rapid dissemination of advanced hull, mechanical, electronic, electrical, and related equipment information. It includes information concerning approved beneficial suggestions, electronics field changes, mechanical alterations (MECHALTS), installation techniques, maintenance notes and practices, and technical manual availabilities, advance change notices, and distribution.

EQUIPMENT TECHNICAL MANUALS.—Technical manuals carry information essential to the proper operation, maintenance, and repair of specific equipment. These manuals may occasionally contain errors. In those cases, change notices are provided to correct the manuals. Updates because of equipment changes are also provided. These changes must be installed in the technical manuals to maintain accuracy and to prevent the loss of man-hours resulting from the use of obsolete data and/or schematics. The *Guide for User Maintenance of NAVSEA Technical Manuals*, NAVSEA S005-AA-GYD-030/TMMP, provides information on identifying, ordering, deficiency reporting, and updating technical manuals.

NAVSUP PUBLICATION 2002.—**NAVSUP 2002** is the *Navy Stock List of Publication and Forms* and provides NSNs for ordering Navy publications and their changes. Each edition is issued quarterly and supersedes the previous one in its entirety. It is produced in *microfiche* only and contains three sections:

Section 1—Forms

Section 2—Publications

Section 3—NAVAIR Technical Directives

DECKPLATE.—The *deckplate* is a technical periodical published monthly by NAVSEA. It contains information on design, construction, conversion, operation, maintenance, and repair of Navy vessels and their equipment. It also includes articles on personnel safety, service hints, and adopted beneficial suggestions.

NAVAL SAFETY CENTER PUBLICATIONS.—The Naval Safety Center publishes bulletins and several periodicals to keep Navy personnel informed on the subject of accident prevention.

- *Ship Safety Bulletin*—This monthly newsletter contains safety notes and accident data.
- *Fathom*—This quarterly magazine contains a review of surface ship and submarine accident prevention and safety articles.
- *Approach*—This magazine is a monthly review of articles concerning aviation safety and accident prevention.
- *MECH*—MECH is a bimonthly review of aviation maintenance related mishaps, material/personnel hazards, and general aviation ground safety.

TRAINING MANUALS (TRAMANS).—Training manuals are designed to give enlisted personnel background knowledge for the proper performance of their assigned jobs. Electrical and electronic theory and operation and maintenance information on pertinent equipment are presented at different rating levels in the TRAMANs written for the technical rates.

APPENDIX I

REFERENCES USED TO DEVELOP THIS NRTC

NOTE: Although the following references were current when this NRTC was published, their continued currency cannot be assured. When consulting these references, keep in mind that they may have been revised to reflect new technology or revised methods, practices, or procedures; therefore, you need to ensure that you are studying the latest references.

Afloat Shopping Guide*, NAVSUP Publication 4400, Naval Supply Systems Command, Washington, D.C., 1991.*

Afloat Supply Procedures*, NAVSUP Publication 485, 0530-LP-185-7600, Naval Supply Systems Command, Washington, D.C., 1989.*

Avionic Cleaning and Corrosion Prevention/Control, *NAVAIR 16-1-540, Naval Air Systems Command, Washington, D.C., 1984.*

Basic Military Requirements*, NAVEDTRA 10054-F, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1986.*

Cable Comparison Handbook*, Military Standard MIL-STD 299, Department of Defense, Washington, D.C., 1989.*

Capacitors, Selection and Use of*, Military Standard MIL-STD-198E, Department of Defense, Washington, D.C., 1984.*

Circuit Breakers, Selection and Use of*, Military Standard MIL-STD-1498B, Department of Defense, Washington, D.C., 1988.*

Design Data Book*, USN Ships NAVSEA 0902-LP-006-0000, Naval Sea Systems Command, Washington, D.C., 1988.*

Electrical Connectors, Plug-in Sockets, and Associated Hardware, Selection and Use of, Military Standard MIL-STD-1353B, Department of Defense, Washington, D.C., 1980.

Electronics Installation and Maintenance Book (EIMB), General*, Naval Sea Systems Command, Washington, D.C., NAVSEA SE000-00-EIM-100, 1983.*

Electronics Installation and Maintenance Book (EIMB), Test Methods and Practices*, NAVSEA 0967-LP-000-0130, Naval Sea Systems Command, Washington, D.C., 1980.*

Electronics Installation and Maintenance Book (EIMB), Reference Data*, Naval Sea Systems Command, Washington, D.C., NAVSHIPS 0967-000-0140, 1972.*

Electronics Installation and Maintenance Book (EIMB), General Maintenance*, Naval Sea Systems Command, Washington, D.C., NAVSEA SE000-00-EIM-160, 1981.*

Fiber Optic Symbols*, Military Standard MIL-STD-1864, Department of Defense, Washington, D.C., 1991.*

Fuses, Fuseholders, and Associated Hardware, Selection and Use of*, Military Standard MIL-STD-1360A, Department of Defense, Washington, D.C., 1979.*

Glossary of Telecommunication Terms*, Federal Standard 1037A, General Services Administration, Washington, D.C., 1986.*

I C Electrician 3*, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., NAVEDTRA 10059-A, 1989.*

Installation Practices-Aircraft Electric and Electronic Wiring*, NAVAIR 01-1A-505, Naval Air Systems Command, Washington, D.C., 1988.*

Installation Standards and Practices, NAVELEX 0280-LP-900-8000, 1977.

Insulation Sleeving, Electrical, Heat-Shrinkable, Polyolefin, Dual-Wall, Outer Wall Crosslinked, Military Specification MIL-I-23053/4C, Department of Defense, Washington, D.C., 1988.

Insulation Sleeving, Electrical, Heat-Shrinkable, Polyolefin, Flexible, Crosslinked, Military Specification MIL-I-23053/5B, Washington, D.C., 1986.

Linear Integrated Circuits Study Booklet, Module 34, CNTT-E-056, 1981.

List of Standard Microcircuits*, Military Standard MIL-STD-1562V, Department of Defense, Washington, D.C., 1991.*

Microcircuits, General Specification for*, Military Specification MIL-M-38510H, Department of Defense, Washington, D.C., 1990.*

Naval Oceanography Command Instruction 3144.1C, 1983.

Naval Ships' Technical Manual (NSTM), Stowage, Handling, and Disposal of Hazardous General Use Consumables*, Chapter 670, NAVSEA S9086-WK-STM-010, Naval Sea Systems Command, Washington, D.C., 1987.*

NEETS, Module 1, Introduction to Matter, Energy, and Direct Current*, NAVEDTRA 172-01-00-88, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1988.*

NEETS, Module 2, Introduction to Alternating Current and Transformers*, NAVEDTRA 172-02-00-91, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1991.*

NEETS, Module 3, Introduction to Circuit Protection, Control, and Measurement*, NAVEDTRA 172-03-00-85, Naval Education and Training Professional Development and Technology Center, Fla., 1995.*

NEETS, Module 4, Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading*, NAVEDTRA 172-04-00-85, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1985.*

NEETS, Module 5, Introduction to Generators and Motors*, NAVEDTRA 172-05-00-79, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1979.*

NEETS, Module 6, Introduction to Electronic Emission, Tubes, and Power Supplies*, NAVEDTRA 172-06-00-82, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1982.*

NEETS, Module 7, Introduction to Solid-State Devices and Power Supplies*, NAVEDTRA 172-07-00-82, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1982.*

NEETS, Module 8, Introduction to Amplifiers*, NAVEDTRA 172-08-00-82, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1982.*

NEETS, Module 9, Introduction to Wave Generation and Wave-Shaping Circuits*, NAVEDTRA 172-09-00-83, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1983.*

NEETS, Module 13, Introduction to Number Systems, Boolean Algebra, and Logic Circuits*, NAVEDTRA 172-13-00-86, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1986.*

NEETS, Module 18, Radar Principles*, NAVEDTRA 172-18-00-84, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1984.*

Resistors, Selection and Use of*, Military Standard MIL-STD-199E, Department of Defense, Washington, D.C., 1991.*

Rf Transmission Lines and Fittings*, MIL-HDBK-216, Department of Defense, Washington, D.C., 1962.*

Shipboard Electronic Material Officer*, NAVEDTRA 10478-A1, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1982.*

Standard General Requirements for Electronics Equipment*, Military Standard MIL-STD-454M, Department of Defense, Washington, D.C., 1989.*

SPECIAL PUBLICATION

Navy Electricity and Electronics Training Series

Module 20—Master Glossary

NAVEDTRA 14192

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

PREFACE

About this course:

This is a self-study course. By studying this course, you can improve your professional/military knowledge, as well as prepare for the Navywide advancement-in-rate examination. It contains subject matter about day-to-day occupational knowledge and skill requirements and includes text, tables, and illustrations to help you understand the information. An additional important feature of this course is its reference to useful information in other publications. The well-prepared Sailor will take the time to look up the additional information.

Training series information:

This is Module 20 of a series.

History of the course:

Sep 1998: *Original edition released. Prepared by FTCM Gilbert J. Coté*
Jan 2004: *Administrative update released. Reviewed by ETC(SW) Jack Weatherford. Technical content was not revised.*

TABLE OF CONTENTS

CHAPTER **PAGE**

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

CHAPTER 1

MASTER GLOSSARY OF TERMS

INTRODUCTION

This module presents the combined glossaries from modules 1 through 18, the primary texts in the NEETS series. Definitions of terms used in NEETS are presented in an alphabetical glossary. Numbers in parenthesis ( ) indicate more than one definition for the same term. Numbers in brackets [ ] indicate the module number in which more information about the term may be found.

ABSORPTION—(1) Dissipation of radio or sound waves as they interact with matter. (2) The absorbing of light waves without reflection or refraction [10].

ABSORPTION, LAW OF—In Boolean algebra, the law which states that the odd term will be absorbed when a term is combined by logical multiplication with the logical sum of that term and another term, or when a term is combined by logical addition with the logical product of one term and another term (for example, $A(A + B) = A + AB = A$) [13].

ABSORPTION WAVEMETER—An instrument used to measure audio frequencies [16].

ACCELERATING ANODE—An electrode charged several thousand volts positive and used to accelerate electrons toward the front of a cathode-ray tube [6].

ACCELERATION SERVOSYSTEM—A servosystem that controls the acceleration (rate of change in velocity) of a load [15].

ACCELEROMETER—A device that measures the acceleration to which it is subjected and develops a signal proportional to it [15].

ACCEPTOR IMPURITY—An impurity which, when added to a semiconductor, accepts one electron from a neighboring atom and creates a hole in the lattice structure of the crystal. Also called TRIVALENT IMPURITY [7].

ACORN TUBE—A very small tube with closely spaced electrodes and no base. The tube is connected to its circuits by short wire pins that are sealed in a glass or ceramic envelope. The acorn tube is used in low-power uhf circuits [6].

ACOUSTICS—The science of sound [10].

ACQUISITION—Operational phase of a fire-control or track radar during which the radar system searches a small volume of space in a prearranged pattern [18].

ACTIVE SATELLITE—A satellite that amplifies the received signal and retransmits it back to earth [17].

ACTUATOR—The part of a switch that is acted upon to cause the switch to change contact connections; for example, toggle, pushbutton, and rocker [3].

AFDS—An abbreviation for the amphibious flagship data system [17].

AIR-CONTROL PANEL—Panel that monitors the dry-air input at each user equipment [18].

AIR-CORE TRANSFORMER—A transformer composed of two or more coils that are wound around a nonmetallic core [2].

ALLOWANCE PARTS LIST (APL)—Repair parts required for units having the equipment/ component listed [14].

ALLOYED JUNCTION—A junction formed by recrystallization of a molten region of P-type material on an N-type substrate, or vice versa [7].

ALPHA—The emitter-to-collector current gain in a common-base circuit [7].

ALTERNATING CURRENT—An electrical current that constantly changes amplitude and changes polarity at regular intervals [2].

ALTITUDE—The vertical distance of an aircraft or object above a given reference, such as ground or sea level [18].

ALUMINUM CREEP—(1) The movement of aluminum wire from a point where pressure is applied. (2) The "retreat" of heated aluminum wire as it cools [4].

AMBIENT TEMPERATURE—The surrounding temperature such as the temperature of air surrounding a conductor in a compartment or within a piece of equipment [4].

AMBIGUOUS RETURNS—Echoes that exceed the prt of a radar and appear at incorrect ranges [18].

AMERICAN WIRE GAUGE (AWG)—The standards adopted in the United States for the measurement of wire sizes [4].

AMMETER—An instrument for measuring the amount of electron flow (in amperes) [1] [3] [6].

AMPERE—The basic unit of electrical current [1].

AMPERE-TURN—The magnetomotive force developed by 1 ampere of current flowing through a coil of one turn [8].

AMPERITE (BALLAST) TUBE—A current-controlling resistance device designed to maintain substantially constant current over a specified range of variation in applied voltage or resistance of a series circuit [6].

AMPLIDYNE—A special dc generator in which a small dc voltage applied to field windings controls a large output voltage from the generator. In effect, an amplidyne is a rotary amplifier that often times produces gain of approximately 10,000 [5].

AMPLIFICATION—(1) The process of enlarging a signal in amplitude (as of voltage or current) [8]. (2) The ratio of output magnitude to input magnitude in a device that is intended to produce an output that is an enlarged reproduction of its input [6] [7].

AMPLIFICATION FACTOR—The voltage gain of an amplifier with no load on the output [6] [7].

AMPLIFIER—The device that provides amplification (the increase in current, voltage, or power of a signal) without appreciably altering the original signal [7] [8].

AMPLITRON—See CROSS-FIELD AMPLIFIER [18].

AMPLITUDE—The size of a signal as measured from a reference line to a maximum value above or below the line. Generally used to describe voltage, current, or power [8] [12].

AMPLITUDE DISTORTION—Distortion that is present in an amplifier when the amplitude of the output signal fails to follow exactly any increase or decrease in the amplitude of the input signal [6] [7].

AMPLITUDE MODULATION—Any method of varying the amplitude of an electromagnetic carrier frequency in accordance with the intelligence to be transmitted [12].

AMPLITUDE STABILITY—Amplitude stability refers to the ability of the oscillator to maintain a constant amplitude in the output waveform [9].

AND CIRCUIT—See AND GATE [13].

AND GATE—(1) An electronic gate whose output is energized only when every input is in its prescribed state. An AND gate performs the function of the logical "AND"; also called an AND circuit. (2) A binary circuit, with two or more inputs and a single output, in which the output is a logic 1 *only* when all inputs are a logic 1 and the output is a logic 0 when *any* one of the inputs is a logic 0 [13].

ANGLE MODULATION—Modulation in which the angle of a sine-wave carrier is varied by a modulating wave [12].

ANGLE OF INCIDENCE—The angle between the incident wave and the normal [10].

ANGLE OF INCLINATION—The angular difference between the equatorial plane of the earth and the plane of orbit of the satellite [17].

ANGLE OF REFLECTION—The angle between the reflected wave and the normal [10].

ANGLE OF REFRACTION—The angle between the normal and the path of a wave through the second medium [10].

ANGSTROM UNIT—The unit used to define the wavelength of light waves [10].

ANISOTROPIC—The property of a radiator that allows it to emit strong radiation in one direction [10].

ANODE—(1) A positive electrode of an electrochemical device (such as a primary or secondary electric cell) toward which the negative ions are drawn [1] [6] [7]. (2) The semiconductor-diode terminal that is positive with respect to the other terminal when the diode is biased in the forward direction [13].

ANTENNA—A conductor or set of conductors used to radiate RF energy into space or to collect RF energy from space or to do both [10].

ANTENNA BEAM WIDTH—Width of a radar beam measured between half-power points [18].

ANTENNA COUPLER—A device used for impedance matching between an antenna and a transmitter or receiver [17].

ANTENNA SYSTEM—Routes RF energy from the transmitter, radiates the energy into space, receives echoes, and routes the echoes to the receiver [18].

ANTIJAMMING CIRCUIT—An electronic circuit used to minimize the effects of enemy countermeasures, thereby permitting radar echoes to be visible on the indicator [18].

ANTISEIZE COMPOUND—A silicon-based, high-temperature lubricant applied to threaded components to aid in their removal after they have been subjected to rapid heating and cooling [4].

ANTITRANSMIT-RECEIVE TUBE (atr)—A tube that isolates the transmitter from the antenna and receiver. Used in conjunction with a tr tube [18].

APERTURE—See SLOT [11].

APOGEE—The point in the orbit of a satellite the greatest distance from the earth [17].

APPARENT DRIFT—The effect of the earth's rotation on a gyro that causes the spinning axis to appear to make one complete rotation in one day. Also called APPARENT PRECESSION or APPARENT ROTATION [15].

APPARENT POWER—That power apparently available for use in an ac circuit containing a reactive element. It is the product of effective voltage times effective current expressed in volt-amperes. It must be multiplied by the power factor to obtain true power available [2].

APPARENT PRECESSION—See APPARENT DRIFT [15].

APPARENT ROTATION—See APPARENT DRIFT [15].

ARC EXTINGUISHER—The part of a circuit breaker that confines and divides the arc which occurs when the contact of the circuit breaker opens [3].

ARMATURE—(1) In a relay, the movable portion of the relay [3]. (2) The windings in which the output voltage is generated in a generator or in which input current creates a magnetic field that interacts with the main field in a motor [5].

ARMATURE LOSSES—Copper losses, eddy current losses, and hysteresis losses that act to decrease the efficiency of armatures [5].

ARMATURE REACTION—The effect in a dc generator of current in the armature creating a magnetic field that distorts the main field and causes a shift in the neutral plane [5].

ARRAY OF ARRAYS—Same as COMBINATION ARRAY [10].

ARTIFICIAL TRANSMISSION LINE—An LC network that is designed to simulate characteristics of a transmission line [18].

ASBESTOS—A noncombustible, nonconductive, fiber-like mineral used as an insulating material [4].

ASBESTOSIS—Fibrosis of the lungs caused by inhalation of asbestos fibers [4].

A-SCOPE—A radar display on which slant range is shown as the distance along a horizontal trace [18].

ASSEMBLY—A number of parts or subassemblies, or any combination thereof, joined together to perform a specific function [17].

ASTABLE MULTIVIBRATOR—A multivibrator that has no stable state. Also called free-running because it alternates between two different output voltage levels during the time it is on. The frequency is determined by the RC time constant of the coupling circuit [9].

ASWTDS—An abbreviation for the antisubmarine warfare tactical data system [17].

ASYMMETRICAL MULTIVIBRATOR—A multivibrator that generates rectangular waves [18].

ASYNCHRONOUS—The teletypewriter operation where the transmitter and receiver do not operate continuously [17].

ASYNCHRONOUS ORBIT—One where the satellite does not rotate or move at the same speed as the earth [17].

ATDS—An abbreviation for the airborne tactical data system [17].

ATTENUATION—The ability of a filter circuit to reduce the amplitude of unwanted frequencies to a level below that of the desired output frequency [9].

ATTRACTION—The force that tends to make two objects approach each other. Attraction exists between two unlike magnetic poles (north and south) or between two unlike static charges [1].

AUDIO AMPLIFIER—An amplifier designed to amplify frequencies between 15 hertz (15 Hz) and 20 kilohertz (20 kHz) [8].

AUDIO-FREQUENCY-TONE SHIFT—A system that uses amplitude modulation to change dc mark and space impulses into audio impulses [17].

AUTOMATIC GAIN CONTROL—A circuit used to vary radar receiver gain for best reception of signals that have widely varying amplitudes [18].

AUTOMATIC TRACKING—Tracking done by equipment that compares the direction of the antenna axis and the direction of the received signal and uses the difference (error) signal to reposition the antenna [17].

AUTOMATIC VOLUME/GAIN CONTROL—A circuit used to limit variations in the output signal strength of a receiver [17].

AVALANCHE EFFECT—A reverse breakdown effect in diodes that occurs at reverse voltages beyond 5 volts. The released electrons are accelerated by the electric field, which results in a release of more electrons in a chain or "avalanche" effect [7].

AVERAGE POWER—(1) The peak power value averaged over the pulse-repetition time [12]. (2) Output power of a transmitter as measured from the start of one pulse to the start of the next pulse [18].

AVERAGE VALUE (OF AC)—The average of all the instantaneous values of one-half cycle of alternating current [2].

AXIS—A straight line, either real or imaginary, passing through a body around which the body revolves [15].

AZIMUTH—Angular measurement in the horizontal plane in a clockwise direction [18].

BACK RESISTANCE—The larger resistance value observed when you are checking the resistance of a semiconductor [16].

BALANCED MIXER—A waveguide arrangement that resembles a T and uses crystals for coupling the output to a balanced transformer [18].

BALANCED PHASE DETECTOR—A circuit that controls the oscillator frequency (afc) [17].

BANDPASS FILTER—A filter that allows a narrow band of frequencies to pass through the circuit. Rejects or attenuates frequencies that are either higher or lower than the desired band of frequencies [9] [16].

BAND-REJECT FILTER—A tuned circuit that does not pass a specified band of frequencies [9] [16].

BANDWIDTH—The difference between the highest usable frequency of a device (upper frequency limit) and the lowest usable frequency of the device (lower frequency limit) - measured at the half-power points [8] [9] [12] [15].

BARRETTER—A type of bolometer characterized by an increase in resistance as the dissipated power rises [16].

BASE—The element in a transistor that controls the flow of current carriers [7].

BASE—(1) A reference value. (2) A number that is multiplied by itself as many times as indicated by an exponent. (3) Same as radix. (4) The region between the emitter and collector of a transistor that receives minority carriers injected from the emitter. It is the element that corresponds to the control grid of an electron tube [13].

BASE-INJECTION MODULATOR—Similar to a control-grid modulator. The gain of a transistor is varied by changing the bias on its base [12].

BATTERY—A device for converting chemical energy into electrical energy [1].

BATTERY CAPACITY—The amount of energy available from a battery. Battery capacity is expressed in ampere-hours [1].

BAUD—A measurement of speed based on the number of code elements or units per second [17].

BAY—Part of an antenna array [10].

BEAM—See LOBE [18].

BEAM-LEAD CHIP—Semiconductor chip with electrodes (leads) extended beyond the wafer [14].

BEAM-POWER TUBE—An electron tube in which the grids are aligned with the control grid. Special beam-forming plates are used to concentrate the electron stream into a beam. Because of this action, the beam-power tube has high power-handling capabilities [6].

BEARING—An angular measurement of the direction of an object from a reference direction, such as true north [11].

BEARING RESOLUTION—Ability of a radar to distinguish between targets that are close together in bearing [18].

BEAT FREQUENCIES—Difference and sum frequencies, which result from the combination of two separate frequencies [18].

BEAT FREQUENCY—The difference between the oscillator frequency and the unknown audio frequency [16].

BEAT-FREQUENCY OSCILLATOR—An additional oscillator used in a receiver when it is receiving a cw signal. It provides an audible tone [17].

BEL—The unit that expresses the logarithmic ratio between the input and output of any given component, circuit, or system [16].

BETA—The ratio of a change in collector current to a corresponding change in base current when the collector voltage is constant in a common-emitter circuit [7].

BEVERAGE ANTENNA—A horizontal, long-wire antenna designed for reception and transmission of low-frequency, vertically polarized ground waves [10].

BIAS—Difference of potential applied to a vacuum tube or transistor to establish a reference operating level [13].

BIAS CURRENT—Current that flows through the base-emitter junction of a transistor and is adjusted to set the operating point of the transistor [13].

BIDIRECTIONAL ARRAY—An array that radiates in opposite directions along the line of maximum radiation [10].

BINARY—(1) A number system that uses a base, or radix, of 2. Two digits (1 and 0) are used in the binary system. (2) Pertaining to a characteristic that involves the selection, choice, or condition in which there are only two possibilities. (3) A bistable multivibrator (flip-flop) is one example of a binary device [13].

BINARY CODE—A method of representing two possible conditions (*on* or *off*, high or low, one or zero, the presence of a signal or absence of a signal). Electronic circuits designed to work in such a way that only two conditions are possible [13].

BINARY-CODED—The state in which conditions are expressed by a series of binary digits (0's and 1's) [13].

BINARY DIGIT—(1) A character that represents one of the two digits in the number system that has a radix of two. (2) Either of the digits 0 or 1 that may be used to represent the binary conditions of *on* or *off* [13].

BINARY NOTATION—See BINARY NUMBER SYSTEM [13].

BINARY NUMBER SYSTEM—A number system using two digits, symbols, or characters (usually 1 and 0) [13].

BINARY POINT—The radix point that separates powers of two and fractional powers of two in a binary number [13].

BISTABLE—A device that is capable of assuming either one of two stable states [13].

BISTABLE MULTIVIBRATOR—A multivibrator that has two stable states. It remains in one of the states until a trigger is applied. It then flips to the other stable state and remains there until another trigger is applied. Also referred to as a FLIP-FLOP [9] [13].

BLACK—The reference color of equipment that passes unclassified information. It normally refers to patch panels [17].

BLEEDER CURRENT—The current through a bleeder resistor. In a voltage divider, bleeder current is usually determined by the 10 percent rule of thumb [1].

BLEEDER RESISTOR—A resistor used to draw a fixed current [1].

BLIP—See PIP [18].

BLOCK DIAGRAM—A diagram in which the major components of an equipment or a system are represented by squares, rectangles, or other geometric figures, and the normal order of progression of a signal or current flow is represented by lines [4].

BLOCKED-GRID KEYING—A method of keying in which the bias is varied to turn plate current on and off [12].

BLOCKING—A condition in an amplifier, caused by overdriving one or more stages, in which the amplifier is insensitive to small signals immediately after reception of a large signal [18].

BOLOMETER—A loading device that undergoes changes in resistance as changes in dissipated power occur [16].

BONDING WIRES—Fine wires connecting the bonding pads of the chip to the external leads of the package [14].

BOOLEAN—(1) Pertaining to the process used in the algebra formulated by George Boole. (2) Pertaining to the operations of formal logic [13].

BOOLEAN ALGEBRA—A system of logic dealing with on-off circuit elements associated by such operators as the AND, OR, NAND, NOR, and NOT functions [13].

BOOLEAN LOGIC—See BOOLEAN ALGEBRA [13].

BOUNDARY CONDITIONS—The two conditions that the E-field and H-field within a waveguide must meet before energy will travel down the waveguide. The E-field must be perpendicular to the walls and the H-field must be in closed loops, parallel to the walls, and perpendicular to the E-field [11].

BRANCH—An individual current path in a parallel circuit [1] [4].

BREAK—In a switch, the number of breaks refers to the number of points at which the switch opens the circuit; for example, single break and double break [3].

BREAKDOWN—The phenomenon occurring in a reverse-biased semiconductor diode. The start of the phenomenon is observed as a transition from a high dynamic resistance to one of substantially lower dynamic resistance. This is done to boost the reverse current [7].

BRIGHTNESS CONTROL—The name given to the potentiometer used to vary the potential applied to the control grid of a CRT [6].

BROADSIDE ARRAY—An array in which the direction of maximum radiation is perpendicular to the plane containing the elements [10] [18].

BRUSHES—Sliding contacts, usually carbon, that make electrical connection to the rotating part of a motor or generator [5].

BUFFER—A voltage amplifier used between the oscillator and power amplifier [12].

BUFFER AMPLIFIER—An amplifier that isolates one circuit from another. It decreases the loading effect on an oscillator by reducing the interaction between the load and the oscillator [9] [18].

BUILT-IN TEST EQUIPMENT (BITE)—A permanently mounted device that is used expressly for testing an equipment or system [14].

BUNCHER CAVITY—The input resonant cavity in a conventional klystron oscillator [11].

BUNCHER GRID—In a velocity-modulated tube, the grid that concentrates the electrons in the electron beam into bunches [11].

BURNISHING TOOL—A tool used to clean and polish contacts on a relay [3].

BUS BAR—A heavy copper strap or bar used to connect several circuits together when a large current-carrying capacity is required [4].

BYPASS CAPACITOR—A capacitor used to transfer unwanted signals out of a circuit; for example, coupling an unwanted signal to ground. Also called a DECOUPLING CAPACITOR [8].

CABLE—Either a stranded conductor (single-conductor cable) or a combination of conductors insulated from one another (multiple conductor cable). Small sizes are commonly referred to as stranded wire or as cords [4].

CABLE HARNESS—A group of wires or ribbons of wiring used to interconnect electronic systems and subsystems [14].

CAPACITANCE—The property of an electrical circuit that opposes changes in voltage [2].

CAPACITIVE REACTANCE—The opposition, expressed in ohms, offered to the flow of an alternating current by capacitance. The symbol for capacitive reactance is X_C [2] [9].

CAPACITOR—An electrical device capable of storing electrical energy in an electrostatic field [2].

CAPACITOR FILTER—This filter is used on extremely high-voltage, low-current power supplies and also where the ripple frequency is not critical [7].

CAPACITOR-START MOTOR—A type of single-phase, ac induction motor in which a starting winding and a capacitor are placed in series to start the motor. The values of X_C and R are such that the main-winding and starting-winding currents are nearly 90 degrees apart and the starting torque is produced as in a two-phase motor [5].

CARBON MICROPHONE—A microphone in which sound waves vary the resistance of a pile of carbon granules. May be single-button or double-button [12].

CARDIOPULMONARY RESUSCITATION—Procedure designed to restore breathing after cardiac arrest. Includes clearing air passages to lungs and heart massage [1].

CARRIER FREQUENCY—The frequency of an unmodulated transmitter output [12] [18].

CARRIER-CONTROLLED APPROACH—A shipboard radar system used to guide aircraft to safe landings in poor visibility conditions [18].

CARRY—(1) One or more digits, produced in connection with an arithmetic operation, that is/are forwarded to another digit place for processing there. (2) The number represented by the digit or digits in (1) above [13].

CATCHER GRID—In a velocity-modulated tube, a grid on which the spaced electron groups induce a signal. The output of the tube is taken from the catcher grid [11].

CATHODE—(1) In an electron tube the electrode that is the source of current flow [6]. (2) The general name for any negative electrode [1]. (3) The negative terminal of a forward-biased semiconductor diode, which is the source of the electrons [7].

CATHODE BIAS—The method of biasing a vacuum tube in which the biasing resistor is placed in the common-cathode return circuit, thereby making the cathode more positive with respect to ground [6].

CATHODE KEYING—A system in which the cathode circuit is interrupted so that neither grid current nor plate current can flow [12].

CATHODE MODULATOR—Voltage on the cathode is varied to produce the modulation envelope [12].

CATHODE-RAY TUBE (CRT)—An electron tube that has an electron gun, a deflection system, and a screen. This tube is used to display visual electronic signals [6].

CATHODE SPUTTERING—A process of producing thin film components [14].

CAVITY RESONATOR—A space totally enclosed by a metallic conductor and supplied with energy in such a way that it becomes a source of electromagnetic oscillations. The size and shape of the enclosure determine the resonant frequency [11].

CAVITY WAVEMETER—An instrument used to measure microwave frequencies [16].

CELL—A single unit that transforms chemical energy into electrical energy. Batteries are made up of cells [1].

CENTER-FEED METHOD—Connecting the center of an antenna to a transmission line which is then connected to the final (output) stage of the transmitter [10].

CENTIMETER CUBE—A unit of volume of large rectangular or square conductors. The cross-sectional area equals 1 square centimeter with a length of 1 centimeter [4].

CHANNEL—A carrier frequency assignment, usually with a fixed bandwidth [12].

CHARACTER—A letter, digit, or other symbol that is used as part of the organization, control, or representation of information [13].

CHARACTERISTIC IMPEDANCE—The ratio of voltage to current at any given point on a transmission line represented by a value of impedance [10].

CHARGE—Represents electrical energy. A material having an excess of electrons is said to have a negative charge. A material having a shortage of electrons is said to have a positive charge [1].

CHARGE CYCLE—The period of time that a capacitor in an electrical circuit is storing a charge [2].

CHOKE—An inductor used to impede the flow of pulsating dc or ac by means of self-inductance [6] [7].

CHOKE JOINT—A joint between two sections of waveguide that provides a good electrical connection without power losses or reflections [11].

CIRCUIT—The complete path of an electric current [1].

CIRCULAR MIL—An area equal to that of a circle with a diameter of 0.001 inch. It is used for measuring the cross-sectional area of wires [1].

CIRCULAR MIL-FOOT—A unit of volume of a conductor having a cross-sectional area of 1 circular mil and a length of 1 foot [4].

CLAMPER—A circuit in which either the upper or lower extremity of a waveform is fixed at a desired value [9].

CLASS A AMPLIFIER OPERATION—The type of operation in which the amplifier is biased so that variations in input signal polarities occur within the limits of cutoff and saturation [7].

CLASS AB AMPLIFIER OPERATION—The type of operation in which the amplifier is biased so that collector current is cut off for a portion of the alternation of the input signal [7].

CLASS B AMPLIFIER OPERATION—The type of operation in which the amplifier is biased so that collector current is cut off for one-half of the input signal [7].

CLASS C AMPLIFIER OPERATION—The type of operation in which the amplifier is biased so that collector current is cut off for more than one-half of the input signal [7] [13].

CLUTTER—Confusing, unwanted echoes that interfere with the observation of desired signals on a radar indicator [18].

COAXIAL CABLE—Cable in which the center conductor is separated from an outer conductor by a dielectric material; used in RF transmission [4].

COAXIAL LINE—A type of transmission line that contains two concentric conductors [10].

CODE—In teletypewriter operation, code is a combination of mark and space conditions representing symbols, figures, or letters [17].

COEFFICIENT OF COUPLING—An expression of the extent to which two inductors are coupled by magnetic lines of force. This is expressed as a decimal or percentage of maximum possible coupling and represented by the letter K [2].

COHERENCE—A definite phase relationship between two energy waves, such as transmitted frequency and reference frequency [18].

COHERENT—Radiation on one frequency [17].

COHERENT OSCILLATOR—In cw radar an oscillator that supplies phase references to provide coherent video from target returns [18].

COIL—An inductive device made by looping turns of wire around a core [2].

COLD-CATHODE TUBE—A gas-filled electron tube that conducts without the use of filaments. Cold-cathode tubes are used as voltage regulators [6].

COLLECTOR—The element in a transistor that collects the current carriers [7].

COLLECTOR-INJECTION MODULATOR—The transistor equivalent of a plate modulator. Modulating voltage is applied to a collector circuit [12].

COLLINEAR ARRAY—An array with all the elements in a straight line. Maximum radiation is perpendicular to the axis of the elements [10].

COMBINATION ARRAY—An array system that uses the characteristics of more than one array [10].

COMBINATION CIRCUIT—A series-parallel circuit [1].

COMBINATION PEAKING—A technique in which a combination of peaking coils in series and parallel (shunt) with the output signal path is used to improve high-frequency response [8].

COMMON BASE—A transistor circuit in which the base electrode is the common element to both input and output circuits [7].

COMMON-BASE DETECTOR—An amplifying detector in which detection occurs in the emitter-base junction and amplification occurs at the output of the collector junction [12].

COMMON COLLECTOR—A transistor circuit configuration in which the collector is the element common to both the input and the output circuits [7].

COMMON EMITTER—A circuit configuration in which the emitter is the element common to both the input and the output circuits [7].

COMMON-EMITTER DETECTOR—Often used in receivers to supply detected and amplified output. The emitter-base junction acts as the detector [12].

COMMON IDENTITIES LAW—In Boolean algebra this law states that anytime the expression $A(A + B) = AB$ or $A + AB = A + B$ appears, it can immediately be simplified to AB without going through the process of using the distributive law, complementary law, or the law of union to simplify [13].

COMMUTATION—The act of a commutator in converting generator output from an ac voltage to a dc voltage [5].

COMMUTATIVE LAW—In Boolean algebra this law states that changing the order of the terms in an equation will not affect the value of the equation. Example: $A + B = B + A$; $A \bullet B = B \bullet A$ [13].

COMMUTATOR—A mechanical device that reverses armature connections in motors and generators at the proper instant so that current continues to flow in only one direction. In effect, the commutator changes ac to dc [5].

COMPARATOR—An equipment that compares incoming signals and selects the strongest to be fed to a teletypewriter through a patch panel. This is used in diversity operation [17].

COMPENSATING WINDINGS—Windings embedded in slots in pole pieces, connected in series with the armature, whose magnetic field opposes the armature field and cancels armature reaction [5].

COMPENSATION—The process of overcoming the problems associated with high frequencies in an amplifier [8].

COMPLEMENT—A number or state that is the opposite of a specified number or state. The negative of a number is often represented by its complement [13].

COMPLEMENTARY (SECONDARY) COLORS OF LIGHT—The colors of light produced when two of the primaries are mixed in overlapping beams of light. The complementary colors of light are magenta, yellow, and cyan [10].

COMPLEMENTARY LAW—In Boolean algebra this law states that the logical addition of a quantity and its complement will result in 1 and the logical multiplication of a quantity and its complement will result in a product of 0 [13].

COMPLEMENT NUMBER—A number that when added to another number gives a sum equal to the base of the number system of operation. For example, in the decimal number system, the complement of 1 is 9 [13].

COMPLEX WAVE—(1) A waveform other than a sine wave [9]. (2) A wave that is produced by combining two or more pure tones at the same time [10] [12].

COMPOUND-WOUND MOTORS AND GENERATORS—Machines that have a series field in addition to a shunt field. Such machines have characteristics of both series- and shunt-wound machines [5].

COMPRESSION WAVES—Longitudinal waves that have been compressed (made more dense) as they move away from the source [10].

COMPUTER—A data processor that can perform substantial computation, including numerous arithmetic or logic operations, without intervention by a human operator during the run [13].

CONCURRENT—Pertaining to the occurrence of two or more events or activities within the same specified interval of time [13].

CONDUCTANCE—The ability of a material to conduct or carry an electric current. It is the reciprocal of the resistance of the material and is expressed in mhos or siemens [1] [4] [10].

CONDUCTION BAND—A partially filled energy band in which electrons can move freely [7].

CONDUCTIVITY—The ease with which a substance transmits electricity [1].

CONDUCTOR—(1) A material with a large number of free electrons. (2) A material that easily permits electric current to flow [1].

CONDUIT—A tubular raceway, usually metal or plastic, for holding wires or cables [4].

CONICAL SCANNING—Scanning in which the movement of the beam describes a cone, the axis of which coincides with that of the reflector [18].

CONNECTED ARRAY—Another term for DRIVEN ARRAY [10].

CONTACT—In radar, an object that reflects RF energy; target [18].

CONTINUITY—An uninterrupted, complete path for current flow [3] [16].

CONTINUOUS-WAVE KEYING—The on-off keying of a carrier [12].

CONTROL DIFFERENTIAL TRANSMITTER (CDX)—A type of synchro that transmits angular information equal to the algebraic sum or difference of the electrical input supplied to its stator, the mechanical input supplied to its stator, and the mechanical input supplied to its rotor. The output is an electrical voltage taken from the rotor windings [15].

CONTROL GRID—The electrode of a vacuum tube, other than a diode, upon which a signal voltage is impressed to regulate the plate current [6].

CONTROL-GRID MODULATOR—Uses a variation of grid bias to vary the instantaneous plate voltage and current. The modulating signal is applied to the control grid [12].

CONTROL SYNCHRO SYSTEMS—Synchro systems that contain control synchros and are used to control large amounts of power with a high degree of accuracy. The electrical outputs of these systems control servosystems, which in turn generate the required power to move heavy loads [15].

CONTROL SYSTEM—A group of components systematically organized to perform a specific control purpose. These systems are categorized as either closed- or open-loop systems. The main difference between the two is that the closed-loop system contains some form of feedback [15].

CONTROL TRANSFORMER (CT)—A type of synchro that compares two signals: the electrical signal applied to its stator and the mechanical signal applied to its rotor. The output is an electrical voltage, which is taken from the rotor winding and is used to control a power-amplifying device. The phase and amplitude of the output voltage depends on the angular position of the rotor with respect to the magnetic field of the stator [15].

CONTROL TRANSMITTER (CX)—A type of synchro that converts a mechanical input, which is the angular position of its rotor, into an electrical output signal. The output is taken from the stator windings and is used to drive either a CDX or CT [15].

CONVERTER—In communications, equipment that changes the audio output of a receiver to dc pulses. These pulses are fed to a tty to indicate marks and spaces [17].

COOKIE-CUTTER TUNER—A mechanical magnetron tuning device that changes the frequency by changing the capacitance of the anode cavities [11].

COPPER LOSS (I^2R LOSS)—The power lost because of the resistance of the conductors. In transformers the power lost because of current flow (I) through the resistance (R) of the windings [2] [10] [11].

CORDWOOD MODULE—A method of increasing the number of discrete components in a given space. Resembles wood stacked for a fireplace [14].

CORE—Any material that affords a path for magnetic flux lines in a coil [2].

CORNER-REFLECTOR ANTENNA—A half-wave antenna with a reflector consisting of two flat metal surfaces meeting at an angle behind the radiator [10] [18].

CORONA—The discharge of electricity from a conductor with a high potential [4].

CORRECTIVE MAINTENANCE—Includes location and repair of equipment failures [16].

CORRESPONDENCE—The term given to the positions of the rotors of a synchro transmitter and a synchro receiver when both rotors are on 0 degree or displaced from 0 degree by the same angle [15].

COULOMB—A measure of the quantity of electricity. One coulomb is equal to 6.28×10^{18} electrons [1].

COULOMB'S LAW—Also called the LAW OF ELECTRIC CHARGES or the LAW OF ELECTROSTATIC ATTRACTION. Coulomb's Law states that charged bodies attract or repel each other with a force that is directly proportional to the product of their individual charges and inversely proportional to the square of the distance between them [1].

COUNTER—A circuit that counts input pulses [9].

COUNTER EMF—The voltage generated within a coil by a moving magnetic field cutting across the coil itself. This voltage is in opposition (counter) to the moving field that created it. Counter emf is present in every motor, generator, transformer, or other inductance winding whenever an alternating current flows [2] [5].

COUNTERPOISE—A network of wire connected to a quarter-wave antenna at one end. The network provides the equivalent of an additional one-fourth wavelength [10].

COUPLING—The process of transferring energy from one point in a circuit to another point, or from one circuit to another [8].

COUPLING CAPACITOR—A capacitor used to couple signals [8].

COUPLING DEVICE—A coupling coil that connects the transmitter to the feeder [10].

COVALENT BOND—A type of linkage between atoms in which the atoms share valence electrons [7].

CPR—Cardiopulmonary Resuscitation [1].

CREST (TOP)—The peak of the positive alternation (maximum value above the line) of a wave [10].

CRITICAL ANGLE—The maximum angle at which radio waves can be transmitted and still be refracted back to earth [10].

CRITICAL FREQUENCY—The maximum frequency at which a radio wave can be transmitted vertically and still be refracted back to earth [10].

CROSSED-FIELD AMPLIFIER—A high-power electron tube that converts dc to microwave power by a combination of crossed electric and magnetic fields [18].

CROSS-SECTIONAL AREA—The area of a "slice" of an object. When applied to electrical conductors it is usually expressed in circular mils [1].

CROWN-OF-THORNS TUNER—See SPROCKET TUNER [11].

CRYSTAL—A natural substance, such as quartz or tourmaline, that is capable of producing a voltage when under physical stress or of producing physical movement when a voltage is applied [9].

CRYSTAL FURNACE—A device for artificially growing cylindrical crystals to be used in the production of semiconductor substrates [14].

CRYSTAL MICROPHONE—A microphone that uses the piezoelectric effect of crystalline matter to generate a voltage from sound waves [12].

CRYSTAL OVEN—A closed oven maintained at a constant temperature in which a crystal and its holder are enclosed to reduce frequency drift [9].

CURRENT—The movement of electrons past a reference point. The passage of electrons through a conductor. Measured in amperes [1].

CURRENT-FEED METHOD—Same as CENTER-FEED METHOD [10].

CURRENT RATING—The safe current-carrying capacity of a wire or cable on a continuous basis [4].

CURRENT REGULATOR—A circuit that provides a constant current output [7].

CURRENT STANDING-WAVE RATIO (ISWR)—The ratio of maximum to minimum current along a transmission line [10].

CUSPS—Sharp phase reversals [12].

CUTOFF—The condition in a tube or transistor whereby the reverse bias prevents current flow [13].

CUTOFF FREQUENCY—The frequency at which the attenuation of a waveguide increases sharply and below which a traveling wave in a given mode cannot be maintained. A frequency with a half-wavelength that is greater than the wide dimension of a waveguide [11].

CW DEMODULATOR—A circuit that detects the presence of RF oscillations and converts them into a useful form [12].

CYCLE—(1) One complete positive and one complete negative alternation of a current or voltage [2] [10]. (2) A 360-degree rotation of a vector generating a sine wave [12].

CYLINDRICAL PARABOLIC REFLECTOR—A parabolically shaped reflector that resembles part of a cylinder [18].

DAMPED WAVE—A sinusoidal wave in which the amplitude steadily decreases with time. Often associated with energy loss [9].

DAMPING—(1) The process of smoothing out oscillations. (2) In a meter, this process is used to keep the pointer of the meter from overshooting the correct reading [3]. (3) A mechanical or electrical technique used in synchro receivers to prevent the rotor from oscillating or spinning. Damping is also used in servosystems to minimize overshoot of the load [15] [16].

D'ARSONVAL METER MOVEMENT—The permanent-magnet moving-coil movement used in most meters [3] [16].

DATA PROCESSING—The execution of a systematic sequence of operations performed upon data. Synonymous with information processing [13].

DATA TRANSMISSION—The transfer of information from one place to another or from one part of a system to another [15].

dBm—An abbreviation used to represent power levels above or below a 1-milliwatt reference [16].

DEAD SHORT—A short circuit having minimum resistance [1].

DECIMAL—Pertaining to the number representation system with a radix of ten [13].

DECIMAL DIGIT—In decimal notation, one of the characters 0 through 9 [13].

DECIMAL NOTATION—A fixed radix notation where the radix is ten [13].

DECIMAL NUMERAL—A decimal representation of a number [13].

DECIMAL POINT—The radix point in decimal representation [13].

DECOUPLING CAPACITOR—A capacitor used to transfer unwanted signals out of a circuit; for example, coupling an unwanted signal to ground. Also called a BYPASS CAPACITOR [8].

DEFLECTION COILS—In a cathode-ray tube, coils used to bend an electron beam a desired amount [18].

DEFLECTION PLATES—Two pairs of parallel electrodes, one pair set forward of the other and at right angles to each other, parallel to the axis of the electron stream within an electrostatic cathode-ray tube [6].

DEGENERATION—The process whereby a part of the output signal of an amplifying device is returned to its input circuit in such a manner that it tends to cancel part of the input [7].

DEGENERATIVE FEEDBACK—Feedback in which the feedback signal is out of phase with the input signal; also called NEGATIVE FEEDBACK [8].

DEGREE-OF-FREEDOM—The number of axes about which a gyro is free to precess [15].

DEIONIZATION POTENTIAL—The potential at which ionization of the gas within a gas-filled tube ceases and conduction stops; also referred to as extinction potential [6].

DEIONIZATION TIME—In a spark gap, the time required for ionized gas to return to its neutral state after the spark is removed [18].

DELTA—A three-phase connection in which windings are connected end-to-end, forming a closed loop that resembles the Greek letter delta. A separate phase wire is then connected to each of the three junctions [5].

DEMODULATION—The removal of intelligence from a transmission medium [12].

DEMODULATOR—A circuit used in servosystems to convert an ac signal to a dc signal. The magnitude of the dc output is determined by the magnitude of the ac input signal, and its polarity is determined by whether the ac input signal is in or out of phase with the ac reference voltage [15].

DeMORGAN'S THEOREM—A theorem which states that the inversion of a series of AND applications is equal to the same series of inverted OR applications, or the inversion of a series of OR applications is equal to the same series of inverted AND applications. In symbols,

$$\overline{A \cdot B \cdot C} = \overline{A} + \overline{B} + \overline{C} \text{ or}$$
$$\overline{A + B + C} = \overline{A} \cdot \overline{B} \cdot \overline{C} \text{ [13].}$$

DENSITY—(1) The compactness of a substance. (2) Mass per unit volume [10].

DEPLETION REGION—The region in a semiconductor where essentially all free electrons and holes have been swept out by the electrostatic field which exists there [7].

DEPOT-LEVEL MAINTENANCE (SM&R CODE D)—Supports SM&R Code I and SM&R Code O activities through extensive shop facilities and equipment and highly skilled personnel [14].

DESIGNATION—Operational phase of a fire-control or track radar during which the radar is directed to the general direction of a desired target [18].

DETECTION—The separation of low-frequency (audio) intelligence from the high-frequency carrier [17].

DETECTOR—A mixer or converter in a superheterodyne receiver [18].

DICE—Uncased chips [14].

DIE BONDING—Process of mounting a chip to a package [14].

DIELECTRIC—An insulator; a term applied to the insulating material between the plates of a capacitor [2].

DIELECTRIC CONSTANT—The ratio of a given dielectric to the dielectric value of air [2] [11].

DIELECTRIC FIELD—The space between and around charged bodies in which their influence is felt. Also called ELECTRIC FIELD OF FORCE or an ELECTROSTATIC FIELD [1].

DIELECTRIC HEATING—The heating of an insulating material by a high-frequency electric field [10].

DIELECTRIC HYSTERESIS LOSS—Power loss of a capacitor because of the changes in orientation of electron orbits in the dielectric; the changes in orientation are caused by rapid reversal in polarity of line voltage. The higher the frequency, the greater the loss [2].

DIELECTRIC LEAKAGE—Power loss of a capacitor because of the leakage of current through the dielectric. Also relates to leakage resistance; the higher the leakage resistance, the lower the dielectric leakage [2].

DIELECTRIC LOSSES—The losses resulting from the heating effect on the dielectric material between conductors [10] [11].

DIELECTRIC STRENGTH—The ability of an insulator to withstand a potential difference without breaking down (usually expressed in terms of voltage) [4].

DIFFERENCE FREQUENCY—See BEAT FREQUENCY [18].

DIFFERENCE OF POTENTIAL—A voltage between two points [6].

DIFFERENTIAL AMPLIFIER—A circuit that amplifies the difference between two input signals [8].

DIFFRACTION—The bending of waves (as light or RF) when the waves are met with some form of obstruction [10].

DIFFUSION—(1) The scattering of reflected light waves from an object, such as white paper [10]. (2) Controlled application of impurity atoms to a semiconductor substrate [14].

DIGIT—A symbol that represents one of the nonnegative integers smaller than the radix. For example, in decimal notation a digit is one of the characters from 0 through 9 [13].

DIGITAL COMPUTER—(1) A computer in which discrete representation of data is used. (2) A computer that operates on discrete data by performing arithmetic and logic processes on these data [13].

DIODE—An electron tube containing two electrodes: a cathode and a plate [6]. (2) A two element, solid-state device made of either germanium or silicon; it is primarily used as a switching device [7] [13].

DIODE DETECTOR—A demodulator that uses one or more diodes to provide a rectified output with an average value that is proportional to the original modulation [12] [18].

DIPOLE—A common type of half-wave antenna made from a straight piece of wire cut in half. Each half operates at a quarter wavelength of the output [10].

DIRECT CURRENT—An electric current that flows in one direction only [1].

DIRECTIONAL ANTENNA—An antenna that radiates most effectively in only one direction [18].

DIRECTIONAL COUPLER—A device that samples the energy traveling in a waveguide in one direction only [11].

DIRECTIVITY—The ability of an antenna to radiate or receive more energy in some directions than in others. The degree of sharpness of the antenna beam [10] [11] [18].

DIRECTLY HEATED CATHODE—A wire, or filament, designed to emit the electrons that flow from cathode to plate. The filament is designed so that a current is passed through it; the current heats the filament to the point where electrons are emitted [6].

DIRECTOR—The parasitic element of an array that reinforces energy coming from the driver element [10].

DIRECT SHORT—Same as SHORT CIRCUIT [3].

DISCRETE COMPONENTS—Individual transistors, diodes, resistors, capacitors, and inductors [14].

DISCRIMINATOR—A circuit in which amplitude variations are derived in response to phase or frequency variations [18].

DISPERSION—The refraction of light waves that causes the different frequencies to bend at slightly different angles [10].

DISPLACEMENT CURRENT—The current that appears to flow through a capacitor [2].

DISTILLED WATER—Water that has been purified through a process of evaporation and condensation [18].

DISTORTION—Any unwanted change between an input signal and output signal [6] [8].

DISTRIBUTED CONSTANTS—The constants of inductance, capacitance, and resistance in a transmission line. They are spread along the entire length of the line and cannot be distinguished separately [10].

DISTRIBUTIVE LAW—In Boolean algebra the law which states that if a group of terms connected by like operators contains the same variable, the variable may be removed from the terms and associated with them by the appropriate sign of operation (for example, $A(B + C) = AB + AC$) [13].

DOMAIN THEORY—A theory of magnetism based upon the electron-spin principle. Spinning electrons have a magnetic field. If more electrons spin in one direction than another, the atom is magnetized [1].

DOMINANT MODE—The easiest mode to produce in a waveguide, and the most efficient mode in terms of energy transfer [11].

DONOR—An impurity that can make a semiconductor material an N-type by donating extra "free" electrons to the conduction band [7].

DONOR IMPURITY—See PENTAVALENT IMPURITY [7].

DOORKNOB TUBE—An electron tube that is similar to the acorn tube but larger. The doorknob tube is designed to operate, at high power, in the uhf frequencies [6].

DOPING—The process of adding impurities to semiconductor crystals to increase the number of free charges that can be moved by an external, applied voltage. Doping produces N-type or P-type material [7] [14].

DOPPLER EFFECT—(1) The apparent change in frequency or pitch when a sound source moves either toward or away from a listener [10]. (2) In radar, the change in frequency of a received signal caused by the relative motion between the radar and the target [18].

DOPPLER FREQUENCY—The difference between transmitted and reflected frequencies; caused by the Doppler effect [18].

DOUBLE-MODING—In a transmitter output tube, the abrupt and random change from one frequency to another [18].

DOUBLE NEGATIVE LAW—In Boolean algebra, the law which states that the complement of a complement is the equivalent of the original term [13].

DOUBLE RECEIVER—A fine and coarse synchro receiver enclosed in a common housing with a two-shaft output (one shaft inside the other) [15].

DOUBLET—Another name for the dipole antenna [10].

DOUBLING UP—This is a type of two-equipment installation where one unit can be substituted for another in the event of failure [17].

DOWN LINK—The frequency used to transmit an amplified signal from a satellite or other craft back to earth [17].

DRIFT SPACE—In an electron, a region free of external fields in which relative electron position depends on velocity [11].

DRIVEN ARRAY—An array in which all of the elements are driven [10].

DRIVEN ELEMENT—The element of an antenna connected directly to the transmission line [10].

DRIVER—The final stage of amplification [8].

DRUM-TYPE ARMATURE—An efficient, popular type of armature designed so that the entire length of the winding is cutting the field at all times. Most wound armatures are of this type [5].

DRY-AIR SYSTEM—Provides dehumidified air for electronic equipment that is moisture critical [18].

DRY CELL—An electrical cell in which the electrolyte is not a liquid. In most dry cells the electrolyte is in the form of a paste [1].

DUAL-GATE MOSFET—A two-gate MOSFET in which either gate can control the conductor independently, a fact which makes this MOSFET very versatile [7].

DUAL IN-LINE PACKAGE (DIP)—IC package having two parallel rows of preformed leads [14].

DUCTILE—Easily drawn out (as to form filaments or wires) [4].

DUCTING—Trapping of an RF wave between two layers of the earth's atmosphere or between an atmospheric layer and the earth [18].

DUMMY ANTENNA—See DUMMY LOAD [16].

DUMMY LOAD—A dissipative but nonradiating device that has the impedance characteristics of an antenna or transmission line. Also called ARTIFICIAL LOAD [11] [16] [17].

DUPLEXER—A radar device that switches the antenna from the transmitter to the receiver and vice versa [18].

DUTY CYCLE—In a transmitter, ratio of time on to time off [12] [18].

DYNAMIC MICROPHONE—A device in which sound waves move a coil of fine wire that is mounted on the back of a diaphragm and located in the magnetic field of a permanent magnet [12].

ECHO—(1) The reflection of the original sound wave as it bounces off a distant surface [10]. (2) The RF signal reflected back from a radar target [18].

ECHO BOX—A resonant cavity device that is used to check the overall performance of a radar system. It receives a portion of the transmitted pulse and retransmits it back to the receiver as a slowly decaying transient [18].

ECLIPSE—A condition in which the satellite is not in view or in direct line of sight with the sun. This happens when the earth is between them [17].

EDDY CURRENT—Induced circulating currents in a conducting material that are caused by a varying magnetic field [2] [5].

EDDY CURRENT LOSS—Losses caused by random current flowing in the core of a transformer. Power is lost in the form of heat [2].

EDISON EFFECT—Also called RICHARDSON EFFECT. The phenomenon wherein electrons emitted from a heated element within a vacuum tube will flow to a second element that is connected to a positive potential [6].

EFFECTIVE VALUE—Same as ROOT-MEAN-SQUARE [2].

EFFICIENCY—The ratio of output-signal power compared to the total input power, generally expressed as a percentage [1] [7].

E-FIELD—Electric field that exists when a difference in electrical potential causes a stress in the dielectric between two points [11].

ELASTICITY—The ability of a substance to return to its original state [10].

ELECTRIC CURRENT—The flow of electrons [1].

ELECTRIC (E) FIELD—The field of force that is produced as a result of a voltage charge on a conductor or antenna [10] [11].

ELECTRICAL CHARGE—Symbol **Q, q**. Electric energy stored on or in an object. The negative charge is caused by an excess of electrons; the positive charge is caused by a deficiency of electrons [1].

ELECTRICAL CHEMICAL—The action of converting chemical energy into electrical energy [1].

ELECTRICAL-LOCK—A synchro zeroing method. This method is used only when the rotors of the synchros to be zeroed are free to turn and their leads are accessible [15].

ELECTRICAL POWER SYSTEM—Provides the necessary input power [18].

ELECTRICAL SYMBOLS—Graphic symbols used to illustrate the various electrical or electronic components of a circuit [4].

ELECTRICAL ZERO—A standard synchro position, with a definite set of stator voltages, that is used as the reference point for alignment of all synchro units [15].

ELECTRODE—The terminal at which electricity passes from one medium into another, such as in an electrical cell where the current leaves or returns to the electrolyte [1].

ELECTRODYNAMIC METER MOVEMENT—A meter movement using fixed field coils and a moving coil; usually used in ammeters and wattmeters [3].

ELECTRODYNAMOMETER—A meter using an electrodynamic movement to measure an electric current [16].

ELECTROLYSIS—The process of changing the chemical composition of a material by passing an electric current through it [4] [11].

ELECTROLYTE—A solution of a substance that is capable of conducting electricity. An electrolyte may be in the form of either a liquid or a paste [1].

ELECTROMAGNET—An electrically excited magnet capable of exerting mechanical force or of performing mechanical work [1].

ELECTROMAGNETIC—The term describing the relationship between electricity and magnetism. A quality that combines both magnetic and electric properties [1].

ELECTROMAGNETIC FIELD—The combination of an electric (E) field and a magnetic (H) field [10].

ELECTROMAGNETIC INDUCTION—The production of a voltage in a coil because of a change in the number of magnetic lines of force (flux linkages) passing through the coil [1] [2].

ELECTROMAGNETIC INTERFERENCE—Man-made or natural interference that degrades the quality of reception of radio waves [10] [17].

ELECTROMAGNETIC RADIATION—The radiation of radio waves into space [10].

ELECTROMAGNETISM—The generation of a magnetic field around a current-carrying conductor [2] [3].

ELECTROMOTIVE FORCE (emf)—The force (voltage) that produces an electric current in a circuit [2].

ELECTRON—The elementary negative charge that revolves around the nucleus of an atom [1].

ELECTRON GUN—An electrode of a CRT that is equivalent to the cathode and control grid of conventional tubes. The electron gun produces a highly concentrated stream of electrons [6].

ELECTRON ORBITAL MOVEMENT—The movement of an electron around the nucleus of an atom [11].

ELECTRON SHELL—A group of electrons which have a common energy level that forms part of the outer structure (shell) of an atom [1].

ELECTRONIC COUNTER-COUNTERMEASURES (ECCM) CIRCUITS—See ANTIJAMMING CIRCUITS [18].

ELECTRONIC-EQUIPMENT DEHYDRATOR—A device that provides an alternate dry-air input in the event of failure of the central dry-air system. It may include a compressor [18].

ELECTRONIC FREQUENCY COUNTER—An instrument that counts the number of cycles (pulses) occurring during a precise time interval [18].

ELECTRONIC SCANNING—Scanning in which the axis of the beam is moved, relative to the antenna axis, in a desired pattern [18].

ELECTRONIC SWITCH—A circuit that causes a start-and-stop switching action by electronic means [13].

ELECTRONICS DRY-AIR BRANCH—A common line for providing dry air to various electronic equipment, such as search radar, fire-control radar, and repeaters [18].

ELECTRONIC TUNING—In a reflex klystron, changing the frequency and output power of the tube by altering the repeller voltage [11].

ELECTRON SPIN—The movement of an electron around its axis [11].

ELECTROSTATIC—Pertaining to electricity at rest, such as charges on an object (static electricity) [1].

ELECTROSTATIC DEFLECTION—The method of deflecting an electron beam by passing it between parallel charged plates mounted inside a cathode-ray tube [6].

ELECTROSTATIC FIELD—The field of influence between two differently charged bodies [2].

ELECTROSTATIC METER MOVEMENT—A meter movement that uses the electrostatic repulsion of two sets of charged plates (one fixed and the other movable). This meter movement reacts to voltage rather than to current and is used to measure high voltage [3].

ELECTROSTATIC STRESS—The force exerted on an insulator by the voltage in a conductor [4].

ELEMENT—(1) A substance, in chemistry, that cannot be divided into simpler substances by any means ordinarily available [1]. (2) A part of an antenna that can be either an active radiator or a parasitic radiator [10].

ELEPHANT TRUNK—Ducting used for ventilation purposes [4].

ELEVATION ANGLE—The angle between the horizontal plane and the line of sight to a target or object [11] [18].

EMERGENCY POWER—Temporary source of limited electrical power used upon the loss of the normal power source [18].

EMF (ELECTROMOTIVE FORCE)—The force that causes electricity to flow between two points with different electrical charges or when there is a difference of potential between the two points. The unit of measurement is volts [1].

EMITTER—The element in a transistor that emits current carriers (electrons or holes) [7] [13].

EMITTER-INJECTION MODULATOR—The transistor equivalent of the cathode modulator. The gain is varied by changing the voltage on the emitter [12].

ENAMEL—A synthetic compound of cellulose acetate (wood pulp and magnesium). Used to insulate wire in meters, relays, and motor windings [4].

ENCAPSULATED—Imbedded in solid material or enclosed in glass or metal [14].

END-FEED METHOD—A method in which one end of an antenna is connected through a capacitor to the final output stage of a transmitter [10].

END-FIRE ARRAY—An array in which the direction of radiation is parallel to the axis of the array [10].

ENERGY—The ability or capacity to do work [1].

EPHEMERIS—A table showing the precalculated position of a satellite at any given time [17].

E_p-I_p CURVE—The characteristic curve of an electron tube used to graphically depict the relationship between plate voltage (E_p) and plate current (I_p) [6].

EPITAXIAL PROCESS—A method of depositing a thin, uniformly doped crystalline region (layer) on a substrate [14].

EQUATORIAL ORBIT—An orbit that occurs when the plane of a satellite coincides with the plane of the earth at the equator [17].

EQUIVALENT RESISTANCE (R_{eq})—A resistance that represents the total ohmic values of a circuit component or group of circuit components. Usually drawn as a single resistor in a simplified circuit [1].

ERECTING (A GYRO)—The positioning of a gyro into a desired position and the maintaining of that position [15].

ERROR DETECTOR—The component in a servosystem that determines when the load has deviated from its ordered position, velocity, and so forth [15].

ERROR REDUCER—The name commonly given to the servomotor in a servosystem. So named because it reduces the error signal by providing feedback to the error detector [15].

ERROR SIGNAL—(1) In servosystems, the signal whose amplitude and polarity or phase are used to correct the alignment between the controlling and the controlled elements. (2) The name given to the electrical output of a control transformer [15].

E-TRANSFORMER—A special form of differential transformer employing an E-shaped core. The secondaries of the transformer are wound on the outer legs of the E, and the primary is wound on the center leg. An output voltage is developed across the secondary coils when its armature is displaced from its neutral position. This device is used as an error detector in servosystems that have limited load movements [15].

E-TYPE T-JUNCTION—A waveguide junction in which the junction arm extends from the main waveguide in the same direction as the E-field in the waveguide [11].

EUTECTIC ALLOY—An alloy that changes directly from a solid to a liquid with no plastic or semiliquid state [14].

EUTECTIC SOLDER—An alloy of 63 percent tin and 37 percent lead. Melts at 361° F [14].

EXCITATION VOLTAGE—The supply voltage required to activate a circuit [15].

EXCITING CURRENT—The current that flows in the primary winding of a transformer when the secondary is open-circuited; it produces a magnetic flux field. Also called magnetizing current [2].

EXCLUSIVE OR—A function whose output is a 1 if one and only one of the input variables is a 1 [13].

EXCLUSIVE-OR GATE—A gate that produces a logic 1 output when the inputs are different, but not when they are the same [13].

EXPONENT—The numeral written in superscript (10^2) which indicates the power to which the base is to be raised [13].

EXPRESSION—A validated series of variables, constants, and functions that can be connected by operating symbols to describe a desired computation [13].

EXTERNALLY EXCITED METER—A term used to describe meters that get their power from the circuit to which they are connected [16].

EXTERNALLY SYNCHRONIZED RADAR—A radar system in which timing pulses are generated by a master oscillator external to the transmitter [18].

EXTREMELY HIGH FREQUENCY—The band of frequencies from 30 gigahertz to 300 gigahertz [17].

EXTREMELY LOW FREQUENCY—The band of frequencies up to 300 hertz [17].

EXTRINSIC—A semiconductor in which impurities have been added to create certain charge carrier concentrations [7].

FACSIMILE—The method for transmitting and receiving still images. These images can be maps, photographs, and handwritten or printed text [17].

FACTOR—Any of the elements, quantities, or symbols that, when multiplied together, form a product [13].

FADING—Variations in signal strength by atmospheric conditions [101 [17].

FARAD—The basic unit of capacitance. A capacitor has a capacitance of 1 farad when a voltage potential of 1 volt across it produces a charge of 1 coulomb [2].

FARADAY ROTATION—The rotation of the plane of polarization of electromagnetic energy when it passes a substance influenced by a magnetic field that has a component in the direction of propagation [11].

FAST-TIME-CONSTANT CIRCUIT—Differentiator circuit in the first video amplifier that allows only the leading edges of target returns, no matter how small or large, to be used [18].

FEEDBACK—The return of a portion of the output of a circuit to its input [8] [18].

FEEDER—A transmission line that carries energy to the antenna [10].

FEEDHORN—A horn radiator used to feed a reflector [18].

FEP—A synthetic type of insulation (fluorinated ethylene propylene) [4].

FERRITE—A powdered and compressed ferric oxide material that has both magnetic properties and light resistance to current flow [11].

FERRITE SWITCH—A ferrite device that blocks the flow of energy through a waveguide by rotating the electric field 90 degrees. The rotated energy is then reflected or absorbed [11].

FERROMAGNETIC MATERIAL—A highly magnetic material, such as iron, cobalt, nickel, or their alloys [1].

FERRULES—The cylindrical metallic ends of a cartridge fuse [3].

FIBER OPTICS—Conductors or optical waveguides that readily pass light [17].

FIBROUS BRAID—An outer covering used to protect a conductor's insulating material. Commonly made from cotton, linen, silk, rayon, or fiberglass [4].

FIDELITY—(1) The faithful reproduction of a signal. (2) The accuracy with which a system reproduces a signal at its output that faithfully maintains the essential characteristics of the input signal [7] [8] [12] [17].

FIELD—The electromagnet which furnishes the magnetic field that interacts with the armature in motors and generators [5].

FIELD-EFFECT TRANSISTOR (FET)—A transistor consisting of a source, a gate, and a drain. Current flow is controlled by the transverse electric field under the gate [7].

FIELD EXCITATION—The creation of a steady magnetic field within the field windings by the application of a dc voltage either from the generator itself or from an external source [5].

FIELD OF FORCE—A term used to describe the total force exerted by an action-at-a-distance phenomenon such as gravity upon matter, electric charges acting upon electric charges, and magnetic forces acting upon other magnets or magnetic materials [1].

FILAMENT—The cathode of a thermionic tube, usually a wire or ribbon, which is heated by current passing through it [6].

FILM ICs—Conductive or nonconductive material deposited on a glass or ceramic substrate. Used for passive circuit components, resistors, and capacitors [14].

FILTER—A selective network of resistors, capacitors, and inductors that offers comparatively little opposition to certain frequencies, while blocking or attenuating other frequencies [6] [9].

FINAL POWER AMPLIFIER (FPA)—The final stage of amplification in a transmitter [12].

FIRST DETECTOR—See MIXER [18].

FIXED BIAS—A constant value of bias voltage [6] [7] [13].

FIXED RESISTOR—A resistor having a definite resistance value that cannot be adjusted [1].

FIXED SPARK GAP—A device used to discharge the pulse-forming network. A trigger pulse ionizes the air between two contacts to initiate the discharge [12].

FLAT LINE—A transmission line that has no standing waves. This line requires no special timing devices to transfer maximum power [10].

FLAT PACK—An IC package [14].

FLEMING VALVE—An earlier name for a diode, or a two-electrode vacuum tube used as a detector [6].

FLEXIBLE COAXIAL LINE—A line made with an inner conductor that consists of flexible wire insulated from the outer conductor by a solid, continuous insulating material [10].

FLIP CHIP—A monolithic IC packaging technique that eliminates the need for bonding wires [14].

FLIP-FLOP—A device having two stable states and two input terminals (or types of input signals), each of which corresponds with one of the two states. The circuit remains in either state until caused to change to the other state by application of a voltage pulse. A similar bistable device with an input that allows it to act as a single-stage binary counter [13].

FLUX—(1) In electrical or electromagnetic devices, a general term used to designate collectively all the electric or magnetic lines of force in a region [1]. (2) A solution that removes surface oxides from metals being soldered [2] [14].

FLUX DENSITY—The number of magnetic lines of force passing through a given area [1].

FLYWHEEL EFFECT—The ability of a resonant circuit to operate continuously because of stored energy or energy pulses [9].

FOCUSING ANODE—An electrode of a CRT that is used to focus the electrons into a tight beam [6].

FOLDED DIPOLE—An ordinary half-wave antenna (dipole) that has one or more additional conductors connected across the ends parallel to each other [10].

FORBIDDEN BAND—The energy band in an atom lying between the conduction band and the valence band. Electrons are never found in the forbidden band but may travel back and forth through it. The forbidden band determines whether a solid material will act as a conductor, a semi-conductor, or an insulator [7].

FORWARD AGC—The type of AGC that causes an amplifier to be driven towards saturation [17].

FORWARD BIAS—An external voltage that is applied to a PN junction in the conducting direction so that the junction offers only minimum resistance to the flow of current. Conduction is accomplished by majority current carriers (holes in P-type material; electrons in N-type material) [7] [13] [14].

FORWARD RESISTANCE—The smaller resistance value observed when you are checking the resistance of a semiconductor [16].

FOSTER-SEELEY DISCRIMINATOR—A circuit that uses a double-tuned RF transformer to convert frequency variations in the received FM signal to amplitude variations. Also known as a phase-shift discriminator [12].

FOUR-ELEMENT ARRAY—An antenna array with three parasitic elements and one driven element [10].

FRAMING—The process of synchronizing a facsimile receiver to a transmitter. This allows proper picture reproduction [17].

FREE CHARGES—Those electrons that can be moved by an externally applied voltage [7].

FREE-SPACE LOSS—The loss of energy of radio waves caused by the spreading of the wavefront as it travels from the transmitter [10].

FREQUENCY (f)—(1) The number of complete cycles per second existing in any form of wave motion, such as the number of cycles per second of an alternating current [2] [10]. (2) The rate at which the vector that generates a sine wave rotates [12].

FREQUENCY COMPENSATION NETWORK—Circuit modification used to improve or broaden the linearity of its frequency response [18].

FREQUENCY CUTOFF—The frequency at which the filter circuit changes from an action of rejecting the unwanted frequencies to an action of passing the desired frequencies. Conversely, the point at which the filter circuit changes from an action in which it passes the desired frequencies to an action in which it rejects the undesired frequencies [9].

FREQUENCY-DETERMINING NETWORK—A circuit that provides the desired response (maximum or minimum impedance) at a specific frequency [8].

FREQUENCY DEVIATION—The amount the frequency varies from the carrier frequency [12].

FREQUENCY DIVERSITY—Transmitting (and receiving) of radio waves on two different frequencies simultaneously [10].

FREQUENCY-DIVISION MULTIPLEXING—Multiplexing that transmits and receives the full 360 degrees of each sine wave [17].

FREQUENCY METER—A meter used to measure the frequency of an ac signal [3] [16].

FREQUENCY MODULATION (fm)—Angle modulation in which the modulating signal causes the carrier frequency to vary. The amplitude of the modulating signal determines how far the frequency changes, and the frequency of the modulating signal determines how fast the frequency changes [12].

FREQUENCY MULTIPLIERS—Special RF power amplifiers that multiply the input frequency [12].

FREQUENCY RESPONSE—The measure of a servo's ability to respond to various input frequencies [15].

FREQUENCY-RESPONSE CURVE—A curve showing the output of an amplifier (or any other device) in terms of voltage or current plotted against frequency with a fixed-amplitude input signal [8].

FREQUENCY SCANNING—Varying the output frequency to achieve electronic scanning [18].

FREQUENCY-SHIFT KEYING (fsk)—Frequency modulation somewhat similar to continuous-wave (cw) keying in AM transmitters. The carrier is shifted between two differing frequencies by opening and closing a key [12].

FREQUENCY SPECTRUM—In a radar, the entire range of frequencies contained in an RF pulse or signal [18].

FREQUENCY STABILITY—Refers to the ability of an oscillator to accurately maintain its operating frequency [9].

FREQUENCY SYNTHESIS—A process that uses hetrodyning and frequency selection to produce a signal [17].

FREQUENCY SYNTHESIZER—(1) A frequency source of high accuracy [17]. (2) A bank of oscillators in which the outputs can be mixed in various combinations to produce a wide range of frequencies [18].

FRONT-TO-BACK RATIO—The ratio of the energy radiated in the principal direction compared to the energy radiated in the opposite direction [10].

FULL-WAVE RECTIFIER—A circuit that uses both positive and negative alternations in an alternating current to produce direct current [6] [7].

FULL-WAVE VOLTAGE DOUBLER—Consists of two half-wave voltage rectifiers and is used to reduce the output ripple amplitude [7].

FUNCTION—A specific purpose of an entity; its characteristic action [13].

FUNDAMENTAL FREQUENCY—The basic frequency or first harmonic frequency [10].

FUSED-ALLOY JUNCTION—See ALLOYED-JUNCTION [7].

GAIN—(1) The ratio between the amount of energy propagated from an antenna that is directional compared to the energy from the same antenna that would be propagated if the antenna were not directional [10]. (2) Any increase in the strength of a signal [18].

GAIN-BANDWIDTH PRODUCT—The number that results when the gain of a circuit is multiplied by the bandwidth of that circuit. For an operational amplifier, the gain-bandwidth product for one configuration will always equal the gain-bandwidth product for any other configuration of the same amplifier [8].

GALENA—A crystalline form of lead sulfide used in early radio receivers [7].

GALVANOMETER—A meter used to measure small values of current by electromagnetic or electrodynamic means [3] [4] [16].

GAMMA (γ)—The emitter-to-base current ratio in a common-collector configuration [7].

GANGED TUNING—The process used to tune two or more circuits with a single control [17].

GAS—One of the three states of matter; it has no fixed form or volume [1].

GATE—As applied to logic circuitry, one of several different types of electronic devices that will provide a particular output when specified input conditions are satisfied. Also, a circuit in which a signal switches another signal on or off [13].

GATED AGC—Circuit that permits automatic gain control to function only during short time intervals [18].

GATED-BEAM DETECTOR—An FM demodulator that uses a special gated-beam tube to limit, detect, and amplify the received FM signal. Also known as a quadrature detector [12].

GATING—The process of selecting those portions of a wave that exist during one or more selected time intervals or that have magnitudes between selected limits. Also, the application of a specific waveform to perform electronic switching [13].

GENERAL PURPOSE ELECTRONIC TEST EQUIPMENT (GPETE)—Test equipment that has the capability, without modification, to generate, modify, or measure a range of electronic functions required to test several equipments or systems of basically different designs [14] [16].

GENERATOR—A machine that converts mechanical energy to electrical energy by applying the principle of magnetic induction. A machine that produces ac or dc voltage, depending on the original design [5].

GENERATOR END—See INPUT END [10].

GERMANIUM—A grayish-white metal having semiconductor properties [7].

GETTER—An alkali metal introduced into a vacuum tube during manufacture. It is fired after the tube has been evacuated to react chemically with (and eliminate) any remaining gases [6].

GIMBAL—A mechanical frame, with two perpendicular intersecting axes of rotation, used to support and furnish a gyro wheel with the necessary freedom to tilt in any direction [15].

GLOW DISCHARGE—Discharge of electricity through a gas in an electron tube [18].

GRAMME-RING ARMATURE—An inefficient type of armature winding in which many of the turns are shielded from the field by its own iron ring [5].

GRAPH—A pictorial presentation of the relation between two or more variable quantities, such as between an applied voltage and the current it produces in a circuit [1].

GRID BIAS—A constant fixed potential applied between the grid and the cathode of a vacuum tube to establish an operating point [6].

GRID CURRENT—The current that flows in the grid-to-cathode circuit of a vacuum tube [6].

GRID-GAP TUNING—A method of changing the center frequency of a resonant cavity by physically changing the distance between the cavity grids [11].

GRID-LEAK BIAS—A self-bias provided by a high resistance connected across the grid capacitor or between the grid and cathode [6].

GROUND—(1) The point in a circuit used as a common reference point for measuring purposes. (2) To connect some point of an electrical circuit or some item of electrical equipment to earth or to the conducting medium used in lieu thereof [13].

GROUND CLUTTER—Unwanted echoes, from surrounding land masses, that appear on a radar indicator [18].

GROUND-CONTROLLED APPROACH—A radar system used to guide aircraft to safe landings in poor visibility conditions [18].

GROUND PLANE—The portion of a ground-plane antenna that acts as ground [10].

GROUND-PLANE ANTENNA—A type of antenna that uses a ground plane as a simulated ground to produce low-angle radiation [10].

GROUND PLANES—Copper planes used to minimize interference between circuits and from external sources [14].

GROUND POTENTIAL—Zero potential with respect to the ground or earth [1].

GROUND RANGE—The distance on the surface of the earth between a radar and its target. Equal to slant range only if both radar and target are at the same altitude [18].

GROUND REFLECTION LOSS—The loss of RF energy each time a radio wave is reflected from the earth's surface [10].

GROUND SCREEN—A series of conductors buried below the surface of the earth and arranged in a radial pattern. Used to reduce losses in the ground [10].

GROUND WAVES—Radio waves which travel near the surface of the earth [10].

GROUP—A collection of units, assemblies, subassemblies, and parts. It is a subdivision of a set or system but is not capable of performing a complete operational function [17].

GROUP VELOCITY—The forward progress velocity of a wave front in a waveguide [11].

GROWN JUNCTION—A method of mixing P-type and N-type impurities into a single crystal while the crystal is being grown [7].

GUIDANCE RADAR—A system which provides information that is used to guide a missile to a target [18].

GYRO—Abbreviation for gyroscope [15].

GYROSCOPE—A mechanical device containing a spinning mass mounted so that it can assume any position in space [15].

HALF-POWER POINT—A point on a waveform or radar beam that corresponds to half the power of the maximum power point [8] [9] [18].

HALF-WAVE DIPOLE ANTENNA—An antenna, consisting of two rods (1/4 wavelength each) in a single line, that radiates electromagnetic energy [10].

HALF-WAVE RECTIFIER—A rectifier using only one-half of each cycle to change ac to pulsating dc [61 [7].

HALF-WAVE VOLTAGE DOUBLER—Two half-wave voltage rectifiers connected to double the input voltage [7].

HAND OVER—The operation where one earth terminal yields control to another as a satellite moves out of its area of coverage [17].

HARD-TUBE MODULATOR—A high-vacuum electron tube modulator that uses a driver for pulse forming [18].

HARMONIC—A frequency that is a whole-number multiple of a smaller base frequency [9] [10] [12] [17].

HEATER—Same as a FILAMENT [6].

HEAT SHUNT—A device (preferably a clip-on type) used to absorb heat and protect heat-sensitive components during soldering [4].

HEIGHT-FINDING RADAR—A radar that provides target altitude, range, and bearing data [18].

HELIX—(1) A spirally wound transmission line used in a traveling-wave tube to delay the forward progress of the input traveling wave [11]. (2) A large coil of wire. It acts as a coil and is used with variable inductors for impedance matching of high-power transmitters [17].

HELIX HOUSE—A building at a transmitter site that contains antenna loading, coupling, and tuning circuits [17].

HENRY (H)—The electromagnetic unit of inductance or mutual inductance. The inductance of a circuit is 1 henry when a current variation of 1 ampere per second induces 1 volt. In electronics, smaller units are used, such as the millihenry (mH), which is one-thousandth of a henry (H), and the microhenry (μH) which is one-millionth of a henry [2].

HERTZ (Hz)—A unit of frequency equal to one cycle per second [2].

HERTZ ANTENNA—A half-wave antenna that is installed some distance above ground and positioned either vertically or horizontally [10].

HETERODYNE DETECTION—The use of an a.f. voltage to distinguish between available signals. The incoming cw signal is mixed with locally generated oscillations to give an a.f. output [12].

HETERODYNING—(1) The process of mixing two frequencies across a nonlinear impedance [12]. (2) The process of mixing the incoming signal with the local oscillator frequency. This produces the two fundamentals and the sum and difference frequencies [17].

HEXADECIMAL—Same as SEXADECIMAL. A number system with a base of sixteen; also pertains to conditions, choices, or selections that have sixteen possible values or states [13].

HEXADECIMAL SYSTEM—Pertaining to the number system with a radix of sixteen. It uses the ten digits of the decimal system and the first six letters of the English alphabet [13].

H-FIELD—Any space or region in which a magnetic force is exerted. The magnetic field may be produced by a current-carrying coil or conductor, by a permanent magnet, or by the earth itself [11].

HIGH FREQUENCY—The band of frequencies from 3 megahertz to 30 megahertz [17].

HIGH-FREQUENCY COMPENSATION—See PEAKING COIL [8].

HIGH-LEVEL MODULATION—Modulation produced in the plate circuit of the last radio stage of the system [12].

HIGH-PASS FILTER—A filter that passes a majority of the high frequencies on to the next circuit and rejects, or attenuates, the lower frequencies. Also called a LOW-FREQUENCY DISCRIMINATOR [9].

HITS PER SCAN—The number of times an RF beam strikes a target per antenna revolution [18].

HOLE FLOW—In the valence band, a process of conduction in which electrons move into holes, thereby creating other holes that appear to move toward a negative potential. (The movement of holes is opposite the movement of electrons.) [7]

HORIZONTAL AXIS—On a graph, the straight line axis that is plotted from left to right [10].

HORIZONTAL-DEFLECTION PLATES—A pair of parallel electrodes that moves the electron beam from side to side in a CRT [6].

HORIZONTALLY POLARIZED—Waves radiated with their E field component parallel to the earth's surface [10].

HORIZONTAL PATTERN—The part of a radiation pattern that is radiated in all directions along the horizontal plane [10].

HORIZONTAL PLANE—An imaginary plane that is tangent (or parallel) to the earth's surface at a given location [11] [18].

HORN—A funnel-shaped section of waveguide used as a termination device and as a radiating antenna [11].

HORN ANTENNA—See HORN RADIATOR [18].

HORN RADIATOR—A tapered, tubular or rectangular microwave antenna that is widest at the open end [18].

HORSEPOWER—The English unit of power equal to work done at the rate of 550 foot-pounds per second; equal to 746 watts of electrical power [1].

HORSESHOE MAGNET—A permanent magnet or electromagnet bent into the shape of a horseshoe or having a U-shape to bring the two poles near each other [1].

HOT CARRIER—A carrier, which may be either a hole or an electron, that has relatively high energy with respect to the carriers normally found in majority-carrier devices [11].

HOT-CARRIER DIODE—A semiconductor diode in which hot carriers are emitted from a semiconductor layer into the metal base. Also called HOT-ELECTRON DIODE. An example is the Schottky barrier diode [11].

HOT-WIRE METER MOVEMENT—A meter movement that uses the expansion of a heated wire to move the pointer of a meter; measures dc or ac [3].

H-TYPE T-JUNCTION—A waveguide junction in which the junction arm is parallel to the magnetic lines of force in the main waveguide [11].

HYBRID CIRCUIT—A circuit where passive components (resistors, capacitors) are deposited onto a substrate made of glass, ceramic, or other insulating material. Then the active components (diodes, transistors) are attached to the substrate and connected to the passive components on the substrate with a very fine wire [7].

HYBRID ICs—Two or more integrated circuit types, or one or more integrated circuit types and discrete components on a single substrate [14].

HYBRID JUNCTION—A waveguide junction that combines two or more basic T-junctions [11].

HYBRID MIXER—See BALANCED MIXER [18].

HYBRID RING—A hybrid-waveguide junction that combines a series of E-type T-junctions in a ring configuration. When properly terminated, energy is transferred from any one branch into any two of the remaining three branches [11] [18].

HYDROMETER—An instrument used to measure specific gravity. In batteries hydrometers are used to indicate the state of charge by the specific gravity of the electrolyte [1].

HYSTERESIS—The time lag of the magnetic flux in a magnetic material behind the magnetizing force producing it. Caused by the molecular friction of the molecules trying to align themselves with the magnetic force applied to the material [2].

HYSTERESIS LOSS—The power loss in an iron-core transformer or other alternating-current device as a result of magnetic hysteresis [2].

IC SYNCHROS—Obsolete synchros with reverse rotation and limited torque capabilities [15].

IDEMPOTENT LAW—In Boolean algebra, combining a quantity with itself either by logical addition or logical multiplication will result in a logical sum or product that is the equivalent of the quantity (for example, A + A = A; A • A = A) [13].

IDENTITY LAW—In Boolean algebra, the law which states that any expression is equal to itself (for example,

$$A = A, \text{ or } A = A \quad [13]$$

IDLER FREQUENCY—In a parametric amplifier, the difference between the input signal and the pump signal frequency. Also called the LOWER-SIDEBAND FREQUENCY [11].

IF AMPLIFIER—Usually a narrow-bandwidth IF amplifier that is tuned to one of the output frequencies produced by the mixer [18].

IGFET—Any field-effect transistor that has an insulated gate [7].

IMAGE FREQUENCY—An undesired frequency capable of producing the desired frequency through heterodyning [17].

IMPEDANCE—The total opposition offered to the flow of an alternating current. It may consist of any combination of resistance, inductive reactance, and capacitive reactance. The symbol for impedance is Z [2] [9].

IMPLOSION—The inward bursting of a CRT because of high vacuum. The opposite of explosion [6].

INCIDENT WAVE—(1) The wave that strikes the surface of a medium. (2) The wave that travels from the sending end to the receiving end of a transmission line [10].

IN-CIRCUIT METER—A meter permanently installed in a circuit; used to monitor circuit operation [3].

INCOHERENT—Refers to radiation on a broad band of frequencies [17].

INDEX OF REFRACTION—The degree of bending of an RF wave when passing from one medium to another [18].

INDICATOR—Equipment in radar that provides a visual presentation of target position information [18].

INDIRECTLY HEATED CATHODE—Same as the directly heated cathode with one exception: The hot filament raises the temperature of the sleeve around the filament; the sleeve then becomes the electron emitter [6].

INDUCED-CHANNEL MOSFET—A MOSFET in which there is no actual channel between the source and the drain. This MOSFET is constructed by making the channel of the same type of material as the substrate [7].

INDUCED CHARGE—An electrostatic charge produced on an object by the electric field that surrounds a nearby object [1].

INDUCED CURRENT—Current caused by the relative motion between a conductor and a magnetic field [1].

INDUCED ELECTROMOTIVE FORCE—The electromotive force induced in a conductor because of the relative motion between the conductor and a magnetic field [1].

INDUCED VOLTAGE—See INDUCED ELECTROMOTIVE FORCE [1].

INDUCTANCE—The property of a circuit that tends to oppose a change in the existing current flow. The symbol for inductance is L [2] [7].

INDUCTANCE BRIDGE—An ac bridge circuit used to measure an unknown value of inductance [16].

INDUCTION—The act or process of producing voltage and current by the relative motion of a magnetic field across a conductor [1].

INDUCTION FIELD—The electromagnetic field that is produced about an antenna when current and voltage are present on the same antenna [10].

INDUCTION LOSSES—The losses that occur when the electromagnetic field around a conductor cuts through nearby metallic objects and induces a current into that object [10].

INDUCTION MOTOR—A simple, rugged, ac motor with desirable characteristics. The rotor is energized by transformer action (induction) from the stator. Induction motors are used more than any other type [5].

INDUCTIVE COUPLING—Coupling of two coils by means of magnetic lines of force. In transformers, coupling applied through magnetic lines of force between the primary and secondary windings [2].

INDUCTIVE REACTANCE—The opposition to the flow of an alternating current caused by the inductance of a circuit, expressed in ohms. Identified by the symbol X_L [2] [9].

INERTIA—The physical tendency of a body in motion to remain in motion and a body at rest to remain at rest unless acted upon by an outside force (Newton's First Law of Motion) [15].

INFINITE—(1) Extending indefinitely, endless. (2) Boundless, having no limits. (3) An incalculable number [1].

INFRALOW FREQUENCY—The band of frequencies from 300 Hz to 3,000 Hz [19].

INFRASONIC (SUBSONIC)—Sounds below 15 Hz [10].

IN PHASE—Applied to the condition that exists when two waves of the same frequency pass through their maximum and minimum values of like polarity at the same instant [2].

INPUT—The current, voltage, power, or driving force applied to a circuit or device [13].

INPUT END—The end of a two-wire transmission line that is connected to a source [10].

INPUT IMPEDANCE—Impedance presented to the transmitter by the transmission line and its load [10].

INPUT/OUTPUT—Pertaining to either input or output or both, especially in data processors [13].

INSTANTANEOUS AMPLITUDE—The amplitude at any given point along a sine wave at a specific instant in time [12].

INSTANTANEOUS AUTOMATIC GAIN CONTROL (IAGC)—A circuit that can vary the gain of the radar receiver with each input pulse to maintain a nearly constant output peak amplitude [18].

INSTANTANEOUS VALUE—The magnitude at any particular instant when a value is continually varying with respect to time [2].

INSULATION—A material used to prevent the leakage of electricity from a conductor and to provide mechanical spacing or support as protection against accidental contact with the conductor [1] [4].

INSULATION RESISTANCE—The resistance offered by an insulating material to current leakage [4].

INSULATOR—(1) Material of such low conductivity that the flow of current through it can usually be neglected. (2) A device having high electrical resistance; used for supporting or separating conductors so as to prevent undesired flow of current from the conductors to other objects [1].

INTEGRATED CIRCUIT (IC)—(1) A circuit in which many elements are fabricated and interconnected by a single process (into a single chip), as opposed to a "nonintegrated" circuit in which the transistors, diodes, resistors, and other components are fabricated separately and then assembled [7]. (2) Elements inseparably associated and formed on or within a single substrate [14].

INTELLIGENCE—In communications any signal that conveys information (voice, teletypewriter, facsimile) [17].

INTENSITY (OF SOUND)—The measurement of the amplitude of sound energy. Generally synonymous with loudness [10].

INTERACTION SPACE—The region in an electron tube where the electrons interact with an alternating electromagnetic field [11].

INTERCEPT—The point where two lines drawn on a graph cross each other [10].

INTERELECTRODE CAPACITANCE—The capacitance between the electrodes of an electron tube [6] [11].

INTERFERENCE—Any disturbance that produces an undesirable response or degrades a signal [10].

INTERMEDIATE FREQUENCY (IF)—A lower frequency to which an RF echo is converted for ease of amplification [18].

INTERMEDIATE-LEVEL MAINTENANCE (SM&R Code I)—Direct support and technical assistance to user organizations. Tenders and shore-based repair facilities [14].

INTERMEDIATE POWER AMPLIFIER—The amplifier between the oscillator and final power amplifier [12].

INTERPOLES—Small auxiliary poles, placed between main field poles, whose magnetic field opposes the armature field and cancels armature reaction. Interpoles accomplish the same thing as compensating windings [5].

INTERSECTION LAW—In Boolean algebra, the law which states that if one input to an AND gate is already TRUE, then the output will depend upon the state of the other inputs only [13].

INVERSELY—Inverted or reversed in position or relationship [1].

INVERT—To change a physical or logical state to its opposite state [13].

INVERTER—A circuit with one input and one output. Its function is to invert or reverse the input. When the input is high, the output is low, and vice versa. The inverter is sometimes called a NOT circuit, since it produces the reverse of the input [13].

ION—An electrically charged atom or group of atoms. Negative ions have an excess of electrons; positive ions have a deficiency of electrons [1].

IONIZATION—(1) The process of producing ions. (2) The electrically charged particles produced by high-energy radiation, such as light or ultraviolet rays, or by the collision of particles during thermal agitation [6] [10].

IONIZATION POINT—The potential required to ionize the gas of a gas-filled tube. Sometimes called firing potential [6].

IONIZE—To make an atom or molecule of an element lose an electron, as by X-ray bombardment, and thus be converted into a positive ion. The free electron may attach itself to a neutral atom or molecule to form a negative ion [1].

IONOSPHERE—The most important region of the atmosphere extending from 31 miles to 250 miles above sea level. Contains four cloud-like layers that affect radio waves [10].

IONOSPHERIC STORMS—Disturbances in the earth's magnetic field that make communications practical only at lower frequencies [10].

IRIS—A metal plate with an opening through which electromagnetic waves may pass. Used as an impedance-matching device in waveguides [11].

I^2R LOSS—See COPPER LOSSES [11].

ISOLATION—The prevention of unwanted interaction or leakage between components [14].

ISOMETRIC DIAGRAM—A diagram showing the outline of a ship, aircraft, or equipment and the location of equipment and cable runs [4].

ISOTROPIC RADIATION—The radiation of energy equally in all directions [10].

JUNCTION—(1) The connection between two or more conductors. (2) The contact between two dissimilar metals or materials, as in a thermocouple [1].

JUNCTION BOX—A box with a cover that serves the purpose of joining different runs of wire or cable and provides space for the connection and branching of the enclosed conductors [4].

JUNCTION DIODE—A two-terminal device containing a single crystal of semiconducting material that ranges from P-type at one terminal to N-type at the other [7].

JUNCTION TRANSISTOR—A bipolar transistor constructed from interacting PN junctions. The term is used to distinguish junction transistors from other types, such as field-effect and point-contact [7].

KEEP-ALIVE CURRENT—See KEEP-ALIVE VOLTAGE [18].

KEEP-ALIVE VOLTAGE—DC voltage applied to a tr gap electrode to produce a glow discharge that allows the tube to ionize faster when the transmitter fires [18].

KEY-CLICK FILTERS—Filters used in keying systems to prevent key-click interference [12].

KEY CLICKS—Interference in the form of "clicks" or "thumps" caused by the sudden application or removal of power [12].

KEYED-OSCILLATOR TRANSMITTER—A transmitter in which one stage is used to produce the RF pulse [18].

KEYER—(1) A device that changes dc pulses to mark and space modulation for teletypewriter transmissions [17]. (2) A synchronizer [18].

KEYING RELAYS—Relays used in radio transmitters where the ordinary hand key cannot accommodate the plate current without excessive arcing [12].

KILO—A prefix meaning one thousand [1].

KINETIC ENERGY—Energy that a body possesses by virtue of its motion [1].

KIRCHHOFF'S LAWS—(1) The algebraic sum of the current flowing toward any point in a circuit and the current flowing away from it is zero. (2) The algebraic sum of the products of the current and resistance in each of the conductors in any closed path in a network is equal to the algebraic sum of the electromotive forces in the path [1].

KLYSTRON POWER AMPLIFIER—A multicavity microwave electron tube that uses velocity modulation [18].

KNEE OF THE CURVE—The point of maximum curvature of a magnetization curve. (Shaped like the knee of a leg that is bent.) [8]

LACING SHUTTLE—A device upon which lacing may be wound to prevent fouling the tape or cord and to aid the lacing process. (Usually made from brass, aluminum, fiber, or plastic) [4].

LAG—The amount one wave is behind another in time; expressed in electrical degrees [2].

LAMINATED CORE—A core built up from thin sheets of metal insulated from each other and used in transformers [2].

LANDS—Conductors or runs on pcbs [14].

LAP WINDING—An armature winding in which opposite ends of each coil are connected to adjoining segments of the commutator so that the windings overlap [5].

LARGE SCALE INTEGRATION (lsi)—An integrated circuit containing 1,000 to 2,000 logic gates or up to 64,000 bits of memory [14].

LASER—An acronym for light amplification by stimulated emission of radiation [17].

LAW OF MAGNETISM—Like poles repel; unlike poles attract [1].

LC CAPACITOR-INPUT FILTER—This is the most common type of filter. It is used in a power supply where output current is low and load current is relatively constant [7].

LC CHOKE-INPUT FILTER—This filter is used in power supplies where voltage regulation is important and where the output current is relatively high and subject to varying load conditions [7].

LEAD—The opposite of lag. Also a WIRE or CONNECTION [2].

LEAD-ACID CELL—A cell in an ordinary storage battery in which electrodes are grids of lead containing an active material consisting of certain lead oxides that change in composition during charging and discharging. The electrodes or plates are immersed in an electrolyte of diluted sulfuric acid [1].

LEAD INDUCTANCE—The inductance of the lead wires connecting the internal components of an electron tube [11].

LEAD SHEATH—A continuous jacket of lead molded around a single conductor or multiple conductor cable. Generally used to ensure conductors are protected from water or extensive moisture [4].

LEAKAGE CURRENT—The small amount of current that flows through the dielectric between the conductors of a transmission line [10].

LEAKAGE FLUX—Magnetic flux lines produced by the primary winding that do not link the turns of the secondary winding [2].

LEAKAGE RESISTANCE—The electrical resistance that opposes the flow of current through the dielectric of a capacitor. The higher the leakage resistance, the slower the capacitor discharges or leaks across the dielectric [2].

LEAST SIGNIFICANT DIGIT (LSD)—The LSD is the digit whose position within a given number expression has the least weighting power [13].

LEFT-HAND RULE FOR GENERATORS—A rule or procedure used to determine the direction of current flow in a generator [2] [5].

LENZ'S LAW—The current induced in a circuit, caused by its motion in a magnetic field or a change in it's magnetic flux, in such a direction as to exert a mechanical force opposing the motion or to oppose the change in flux [2].

LIGHT-EMITTING DIODE (LED)—A PN-junction diode that emits visible light when it is forward biased. Depending on the material used to make the diode, the light may be red, green, or amber [7].

LIGHTHOUSE TUBE—An electron tube shaped like a lighthouse that is designed to handle large amounts of power at uhf frequencies [6].

LIGHT RAYS—Light waves emitting from a source in straight lines [10].

LIMITER—A device that prevents (limits) a waveform from exceeding a specified value [9].

LINEAR—Having an output that varies in direct proportion to the input [6].

LINEAR IMPEDANCE—An impedance in which a change in current through a device changes in direct proportion to the voltage applied to the device [12].

LINE OF FORCE—A line in an electric or magnetic field that shows the direction of the force [1].

LINE OF SIGHT—Straight line from a radar antenna to a target [18].

LINE-PULSING MODULATOR—Circuit that stores energy and forms pulses in the same circuit element, usually the pulse-forming network (pfn) [18].

LIN-LOG AMPLIFIER—An amplifier in which the response is linear for weak signals and logarithmic for large signals [18].

LIQUID—One of the three states of matter. It has a definite volume but no definite form (water is a liquid) [1].

LIQUID-COOLING SYSTEM—Source of cooling for high-heat producing equipments, such as microwave components, radar repeaters, and transmitters [18].

LISSAJOUS PATTERN—A combined, simultaneous display of the amplitude and phase relationships of two input signals on a CRT [17].

LOAD—(1) A device through which an electric current flows and which changes electrical energy into another form. (2) Power consumed by a device or circuit in performing its function [1] [13].

LOAD END—See OUTPUT END [10].

LOADING—See LUMPED-IMPEDANCE TUNING [10].

LOADING EFFECT—The effect of a voltmeter upon the circuit being measured that results in an inaccurate measurement. Loading effect is minimized by using a voltmeter with an internal resistance many times higher than the resistance of the circuit being measured [3].

LOAD ISOLATOR—A passive attenuator in which the loss in one direction is much greater than that in the opposite direction. One example is a ferrite isolator for waveguides that allows energy to travel in only one direction [11].

LOBE—An area of greater signal strength in the transmission pattern of an antenna [10] [18].

LOCAL ACTION—A continuation of current flow within a battery cell when there is no external load. Caused by impurities in the electrode [1].

LOGARITHMIC RECEIVER—Receiver that uses a linear logarithmic amplifier (lin-log) instead of a normal linear amplifier [18].

LOGIC—The basic principles and applications of truth tables, interconnections of off-on circuit elements, and other factors involved in mathematical computation in automatic data processing systems and other devices [13].

LOGIC CIRCUIT—The primary control information processor in digital equipment; made up of electronic gates and so named because their operation is described by simple equations of a specialized logic algebra [13].

LOGIC DIAGRAM—In computers and data processing equipment, a diagram representing the logical elements and their interconnections without necessarily expressing construction or engineering details [13].

LOGIC ELEMENT—The smallest building blocks that can be represented by operators in an appropriate system of symbolic logic. Typical logic elements are the AND-gate and the flip-flop, which can be represented as operators in a suitable symbolic logic. Also a device that performs the logic function [13].

LOGIC INSTRUCTION—Any instruction that executes a logic operation that is defined in symbolic logic, such as AND, OR, NAND, or NOR [13].

LOGIC OPERATION—A nonarithmetical operation in a computer, such as comparing, selecting, making references, matching, sorting, and merging, where the logical YES or NO quantities are involved [13].

LOGIC SWITCH—A diode matrix (See MATRIX) or other switching arrangement that is capable of directing an input signal to one of several outputs [13].

LOGIC SYMBOL—A symbol used to represent a logic element graphically. Also a symbol used to represent a logic operator [13].

LONGITUDINAL WAVES—Those waves in which the disturbance (back and forth motion) takes place in the direction of propagation. Sometimes called compression waves [10].

LONG-WIRE ANTENNA—An antenna that is a wavelength or more long at its operating frequency [10].

LOOP—A curved conductor that connects the ends of a coaxial cable or other transmission line and projects into a waveguide or resonant cavity for the purpose of injecting or extracting energy [10] [11].

LOOSE COUPLING—Inefficient coupling of energy from one circuit to another that is desirable in some applications. Also called weak coupling [11].

LOWER-FREQUENCY CUTOFF—The lowest frequency a circuit will pass [9].

LOWER SIDEBAND—All difference frequencies below that of the carrier [12].

LOWEST USABLE FREQUENCY—The minimum operating frequency that can be used for communications between two points [10].

LOW FREQUENCY—The band of frequencies from 30 kHz to 300 kHz [17].

LOW-LEVEL MODULATION—Modulation produced in an earlier stage than the final [12].

LOW-NOISE AMPLIFIER—See PREAMPLIFIER [18].

LOW-PASS FILTER—A filter that passes a majority of the low frequencies on to the next circuit and rejects, or attenuates, the higher frequencies. Also called a high-frequency discriminator [9] [12].

LSD—See LEAST SIGNIFICANT DIGIT [13].

LUMPED CONSTANTS—The properties of inductance, capacitance, and resistance in a transmission line [10].

LUMPED IMPEDANCE TUNING—The insertion of an inductor or capacitor in series with an antenna to electrically lengthen or shorten the antenna [10].

MACHINE KEYING—A method of cw keying using punched tape or other mechanical means to key a transmitter [12].

MAGIC T—See BALANCED MIXER [18].

MAGIC-T JUNCTION—A combination of H-type and E-type T-junctions [11].

MAGNET WIRE—Wire coated with an enamel insulation and used in coils, relays, transformers, motor windings, and so forth [4].

MAGNETIC AMPLIFIER—An electromagnetic device that uses one or more saturable reactors to obtain a large power gain. This device is used in servosystems requiring large amounts of power to move heavy loads [8] [15].

MAGNETIC FIELD—(1) The region in which the magnetic forces created by a permanent magnet or by a current-carrying conductor or coil can be detected [1] [2]. (2) The field that is produced when current flows through a conductor or antenna [10] [11].

MAGNETIC INDUCTION—Generating a voltage in a circuit by the creation of relative motion between a magnetic field and the circuit. The relative motion can be the result of physical movement or the rise and fall of a magnetic field created by a changing current [5].

MAGNETIC LINES OF FORCE—Imaginary lines used for convenience to designate the direction in which magnetic forces are acting as a result of magnetomotive force [2].

MAGNETIC MICROPHONE—A microphone in which the sound waves vibrate a moving armature. The armature consists of a coil wound on the armature and located between the pole pieces of a permanent magnet. The armature is mechanically linked to the diaphragm [12].

MAGNETIC POLES—The section of a magnet where the flux lines are concentrated; also where they enter and leave the magnet [1].

MAGNETIC TRIP ELEMENT—A circuit breaker trip element that uses the increasing magnetic attraction of a coil with increased current to open the circuit [3].

MAGNETISM—The property possessed by certain materials by which these materials can exert mechanical force on neighboring masses of magnetic materials and can cause currents to be induced in conducting bodies moving relative to the magnetized bodies [1].

MAGNETRON OSCILLATOR—An electron tube that provides a high power output. Theory of operation is based on interaction of electrons with the crossed electric and magnetic fields in a resonant cavity [18].

MAINTENANCE—Work done to correct, reduce, or counteract wear, failure, and damage to equipment [16].

MAJOR LOBE—The lobe in which the greatest amount of radiation occurs [10].

MAJORITY CARRIERS—The mobile charge carriers (hole or electron) which are predominate in a semiconductor material; for example, electrons in an N-type region [7].

MARCONI ANTENNA—A quarter-wave antenna that is operated with one end grounded and is positioned perpendicular to the earth [10].

MARK—An interval during which a signal is present. Also the presence of an RF signal in cw keying. The key-closed condition (presence of data) in communications systems [12].

MARKING—The state where a circuit is closed and current flows in teletypewriter operation [17].

MASK—A device used to deposit materials on a substrate in the desired pattern [14].

MASTER OSCILLATOR—In a transmitter, the oscillator that establishes the carrier frequency of the output [18].

MASTER OSCILLATOR POWER AMPLIFIER (MOPA)—A transmitter in which the oscillator is isolated from the antenna by a power amplifier [12].

MATRIX—In computers, a logic network in the form of an array of input leads and output leads with logic elements connected at some of their intersections [13].

MATTER—Any physical entity that possesses mass [1].

MAXIMUM USABLE FREQUENCY—Maximum frequency that can be used for communications between two locations for a given time of day and a given angle of incidence [10].

MEASURE (METROLOGY AUTOMATED SYSTEM FOR UNIFORM RECALL AND REPORTING)—The Navy data processing system designed to provide a standardized system for the recall, scheduling, and documenting of test equipment into calibration facilities [16].

MECHANICAL-ROTATION FREQUENCY—The speed in revolutions per minute of armatures in electric motors and engine-driven generators; blade speed in turbines [16].

MECHANICAL SCANNING—The reflector, its feed source, or the entire antenna is moved in a desired pattern [18].

MECHANIZATION—Using electric or electro-mechanical switches to represent logic circuits (AND, OR, NOT, NOR, NAND) [13].

MEDIUM—The vehicle through which a wave travels from one point to the next. Air, water, and wood are examples [10].

MEDIUM ALTITUDE ORBIT—An orbit from 2,000 to 12,000 miles above the earth. The rotation rate of the earth and satellite are quite different, and the satellite moves quickly across the sky [17].

MEDIUM FREQUENCY—The band of frequencies from 300 kHz to 3 MHz [17].

MEGA—A prefix meaning one million; also MEG [1].

MEGGER—Common name for a megohmmeter [3] [16].

MEGOHMMETER—A meter that measures very large values of resistance; usually used to check for insulation breakdown in wires [3].

METALLIC ARMOR—A protective covering for wires or cables. Made as a woven wire braid, metal tape, or interlocking metal cover. Made from steel, copper, bronze, or aluminum [4].

METALLIC, INSULATOR—A shorted quarter-wave section of transmission line [11].

METALLIC RECTIFIER—Also known as a DRY-DISC RECTIFIER. A metal-to-semiconductor, large-area, contact device in which a semiconductor is sandwiched between two metal plates. This asymmetrical construction permits current to flow more readily in one direction than the other [7].

METAL-OXIDE SEMICONDUCTOR FIELD-EFFECT TRANSISTOR—See MOSFET [7].

METER—A device used to measure a specific quantity, such as current, voltage, or frequency [3].

METER MOVEMENT—The part of the meter that moves to indicate some value [3] [16].

METER SHUNT—A resistor placed in parallel with the meter terminals; used to provide increased range capability [16].

MHO—Unit of conductance; the reciprocal of the ohm [1].

MICRO—A prefix meaning one-millionth [1].

MICROCIRCUIT—A small circuit having high equivalent-circuit-element density, which is considered as a single part composed of interconnected elements on or within a single substrate to perform an electronic-circuit function [14].

MICROCIRCUIT MODULE—An assembly of microcircuits or a combination of microcircuits and discrete components that perform one or more distinct functions [14].

MICROELECTRONICS—The solid-state concept of electronics in which compact semiconductor materials are designed to function as an entire circuit or subassembly rather than as circuit components [7] [14].

MICROPHONE—An energy converter that changes sound energy into electrical energy [12].

MICROWAVE REGION—The portion of the electromagnetic spectrum from 1,000 MHz to 100,000 MHz [11].

MIL—The diameter of a conductor equal to 1/1000 (.001) inch [4].

MIL FOOT—A unit of measurement for conductors (diameter of 1 mil, 1 foot in length.) [4].

MILITARY SPECIFICATIONS (MIL-SPEC)—Technical requirements and standards adopted by the Department of Defense that must be met by vendors selling materials to DOD [4].

MILITARY STANDARDS (MILSTD)—Standards of performance for components or equipment that must be met to be acceptable for military systems [14].

MILLI—A prefix meaning one-thousandth [1].

MINIATURE ELECTRONICS—Modules, packages, pcbs, and so forth, composed exclusively of discrete components [14].

MINIMUM DISCERNIBLE SIGNAL (MDS)—The weakest input signal that produces a usable signal at the output of a receiver. The weaker the input signal, the more sensitive the receiver [18].

MINORITY CARRIERS—Either electrons or holes, whichever is the less dominant carrier in a semiconductor device. In P-type semiconductors, electrons are the minority carriers; in N-type semiconductors, the holes are the minority carriers [7].

MINORITY CURRENT—A very small current that passes through the base-to-collector junction when this junction is reverse biased [7].

MINOR LOBE—The lobe in which the radiation intensity is less than that of a major lobe [10].

MIXER—In radar, a circuit that combines the received RF signal with a local-oscillator signal to effectively convert the received signal to a lower IF frequency signal [18].

MODE SHIFTING—In a magnetron, the inadvertent shifting from one mode to another during a pulse [18].

MODE SKIPPING—Operation in which the magnetron fires randomly, rather than firing on each successive pulse as desired [18].

MODIFIED TRANSISTOR OUTLINE (TO)—An IC package resembling a transistor [14].

MODULAR CIRCUITRY—A technique where printed circuit boards are stacked and connected together to form a module [7].

MODULAR PACKAGING—Circuit assemblies or subassemblies packaged to be easily removed for maintenance or repair [14].

MODULATED WAVE—A complex wave consisting of a carrier and a modulating wave that is transmitted through space [12].

MODULATING WAVE—An information wave representing intelligence [12].

MODULATION—The process of impressing intelligence upon a transmission medium, such as radio waves [12].

MODULATION FACTOR (M)—An indication of relative magnitudes of the RF carrier and the modulating signal [12].

MODULATION INDEX—The ratio of frequency deviation to the frequency of the modulating signal [12].

MODULATOR—(1) A device that produces modulation; that is, a device that varies the amplitude, frequency, or phase of an ac signal [11] [12]. (2) A circuit used in servosystems to convert a dc signal to an ac signal. The output ac signal is a sine wave at the frequency of the ac reference voltage. The amplitude of the output is directly related to the amplitude of the dc input. The circuit's function is opposite to that of a DEMODULATOR [15]. (3) In radar, it produces a high-voltage pulse that turns the transmitter on and off [18].

MODULATOR SWITCHING DEVICE—Controls the on (discharge) and off (charge) time of the modulator [18].

MODULE—A circuit or portion of a circuit packaged as a removable unit. A separable unit in a packaging scheme displaying regularity of dimensions [14].

MOISTURE LAPSE—Abnormal variation of moisture content at different altitudes because of high moisture located just above large bodies of water [18].

MONOLITHIC CIRCUIT—A circuit where all elements (resistors, transistors, and so forth) associated with the circuit are fabricated inseparably within a continuous piece of material (called the substrate), usually silicon [7].

MONOLITHIC IC—ICs that are formed completely within a semiconductor substrate. Silicon chips [14].

MONOPULSE (SIMULTANEOUS) LOBING—A radar receiving method using two or more (usually four) partially overlapping lobes. Sum and difference locate the target with aspect to the axis of the antenna [18].

MONOPULSE RADAR—A radar that gets the range, bearing, and elevation position data of a target from a single pulse [18].

MONOPULSE RECEIVER—See MONOPULSE RADAR [18].

MONOSTABLE MULTIVIBRATOR—A multivibrator that has one steady state. A signal (trigger) must be applied to cause change of states [9].

MOSFET—A semiconductor device that contains diffused source and drain regions on either side of a P- or N-channel area. Also contains a gate insulated from the channel area by silicon-oxide. Operates in either the depletion or the enhancement mode [7].

MOST SIGNIFICANT DIGIT (MSD)—The MSD is the digit whose position within a given number expression has the greatest weighting power [13].

MOTOR—A machine that converts electrical energy to mechanical energy. It is activated by ac or dc voltage, depending on the design [5].

MOTOR LOAD—Any device driven by a motor. Typical loads are drills, saws, water pumps, rotating antennas, generators, and so forth. The speed and power capabilities of a motor must be matched to the speed and power capabilities of the motor load [5].

MOTOR REACTION—The force created by generator armature current that tends to oppose the normal rotation of the armature [5].

MOTOR STARTERS—Large resistive devices placed in series with dc motor armatures to prevent the armature from drawing excessive current until armature speed develops counter emf. The resistance is gradually removed from the circuit either automatically or manually as motor speed increases [5].

MOVING-IRON METER MOVEMENT—Same as MOVING-VANE METER MOVEMENT [3].

MOVING TARGET INDICATOR—A device that limits the display of radar information to moving targets [18].

MOVING-VANE METER MOVEMENT—A meter movement that uses the magnetic repulsion of the like poles created in two iron vanes by current through a coil of wire; most commonly used movement for ac meters [3].

MSD—See MOST SIGNIFICANT DIGIT [13].

MTDS—An abbreviation for the marine tactical data system [17].

MU—Symbol for amplification factor [6] [7].

MULTICONDUCTOR—More than one conductor, as in a cable [4].

MULTICOUPLERS—Couplers that patch receivers or transmitters to antennas. They also filter out harmonics and spurious responses and impedance-match the equipment [17].

MULTIELECTRODE TUBE—An electron tube normally classified according to its number of electrodes (the multielectrode tube contains more than three electrodes) [6].

MULTIELEMENT ARRAY—An array that consists of one or more arrays and is classified as to directivity [10].

MULTIELEMENT PARASITIC ARRAY—An array that contains two or more parasitic elements and a driven element [10].

MULTILOOP SERVOSYSTEM—A servosystem that contains more than one servo loop; each loop is designed to perform its own function [15].

MULTIMETER—A single meter combining the functions of an ammeter, a voltmeter, and an ohmmeter [3].

MULTIPATH—The multiple paths a radio wave may follow between transmitter and receiver [10].

MULTIPHASE—See POLYPHASE [5].

MULTIPLICATION FACTOR—The number of times an input frequency is multiplied [12].

MULTIPLEXING—A method for simultaneous transmission of two or more signals over a common carrier wave [17].

MULTISPEED SYNCHRO SYSTEMS—Systems that transmit data at different transmission speeds; for example, dual-speed and tri-speed synchro systems [15].

MULTIUNIT TUBE—An electron tube containing two or more units within the same envelope. The multiunit tube is capable of operating as a single-unit tube, or each unit can operate as a separate tube [6].

MULTIVIBRATOR—A form of relaxation oscillator which comprises two stages that are coupled so that the input of one is derived from the output of the other [9] [13].

MULTIVIBRATOR MODULATOR—An astable multivibrator used to provide frequency modulation. The modulating af voltage is inserted in series with the base return of the multivibrator transistors to produce the frequency modulation [12].

MUTUAL FLUX—The total flux in the core of a transformer that is common to both the primary and secondary windings. The flux links both windings [2].

MUTUAL INDUCTANCE—A circuit property existing when the relative position of two inductors causes the magnetic lines of force from one to link with the turns of the other. The symbol for mutual inductance is M [2].

NAND—A logic function of A and B that is true if either A or B is false [13].

NAND CIRCUIT—A combination of a NOT function and an AND function in a binary circuit that has two or more inputs and one output. The output is logic 0 only if ALL inputs are logic 1; it is logic 1 if ANY input is logic 0 [13].

NATURAL FREQUENCY—See RESONANT FREQUENCY [9].

NATURAL HORIZON—The line-of-sight horizon [10].

NAUTICAL MILE—The length of a minute of arc of a great circle of the earth (6,076 ft) [18].

NAUTICAL RADAR MILE—See RADAR MILE [18].

NEAR SYNCHRONOUS ORBIT—An orbit in which the satellite rotates close to but not exactly at the same speed as the earth [17].

NEGATION—The process of inverting the value of a function or variable [13].

NEGATIVE ALTERNATION—That part of a sine wave that is below the reference level [2] [10] [12].

NEGATIVE CLAMPER—A circuit that clamps the upper extremity of the output waveshape to a dc potential of 0 volts [9].

NEGATIVE ELECTRODE—A terminal or electrode having more electrons than normal. Electrons flow out of the negative terminal of a voltage source [1].

NEGATIVE FEEDBACK—Feedback in which the feedback signal is out of phase with the input signal. Also called DEGENERATIVE FEEDBACK [8].

NEGATIVE LOGIC—The form of logic in which the more positive voltage level represents a logic 0, FALSE, or LOW and the more negative voltage represents a logic 1, TRUE, or HIGH [13].

NEGATIVE-RESISTANCE ELEMENT—A component having an operating region in which an increase in the applied voltage increases the resistance and produces a proportional decrease in current. Examples include tunnel diodes and silicon unijunction transistors [11].

NEGATIVE TEMPERATURE COEFFICIENT—A characteristic of a semiconductor material, such as silver sulfide, in which resistance to electrical current flow decreases as temperature increases [1] [4] [7].

NETWORK—A combination of electrical components. In a parallel circuit it is composed of two or more branches [1].

NEUTRAL—(1) In a normal condition, hence neither positive nor negative. A neutral object has a normal number of electrons (the same number as protons) [1]. (2) The teletypewriter operation where current flow represents a mark and no flow represents a space [17].

NEUTRALIZATION—The process of counteracting or "neutralizing" the effects of interelectrode capacitance [8].

NEWTON'S SECOND LAW OF MOTION—If an unbalanced outside force acts on a body, the resulting acceleration is directly proportional to the magnitude of the force, is in the direction of the force, and is inversely proportional to the mass of the body [15].

NODE—The fixed minimum points of voltage or current on a standing wave or antenna [10].

NOISE—(1) In reference to sound, an unwanted disturbance caused by spurious waves that originate from man-made or natural sources [10]. (2) In radar, erratic or random deflection or intensity of the indicator sweep that tends to mask small echo signals [18].

NOISE FIGURE—The ratio of output noise to input noise in a receiver [18].

NOISE LIMITER—Circuit that clips the peaks of the noise spikes in a receiver [17].

NOISE SILENCER—See NOISE LIMITER [17].

NOISE SUPPRESSOR—See NOISE LIMITER [17].

NO-LOAD CONDITION—The condition that exists when an electrical source or secondary of a transformer is operated without an electrical load [2].

NONDEGENERATIVE PARAMETRIC AMPLIFIER—A parametric amplifier that uses a pump signal frequency that is higher than twice the frequency of the input signal [11].

NONDIRECTIONAL—See OMNIDIRECTIONAL [10].

NONLINEAR DEVICE—A device in which the output does not rise and fall in direct proportion to the input [6] [7] [12].

NONLINEAR IMPEDANCE—An impedance in which the resulting current through the device is not proportional to the applied voltage [12].

NONLUMINOUS BODIES—Objects that either reflect or diffuse light that falls upon them [10].

NONRESONANT LINE—A transmission fine that has no standing waves of current or voltage [10].

NONTRIP-FREE CIRCUIT BREAKER—A circuit breaker that can be held in the ON position during an overcurrent condition [3].

NOR—A logic function of A and B that is true if both A and B are false [13].

NOR GATE—An OR gate that is followed by an inverter to form a binary circuit in which the output is a logic 0 if ANY of the inputs is a logic 1 and the output is a logic 1 only if ALL the inputs are a logic 0 [13].

NORMAL—The imaginary line perpendicular to the point at which the incident wave strikes the reflecting surface. Also called the perpendicular [10].

NOT CIRCUIT—A binary circuit with a single output that is always the opposite of the input. Also called an INVERTER CIRCUIT [13].

NPN—An NPN transistor is formed by introducing a thin region of P-type material between two regions of N-type material [7].

NULL—On a polar-coordinate graph, the area that represents minimum or 0 radiation [10].

NUMBER—(1) A mathematical entity that may indicate quantity or amount of units. (2) Loosely, a numeral. An abstract mathematical symbol for expressing a quantity. In this sense, the manner of representing the number is immaterial. Take 26, for example; this is its decimal form - but it could be expressed as a binary (base 2), octal (base 8), or hexadecimal (base 16) number [13].

NUMBER REPRESENTATION—The representation of numbers by agreed sets of symbols according to agreed rules [13].

NUMBER SYSTEM—Loosely, a number representation system. Any system for the representation of numbers (see POSITIONAL NOTATION) [13].

NUMERAL—(1) A discrete representation of a number. For example, twelve, 12, XII, 1100_2 are four different numerals that represent the same number. (2) A numeric word that represents a number [13].

NUTATING—Moving an antenna feed point in a conical pattern so that the polarization of the beam does not change [18].

OCTAL NUMBER SYSTEM—A number system based on powers of eight. This system is used extensively in computer work [13].

OFF-LINE TEST EQUIPMENT—Equipment that tests and isolates faults in modules or assemblies removed from systems [14].

OHM—The unit of electrical resistance. That value of electrical resistance through which a constant potential difference of 1 volt across the resistance will maintain a current flow of 1 ampere through the resistance [1].

OHMIC VALUE—Resistance in ohms [1].

OHMMETER—A meter used to measure resistance [3] [16].

OHM'S LAW—The current in an electrical circuit is directly proportional to the electromotive force in the circuit. The most common form of the law is $E = IR$, where E is the electromotive force or voltage across the circuit, I is the current flowing in the circuit, and R is the resistance of the circuit [1].

OHMS PER SQUARE—The resistance of any square area of thin film resistive material as measured between two parallel sides [14].

OILCAN TUBE—A type of planar tube, similar to the lighthouse tube, which has cooling fins. The oilcan tube is designed to handle large amounts of power at uhf frequencies [6].

OMNIDIRECTIONAL ANTENNA—An antenna that radiates and receives equally in all directions (nondirectional) [10] [18].

ON-LINE TEST EQUIPMENT—Continuously monitors the performance of electronic systems [14].

OPAQUE—Those substances that do not transmit (pass) any light rays; that is, the light rays are either absorbed or reflected [10].

OPEN CIRCUIT—(1) The condition of an electrical circuit caused by the breaking of continuity of one or more conductors of the circuit; usually an undesired condition. (2) A circuit that does not provide a complete path for the flow of current [1].

OPEN-ENDED LINE—A transmission fine that has a terminating impedance that is infinitely large [10].

OPERATIONAL AMPLIFIER (OP AMP)—An amplifier designed to perform computing or transfer operations and that has the following characteristics: (1) very high gain, (2) very high input impedance, and (3) very low output impedance [8].

OPTICAL COUPLER—A coupler composed of an LED and a photodiode and contained in a light-conducting medium. Suitable for frequencies in the low-megahertz range [7].

OPTIMUM WORKING FREQUENCY—The most practical operating frequency that can be used with the least amount of problems and is roughly 85 percent of the maximum usable frequency [10].

OPTOELECTRONIC DEVICES—Devices that either produce or use light in their operation [7].

ORDER-WIRE CIRCUIT—A circuit between operators used for operations control and coordination [17].

ORGANIZATIONAL-LEVEL MAINTENANCE (SM & R CODE O)—Responsibility of the user organization [14].

OR GATE—A gate that performs the logic OR function. It produces an output 1 whenever any or all of its inputs is/are 1 [13].

ORIGIN—The point on a graph where the vertical and horizontal axes cross each other [10].

OSCILLATOR—An oscillator is a nonrotating device that produces alternating current. The frequency is determined by the characteristics of the device [9].

OUT-OF-CIRCUIT METER—A meter that is not permanently installed in a circuit. Usually portable and self-contained, these meters are used to check the operation of a circuit or to isolate troubles within a circuit [3].

OUTPUT END—The end of a transmission line that is opposite the source; receiving end [10].

OUTPUT IMPEDANCE—The impedance that is presented to the load by the transmission line and its source [10].

OVERDRIVEN—When the input signal amplitude is increased to the point that the transistor goes into saturation and cutoff [7].

OVERMODULATION—A condition that exists when the peaks of the modulating signal are limited [12].

PACKAGING LEVELS—A system developed to assist maintenance personnel in determining the repairability of components, printed circuit boards, modules, and so forth [14].

PAGE PRINTER—A high-speed printer that prints teletypewriter characters one at a time in a full-page format [17].

PARABOLIC REFLECTOR—An antenna reflector in the shape of a parabola. It converts spherical wavefronts from the radiating element into plane wavefronts [18].

PARALLAX ERROR—The error in meter readings that results when you look at a meter from some position other than directly in line with the pointer and meter face. A mirror mounted on the meter face aids in eliminating parallax error [3].

PARALLEL CIRCUIT—Two or more electrical devices connected to the same pair of terminals so separate currents flow through each; electrons have more than one path to travel from the negative to the positive terminal [1].

PARALLEL-CONNECTED DUPLEXER—Configuration in which the tr spark gap is connected across the two legs of the transmission line one-quarter wavelength from the T-junction [18].

PARALLEL LIMITER—A resistor and diode, connected in series with the input signal, in which the output is taken across (parallel to) the diode [9].

PARALLEL-NEGATIVE LIMITER—A resistor and diode, connected in series with the input signal, in which the output is taken across the diode and the negative alternation is eliminated [9]

PARALLEL-POSITIVE LIMITER—A resistor and diode connected in series with the input signal, in which the output is taken across the diode and the positive alternation is eliminated [9].

PARALLEL-RESONANT CIRCUIT—A resonant circuit in which the source voltage is connected across a parallel circuit (formed by a capacitor and an inductor) to furnish a high impedance to the frequency at which the circuit is resonant. Often referred to as a tank circuit [9] [10].

PARALLEL-WIRE—A type of transmission line consisting of two parallel wires [10].

PARASITIC ARRAY—An antenna array containing one or more elements not connected to the transmission line [10] [18].

PARASITIC ELEMENT—The passive element of an antenna array that is connected to neither the transmission line nor the driven element [10].

PART—A part is one component or two or more components joined together. It is not normally subject to disassembly without destruction [17].

PASSIVE SATELLITE—A satellite that reflects radio signals back to earth [17].

PATCH PANEL—A panel used to tie a receiver or transmitter to its associated equipment [17].

PEAK AMPLITUDE—The maximum value above or below the reference line [12].

PEAK CURRENT—The maximum current that flows during a complete cycle [6].

PEAK DETECTION—Detection that uses the amplitude of pam or the duration of pdm to charge a holding capacitor and restore the original waveform [12].

PEAKING COIL—An inductor used in an amplifier to provide high-frequency compensation, which extends the high-frequency response of the amplifier [8].

PEAK POWER—The maximum value of the transmitted pulse [12].

PEAK-REVERSE VOLTAGE—The peak ac voltage that a rectifier tube will withstand in the reverse direction [6].

PEAK-TO-PEAK—The measure of absolute magnitude of an ac waveform, measured from the greatest positive alternation to the greatest negative alternation [2].

PEAK VALUE—The maximum instantaneous value of a varying current, voltage, or power. It is equal to 1.414 times the effective value of a sine wave [2].

PEAK VOLTAGE—The maximum value present in a varying or alternating voltage. This value may be positive or negative [6].

PENTAVALENT IMPURITY—A type of impurity that contains five valence electrons and donates one electron to the doped material. Also called DONOR IMPURITY [7].

PENTODE TUBE—A five-electrode electron tube containing a plate, a cathode, a control grid, and two grids [6].

PERCENT OF MODULATION—The degree of modulation defined in terms of the maximum permissible amount of modulation [12].

PERFORATOR—A device that stores a teletypewriter message on a paper tape [17].

PERIGEE—The point in the orbit of a satellite closest to the earth [17].

PERIOD TIME—The time required to complete one cycle of a waveform [2] [10] [12].

PERIODIC WAVE—A waveform that undergoes a pattern of changes, returns to its original pattern, and then repeats the same pattern of changes. Examples are square waves, rectangular waves, and sawtooth waves [9].

PERMANENT MAGNET SPEAKER—A speaker with a permanent magnet mounted on soft iron pole pieces [17].

PERMEABILITY—The measure of the ability of a material to act as a path for magnetic lines of force [1] [8].

PERSISTENCE—The length of time a phosphor dot glows on a CRT before disappearing [6] [18].

PHANTASTRON—A variable-length sawtooth generator often used to produce a sweep on an A-scope [18].

PHASE—The angular relationship between two alternating currents or voltages when the voltage or current is plotted as a function of time. When the two are in phase, the angle is zero; both reach their peak simultaneously. When out of phase, one will lead or lag the other; that is, at the instant when one is at its peak, the other will not be at peak value and (depending on the phase angle) may differ in polarity as well as magnitude [2].

PHASE ANGLE—The number of electrical degrees of lead or lag between the voltage and current waveforms in an ac circuit [2] [12].

PHASE MODULATION (pm)—Angle modulation in which the phase of the carrier is controlled by the modulating waveform. The amplitude of the modulating wave determines the amount of phase shift, and the frequency of the modulation determines how often the phase shifts [12].

PHASE-SHIFT DISCRIMINATOR—See FOSTER-SEELEY DISCRIMINATOR [12].

PHASE SHIFTER—A device used to change the phase relationship between two ac signals [11].

PHASE-SHIFT KEYING—Similar to ON-OFF cw keying in AM systems and frequency-shift keying in FM systems. Each time a mark is received, the phase is reversed. No phase reversal takes place when a space is received [12].

PHASE SPLITTER—A device that provides two output signals from a single input signal. The two output signals differ from each other in phase (usually by 180 degrees) [8].

PHOSPHOR—The material used to convert the energy of electrons into visible light [6].

PHOTOCELL—A light-controlled variable resistor which has a light-to-dark resistance ratio of 1:1000. Used in various types of control and timing circuits [7].

PHOTODIODE—A light-controlled PN junction. Current flow increases when the PN junction is exposed to an external light source [7].

PHOTOELECTRIC VOLTAGE—A voltage produced by light [1].

PHOTOETCHING—Chemical process of removing unwanted material in producing printed circuit boards [14].

PHOTOTRANSISTOR—An optoelectronic device that conducts current when exposed to light. Produces more current and is much more sensitive to light than the photodiode [7].

PHOTOVOLTAIC CELL (SOLAR CELL)—A device that acts much like a battery when exposed to light and converts light energy into electrical energy [7].

PICO—A prefix adopted by the National Bureau of Standards meaning 10^{-12} [1].

PICTORIAL DIAGRAM—A diagram showing actual pictorial sketches of the various parts of an equipment and the electrical connections between the parts [4].

PIEZOELECTRIC EFFECT—The effect of producing a voltage by placing a stress, either by compression, expansion, or twisting, on a crystal and, conversely, producing a stress in a crystal by applying a voltage to it [1].

PIP (BLIP)—On a CRT display, a spot of light or a baseline irregularity representing the radar echo [18].

PITCH—A term used to describe the frequency of a sound heard by the human ear [10].

PLANAR TUBE—An electron tube, constructed with parallel electrodes and a ceramic envelope, that is used at uhf frequencies. It is commonly referred to as a lighthouse tube [6].

PLANE OF POLARIZATION—The plane (vertical or horizontal), with respect to the earth, in which the E field propagates [10].

PLANE WAVEFRONTS—Waves of energy that are flat, parallel planes and are perpendicular to the direction of propagation [18].

PLANNED-POSITION INDICATOR—A radar display in which range is indicated by the distance of a bright spot or pip from the center of the screen and the bearing is indicated by the radial angle of the spot [18].

PLATE—(1) One of the electrodes in a storage battery [1]. (2) One of the electrodes in a capacitor [2]. (3) The principal electrode to which the electron stream is attracted in an electron tube [6].

PLATE DISSIPATION—The amount of power lost as heat in the plate of a vacuum tube [6].

PLATE KEYING—A keying system in which the plate supply is interrupted [12].

PLATE MODULATOR—An electron-tube modulator in which the modulating voltage is applied to the plate circuit of the tube [12].

PLATE RESISTANCE—The plate voltage change divided by the resultant plate current change in a vacuum tube, all other conditions being fixed [6].

POINT BENDER—A tool used to adjust the contact spacing on a relay [3].

POINT-CONTACT DIODE—A diode in which the end of a fine wire is pressed against a semiconductor. Used as a detector or mixer over the microwave region [7].

POINT OF ZERO DISPLACEMENT—See REFERENCE LINE [10].

POINT-TO-POINT WIRING—Individual wires run from terminal to terminal to complete a circuit [14].

POLAR—The teletypewriter operation in which current flow of one polarity represents a mark, and current of the opposite polarity represents a space [17].

POLAR-COORDINATE GRAPH—A graph with a pair of axes. One consists of a series of circles with a common center, and the other consists of a rotating radius extending from the center of the concentric circles [10].

POLARITY—(1) The condition in an electrical circuit by which the direction of the flow of current can be determined. Usually applied to batteries and other direct voltage sources. (2) Two opposite charges, one positive and one negative. (3) A quality of having two opposite magnetic poles, one north and the other south [1] [13].

POLARIZATION—(1) The effect of hydrogen surrounding the anode of a cell, which increases the internal resistance of the cell [1]. (2) The magnetic orientation of molecules in a magnetizable material in a magnetic field, whereby tiny internal magnets tend to line up in the field [2].

POLAR ORBIT—An orbit that has an angle of inclination of or near 90 degrees [17].

POLE—(1) The number of points at which current can enter a switch; for example, single pole, double pole, and three pole [3]. (2) The sections of a field magnet where the flux lines are concentrated; also where they enter and leave the magnet [5].

POLE PIECE—(1) A piece of ferromagnetic material used to control the distribution of magnetic lines of force; that is, it concentrates the lines of force in a particular place or evenly distributes the lines of force over a wide area [3]. (2) The shaped magnetic material upon which the stator windings of motors and generators are mounted or wound [5].

POLYMER FUME FEVER—A flu-like condition caused by a person breathing the vapors of fluorocarbons when they are heated. Sometimes called FOUNDRYMAN'S FEVER [4].

POLYPHASE—A term that describes systems or units of a system that are activated by or which generate separate out-of-phase voltages. Typical polyphase systems are 2-phase and 3-phase; their voltages are 90- and 120-degrees out of phase, respectively. This term means the same as MULTIPHASE [5].

POSITIONAL NOTATION—A numbering system in which a number is represented by means of a stated set of symbols or digits, such that the value contributed by each symbol or digit depends upon its position as well as upon its value [13].

POSITIONAL WEIGHTING—The value given a digit based on the digit's position within a given number [13].

POSITION SENSOR—A component in a servosystem that measures position and converts the measurement into a form convenient for transmission as a feedback signal [15].

POSITION SERVOSYSTEM—A servosystem whose end function is to control the position of the load it is driving [15].

POSITIVE ALTERNATION—The part of a sine wave that is above the reference line [2] [10] [12].

POSITIVE CLAMPER—A circuit that clamps the lower extremity of the output waveshape to a dc potential of 0 volts [9].

POSITIVE FEEDBACK—Feedback in which the feedback signal is in phase with the input signal. Also called REGENERATIVE FEEDBACK [8].

POSITIVE LOGIC—The form of logic in which the more positive logic level represents 1 and the more negative level represents 0 [13].

POSITIVE TEMPERATURE COEFFICIENT—The characteristic of a conductor in which the resistance increases as temperature increases [7].

POTENTIAL ENERGY—Energy caused by the position of one body with respect to another body or to the relative parts of the same body [1].

POTENTIOMETER—A variable resistor, used as a position sensor in servosystems, having a terminal connected to each end of a resistive element and a third terminal connected to a wiper contact. The output is a voltage that is variable depending upon the position of the wiper contact. The potentiometer is commonly referred to as a variable voltage divider. It, in effect, converts mechanical information into an electrical signal [1] [15].

POWER—The rate of doing work or the rate of expending energy. The unit of electrical power is the watt [1].

POWER AMPLIFIER—An amplifier in which the output-signal power is greater than the input-signal power [8].

POWER-AMPLIFIER (CHAIN) TRANSMITTER—Transmitter that uses a series of power amplifiers to create a high level of power [18].

POWER FACTOR—The ratio of the actual power of an alternating or pulsating current, as measured by a wattmeter, to the apparent power, as indicated by ammeter and voltmeter readings. The power factor of an inductor, capacitor, or insulator is an expression of their losses [2] [16].

POWER GAIN—In an antenna, the ratio of its radiated power to that of a reference [11] [18].

POWER LOSS—(1) The electrical power, supplied to a circuit, that does no work and is usually dissipated as heat [2] [4]. (2) The heat loss in a conductor as current flows through it [10].

POWER PENTODE—A special purpose tube used to provide high-current gain or power amplification. Each grid wire is directly in line with the one before and after it, a fact which allows more electrons to reach the plate [6].

POWER RATIO—See POWER GAIN [11].

POWER STANDING-WAVE RATIO (PSWR)—The ratio of the square of the maximum and minimum voltages of a transmission line [10].

POWER SUPPLY—A unit that supplies electrical power to another unit. It changes ac to dc and maintains a constant voltage output within limits [6] [7].

PREAMPLIFIER (PREAMP)—An amplifier that raises the output of a low-level source for further processing without appreciable degradation of the signal-to-noise ratio [18].

PRECESSION—The rotation of the spin axis of a gyro in response to an applied force. The direction of precession is always perpendicular to the direction of applied force [15].

PRECESSION VECTOR—In a gyro, a vector representing the angular change of the spin axis when torque is applied. The precession vector represents the axis about which precession occurs [15].

PRESTANDARD NAVY SYNCHROS—Synchros that are designed to meet Navy, rather than servicewide, specifications [15].

PREVENTIVE MAINTENANCE—Visual, mechanical, electrical, and electronic checks that are made to determine whether or not equipment is functioning properly [16].

PRIMARIES (OF LIGHT)—The three primary colors of light from which all other colors can be derived. The colors are red, green, and blue [10].

PRIMARY CELL—An electrochemical cell in which the chemical action eats away one of the electrodes, usually the negative electrode [1].

PRIMARY LOOP—In a cooling system, the primary source of cooling for the distilled water [18].

PRIMARY WINDING—The winding of a transformer connected to the electrical source [2].

PRIME MOVER—The source of the turning force applied to the rotor of a generator. This may be an electric motor, a gasoline engine, a steam turbine, and so forth [5].

PRINTED CIRCUIT BOARD—A flat, insulating surface upon which printed wiring and miniaturized components are connected in a predetermined design and attached to a common base [7] [14].

PRISM—A triangular-shaped glass that refracts and disperses light waves into component wavelengths [10].

PROBE COUPLER—A resonant conductor placed in a waveguide or cavity to insert or extract energy [18].

PROGRAMMED TRACKING—The method that uses known satellite orbital parameters to generate antenna pointing angles [17].

PROPAGATION—Waves traveling through a medium [10].

PULSE—Signal characterized by a steep rise from and decay toward an initial level [9] [12].

PULSE-AMPLITUDE MODULATION (PAM)—Pulse modulation in which the amplitude of the pulses is varied by the modulating signal [12].

PULSE-CODE MODULATION (PCM)—A modulation system in which the standard values of a quantized wave are indicated by a series of coded pulses [12].

PULSE DURATION (PD)—The period of time during which a pulse is present [12].

PULSE-DURATION MODULATION (PDM)—Pulse modulation in which the time duration of the pulses is changed by the modulating signal [12].

PULSE-FORMING NETWORK (PFN)—An lc network that alternately stores and releases energy in an approximately rectangular wave [12] [18].

PULSE-FREQUENCY MODULATION (PFM)—Pulse modulation in which the modulating voltage varies the repetition rate of a pulse train [12].

PULSE MODULATION—A form of modulation in which one of the characteristics of a pulse train is varied [12].

PULSE OSCILLATOR—A sine-wave oscillator that is turned on and off at specific times. Also known as a ringing oscillator [9].

PULSE-POSITION MODULATION (PPM)—Pulse modulation in which the position of the pulses is varied by the modulating voltage [12].

PULSE-REPETITION FREQUENCY (PRF)—The rate, in pulses per second, at which the pulses occur [9] [12] [18].

PULSE-REPETITION RATE (PRR)—Same as PULSE-REPETITION FREQUENCY (PRF) [9] [12] [18].

PULSE-REPETITION TIME (PRT)—Interval between the start of one pulse and the start of the next pulse; reciprocal of pulse-repetition frequency [18].

PULSE-TIME MODULATION (PTM)—Pulse modulation that varies one of the time characteristics of a pulse train (pwm, pdm, ppm, or pfm) [12].

PULSE WIDTH—Duration of time between the leading and trailing edges of a pulse [12] [18].

PULSE-WIDTH MODULATION (PWM)—Pulse modulation in which the duration of the pulses is varied by the modulating voltage [12].

PULSING—Allowing oscillations to occur for a specific period of time only during selected intervals [12].

PUMP—Electrical source of the energy required to vary the capacitance of a parametric amplifier [11].

PUSH-PULL AMPLIFIER—An amplifier that uses two transistors (or electron tubes) whose output signals are in phase opposition [8].

Q—(1) Figure of merit of efficiency of a circuit or coil. (2) Ratio of inductive reactance to resistance in servos. (3) Relationship between stored energy (capacitance) and rate of dissipation in certain types of electric elements, structures, or materials [2] [9].

QUALITY (OF SOUND)—The factor that distinguishes tones of pitch and loudness [10].

QUANTIZED WAVE—A wave created by the arbitrary division of the entire range of amplitude (or frequency, or phase) values of an analog wave into a series of standard values. Each sample takes the standard value nearest its actual value when modulated [12].

QUANTIZING NOISE—A distortion introduced by quantizing the signal [12].

QUANTUM-MECHANICAL TUNNELING—The action of an electron crossing a PN junction because of tunnel effect [7].

QUARTER-WAVE ANTENNA—Same as the MARCONI ANTENNA [10].

QUIESCENCE—(1) The state of an amplifier with no signal applied. (2) The operating conditions that exist in a circuit when no input signal is applied to the circuit [6] [71 [13].

QUIESCENT STATE—The period during which a transistor, tube, or other circuit element is not performing an active function in the circuit [9] [13].

RADAR—An acronym for **RA**dio **D**etecting **A**nd **R**anging [18].

RADAR ALTIMETER—Airborne radar that measures the distance of the aircraft above the ground [18].

RADAR BEAM—The space in front of a radar antenna where a target can be effectively detected or tracked. Defined by areas that contain half or more of the maximum power transmitted [18].

RADAR DETECTOR—A detector that, in its simplest form, only needs to be capable of producing an output when RF energy (reflected from a target) is present at its input [12].

RADAR DISTRIBUTION SWITCHBOARD—An electrical switching panel used to connect inputs from any of several radars to repeaters (indicators) [18].

RADAR MILE—Time interval (12.36 microseconds) for RF energy to travel out from a radar to a target and back to the radar; radar nautical mile [18].

RADAR TEST SET—A combination of several test circuits and equipment used to test various characteristics of a radar [18].

RADIATION FIELD—The electromagnetic field that radiates from an antenna and travels through space [10].

RADIATION LOSSES—The losses that occur when magnetic lines of force about a conductor are projected into space as radiation and are not returned to the conductor as the cycle alternates [10].

RADIATION PATTERN—A plot of the radiated energy from an antenna [10].

RADIATION RESISTANCE—The resistance that if inserted in place of the antenna would consume the same amount of power as that radiated by the antenna [10].

RADIO COMMUNICATIONS—The term describing teletypewriter, voice, telegraphic, and facsimile communications. [17].

RADIO FREQUENCY (RF)—(1) Any frequency of electromagnetic energy capable of propagation into space [2]. (2) The frequencies that fall between 3 kilohertz and 300 gigahertz used for radio communications [10].

RADIO FREQUENCY CARRIER SHIFT—The system that uses a keyer to shift a radio frequency signal above or below an assigned frequency. These shifts correspond to marks and spaces [17].

RADIO HORIZON—The boundary beyond the natural horizon in which radio waves cannot be propagated over the earth's surface [10].

RADIO SET CONTROL UNIT—Equipment used to remotely control certain transmitter and receiver functions [17].

RADIO WAVES—(1) A form of radiant energy that can neither be seen nor felt. (2) An electromagnetic wave that is generated by a transmitter [10].

RADIX—Also called the base. The number of distinct symbols used in a number system. For example, since the decimal number system uses ten symbols (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), the radix is 10. In the binary number system, the radix is 2 because it uses only two symbols (0, 1) [13].

RADIX POINT—Also called BINARY POINT, OCTAL POINT, DECIMAL POINT, and so forth, depending on the number system [13].

RANGE—The length of a straight line between a radar set and a target [11] [18].

RANGE-GATE—A movable gate used to select radar echoes from a very short range interval [18].

RANGE-HEIGHT INDICATOR—A radar display on which slant range is shown along the X axis and height along the Y axis [18].

RANGE MARKER—A movable vertical pulse on an A-scope or a ring on a PPI scope used to measure the range of an echo or to calibrate the range scale [18].

RANGE RESOLUTION—Ability of a radar to distinguish between targets that are close together [18].

RANGES—The several upper limits a meter will measure as selectable by a switch or by jacks; for example, a voltmeter may have ranges of 1 volt, 2.5 volts, 10 volts, 25 volts, and 100 volts [3].

RANGE STEP—On an A-scope sweep, a vertical displacement used to measure the range of an echo [18].

RAREFIED WAVE—A longitudinal wave that has been expanded or rarefied (made less dense) as it moves away from the source [10].

RATE GYRO—A gyro used to detect and measure angular rates of change [15].

RATIO—The value obtained when one number is divided by another. This value indicates the relative proportions of the two numbers [2].

RATIO DETECTOR—A detector that uses a double-tuned transformer to convert the instantaneous frequency variations of the FM input signal to instantaneous amplitude variations [12].

RATIO OF TRANSMITTED POWERS—The power ratio (FSK versus AM) that expresses the overall improvement of FSK transmission when compared to AM under rapid-fading and high-noise conditions [12].

RC CONSTANT—Time constant of a resistor-capacitor circuit; equal in seconds to the resistance value in ohms multiplied by the capacitance value in farads [2] [9].

RC DIFFERENTIATOR—An RC circuit in which the output is taken from the resistor [9].

RC FILTER—A filter used in applications where load current is low and constant, and voltage regulation is not necessary [7].

RC INTEGRATOR—An RC circuit in which the output is taken from the capacitor [9].

RC NETWORK—A circuit containing resistance and capacitance arranged in a particular manner to perform a specific function [9].

RC OSCILLATOR—An oscillator in which the frequency is determined by resistive and capacitive elements [9].

REACTANCE—The opposition offered to the flow of an alternating current by the inductance, capacitance, or both, in any circuit [2].

REACTANCE AMPLIFIER—A low-noise amplifier that uses a nonlinear variable reactance as the active element instead of a variable resistance. Also called a parametric amplifier [11].

REACTANCE TUBE—A tube connected in parallel with the tank circuit of an oscillator. Provides a signal that will either lag or lead the signal produced by the tank [12].

REACTANCE-TUBE MODULATOR—An FM modulator that uses a reactance tube in parallel with the oscillator tank circuit [12].

RECEIVER—(1) The object that responds to the wave or disturbance. Same as DETECTOR [10]. (2) Equipment that converts electromagnetic energy into a visible or an audible form [17]. (3) In radar, a unit that converts RF echoes to video and/or audio signals [18].

RECEIVER SENSITIVITY—(1) The degree to which a receiver can usefully detect a weak signal. (2) The lower limit of useful signal input to the receiver [18].

RECEIVER TRANSFER SWITCHBOARD—Equipment used to transfer receiver audio outputs to remote control station audio circuits [17].

RECEIVING ANTENNA—The device used to pick up the RF signal from space [10].

RECEIVING END—See OUTPUT END [10].

RECEPTION—The instant when an electromagnetic wave passes through a receiver antenna and induces a voltage in that antenna [17].

RECIPROCAL (OF A QUANTITY)—The value obtained by dividing the number 1 by that quantity [1].

RECIPROCITY—The property of interchangeability of the same antenna for transmitting and receiving [10] [11].

RECOVERY TIME—In a radar, the time interval between the end of the transmitted pulse and the time when echo signals are no longer attenuated by the tr gap [18].

RECTIFIER—A device used to convert ac to pulsating dc [3] [6] [7].

RECTANGULAR-COORDINATE GRAPH—A graph in which straight-line axes (horizontal and vertical) are perpendicular [10].

RED—The reference color of equipment that passes classified information. It normally refers to patch panels [17].

REFERENCE LINE—The position of zero displacement in a wave [10].

REFERENCE POINT—A point in a circuit to which all other points in the circuit are compared [1].

REFLECTED WAVE—(1) The wave that reflects back from a medium. (2) The wave moving back to the source from the termination of a transmission line after reflection has occurred [10].

REFLECTING OBJECT—In radar a air or surface contact that provides an echo [18].

REFLECTION WAVES—Waves that are neither transmitted nor absorbed, but are reflected from the surface of the medium they encounter [10].

REFLECTOR—The parasitic element of an array that causes maximum energy radiation in a direction toward the driven element [10].

REFLEX KLYSTRON—A klystron with a reflector (repeller) electrode in the place of a second resonant cavity used to redirect the velocity-modulated electrons back through the cavity that produced the modulation [11]. (2) A microwave oscillator that is tuned by changing the repealer voltage [18].

REFRACTION—The changing of direction of a wave as it leaves one medium and enters another medium of a different density [10] [18].

REFRACTIVE INDEX—In a wave-transmission medium, the ratio between the phase velocity in free space and in the medium [11] [18].

REGENERATION—See FEEDBACK [18].

REGENERATIVE DETECTOR—A detector circuit that produces its own oscillations, heterodynes them with an incoming signal, and deflects them [12].

REGENERATIVE FEEDBACK—The process by which a portion of the output signal of an amplifying device is fed back in phase to reinforce the input. Also called POSITIVE FEEDBACK [8] [9].

REGULATOR—The section in a basic power supply that maintains the output of the power supply at a constant level in spite of large changes in load current or input line voltage [6] [7].

RELATIVE BEARING—Bearing of a target measured in a clockwise direction from "dead ahead" of a ship or plane [18].

RELAY—An electromagnetic device with one or more sets of contacts that change position by the magnetic attraction of a coil to an armature [3].

RELUCTANCE—A measure of the opposition that a material offers to magnetic lines of force [1].

REMOTE-CUTOFF TUBE—An electron tube in which the control grid wires are farther apart at the centers than at the ends. This arrangement allows the tube to amplify large signals without being driven into cutoff. This tube is also called a VARIABLE-MU TUBE [6].

REPEATER—(1) Another name for an active satellite [17]. (2) Also, a common name for remote radar indicators.

REPELLER—Sometimes called a REFLECTOR. An electrode in a reflex klystron with the primary purpose of reversing the direction of the electron beam [11].

REPERFORATOR—Equipment that converts the incoming TTY signal and stores it on paper tape [17].

REPRODUCTION—The process of converting electrical signals to sound waves. This sound is speech, music, and so on [17].

REPULSION—The mechanical force tending to separate bodies having like electrical charges or like magnetic polarity [1].

RERADIATION—The reception and retransmission of radio waves that is caused by turbulence in the troposphere [10].

RESIDUAL MAGNETISM—Magnetism remaining in a substance after removal of the magnetizing force [1].

RESISTANCE—(1) The opposition a device or material offers to the flow of current. The effect of resistance is to raise the temperature of the material or device carrying the current. (2) A circuit element designed to offer a predetermined resistance to current flow. A resistance of 1 ohm will allow a current of 1 ampere to flow through it when a potential of 1 volt is applied. [1].

RESISTIVITY—See SPECIFIC RESISTANCE. The reciprocal of conductivity [4].

RESISTOR—The electrical component that offers resistance to the flow of current. It may be a coil of fine wire or a composition rod [1].

RESOLVER—A rotary, electromechanical device used to perform trigonometric computations by varying the magnetic couplings between its primary and secondary windings. It is generally used in circuits that solve vector problems, such as analog computers and conversion equipment. The resolver solves three different type problems: (1) Resolution - separating a vector into two mutually perpendicular components; (2) Composition - combining two components of a vector to produce a vector sum; and (3) Combination - the process of resolution and composition taking place simultaneously [15].

RESONANCE—The condition in a circuit containing inductance and capacitance in which the inductive reactance is equal and opposite to the capacitive reactance. This condition occurs at only one frequency and the circuit in that condition is said to be in resonance [2] [9] [10].

RESONANCE CHAMBER—See ECHO BOX [18].

RESONANT CIRCUIT—A circuit that contains both inductance and capacitance and is resonant at one frequency ($X_L = X_C$) [9].

RESONANT FREQUENCY—That frequency in a given resonant circuit at which the inductive and capacitive reactance values are equal and cancel each other [9].

RESONANT LINE—A transmission line that has standing waves of current and voltage [10].

REST FREQUENCY—The carrier frequency during the constant-amplitude portions of a phase modulation signal [12].

REST POSITION—See REFERENCE LINE [10].

REST TIME (RT)—The time when there is no pulse; nonpulse time [12].

RESULTANT MAGNETIC FIELD—The magnetic field produced in a synchro by the combined effects of the three stator magnetic fields [15].

RETENTIVITY—The ability of a material to retain its magnetism [1].

RETURN—The RF signal reflected back from a radar target; echo [18].

REVERBERATION—The multiple reflections of sound waves [10].

REVERSE AGC—The type of AGC that causes an amplifier to be driven toward cut-off [17].

REVERSE BIAS—An external voltage applied to a diode or semiconductor junction to reduce the flow of electrons across the junction. Also called BACK BIAS [7] [13].

RF RADIATION HAZARD—A health hazard caused by exposure to electromagnetic radiation or high-energy particles (ions). Abbreviated RADHAZ [18].

RF (RADIO FREQUENCY) AMPLIFIER—An amplifier designed to amplify signals with frequencies between 10 kilohertz (10 kHz) and 100,000 megahertz (100,000 MHz) [8].

RF (RADIO FREQUENCY) TRANSFORMER—A transformer specially designed for use with RF (radio frequencies). An RF transformer is wound onto a tube of nonmagnetic material and has a core of either powdered iron or air [8].

RGK—The symbol used to express the resistance between the grid and the cathode of an electron tube [6].

RHEOSTAT—A variable resistor used for the purpose of adjusting the current in a circuit [1] [4].

RHO—Greek letter "rho" (ρ). Used in the field of electricity and electronics to represent the specific resistance of a substance [4].

RHOMBIC ANTENNA—A diamond-shaped antenna used widely for long-distance, high-frequency transmission and reception [10].

RIGID COAXIAL LINE—A coaxial line consisting of a central insulated wire (inner conductor) mounted inside of a tubular outer conductor [10].

RIGIDITY—The tendency of the spin axis of a gyro wheel to remain in a fixed direction in space if no force is applied to it [15].

RINGING—RF oscillations caused by shock excitation of a resonant circuit or cavity [18].

RING TIME—In radar, the time during which the output of an echo box remains above a specified level [18].

RIPPLE FREQUENCY—The frequency of the ripple current. In a full-wave rectifier it is twice the input-line frequency [6].

RIPPLE VOLTAGE—The alternating component of unidirectional voltage. (This component is small compared to the direct component.) [6]

RLC CIRCUIT—An electrical circuit that has the properties of resistance, inductance, and capacitance [2].

RL DIFFERENTIATOR—An RL circuit in which the output is taken from the inductor [9].

RL INTEGRATOR—An RL circuit in which the output is taken from the resistor [9].

RMS—Abbreviation of root mean square [2].

ROOT MEAN SQUARE (RMS)—The equivalent heating value of an alternating current or voltage, as compared to a direct current or voltage. It is 0.707 times the peak value of a sine wave [2].

ROTARY CAP—A spark gap, similar to a mechanically driven rotary switch, used to discharge a pulse-forming network [12].

ROTARY SWITCH—A multicontact switch with contacts arranged in a circular or semicircular manner [3].

ROTATING FIELD—The magnetic field in a multiphase ac motor that is the result of field windings being energized by out-of-phase currents. In effect, the magnetic field is made to rotate electrically rather than mechanically [5].

ROTATING JOINT—A joint that permits one section of a transmission line or waveguide to rotate continuously with respect to another while passing energy through the joint. Also called a rotary coupler [11].

ROTOR—(1) The revolving part of a rotating electrical machine. The rotor may be either the field or the armature, depending on the design of the machine [5]. (2) The rotating member of a synchro that consists of one or more coils of wire wound on a laminated core. Depending on the type of synchro, the rotor functions similarly to the primary or secondary winding of a transformer [15].

RPK—The symbol used to represent the resistance between the cathode and plate of a tube [6].

RUNNING OPEN—The teletypewriter condition where the type hammer constantly strikes the type box but does not print or move across the page [17].

SATELLITE ECLIPSE—An eclipse where the rays of the sun do not reach the satellite. This prevents recharging of the solar cells of the satellite and decreases the power to the transmitter [17].

SATELLITE-SUN CONJUNCTION—A period when the satellite and sun are close together and the noise from the sun prevents or hampers communications [17].

SATURABLE-CORE REACTOR—A coil in which the reactance is controlled by changing the permeability of the core [8].

SATURATION—(1) The condition existing in any circuit in which an increase in the input signal produces no further change in the output [13]. (2) The operating point of a vacuum tube or transistor at which a further increase in grid or base current no longer produces an increase in plate or collector current [6] [7]. (3) In a magnetic core, the condition in which a magnetic material has reached a maximum flux density and the permeability has decreased to a value of (approximately) 1 [8].

SCALING FACTOR—The term used to describe the use of unequal resistors in a servo's summing network to compensate for differences between input and output signal levels [15].

SCANNING—(1) The process of subdividing a picture in an orderly manner into segments. This is used in facsimile transmission [17]. (2) Systematic movement of a radar beam to cover a definite pattern or area in space [18].

SCAT CODE—A four-digit subcategory code used to identify the functional measurement parameters that can be satisfied by any one of many pieces of test equipment [16].

SCATTER ANGLE—The angle at which the receiving antenna must be aimed to capture the scattered energy of tropospheric scatter [10].

SCHEMATIC—A diagram which shows, by means of graphic symbols, the electrical connections and functions of a specific circuit arrangement [1] [4].

SCHEMATIC SYMBOLS—A letter, abbreviation, or design used to represent specific characteristics or components on a schematic diagram [1].

SCINTILLATION—Apparent change in target reflectivity. Motion of the target causes successive radar pulses to bounce off different parts of the target, such as fuselage and wingtip [18].

SCREEN GRID—A grid placed between a control grid and the plate and usually maintained at a fixed positive potential [6].

SCREENING—Process of applying nonconductive or semiconductive materials to a substrate to form thick film components [14].

SEA CLUTTER—Unwanted echoes from the irregular surface of the sea that appear on a radar indicator [18].

SEARCH RADAR SYSTEM—An early-warning device that searches a fixed volume of space [18].

SECAS (SHIP EQUIPMENT CONFIGURATION ACCOUNTING SYSTEM)—The Navy system that keeps track of the configuration of equipment in the fleet [16].

SECONDARY—The output coil of a transformer [2].

SECONDARY CELL—A cell that can be recharged by a current being passed through the cell in a direction opposite to the discharge current [1].

SECONDARY EMISSION—The liberation of electrons from an element, other than the cathode, as a result of being struck by other high-velocity electrons [6].

SECONDARY LOOP—In a cooling system, the loop that transfers the heat from the heat source, such as electronic equipment, to the primary loop; usually distilled water [18].

SECOND DETECTOR (DEMODULATOR)—The part of the receiver that separates the audio or video component from the modulated intermediate frequency [18].

SECOND-SWEEP ECHOES—See AMBIGUOUS RETURNS [18].

SELECTIVITY—The ability of a receiver to select the desired signal and reject unwanted signals [9] [17].

SELENIUM—A chemical element with light-sensitive properties that makes it useful as a semiconductor material in metallic rectifiers [7].

SELF-BIAS—In a vacuum tube circuit, the voltage developed by the flow of current through a resistor in the grid or cathode leads [6].

SELF-EXCITED GENERATORS—DC generators in which the generator output is fed to the field to produce field excitation [5].

SELF-EXCITED METER—A term used to describe meters that operate from their own power sources [16].

SELF-INDUCTION—(1) The production of a counterelectromotive force in a conductor when its own magnetic field collapses or expands with a change in current in the conductor [2]. (2) The phenomenon caused by the expanding and collapsing fields of an electron that encircle other electrons and retard the movement of the encircled electrons [10].

SELF-LUMINOUS BODIES—Objects that produce their own light [10].

SELF-SYNCHRONIZED RADAR—A type of radar in which the timing pulses are generated within the transmitter [18].

SENDING END—See INPUT END [10].

SENSITIVITY—(1) For an ammeter, the amount of current that will cause full-scale deflection of the meter. (2) For a voltmeter, the ratio of the voltmeter resistance divided by the full-scale reading of the meter; expressed in ohms per volt [3] [16]. (3) The ability of a receiver to reproduce very weak signals. The greater the receiver sensitivity, the weaker the signal that can be reproduced [17]. (4) Efficiency of a microphone. Describes microphone power delivered to a matched-impedance load as compared to the sound level being converted. Usually expressed in terms of the electrical power level [12].

SENSITIVITY TIME CONTROL (STC)—A circuit that varies the gain of a receiver as a function of time [18].

SERIES CIRCUIT—An arrangement where electrical devices are connected so that the total current must flow through all the devices; electrons have one path to travel from the negative terminal to the positive terminal [1].

SERIES-CONNECTED DUPLEXER—A configuration in which the tr spark gap is connected in series in one leg of the transmission line one-half wavelength away from the T-junction [18].

SERIES-DIODE DETECTOR—The semiconductor diode in series with the input voltage and the load impedance. Sometimes called a VOLTAGE-DIODE DETECTOR [12].

SERIES-FED OSCILLATOR—An oscillator in which dc power is supplied to the amplifier through the tank circuit or a portion of the tank circuit [9].

SERIES LIMITER—A diode connected in series with the output, in which the output is taken across the resistor. Either the positive or negative alternation of the input wave is eliminated [9].

SERIES-NEGATIVE LIMITER—A diode connected in series with the output, in which the output is taken across the resistor. It eliminates the negative alternation of the input wave [9].

SERIES-PARALLEL CIRCUIT—A circuit that consists of both series and parallel networks [1] [9].

SERIES PEAKING—A technique used to improve high-frequency response in which a peaking coil is placed in series with the output signal path [8].

SERIES-POSITIVE LIMITER—A diode connected in series with the output, in which the output is taken across a resistor. It eliminates the positive alternation of the input wave [9].

SERIES-RESONANT CIRCUIT—A resonant circuit in which the source voltage is connected in series with a capacitor and an inductor (also in series) to furnish a low impedance at the frequency at which the circuit is resonant [9] [10].

SERIES VOLTAGE REGULATOR—A regulator with a regulating device that is in series with the load resistance [7].

SERIES-WOUND MOTORS AND GENERATORS—Machines in which the armature and field windings are connected in series with each other [5].

SERVOAMPLIFIER—Either ac or dc amplifiers used in servosystems to build up signal strength. These amplifiers usually have relatively flat gain versus frequency response, minimum phase shift, low output impedance, and low noise level [15].

SERVOMOTOR—An ac or dc motor used in servosystems to move a load to a desired position or at a desired speed. The ac motor is usually used to drive light loads at a constant speed, while the dc motor is used to drive heavy loads at varying speeds [15].

SERVOSYSTEM—An automatic feedback control system that compares a required condition (desired value, position, and so forth) with an actual condition and uses the difference to drive a control device to achieve the required condition [15].

SET—A unit or units and the assemblies, subassemblies, and parts connected or associated together to perform a specific function [17].

SEXADECIMAL—Same as HEXADECIMAL [13].

SHADOW—A dead spot (minimum radiation) caused by the physical obstruction of transmitted waves by a feed horn [18].

SHAPING CIRCUIT—A circuit that alters the shapes of input waveforms [9].

SHARP-CUTOFF TUBE—The opposite of a remote-cutoff tube. An electron tube that has evenly spaced grid wires. The amplification of the sharp-cutoff tube is limited by the bias voltage and tube characteristics [6].

SHELF LIFE—The period of time that a cell or battery may be stored and still be useful [1].

SHIELDED PAIR—A line consisting of parallel conductors separated from each other and surrounded by a solid dielectric [10].

SHIELDING—(1) A metallic covering used to prevent magnetic or electromagnetic fields from affecting an object [1]. (2) Technique designed to minimize internal and external interference [14].

SHORT CIRCUIT—An unintentional current path between two components in a circuit or between a component and ground; usually caused by a circuit malfunction [1] [3] [16].

SHORT-CIRCUITED LINE—A transmission line that has a terminating impedance equal to 0 [10].

SHUNT—A resistive device placed in parallel with another component. Appreciable current may flow through it and an appreciable voltage may exist across it [12].

SHUNT-DIODE DETECTOR—A diode detector in which the diode is in parallel with the input voltage and the load impedance. Also known as a current detector because it operates with smaller input levels [12].

SHUNT-FED OSCILLATOR—An oscillator that receives its dc power for the transistor or tube through a path both separate from and parallel to the tank circuit [9].

SHUNT PEAKING—A technique used to improve high-frequency response in which a peaking coil is placed in parallel (shunt) with the output signal path [8].

SHUNT RESISTOR—A resistor in parallel. In an ammeter, shunt resistors are used to provide multiple ranges [3].

SHUNT VOLTAGE REGULATOR—A regulator whose regulating device is in parallel with the load resistance [7].

SHUNT-WOUND MOTORS AND GENERATORS—Machines in which the armature and field windings are connected in parallel (shunt) with each other [5].

SIEMENS—The new and preferred term for MHO [1].

SIGNAL—A general term used to describe any ac or dc of interest in a circuit; for example, input signal [8] [15].

SIGNAL DISTORTION—Any unwanted change to the signal [12].

SIGNIFICANT SIDEBANDS—Those sidebands with significantly large amplitude [12].

SILICON—A metallic element which, in its pure state, is used as a semiconductor [7].

SILICON-CONTROLLED RECTIFIER (SCR)—A semiconductor device that functions as an electrically controlled switch [7].

SINE WAVE—(1) The curve traced by the projection on a uniform time scale of the end of a rotating arm, or vector. Also known as a sinusoidal wave [2]. (2) The basic synchronous alternating waveform for all complex waveforms [12].

SINGLE-ENDED MIXER—See UNBALANCED CRYSTAL MIXER [18].

SINGLE LINE DIAGRAM—A diagram which shows, by means of single lines and graphic symbols, the course of an electric circuit or system of circuits and the component devices or parts used therein [4].

SINGLE, STATIONARY-LOBE SCANNING SYSTEM—Antenna (with a single, stationary beam) that is rotated to obtain 360-degree coverage [18].

SINK—See OUTPUT END [10].

SKIN EFFECT—The tendency for alternating current to concentrate in the surface layer of a conductor. The effect increases with frequency and serves to increase the effective resistance of the conductor [10] [11].

SKIP DISTANCE—The distance from a transmitter to the point where the sky wave is first returned to earth [10].

SKIP ZONE—A zone of silence between the point where the ground wave becomes too weak for reception and the sky wave is first returned to earth [10].

SKY WAVES—Radio waves reflected back to earth from the ionosphere [10].

SLANT RANGE—See RANGE [18].

SLIP—The difference between rotor speed and synchronous speed in an ac induction motor [5].

SLIP RINGS—Contacts that are mounted on the shaft of a motor or generator to which the rotor windings are connected and against which the brushes ride [5]. Devices for making electric connections between stationary and rotating contacts.

SLOPE DETECTOR—A tank circuit tuned to a frequency, either slightly above or below an FM carrier frequency, that is used to detect intelligence [12].

SLOT—Narrow opening in a waveguide wall used to couple energy in or out of the waveguide. Also called an aperture or a window [11].

SNAP-ACTING—Changing position quickly with the aid of a spring [3].

SOLENOID—An electromagnetic device that changes electrical energy into mechanical motion; based upon the attraction of a movable iron plunger to the core of an electromagnet [3].

SOLID—One of the three states of matter; it has definite volume and shape (ice is a solid) [1].

SOLID-STATE DEVICE—An electronic device that operates by the movement of elections within a solid piece of semiconductor material [7].

SONIC—Pertaining to sounds capable of being heard by the human ear [10].

SOURCE—(1) The object that produces the waves or disturbance. (2) The name given to them end of a two-wire transmission line that is connected to a source [10]. (3) The device which furnishes the electrical energy used by a load [1].

SOURCE, MAINTENANCE, AND RECOVERABILITY CODE (SM & R CODE)—Specifies maintenance level for repair of components or assemblies [14].

SPACE—Absence of an RF signal in cw keying. Key-open condition or lack of data in communications systems. Also a period of no signal [12].

SPACE CHARGE—An electrical charge distributed throughout a volume or space [6].

SPACE DIVERSITY—Reception of radio waves by two or more antennas spaced some distance apart [10].

SPACE WAVE—Radio waves that travel directly from the transmitter to the receiver and remain in the troposphere [10].

SPACING—The condition in teletypewriter operation where a circuit is open and no current flows [17].

SPARK-GAP MODULATOR—A modulator that consists of a circuit for storing energy, a circuit for rapidly discharging the storage circuit (spark gap), a pulse transformer, and a power source [12].

SPECIAL PURPOSE ELECTRONIC TEST EQUIPMENT (SPETE)—Test equipment that is specifically designed to generate, modify, or measure a range of electronic functions of a specific or peculiar nature on a single system or equipment [16].

SPECIFIC GRAVITY—The ratio between the density of a substance and that of pure water at a given temperature [1].

SPECIFIC RESISTANCE—The resistance measured in ohms of a unit volume of a substance to the flow of electric current. (The unit volume used is generally the circular mil-foot.) [4]

SPECTRUM—(1) The entire range of electromagnetic waves arranged in order of their frequencies. (2) The range of frequencies considered in a system [10].

SPECTRUM ANALYSIS—The display of electromagnetic energy arranged according to wavelength or frequency [12].

SPECTRUM ANALYZER—A test instrument that provides a visual display of the frequency distribution of an RF signal such as a transmitter output [18].

SPIN VECTOR—In a gyro, a vector representing the angular velocity of the gyro rotor. The spin vector lies along the spin axis of the rotor [15].

SPHERICAL WAVEFRONTS—Waves of energy that spread out in concentric circles [18].

SPLATTER—Unwanted sideband frequencies that are generated from overmodulation [12].

SPLICE—A joint formed by the connecting of two or more conductors [4].

SPORADIC E LAYER—Irregular, cloud-like patches of unusually high ionization. Often forms at heights near the normal E layer [10].

SPREADER—Insulator used with transmission lines and antennas to keep the parallel wires separated [10].

SPROCKET TUNER—A mechanical tuning device for magnetron tubes that changes the frequency of the cavities by changing the inductance. Also called a CROWN-OF-THORNS TUNER [11].

SQUARE MIL—The area of a square, the sides of which are each equal to 1 mil. One square mil is equal to 1.2732 circular mils [4].

SQUELCH—A circuit that cuts off the output of a receiver when there is no input [17].

SQUIRREL-CAGE WINDINGS—A type of rotor winding in which heavy conductors are imbedded in the rotor body. The conductors are shorted together at the ends by continuous rings. It is widely applied in ac induction motors. Physically, it appears as a rotating squirrel-cage, thus the name [5].

STABILITY—In a magnetron, the ability to maintain normal operating characteristics [18].

STAGE—One of a series of circuits within a single device; for example, first stage of amplification [8].

STAGGER TUNING—A method of klystron tuning in which the resonant cavities are tuned to slightly different frequencies to increase the bandwidth of the amplifier [11].

STANDING WAVE—The distribution of voltage and current, formed by the incident and reflected waves, that has minimum and maximum points on a resultant wave that appear to stand still [10].

STANDING-WAVE RATIO (SWR)—The ratio of the maximum (voltage, current) to the minimum (voltage, current) points of a transmission line. Indicates the impedance matching quality of the termination of the line [10] [11].

START—The first unit of a teletypewriter signal. It is always a space [17].

STATIC—(1) A fixed nonvarying condition, without motion [13]. (2) Atmospheric noise, as in a receiver.

STATIC ELECTRICITY—Stationary electricity that is in the form of a charge. The accumulated electric charge on an object [1].

STATOR—(1) The stationary part of a rotating electrical machine. The stator may be either the field or the armature, depending on the design of the machine [5]. (2) The stationary member of a synchro that consists of a cylindrical structure of slotted laminations on which three Y-connected coils are wound with their axes 120 degrees apart. Depending on the type of synchro, the stator's functions are similar to the primary or secondary windings of a transformer [15].

STATUTE MILE—5,280 feet [18].

STEP-BY-STEP COUNTER—A counter that provides an output for each cycle of the input in one-step increments [9].

STEP-TRANSMISSION SYSTEM—A data transmission system that operates on direct current. It consists of a step transmitter (rotary switch) and a step motor interconnected to transmit data (information) between remote locations [15].

STICKOFF VOLTAGE—A low voltage used in multispeed synchrosystems to prevent false synchronizations [15].

STOP—The last unit of a teletypewriter signal. It is always a mark [17].

STRANDED CONDUCTOR—A conductor composed of a group of wires. The wires in a stranded conductor are usually twisted together and not insulated from each other [4].

STRANDS—Fine metallic filaments twisted together to form a single wire [4].

STRATOSPHERE—Located between the troposphere and the ionosphere; it has little effect on radio waves [10].

STROBOSCOPE—An instrument that allows viewing of rotating or reciprocating objects by producing the optical effect of a slowing down or stopping motion [16].

STUB—Short section of a transmission line used to match the impedance of a transmission line to an antenna. Can also be used to produce desired phase relationships between connected elements of an antenna [10] [18].

SUBASSEMBLY—Consists of two or more parts that form a portion of an assembly or a unit [17].

SUBHARMONIC—An exact submultiple of the fundamental frequency. Even subharmonics are one-half, one-quarter, and so on. Odd subharmonics are one-third, one-fifth, and so on of the fundamental frequency [17].

SUBSTRATE—Mounting surface for integrated circuits. May be semiconductor or insulator material depending on type of IC [14].

SUDDEN IONOSPHERIC DISTURBANCE—An irregular ionospheric disturbance that can totally blank out hf radio communications [10].

SUMMING NETWORK—A combination of two or more parallel resistors used in servosystems as an error detector. The output of the network is the algebraic sum of the inputs [15].

SUPERHETERODYNE RECEIVER—A type of receiver that uses a mixer to convert the RF echo to an IF signal for amplification [18].

SUPERHIGH FREQUENCY—The band of frequencies from 3 gigahertz to 30 gigahertz [17].

SUPERSONIC—(1) Speed greater than the speed of sound [10]. (2) Ultrasonic.

SUPPORT SYSTEM—For a radar, a system that provides an auxiliary input, such as dry air, electrical power, or liquid cooling [18].

SUPPRESSION—The process of eliminating an undesired portion of a signal [17].

SURFACE WAVE—Radio waves that travel along the contours of the earth, thereby being highly attenuated [10].

SWAMPING RESISTOR—A resistor used to increase or "broaden" the bandwidth of a circuit [8].

SWITCH—(1) A device used to connect, disconnect, or change the connections in an electrical circuit [1]. (2) A device used to open or close a circuit [3].

SYMMETRICAL MULTIVIBRATOR—A circuit that generates square waves [18].

SYMPTOM ELABORATION—Using built-in indicating instruments or other aids to define equipment malfunction [16].

SYMPTOM RECOGNITION—Recognition of a situation in equipment operation that is not normal [16].

SYNCHRO—A small motorlike analog device that operates like a variable transformer and is used primarily for the rapid and accurate transmission of data among equipments and stations [15].

SYNCHRO CAPACITOR—A unit containing three delta-connected capacitors. The synchro capacitor is used in synchro systems to increase the system's accuracy by cancelling or reducing the phase shift introduced by synchro inductance [15].

SYNCHRONIZER—A circuit that supplies timing signals to other radar components [18].

SYNCHRONIZING NETWORK—A circuit, also called a crossover or switching network, used in servosystems to sense how far the load is from the point of correspondence; it then functions to switch the appropriate signal into control [15].

SYNCHRONOUS—A type of teletypewriter operation where both transmitter and receiver operate continuously [17].

SYNCHRONOUS MOTOR—An ac motor whose rotor is activated by dc. It is characterized by constant speed and requires squirrel-cage windings or some other method to be self-starting [5].

SYNCHRONOUS ORBIT—An orbit in which the satellite moves or rotates at the same speed as the earth [17].

SYNCHRONOUS SPEED—The speed at which the rotating field in an ac motor revolves. This speed is a function of the number of poles in the field and the frequency of the applied voltage [5].

SYNCHRONOUS TUNING—In a klystron amplifier, a method of tuning that tunes all the resonant cavities to the same frequency. High gain is achieved, but the bandwidth is narrow [11].

SYNCHRO SYSTEM—Two or more synchros interconnected electrically. The system is used to transmit data among equipments and stations [15].

SYNCHRO TESTER—A synchro receiver with a calibrated dial. This receiver is used primarily for locating defective synchros. It can also be used for zeroing synchros [15].

SYSTEM—A combination of sets, units, assemblies, subassemblies, and parts joined together to form a specific operational function or several functions [17].

TACHOMETER—(1) A small ac or dc generator, sometimes referred to as a rate generator, that converts its shaft speed into an electrical output. The tachometer is frequently used in servosystems to sense the velocity of a load [15]. (2) An instrument that measures the rate at which a shaft is turning [16].

TANK CIRCUIT—A tuned circuit used to temporarily store energy. Also referred to as a parallel-resonant circuit [9].

TAPPED RESISTOR—A wire-wound, fixed resistor having one or more additional terminals along its length, generally for voltage-divider applications [1].

TARGET—In radar, a specific object of radar search or detection [18].

TARGET RESOLUTION—The ability of a radar to distinguish between two or more targets that are close to each other [18].

TELECOMMUNICATIONS—The transmission, emission, or reception of signs, signals, writings, images, or sounds. This is done by visual, oral, wire, radio, or other means [17].

TELETYPEWRITER—A machine that can transmit and/or receive letters, numbers, or symbols. It may have a keyboard similar to a typewriter [17].

TEMPERATURE COEFFICIENT—The amount of change of resistance in a material per unit change in temperature [1] [4].

TEMPERATURE INVERSION—The condition in which warm air is formed above a layer of cool air that is near the earth's surface [10].

TEMPEST—A term normally used to describe compromising emanations. These emanations are unintentionally radiated signals that could disclose classified information [17].

TENSILE STRENGTH—The greatest stress a substance can withstand along its length without tearing apart [4].

TERMINAL—An electrical connection [1] [4].

TERMINAL BOARD—Also called a terminal strip. An insulating base or slab equipped with terminals for connecting wiring [4].

TERMINAL DIAGRAM—A diagram of a switch, relay, terminal board, or other component showing the connections to the equipment [4].

TERMINAL LUG—A device attached to a conductor to permit connection to a terminal [4].

TEST EQUIPMENT—A general term applied to devices used to test electrical and electronic circuits [3].

TEST EQUIPMENT INDEX—The Navy guide used to assist in identifying portable electrical/ electronic test equipment required for support of prime electrical/electronic, IC, weapons, and reactor instrumentation systems [16].

TEST POINTS—Locations in equipment that are accessible to the technician's test probes where operating voltages or signals can be monitored [16].

TETRODE TUBE—A four-electrode electron tube containing a plate, a cathode, a control grid, and a screen grid [6].

THERMAL INERTIA—The capacity of a soldering iron to generate and maintain a satisfactory soldering temperature while giving up heat to the material being soldered [4].

THERMAL-MAGNETIC TRIP ELEMENT—A single circuit breaker trip element that combines the action of a thermal and a magnetic trip element [3].

THERMAL RUNAWAY—A conduction that exists when heat causes more electron-hole pairs to be generated, which, in turn, causes more heat and which may eventually cause diode destruction [7].

THERMAL TRIP ELEMENT—A circuit breaker trip element that uses the increased bending of a bimetallic strip caused by increased current to open a circuit [3].

THERMIONIC EMISSION—Emission of electrons from a solid body as a result of elevated temperature [6].

THERMISTOR—(1) A semiconductor device whose resistance varies with temperature [4]. (2) A type of bolometer characterized by a decrease in resistance as the dissipated power increases [16].

THERMOCOUPLE—A junction of two dissimilar metals that produces a voltage when heated [1].

THERMOCOUPLE METER MOVEMENT—A meter movement that uses the current induced in a thermocouple by the heating of a resistive element to measure the current in a circuit; used to measure ac or dc [3].

THERMOPLASTIC—A synthetic mixture of rosins that is flexible and used as an insulting material. Generally used as an insulator for low- and medium-range voltages [4].

THETA—The Greek letter (!) used to represent phase angle [2].

THICK FILM COMPONENTS—Passive circuit components (resistors and capacitors) having a thickness of 0.001 centimeter [14].

THIN FILM COMPONENTS—Passive circuit elements (resistors and capacitors) deposited on a substrate to a thickness of 0.0001 centimeter [14].

THREE-ELEMENT ARRAY—An array with two parasitic elements (reflector and director) and a driven element [10].

THREE-DIMENSIONAL RADAR (3D)—A radar set that measures the range, bearing and altitudes of a target [18].

THROW—In a switch, the number of different circuits each pole can control; for example, single throw and double throw [3].

THYRATRON—A gas tube used as a modulator switching device [18].

THYRATRON TUBE—A gas-filled triode in which a sufficiently large positive pulse applied to the control grid ionizes the gas and causes the tube to conduct, after which the control grid has no effect in conduction [6] [12].

TICKLER COIL—A small coil connected in series with the collector or plate circuit of a transistor or tube and inductively coupled to the base or grid-circuit coil to establish feedback (regeneration) [9].

TIME CONSTANT—Time required for an exponential quantity to change by an amount equal to 63.2 percent of the total change that can occur [2] [9].

TIME-DIVISION MULTIPLEXING—The process that periodically samples the full 360 degrees of each sine wave. The sample can be of a received signal or of a signal to be transmitted [17].

TIME LAG—The delay in a servosystem between the application of the input signal and the actual movement of the load [15].

TIMER—See SYNCHRONIZER [18].

TINNING—The process of applying a thin coat of solder to materials prior to their being soldered; for example, application of a light coat of solder to the filaments of a conductor to hold the filaments in place prior to soldering of the conductor [4].

TOLERANCE—(1) The maximum permissible error or variation from the standard in a measuring instrument. (2) A maximum electrical or mechanical variation of specifications that can be tolerated without impairing the operation of a device [1].

TONES—Musical sounds [10].

TONE-TERMINAL SET—Equipment that converts TTY dc pulses into audio tones for modulation of a transmitter in audio-frequency-tone shift transmissions [17].

TOP-HAT—An antenna that is center-fed and capacitively loaded [17].

TORQUE—A measure of how much load a machine can turn. This measurement is expressed either in ounce-inches for torque synchro systems or in pound-feet for heavy machinery [15].

TORQUE DIFFERENTIAL RECEIVER (TDR)—A type of differential synchro that takes two electrical inputs, one to the rotor and one to the stator, and produces a mechanical output. The output is the angular position of the rotor that represents the algebraic sum or difference of the two electrical inputs [15].

TORQUE DIFFERENTIAL SYNCHRO SYSTEM—A synchro system containing either a TDX or a TDR. This system is used in applications where it is necessary to compare two signals, add or subtract the signals, and furnish an output proportional to the sum or difference between the two signals [15].

TORQUE DIFFERENTIAL TRANSMITTER (TDX)—This type of synchro is functionally the same as the CDX, except that it is used in torque systems rather than control systems [15].

TORQUE GRADIENT—A term used in the rating of torque synchros. It is expressed in the number of inch-ounces of torque required to pull a specific synchro 1 degree away from its normal position; for example, 0.4 inch-ounce per degree [15].

TORQUE RECEIVER (TR)—A type of synchro that converts the electrical input supplied to its stator back to a mechanical angular output through the movement of its rotor [15].

TORQUE SYNCHRO SYSTEM—A synchro system that uses torque synchros to move light loads such as dials, pointers, and other similar devices [15].

TORQUE TRANSMITTER (TX)—This type of synchro is functionally the same as the CX, except that it is used in torque synchro systems [15].

TORQUE VECTOR—In a gyro, a vector representing the rotary motion applied to change the direction of the rotor axis. The torque vector represents the axis about which the applied force is felt [15].

TOTAL RESISTANCE—(R_T) The equivalent resistance of an entire circuit. For a series circuit: $R_T = R_1 + R_2 + R_3 \ldots R_n$. For parallel circuits:

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots \frac{1}{R_n} \quad [1].$$

TOXIC VAPORS—Vapors emitted by a substance that can do bodily harm [4].

TR RECOVERY TIME—Time required for a fired tr or atr tube to deionize to a normal level of conductance [18].

TRACK—Operational phase of a fire-control or track radar during which the radar beam is kept on the target [18].

TRACK RADAR—Radar that provides continuous range, bearing, and elevation data by keeping the RF beam on the target [18].

TRANSCONDUCTANCE—Transconductance is a ratio of the change in plate current to a change in grid voltage with the plate voltage held constant. Transconductance (g_m) is usually expressed in micromhos or millimhos. Mathematically,

$$g_m = \frac{I_p}{E_g} \quad [6][16].$$

TRANSFORMER—A device composed of two or more coils, linked by magnetic lines of force, used to transfer energy from one circuit to another [2].

TRANSFORMER EFFICIENCY—The ratio of output power to input power, generally expressed as a percentage.

$$\text{Efficiency} = \frac{P_{out}}{P_{in}} \times 100 \quad [2].$$

TRANSFORMER, STEP-DOWN—A transformer so constructed that the number of turns in the secondary winding is less than the number of turns in the primary winding. This construction will provide less voltage in the secondary circuit than in the primary circuit [2].

TRANSFORMER, STEP-UP—A transformer so constructed that the number of turns in the secondary winding is more than the number of turns in the primary winding. This construction will provide more voltage in the secondary circuit than in the primary circuit [2].

TRANSISTOR—A semiconductor device with three or more elements [7].

TRANSITION—The time it takes to shift from a mark to a space condition or from a space to a mark condition [17].

TRANSIT TIME—The time an electron takes to cross the distance between the cathode and the plate [6] [11].

TRANSLATION—In a gyro, a force acting through the center of gravity of the gyro that causes no torque on the gyro rotor. Translation forces do not change the angle of the plane of rotation but move the gyroscope as a unit [15].

TRANSLUCENT—Those substances, such as frosted glass, through which some light rays can pass but through which objects cannot be seen clearly [10].

TRANSMISSION LINE—A device designed to guide electrical or electromagnetic energy from one point to another [10].

TRANSMISSION MEDIUM—A means of transferring intelligence from point to point; includes light, smoke, sound, wire lines, and radio-frequency waves [10] [12].

TRANSMIT-RECEIVE TUBE (TR)—A gas-filled RF switch that is used as a duplexer [18].

TRANSMITTER—Equipment that generates and amplifies an RF carrier, modulates the RF carrier with intelligence, and radiates the signal into space [17] [18].

TRANSMITTER DISTRIBUTOR—A device that reads Baudot code from paper tape and allows a message to be printed on a page printer [17].

TRANSMITTER END—See INPUT END [10].

TRANSMITTER FREQUENCY (CARRIER FREQUENCY)—The frequency of the unmodulated output of a transmitter [18].

TRANSMITTER TRANSFER SWITCHBOARD—Equipment that selectively transfers remote station functions and signals to transmitters [17].

TRANSMITTING ANTENNA—The device used to send the transmitted signal energy into space [10].

TRANSPARENT—Those substances, such as glass, that pass almost all of the light waves falling upon them [10].

TRANSVERSE ELECTRIC MODE—A waveguide mode in which the entire electric field is perpendicular to the wide dimension and the magnetic field is parallel to the length. Also called the TE mode [11].

TRANSVERSE MAGNETIC MODE—A waveguide mode in which the entire magnetic field is perpendicular to the wide dimension and some portion of the electric field is parallel to the length. Also called the TM mode [11].

TRANSVERSE WAVE MOTION—The up and down motion of a wave as the wave moves outward [10].

TRAVERSE (BEARING) SIGNAL—In a monopulse radar system, the combination of individual lobe signals that represents target-offset direction and amplitude from the antenna axis [18].

TREMENDOUSLY HIGH FREQUENCY—The band of frequencies from 300 gigahertz to 3,000 gigahertz.

TRIAC—A three-terminal device that is similar to two SCRs back to back with a common gate and common terminals. Although similar in construction and operation to the SCR, the triac controls and conducts current flow during both alternations of an ac cycle [7].

TRIATIC—A special type of monopole antenna array [17].

TRIGGER—A short pulse, either positive or negative, that can be used to cause an electronic function to take place [9].

TRIGGER GENERATOR—See SYNCHRONIZER [18].

TRIGGER PULSES—In radar, pulses that are used to initiate specific events [18].

TRIODE TUBE—A three-electrode electron tube containing a plate, a cathode, and a control grid [6].

TRIP-ELEMENT—The part of a circuit breaker that senses any overload condition and causes the circuit breaker to open the circuit [3].

TRIP-FREE CIRCUIT BREAKER—A circuit breaker that will open a circuit even if the operating mechanism is held in the ON position [3].

TRIVALENT IMPURITY—Acceptor impurities containing only three valence electrons [7].

TROPOSPHERE—The portion of the atmosphere, closest to the earth's surface, where all weather phenomena take place [10].

TROPOSPHERIC SCATTER—The propagation of radio waves in the troposphere by means of scatter [10].

TROUBLE INDICATORS—Signal lights used to aid maintenance personnel in locating troubles quickly [15].

TROUBLESHOOTING—The process of locating and diagnosing faults in equipment by means of systematic checking or analysis [3] [15] [16].

TROUBLE TABLES—Tables of trouble symptoms and probable causes, furnished by many manufacturers to help technicians isolate problems [15].

TROUGH (BOTTOM)—The peak of the negative alternation (maximum value below the line) of a sine wave [10].

TRUE BEARING—Angle between a target and true north measured clockwise in the horizontal plane [18].

TRUE NORTH—Geographic north [18].

TRUE POWER—The power dissipated in the resistance of the circuit, or the power actually used in the circuit [2].

TRUNCATED PARABOLOID—A paraboloid reflector that has been cut away at the top and bottom to increase beam width in the vertical plane [18].

TRUTH TABLE—A table that describes a logic function by listing all possible combinations of input values and indicating, for each combination, the true output values [13].

TUBE DYNAMIC CONDITION—Refers to the testing condition in which a vacuum tube is actually performing its function [16].

TUBE STATIC CONDITION—Refers to the testing condition in which a tube has certain voltages applied but is not in its normal operating condition [16].

TUNED CIRCUIT—(1) A circuit consisting of inductance and capacitance that can be adjusted for resonance at a desired frequency [9]. (2) A circuit that is used as a filter which passes or rejects specific frequecies [16]. (3) An LC circuit used as a frequency-determining device [8].

TUNED LINE—Another name for the resonant line. This line uses tuning devices to eliminate the reactance and transfer maximum power from the source to the line [10].

TUNNEL DIODE—A heavily doped semiconductor device that has high gain and fast switching capabilities [7]. See NEGATIVE-RESISTANCE ELEMENT [11].

TUNNELING—The piercing of a potential barrier in a semiconductor by a particle (current carrier) that does not have sufficient energy to go over the barrier [11].

TURN—One complete loop of a conductor about a core [2].

TURNS RATIO—The ratio of the number of turns in the primary winding to the number of turns in the secondary winding of a transformer [2].

TURNSTILE ANTENNA—A type of antenna used in vhf communications that is omnidirectional and consists of two horizontal half-wave antennas mounted at right angles to each other in the horizontal plane [10].

TWISTED PAIR—A line consisting of two insulated wires twisted together to form a flexible line without the use of spacers [10].

TWO-DIMENSIONAL RADAR (2D)—Measures the range and bearing to a target [18].

TWO-M (2M)—Miniature/microminiature repair program [14].

TWO-WIRE OPEN LINE—A parallel line consisting of two wires that are generally spaced from 2 to 6 inches apart by insulating spacers [10].

TWO-WIRE RIBBON (TWIN LEAD)—A parallel two-wire line in which uniform spacing is assured by two wires imbedded in a low-loss dielectric [10].

ULTRAHIGH FREQUENCY—The band of frequencies from 300 megahertz to 3 gigahertz [17].

ULTRASONIC—(1) Sounds above 20,000 hertz [10]. (2) Supersonic.

UNBALANCED CRYSTAL MIXER—A circuit consisting of a section of coaxial transmission line one-half wavelength long that is tuned to the difference (intermediate) frequency between the local oscillator and RF echo signals [18].

UNIDIRECTIONAL—In one direction only [1].

UNIDIRECTIONAL ARRAY—An antenna array that radiates in only one general direction [10].

UNIJUNCTION TRANSISTOR (UJT)—A three-terminal, semiconductor device with a negative resistance characteristic that is used in switching circuits, oscillators, and wave-shaping circuits [7].

UNIT—(1) An assembly or any combination of parts, subassemblies, and assemblies mounted together. Normally capable of independent operation [17]. (2) A single object or thing [13].

UNIT SIZE—The standards adopted to make comparisons between things of like value (for example, the unit size for conductors is the mil-foot) [4].

UNIVERSAL TIME CONSTANT CHART—A chart used to find the time constant of a circuit if the impressed voltage and the values of R and C or R and L are known [2].

UNTUNED LINE—Another name for the flat or nonresonant line [10].

UP LINK—The frequency used to transmit a signal from earth to a satellite [17].

UPPER-FREQUENCY CUTOFF—The highest frequency a circuit can pass [9].

UPPER SIDEBAND—All of the sum frequencies above the carrier [12].

VACUUM EVAPORATION—Process of producing thin film components [14].

VALENCE—The measure of the extent to which an atom is able to combine directly with other atoms. It generally depends on the number and arrangement of the electrons in the outermost shell of the atom [1].

VALENCE SHELL—The electrons that form the outermost shell of an atom [1].

V ANTENNA—A bidirectional antenna, shaped like a V, which is widely used for communications [10].

VAR—Abbreviation for volt-amperes reactive [2].

VARACTOR—A PN junction semiconductor, designed for microwave frequencies, in which the capacitance varies with the applied voltage [7] [11] [12].

VARACTOR FM MODULATOR—An FM modulator that uses a voltage-variable capacitor (varactor) [12].

VARIABLE—A representative symbol that can assume any of a given set of values [13].

VARIABLE ATTENUATOR—An attenuator for reducing the strength of an ac signal either continuously or in steps, without causing signal distortion [11].

VARIABLE-MU-TUBE—Same as REMOTE-CUTOFF TUBE [6].

VARIABLE RESISTOR—A wire-wound or composition resistor, the value of which may be changed over a designed range [1].

VARNISHED CAMBRIC—Cotton cloth coated with insulation varnish. An insulation used on high-voltage conductors [4].

VECTOR—A line used to represent both direction and magnitude [2] [12].

VEITCH DIAGRAM—A diagram consisting of joined squares, which is used to give a graphic representation of basic logic relations [13].

VELOCITY—The rate at which a disturbance travels through a medium [10].

VELOCITY MODULATION—Modification of the velocity of an electron beam by the alternate acceleration and deceleration of electrons [11].

VELOCITY SERVOSYSTEM—A servosystem which controls the speed of the load it is driving [15].

VERTICAL AXIS—On a graph, the straight line axis that is plotted from bottom to top [10].

VERTICAL DEFLECTION PLATES—A pair of parallel electrodes in a CRT that moves the electron beam up and down [6].

VERTICAL PATTERN—The part of a radiation pattern that is radiated in the vertical plane [10].

VERTICAL PLANE—An imaginary plane that is perpendicular to the horizontal plane [11] [18].

VERTICALLY POLARIZED—Waves that are radiated with the E field component perpendicular to the earth's surface [10].

VERY HIGH FREQUENCY—The band of frequencies from 30 megahertz to 300 megahertz [17].

VERY LARGE SCALE INTEGRATION (vlsi)—An integrated circuit containing over 2,000 logic gates or 64,000 bits of memory [14].

VERY LOW FREQUENCY—The band of frequencies from 3 kilohertz to 30 kilohertz [17].

VIDEO AMPLIFIER—An amplifier designed to amplify the entire band of frequencies from 10 hertz (10 Hz) to 6 megahertz (6 MHz). Also called a WIDE-BAND AMPLIFIER [8].

VIDEO ENHANCEMENT FEATURES—See ANTIJAMMING CIRCUITS [18].

VINCULA—Plural of vinculum (see below) [13].

VINCULUM—A straight horizontal line placed over one or more members of a compound logic expression to negate or complement. Also, used to join two or more members together [13].

VIRTUAL GROUND—A point in a circuit that is at ground potential (0 V) but is not connected to ground [8].

VOLT—The unit of electromotive force or electrical pressure. One volt is the pressure required to send 1 ampere of current through a resistance of 1 ohm [1].

VOLTAGE—(1) The term used to signify electrical pressure. Voltage is a force that causes current to flow through an electrical conductor. (2) The voltage of a circuit is the greatest effective difference of potential between any two conductors of the circuit [1].

VOLTAGE AMPLIFIER—An amplifier in which the output-signal voltage is greater than the input-signal voltage [8].

VOLTAGE-DIODE DETECTOR—A series-diode detector in which the diode is in series with the input voltage and the load impedance [12].

VOLTAGE DIVIDER—A series network in which desired portions of the source voltage may be tapped off for use in the circuit [1].

VOLTAGE DROP—The difference in voltage between two points. It is the result of the loss of electrical pressure as a current flows through a resistance [1] [4].

VOLTAGE-FEED METHOD—Same as END-FEED METHOD [10].

VOLTAGE GAIN—The ratio of output voltage to input voltage in an amplifier [6].

VOLTAGE MULTIPLIERS—Methods of increasing voltages; used primarily where low current is required [7].

VOLTAGE REGULATION—A measure of the ability of a generator or power supply to maintain a constant output voltage from no-load to full-load operation. Expressed as a percentage of full-load voltage; the better the regulation, the lower the percent [5].

VOLTAGE STANDING WAVE RATIO (VSWR)—In a waveguide, the ratio of the electric field (voltage) at a maximum point to that of an adjacent minimum point [10] [18].

VOLTMETER—A meter used to measure voltage [3] [16].

WAFER—A slice of semiconductor material upon which monolithic ICs are produced [14].

WAFER SWITCH—A rotary switch in which the contacts are arranged on levels. Each level (wafer) is electrically independent but mechanically connected by the shaft of the switch [3].

WATT—The unit of electrical power that is the product of voltage and current [16].

WATTAGE RATING—A rating expressing the maximum power that a device can safely handle [1].

WATT-HOUR—A practical unit of electrical energy equal to one watt of power for one hour [1].

WATT-HOUR METER—A meter used to measure electrical energy [3].

WATTMETER—A meter used to measure electrical power [3] [16].

WAVE ANTENNA—Same as BEVERAGE ANTENNA [10].

WAVEFORM—The shape of the wave obtained when instantaneous values of an ac quantity are plotted against time in rectangular coordinates [2].

WAVEFORM ANALYSIS—Observation displays of voltage and current variations with respect to time or by harmonic analysis of complex signals [16].

WAVEFRONT—A small section of an expanding sphere of electromagnetic radiation that is perpendicular to the direction of travel of the energy [10].

WAVEGUIDE—A rectangular, circular, or elliptical metal pipe designed to transport electromagnetic waves through its interior [10] [11].

WAVEGUIDE DUPLEXER—TR and atr tubes housed in a resonant cavity attached to a waveguide system.

WAVEGUIDE MODE OF OPERATION—Any particular field configuration in a waveguide that satisfies the boundary conditions. Usually divided into two broad types: the transverse electric (TE) and the transverse magnetic (TM) modes [11].

WAVEGUIDE POST—A rod of conductive material used as impedance changing devices in waveguides [11].

WAVEGUIDE SCREW—A screw that projects into a waveguide for the purpose of changing the impedance [11].

WAVELENGTH—The distance, usually expressed in meters, traveled by a wave during the time interval of one complete cycle. It is equal to the velocity divided by the frequency [2] [10] [12].

WAVEMETERS—Calibrated resonant circuits that are used to measure frequency [16]. An instrument for measuring the wavelength of an RF wave [18].

WAVE MOTION—A recurring disturbance advancing through space with or without the use of a physical medium [10].

WAVE TRAIN—A continuous series of waves with the same amplitude and wavelength [10].

WAVE WINDING—An armature winding in which the two ends of each coil are connected to commutator segments separated by the distance between poles [5].

WEBER'S THEORY—A theory of magnetism which assumes that all magnetic material is composed of many tiny magnets. A piece of magnetic material that is magnetized has all of the tiny magnets aligned so that the north pole of each magnet points in one direction [1].

WHEATSTONE BRIDGE—An ac bridge circuit used to measure unknown values of resistance, inductance, or capacitance [16].

WIDE-BAND AMPLIFIER—An amplifier designed to pass an extremely wide band of frequencies, such as a video amplifier [8].

WINDOW—See SLOT [11].

WIRE—An insulated conductor, with low resistance to current flow, that is either solid or stranded [1] [4].

WIRING DIAGRAM—A diagram that shows the connections of an equipment or its component devices or parts. It may cover internal or external connections, or both, and contains such detail as is needed to make or trace connections that are involved [4].

WOBBLE FREQUENCY—The frequency at which an electron wobbles on its axis under the influence of an external magnetic field of a given strength [11].

WORDS-PER-MINUTE—An approximate rate of speed. It means the number of five letter words with a space between them that can be transmitted or received in a one-minute period [17].

WORK—The product of force and motion [1].

WORKING VOLTAGE—The maximum voltage that a capacitor may operate at without the risk of damage [2].

WYE (Y)—A 3-phase connection in which one end of each phase winding is connected to a common point. Each free end is connected to a separate phase wire. The diagram of this connection often resembles the letter Y [5].

X-AXIS—In a gyro, the spin axis of the gyro [15].

X-RAY EMISSION—Penetrating radiation similar to light, but with shorter wavelength, that can penetrate human tissue [18].

YAGI ANTENNA—A multielement parasitic array. The elements lie in the same plane as those of the end-fire array [10].

Y-AXIS—In a gyro, an axis through the center of gravity and perpendicular to the spin axis [15].

Z-AXIS—In a gyro, an axis through the center of gravity and mutually perpendicular to both the X (spin) and Y axes [15].

ZENER DIODE—A PN-junction diode designed to operate in the reverse-bias breakdown region [7].

ZENER EFFECT—A reverse breakdown effect in diodes in which breakdown occurs at reverse voltages below 5 volts. The presence of a high energy field at the junction of a semiconductor produces the breakdown [7].

ZEROING—The process of adjusting a synchro to its electrical zero position [15].

ZONE OF MUTUAL VISIBILITY—The area where the satellite can be seen by both the up- and down-link earth terminals [17].

NONRESIDENT
TRAINING
COURSE
SEPTEMBER 1998

Navy Electricity and Electronics Training Series

Module 21—Test Methods and Practices

NAVEDTRA 14193

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Although the words “he,” “him,” and “his” are used sparingly in this course to enhance communication, they are not intended to be gender driven or to affront or discriminate against anyone.

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: To introduce the student to the subject of Test Methods and Practices who needs such a background in accomplishing daily work and/or in preparing for further study.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and either the occupational or naval standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards,* NAVPERS 18068.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
ETC Richard L. Baker

NAVSUP Logistics Tracking Number
0504-LP-026-8440

Sailor's Creed

"I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country's Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all."

TABLE OF CONTENTS

CREDITS

Many of the figures included in this edition of NEETS, Module 21, *Test Methods and Practices*, were provided by the companies listed below. Permission to use these illustrations is gratefully acknowledged.

SOURCE	FIGURE
Hewlett-Packard Company	2-24, 2-25, 2-27B, 2-28, 2-30, 3-25, 5-30
Huntron Instruments, Inc.	Huntron Tracker
Tektronix, Inc.	5-13, 5-27

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 6 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

THIS PAGE LEFT BLANK INTENTIONALLY.

Student Comments

Course Title: *NEETS Module 21*
Test Methods and Practices

NAVEDTRA: 14193 **Date**: ______________

We need some information about you:

Rate/Rank and Name: ______________ SSN: __________ Command/Unit ______________

Street Address: ______________ City: __________ State/FPO: ________ Zip ________

Your comments, suggestions, etc.:

Privacy Act Statement: Under authority of Title 5, USC 301, information regarding your military status is requested in processing your comments and in preparing a reply. This information will not be divulged without written authorization to anyone other than those within DOD for official use in determining performance.

NETPDTC 1550/41 (Rev 4-00)

CHAPTER 1

BASIC MEASUREMENTS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions and answers are based on the objectives and enable you to check your progress through the reading assignments. By successfully completing the OCC/ECC, you demonstrate that you have met the objectives and have learned the information. The learning objectives for this chapter are listed below.

Upon completion of this chapter, you will be able to do the following:

1. Explain the importance of performing basic electronic measurements.
2. Explain the importance of voltage measurements in troubleshooting.
3. Identify the various methods of performing voltage measurements.
4. Identify the various methods of performing current measurements.
5. Identify the various methods of performing resistance measurements.
6. Identify the various methods of performing capacitance measurements.
7. Identify the various methods of measuring inductance.

INTRODUCTION TO MEASUREMENTS

In today's modern Navy, a large part of a ship's, submarine's, or aircraft's ability to complete its mission depends on the efficiency of sophisticated electronic systems. As the technician responsible for these systems, you are the focal point in ensuring their reliability. In the event of a system failure, it is your responsibility to repair the system and to do so in a timely manner. Whether you are troubleshooting a faulty system or performing preventive maintenance, you are required to perform basic electronic measurements on a regular basis. This chapter will acquaint you with various alternative methods of performing measurements and discuss the relative merits and demerits of each method.

No discussion of electronic test equipment or electronic measurements would be complete without mentioning the Navy's Metrology Calibration (METCAL) program. Figure 1-1 shows the METCAL structure. Basically, the METCAL program is an elaborate quality control system designed to compare your electronic test equipment with test equipment of much greater accuracy. When you submit your piece of test equipment for calibration, it is compared with the calibration laboratory's equipment (referred to as STANDARDS), which are generally at least four times more accurate than yours. If your equipment does not meet specifications, it is either repaired, adjusted, or rejected with an explanation of why the calibration laboratory was unable to calibrate it. The accuracy of equipment at your local calibration laboratory is ensured by calibration of the test equipment to the standards of the next higher echelon calibration laboratory. The accuracies of test equipment at each higher echelon is increased by a ratio of approximately 4 to 1.

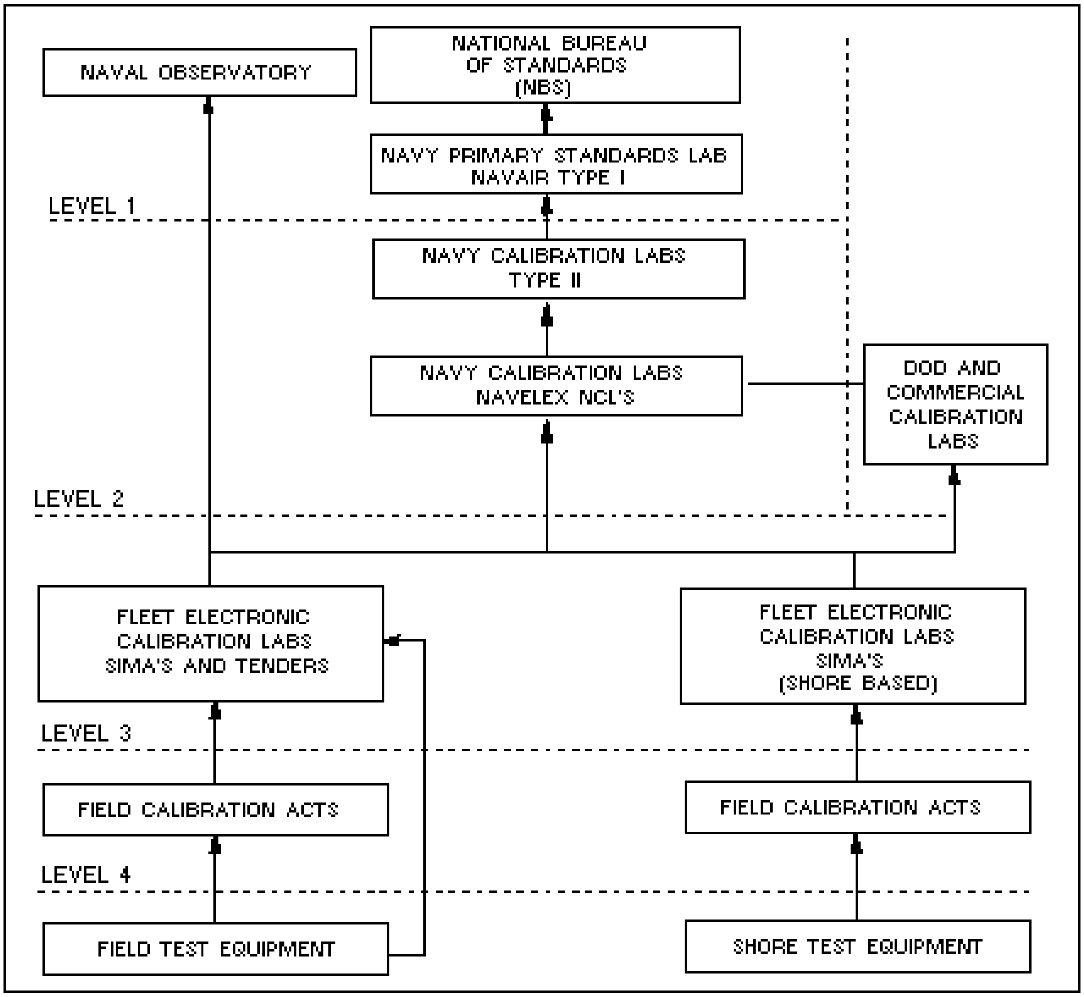


Figure 1-1.—Calibration laboratory structure.

METCAL provides assurance that your test equipment is in top-notch shape. Remember, your measurements are only as accurate as your test equipment; be fully aware of the limitations of your test equipment and never use equipment that isn't properly calibrated when performing measurements or adjustments.

Q-1. What assures the accuracy of your electronic test equipment?

Now that we have discussed the advantages of calibrated test equipment, let's review the reason for all this concern. The fundamental electrical quantities of a circuit are voltage and current and are dependent on the circuit characteristics of resistance, capacitance, and inductance. In addition to these three individual characteristics, don't forget that many electronic components exhibit more than one circuit characteristic at the same time. An example would be a piece of coaxial cable that is engineered by its manufacturer to meet characteristic specifications for impedance, capacitance, and inductance. But let's keep it simple and begin by covering voltage measurements.

Operation and use of common test equipment was covered in NEETS Module 16, *Introduction to Test Equipment,* NAVEDTRA B72-16-00-95. It is recommended that you review this module before continuing.

VOLTAGE MEASUREMENTS

Most Navy technical manuals provide voltage charts that list correct voltages at all primary test points in a piece of equipment. Voltage measurements, when compared with these charts, provide a valuable aid in locating troubles quickly and easily. However, if the sensitivity of the test equipment differs from that of the test equipment used in preparing the chart, the voltage measurements may not reflect true circuit conditions. You must keep in mind that a voltmeter with low sensitivity used on a low range may disturb circuits under test or provide a false indication. Most technical manuals will tell you what type and model of test equipment was used to prepare the voltage charts. As a rule of thumb, the input impedance of the voltmeter should exceed the impedance of the circuit by a ratio of at least 10 to 1. Technicians have spent uncounted hours of wasted time because they have selected improper test equipment.

Q-2. The input impedance of your test equipment should exceed the impedance of the circuit under test by what ratio?

DC VOLTAGE MEASUREMENTS

Direct current voltage may be steady, pulsating, or have ac superimposed on it. The average value of a dc waveform depends on the symmetry of the wave and other aspects of the wave shape. It can vary from 63.6% of peak value for a rectified full sine wave to 50% of peak value for a triangular wave. For a superimposed sine wave, the average value can be zero. Regardless of whether the dc is steady, pulsating, or the ac is superimposed on the dc, a rectifier form of measuring device will indicate its *average value*.

Voltages are usually measured by placing the measuring device in parallel with the component or circuit (load) to be measured. The measuring device should have an infinite internal resistance (input impedance) so that it will absorb *no* energy from the circuit under test and, therefore, measure the true voltage. The accuracy of the voltage measurement depends on the total resistance of the measuring device compared to the load being measured. When the input impedance of the measuring device is 10 times greater than the load being measured, the error usually can be tolerated. If this error cannot be tolerated, a high input impedance measuring device, such as a vacuum tube voltmeter (vtvm), should be used. Alternatively, using two voltmeters in series increases the voltage range and, because of the increase in total voltmeter resistance, provides a more accurate measurement of voltage across the load. If the voltage to be measured is sufficiently high, more than two similar voltmeters can be connected in series across the load to provide greater accuracy; the total voltage measurement is the sum of the individual meter indications.

Q-3. What are the advantages of using two voltmeters in series?

Multimeter Method

A common piece of test equipment used in the Navy is the Simpson 260 analog multimeter, as shown in figure 1-2. It is capable of measuring both ac and dc voltages of up to 5,000 volts.

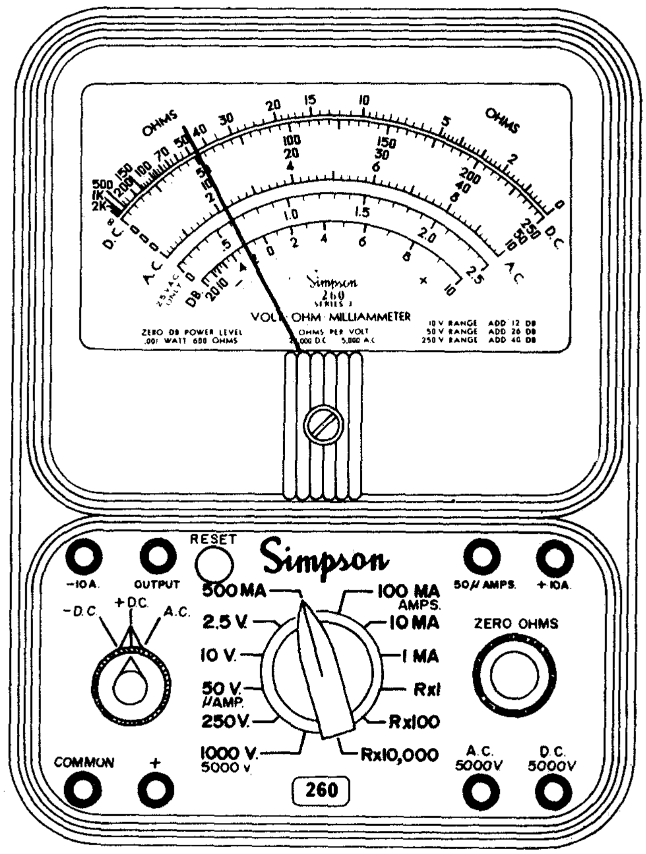


Figure 1-2.—Simpson 260 multimeter.

Two obvious advantages of the Simpson 260 are its portability and ease of operation. Among its disadvantages are its low input impedance and the inherent low accuracy associated with D'Arsonval meter movements, which are used in the meter. When performing measurements with any analog multimeter, remember that the most accurate readings are taken with the pointer midscale. You should also be aware of inaccuracies introduced as a result of parallax. PARALLAX is defined as the apparent displacement of the position of an object because of the difference between two points of view. In the case of meters, this means the position of a meter's pointer will appear to be at different positions on the scale depending on the angle from which the meter is viewed. Some of the Simpson 260 and 270 series multimeters have effectively eliminated the problem of parallax by incorporating a mirror on the scale that accurately reflects the position of the pointer of the meter movement.

Q-4. At what point on a meter movement are the most accurate readings taken?

Oscilloscope Method

A dc voltage measurement can be made with an oscilloscope, as shown in figure 1-3, that has a direct-coupled deflection amplifier or terminals for connection directly to the deflection plates of the cathode-ray tube. Measuring a dc voltage with an oscilloscope is convenient only under certain circumstances; for example, when other measurements are being made on the same equipment with the oscilloscope or when a vacuum tube voltmeter is not available and a high-impedance measuring device is required.

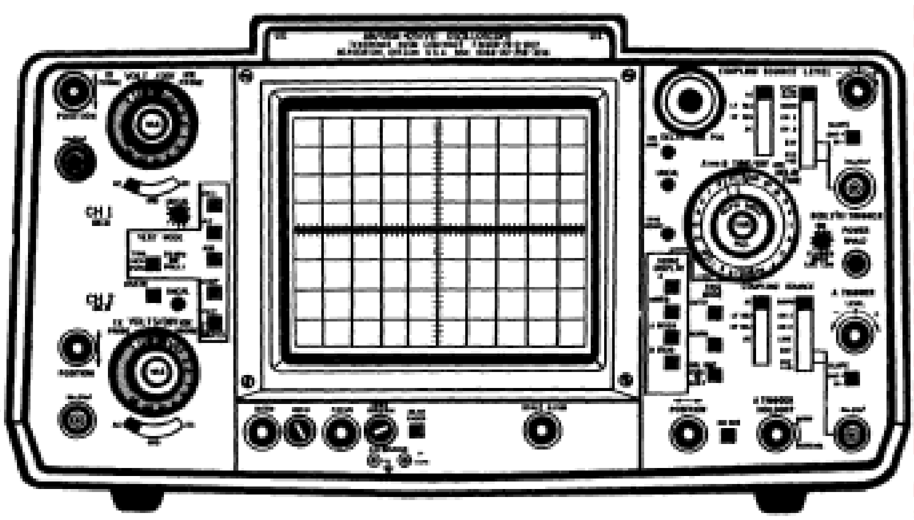


Figure 1-3.—Dual-trace oscilloscope.

Oscilloscopes have a high input impedance and normally will not load down the circuit under test. However, oscilloscopes are primarily designed for waveform observation and are typically less accurate than other pieces of test equipment used to measure dc voltages. A distinct advantage of the oscilloscope is its ability to monitor the level of ac ripple voltage riding the dc voltage. This feature makes the oscilloscope an indispensable aid in troubleshooting dc power supplies with excessive ripple caused by component failure.

Digital Multimeter Method

Most analog voltmeters (that use D'Arsonval meter movements) in common use today are accurate to approximately ±2% of full-scale reading. Most digital multimeters, as shown in figure 1-4, have a high input impedance and are not likely to disturb the circuit being tested. The digital multimeter in most cases provides an accuracy of at least ±0.1%.

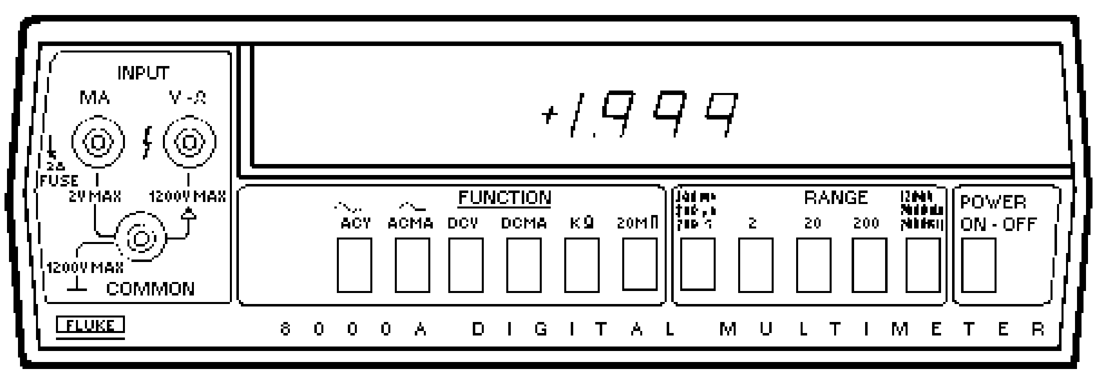


Figure 1-4.—Digital multimeter.

Digital multimeters display the reading numerically. These direct-reading displays, along with automatic range- and polarity-changing features, eliminate the problem of parallax, reduce error and tedium, and increase measurement speed. Data from these meters in digital format can also be processed by computers, printers, tape and card punches, and magnetic-tape equipment. Digital multimeters are typically compact and lightweight; many come with rechargeable batteries, making them ideal for portable field use. The disadvantages are that they are not rugged and will not tolerate abuse and that some models do not produce sufficient bias voltage to test a diode or transistor junction. The John Fluke Model 77 A/N digital multimeter is presently being purchased by the Navy and will eventually phase out the older and less accurate analog meters.

Differential Voltmeter Method

Using the differential voltmeter, as shown in figure 1-5, provides one of the most accurate methods of measuring dc voltage. Typical accuracies attained by this method are ±0.005%. These extremely high accuracies are achieved by the design of the voltmeter with precision internal reference voltages and precision resistors. As discussed earlier in NEETS, module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies*, most differential voltmeters can be operated as transistor voltmeters (tvm) or as differential null voltmeters. The tvm mode is used to measure the *approximate* voltage and polarity of the unknown voltage being measured. The approximate voltage, as measured in the tvm mode, is then used to make the initial range and mode switch selections for nulling the input voltage.

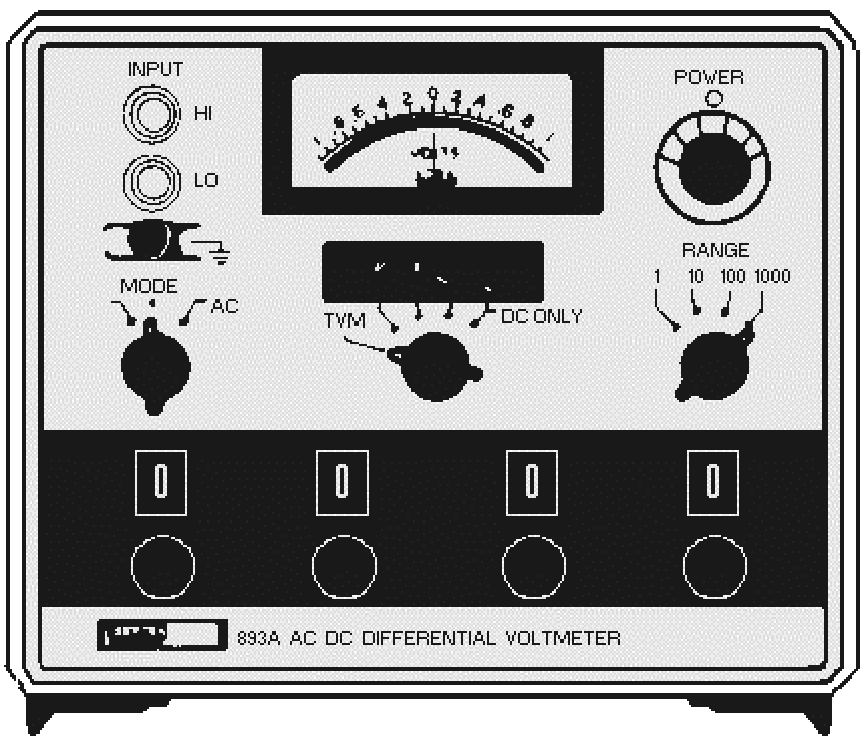


Figure 1-5.—Ac-dc differential voltmeter.

The advantages of using a differential voltmeter for measuring dc voltages are the extreme accuracy and minimal circuit loading made possible by the high input impedance of the meter. However, differential voltmeters are less portable, heavier, and require greater skill and time when performing measurements than other types of voltmeters. Additionally, they require long warm-up periods and are susceptible to variations in temperature and humidity.

Q-5. What are the advantages of using a differential voltmeter?

AC VOLTAGE MEASUREMENTS

When ac voltage measurements are performed, the input impedance of the selected test equipment determines the amount of energy removed from the circuit under test. If an ac meter is placed across a high-impedance circuit, the meter may load the high-impedance circuit and disturb circuit conditions, possibly to the point of causing the circuit to cease functioning. A dc electronic voltmeter, used in conjunction with a rectifying probe, extracts only a small amount of energy from the circuit under test. Another advantage of an electronic voltmeter over the analog voltmeter is that voltages of low values can be accurately measured.

If the circuit being measured is a relatively high-frequency circuit, the internal capacitance of an analog voltmeter rectifier could produce a disturbance by detuning the circuit. Figure 1-6 depicts the frequency response of a Simpson 260. Note the percent of error introduced at different frequencies. For high-frequency voltage measurements, an electronic voltmeter or an oscilloscope should be used. The sensitivity of the meter (or oscilloscope) determines the lowest voltage it can measure accurately, and the shunt capacitance of its input determines the upper frequency limits. It should be clear that the frequency response of a piece of test equipment is just as important as its range limitations. If you exceed the *range limitations* of a meter, it will either "peg" the meter or belch out the smell of smoke that many of us are intimately acquainted with. This, however, is not the case when you exceed the *frequency limitations* of your test equipment. Your test equipment will normally show a response, but that response will be grossly inaccurate. The lesson to be learned here is that you should be fully aware of the limitations of your test equipment and adhere to them.

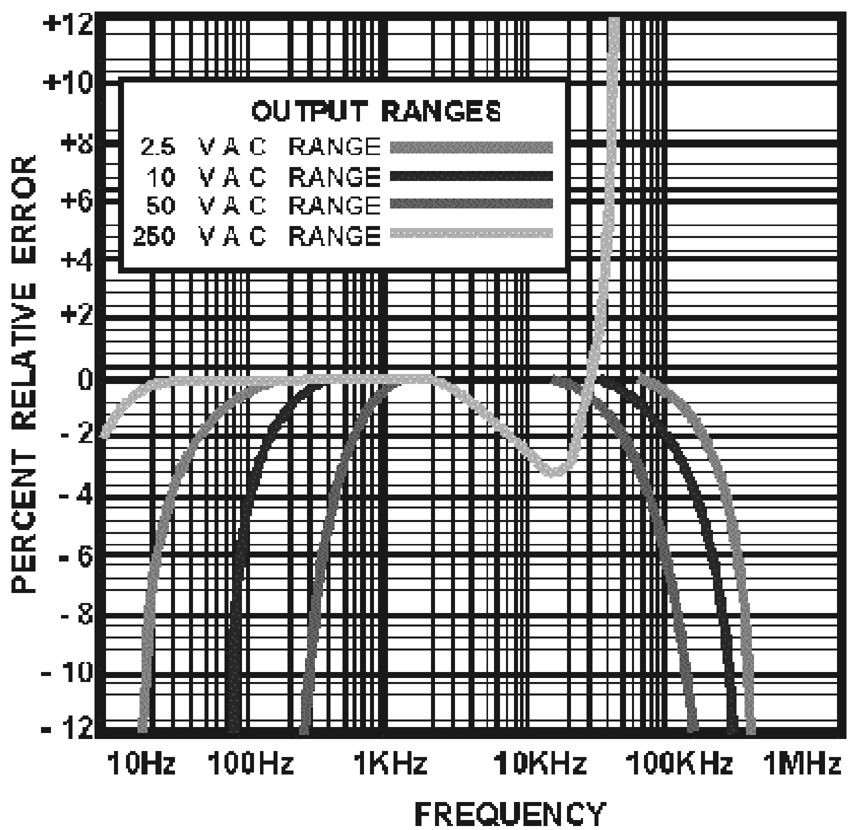


Figure 1-6.—Simpson 260 frequency response for ac voltage ranges.

Q-6. The frequency response of test equipment refers to what aspect of ac voltage measurements?

Multimeter Method

As previously stated, an analog multimeter's usefulness is limited by its low input impedance and poor accuracy (typically ±2%). However, rugged construction and ease of operation make analog multimeters extremely useful whenever poor accuracy and low input impedance can be tolerated. When performing ac voltage measurements with a multimeter, be certain that the frequency of the signal being measured falls within the upper and lower frequency limitations of the meter.

Oscilloscope Method

A major advantage of using an oscilloscope for ac voltage measurements is that the waveform can be observed; consequently, errors in measuring complex peak voltages are minimized. An oscilloscope may be used as a high-impedance ac voltmeter. In standard oscilloscopes, the vertical amplifier input impedance is generally greater than 1 megohm, making it possible to measure voltages in high-impedance circuits. If the signal is applied directly to the plates, rather than at the vertical amplifier input, the input impedance is increased considerably.

Voltage measurements are most easily made when the deflection of the trace extends across the major portion of the oscilloscope screen; whenever possible, the trace should cover at least 60% of the vertical viewing area of the screen. If the amplitude of the measured voltage is very low, the trace dimensions may be small. If a voltage to be measured is large and cannot be attenuated to a usable value by attenuation circuits within the oscilloscope, an external resistive or capacitive voltage divider can be used. Such voltage dividers are often furnished with oscilloscope test sets and are called HIGH VOLTAGE PROBES. When the voltage of pulses or other complete waveforms is being measured, the high voltage probe selected must be so designed as not to distort the measured signal. Most probes have adjustable (compensating) capacitors that are used to adjust the symmetry of the displayed waveform. You adjust the probe by monitoring either the calibrator output of the oscilloscope or a known good signal and adjusting the probe for a symmetrical display. Oscilloscopes are calibrated to display peak-to-peak values. To determine the rms voltage of a sinusoidal signal, divide the number of graticule units from the positive to the negative peaks by two and multiply this value by 0.707. When using the oscilloscope for ac voltage measurements, ensure the upper frequency range of the oscilloscope is not exceeded; otherwise, inaccurate values will be displayed. Most commonly used oscilloscopes have a frequency response from dc up to 100 megahertz.

Q-7. Ideally, an oscilloscope presentation should cover what vertical portion of the screen?

Digital Multimeter Method

As previously mentioned, digital multimeters present a high input impedance to the circuit under test and are fairly accurate. Many earlier models had very limited frequency responses. Even today the upper frequency limitations of digital multimeters vary from 20 kilohertz to over 300 kilohertz, depending on the model. Their upper frequency limitations can, however, be significantly extended by using optional rf probes. When you perform ac voltage measurements with a digital multimeter, remember that they are *true rms* indicating devices.

Differential Voltmeter Method

Most differential voltmeters can be used to measure both ac and dc voltages. The differential voltmeter method of measuring ac voltage is the most accurate of the common measurement techniques. Typical accuracies are ±0.05% when operated in the ac mode.

CURRENT MEASUREMENTS

Unless an ammeter is already an integral part of the circuit under test, current measurements are rarely taken. In the case of a high-resistance circuit, it will contain such a small amount of current that it cannot be measured accurately with ordinary field test equipment. In lower resistance circuits, current measurements can be taken only if the ammeter is placed in series with the circuit under test. These measurements require that a circuit connection be unsoldered or otherwise opened to insert the meter in series with the circuit. An easier method you may use to obtain a current measurement is to take a voltage measurement across a known resistance and calculate the current with Ohm's law. The accuracy of current measurements depends on the internal resistance of the meter as compared with the resistance of the external circuit. If the total circuit current is decreased by increasing the load, then the percentage of error will decrease. Therefore, greater accuracy is obtained if the meter resistance is considerably less than the load resistance. A method of obtaining greater accuracy of current measurement is to decrease the total internal meter resistance with respect to load resistance. This is accomplished by connecting two ammeters in parallel with each other and in series with the circuit in which the current is being measured. Additional ammeters may be connected in parallel in the same manner for increased accuracy. This method also increases the range of measurements that can be taken. The arithmetical sum of the indications of all the parallel meters represents the total current flow in the circuit. You should note that this is not a common test method and that your test equipment may be damaged if connected incorrectly.

MULTIMETER METHOD

As previously mentioned, current measurements are usually taken by breaking the current path of the circuit under test and electrically inserting a meter in series. This is normally accomplished by disconnecting a wire from a terminal or unsoldering one end of a component and electrically inserting the meter in series using the meter leads. This method is both time consuming and usually requires the use of a soldering iron, which can damage components. Most analog multimeters cannot be used for measuring ac current and are only accurate to within ±2% on dc ranges.

Q-8. What are the advantages of connecting ammeters in parallel when performing current measurements?

DIGITAL MULTIMETER METHOD

Unlike the analog multimeter, the digital multimeter will measure ac current as well as dc current. Again, current measurements are taken by breaking the current path and inserting the meter in series. Regardless of whether you're using an analog multimeter or digital multimeter, this procedure for measuring current is time consuming. However, there is a major advantage to be gained by using the digital multimeter — its high degree of accuracy. The Fluke 8000A digital multimeter, for example, is accurate to within ±0.3% when measuring dc current and ±1% when measuring ac current. These accuracies are representative of most medium-priced digital multimeters.

CURRENT TRACERS

For the purpose of discussion, we have selected the Hewlett-Packard 547A, shown in figure 1-7, as a representative current tracer. A current tracer will not actually measure current; it is designed to indicate the presence of current and the relative magnitude of one source of current as compared to another. The Hewlett-Packard 547A is a hand-held probe that enables you to precisely localize low-impedance faults in a circuit. The probe senses the magnetic field generated by a pulsing current and lights an indicator lamp near the current tracer tip. The brightness of the indicator lamp is proportional to the magnitude of the current. The sensitivity of the indicator lamp can be adjusted with a thumb-wheel potentiometer located on the probe. Figure 1-8 depicts a typical logic circuit application for a current tracer. Current tracers are ideally suited for locating shorted or opened printed-circuit-board runs, wires, or components. In the absence of a suitable pulsing current to drive the current tracer, a logic pulser or pulse generator may be used as a signal source. The inherent disadvantage of a current tracer is that it requires an external power supply. They can, however, be connected to the power supply of the equipment under test if the voltage is correct.

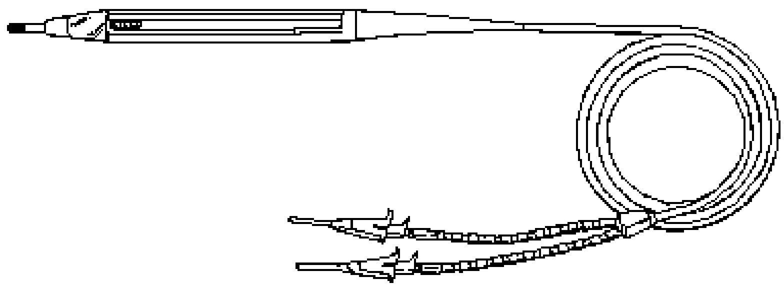


Figure 1-7.—Current tracer.

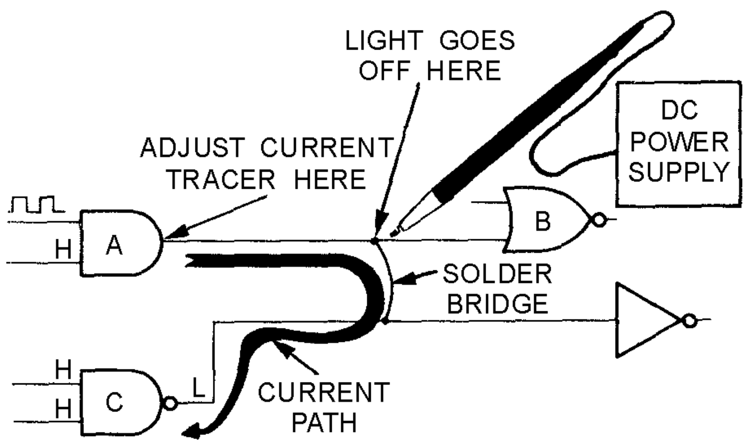


Figure 1-8.—Current tracer application.

CURRENT PROBES

Current probes, as shown in figure 1-9, are primarily designed to be used with an oscilloscope or milliammeter for measuring current. Although not used very often by Navy technicians, current probes are available. The primary advantage in using a current probe is that it does not need to be in series with the current being measured. Unsoldering wires or connections to terminals is not necessary; current probes are designed to be clamped onto insulated conductors. They are able to sense, through inductive action, the magnitude of the current flowing in the conductor. Current probes are designed for performing small ac current measurements. Also, when you use them in conjunction with current probe amplifiers, the capabilities of the current probe are extended to measurement of both ac and dc currents with large magnitudes. Current probes are extremely useful when you measure the current drain on a power supply, start-up current of a motor, or current flow in relays. These probes can be divided into three basic types: passive, active, and Hall effect. Each type has advantages and disadvantages peculiar to its method of operation. Prior to using a current probe, you should thoroughly understand its instructions.

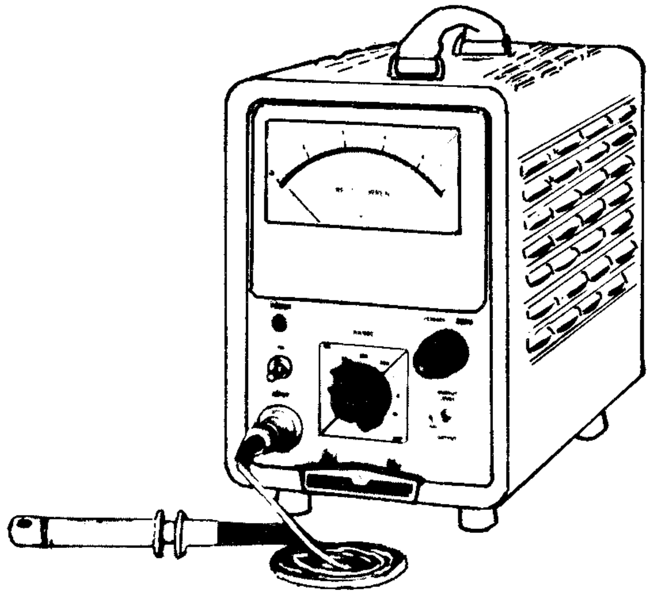


Figure 1-9.—Current probe used with electronic ammeter.

Q-9. What is the primary advantage of using a current probe?

OSCILLOSCOPE METHOD

Current can be measured with an oscilloscope by shunting the input terminals with a low-value resistor. The input terminals must then be connected in series with the circuit being tested. The value of the resistor must be small enough not to interfere with the operation of the circuit under test. At the same time, it must be large enough that the voltage developed will cause adequate deflection of the oscilloscope trace. For example, if an oscilloscope with a vertical deflection sensitivity of 0.1volt rms per centimeter (cm) is used in conjunction with a 10-ohm shunt resistor to measure a 25-milliamp current, the vertical trace will be deflected 2.5 centimeters, as shown in the following example:

Where:

$$I = 25 \text{ mA}$$

$$R = 10 \text{ ohms}$$

$$\text{Sensitivity} = 0.1 \text{ volt/cm}$$

First we'll figure the applied voltage:

$$E = IR$$

$$E = 0.025 \text{ A} \times 10 \text{ ohms} = 0.25 \text{ volts}$$

Now let's figure the deflection in cm:

$$\text{Deflection} = \frac{\text{voltage applied}}{\text{sencitivity}}$$

$$\text{Deflection} = \frac{0.25 \text{ volt}}{0.1 \text{ volt/cm}} = 2.5 \text{ cm}$$

For current measurements the oscilloscope can be calibrated by connecting an ammeter in series with the input terminals and the calibration signal source. An alternate method is to determine the value of the shunt resistor and measure the calibration signal voltage developed across it with an accurate voltmeter. The calibration signal current can then be calculated by means of Ohm's law. Since the oscilloscope merely indicates the voltage developed across the shunt resistor, the measurements for alternating or direct current will be similar to voltage measurements using an oscilloscope.

RESISTANCE MEASUREMENTS

A high percentage of technical manuals contain point-to-point resistance charts that list correct resistance readings for major test points. These resistance charts are extremely useful when you troubleshoot faulty equipment. Without them, equipment resistance measurements within a complicated circuit would not mean much. Many circuits contain other circuit elements, such as capacitors, coils, or other resistors in parallel with the resistances being measured. This, of course, is a possible source of measurement error that you eliminate when you disconnect or unsolder one side of the resistor or a group of resistors under test.

You should be thoroughly familiar with the calibration of your ohmmeter. Analog meters are typically more accurate and easier to read at midscale. With the exception of bridge circuits, a meter may provide only approximate resistance readings. However, these readings may be adequate when you also consider the wide tolerances of resistors themselves. An ohmmeter that you use in field testing should be portable, convenient, and simple to operate - factors that usually are more important than extreme accuracy.

When an ohmmeter is used, completely de-energize the circuit under test and remove any current-sensitive elements before the resistance measurement is performed. Low-resistance measurements that require precision readings should be taken with a bridge type of instrument.

An ohmmeter consists of a galvanometer, batteries, and resistors of known value that are connected in such a way that unknown resistors to be measured are compared with standard values. Figure 1-10 illustrates three basic ohmmeter circuits: (A) single range type, (B) series multirange type, and (C) shunt type.

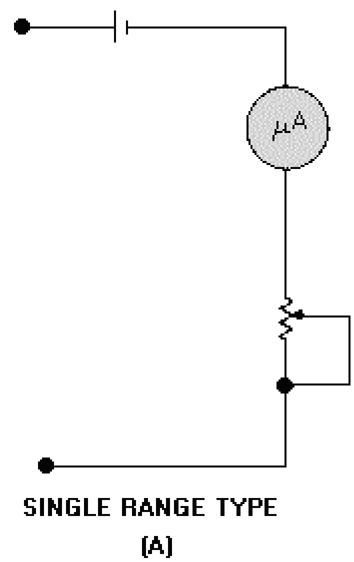


Figure 1-10A.—Basic ohmmeter circuits.

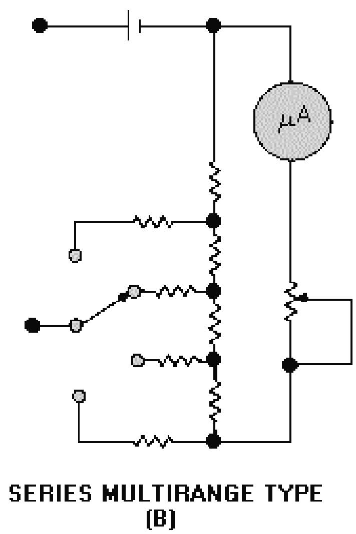


Figure 1-10B.—Basic ohmmeter circuits.

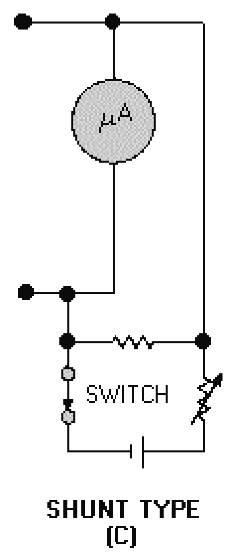


Figure 1-10C.—Basic ohmmeter circuits.

MULTIMETER METHOD

When you use an analog multimeter to perform resistance measurements, the first thing you do is zero the meter. The meter indication varies greatly depending on the resistance of the test leads, the condition of the batteries within the meter, and the resistance range selected on the multimeter. The meter should be zeroed every time you change range settings. To zero a multimeter you short the leads together and adjust the meter for a full-scale deflection. Scale markings are spaced closer together toward the infinity point on the meter; therefore, more accurate readings are obtained near center scale. You should select a range setting that will give you a mid-scale indication.

Ohmmeter applications include resistance measurements; continuity checks; and inductor, capacitor, and transformer checks. A transformer, for example, may be tested by checking whether there is an open or short, low-insulation resistance to ground, or improper continuity between transformer windings. A capacitor may be tested to determine whether it is open or shorted. Ensure that capacitors are properly discharged before you test them; otherwise, damage to the multimeter may occur. When an ohmmeter is placed in series with a capacitor, the changing current will cause a meter deflection that is proportional to the capacitance. The deflection obtained is compared with the deflection from a similar capacitor of known value. This deflection may be small or large, depending on the type and size of the capacitor and the voltage of the battery within the meter. An external series battery will increase the sensitivity of the instrument.

Q-10. How do you compensate for the resistance of the test leads of a meter?

DIGITAL MULTIMETER

The two major advantages of using a digital multimeter are its ease of operation and accuracy. Most digital multimeters can be ordered with an optional battery pack, which makes them just as portable as an analog multimeter. Another advantage is that their LED or LCD readouts are much easier to read than the scale on an analog multimeter. Digital multimeters also are ideally suited for measuring sensitive devices that might otherwise be damaged by the excessive current associated with analog multimeters — maximum current flow through the component being tested is typically limited to less than 1 milliamp. When measuring small values of resistances, remember to consider the resistance of your test leads. Most digital multimeters cannot be zeroed in the way analog multimeters can. With digital multimeters, you have to short the leads, read the lead resistance displayed, and then subtract the reading from subsequent component measurements that you make.

Q-11. Why are digital multimeters well suited for testing sensitive devices?

RCL BRIDGES

The 250DE+1325 is a typical resistance, capacitance, inductance (rcl) bridge. Like the vtvm, the rcl bridge has several disadvantages. It requires ac power and a lengthy warm-up period, and its accuracy is limited to ±2%. The rcl bridge uses a tuning indicator electron tube, commonly referred to as the bridge's "eye," and an internal arrangement of resistors that form a Wheatstone bridge. As discussed in NEETS, module 16, the rcl bridge can be a time-consuming method of performing resistance measurements.

Difficulty may be experienced when you attempt to measure wire-wound resistors. To obtain a sharp balance on the indicator, you can shunt the resistor with a variable capacitor and adjust the capacitor for the clearest indication. The resistance measurement will not be affected by this reactance neutralization.

MEGGERS

Meggers produce the large voltages that are required to measure resistances as high as 10,000 megohms — only high resistance values can be measured. The unknown resistance is connected between the megger terminals, and the hand generator part of the meter is cranked. Some meggers are capable of producing in excess of 500 volts, so use caution when you operate them. Typical applications for a megger are testing unterminated transmission lines and ac power cords for insulation breakdown.

DIFFERENTIAL VOLTMETERS

It is a seldom-known fact that the Fluke 893 ac-dc differential voltmeter can be used for measuring extremely high resistances from 10 megohms to 10^6 megohms with a typical accuracy of ±5%. This measurement method, however, requires some basic calculations on your part. The obvious advantage of the differential voltmeter is its capability of measuring extremely high resistances. Consult the Fluke 893 technical manual for initial switch settings and a more detailed explanation of its operation.

CAPACITOR MEASUREMENTS

Capacitance is that property of a circuit that produces an electrostatic field when two conducting bodies separated by a dielectric material have a potential applied to them. Capacitors are made by compressing an insulating material (dielectric) between two conductors (plates). The farad is the basic measurement of capacitance. It is dependent upon the area of the plates, the distance between the plates, and the type of dielectric used. Electrically, the farad is a measure of 1 coulomb of potential charged by 1 volt. A coulomb (the amount of current flow maintained at 1 ampere that passes a given point of a circuit in 1 second) is a large charge. Most capacitors are measured in millionths of a farad (microfarad), expressed as ∝F, or in one-millionth of a microfarad (picofarad), expressed as pF.

Capacitors incur various losses as a result of such factors as resistance in the conductors (plates) or leads, current leakage, and dielectric absorption, all of which affect the power factor of the capacitor. Theoretically, the power factor of an ideal capacitor should be zero; however, the losses listed above cause the power factors of practical capacitors to range from near 0 to a possible 100%. The average power factor for good capacitors, excluding electrolytics, is 2% to 3%. Current leakage, which is an inverse function of frequency, is important only at the lower frequencies and becomes negligible at higher frequencies. Dielectric absorption (sometimes referred to as dielectric viscosity) results in losses that produce heat. The effect of this type of loss is the same as resistance in series with the capacitor.

You have probably learned the hard way that some capacitors can retain a charge long after the voltage has been removed. The electrical charge retained by capacitors in de-energized electronic circuits is, in many cases, sufficient to cause a lethal shock. Be sure you and those working with you consider this hazard before performing any type of maintenance on any electrical or electronic circuit and before making connections to a seemingly dead circuit. Use extreme caution prior to working on or near de-energized circuits that employ large capacitors. Be safe—discharge and ground all high-voltage capacitors and exposed high-voltage terminal leads by using only an authorized shorting probe, as shown in figure 1-11. Repeat discharge operations several times to make sure that all high-voltage terminations are completely discharged. It is of the utmost importance that you use only an *authorized* safety shorting probe to discharge the circuits before performing any work on them. An authorized *general-purpose* safety shorting probe for naval service application may be requisitioned using the current stock number listed in the ELECTRONICS INSTALLATION AND MAINTENANCE BOOK (EIBM), *General* NAVSEA 0967-LP-000-0100, Section 3, Safety Equipment. Certain electronic equipment are provided with built-in, special-purpose safety shorting probes. These probes are *not* considered general purpose. Use them only with the equipment for which they are provided and only in a manner specified by the technical manuals for the equipment. It is considered to be poor practice to remove them for use elsewhere.

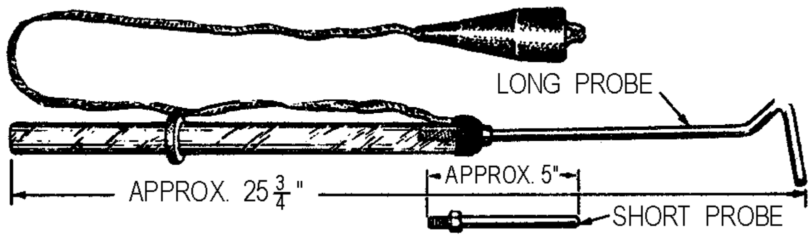


Figure 1-11.—Safety shorting probe.

When using the general-purpose safety shorting probe, always be sure first to connect the grounding clip to a good ground connection (if necessary, scrape the paint off the grounding metal to make a good contact — paint can be replaced, lives can't). Then, while holding the safety shorting probe by the handle behind the protective shield, touch the end of the metal rod to the points to be discharged. Touch each point several times to ensure that the circuit is completely discharged. Be extremely careful that *you* do not touch any of the metal parts of the safety shorting probe while touching the probe to the exposed "hot" terminal. Don't develop a nonchalant or routine attitude about these procedures. It pays to be safe; use the safety shorting probe with care.

Large capacitors, dormant in storage, can also develop a large static charge. This charge is caused by environmental conditions such as a close proximity to an rf field. An easy way to avoid this condition is to short the stored capacitor's terminals with a piece of wire before putting it in storage. Remember to remove the wire before installing the capacitor. If you receive a large capacitor that is not shorted, short the terminals together. Remember, CHARGED CAPACITORS CAN KILL.

Q-12. Charged capacitors can kill. True or false?

BRIDGE-TYPE MEASUREMENTS

Capacitor tests involving quality and value must be made in the course of everyday troubleshooting. You must make the important decision of whether to reject or continue to use a certain capacitor after it has been tested. Capacitance measurements are usually accomplished by either a bridge-type or a reactance-type capacitance meter. The bridge-type capacitance meter is much more accurate than the reactance-type meter. You may want to review rcl bridges in chapter 1 of NEETS, module 16, before reading further. Capacitance tolerances vary more widely than resistance tolerances and are dependent upon the type of capacitor, the capacitance value, and the voltage rating. The results of capacitance tests must be evaluated to determine whether a particular capacitor will fulfill the requirements of the circuit in which it is used.

The power factor of a capacitor is important because it is an indication of the various losses attributable to the dielectric, such as current leakage and dielectric absorption. Current leakage is of considerable importance, especially in electrolytic capacitors.

Figure 1-12 is a simplified schematic of a capacitance bridge. As you can see, a capacitance bridge is very similar in construction to a resistance bridge with the exception of the standard capacitor (C_S) and the unknown capacitor (C_X). Because current varies inversely with resistance and directly with capacitance, an inverse proportion exists between the four arms of the bridge. The following expression shows the inverse proportion between resistors A and B and capacitors C_S and C_X:

$$\frac{A}{B} = \frac{C_X}{C_S}$$

Solving for C_X:

$$C_X = \frac{AC_S}{B}$$

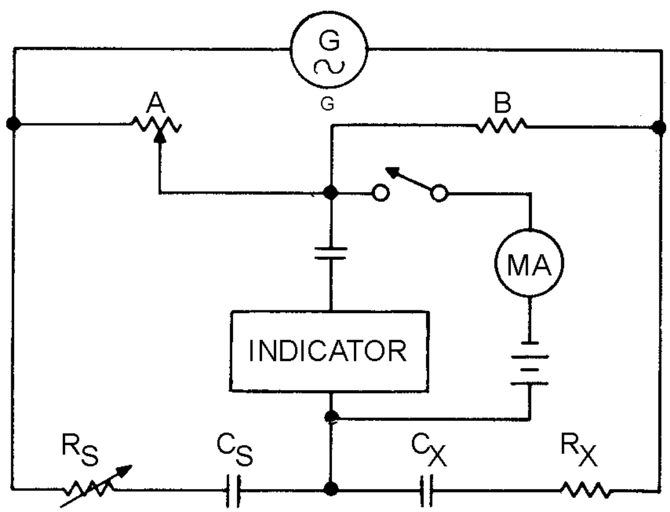


Figure 1-12.—Simplified capacitance bridge.

It is actually the capacitive reactance, rather than the capacitance, that is balanced in this circuit.

In addition to its reactive properties, the capacitor under test always exhibits some loss. This loss may have the characteristics of either a shunt or series resistance, or it may be a combination of both. Regardless of its true nature, the loss can always be represented as a simple series resistance, which is shown in figure 1-12 as R_X. This loss is balanced by the calibrated resistor R_S. Rather than calibrate this control in terms of resistance, it is convenient to calibrate it in terms of the dissipation factor (the ratio of the energy dissipated to the energy stored in a capacitor). The R_S control then provides the means for completing the capacitance balance, and its dial reading indicates a loss figure for the capacitor under test.

Q-13. Which is more accurate, the bridge- or reactance-type meter?

REACTANCE-TYPE MEASUREMENTS

The reactance type of capacitance measuring equipment makes use of the following principle: If an ac voltage (usually 6.3 volts) at a fixed frequency is applied across a capacitor and resistor in series, the voltage drop produced across the reactance of the capacitor by the resulting current flow is inversely proportional to the capacitance. The voltage drop is used to actuate a meter that is calibrated in capacitance values. This test equipment gives approximate values only and, like the ohmmeter, is used mostly when portability and speed are more important than precision. The accuracy of the reactance-type measurement is less for capacitors that have a high power factor. In capacitors with high power factors, the losses incurred effectively place a certain amount of resistance in series with the capacitive reactance. The effect of this resistance, when the capacitor is measured, is to cause a greater voltage drop across the capacitor. This drop is not because of the reactance above, but is the result of the impedance, which of course is made up of both the reactance and the resistance. Therefore, the capacitance indicated by the analyzer will be lower than the actual value.

Figure 1-13 shows a simplified schematic diagram of the capacitance-measuring section of a typical reactance-type electronic volt-ohm-capacitance milliammeter. A 6.3-vac voltage is taken from the filament source and applied across the resistive voltage divider network to determine the designated value of the capacitor. Because of a particular use or circuit application, some capacitors are permitted an even wider variation of capacitance value than is indicated by their rated tolerances.

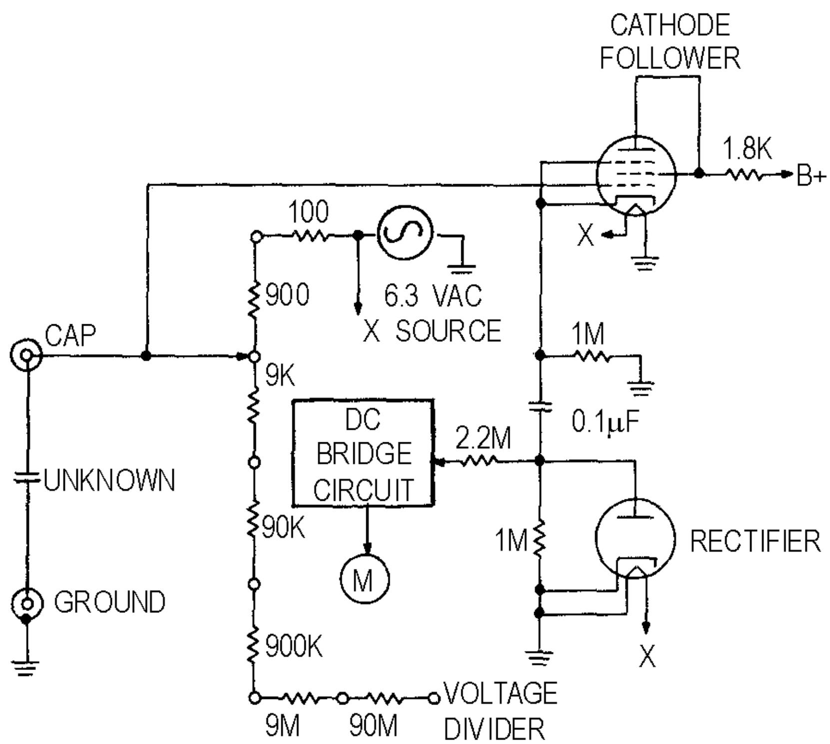


Figure 1-13.—Reactance-type capacitance meter.

INDUCTANCE MEASUREMENT

A current flowing through a conductor produces a magnetic field around that conductor. If the conductor is formed into a coil, a stronger magnetic field is set up. The relationship between the strength of the field and the intensity of the current causing it is expressed by the inductance of the coil (or conductor). When the current producing the magnetic field ceases, the energy of the magnetic field is returned in part to the circuit source in the form of a reverse current. Inductance, then, is the ability of a coil to function as a storehouse of energy in magnetic form and is determined by the shape and dimensions of the coil. Inductance is measured in henries, millihenries, or microhenries. Inductors can be described generally as circuit elements used to introduce inductive reactance into ac circuits.

An inductor is essentially a coil of wire wound around a form using a core of air, magnetic metal, or nonmagnetic metal. A core of magnetic metal produces greater inductance (for a coil of given size and number of turns) than does an air core; a core of nonmagnetic metal produces less inductance than does an air core. At frequencies in the hf and higher regions of the frequency spectrum, coils of small size and high *Q* (discussed briefly at the end of this section) are generally required. These coils usually are single-layered with air or metallic cores. Since comparatively low values of inductance are required, this type of coil is very compact, and relatively high values of *Q* are obtained.

At frequencies in the lf and mf regions of the frequency spectrum, single-layered, universal, spiral, and other types of windings are used. When size is a factor, the more compact windings are preferred to the single-layered type of coil. At frequencies below 500 kilohertz, the single-layered type is too large for practical use; therefore, the more compact types are used exclusively.

The inherent resistance of the conductor with which an inductor is wound is the most important factor contributing to the losses of the inductor. Losses caused by this resistance increase with frequency. This results in a concentration of current near the outer surface of the wire, called SKIN EFFECT. Skin effect is negligible at low frequencies, but can be an important factor at high frequencies. Other contributing factors to inductor losses are (1) eddy currents set up in the core and surrounding objects (if they are conductors); (2) the dielectric properties of the form used for the coil and surrounding objects; and (3) hysteresis in the core and surrounding objects, if they are magnetic metals. Losses occur as a result of the dielectric properties of the coil form because of the distributed capacitance of the inductor (for example, between turns and between the terminals and leads). To some extent the core and surrounding objects serve as a dielectric of the distributed capacitance, and the resulting dielectric losses contribute to the overall losses of the inductor.

As we discussed earlier, an inductor has the ability to act as a storehouse of magnetic energy. However, because of the various loss factors described above, all of the energy stored in the magnetic field is not returned to the source when the applied voltage decreases to zero. The losses of an inductor may be represented by an equivalent series resistance. The value that it would dissipate would be an amount of energy equal to the total amount dissipated by the inductor. The losses of an inductor may be expressed in terms of the ratio of its inductive reactance to its equivalent series resistance. This ratio is referred to as the Q of the inductor and is stated in equation form as shown below:

$$Q = \frac{X_L}{R}$$

Q-14. What type of core produces the greatest inductance?

HAY BRIDGE

Inductance measurements are seldom required in the course of troubleshooting. However, in some cases inductance measurements are useful and instruments are available for making this test. Many capacitance test sets can be used to measure inductance. Most manufacturers of capacitance test sets furnish inductance conversion charts if the test equipment scale is not calibrated to read the value of inductance directly. For the measurement of inductance, the following basic types of test equipment circuitry are used: (1) the bridge-circuit type, which is the most accurate, and (2) the reactance type, which is often an additional test circuit incorporated into another piece of test equipment to increase its utility. The measurement of capacitance using the capacitance-inductance-resistance bridge instrument was discussed. Since the measurement of capacitance and inductance are interrelated, the existing capacitance standards and loss controls of this test equipment are used whenever possible. A wider range of dissipation must be provided to accommodate the practical value of inductors. The 250DE+1325 (view A of fig. 1-14), a typical rcl bridge and our reference in this discussion, uses two basic bridge circuits (Hay bridge and Maxwell bridge) to accommodate the extensive range in inductor loss factors. You should take time to review the bridges in NEETS, module 16, or other bridge-circuit descriptions before continuing.

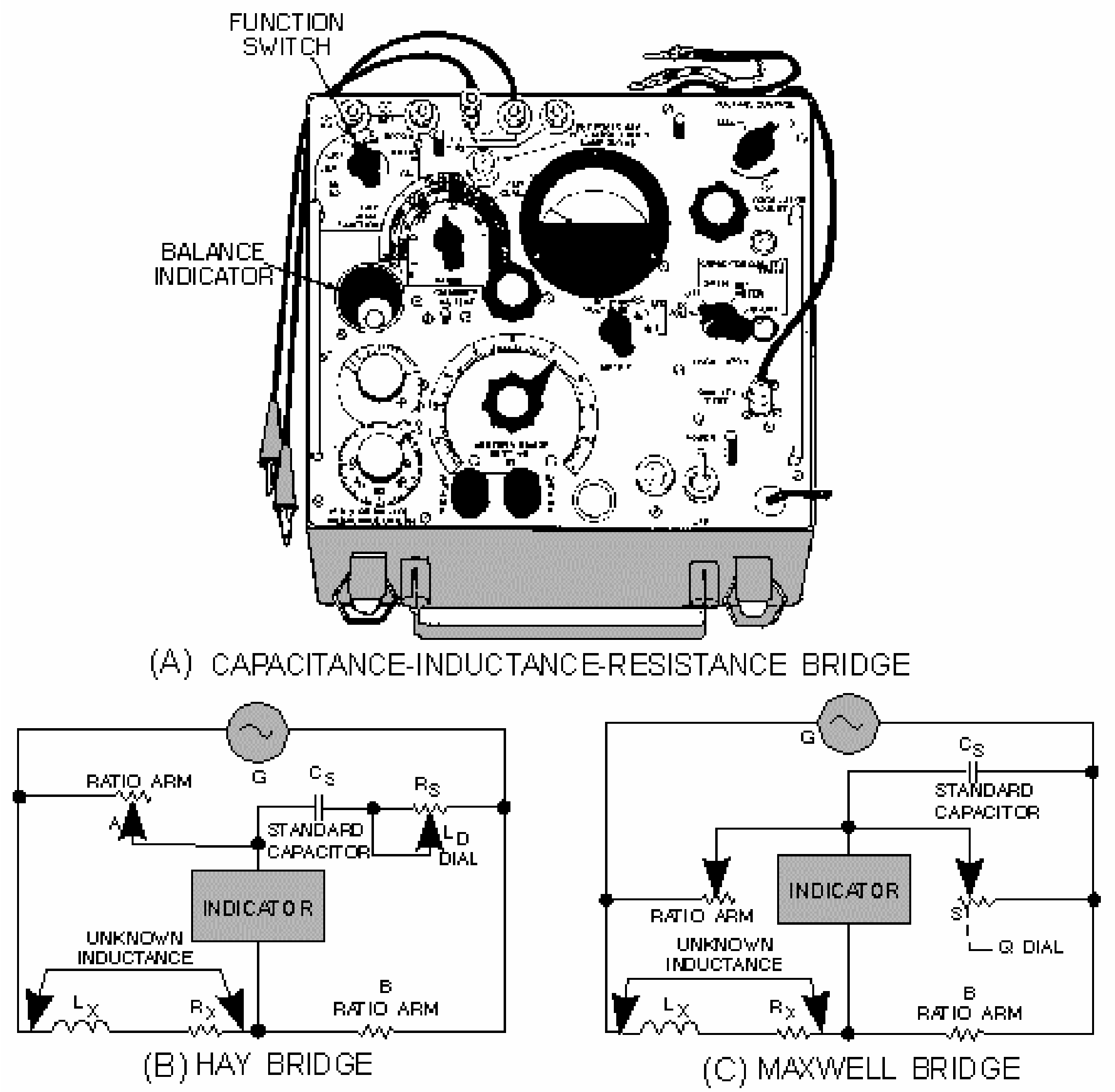


Figure 1-14.—Bridge circuits.

The Hay bridge (view B of fig. 1-14) measures inductance by comparing it with a capacitance; it differs from the Maxwell bridge (view C) in that the resistance associated with the capacitance is a series instead of a shunt resistance. The inductance balance depends upon the losses (*Q*) of the inductor. The Hay bridge is used for inductors with low losses low D dial reading or high *Q*) at 1 kilohertz. This circuit is in effect when the FUNCTION switch is turned to the L(D) position. For a D dial reading up to 0.05, the error is 0.25%. Above this point the error increases rapidly and affects the basic accuracy of the test equipment. This limitation is expressed on the front panel of the test equipment as follows: IF D>0.05 ON L(D)—REBALANCE ON L(Q) . In other words, if the dissipation of an inductor, as read on the D dial when using the Hay bridge (FUNCTION switch set to L(D) position), exceeds 0.05, then you should change to the Maxwell bridge (FUNCTION switch set to L(Q) position), which is discussed in the following paragraph. The loss factor of the inductor under test is then balanced in terms of the *Q* of the inductor.

Q-15. A Hay bridge measures inductance by comparing an inductor to what component?

MAXWELL BRIDGE

The Maxwell bridge, shown in view C of figure 1-14, measures inductance by comparing it with a capacitance and (effectively) two resistances.] This bridge circuit is employed for measuring inductances having losses greater than 0.05 (expressed by the D dial reading). For such inductors it is necessary to introduce, in place of the series control (D dial), a new loss control (Q dial), which shunts the standard capacitor. This control, which becomes effective when the FUNCTION switch is turned to the L(Q) position, is conveniently calibrated in values of Q, the storage factor of the inductor under measurement. The balance for inductance is the same for either bridge circuit. This permits the use of the same markings on the RANGE switch for both the L(D) and L(Q) positions of the FUNCTION switch.

REACTANCE MEASURING EQUIPMENT

The reactance type of inductance measuring equipment makes use of the following principle: If an ac voltage of fixed frequency is applied across an inductor (and a resistor in series), the voltage drop produced across the reactance of the inductor by the resulting current flow is directly proportional to the value of the inductance. An inductance measurement using the reactance method is identical to capacitance measurements using the same method, except that current flow is directly proportional to the value of inductance, rather than inversely proportional as in the case of capacitance. It follows then that if a reactance-type capacitance measuring equipment is provided with a chart that converts the capacitance readings to equivalent inductance values and a proper range multiplying factor, the same test setup can be used to measure both capacitance and inductance. In practice, test equipment using the reactance method for capacitance measurements usually provides an inductance conversion chart. Because the current flowing through the inductance under test is directly proportional to the value of inductance, the reciprocals of the capacitance range multipliers must be used; for example, a multiplier of 0.1 becomes

$$\frac{1}{0.1} \text{ or } 10$$

and a multiplier of 100 becomes

$$\frac{1}{100} \text{ or } 0.01$$

The reactance-type equipment gives approximate values only. Like the analog multimeter, it is used only when portability and speed are more important than precision. If the ohmic resistance of the inductor is low, the inductance value obtained from the conversion chart can be used directly. If the ohmic value (as measured with an ohmmeter) is appreciable, a more accurate value of inductance can be obtained by use of the following formula:

$$L = \frac{(Z_L)^2 - (R_L)^2}{2\pi f}$$

Where:

L = the inductance

Z_L = the impedance of the inductance under test

f = the frequency

R_L = the ohmic resistance

Q-16. Is the current flow through an inductor directly proportional or inversely proportional to its inductance value?

MEASUREMENT OF INDUCTANCE USING THE VTVM

If you do not have a 250DE+1325 at your disposal, the inductance of a coil can be determined by using a vtvm and a decade resistance box, as shown in figure 1-15. In the following example the inductance of an unknown coil in the secondary winding of a 6.3-volt filament transformer will be determined with a vtvm and decade resistance box. The unknown coil must be connected in series with the decade resistance box. The voltage across the decade box and across the coil must be monitored as the decade box is adjusted. When equal voltages are reached, read the resistance of the decade box. Since the voltage across the inductor equals the voltage across the decade box, the X_L of the coil must be equal to the resistance read on the decade box. For example, assume that the resistance reading on the decade box is 4 kilohms and the frequency is 60 hertz. This must mean that the X_L of the coil is also equal to 4,000 ohms. The inductance formula $L = X_L/2\pi f$ can be used to find the inductance of the coil in henries:

$$L = \frac{4000\Omega}{(6.28)(60)}$$

$$L = \frac{4000\Omega}{376.8}$$

$$L = 10.62H$$

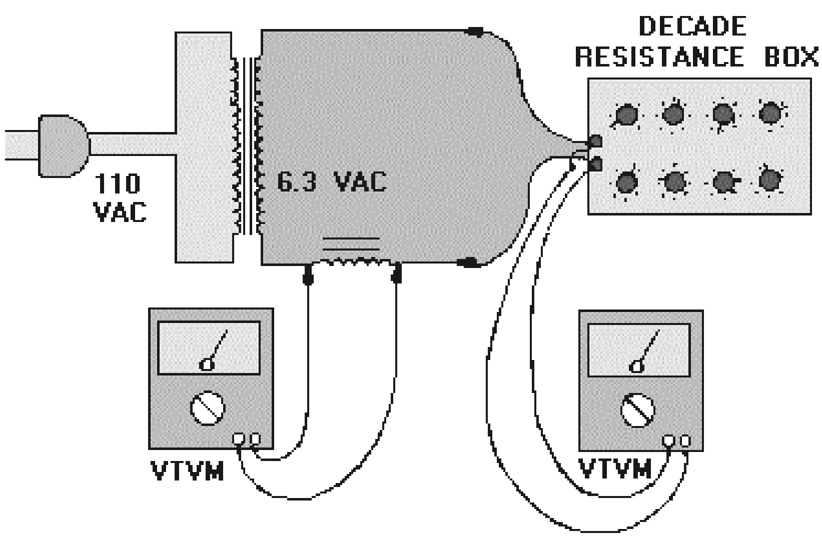


Figure 1-15.—Determining inductance with a vtvm and decade resistance box.

SUMMARY

This chapter has presented information on basic measurements. The information that follows summarizes the important points of this chapter.

The five basic measurements are **VOLTAGE, CURRENT, RESISTANCE, CAPACITANCE**, and **INDUCTANCE**. The accuracy of all measurements depends upon YOUR SKILL as a technician and the accuracy of your TEST EQUIPMENT.

Accuracy of different types of test equipment varies greatly and depends on design characteristics, tolerances of individual components, and YOUR KNOWLEDGE of test equipment applications.

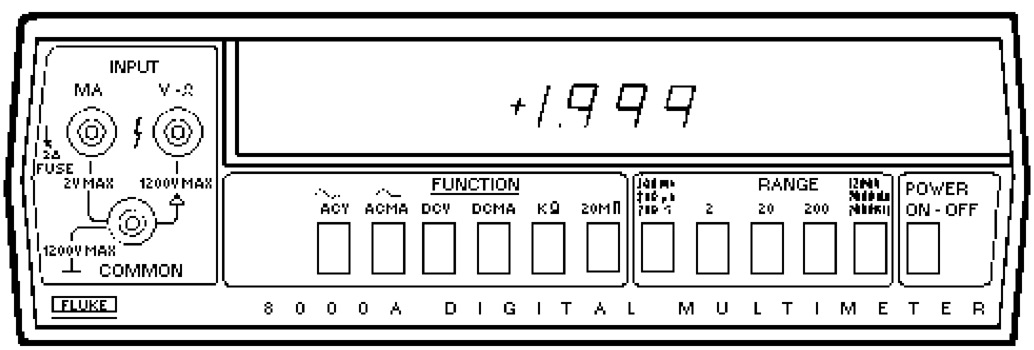
The **METCAL** program ensures that your calibrated test equipment meets established specifications.

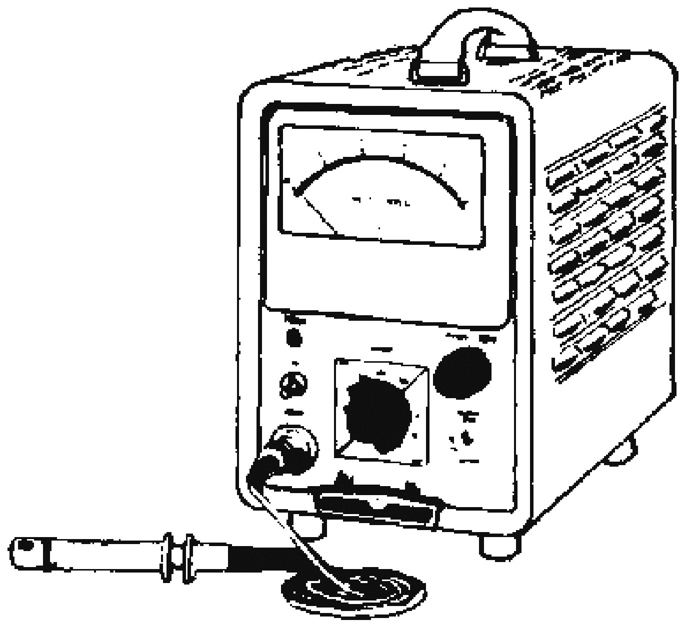
Most equipment technical manuals contain VOLTAGE CHARTS which list correct voltages that should be obtained at various test points.

It is important to remember that the **INPUT IMPEDANCE** of your test equipment must be high enough to prevent circuit loading.

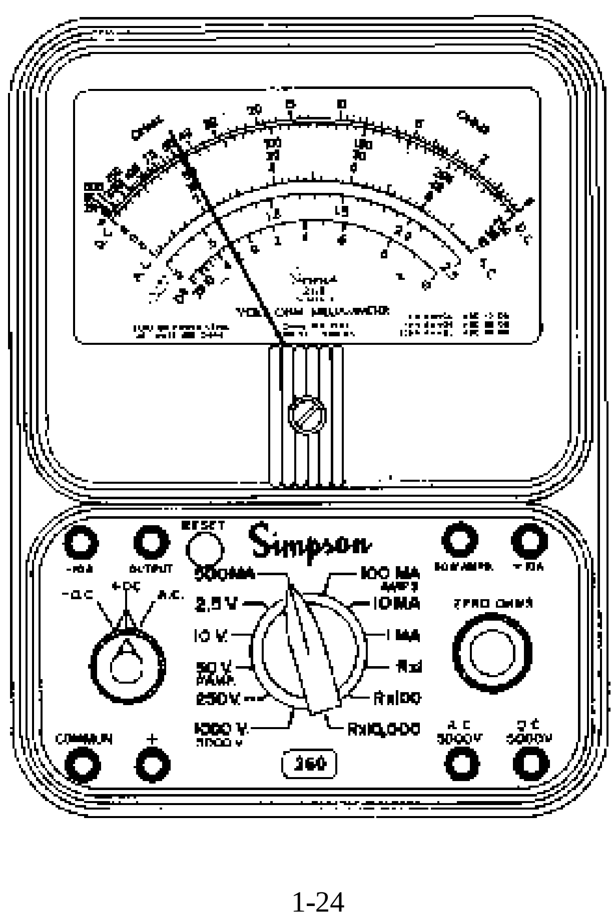
When you are performing ac voltage measurements, an additional consideration that greatly affects the accuracy of your measurements is the **FREQUENCY LIMITATIONS** of your test equipment.

Ac and dc CURRENT MEASUREMENTS can be performed using a wide variety of test equipment. Most current measurements require you to break the current path by unsoldering components and wires and inserting an ammeter in series with the current path. One alternative method is to compute (using OHM'S LAW) the current through a circuit by measuring the voltage drop across a known resistance. Another alternative is to use a CURRENT PROBE that requires no unsoldering.



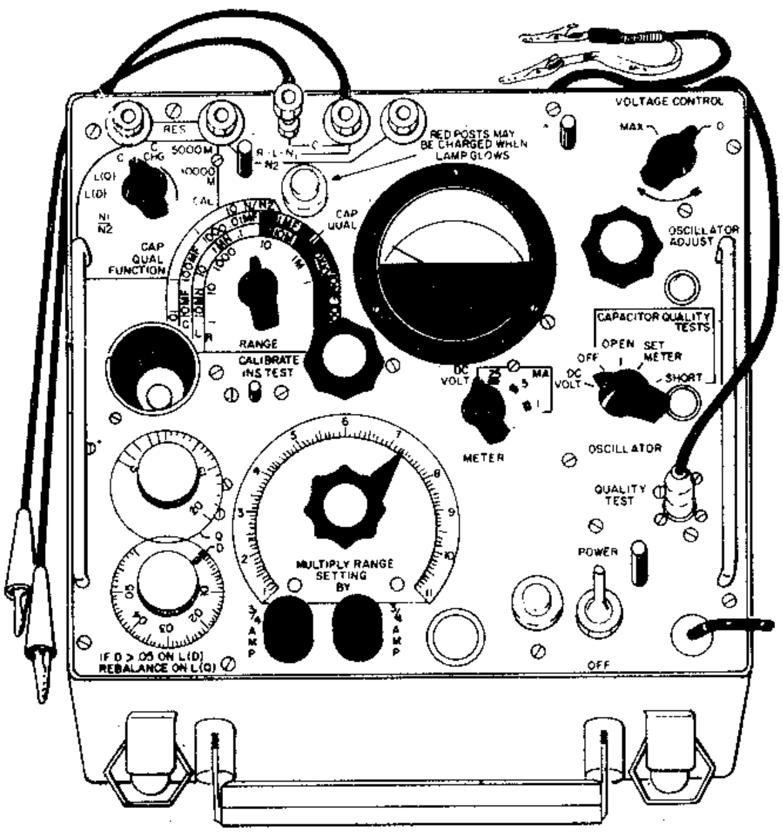


When performing resistance measurements, your primary concerns are the **RANGE AND DEGREE OF ACCURACY** of your test equipment. In most instances, an analog multimeter is accurate enough to perform basic troubleshooting. When measuring extremely large resistances, you are sometimes required to use a MEGGER or a DIFFERENTIAL VOLTMETER.



When testing current-sensitive devices, you must be certain that the current produced by your test equipment does not exceed the current limitations of the device being tested.

Capacitance and inductance measurements are seldom required in the course of troubleshooting. These measurements are usually performed with various types of BRIDGES or with a reactance type of measuring device. The bridge -measuring techniques are more commonly used and are more accurate than reactance types of measurements.



REFERENCES

8000A Digital Multimeter, *NAVSEA 0969-LP-279-9010, Naval Sea Systems Command, Washington, D.C., undated.*

EIMB, Test Methods and Practices Handbook, *NAVSEA 0967-LP-000-0130, Naval Sea Systems Command, Washington, D.C., 1980.*

EIBM, General, *NAVSEA 0967-000-0100, Naval Sea Systems Command, Washington, D.C., 1983.*

Instruction Manual for Universal Impedance Bridge, Model 250DE, *13202, Electro Scientific Industries, 13900 N. W. Science Park Drive, Portland, Oregon 97229, March 1971.*

Instruction Manual, Model 893A/AR AC-DC Differential Voltmeter, *NAVSEA 0969-LP-279-7010, Naval Sea Systems Command, Washington, D.C., 1969.*

Operation and Maintenance Instruction, *Current Tracer 547A, NAVAIR 16-45-3103, Naval Air Systems Command, Washington, D.C., 1979.*

Operation and Maintenance Instructions, Volt-Ohm-Milliammeter, *260 Series 6P, NAVSEA 0969-LP-286-1010, Naval Sea Systems Command, Washington, D.C., 1974.*

ANSWERS TO QUESTIONS Q1. THROUGH Q16.

A-1. *Its calibration.*

A-2. *10 to 1*

A-3. *Increased input impedance, greater accuracy, and increased voltage range.*

A-4. *Midscale.*

A-5. *Accuracy and high input impedance.*

A-6. *The range of frequencies that can accurately be measured.*

A-7. *At least 60% of the vertical trace.*

A-8. *Decreased internal meter resistance, greater accuracy, and greater current range.*

A-9. *Current probes enable you to perform current measurements without disconnecting wires. Current probes are clamped around the insulated wire.*

A-10. *By zeroing the meter with the test leads shorted.*

A-11. *The current flow through the component is limited to 1 milliamp.*

A-12. *True.*

A-13. *Bridge type.*

A-14. *Magnetic-metal core.*

A-15. *A capacitor.*

A-16. *Directly proportional.*

CHAPTER 2

COMPONENT TESTING

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to do the following:

1. Explain the importance of testing individual electronic components.
2. Identify the various methods of testing electron tubes.
3. Identify the various methods of testing semiconductors.
4. Identify the various methods of testing integrated circuits.
5. Identify the various types of testing batteries and their characteristics.
6. Identify the various methods of testing rf attenuators and resistive loads.
7. Identify the various methods of testing fiber-optic devices.

INTRODUCTION TO COMPONENT TESTING

It is imperative that you be able to troubleshoot an equipment failure to the component level. In the majority of cases, Navy technicians are expected to troubleshoot and identify faulty components. This chapter, "Component Testing," will acquaint you with alternative methods of testing various components and their parameters. A quick glance at the Navy's mission and concept of operation explains why we, in most cases, must be able to troubleshoot to the faulty component level. A ship must be a self-sustaining unit when deployed. Storage space is a primary consideration on most ships and a limiting factor for storage of bulky items or electronic modules as ready spares. Therefore, it is practical to store only individual components common to a great number of equipment types. This of course, limits the larger replacement modules available to you during troubleshooting.

Q-1. Why are most ships limited in their ability to stock replacement modules for repair of electronic equipment?

TESTING ELECTRON TUBES

In equipment that uses vacuum tubes, faulty tubes are responsible for more than 50% of all electronic equipment failures. As a result, testing of electronic tubes is important to you. You can determine the condition of a tube by substituting an identical tube known to be good for the questionable one. However, indiscriminate substitution of tubes is to be avoided for at least the following two reasons: (1) detuning of circuits may result and (2) a tube may not operate properly in a high-frequency circuit even though it performs well in a low-frequency circuit. Therefore, your knowledge of tube-testing devices and their limitations, as well as correct interpretation of the test results obtained, is indispensable for accurate and rapid maintenance.

Because the operating capabilities and design features of a tube are demonstrated by its electrical characteristics, a tube is tested by measuring those characteristics and comparing them with representative values established for that type of tube. Tubes that read abnormally high or low with respect to the standard are suspect. Practical considerations, which take into account the limitations of the tube test in predicting actual tube performance in a particular circuit, make it unnecessary to use complex and costly test equipment with laboratory accuracy. For most applications, testing of a single tube characteristic is good enough to determine tube performance. Some of the more important factors affecting the life expectancy of an electron tube are listed below:

- The circuit function of the tube
- Deterioration of the cathode coating
- A decrease in emission of impregnated emitters in aging filament-type tubes
- Defective seals that permit air to leak into the envelope and oxidize the emitting surface
- Internal short circuits and open circuits caused by vibration or excessive voltage

If the average receiving tube is not overdriven or operated continuously at maximum rating, it can have a life of at least 2,000 hours before the filament opens. Because of the expansion and contraction of tube elements during the process of heating and cooling, electrodes may lean or sag, which causes excessive noise or microphonics to develop. Other electron-tube defects are cathode-to-heater leakage and nonuniform electron emission of the cathode. These common tube defects contribute to about 50% of all electronic equipment failures. For this reason you should immediately eliminate any tube known to be faulty. However, avoid blind or random replacement of good tubes with fresh spares. The most common cause of tube failure is open filaments. Evidence of a tube defect is often obvious when the filament is open in glass-envelope tubes. You will also notice the brighter-than-normal cherry-red glow of the plate when the plate current is excessive. Also, when the tube becomes gassy or when arcing occurs between electrodes, you will probably have visual indication. Metal-encased tubes can be felt for warmth to determine if the heater is operating. You can tap a tube while it is operating in a circuit to reveal an aural indication of loose elements within the tube or microphonics, which are produced by loose elements.

Most tubes are extremely fragile and subject to damage during shipment. When you replace a tube, never make the assumption that the new tube is good because it’s new. You should always test tubes before installing them.

Q-2. What is the most common cause of electron tube failure?

SUBSTITUTION METHODS

Substituting with a tube known to be in good condition is a simple method of testing a questionable tube. However, in high-frequency circuits tube substitution should be carried out in a logical sequence. Replace tubes one at a time so that you can observe the effect of differences in interelectrode capacitance in the substituted tubes on tuned circuits. The tube substitution test method cannot be used to advantage in locating more than one faulty tube in a single circuit for two reasons: (1) If both an rf amplifier tube and IF amplifier tube are defective in a receiver, replacing either one will not correct the trouble; and (2) if all the tubes are replaced, there is no way for you to know what tubes were defective. Under these conditions, using test equipment designed for testing the quality of a tube saves you valuable time.

Q-3. What is the most accurate method of determining the condition of an electron tube?

NOTE ON SYMBOLS USED IN THE FOLLOWING SECTIONS: IEEE and ANSI standards (see inside front cover) are used to define various terms, such as anode (plate) current, anode voltage, and anode resistance. This book uses E_a for anode voltage, I_a for anode current, and r_a for anode resistance. These are the same as E, I_p, and $_{rp}$ that you will see elsewhere. This module uses the terms anode and plate interchangeably.

ELECTRON TUBE TESTERS

A representative field type of electron tube tester designed to test all common low-power tubes is shown in figure 2-1. The tube test conditions are as close as possible to actual tube operating conditions and are programmed on a prepunched card. The card switch (S101, fig. 2-1) automatically programs the tube test conditions when it is actuated by a card. A card compartment on the front panel of the tester provides storage for the most frequently used cards. The cover of the tester (not shown) contains the operating instructions, the brackets for storing the technical manual, the power cord, the calibration cell for checking the meter and short tests, the calibration cards, the blank cards, and a steel hand punch.

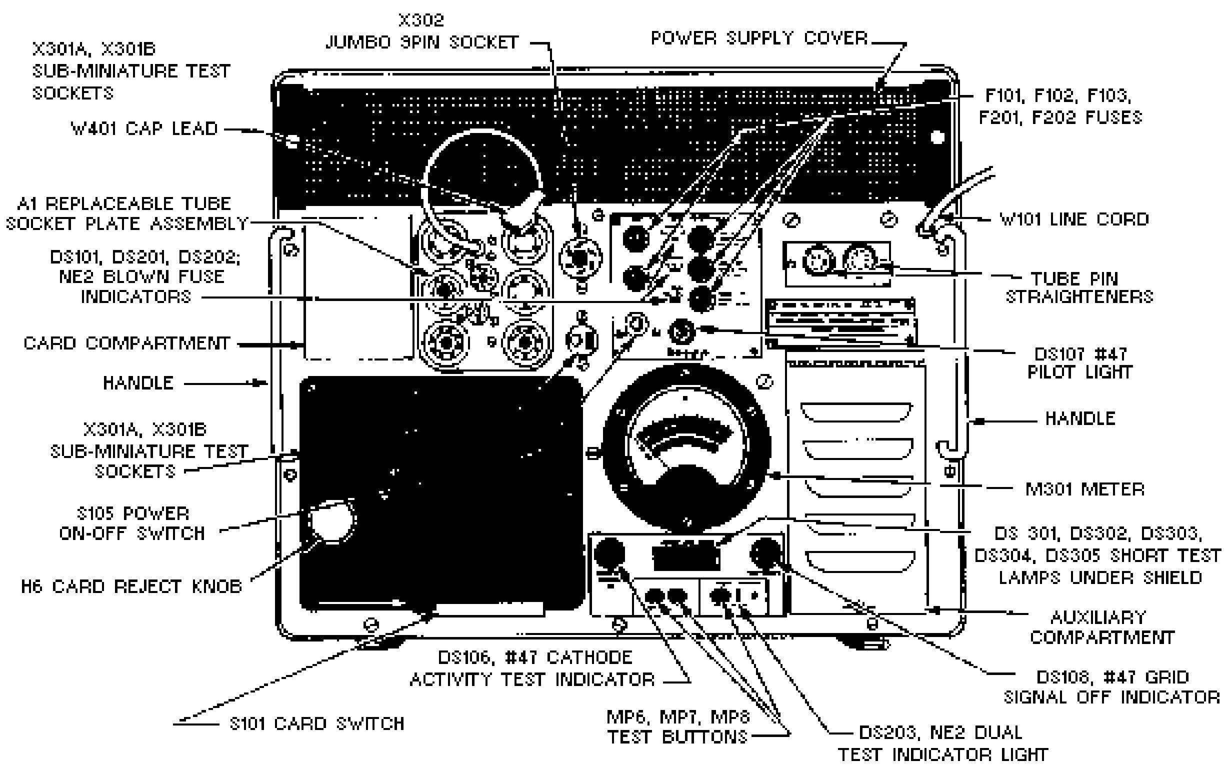


Figure 2-1.—Electron tube tester.

Front Panel

When a prepunched card is fully inserted into the card switch (S101), a microswitch is actuated that energizes a solenoid, causing the card switch contacts to complete the circuit. The card switch has 187 single-pole, single-throw switches arranged in 17 rows with 11 switches in each row. The card is used to push the switches closed; thus, the absence of a hole in the card is required to actuate a switch.

The meter (M301) contains four scales. The upper scale is graduated from 0 to 100 for direct numerical readings. The three lower scales, numbered 1, 2, and 3, are read for LEAKAGE, QUALITY, and GAS, respectively. Each numbered scale includes green and red areas marked GOOD and REPLACE. Inside a shield directly in front of the meter are five neon lamps (DS301 through DS305), which indicate shorts between tube elements.

The number 2 pushbutton (MP6) is used for transconductance, emission, and other quality tests (described later). The number 3 pushbutton (MP7) is used to test for the presence of gas in the tube envelope. The number 4 pushbutton (MP8) is used for tests on dual tubes. A neon lamp (DS203) lights when pushbutton number 4 is to be used. Eleven tube test sockets are located on the panel, plus tube pin straighteners for the 7- and 9-pin miniature tubes.

The power ON-OFF spring-return toggle switch (S105) turns the tester on by energizing a line relay. The pilot light (DS107) lights when this relay closes. Above the power ON-OFF switch are five fuses. Fuses F101, F201, and F202 protect circuits in the tester not protected by other means and have neon lamps to indicate when they have blown. Fuses F102 and F103 protect both sides of the power line.

Auxiliary Compartment

A group of auxiliary controls covered by a hinged panel is used for special tests and for calibration of the tester. Two of these controls, labeled SIGNAL CAL (R152 and R155, fig. 2-2), are used with special test cards for adjusting the regulation and amplitude of the signal voltage. A pushbutton labeled CATH ACT (S302D) is used for making cathode activity tests. When this button is pressed, DS106 on the front panel (fig. 2-1) lights, and the filament voltage of the tube under test is reduced by 10%. Results of the test are read as a change in reading on the numerical meter scale.

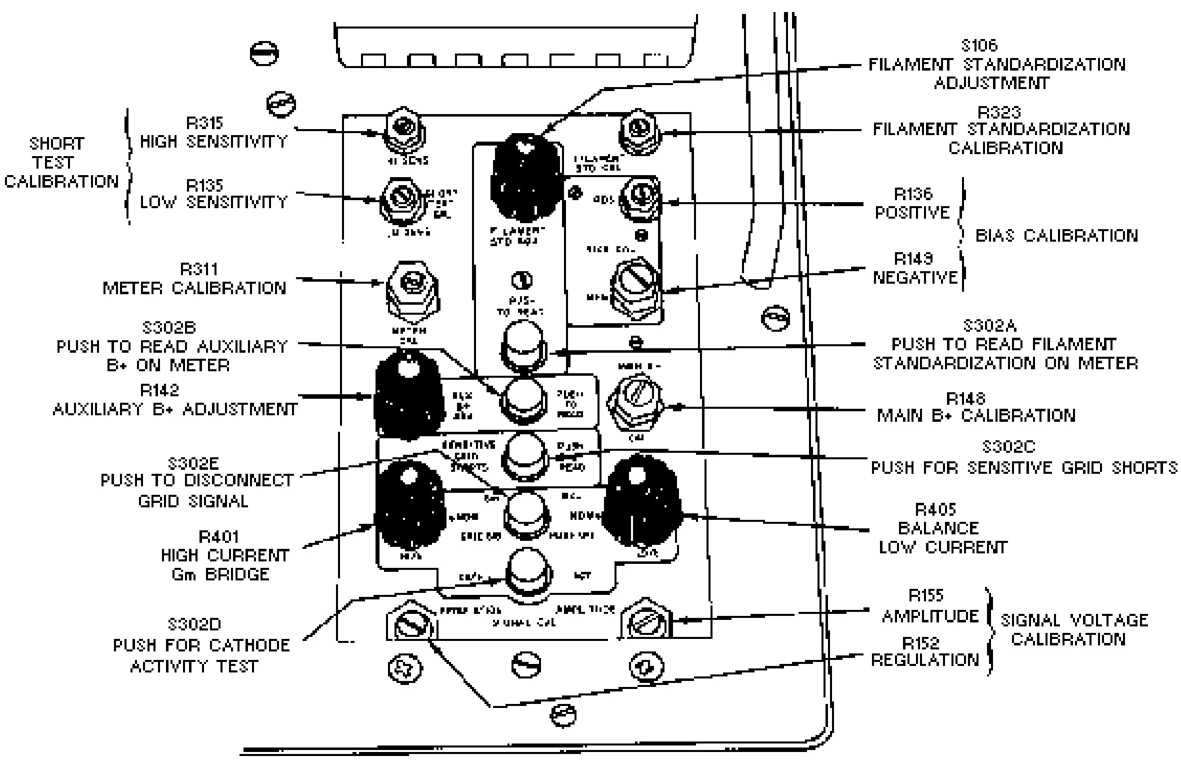


Figure 2-2.—Auxiliary compartment.

Pushbutton S302E and potentiometers R401 and R405 (fig. 2-2) are used for balancing the transconductance (Gm) bridge circuit under actual tube operating current. Pressing S302E removes the grid signal and allows a zero balance to be made with one potentiometer or the other, depending upon whether the tube under test is passing high or low plate current. Lamp DS108 on the front panel lights when S302E is pressed. Pushbutton S302C is used for checking grid-to-cathode shorts at a sensitivity much higher than the normal tests. Results of this test are indicated by the short test lamps on the front panel.

Certain special tests require the use of a continuously adjustable auxiliary power supply. By pressing pushbutton S302B, you may use meter M301 to read the voltage of the auxiliary power supply on meter M301. This voltage may be adjusted by the use of the potentiometer R142. The rest of the potentiometer controls are calibration controls and are adjusted by the use of special calibration cards and a calibration test cell.

All circuits in the tester, except the filament supply, are electronically regulated to compensate for line voltage fluctuations. The filament supply voltage is adjusted by pressing pushbutton S302A and rotating the filament standardization adjustment switch S106 until meter M301 reads midscale.

Program Cards

The circuits to be used in testing are selected by a prepunched card. These cards are made of tough vinyl plastic material. The tube numbers are printed in color on the tabs of the cards and also at the edge of the card for convenience in filing. A special card is provided to use as a marker when a card is removed for use. Blank cards are provided so that additional test cards may be punched for new tubes that are developed or to replace cards that have become unserviceable.

Operation

Before operating the tester for the first time, and periodically thereafter, you should calibrate it using the calibration test cards as described in the equipment technical manual.

NORMAL TESTS.—The tester is equipped with a three-conductor power cord, one wire of which is chassis ground. It should be plugged into a grounded 105- to 125-volt, 50- to 400-hertz outlet.

Before operating the tester, open the auxiliary compartment (fig. 2-2) and ensure that the FILAMENT STD ADJ and the Gm BAL knobs are in the NOM position. The GRID SIG and CATH ACT buttons (S302E and S302D) should be up and lamps DS108 and DS106 on the front panel should be out.

Turn on the tester and allow it to warm up for 5 to 10 minutes, then press the CARD REJECT KNOB (fig. 2-1) down until it locks. If a nontest card is installed in the card switch, remove it. This card is used to keep the switch pins in place during shipment and should be inserted before transporting the tester.

Plug the tube to be tested into its proper socket. (Use the pin straighteners before plugging in 7- and 9-pin miniature tubes.) Select the proper card or cards for the tube to be tested. Insert the card selected into the slot in the card switch until the CARD REJECT KNOB pops up. The card will operate the tester only if it is fully inserted and the printing is up and toward the operator. Do not put paper or objects other than program cards into the card switch, because they will jam the switch contacts. If the overload shuts off the tester when the card is inserted in the switch, check to see that the proper card is being used for the tube under test and that the tube under test has a direct interelement short.

As soon as the card switch is actuated, the tube under test is automatically subjected to an interelement short test and a heater-to-cathode leakage test. A blinking or steady glow of any of the short test lamps is an indication of an interelement short. If the short test lamps remain dark, no interelement shorts exist within the tube. If a short exists between two or more elements, the short test lamp or lamps connected between these elements remain dark, and the remaining lamps light. The abbreviations for the tube elements are located on the front panel just below the short test shield so that the neon lamps are between them. This enables the operator to tell which elements are shorted. Heater-to-cathode shorts are indicated as leakage currents on the #1 meter scale. If the meter reads above the green area, the tube should be replaced. A direct heater-to-cathode short causes the meter to read full scale.

To make the QUALITY test, push the number 2 button (fig. 2-1) and read the number 2 scale on meter M301 to determine if the tube is good. (This test may be one of various types, such as transconductance, emission, plate current, or voltage drop, depending upon the type of tube under test.)

To test the tube for GAS, press the number 3 button and read the number 3 meter scale. The number 2 button also goes down when number 3 is pressed. If a dual tube having two identical sections is being tested, the neon lamp (DS203) will light, indicating that both sections of the tube may be tested with one card. To do this, check the tube for shorts, leakage, quality, and gas as described previously; then hold down button number 4 and repeat these tests to test the second section of the tube. Dual tubes with sections that are not identical require two cards for testing. A second card is also provided to make special tests on certain tubes.

AUXILIARY TEST.—As mentioned previously, two special tests (cathode activity and sensitive grid shorts) may be made by use of controls located in the auxiliary compartment (fig. 2-2). The cathode activity test (CATH ACT) is used to indicate the amount of useful life remaining in the tube. By reducing the filament voltage by 10 percent and allowing the cathode to cool off slightly, the ability of the cathode as an emitter of electrons can be estimated. This test is made in conjunction with the normal quality test.

To make the CATH ACT test, allow the tube under test to warm up, press button number 2 (fig. 2-1), and note the reading of scale number 2 on meter M301. Note also the numerical scale reading on M301. Next, lock down the CATH ACT button (fig. 2-2), wait for about 1.5 minutes, then press button number 2 (fig. 2-1) again and note the numerical and number 2 scale readings on meter M301. The tube should be replaced if the numerical reading on M301 differs from the first reading by more than 10 percent or if the reading is in the red area on the number 2 scale.

It is sometimes desirable to check certain tubes for shorts at a sensitivity greater than normal. To make the SENSITIVE GRID SHORTS test, push S302C (fig. 2-2) and note if any short test lamps (fig. 2-1) light.

HIGH-POWER HF AMPLIFIER TUBE TESTS

You normally test high-power amplifier tubes, which operate in the low-to-high frequency range, in the transmitter in which they are to be used. When you operate the tube in a transmitter, its condition can be determined by using built-in meters to measure the grid current, plate current, and power output and comparing those values with those obtained when using tubes known to be good.

Q-4. Normally, how are high-power rf tubes tested?

Klystron Tube Tests

You can check low-power klystron tubes for gas, frequency of the output signal, and output power by placing them in the equipment where they are to be used. You measure the beam current, output frequency, and output power with the transmitter's built-in test equipment. You can check the output of klystrons used as receiver local oscillators by measuring the current in the crystal mixer unit.

Klystron tubes that remain inoperative for more than 6 months may become gassy. This condition occurs in klystrons installed in stored or spare equipment as well as in klystrons stored as stock supplies. Operation of a gassy klystron at its rated voltages will ionize the gas molecules and may cause excessive beam current to flow. This excessive beam current may shorten the life of the klystron or produce immediate failure. You can detect gas in a klystron tube by setting the applied reflector voltage to zero and slowly increasing the beam voltage while observing a meter that indicates the beam current - excessive beam current for a specific value of voltage indicates that the tube is gassy.

A gassy klystron tube can usually be restored to serviceable condition if you temporarily operate it at reduced beam voltage. Eight hours or more of reduced voltage operation may be required for klystrons that have been inoperative for periods in excess of 6 months.

The beam current is also an indication of the power output of the klystron. As klystrons age they normally draw less beam current; when this current decreases to a minimum value for a specific beam voltage, the tube must be replaced. You can usually determine the power output of transmitter klystrons by measuring the transmitter power output during equipment performance checks.

Q-5. What should you do if a klystron becomes gassy?

Traveling-Wave Tube

You can usually test a traveling-wave tube (twt) in the equipment in which it is used. When the twt is installed, you can usually measure the collector current and voltage and check the power output for various inputs. Any deviation greater than 10% from normal specifications may be considered to be an indication of a defective tube. Most amplifiers are supplied with built-in panel meters and selector switches so that the cathode, anode, helix, focus, and collector currents may be measured. Thus, continuous monitoring of amplifier operation and tube evaluation is possible. Adjustments usually are provided for you to set the helix, grid bias, and collector voltages for optimum operation. If variation of these controls will not produce normal currents and if all voltages are normal, you should consider the tube to be defective and replace it with a new tube or one known to be in good operating condition. To avoid needless replacement of tubes, however, you should make an additional check by measuring the input power and output power and determining the tube gain. If, with normal operating conditions, the gain level drops below the minimum indicated value in the equipment technical manual, the tube is defective.

Q-6. When used as an amplifier, what is the best indication that a twt is operating properly?

In the absence of special field-test sets, you may construct a laboratory test mock-up similar to that shown in figure 2-3. Because of the variations in power and gain between tubes and the large frequency ranges offered, we can illustrate only a general type of equipment. The equipment you select must have the proper range, impedance, and attenuation to make the test for a specific type of twt. To make gain measurements, you turn the switch shown in figure 2-3 to position 1 and set the precision attenuator to provide a convenient level of detector output. Then turn the switch to position 2 and insert attenuation until the detector output level is identical to that obtained without the twt in the circuit. The gain of the traveling-wave tube is equal to the amount of added attenuation.

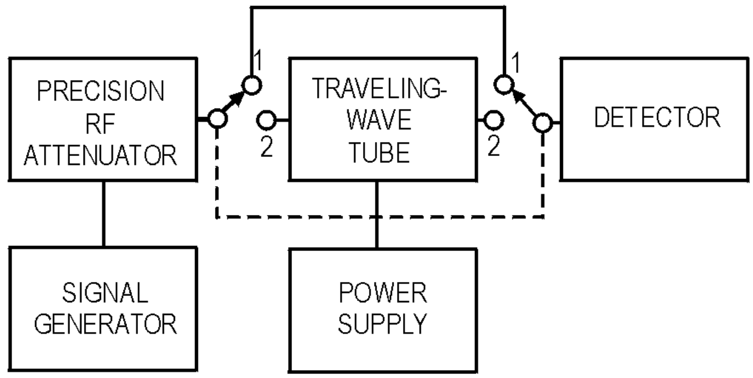


Figure 2-3.—Traveling-wave tube test arrangement.

When you use the twt as an oscillator, failure of the tube to break into oscillations when all other conditions are normal usually indicates a defective tube. In the case of a tube used as a receiving amplifier, an increase of noise with a normal or reduced output can indicate that the tube is failing but is still usable. All the general rules applying to klystron tubes mentioned previously are also applicable to the twt.

Magnetron Tube Tests

You test a magnetron tube while it is in the transmitter equipment in which it is to be used. When you install the magnetron in the transmitter, the condition of the tube can be determined by the normal plate-current measurement and the power, frequency spectrum, and standing-wave-ratio tests of the output signal. An unusual value for any of these measurements may indicate a defective tube.

Crossed-Field Amplifier

You usually test a crossed-field amplifier (cfa) tube while it is in the equipment in which it is used. Like the klystron, if you do not operate the cfa for more than a few months, the tube may become gassy. If a cfa tube is suspected of being gassy, we recommend that you consult the technical manual for the particular piece of equipment in which the crossed-field amplifier is used.

TESTING SEMICONDUCTORS

Unlike vacuum tubes, transistors are very rugged in that they can tolerate vibration and a rather large degree of shock. Under normal operating conditions, they will provide dependable operation for a long period of time. However, transistors are subject to failure when they are subjected to relatively minor overloads. Crystal detectors are also subject to failure or deterioration when subjected to electrical overloads and will deteriorate from a long period of normal use. To determine the condition of semiconductors, you can use various test methods. In many cases you may substitute a transistor of known good quality for a questionable one to determine the condition of a suspected transistor. This method is highly accurate and sometimes efficient. However, you should avoid indiscriminate substitution of semiconductors in critical circuits. When transistors are soldered into equipment, substitution becomes impractical - generally, you should test these transistors while they are in their circuits.

Q-7. What is the major advantage of a transistor over a tube?

Since certain fundamental characteristics indicate the condition of semiconductors, test equipment is available that allows you to test these characteristics with the semiconductors in or out of their circuits. Crystal-rectifier testers normally allow you to test only the forward-to-reverse current ratio of the crystal. Transistor testers, however, allow you to measure several characteristics, such as the collector leakage current ($_{Ic}$), collector to base current gain (β), and the four-terminal network parameters. The most useful test characteristic is determined by the type of circuit in which the transistor will be used. Thus, the alternating-current beta measurement is preferred for ac amplifier or oscillator applications; and for switching-circuit applications, a direct-current beta measurement may prove more useful.

Many common transistors are extremely heat sensitive. Excess heat will cause the semiconductor to either fail or give intermittent operation. You have probably experienced intermittent equipment problems and know them to be both time consuming and frustrating. You know, for example, that if a problem is in fact caused by heat, simply opening the equipment during the course of troubleshooting may cause the problem to disappear. You can generally isolate the problem to the faulty printed-circuit board (pcb) by observing the fault indications. However, to further isolate the problem to a faulty component, sometimes you must apply a minimal amount of heat to the suspect pcb by carefully using a low wattage, heat shrink gun; an incandescent drop light; or a similar heating device. Be careful not to overheat the pcb. Once the fault indication reappears, you can isolate the faulty component by spraying those components suspected as being bad with a nonconductive circuit coolant, such as Freon. If the alternate heating and cooling of a component causes it to operate intermittently, you should replace it.

Q-8. Name two major disadvantages of transistors.

TRANSISTOR TESTING

When trouble occurs in solid-state equipment, you should first check power supplies and perform voltage measurements, waveform checks, signal substitution, or signal tracing. If you isolate a faulty stage by one of these test methods, then voltage, resistance, and current measurements can be made to locate defective parts. When you make these measurements, the voltmeter impedance must be high enough that it exerts no appreciable effect upon the voltage being measured. Also, current from the ohmmeter you use must not damage the transistors. If the transistors are not soldered into the equipment, you should remove the transistors from the sockets during a resistance test. Transistors should be removed from or reinserted into the sockets only after power has been removed from the stage; otherwise damage by surge currents may result.

Transistor circuits, other than pulse and power amplifier stages, are usually biased so that the emitter current is from 0.5 milliampere to 3 milliamperes and the collector voltage is from 3 to 15 volts. You can measure the emitter current by opening the emitter connector and inserting a milliammeter in series. When you make this measurement, you should expect some change in bias because of the meter resistance. You can often determine the collector current by measuring the voltage drop across a resistor in the collector circuit and calculating the current. If the transistor itself is suspected, it can be tested by one or more of the methods described below.

Resistance Test

You can use an ohmmeter to test transistors by measuring the emitter-collector, base-emitter, and base-collector forward and reverse resistances. A back-to-forward resistance ratio on the order of 100 to 1 or greater should be obtained for the collector-to-base and emitter-to-base measurements. The forward and reverse resistances between the emitter and collector should be nearly equal. You should make all three measurements for each transistor you test, because experience has shown that transistors can develop shorts between the collector and emitter and still have good forward and reverse resistances for the other two measurements. Because of shunting resistances in transistor circuits, you will normally have to disconnect at least two transistor leads from the associated circuit for this test. Exercise caution during this test to make certain that current during the forward resistance tests does not exceed the rating of the transistor — ohmmeter ranges requiring a current of more than 1 milliampere should not be used for testing transistors. Many ohmmeters are designed such that on the R × 1 range, 100 milliamperes or more can flow through the electronic part under test. For this reason, you should use a digital multimeter. Be sure you select a digital multimeter that produces enough voltage to properly bias the transistor junctions.

Q-9. When you are using an ohmmeter to test a transistor, what range settings should be avoided?

Transistor Testers

Laboratory transistor test sets are used in experimental work to test all characteristics of transistors. For maintenance and repair, however, it is not necessary to check all of the transistor parameters. A check of two or three performance characteristics is usually sufficient to determine whether a transistor needs to be replaced. Two of the most important parameters used for transistor testing are the transistor current gain (beta) and the collector leakage or reverse current (I_c).

The semiconductor test set (fig. 2-4) is a rugged, field type of tester designed to test transistors and semiconductor diodes. The set measures the beta of a transistor, resistance appearing at the electrodes, reverse current of a transistor or semiconductor diode, shorted or open conditions of a diode, forward transconductance of a field-effect transistor, and condition of its own batteries.

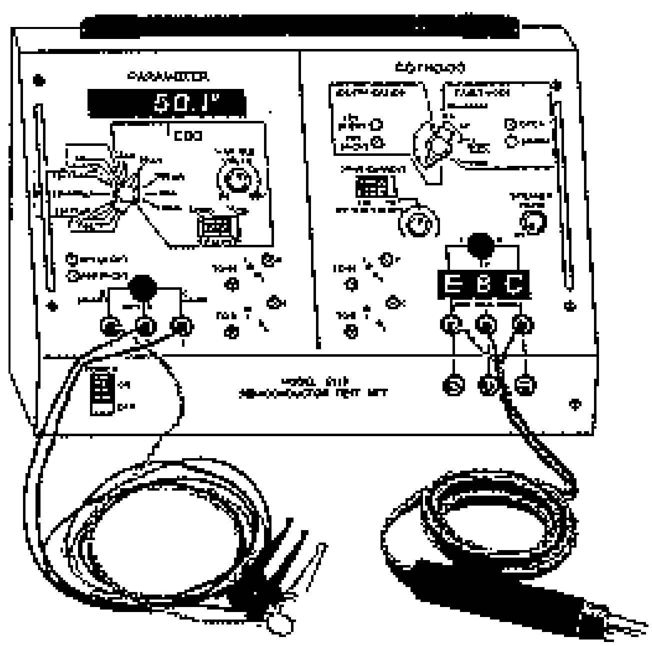


Figure 2-4.—Semiconductor test set.

In order to assure that accurate and useful information is gained from the transistor tester, the following preliminary checks of the tester should be made prior to testing any transistors.

With the POLARITY switch (fig. 2-4) in the OFF position, the meter pointer should indicate exactly zero. (When required, rotate the meter adjust screw on the front of the meter to fulfill this requirement.) When measurements are not actually being made, the POLARITY switch must always be left in the OFF position to prevent battery drain.

Always check the condition of the test set batteries by disconnecting the test set power cord, placing the POLARITY switch in the PNP position and placing the FUNCTION switch first to BAT.1, then to BAT.2. In both BAT positions the meter pointer should move so as to indicate within the red BAT range.

BETA MEASUREMENTS.—If the transistor is to be tested out of the circuit, plug it into the test jack located on the right-hand side below the meter shown in figure 2-4. If the transistor is to be tested in the circuit, it is imperative that at least 300 ohms exist between E-B, C-B, and C-E for accurate measurement. Initial settings of the test set controls are as follows:

1. FUNCTION switch to BETA
2. POLARITY switch to PNP or NPN (dependent on type of transistor under test)
3. RANGE switch to X10
4. Adjust METER ZERO for zero meter indication (transistor disconnected)

NOTE: The POLARITY switch should remain OFF while the transistor is connected to or disconnected from the test set. If you determine that the beta reading is less than 10, reset the RANGE switch to X1 and reset the meter to zero.

After connecting the yellow test lead to the emitter, the green test lead to the base, and the blue test lead to the collector, plug the test probe (not shown) into the jack located at the lower right-hand corner of the test set. When testing grounded equipment, unplug the 115 vac line cord and use battery operation. The beta reading is attained by multiplying the meter reading times the RANGE switch setting. Refer to the transistor characteristics book provided with the tester to determine if the reading is normal for the type of transistor under test.

ELECTRODE RESISTANCE MEASUREMENTS.—Connect the in-circuit probe test leads to the transistor with the yellow lead to the emitter, the green lead to the base, and the blue lead to the collector. Set the FUNCTION switch to the OHMS E-B position, and read the resistance between the emitter and base electrode on the center scale of the meter.

To read the resistance between the collector and base and the collector and emitter, set the FUNCTION switch to OHMS C-B and OHMS C-E. These in-circuit electrode resistance measurements are used to correctly interpret the in-circuit beta measurements. The accuracy of the BETA X1, X10 range is ±15 percent only when the emitter-to-base load is equal to or greater than 300 ohms.

I_c MEASUREMENTS.—Adjust the METER ZERO control for zero meter indication. Plug the transistor to be tested into the jack or connect test leads to the device under test. Set the PNP/NPN switch to correspond with the transistor under test. Set the FUNCTION switch to I_c and the RANGE switch to X0.1, X1, or X10 as specified by the transistor data book for allowable leakage. Read the amount of leakage on the bottom scale, and multiply this by the range setting figure as required.

DIODE MEASUREMENTS.—Diode qualitative in-circuit measurements are attained by connecting the green test lead to the cathode and the yellow test lead to the anode. Set the FUNCTION switch to DIODE IN/CKT and the RANGE switch to X1. (Ensure that the meter has been properly zeroed on this scale.) If the meter reads down scale, reverse the POLARITY switch. If the meter reads less than midscale, the diode under test is either open or shorted. The related circuit impedance of this test is less than 25 ohms.

PRECAUTIONS.—Transistors, although generally more rugged mechanically than electron tubes, are susceptible to damage by excessive heat and electrical overload. The following precautions should be taken in servicing transistorized equipment:

1. Test equipment and soldering irons must be checked to make certain that there is no leakage current from the power source. If leakage current is detected, isolation transformers must be used.
2. Ohmmeter ranges that require a current of more than 1 milliampere in the test circuit are not to be used for testing transistors.
3. Battery eliminators should not be used to furnish power for transistor equipment because they have poor voltage regulation and, possibly, high ripple voltage.
4. The heat applied to a transistor, when soldered connections are required, should be kept to a minimum by using a low-wattage soldering iron and heat shunts (such as long-nose pliers) on the transistor leads.
5. All circuits should be checked for defects before a transistor is replaced.
6. The power should be removed from the equipment before replacing a transistor or other circuit part.
7. When working on equipment with closely spaced parts, you will find that conventional test probes are often the cause of accidental short circuits between adjacent terminals. Momentary short circuits, which rarely cause damage to an electron tube, may ruin a transistor. To avoid accidental shorts, a test probe can be covered with insulation for all but a very short length of the tip.

Electrostatic Discharge Sensitive (ESDS) Care

Devices that are sensitive to electrostatic discharge (ESD) require special handling. You can readily identify ESD-sensitive (ESDS) devices by the symbols shown in figure 2-5. Static electricity is created whenever two substances (solid or fluid) are rubbed together or separated. The rubbing or separating of substances causes the transfer of electrons from one substance to the other; one substance then becomes positively charged, and the other becomes negatively charged. When either of these charged substances comes in contact with a grounded conductor, an electrical current flows until that substance is at the same electrical potential as ground.

Figure 2-5.—Warning symbols for ESDS devices.

You commonly experience static build-up during the winter months when you walk across a vinyl or carpeted floor. (Synthetics, especially plastics, are excellent generators of static electricity.) If you then touch a doorknob or any other conductor, an electrical arc to ground may result, and you may receive a slight shock. For you to experience such a shock, the electrostatic potential created must be 3,500 to 4,000 volts. Lesser voltages, although present and similarly discharged, normally are not apparent to your nervous system. Some typical measured static charges caused by various actions are shown in table 2-1.

Table 2-1.—Typical Measured Static Charges (in volts)

ITEM	RELATIVE HUMIDITY	
	LOW (10 - 20%)	HIGH (65 - 90%)
WALKING ACROSS CARPET	35,000V	1,500V
WALKING OVER VINYL FLOOR	12,000V	250V
WORKER AT BENCH	6,000V	100V
VINYL ENVELOPES FOR WORK INSTRUCT.	7,000V	600V
POLY BAG PICKED UP FROM BENCH	20,000V	1,200V
WORK CHAIR PADDED WITH URETHANE FORM	18,000 V	1,500 V

Q-10. At approximately what minimum voltage potential should you be able to feel an electrostatic discharge?

Metal oxide semiconductor (MOS) devices are the most susceptible to damage from ESD. For example, an MOS field-effect transistor (MOSFET) can be damaged by a static voltage potential of as little as 35 volts. Commonly used discrete bipolar transistors and diodes (often used in ESD-protective circuits), although less susceptible to ESD, can be damaged by voltage potentials of less than 3,000 electrostatic volts. Damage does not always result in sudden device failure but sometimes results in device degradation and early failure. Table 2-1 clearly shows that electrostatic voltages well in excess of

3,000 volts can be easily generated, especially under low-humidity conditions. ESD damage of ESDS parts or circuit assemblies is possible whenever two or more pins of any of these devices are electrically exposed or have low impedance paths. Similarly, an ESDS device in a printed-circuit board or even in another pcb that is electrically connected in a series can be damaged if it provides a path to ground. ESD damage can occur during the manufacture of equipment or during the servicing of the equipment. Damage can occur anytime devices or assemblies are handled, replaced, tested, or inserted into a connector.

Q-11. A MOSFET can be damaged by an electrostatic discharge at approximately what minimum potential?

ESD-sensitive devices can be grouped by their sensitivity to ESD. Semiconductors fall within the following categories:

- VERY SENSITIVE DEVICES. These include MOS and CMOS devices without input diode protection circuitry on all input circuits; dielectrically isolated semiconductors with internal capacitor contacts connected to external pins; and microcircuits using N + guard-ring construction (with metalization crossing over the guard ring).
- SENSITIVE DEVICES. These include all low-power Schottky-barrier and Schottky-TTL devices; all ECL devices; high input-impedance linear microcircuits; all small-signal transistors that operate at 500 MHz or higher; all discrete semiconductors that use silicon dioxide to insulate metal paths over other active areas; MOS or CMOS devices with input diode protection on all input terminals; junction field-effect transistors; and precision resistive networks.
- MODERATELY SENSITIVE DEVICES. These include all microcircuits and small-signal discrete semiconductors with less than 10 watts dissipation at 25º C, and thick-film resistors.

The following procedure is an example of some of the protective measures used to prevent ESD damage:

1. Before servicing equipment, you should be grounded to discharge any static electricity from your body. This can be accomplished with the use of a test lead (a single-wire conductor with a series resistance of 1 megohm) equipped with alligator clips on each end. After the equipment has been completely de-energized, one clip end is connected to the grounded equipment frame; the other clip end is touched with your bare hand. Figure 2-6 shows a more refined ground strap, which frees both hands for work.

NOTE: When wearing a wrist strap, you should *never use ac-powered test equipment* because of your increased chance of receiving an electrical shock.

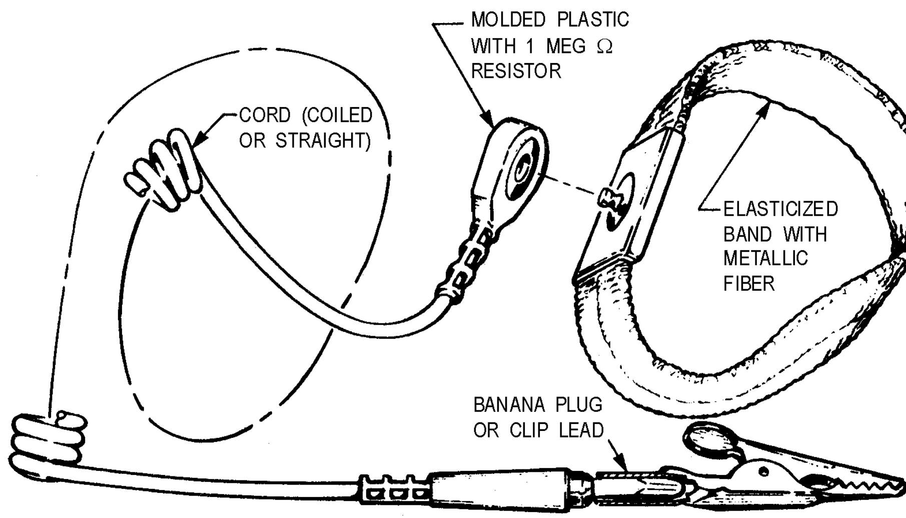


Figure 2-6.—ESD wrist strap.

2. Equipment technical manuals and packaging material should be checked for ESD warnings and instructions.

3. Prior to opening an electrostatic unit package of an ESDS device or assembly, clip the free end of the grounded test lead to the package. This will cause any static electricity that may have built up on the package to discharge. The other end remains connected to the equipment frame or other ESD ground. Keep the unit package grounded until the replacement device or assembly is placed in the unit package.

4. Minimize handling of ESDS devices and assemblies. Keep replacement devices or assemblies, with their connector-shorting bars, clips, and so forth, intact in their electrostatic-free packages until needed. Place removed repairable ESDS devices or assemblies, with their connector shorting bars or clips installed, in electrostatic-free packages as soon as they are removed from the equipment. ESDS devices or assemblies should be transported and stored only in protective packaging.

5. Always avoid unnecessary physical movement, such as scuffing the feet, when handling ESDS devices or assemblies. Such movement will generate additional charges of static electricity.

6. When removing or replacing an ESDS device or assembly in the equipment, hold the device or assembly through the electrostatic-free wrap if possible. Otherwise, pick up the device or assembly by its body only. DO NOT TOUCH component leads, connector pins, or any other electrical connections or paths on boards, even though they are covered by conformal coating.

7. Do not permit ESDS devices or assemblies to come in contact with clothing or other ungrounded materials that could have an electrostatic charge. The charges on a nonconducting material are not equal. A plastic storage bag may have a −10,000 volt potential one-half inch from a +15,000 volt potential, with many other such charges all over the bag. Placing a circuit card inside the bag allows the charges to equalize through the pcb conductive paths and components, thereby causing failures. Do not hand an ESDS device or assembly to another person until the device or assembly is protectively packaged.

8. When moving an ESDS device or assembly, always touch (with your bare skin) the surface on which it rests for at least 1 second before picking it up. Before placing it on any surface, touch the

surface with your free hand for at least 1 second. The bare skin contact provides a safe discharge path for electrostatic charges accumulated while you are moving around.

9. While servicing equipment containing ESDS devices, do not handle or touch materials such as plastic, vinyl, synthetic textiles, polished wood, fiber glass, or similar items that could create static charges; or, be sure to repeat the grounding action with the bare hands after contacting these materials. These materials are prime electrostatic generators.

10. If possible, avoid repairs that require soldering at the equipment level. Soldering irons must have heater and tip assemblies grounded to ac electrical ground. Do not use ordinary plastic solder suckers (special antistatic solder suckers are commercially available).

11. Ground the leads of test equipment momentarily before you energize the test equipment and before you probe ESDS items.

Q-12. Why should you avoid using ac-powered test equipment when wearing a wrist strap?

DIODE TESTING

Because of the reliability of semiconductor devices, servicing techniques developed for transistorized equipment differ from those used for electron-tube circuits. Electron tubes are usually considered to be the circuit component most susceptible to failure and are normally the first to be tested. Transistors, however, are capable of operating in excess of 30,000 hours at maximum ratings without appreciable degradation. They are often soldered into equipment in the same manner as resistors and capacitors. Substitution of a diode or transistor known to be in good condition is a simple method of determining the quality of a questionable semiconductor device. You should use this technique only after voltage and resistance measurements indicate that no circuit defect exists that might damage the substituted semiconductor device. If more than one defective semiconductor is present in the equipment section where trouble has been localized, substitution becomes cumbersome since several semiconductors may have to be replaced before the trouble is corrected. To determine which stages failed and which semiconductors are not defective, you must test all of the removed semiconductors. This can be accomplished by observing whether the equipment operates correctly as each of the removed semiconductor devices is reinserted into the equipment.

Q-13. Prior to substituting a diode, what measurements should you take to determine its condition?

DIODE TESTERS

Diodes, such as general-purpose germanium and silicon diodes, power silicon diodes, and microwave silicon diodes, are most effectively tested under actual operating conditions. However, rectifier testers are available for you to determine direct-current characteristics that provide an indication of diode quality.

Rf Diode Test

A common type of diode test set is a combination ohmmeter-ammeter. You can make measurements of forward resistance, back resistance, and reverse current with this equipment. You can determine the condition of the rectifier under test by comparing its actual values with typical values obtained from test information furnished with the test set or from the manufacturer’s data sheets. Comparing the diode’s back and forward resistance at a specified voltage provides you with a rough indication of the rectifying property of a diode. A typical back-to-forward resistance ratio is on the order of 10 to 1, and a forward-resistance value of 50 to 80 ohms is common.

Switching Diode Test

To effectively test diodes used for computer applications, you must obtain back-resistance measurements at a large number of different voltage levels. This can be done efficiently by using a dynamic diode tester in conjunction with an oscilloscope, which is used to display the diode's back-current-versus-voltage curve. You can easily interpret diode characteristics, such as flutter, hysteresis, and negative resistance, through use of the dynamic current and voltage display.

DIODE CHARACTERISTIC GRAPHICAL DISPLAY

You can use an oscilloscope to graphically display the forward- and back-resistance characteristics of a diode. A test circuit used in conjunction with an oscilloscope is shown in figure 2-7. This circuit uses an audio-signal generator as the test signal. It should be adjusted for an approximate 2-volt, 60-hertz signal, as measured across R1.

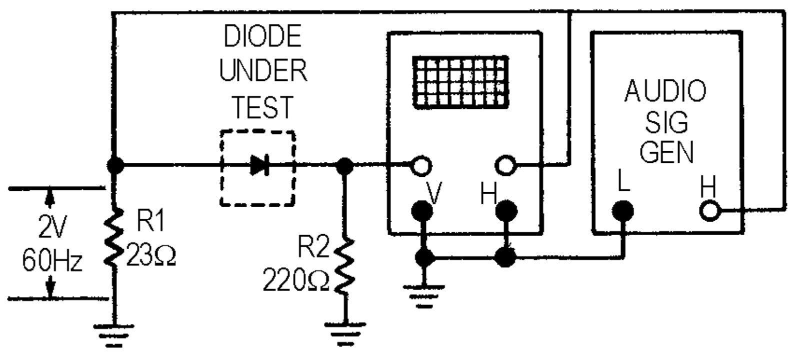


Figure 2-7.—Display circuit used with an oscilloscope.

The test signal you apply to the diode is also connected to the horizontal input of the oscilloscope. The horizontal sweep will then display the voltage applied to the diode under test. The voltage developed across current-measuring resistor R2 is applied to the vertical input of the oscilloscope. Since this voltage is proportional to the current through the diode under test, the vertical deflection will indicate diode current. The resulting oscilloscope trace will be similar to the curve shown in figure 2-8.

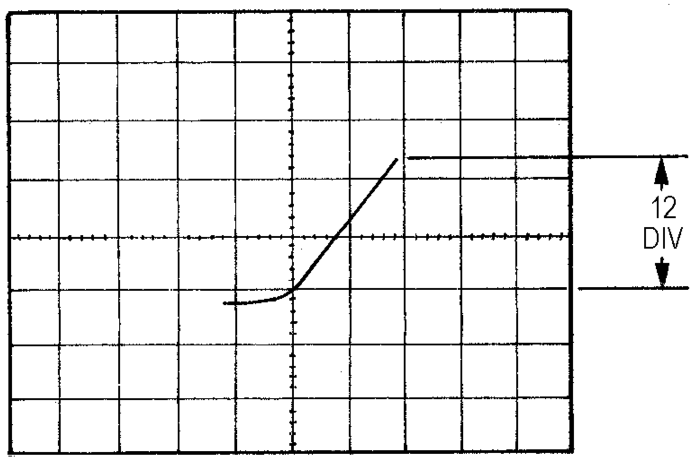


Figure 2-8.—Typical characteristic curve of a silicone diode.

Reverse Voltage-Current Analysis

You can make an analysis of the reverse voltage-current portion of the characteristic curve for a diode with the method described above or with a diode test set. This test is very important for diodes used in computer applications, where stability of operation is essential. Various diode conditions that may be detected by this test are shown in figure 2-9, view A, view B, view C, and view D.

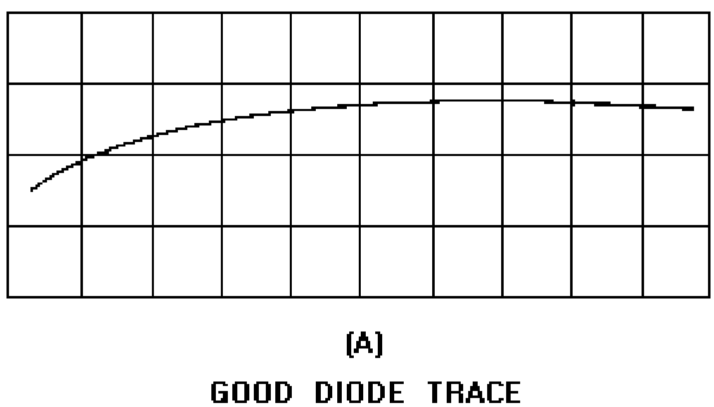


Figure 2-9A.—Diode reverse current-voltage characteristics. GOOD DIODE TRACE.

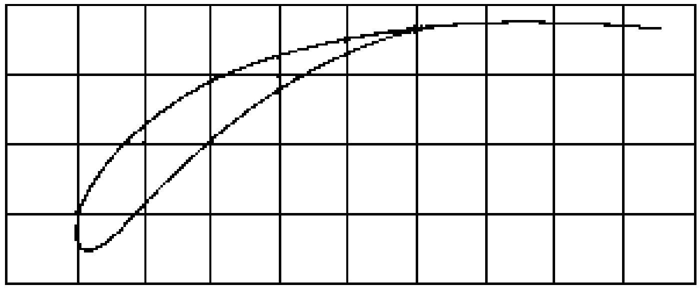


Figure 2-9B.—Diode reverse current-voltage characteristics. HYSTERESIS CHARACTERISTIC.

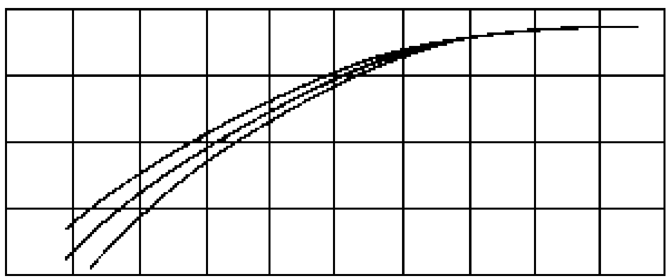


Figure 2-9C.—Diode reverse current-voltage characteristics. FLUTTER (OR DRIFT) CHARACTERISTIC.

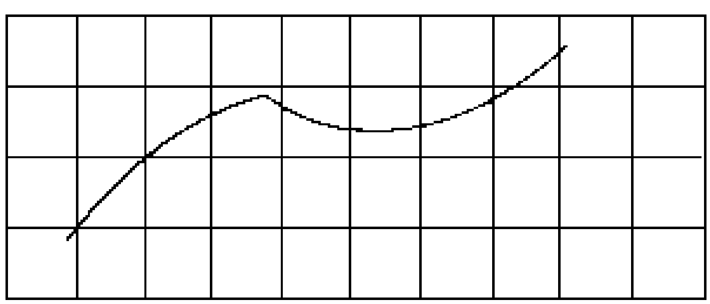


Figure 2-9D.—Diode reverse current-voltage characteristics. NEGATIVE RESISTANCE TRACE.

Zener Diode Test

An audio signal generator may not be able to produce a high enough voltage for you to test Zener diodes. You can, however, make this test with a diode test set or with the circuit shown in figure 2-10. In this circuit, R1 is used to adjust the input voltage to a suitable value for the Zener diode being tested. Resistor R2 limits the current through the diode. The signal voltage applied to the diode is also connected to the horizontal input of the oscilloscope. The voltage developed across current-measuring resistor R3 is applied to the vertical input of the oscilloscope. The horizontal sweep will therefore represent the applied voltage, and the vertical deflection will indicate the current through the diode under test. Figure 2-11 shows the characteristic pattern of a Zener diode (note the sharp increase in current at the avalanche breakdown point). For the Zener diode to be acceptable, this voltage must be within the limits specified by the manufacturer.

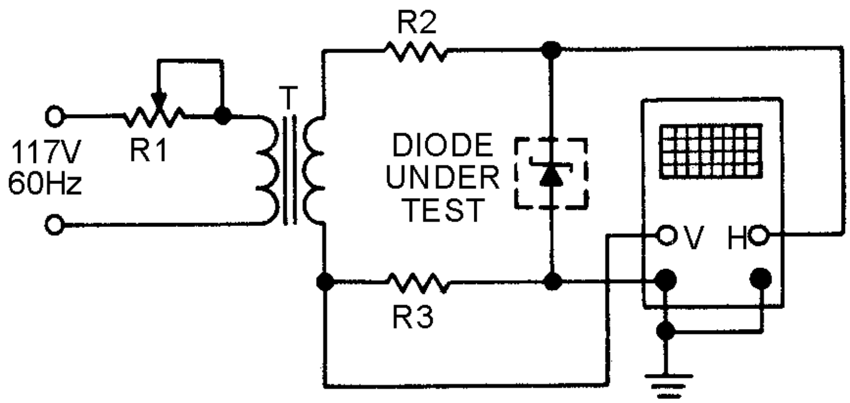


Figure 2-10.—Zener diode test circuit.

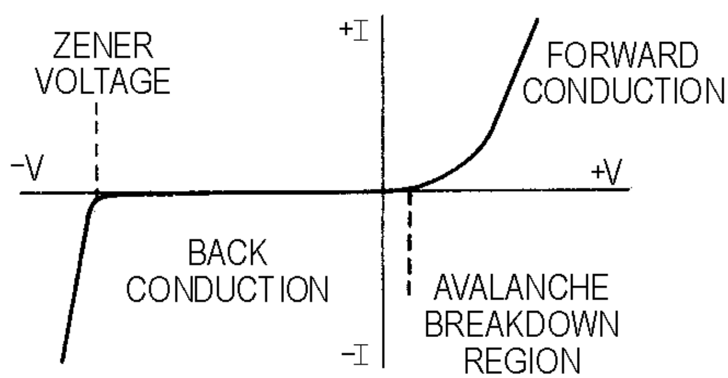


Figure 2-11.—Zener diode characteristic pattern.

STATIC RESISTANCE MEASUREMENTS

One convenient method of testing a diode requires only your ohmmeter. The forward and back resistances can be measured at a voltage determined by the battery potential of the ohmmeter and the resistance range at which the meter is set. When the test leads of the ohmmeter are connected to the diode, a resistance will be measured that is different from the resistance indicated if the leads are reversed. The smaller value is called the forward resistance, and the larger value is called the back resistance. If the ratio of back-to-forward resistance is greater than 10 to 1, the diode should be capable of functioning as a rectifier. This is a very limited test, which does not take into account the action of the diode at voltages of different magnitudes and frequencies. Some diodes may be damaged by the excessive current produced by some range settings of a standard multimeter. Therefore, you should use a digital multimeter when performing this measurement.

Q-14. As a rule of thumb, what is an acceptable ratio of back-to-forward resistance for a diode?

SILICON-CONTROLLED RECTIFIERS (SCR)

Many naval electronic equipments use silicon-controlled rectifiers (SCRs) for the control of power. Like other solid-state components, SCRs are subject to failure. You can test most SCRs with a standard ohmmeter, but you must understand just how the SCR functions.

As shown in figure 2-12, the SCR is a three-element, solid-state device in which the forward resistance can be controlled. The three active elements shown in the figure are the anode, cathode, and gate. Although they may differ in outward appearance, all SCRs operate in the same way. The SCR acts like a very high-resistance rectifier in both forward and reverse directions without requiring a gate signal. However, when the correct gate signal is applied, the SCR conducts only in the forward direction, the same as any conventional rectifier. To test an SCR, you connect an ohmmeter between the anode and cathode, as shown in figure 2-12. Start the test at $R \times 10,000$ and reduce the value gradually. The SCR under test should show a very high resistance, regardless of the ohmmeter polarity. The anode, which is connected to the positive lead of the ohmmeter, must now be shorted to the gate. This will cause the SCR to conduct; as a result, a low-resistance reading will be indicated on the ohmmeter. Removing the anode-to-gate short will not stop the SCR from conducting; but removing either of the ohmmeter leads will cause the SCR to stop conducting — the resistance reading will then return to its previous high value. Some SCRs will not operate when you connect an ohmmeter. This is because the ohmmeter does not supply enough current. However, most of the SCRs in Navy equipment can be tested by the ohmmeter method. If an SCR is sensitive, the $R \times 1$ scale may supply too much current to the device and damage it. Therefore, try testing it on the higher resistance scales.

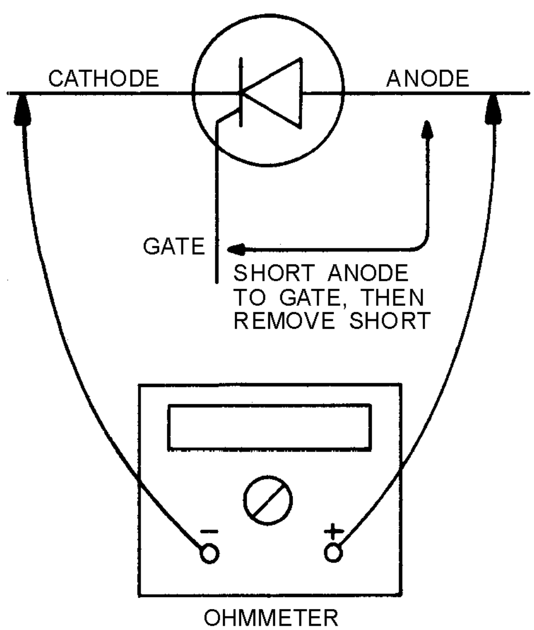


Figure 2-12.—Testing an SCR with an ohmmeter.

Q-15. When testing an SCR with an ohmmeter, the SCR will conduct if what two elements are shorted together?

TRIAC

Triac is General Electric's trade name for a silicon, gate-controlled, full-wave, ac switch, as shown in figure 2-13. The device is designed to switch from a blocking state to a conducting state for either polarity of applied voltages and with either positive or negative gate triggering. Like a conventional SCR, the Triac is an excellent solid-state device for controlling current flow. You can make the Triac conduct by using the same method used for an SCR, but the Triac has the advantage of being able to conduct equally well in either the forward or reverse direction.

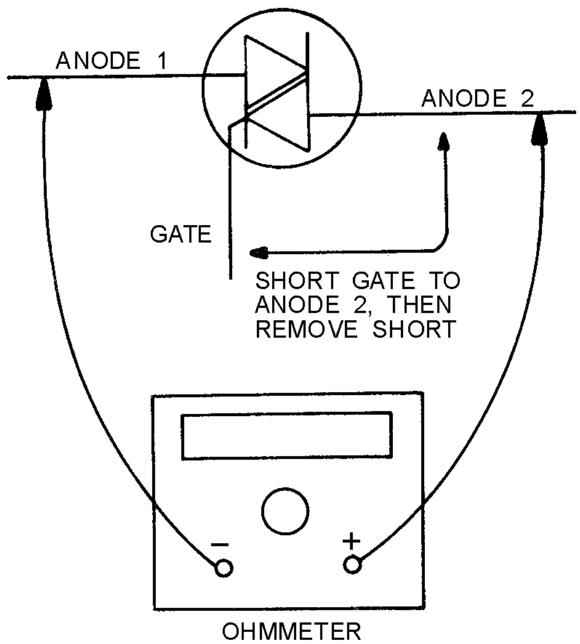


Figure 2-13.—Testing a Triac with an ohmmeter.

To test the Triac with an ohmmeter (R × 1 scale), you connect the ohmmeter's negative lead to anode 1 and the positive lead to anode 2, as shown in figure 2-13. The ohmmeter should indicate a very high resistance. Short the gate to anode 2; then remove it. The resistance reading should drop to a low value and remain low until either of the ohmmeter leads is disconnected from the Triac. This completes the first test.

The second test involves reversing the ohmmeter leads between anodes 1 and 2 so that the positive lead is connected to anode 1 and the negative lead is connected to anode 2. Again, short the gate to anode 2; then remove it. The resistance reading should again drop to a low value and remain low until either of the ohmmeter leads is disconnected.

Q-16. When a Triac is properly gated, what is/are the direction(s) of current flow between anodes 1 and 2?

UNIJUNCTION TRANSISTORS (UJTs)

The unijunction transistor (UJT), shown in figure 2-14, is a solid-state, three-terminal semiconductor that exhibits stable open-circuit, negative-resistance characteristics. These characteristics enable the UJT to serve as an excellent oscillator. Testing a UJT is a relatively easy task if you view the UJT as being a diode connected to the junction of two resistors, as shown in figure 2-15. With an ohmmeter, measure the resistance between base 1 and base 2; then reverse the ohmmeter leads and take another reading. Readings should show the same high resistance regardless of meter lead polarity. Connect the negative lead of the ohmmeter to the emitter of the UJT. Using the positive lead, measure the resistance from the emitter to base 1 and then from the emitter to base 2. Both readings should indicate high resistances that are approximately equal to each other. Disconnect the negative lead from the emitter and connect the positive lead to it. Using the negative lead, measure the resistance from the emitter to base 1 and then from the emitter to base 2. Both readings should indicate low resistances approximately equal to each other.

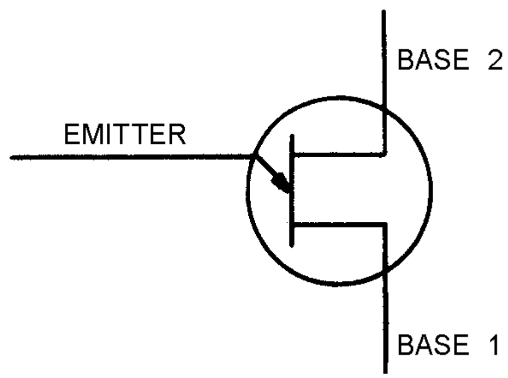


Figure 2-14.—Unijunction transistor.

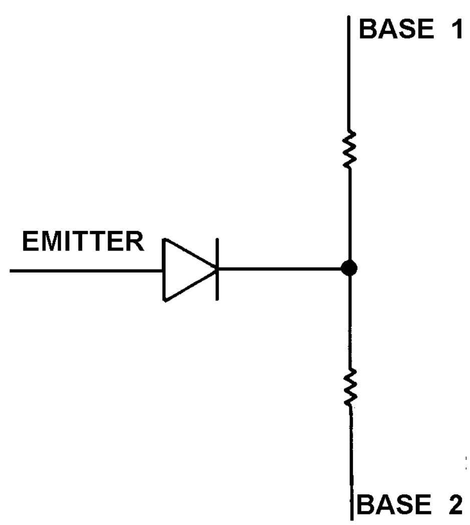


Figure 2-15.—Unijunction transistor equivalent circuit.

JUNCTION FIELD-EFFECT TRANSISTOR (JFET) TESTS

The junction field-effect transistor (JFET) has circuit applications similar to those of a vacuum tube. The JFET has a voltage-responsive characteristic with a high input impedance. Two types of JFETs that you should become familiar with are the junction p-channel and the junction n-channel types, as shown in figure 2-16. Their equivalent circuits are shown in figures 2-17 and 2-18, respectively. The only difference in your testing of these two types of JFETs involves the polarity of the meter leads.

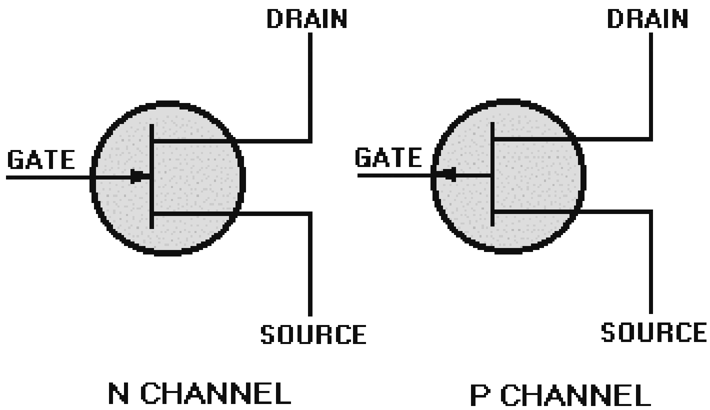


Figure 2-16.—Junction FETs.

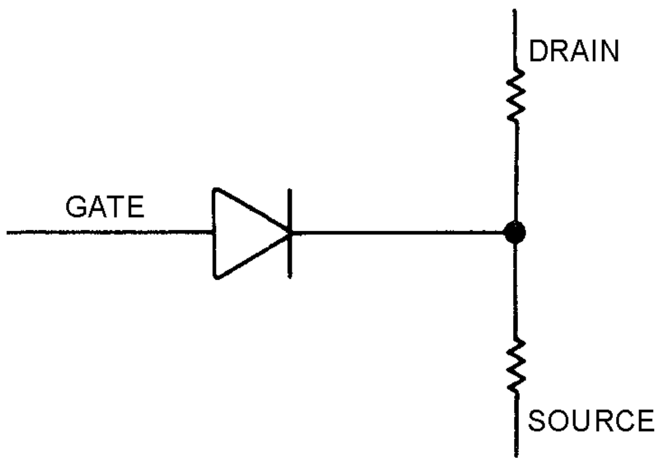


Figure 2-17.—N-channel JFET equivalent circuit.

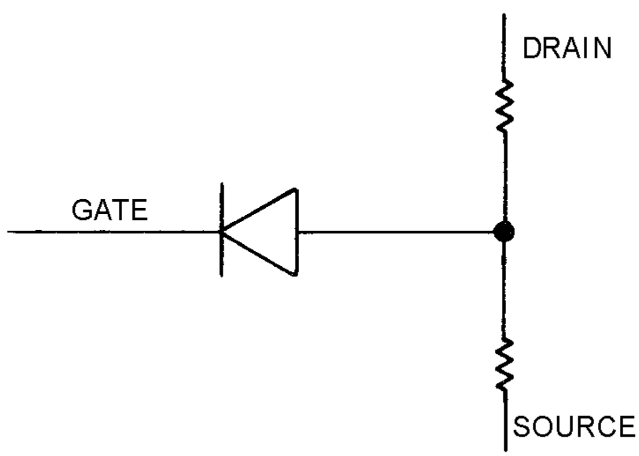


Figure 2-18.—P-channel JFET equivalent circuit.

N-Channel Test

Using an ohmmeter set to the R $\times$ 100 scale, measure the resistance between the drain and the source; then reverse the ohmmeter leads and take another reading. Both readings should be equal (in the 100- to 10,000-ohm range), regardless of the meter lead polarity. Connect the positive meter lead to the gate. Using the negative lead, measure the resistance between the gate and the drain; then measure the resistance between the gate and the source. Both readings should indicate a low resistance and be approximately the same. Disconnect the positive lead from the gate and connect the negative lead to the gate. Using the positive lead, measure the resistance between the gate to the drain; then measure the resistance between the gate and the source. Both readings should show infinity.

P-Channel Test

Using an ohmmeter set to the R $\times$ 100 scale, measure the resistance between the drain and the source; then reverse the ohmmeter leads and take another reading. Both readings should be the same (100 to 10,000 ohms), regardless of meter lead polarity. Next, connect the positive meter lead to the gate. Using the negative lead, measure the resistance between the gate and the drain; then measure it between the gate and the source. Both readings should show infinity. Disconnect the positive lead from the gate and connect the negative lead to the gate. Using the positive lead, measure the resistance between the gate and the drain; then measure it between the gate and the source. Both readings should indicate a low resistance and be approximately equal.

MOSFET TESTING

Another type of semiconductor you should become familiar with is the metal oxide semiconductor field-effect transistor (MOSFET), as shown in figures 2-19 and 2-20. You must be extremely careful when working with MOSFETs because of their high degree of sensitivity to static voltages. As previously mentioned in this chapter, the soldering iron should be grounded. A metal plate should be placed on the workbench and grounded to the ship's hull through a 250-kilohm to 1-megohm resistor. You should also wear a bracelet with an attached ground strap and ground yourself to the ship's hull through a 250-kilohm to 1-megohm resistor. You should not allow a MOSFET to come into contact with your clothing, plastics, or cellophane-type materials. A vacuum plunger (solder sucker) must not be used because of the high electrostatic charges it can generate. Solder removal by wicking is recommended. It is also good practice to wrap MOSFETs in metal foil when they are out of a circuit. To ensure MOSFET safety under test, use a portable volt-ohm-milliammeter (vom) to make MOSFET resistance measurements. **A vtvm must never be used in testing MOSFETs.** You must be aware that while you are testing a MOSFET, you are grounded to the ship's hull or station's ground. Use of a vtvm would cause a definite safety hazard because of the 115-volt, 60-hertz power input. When the resistance measurements are complete and the MOSFET is properly stored, unground both the plate on the workbench and yourself. You will understand MOSFET testing better if you visualize it as equivalent to a circuit using diodes and resistors, as shown in figures 2-21 and 2-22.

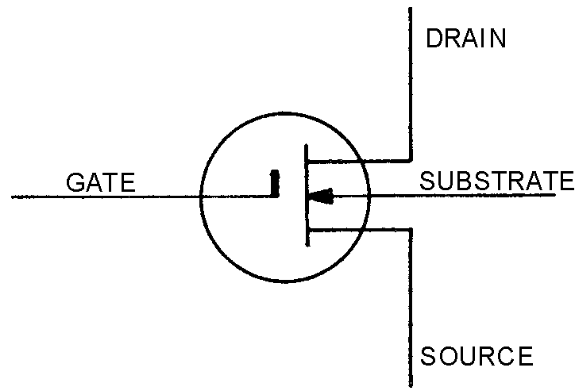


Figure 2-19.—MOSFET (depletion/enhancement type).

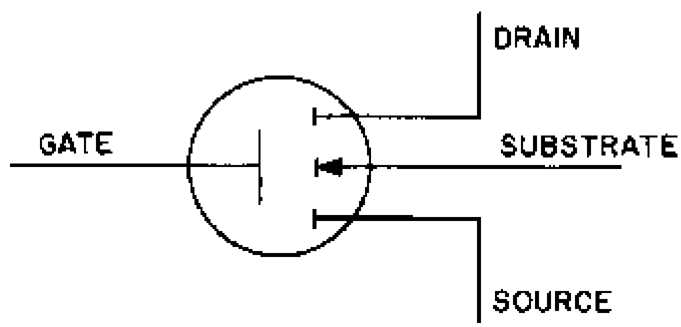


Figure 2-20.—MOSFET (enhancement type).

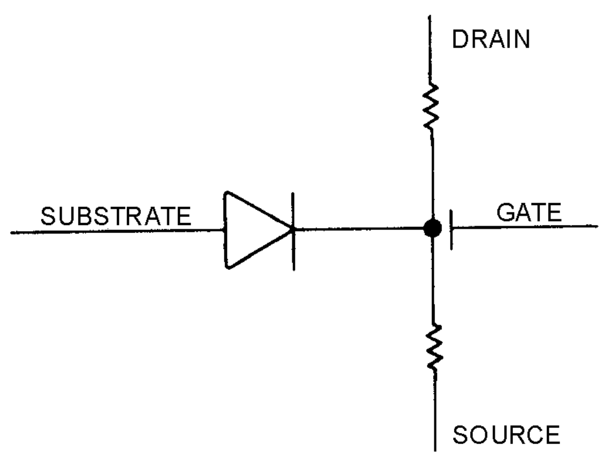


Figure 2-21.—MOSFET (depletion/enhancement type) equivalent circuit.

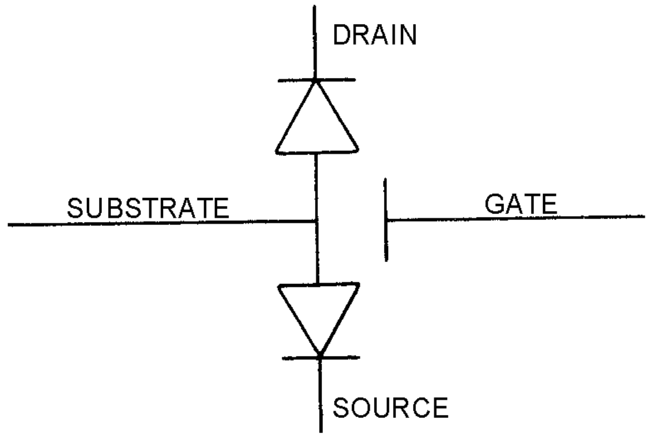


Figure 2-22.—MOSFET (enhancement type) equivalent circuit.

Q-17. Why is it not advisable to use a solder sucker when working on MOSFETs?

MOSFET (Depletion/Enhancement Type) Test

Using an ohmmeter set to the R × 100 scale, measure the resistance between the MOSFET drain and the source; then reverse the ohmmeter leads and take another reading. The readings should be equal, regardless of meter lead polarity. Connect the positive lead of the ohmmeter to the gate. Using the negative lead, measure the resistance between the gate and the drain and between the gate and the source. Both readings should show infinity. Disconnect the positive lead from the gate and connect the negative lead to the gate. Using the positive lead, measure the resistance between the gate and the drain; then measure it between the gate and the source. Both readings should show infinity. Disconnect the negative lead from the gate and connect it to the substrate. Using the positive lead, measure the resistance between the substrate and the drain and between the substrate and the source. Both of these readings should indicate infinity. Disconnect the negative lead from the substrate and connect the positive lead to the substrate. Using the negative lead, measure the resistance between the substrate and the drain and between the substrate and the source. Both readings should indicate a low resistance (about 1,000 ohms).

MOSFET (Enhancement Type) Test

Using an ohmmeter set to the R × 100 scale, measure the resistance between the drain and the source; then reverse the leads and take another reading between the drain and the source. Both readings should show infinity, regardless of meter lead polarity. Connect the positive lead of the ohmmeter to the gate. Using the negative lead, measure the resistance between the gate and the drain and then between the gate and the source. Both readings should indicate infinity. Disconnect the positive lead from the gate and connect the negative lead to the gate. Using the positive lead, measure the resistance between the gate and the drain and then between the gate and the source. Both readings should indicate infinity. Disconnect the negative lead from the gate and connect it to the substrate. Using the positive lead, measure the resistance between the substrate and the drain and between the substrate and the source. Both readings should indicate infinity. Disconnect the negative lead from the substrate and connect the positive lead to the substrate. Using the negative lead, measure the resistance between the substrate and the drain and between the substrate and the source. Both readings should indicate a low resistance (about 1,000 ohms).

INTEGRATED CIRCUIT (IC) TESTING

Integrated circuits (ICs) constitute an area of microelectronics in which many conventional electronic components are combined into high-density modules. Integrated circuits are made up of active and passive components, such as transistors, diodes, resistors, and capacitors. Because of their reduced size, use of integrated circuits can simplify otherwise complex systems by reducing the number of separate components and interconnections. Their use can also reduce power consumption, reduce the overall size of the equipment, and significantly lower the overall cost of the equipment concerned. Many types of integrated circuits are ESDS devices and should be handled accordingly.

Q-18. Name two advantages in using ICs.

Your IC testing approach needs to be somewhat different from that used in testing vacuum tubes and transistors. The physical construction of ICs is the prime reason for this different approach. The most frequently used ICs are manufactured with either 14 or 16 pins, all of which may be soldered directly into the circuit. It can be quite a job for you to unsolder all of these pins, even with the special tools designed for this purpose. After unsoldering all of the pins, you then have the tedious job of cleaning and straightening all of them.

Although there are a few IC testers on the market, their applications are limited. Just as transistors must be removed from the circuit to be tested, some ICs must also be removed to permit testing. When ICs are used in conjunction with external components, the external components should first be checked for proper operation. This is particularly important in linear applications where a change in the feedback of a circuit can adversely affect operating characteristics of the component.

Any linear (analog) IC is sensitive to its supply voltage. This is especially the case among ICs that use bias and control voltages in addition to a supply voltage. If you suspect a linear IC of being defective, all voltages coming to the IC must be checked against the manufacturer's circuit diagram of the equipment for any special notes on voltages. The manufacturer's handbook will also give you recommended voltages for any particular IC.

When troubleshooting ICs (either digital or linear), you cannot be concerned with what is going on inside the IC. You cannot take measurements or conduct repairs inside the IC. You should, therefore, consider the IC as a black box that performs a certain function. You can check the IC, however, to see that it can perform its design functions. After you check static voltages and external components associated with the IC, you can check it for dynamic operation. If it is intended to function as an amplifier, then you can measure and evaluate its input and output. If it is to function as a logic gate or combination of gates, it is relatively easy for you to determine what inputs are required to achieve a desired high or low output. Examples of different types of ICs are provided in figure 2-23.

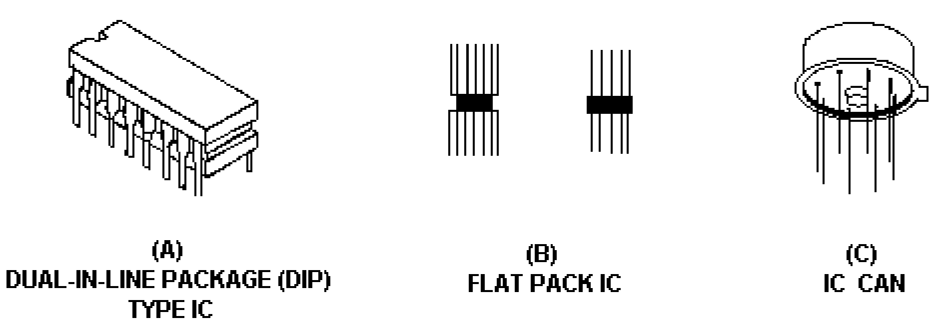


Figure 2-23.—Types of ICs.

Q-19. Why should you consider an IC as a black box?

Digital ICs are relatively easy for you to troubleshoot and test because of the limited numbers of input/output combinations involved. When using positive logic, the logic state of the inputs and outputs of a digital IC can only be represented as either a high (also referred to as a 1 state) or as a low (also referred to as a 0 state). In most digital circuitry, a high is a steady 5-vdc level, and a low is a 0-vdc level. You can readily determine the logic state of an IC by using high-input-impedence measuring devices, such as an oscilloscope. Because of the increased use of ICs in recent years, numerous pieces of test equipment have been designed specifically for testing ICs. They are described in the following paragraphs.

Q-20. What are the two logic states of an IC?

LOGIC CLIPS

Logic clips, as shown in figure 2-24, are spring-loaded devices that are designed to clip onto a dual-in-line package IC while the IC is mounted in its circuit. It is a simple device that usually has 16 light emitting diodes (LEDs) mounted at the top of the clips. The LEDs correspond to the individual pins of the IC, and any lit LED represents a high logic state. An unlit LED represents a low logic state. Logic clips require no external power connections, and they are small and lightweight. Their ability to simultaneously monitor the input and output of an IC is very helpful when you are troubleshooting a logic circuit.

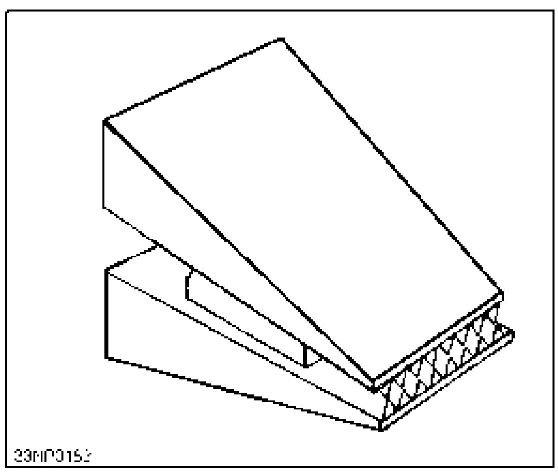


Figure 2-24.—Logic clip.

Q-21. A lighted LED on a logic clip represents what logic level?

LOGIC COMPARATORS

The logic comparator, as shown in figure 2-25, is designed to detect faulty, in-circuit-DIP ICs by comparing them with ICs that are known to be good (reference ICs). The reference IC is mounted on a small printed-circuit board and inserted into the logic comparator. You then attach the logic comparator to the IC under test by a test lead, which is connected to a spring-loaded device similar in appearance to a logic clip. The logic comparator is designed to detect differences in logic states of the reference IC and the IC being tested. If any difference in logic states does exist on any pin, an LED corresponding to the pin in question will be lit on the logic comparator. The logic comparator is powered by the IC under test.

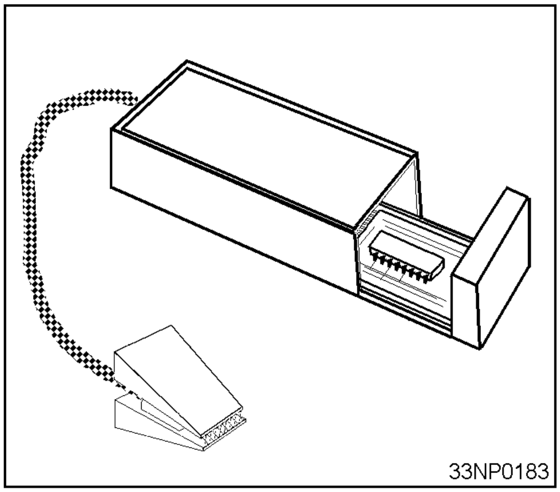


Figure 2-25.—Logic comparator.

Q-22. What does a lighted LED indicate on a logic comparator?

LOGIC PROBES

Logic probes, as shown in figure 2-26, are extremely simple and useful devices that are designed to help you detect the logic state of an IC. Logic probes can show you immediately whether a specific point in the circuit is **low, high, open,** or **pulsing**. A **high** is indicated when the light at the end of the probe is lit and a **low** is indicated when the light is extinguished. Some probes have a feature that detects and displays high-speed transient pulses as small as 5 nanoseconds wide. These probes are usually connected directly to the power supply of the device being tested, although a few also have internal batteries. Since most IC failures show up as a point in the circuit stuck either at a **high** or **low** level, these probes provide a quick, inexpensive way for you to locate the fault. They can also display that single, short-duration pulse that is so hard to catch on an oscilloscope. The ideal logic probe will have the following characteristics:

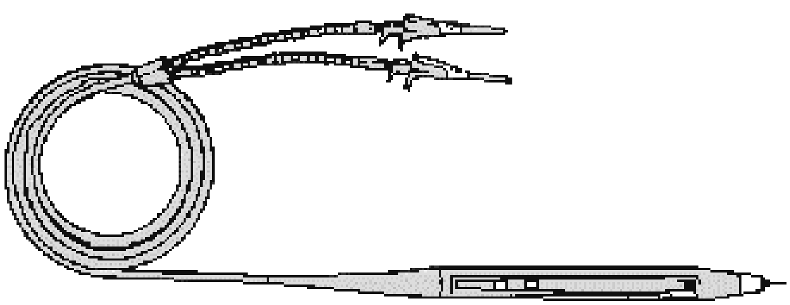


Figure 2-26.—Logic probe.

1. Be able to detect a steady logic level
2. Be able to detect a train of logic levels
3. Be able to detect an open circuit
4. Be able to detect a high-speed transient pulse

5. Have overvoltage protection

6. Be small, light, and easy to handle

7. Have a high input impedance to protect against circuit loading

Q-23. What is the purpose of a logic probe?

LOGIC PULSERS

Another extremely useful device for troubleshooting logic circuits is the logic pulser. It is similar in shape to the logic probe and is designed to inject a logic pulse into the circuit under test. Logic pursers are generally used in conjunction with a logic clip or a logic probe to help you trace the pulse through the circuit under test or verify the proper operation of an IC. Some logic pursers have a feature that allows a single pulse injection or a train of pulses. Logic pursers are usually powered by an external dc power supply but may, in some cases, be connected directly to the power supply of the device under test. View A of figure 2-27 shows a typical logic pulser. View B shows a logic pulser (right) used with a logic probe (left).

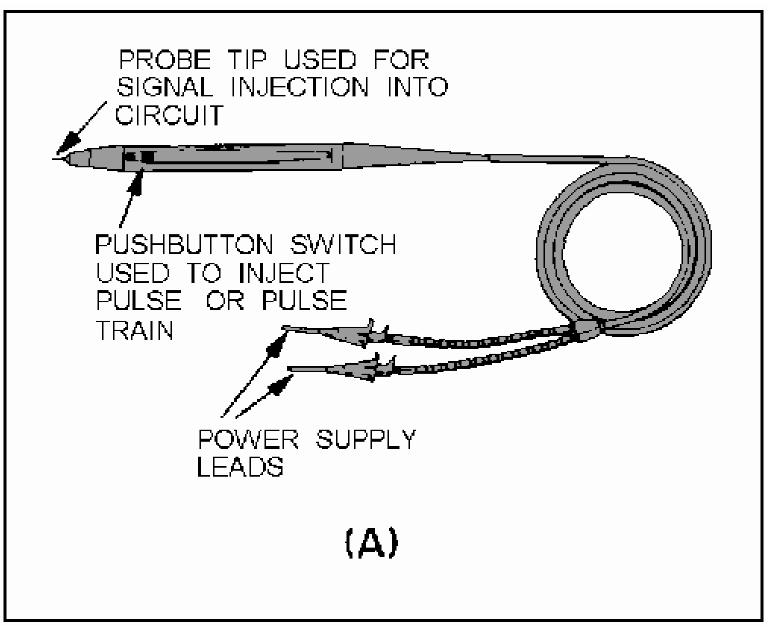


Figure 2-27A.—Logic pulser.

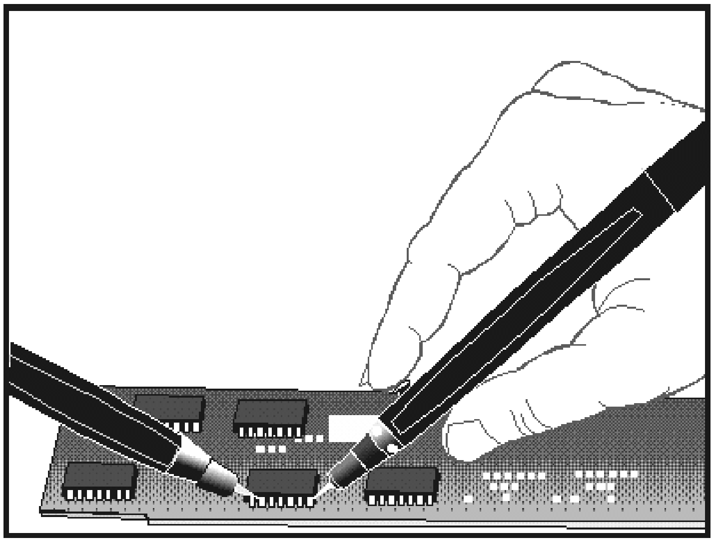


Figure 2-27B.—Logic pulser.

LOGIC ANALYZER

A relatively new device on the test equipment scene is the logic analyzer. A logic analyzer provides various functions that can assist you in maintenance, testing, and troubleshooting of equipment using digital circuitry. From your standpoint, they are extremely useful in performing timing analysis. Most logic analyzers have crt displays that can monitor up to 32 timing signals at the same time. A large percentage of today's digital equipment is designed with the logic analyzer in mind and have built-in status or bus lines for your convenience in monitoring multiple signals at the same time. When monitoring a bus line, you can readily determine, through visual displays, such things as the presence of master clock signals or sequential timing events.

BATTERY MEASUREMENTS

As a technician, you are primarily concerned with the *uses* of batteries; however, checking or testing of storage and dry cell batteries is an important part of your maintenance program. Proper preventive maintenance of batteries can significantly extend the useful life of a battery.

STORAGE BATTERIES

When you check a lead-acid type of storage battery for its condition of charge or discharge, you take a specific gravity reading of the electrolyte by using a hydrometer. A specific gravity reading between 1.275 and 1.300 indicates a full-charge condition and assures you that the battery is in good condition. A hydrometer reading of approximately 1.175 indicates a normal discharge condition, and a reading of approximately 1.250 indicates that the battery is half-discharged. Since the acids used in various batteries do not always have the same specific gravity and since electrode composition may differ, the hydrometer reading you obtain at the charged and discharged conditions will vary with the type of electrolyte and battery composition. A general rule for you to follow is not to discharge a battery more than 100 points (.100 specific gravity) before recharging.

Although readings of specific gravity are a reliable measure of the condition of a storage battery, cells that indicate normal may prove useless under load. This is usually caused by a high internal resistance. A load-voltage check of the cells with the use of a cell tester indicates the actual voltage charge held by each battery cell. Cell voltages should not differ by more than 0.15 volt for 6-volt or 12-volt batteries.

Use extreme caution whenever testing or working around lead-acid storage batteries. OPNAVINST 5100.23B emphatically states that you must wear eye protection devices at all times and that emergency eyewash facilities must be immediately adjacent to, or within 10 feet of, any eye-hazard area. Smoking and spark-producing tools or devices are also prohibited in enclosed spaces that contain lead-acid storage batteries. When charging, these batteries produce sufficient quantities of hydrogen to produce large explosions. **Lead-acid storage batteries should only be charged in well-ventilated spaces.**

Q-24. Emergency eyewash facilities must be located within what minimum number of feet of an eye-hazard area?

DRY BATTERIES

You must periodically check dry cell batteries that are used for test instruments and portable or field equipments for loss of power. For actual voltages of dry batteries, you should measure with a battery tester for a minimum acceptable voltage before installation. The TS-183/U series of battery testers incorporate a multiple-range voltmeter, battery-loading resistors, multiplier resistors, and a jack-switching arrangement that connects the load resistors across the voltmeter for a total of 32 different voltmeter-load resistor combinations. This type of tester permits you to complete a rapid and accurate measurement of battery potentials under load conditions, ranging in voltages from 1.5 to 180 volts. A data chart supplied with the battery tester provides information regarding the jack to be used and minimum acceptable voltages of various batteries used in Navy equipments.

Q-25. What is the advantage of using a battery test set versus a voltmeter to test batteries?

Table 2-2 shows general standards of tolerance for dry batteries. Whenever practical, dry cell batteries that are not in use should be stored in a refrigerated area to extend their shelf life.

Table 2-2.—Typical voltage Tolerances for Dry Cell Batteries

RATED VOLTAGE	MAX. VOLTAGE TOLERANCE
1 to 2	0.1
3 to 10	0.3
11 to 15	0.5
16 to 25	1.0
26 to 50	2.0
50 to 70	3.0
70 to 99	5.0

CARBON-ZINC AND ALKALINE BATTERIES

Carbon-zinc and alkaline cells are used primarily in portable test equipment, vom's, flashlights, some portable radios, and beacon equipment. The carbon-zinc cell provides 1.5 volts and holds its charge for approximately 1 year in normal service. The alkaline cell provides 1.2 volts and has about twice the stored energy of the carbon-zinc cell of the same size. It also has a longer life at a higher discharge rate than the carbon-zinc cell. You should discard both types of batteries at the first indication of weakness.

MERCURY CELLS

The storage life of a mercury cell varies but is generally classified as *long*. The working life of the cell is extremely long relative to other types of batteries; and it maintains its full rated voltage (1.34 volts) until just before it is ready to expire, at which point its voltage will drop off sharply. Recharging of mercury cells is possible, but is not recommended because the recharging cycle can vary from one cell to another; and, after being recharged, their operating lifetime is uncertain.

NICKEL-CADMIUM BATTERIES (NICAD)

Nickel-cadmium batteries have very high efficiency. They can be recharged hundreds of times; given the proper conditions, they may even be recharged thousands of times. They can be stored for a number of years with no significant loss of performance. After just a few charge and discharge cycles, NICAD cells can be recharged to the point that they are just as good as new batteries. Since they are sealed, they are maintenance free and can be installed in any position. There are two types of nickel-cadmium batteries — vented and nonvented. This description deals with the nonvented exclusively because a vented NICAD would have extremely limited application in a shipboard environment.

The voltage at the terminals of a NICAD will normally be between 1.25 and 1.30 volts in an open-circuit condition. This value will vary, of course, depending on the state of charge. If the charge has dropped to a low of 1.1 volts, the NICAD should be regarded as being completely discharged and should not be permitted to be discharged further. The majority of small NICADs are rated in milliampere hours; the large ones are rated in ampere hours. The small NICAD is the one the technician will almost always be concerned with.

Q-26. At what voltage is a NICAD battery considered to be fully discharged?

As a general rule, if the charging current is held to 10% of the milliampere-hour rating for the NICAD and the time of charge is held at 150% of the time required to establish its full milliampere-hour rating, you will encounter no difficulty in maintaining NICADs at their maximum charge. For example, you should charge a battery rated at 300 milliampere hours for 15 hours at 30 milliamperes. You can leave the battery on extended charge for years, provided the charge rate is lowered to less than 10% of the NICAD's milliampere rating.

You should never place a NICAD in your pocket, because metal objects (such as keys) can short the cell and cause extreme heat. Never dispose of a NICAD by fire, because it can explode. Never solder a connection directly to the cell, because the heat of an iron can damage it. Never overcharge a NICAD cell, because an accumulation of gases within its case can destroy it.

NICADs are also subject to a phenomenon commonly referred to as cell *memory*. If a NICAD is consistently discharged to a minor extent (for example, 30 minutes per day) and then recharged after each use, the useful capacity of the cell will eventually be reduced to that level. To keep this from happening, you should fully discharge (1.1 volts) NICADs on a regular basis. In fact, some maintenance requirement cards and calibration laboratory procedures require this periodic full discharge of equipment containing NICADs.

RF ATTENUATORS AND RESISTIVE LOAD TESTS

All rf attenuators, decade or step attenuators, decade resistors, and 50/75-ohm loads are clearly marked to show their attenuation factor or resistance. In the case of precision rf attenuators, they are usually marked to show their useful frequency ranges. They are all basically resistive devices and are designed for a multitude of applications. None of these devices are user-repairable; however, you should be aware of the different methods of determining whether or not they are functioning properly.

FIXED RF ATTENUATORS

Fixed rf attenuators (shown in fig. 2-28), such as the ones commonly found in power-measuring sets, are designed to provide a fixed-signal attenuation over a specific frequency range. Frequency ranges can be in excess of 30 gigahertz, and attenuation factors are typically in 1-, 3-, 6-, and 10-dB steps. Fixed attenuators can be connected in series to provide you with the desired attenuation. Most fixed rf attenuators are designed to handle only small amounts of rf power and are extremely susceptible to damage because of overloading. To test a fixed rf attenuator, you can either substitute it with a known good attenuator or perform basic measurements on the attenuator itself. With the rf substitution method, you connect an rf signal generator to a power meter and establish a suitable reference point on the meter by adjusting the power output of the signal generator. Once you establish the reference point, insert the rf attenuator between the signal generator and the power meter. You then determine the attenuation by noting the difference between the power meter reading and the initial reference point.

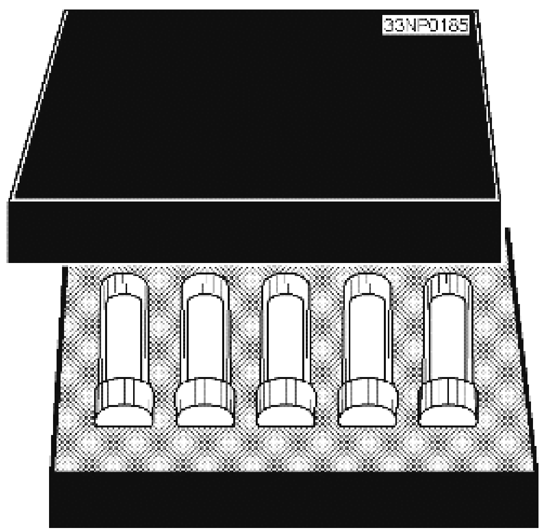


Figure 2-28.—Fixed attenuator set.

Q-27. What is the most common method of testing a fixed rf attenuator?

DECADE RESISTORS

Decade resistors (also referred to as decade boxes) typically are precision devices. Depending on the make and model of the decade resistor, it may be capable of providing you with a selection of resistors ranging in value from a small fraction of an ohm to hundreds of megohms. Decade resistors are commonly used in calibration laboratories and in engineering design applications. Like the fixed rf attenuator, most decade resistors are capable of handling only small amounts of current. They are very limited in respect to frequency capabilities and are commonly used in dc-circuit applications. You may encounter specific equipment that requires the use of a decade resistor in performing your maintenance tests or alignments. To test a decade resistor, you can connect a standard multimeter or digital multimeter directly across its resistance terminals and read its resistance on the meter. This test will only indicate gross errors in the decade resistor such as an open or a badly damaged resistor. If you are performing a precision measurement or an alignment using a decade resistor and have any doubt as to its accuracy, you should submit it to your servicing calibration laboratory. Figure 2-29 shows a typical decade resistor.

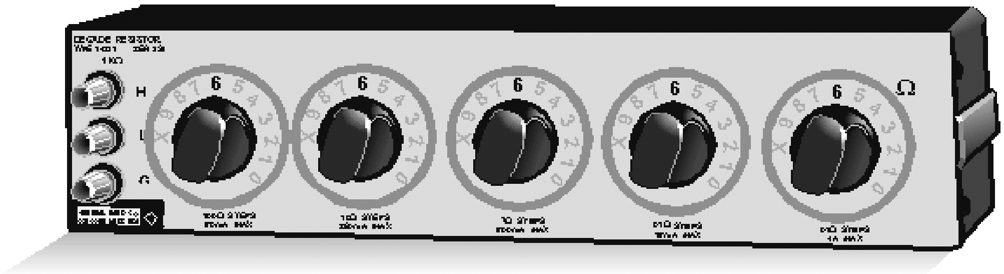


Figure 2-29.—Decade resistor.

DECADE (STEP) ATTENUATORS

Decade attenuators (also referred to as step attenuators) are common devices that may be designed as either a stand-alone piece of test equipment or as an integral part of an operational piece of electronic equipment. As the name implies, they are used to attenuate rf signals in incremental steps. Like the fixed rf attenuator, you can easily test them by using the rf substitution method, as previously described. Views A and B of figure 2-30 show two types of decade attenuators.

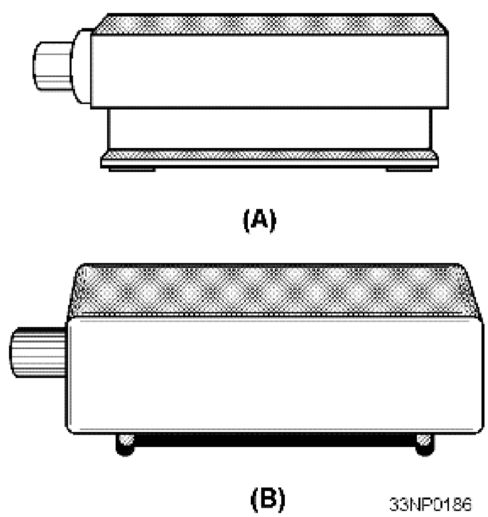


Figure 2-30.—Step attenuators.

50/75-OHM TERMINATIONS

Terminations of 50 and 75 ohms are designed as either feedthrough, impedance-matching devices, or as rf loading devices. They are precision resistors sealed in small plastic or metal enclosures and are designed to be mounted on various rf connectors. In the case of feedthrough terminations, they are designed with rf connectors at both ends, which allows the rf signal to pass through them. They are impedance-matching devices designed primarily to reduce the voltage standing-wave ratio (vswr) that is produced when two pieces of equipment with dissimilar impedances are connected together.

You can test a feedthrough termination by measuring the resistance between the center conductor and the shield of either rf connector with an ohmmeter. As mentioned above, some terminations are manufactured as loading devices that are designed to shunt an rf signal to ground. A perfectly matched termination can be compared to a transmitting antenna in that it absorbs all of the rf signal with only a small amount of power being reflected back to the transmitting device. When using a termination as a load, you should ensure that its wattage rating exceeds the power output of the equipment to which it is connected. You can also measure this type of termination by using a standard ohmmeter to read the resistance between the center conductor and the shield of the rf connector.

Q-28. What is the most common method of testing resistive terminations?

FIBER-OPTIC TESTING

Fiber optics are a relatively new type of transmission media. Figure 2-31 depicts a typical fiber-optic cable design. The core of the fiber-optic cable is the optical transmission path, which carries data from the optical transmitter to the optical receiver. The core is usually made of plastic, glass, or plastic-clad silica (PCS). Glass-core fibers are usually smaller in diameter than plastic or PCS cores. The major disadvantages of glass cores are that they have high attenuation (25 dB/km), require precision tools and connectors, and are extremely susceptible to mechanical damage. Plastic cores are typically more rugged than other types of cores, but their attenuation is high (35 dB/km). PCS cores are fairly rugged and have a relatively low attenuation (10 dB/km). A fiber-optic cable may consist of one fiber, multiples of single-optical fibers, or bundles of optical fibers. Fiber-optic cables are well suited for the transmission of high-speed data over relatively short distances. They are virtually immune to crosstalk or interference through inductance. (Interference is a characteristic of metallic cables.)

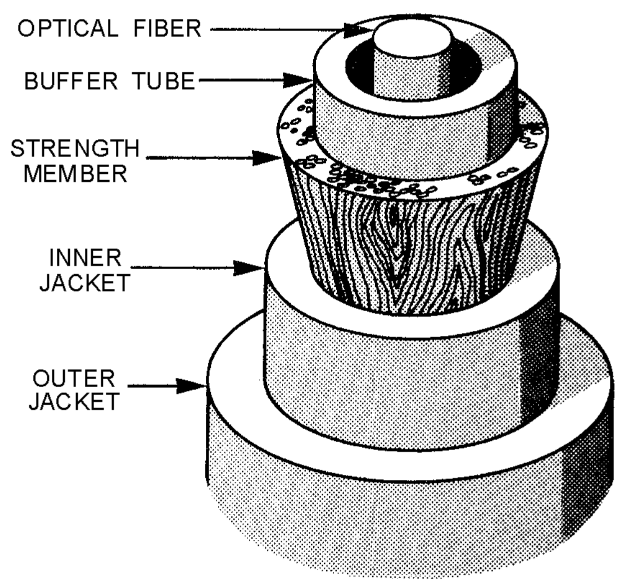


Figure 2-31.—Typical fiber-optic cable.

Testing techniques and the principles of measurement for fiber-optic and conventional cable are similar. For example, if both ends of the cable are exposed and can be used for testing, relatively unsophisticated equipment can be used to measure cable parameters, such as continuity and attenuation. This includes equipment such as optical multimeters and optical power meters (OPM). If only one cable end is available, then more sophisticated equipment such as an optical time-domain reflectometer (OTDR), is used. The following section lists and defines some common optical test equipment.

Q-29. What is the main disadvantage of using fiber-optic cables?

OPTICAL TIME-DOMAIN REFLECTOMETER (OTDR)

The portable optical time-domain reflectometer (OTDR) is used to check loss at each splice, at each connector, and of the entire system. Loss measurements are figured by using the same methods you would use for wire loss measurements. The OTDR injects a short, intense laser pulse into the fiber and monitors reflections caused by breaks, inclusions, microcracks, and discontinuities. Discontinuities appear as a spike on the OTDR display. The loss at the discontinuity point is directly related to the distance between the major pulse triggered by the laser and the spike. The manufacturer's manual provides you with conversion factors to figure actual losses and locations of the discontinuities.

OSCILLOSCOPE

An oscilloscope is used with an OTDR to provide visual evidence of fiber faults, connector and splice locations, and attenuation locations.

OPTICAL MULTIMETER

The optical multimeter measures light sources and light in cable and at the detector, fiber cable transmission loss, and connector splice loss. For cable transmission measurements, transmission through a short length of cable is compared with transmission through a known longer length.

OPTICAL OHMMETER

The optical ohmmeter measures the input versus the output of light in an optical fiber. It displays attenuation losses based on a comparison of known and unknown cable signals. It can be used in manufacturing, connecting, and installing cable. It is as simple to use as a digital voltmeter.

OPTICAL POWER METER

The optical power meter measures current by converting light power from plug-in units, such as light emitting diodes, into electrical current. In some models, the readout is in power units, watts. In other models, the readout is in absolute power levels and attenuation. Some units operate with a variety of power sensors for conventional coaxial and waveguide systems and fiber-optic systems.

RADIOMETER/PHOTOMETER

The radiometer/photometer measures light power in watts from dc to unlimited ac response. It uses plug-in sensor heads and, for low-light displays, it uses spectrometers and fiber-optic measurements.

AUTOMATIC TEST EQUIPMENT

Automatic Test Equipment (ATE) is test equipment designed to evaluate the operational performance of a piece of equipment or printed circuit board (pcb). ATE assists you in troubleshooting a fault to the defective component. Basically, ATEs are state-of-the-art, computer devices in which software programs are specifically tailored to meet the requirements of the device being tested.

The AN/USM-465 Portable Service Processor (psp), shown in figure 2-32, is the Navy's standard ATE for testing digital pcb's.

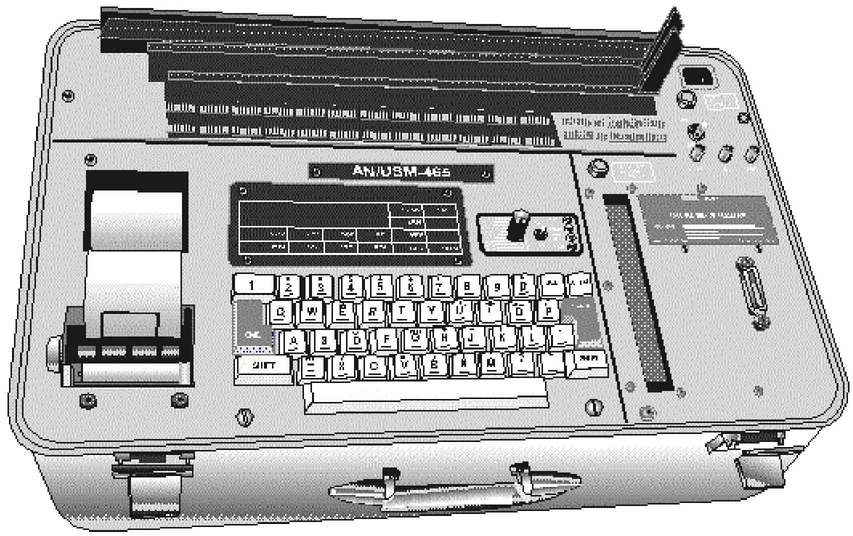


Figure 2-32.—AN/USM-465 Portable Service Processor.

The AN/USM-465 is part of the Support and Test Equipment Engineering Program (STEEP). It provides on-site screen testing and fault isolation of digital pcb's and modules. The psp is presently available on most ships and shore intermediate maintenance activities (SIMA) with Mini/Micro maintenance stations (2M). Psp's come with maintenance-assist modules (spare parts kit) and diagnostic kits.

The psp is easy to use. You have a choice of three pcb connectors (located on the top panel of the test set) into which you insert the pcb being tested. The software program, which is provided on magnetic tape cartridges, is then loaded into the test set. The test set automatically tests the pcb by applying input signals to the appropriate pins while monitoring the output signal for a correct indication. An LED display will give you a pass or fail indication. If a pcb fails the operational test, the psp tells you (via LED display) what troubleshooting steps must be taken. The psp uses a *guided probe* fault isolation technique that tells you what test points to check on the faulty pcb. The software program guides you from the faulty output backwards toward the input until the fault is located. The probe is a standard 10 megohm, 10 to 1 oscilloscope probe. The guided probe circuitry and software is also unique because it is capable of locating faults within feedback loops and can sense when you have placed the probe at an incorrect test point.

An interesting advantage is that if the psp itself fails, the faulty board inside the psp can be identified by the test set's own capability. After you replace the faulty pcb with a good one from the spare parts kit, you can use the psp to identify the faulty component on its own pcb.

HUNTRON TRACKER 2000

The Huntron Tracker 2000, shown in figure 2-33, is a versatile troubleshooting tool used to statically test resistors, capacitors, inductors, diodes, transistors, multiple-component circuits, and integrated circuits. Its built-in features eliminate the use of multiple pieces of test equipment. These features and its lightweight portability make the 2000 a widely used tool for troubleshooting.

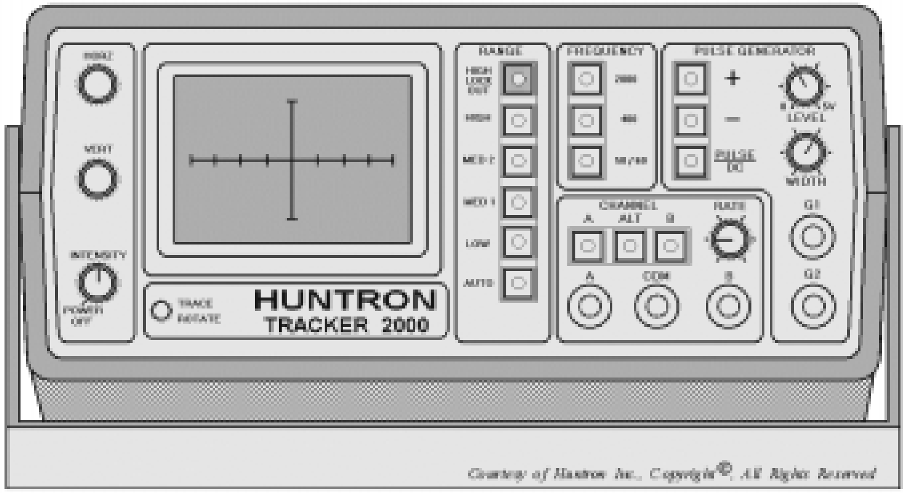


Figure 2-33.—Huntron Tracker 2000.

We recommend you review setup and operating procedures discussed in NEETS Module 16, *Introduction to Test Equipment*, NAVEDTRA B72-16-00-95, before continuing with this chapter. Since the 2000 was covered in depth in module 16, we will cover only the most common troubleshooting procedures and provide a few troubleshooting tips.

Q-30. What two features make the Huntron Tracker 2000 a widely used troubleshooting tool?

The Huntron Tracker 2000 has the following features:

- Multiple-test signal frequencies (2000 Hz, 400 Hz, and 50/60 Hz).
- Four impedance ranges (low, medium 1, medium 2, high).
- Automatic range scanning.
- Range control: High Lockout.
- Rate-of-channel alteration and/or range scanning is/are adjustable.
- Dual-polarity pulse generator for dynamic testing of three terminal devices.
- LED indicators for all functions.
- Dual-channel capability for easy comparison.
- Large CRT display with easy-to-operate controls.

CAUTION

The device to be tested must have all power turned off, and have all high voltage capacitors discharged before connecting the Tracker 2000 to the device.

Testing Components by Comparison

Testing components by comparison is the most preferred method for troubleshooting. The ALT (alternate) mode setup is the most commonly used mode for this method. This mode allows the technician to compare a known good component to a suspect component. This is accomplished by connecting channel A to a known good device, channel B to the device under test, and a common test lead to COM as illustrated in figure 2-34. Select the ALT button, and the 2000 will alternately display the signature of the known good device and the device under test. By examining the signature differences, you can detect a defective component. Figure 2-35 is a typical example of the CRT display on the 2000 while testing the base to emitter on a good transistor. Figure 2-36 illustrates a defective transistor under the same test setup. Note that in the low range, the transistor appears to be good. Sometimes component defects are more obvious in one range than another, so is a suspect device appears normal for one range, try the other ranges.

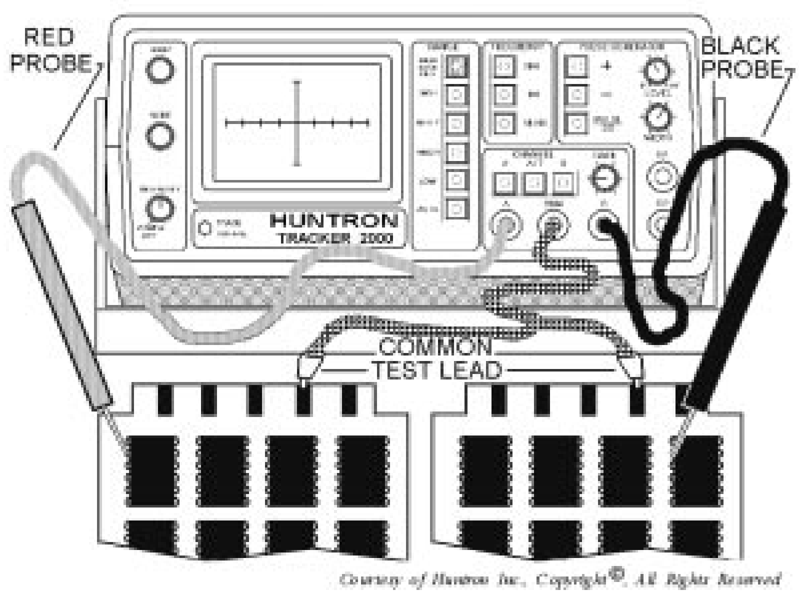


Figure 2-34.—Alternate mode setup.

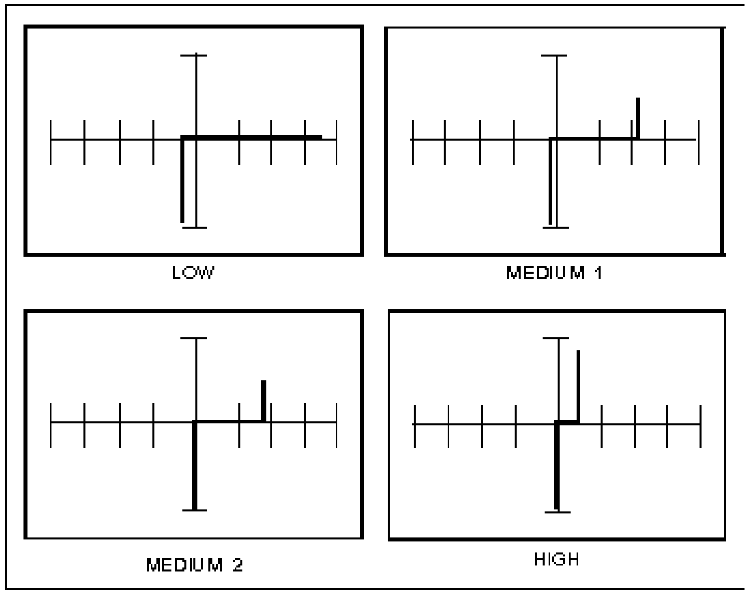


Figure 2-35.—Signatures between base-emitter of a good transistor.

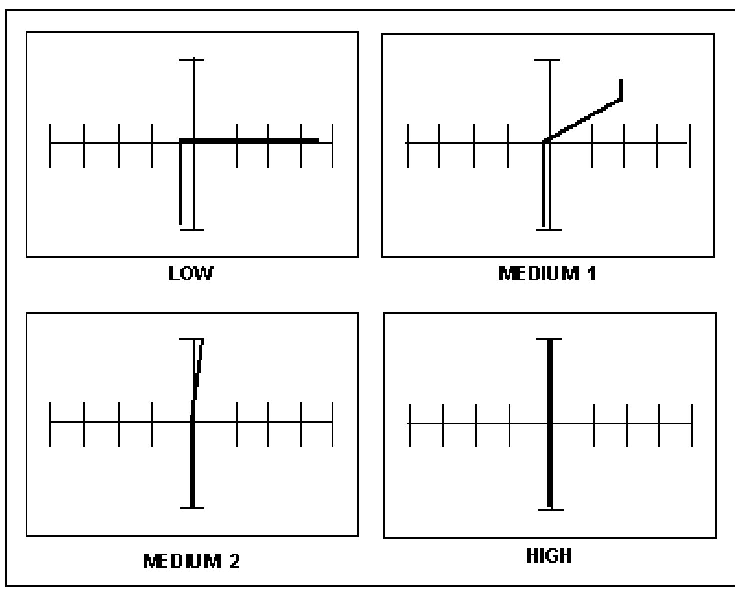


Figure 2-36.—Signatures between base-emitter of a defective transistor.

Q-31. What is the most preferred method of troubleshooting?

Q-32. Why is it recommended to use more than one range while troubleshooting a device?

Troubleshooting Tips

When you are testing individual components in a circuit, a parallel resistor or diode of similar value may cause a defective component to appear good. Therefore, you should, in most cases, electrically isolate the suspected component from the circuit while testing individual components. The best way to do this is to desolder all but one lead on the suspected component.

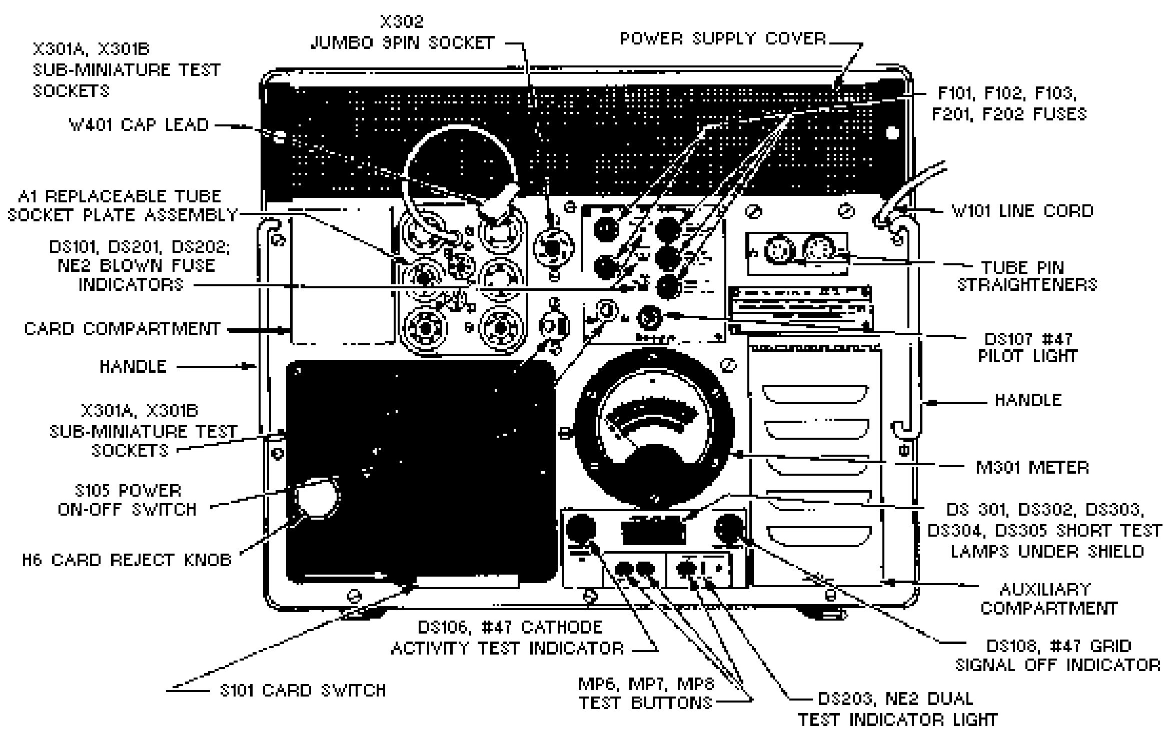
Q-33. When you are testing individual components in a circuit, what may cause a defective component to appear good?

You should be aware that devices made by different manufacturers may appear to have slightly different signatures. This is normal, especially with digital integrated circuits, and does not necessarily indicate a failed device. When this occurs, the best way to verify this is to compare the outputs of the device under test with the equipment specifications to ensure the signals are adequate for proper equipment operation.

SUMMARY

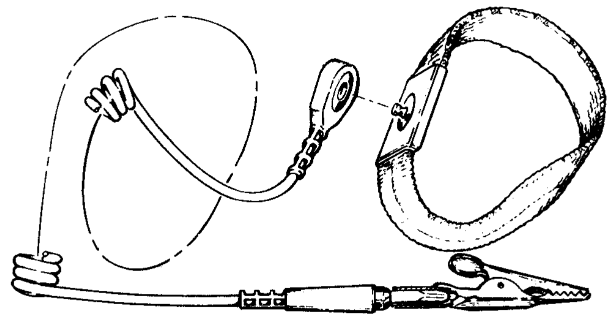
The information that follows summarizes the important points of this chapter.

ELECTRON TUBES are usually tested for **SHORTS, TRANSCONDUCTANCE**, and the presence of GAS. Several different types of tubes (i.e., twt's, magnetrons, and klystrons) are normally tested in-circuit.



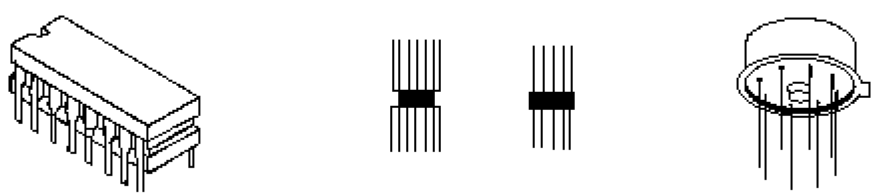
Most **TRANSISTORS** can be tested by measuring the forward-to-back resistance of their junctions using a standard ohmmeter. The resistance scale of the ohmmeter must be carefully selected to ensure that the current rating of the transistor is not exceeded.

ESD-SENSITIVE DEVICES are components that require special handling. Some of the more sensitive devices can be damaged by static charges as small as 35 volts.



Most **DIODES** and **MOSFETs** can be tested by measuring the forward-to-back resistance of their junctions using a standard ohmmeter. MOSFETS, however, are classed as ESD-sensitive devices; and care should be exercised when handling or testing them.

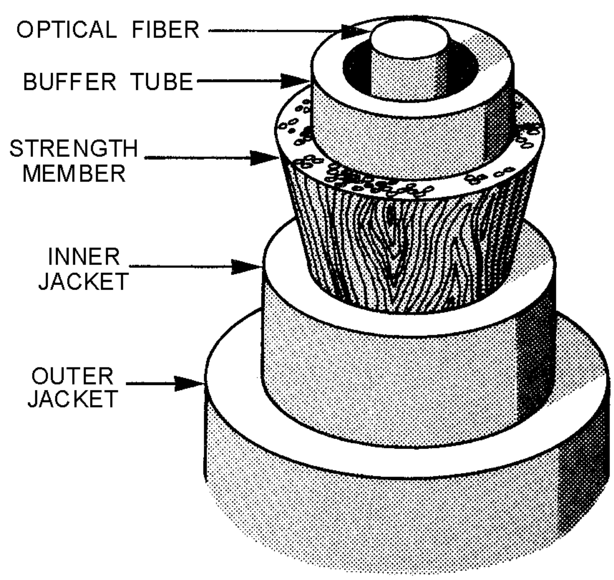
INTEGRATED CIRCUITS (ICs) have revolutionized the electronics industry. They are rugged, compact, and inexpensive. There is a wide assortment of equipment on the market designed for testing ICs.



BATTERIES are common to a large number of both electronic test equipment and operational equipment. You should be familiar with the different types of batteries, their test requirements, and the safety precautions to be followed.

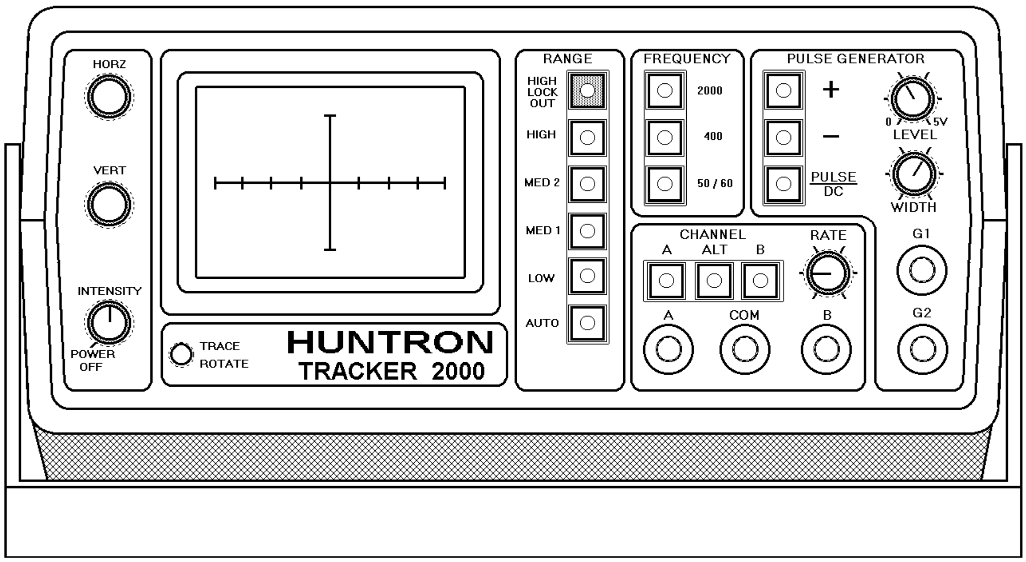
RF ATTENUATORS and **RESISTIVE LOADS** are common devices that are widely used for attenuating rf signals and impedance matching. Resistive loads can be tested with a standard ohmmeter, and rf attenuators are normally tested through the rf substitution method.

FIBER-OPTIC CABLES are used primarily for the transmission of high-speed data over short distances. Their construction and theory of operation require that they be tested with a light source, usually a laser beam. There is a wide assortment of test equipment designed specifically for testing fiber-optic cables.



AUTOMATIC TEST EQUIPMENT (ATE) is test equipment designed to evaluate the operational performance of a piece of equipment or printed circuit board (pcb).

The **HUNTRON TRACKER 2000** is a versatile troubleshooting tool commonly used for statically testing resistors, capacitors, inductors, diodes, transistors, multiple-component circuits, and integrated circuits.



REFERENCES

EIMB, *Test Methods and Practices*, NAVSEA 0967-LP-000-0130, Naval Sea Systems Command, Washington, D.C., 1980.

Fiber Optic Communication Cables and Connectors, (Navy) EE169-CA-GYD-010/E110 TSER E & I, (Published under the authority of the Secretaries of the Air Force, Army, and Navy), 1983.

Fire Control Technician G 3 & 2, NAVEDTRA 10207-B, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1981.

Huntron Tracker 2000 Operation and Maintenance Manual, P/N 21-1052, Huntron Instruments, Inc., 15720 Mill Creek Blvd., Mill Creek, WA 98012.

Introduction to Microelectronics, NAVEDTRA 172-14-00-84, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1984.

Logic Clip 548A, NAVAIR 16-45-3102, Naval Air Systems Command, Washington, D.C., 1979.

Logic Comparator 10529A, NAVAIR 16-45-3100, Naval Air Systems Command, Washington, D.C., 1979.

Logic Probe 545A, NAVAIR 16-45-3105, Naval Air Systems Command, Washington, D.C., 1979.

Logic Pulser 546A, NAVAIR 16-45-3104, Naval Air Systems Command, Washington, D.C., 1979.

ANSWERS TO QUESTIONS Q1. THROUGH Q33.

A-1. Lack of adequate storage space.

A-2. Open filaments.

A-3. Testing the tube in its circuit.

A-4. In their circuit.

A-5. Restore it to serviceable condition by operating it temporarily at reduced beam voltage.

A-6. Correct gain figure.

A-7. Rugged design.

A-8. Sensitive to heat and minor overloads.

A-9. Any range setting that produces a current flow through the transistor that exceeds 1 milliamp (usually R x 1 range).

A-10. 3,500 to 4,000 volts.

A-11. 35 volts.

A-12. For your own safety.

A-13. Voltages and resistances.

A-14. Greater than 10 to 1.

A-15. Gate and anode.

A-16. Current is allowed to flow in either direction.

A-17. Solder suckers create an electrostatic charge capable of damaging a MOSFET.

A-18. Low power consumption, compact size, and lower cost.

A-19. ICs cannot be repaired. All you need to test is output versus input.

A-20. A "1" or "0."

A-21. A "1" state.

A-22. A difference in logic states between the reference IC and the IC under test.

A-23. They provide you with a visual indication of the logic state at any point you choose in the circuit.

A-24. 10 feet.

A-25. A battery test set will test batteries under load conditions.

A-26. At 1.1 volts.

A-27. Rf substitution method.

A-28. Reading their resistances with a standard ohmmeter.

A-29. High attenuation.

A-30. It eliminates the need for multiple pieces of test equipment and it is lightweight and portable.

A-31. Testing components by comparison.

A-32. Some defective devices may appear to be good in certain ranges.

A-33. A parallel resistor or diode of similar value.

CHAPTER 3

QUANTITATIVE MEASUREMENTS

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to do the following:

1. Explain the purposes and benefits of performing quantitative measurements.
2. Identify the various methods of performing impedance measurements.
3. Identify the various methods of performing power measurements.
4. Identify the various methods of performing frequency measurements.

INTRODUCTION TO QUANTITATIVE MEASUREMENTS

You have already studied the basics of performing electronics measurements and how to determine if a component is or is not functioning properly. This chapter will cover techniques used in measurements of specific impedance, frequency, and power. These measurements are extremely important to you in evaluating the performance of a piece of electronic equipment.

IMPEDANCE MEASUREMENTS

Impedance measurements are often used during routine test procedures. Impedance-measuring equipment, such as impedance bridges, are mainly used in determining the capacitance and inductance of component parts. However, the values of combined circuit constants also may be obtained and used in direct calculations of impedance. An impedance measurement effectively totals the inductive and capacitive reactance together with the resistance in a circuit. In addition, impedance measurements are useful in testing and analyzing antenna and transmission line performance and for determining the figure of merit (Q) of electrical parts and resonant circuits.

Q meters are impedance-measuring instruments that determine the ratio of reactance to resistance of capacitors or inductors and resistors. Details of Q meters and impedance bridges as well as a number of other methods of measuring circuit impedance are described in the following paragraphs. Also discussed are methods of measuring the impedance of antennas and transmission lines.

BRIDGE METHODS

Bridges are among the most accurate types of measuring devices used in the measurement of impedance. In addition, bridges are also used to measure dc resistance, capacitance, and inductance. Certain types of bridges are more suitable for measuring a specific characteristic, such as capacitance or inductance. Basic schematics for the various bridge circuits are shown in figure 3-1. The bridge circuits shown are similar in that they usually contain two branches in the measuring circuit, two branches in the comparing circuit, a detector circuit, and a power circuit, as shown in figure 3-2. The bridge shown in figure 3-2 is actually the dc Wheatstone bridge; however, the general principles of circuit operation for ac remain the same.

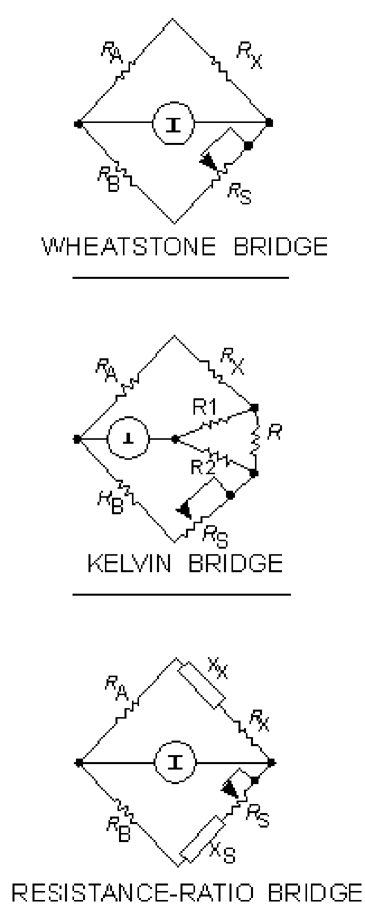


Figure 3-1.—Basic bridge circuits.

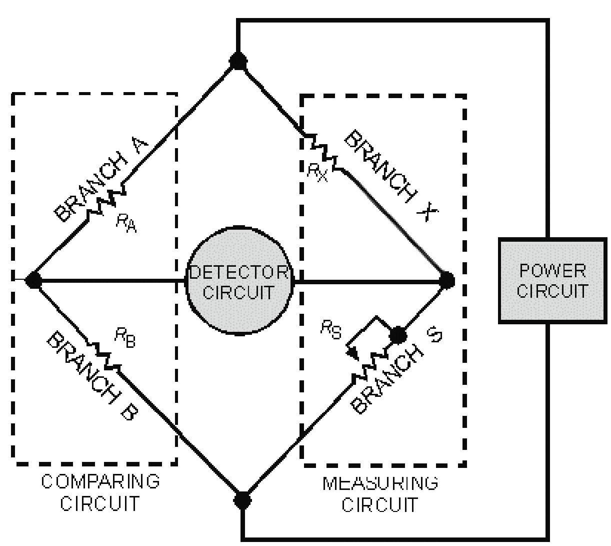


Figure 3-2.—Typical bridge circuit configuration.

The comparing circuit contains branches A and B and has provisions for changing the ratios of the branches with respect to each other, which enables various measuring ranges to be obtained. Comparison of figures 3-1 and 3-2 shows that either or both branches of the comparing circuit do not necessarily contain resistors alone. Branch B of the Hay bridge, containing C_B and R_B in series connection, provides a striking contrast with the parallel connection of C_B and R_B of the Maxwell bridge.

The measuring circuit in figure 3-2 also contains two branches. The resistance, capacitance, or inductance to be measured is connected to branch X of the bridge-measuring circuit. The subscript X is also used in figure 3-1 to designate the circuit parameters involved in computing the values of various electronic parts. Branch S contains the variable control used to bring the bridge into a balanced condition. A potentiometer is used for this purpose in most bridge equipment, because it offers a wide range of smoothly variable current changes within the measuring circuit.

The third arm of the bridge is the detector circuit. The detector circuit may use a galvanometer for sensitive measurements that require high accuracy. In the case of bridges using ac as the power source, the galvanometer must be adapted for use in an ac circuit. In many practical bridge circuits using ac to operate the bridge, an electron-ray indicating tube is used to indicate the balanced condition by opening and closing the shadow area of the tube. Headsets are also used for audible balance detection, but this method reduces the accuracy obtainable with the bridge.

Switches are used in bridge circuits to control the application of operating power to the bridge and to complete the detector circuit. Frequently, the two switching functions are combined into a single key, called a bridge key, so that the operating power is applied to the bridge prior to the detector circuit. This sequence reduces the effects of inductance and capacitance during the process of measurement.

The most unfavorable condition for making a measurement occurs when the resistance, capacitance, or inductance to be measured is completely unknown. In these cases, the galvanometer cannot be protected by setting the bridge arms for approximate balance. To reduce the possibility of damage to the galvanometer, you should use an adjustable shunt circuit across the meter terminals. As the bridge is brought closer to the balanced condition, the resistance of the shunt can be increased; when the bridge is in balance, the meter shunt can be removed to obtain maximum detector sensitivity.

Bridges designed specifically for capacitance measurements provide a dc source of potential for electrolytic capacitors. The electrolytic capacitors often require the application of dc polarizing voltages in order for them to exhibit the same capacitance values and dissipation factors that would be obtained in actual circuit operation. The dc power supply and meter circuits used for this purpose are connected so that there is no interference with the normal operation of the capacitance-measuring bridge circuit. The dissipation factor of the capacitor may be obtained while the capacitor is polarized. In figure 3-2, the signal voltage in the A and B branches of the bridge will be divided in proportion to the resistance ratios of its component members, R_A and R_B, for the range of values selected. The same signal voltage is impressed across the branches S and X of the bridge. The variable control, R_S, is rotated to change the current flowing through the S and X branches of the bridge. When the voltage drop across branch S is equal to the voltage drop across branch A, the voltage drop across branch X is equal to the voltage drop across branch B. At this time the potentials across the detector circuit are the same, resulting in no current flow through the detector circuit and an indication of zero-current flow. The bridge is balanced at these settings of its operating controls, and they cannot be placed at any other setting and still maintain this balanced condition.

The ability of the bridge circuit to detect a *balanced* condition is not impaired by the length or the leads connecting the bridge to the electronic part to be measured. However, the *accuracy* of the measurement is not always acceptable, because the connecting leads exhibit capacitive and inductive characteristics, which must be subtracted from the total measurement. Hence, the most serious errors affecting accuracy of a measurement are because of the connecting leads.

Stray wiring capacitance and inductance, called residuals, that exist between the branches of the bridge also cause errors. The resistance-ratio bridge, for example, is redrawn in figure 3-3 to show the interfering residuals that must be eliminated or taken into consideration. Fortunately, these residuals can be reduced to negligible proportions by shielding and grounding. A method of shielding and grounding a bridge circuit to reduce the effects of interfering residuals is through the use of a Wagner ground, as shown in figure 3-4. Observe that with switch S in position Y, the balanced condition can be obtained by adjusting Z^1 and Z^2. With switch S in position X, the normal method of balancing the bridge applies. You should be able to reach a point where there is no deflection of the meter movement for either switch position (X or Y) by alternately adjusting Z^1 and Z^2 when the switch is at position Y and by adjusting R_S when the switch is at position X. Under these conditions, point 1 is at ground potential; and the residuals at points 2, 3, and 4 are effectively eliminated from the bridge. The main disadvantage of the Wagner ground is that two balances must be made for each measurement. One is to balance the bridge, and the other is to balance the Wagner ground. Both adjustments are interacting because R_A and R_B are common to both switch positions X and Y.

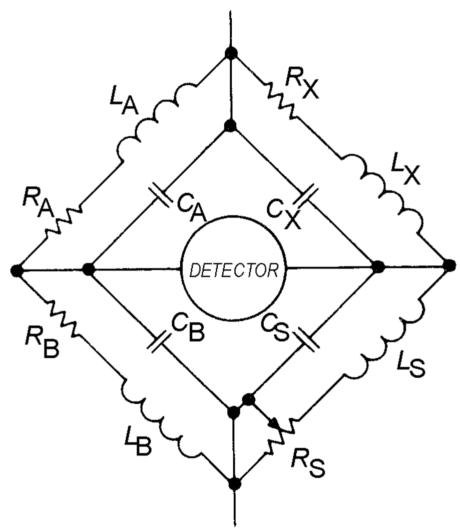


Figure 3-3.—Resistance-ratio bridge residual elements.

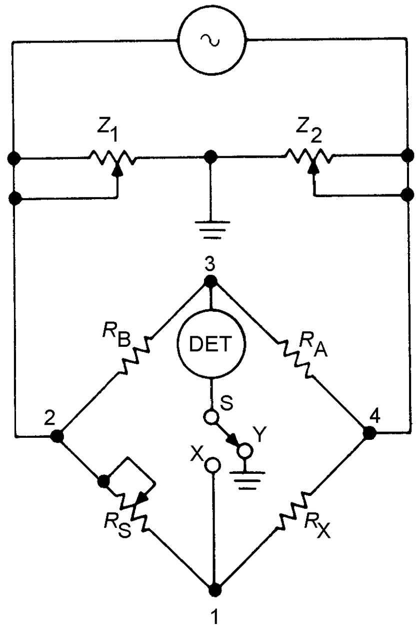


Figure 3-4.—Wagner ground.

Many bridge instruments provide terminals for external excitation potentials; however, do not use a voltage in excess of that needed to obtain reliable indicator deflection because the resistivity of electronic parts varies with heat, which is a function of the power applied.

Q-1. What conditions must be met in order to balance a bridge circuit?

Q-2. When you are measuring a component using a bridge, what is the most common cause of inaccurate measurements?

Wheatstone Bridge

The Wheatstone bridge, shown in figure 3-1, is often used to measure resistance. These instruments are usually portable because they require only a small, dc source to power the bridge, which is easily obtained from flashlight batteries. In those cases where an external supply voltage is desirable for the operation of the bridge, use the minimum voltage that will give a reliable indication by the galvanometer. Increasing the supply voltage any further results in uncompensated thermal variations and decreased bridge accuracy. If greater bridge sensitivity is needed, use a galvanometer with greater sensitivity.

A number of other considerations are involved in the choice of a galvanometer. For example, the galvanometer should not be subjected to false or erratic indications because of external magnetic fields. This requirement dictates the choice of a shielded meter mechanism. It is also desirable to use a critically dampened meter movement to ensure decisive movement of the meter pointer during conditions of bridge unbalance. Thermal agitation sometimes produces voltages that interfere with the balancing of the bridge. For this reason, the Wheatstone bridge usually includes a polarity-reversing switch in the detector circuit. When a measurement is required, note the reading for both positive and negative indications, and figure the average of both readings. With the exception of inaccuracies introduced by thermal variations (caused by excessive supply voltages), the accuracy of the Wheatstone bridge is, otherwise, independent of the value of supply voltages. The units used in calibrating the galvanometer are unimportant to the accuracy of the bridge, since a 0 indication is desired at the balanced condition.

Resistance values ranging from 1 ohm to 1 megohm can be measured with an accuracy of approximately 0.1%. However, difficulties are encountered when very high and very low resistances are measured. Resistances less than 1 ohm are difficult to measure accurately because of uncertainty arising from the contact resistance present between the resistor to be measured and the binding posts of the bridge. Measurement of resistances greater than 1 megohm becomes difficult because of two factors: (1) The ratio of standard resistances R_A and R_B involve a ratio on the order of 1,000 to 1, and (2) the voltage applied to the bridge must be substantially increased to obtain definite galvanometer action. The result is that an increase in the supply voltage increases the power dissipation (heat) of the bridge resistors. The change in resistance R_B, because of the heat, is sufficient to produce an appreciable error. A Kelvin bridge is recommended for measuring resistances lower than 1 ohm. An electronic multimeter is recommended for the indicating device in bridges used for the measurement of very high resistances.

One of the most elementary precautions concerning the use of a bridge, when measuring low resistance, is to tighten the binding posts securely so that the contact resistance between the binding posts and the resistance to be measured is minimum. Leakage paths between the resistor leads along the outside surface of the resistor body must be avoided when resistances greater than 0.1 megohm are measured. Search for defective solder joints or broken strands in stranded wire leads; these defects can cause erratic galvanometer indications. In those cases where wire leads must be used to reach from the resistance under test to the bridge terminals, measure the ohmic value of those leads prior to further measurements.

Q-3. How does the supply voltage affect the accuracy of Wheatstone bridge measurements?

Kelvin Bridge

It is often necessary to make rapid measurements of low resistances, such as samples of wire or low values of meter shunt resistors. A frequently used instrument that is capable of good precision is the Kelvin bridge, shown in figure 3-1. Note the similarity between this and the Wheatstone bridge. Two additional resistances, R1 and R2, are connected in series and shunted across resistance R, which is the circuit resistance existing between the standard and unknown resistances, R_S and R_X, respectively. In performing the adjustment for balance, you must make the ratio of R1 to R2 equal to the ratio of R_A to R_B. When this is done, the unknown resistance can be computed in the same manner as that for the Wheatstone bridge, because resistance R is effectively eliminated.

In using a Kelvin bridge, you must follow precautions similar to those given for the Wheatstone bridge. A rheostat is usually placed in series with the battery so that bridge current can be conveniently limited to the maximum current allowable. This value of current, which affects the sensitivity of the bridge, is determined by the largest amount of heat that can be sustained by the bridge resistances without causing a change in their values. All connections must be firm and electrically perfect so that contact resistances are held to a minimum. The use of point and knife-edge clamps is recommended. Commercially manufactured Kelvin bridges have accuracies of approximately 2% for resistance ranges from 0.001 ohm to 25 ohms.

Q-4. Kelvin bridges are well suited for what type of measurements?

Resistance-Ratio Bridge

The resistance-ratio bridge, shown in figure 3-1, may be used to measure capacitance, inductance, or resistance so long as the electronic part to be measured is compared with a similar standard. The measurement of the value of a capacitor must be made in terms of another capacitor of known characteristics, termed the STANDARD CAPACITOR. The same requirement is necessary for an inductance measurement. The standard of comparison is designated as X_X, and the losses of the standard are represented as R_X. If you experience difficulty in obtaining a balanced bridge condition, insert additional resistance in series with branch S of the bridge. This adjustment becomes necessary because the *Q* of the unknown capacitor or inductor in branch X is higher than the comparable *Q* of the standard in branch S.

Schering Bridge

The Schering bridge, shown in figure 3-1, is a commonly used type of bridge for the measurement of capacitors and dielectric losses. The *Q* of a capacitor is defined as the reciprocal of the dissipation factor, which is the ratio of the capacitor's dielectric constant to its conductivity at a given frequency. Accordingly, capacitor *Q* is determined by the frequency used to conduct the measurement and the value of the capacitor, C_B, required to obtain bridge balance. The accuracy of this type of bridge is excellent, about 2% for dissipation factors ranging from 0.00002 to 0.6. Typical accuracies for capacitive reactances in the range of 100 picofarads to 1 microfarad are 0.2%.

Hay Bridge

The Hay bridge, shown in figure 3-1, is used for the measurement of inductance and the *Q* of the inductor. It is interesting to note that this type of bridge measures inductance by comparing it with a standard capacitor of known characteristics. This arrangement provides the advantage of a wide measurement range with the minimum use of electronic parts as comparison standards. A typical range of values that can be measured with the Hay bridge is from 1 microhenry to 100 henries. The accuracy of the measurements made with this bridge is about 2%. The frequency used in conducting the inductance measurement must be taken into account because of the series reactance of capacitor C_B. The loss factor of the inductor under test is balanced in terms of the *Q* of the inductor. The Hay bridge, then, is used for measurement of inductances having a *Q* greater than 10. For instance, a *Q* of 10 gives a calibration error of 1%, whereas a *Q* of 30 gives a calibration error of 0.1%.

Q-5. When you are testing an inductor with a Hay bridge, the characteristics of the inductor are compared with what type of device?

Maxwell Bridge

The Maxwell bridge, shown in figure 3-1, is used for the measurement of inductance and inductive *Q*. This bridge is similar to the Hay bridge because it also measures inductance by comparison with a standard capacitor of known characteristics. Notice, in particular, that capacitor C_B is connected in parallel with resistor R_B. In connection with this difference, the requirement of an accurately known frequency is removed. This bridge circuit is employed for measuring the inductance of inductors having large losses; i.e., low *Q*. The range of this type of instrument is much greater than that of the Hay bridge; values ranging from 1 microhenry to 1,000 henries are measurable, with an error of only 2%.

VECTOR BRIDGES

The basic bridges described up to now determined the resistive and reactive components of the unknown impedance; however, the vector bridge indicates the magnitude and phase angle. Typically, vector bridges require two null readings. Consider the basic bridge circuit of figure 3-5. The magnitude of the unknown impedance (Z_X) is determined by the voltages applied across *R* and Z_X and to the bases of emitter followers Q1 and Q2, which bias the balanced rectifiers, CR1 and CR2. Resistors *A* and *B* are equal in value. When *R* is adjusted to equal Z_X, the voltages between points 1 and 2 and between points 1 and 4 are equal in magnitude, and the vtvm will indicate 0 volts.

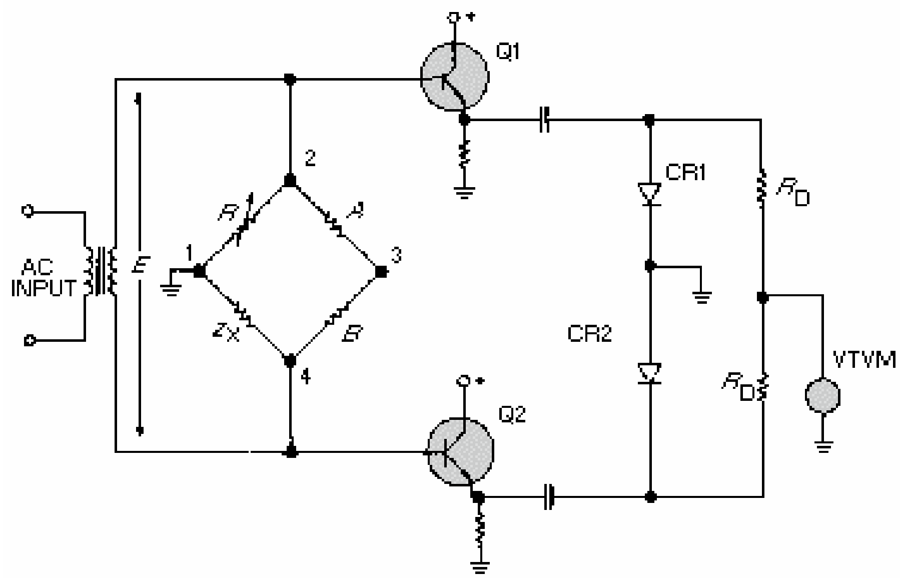


Figure 3-5.—Typical vector-bridge configuration (amplitude).

The absolute value of Z_X is determined from the dial calibration of *R*. Without altering the amplitude balance, you reconnect the external circuits as shown in figure 3-6. Note that the voltage between points 1 and 3 is being compared to the voltage between points 1 and 2. Potentiometer *R*, calibrated in degrees, is adjusted for a null indication on the vtvm; and the phase angle is read directly. If Z_X is purely resistive, the voltage between points *1* and *3* will be zero and the setting of *R* will be 0 volts. If Z_X is purely reactive (capacitive or inductive), the setting of *R* will be at maximum voltage. For phase angles between 0º and 90º, the scale of *R* may be calibrated directly in degrees. The sign of the phase angle can be determined by changing the signal frequency slightly and observing the change in impedance. The presence of harmonics in the signal input will severely hamper the measurements. If a pure frequency source is not available, suitable low-pass filters will have to be employed in the output leads from the bridge.

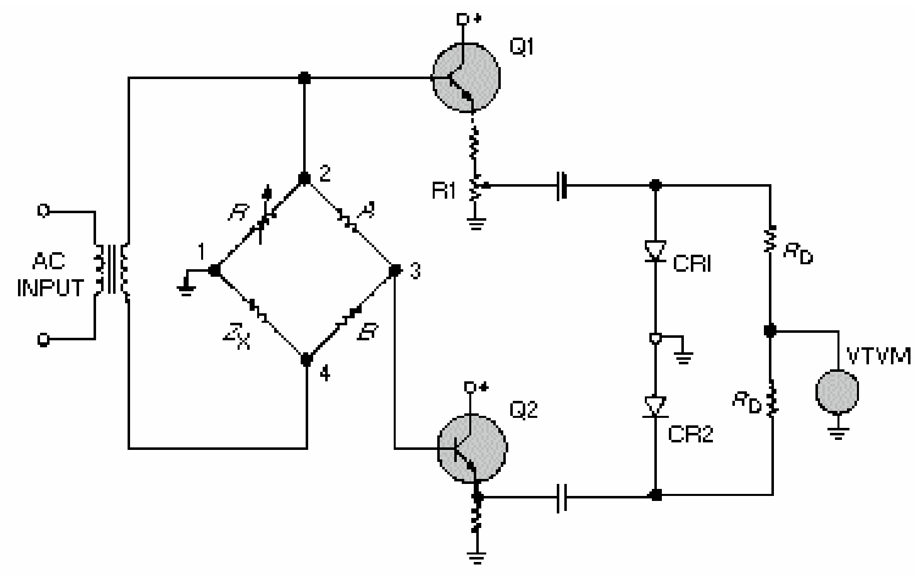


Figure 3-6.—Typical vector-bridge configuration (phase).

CONSTANT-CURRENT, IMPEDANCE-MEASURING TECHNIQUE

This technique employs an oscillator circuit and a vtvm, as shown in figure 3-7.

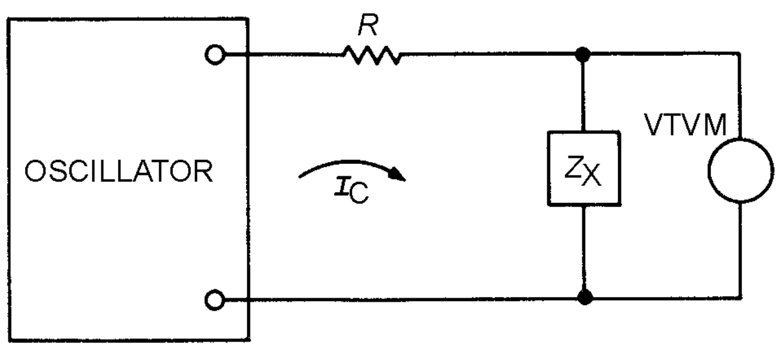


Figure 3-7.—Constant-current, impedance-measuring method.

A large value of resistance, R, is selected so that I_C is virtually independent of the range of Z_X to be measured. Thus, $I_C Z_X$ represents the value of voltage measured by the vtvm. If R is chosen so that the voltage drop across Z_X corresponds to a full-scale reading on the vtvm, a direct reading impedance meter is realized. For example, assume that the audio oscillator open-circuit voltage is 10 volts (rms) and that the full-scale reading of the vtvm is 0.05 volt. If you want to measure Z_X values ranging up to a maximum of 5,000 ohms, you should use a 1-megohm resistor for R. This will result in a full-scale, 0.05-volt deflection. An oscillator that does not produce harmonics should be used.

IMPEDANCE-ANGLE METER

Like vector bridges, impedance-angle meters determine an unknown impedance in terms of magnitude and phase angle. However, a non-bridge technique is used. The simplified circuit of a commercial instrument is shown in figure 3-8. With switches S1 and S2 at the BAL position, the variable standard resistor, *R*, is adjusted until the balanced rectifier outputs of Q1 and Q2 are equal (indicated by a null in the deflection of the voltmeter connected between the emitters of Q3 and Q4). The dial setting of *R* gives the value of Z_X. For phase angle determination, the circuit is switched to CAL and the input voltage is adjusted for full-scale voltmeter deflection. The circuit is then switched to PHASE; thus, the paralleled outputs of Q1 and Q2 are applied to rectifier CR1 only. With S2 in the phase position, there is no input to the base of Q4. If *Z* is purely resistive, the outputs of Q1 and Q2 cancel, and the voltmeter indicates zero deflection. For a complex impedance, the base of Q3 will be unbalanced with respect to the base of Q4; and the voltmeter deflection, calibrated in degrees, determines the phase angle of the unknown impedance. Typical commercial impedance angle meters, operating at 2 MHz, are accurate to within 4% for impedances of from 10 to 500 ohms.

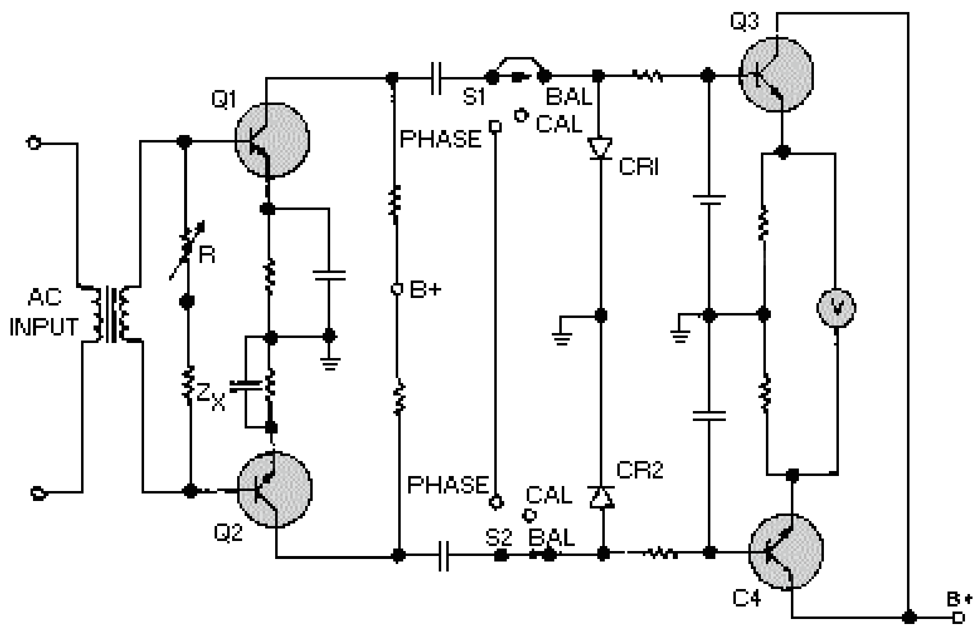


Figure 3-8.—Impedance-angle meter.

Q-6. What do impedance-angle meters and vector bridges have in common?

IMPEDANCE TESTING OF ANTENNAS AND TRANSMISSION LINES

The amount of current that flows in an antenna is one of the most important factors affecting the performance of transmitter equipment. As much of the rf energy generated as possible must be efficiently transferred to the antennas to secure the maximum radiated power from a transmitter. Also, for best reception, maximum transfer of energy from the antenna to the receiver must occur. Efficient transmission and reception conditions prevail whenever the transmitter (or receiver) is properly matched to the transmission line and the transmission line is properly matched to the antenna. Normally, performance tests concerning impedance match consist primarily of taking standing-wave measurements. In certain instances, it may be found that a change in antenna impedance has resulted in an undesirably high standing-wave ratio. This could be the result of a new antenna installation or an interfering structure near the antenna that influences antenna characteristics.

In practice, the antenna-matching network is varied to match the new antenna characteristics, since the transmission line is designed to match equipment impedance. This can best be done by making a series of standing-wave-ratio checks and antenna-matching adjustments until an acceptable standing-wave ratio is reached. It must be understood, however, that the antenna does have a specific impedance at a given frequency and that, when necessary, this impedance may be determined by use of an rf impedance bridge.

A typical rf impedance bridge circuit is shown in figure 3-9. Rf impedance bridge measurements require an rf signal generator, a detector, and a calibrated rf bridge to determine transmission-line impedance. The bridge compares the parallel resistive-reactive combination with the series combination and can typically measure impedance over a frequency range of 500 kHz to 60 MHz.

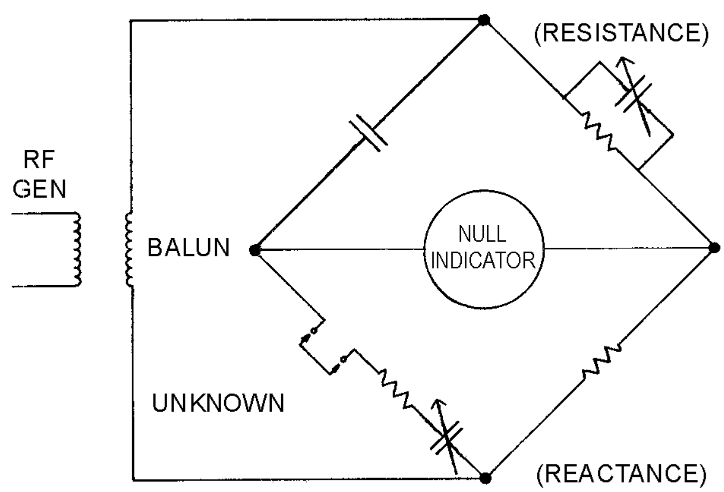


Figure 3-9.—Typical rf bridge.

Basically, the bridge is balanced with a known capacitance under short-circuit conditions. The unknown impedance is then inserted in lieu of the short bus, and the bridge is rebalanced. The difference between the known impedance under short-circuit conditions and the balance measurements obtained with the unknown impedance inserted in lieu of the short is the value of the unknown impedance.

Q-7. What is the result of an impedance mismatch between a receiver or transmitter and its transmission line or antenna?

POWER MEASUREMENTS

It is often necessary to check the input and output signal power levels of electronic equipment. The determination of dc power is computed by using a derivative of Ohm's law ($P = IE = I^2R = E^2/R$). However, the presence of a reactive component in ac circuits means that apparent power is being measured or calculated unless the rms voltage-current value is multiplied by a power factor to obtain true lower. The measurement of ac power is further complicated by the frequency limitations of various power meters. If there is no phase difference, ac power may be computed in the same manner as dc power by determining the average value of the product of the voltage and current. In practical ac circuits, the apparent power must be multiplied by the cosine of the phase angle between the voltage and current in order to compute true power.

In the repeated measurement of audio-frequency (af) power, you may use a normal power meter calibrated directly in watts. However, when reactive components of dissipative impedance introduce a phase angle, a device that is proportional to both the power factor and the apparent power must be used. Because power-level measurements are concerned with decibel units, a working knowledge of decibels is required for proper interpretation of power tests. The decibel is used to determine the ratio of power changes or to indicate the power level in a circuit with respect to either 0 or a standard reference level.

AF POWER

In the electrical transmission of speech or music, rapidly fluctuating amplitudes and frequencies are involved. The average power-level measurement and its variation rate depend on the signal characteristics and time interval over which this average is taken. Power measurements for af circuits are usually indicated in terms of decibels (dB), decibels referenced to 1 milliwatt (dBm), or volume units (vu). For example, the power gain of an amplifier can be expressed in dB; the power level of a sinusoidal signal compared to a 1-milliwatt reference is indicated in dBm; and the power level of a complex signal, such as voice, music, or multiplexed information, compared to a reference level of 1 milliwatt, is indicated in vu.

Q-8. What are the three units of measure most commonly used when referring to af power measurements?

DECIBEL METERS

A dB meter is a form of ac electronic voltmeter calibrated in dB's. These meters are useful for making measurements where direct indication in decibels is desired. However, remember that these are voltmeters, and power measurements are not meaningful unless the circuit impedance is known. When the dB meter is calibrated, a reference point, based on a specific power or value of voltage across a specified resistance, is selected to represent 0 dB. Many electronic voltmeters use a single dB scale based on 1 milliwatt into a 600-ohm load to represent 0 dBm. Based on this reference point, various voltage readings could be made on the low ac-voltage scale. The +dB numbers corresponding to voltage ratios that exist between successive ranges and the low ac range have been computed for each range. These numbers, shown on the front panel of the instrument, are added algebraically to each successive range reading to produce the correct value for the range. The term decibel does not, in itself, indicate power. It indicates a ratio or comparison between two power levels that permits you to calculate the power. Often, it is more desirable to express performance measurements in terms of decibels using a fixed power level as a reference. The original standard reference level was 6 milliwatts, but to simplify calculations a standard reference level of 1 milliwatt has been adopted.

Q-9. In reference to dB meters, 0 dBm represents 1 milliwatt into what value of load?

VOLUME UNIT METERS

The volume unit (vu) meter is used in audio equipment to indicate input power to a transmitter or to a transmission line. This type of meter has special characteristics, such as a standardized speed of pointer movement, speed of return, and calibration. The measurement of the average power level and its rate of variation with respect to time depends not only on the signal characteristics, but also on the time interval over which the average is being taken. Accordingly, the speed of response of the instrument used to measure average power is of particular concern. The unit of measurement is the volume unit (vu), which is numerically equal to the number of dB above or below the reference level of 1 milliwatt into a 600-ohm load (provided the standard instrument was calibrated under constant-amplitude, sine-wave conditions). A change of one vu is the same as a change of one decibel. Therefore, the vu value obtained represents averages of instantaneous power of speech or music obtained by an instrument having particular dynamic characteristics. The vu readings are equivalent to the power level in decibels only if the sinusoidal waveform is of constant amplitude.

Q-10. What is the main difference between a vu and a dB meter?

ELECTRODYNAMIC WATTMETER

The electrodynamic wattmeter is used to measure power taken from ac or dc power sources. The electrodynamic wattmeter, shown in figure 3-10, uses the reaction between the magnetic fields of two current-carrying coils (or sets of coils), one fixed and the other movable. When the current through the fixed-position field winding(s) is the same as current through the load and the current through the moving coil is proportional to the load voltage, then the instantaneous pointer deflection is proportional to the instantaneous power. Since the moving pointer cannot follow the rapid variations in torque because of its momentum, it assumes a deflection proportional to the average power. The dynamometer-type wattmeter automatically compensates for the power factor error of the circuit under test. It indicates only the instantaneous power resulting from in-phase values of current and voltage. With out-of-phase relationships, a current peak through the moving coil never occurs at the same instant as the voltage peak across the load, resulting in less pointer deflection than when the current and voltage are in phase. The simple meter shown in figure 3-10 is not compensated. When the load is disconnected, this meter will still indicate that power is being consumed in the circuit. This difficulty can be eliminated by incorporating two compensating windings, mounted with the primary fixed-coil current windings, as shown in figure 3-11. These stationary windings are used to produce a magnetic flux proportional to the current through the movable coil. As shown by the arrows, the currents through the primary movable coil and the compensating coil flow in opposite directions, producing a torque caused by the opposing magnetic fields. These opposing fields cancel. Hence, with the load removed from the circuit, the meter will indicate zero power through the load.

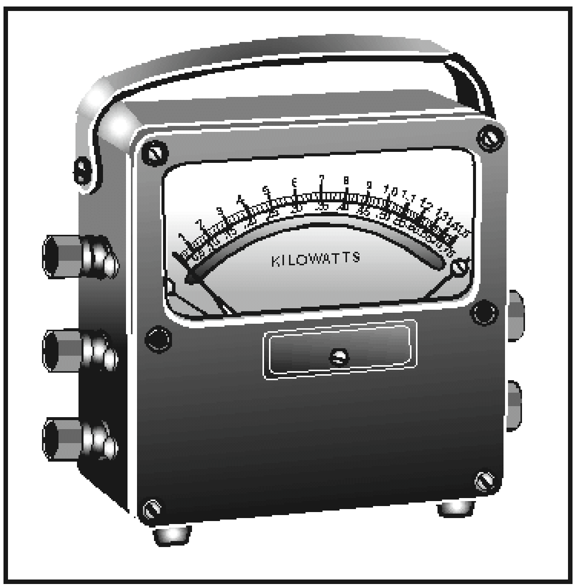


Figure 3-10.—Typical electrodynamic wattmeter.

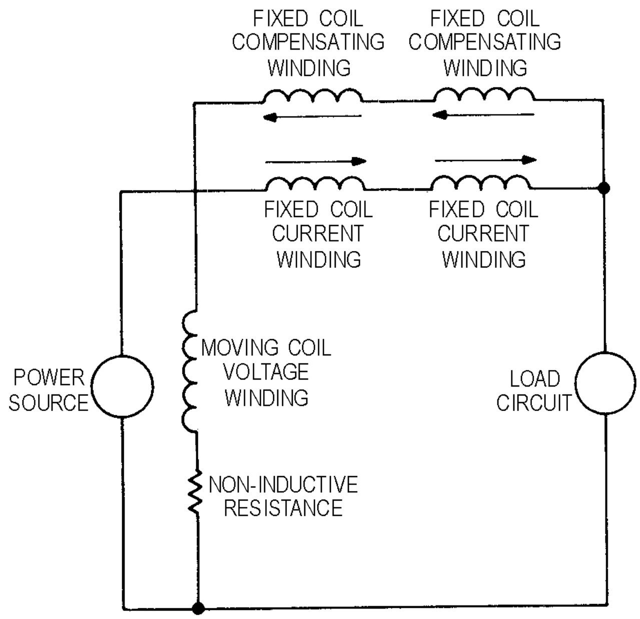


Figure 3-11.—Electrical equivalent of the compensated electrodynamic wattmeter.

Electrodynamic wattmeters are subject to errors arising from various factors, such as temperature and frequency characteristics and vibration. Heat through the control mechanism can cause the springs to lengthen and lose tension; as a result, deflection errors are produced. Figure 3-12 illustrates the mechanical equivalent of the electrodynamic wattmeter. Large currents within the circuit will also produce errors. Therefore, the maximum current range of electrodynamic wattmeters is normally restricted to about 20 amperes. When larger load currents are involved, a current transformer of suitable range is used in conjunction with the wattmeter. However, a current transformer cannot be used if the ac circuit under test contains a dc component.

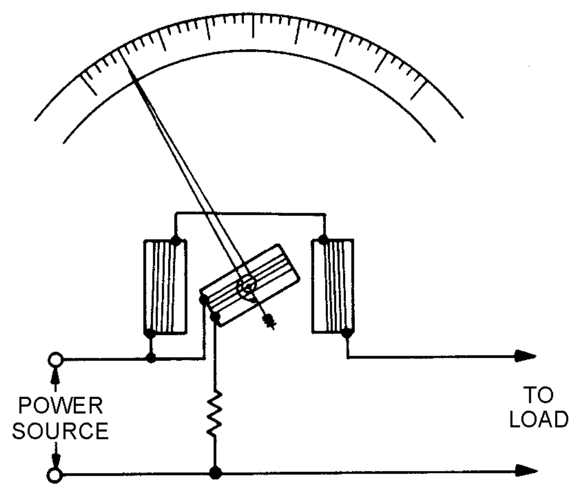


Figure 3-12.—Mechanical equivalent of the electrodynamic wattmeter.

The voltage range of wattmeters is generally limited to several hundred volts because of heat dissipation within the voltage circuit. However, the voltage range can be extended by using external voltage dividers. Wattmeters used as laboratory standards have an accuracy of 0.1%, high-grade portable wattmeters an accuracy of 0.2% to 0.25%, and high-grade switchboard wattmeters an accuracy of 1% of full-scale value. Because electrodynamic wattmeter errors increase with frequency, they are used primarily for measuring 60-hertz line power. Unshielded electrodynamic wattmeters should not be placed in the vicinity of stray magnetic fields. A wattmeter has current, voltage, and power ratings; therefore, damage may result when any of these ratings is exceeded.

The electrodynamic wattmeter may be converted into an instrument for measuring reactive power by replacing the resistance normally in series with the voltage coil with a large inductance. A 90-degree current lag within the voltage coil provides a direct reading proportional to the reactive power in the circuit. Compensating networks must be used to cause the phase shift to be exactly 90º.

Q-11. What type of device is used to extend the current-measuring capability of electrodynamic wattmeters?

IRON-CORE, COMPOSITE-COIL, AND TORSION-HEAD WATTMETERS

Iron-core wattmeters are primarily used as switchboard instruments and employ the induction principle. Voltage and current coils are wound around a laminated iron core shaped to produce a mutually perpendicular magnetic field across an air gap. Eddy currents induced in a thin metal cylinder rotating in this air gap interact with the magnetic field to produce a torque proportional to the instantaneous power. This type of construction provides the advantages of increased operating torque, larger angles of rotation, ruggedness, compactness, and freedom from errors caused by stray fields. It has the disadvantage of a very narrow frequency range.

The composite-coil wattmeter uses the upscale torque, produced by the ac power being measured, in opposition to the torque produced by an adjustable dc current in a set of windings intermingled or wound within the ac windings. Greater reading precision is obtained with this method than is possible with straightforward wattmeters, and errors caused by elasticity of the spring suspension carrying the moving-coil system are avoided. The torsion-head wattmeter is used to restore the movable coil to its original position after deflection and to remove the mutual inductance error.

ELECTRONIC WATTMETER

Electronic wattmeters are used for direct, small power measurements or for power measurements at frequencies beyond the range of electrodynamometer-type instruments. A simplified electronic wattmeter circuit is shown in figure 3-13. The matched triodes are operated in the nonlinear portion of their characteristic grid-voltage, plate-current curves. The symmetrical resistive T network between the generator and load will provide V1 and V2 voltages proportional to, and in phase with, the load current and voltage, respectively. A source of ac power is connected to the load through the series resistors R1 and R2. These two resistors are of equal value and are made small to prevent the voltage drop across them from reducing the load voltage appreciably. R3 is made large enough to have negligible power consumption. Therefore, the R3 voltage is equal to the load voltage, and the voltage across either series resistor is proportional to the difference in the output currents of the tubes. The average value of the difference could be measured by a dc meter connected to read the voltage potential between the grids of V1 and V2. This method is adequate only at low frequencies. As the frequency increases, the stray capacitances and inductances also increase. The frequency range of the electronic wattmeter can be extended up to 20 megahertz by using pentodes instead of triode tubes. The operating conditions in a pentode are adjusted so that plate current is proportional to the product of a linear function of plate voltage and an exponential function of grid voltage.

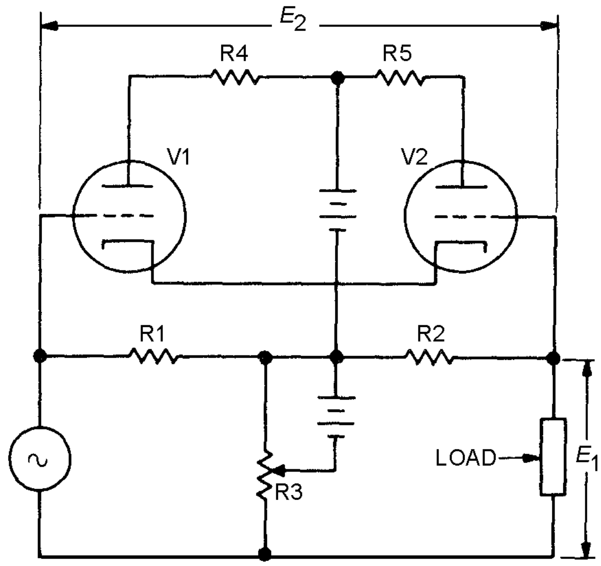


Figure 3-13.—Simple electronic wattmeter circuit.

Q-12. For power measurements, what advantage does an electronic wattmeter have over an electrodynamic wattmeter?

ABSORPTION POWER METERS

Absorption power meters absorb either all or part of the source power. They require means of dissipating the absorbed power, sensing the power thus dissipated, and indicating the amount of power absorbed by the sensing network. Output power meters, in-line wattmeters, and meters employing bolometers are examples of absorption power meters used by the Navy.

Output Power Meters

Figure 3-14 shows a common output power meter used in vhf-uhf applications. It has a 0- to 150-watt range covered in two steps: 0-50 watts and 0-150 watts. Attenuator AT1 provides a 50-ohm nominal resistive (dummy) load and uses metal film on glass construction. This dummy load is tapped to provide the proper operating voltage to the meter. Resistors R3 and R5 form a calibration network at 50 watts; R7 and R8 form a calibration network at 150 watts. Accuracy, at approximately 20º C, is ±5% for frequencies between 30 MHz and 600 MHz, ±10% for frequencies between 0.6 GHz and 0.8 GHz, and ±20% for frequencies between 0.8 and 1.0 GHz. When radio-frequency (rf) power is applied to AT1, this attenuator minimizes the effects of power factors generated by any reactive components. The rf energy is then detected and filtered by CR1 and C1, respectively. The resultant dc voltage, which is proportional to the input power, is applied to a sensitive microammeter via one of the calibration networks. This meter has a scale provided with two ranges: 0-50 watts and 0-150 watts. To protect the meter, you should always try the higher range first. If the value proves to be under 50 watts, a shift to the lower scale would provide improved accuracy.

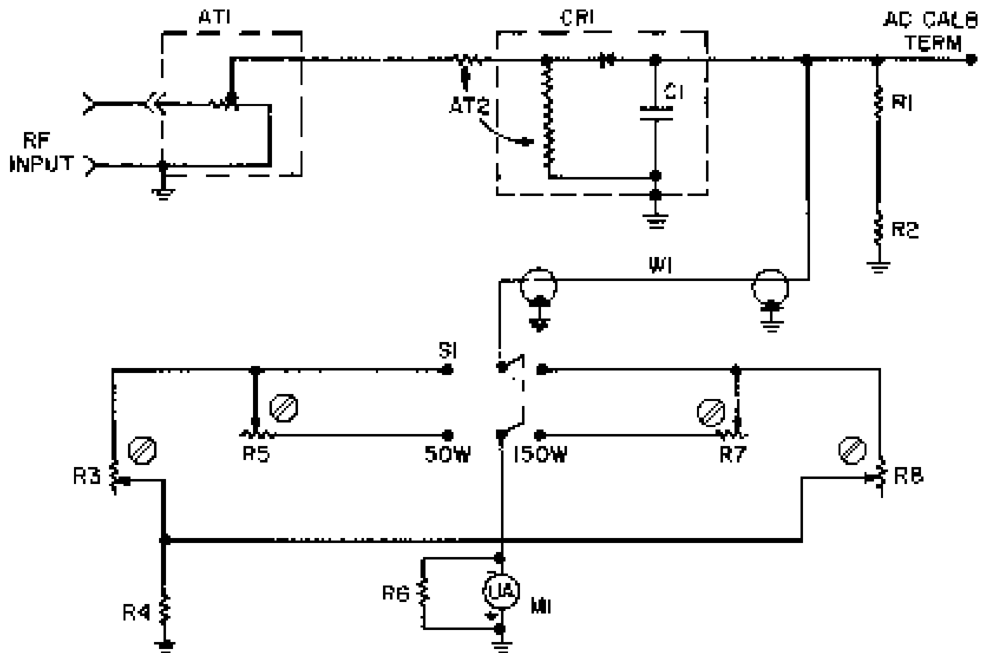


Figure 3-14.—Vhf-uhf wattmeter.

In-Line Wattmeters

The AN/URM-120 in-line wattmeter, shown in figure 3-15, measures power applied to a 50-ohm impedance load and the power reflected from that load. The internal directional coupler is oriented such that it responds only to a wave traveling in one direction on the transmission line. The coupler can be rotated to accommodate either incidental or reflected power. The rf is then rectified, filtered, and applied to the meter, which is scaled in watts. The rf power of 50 to 1,000 watts can be measured between the frequencies of 2 MHz to 30 MHz; and 10 to 500 watts, between the frequencies of 30 MHz to 1,000 MHz.

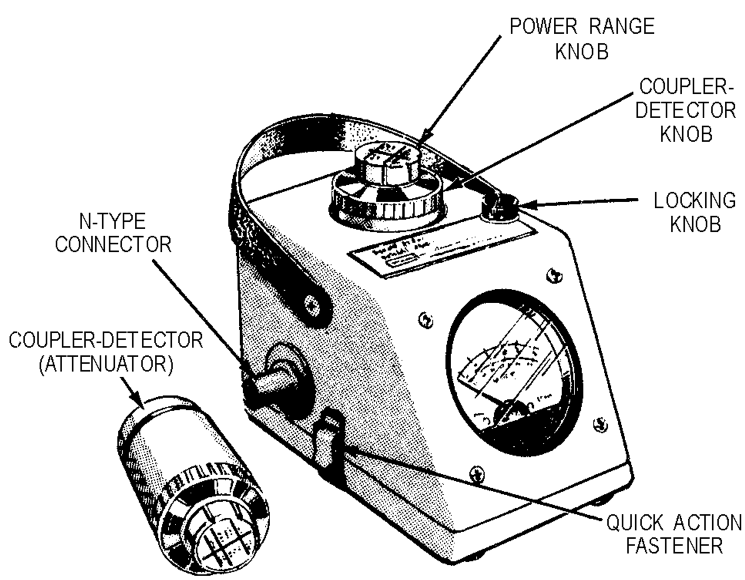


Figure 3-15.—Typical in-line wattmeter.

Q-13. What is the advantage of using in-line wattmeters over output power meters?

Bolometer

A bolometer features a specially constructed element of temperature-sensitive material. The active material is a semiconductor bead supported between two pigtail leads. When rf power is applied to a bolometer element, the power absorption by the element heats the element and causes a change in its electrical resistance. Thus, a bolometer can be used in a bridge circuit so that small resistance changes can be easily detected and power measurement can be accomplished by the substitution method (that is, substitution of dc or low-frequency power to produce an equivalent heating effect). A D'Arsonval meter movement is usually employed as the null indicator.

According to one principle of measurement (the principle used in the balanced bridge), the bridge is initially balanced with low-frequency bias power. Rf power is then applied to the bolometer and the bias power is gradually removed until the bridge is again balanced. The actual rf power is then equal to the bias power removed.

According to another principle of measurement (the principle used in the unbalanced bridge), the bridge is not rebalanced after the rf power is applied. Rather, the indicator reading is converted directly into power by calibration previously performed.

Figure 3-16 illustrates the basic bolometer bridge circuit. The bolometer element must be physically small to be highly sensitive; it must be equally responsive to low-frequency and rf power; and it must be matched to the rf-input power line. The cross-sectional dimension of the bolometer element is approximately equal to the skin depth of rf current penetration at the highest frequency of operation. This condition permits the dc and rf resistivities to be essentially equal with the reactive component of the bolometer impedance at a minimum. Thermistors, which are a type of bolometer, use semiconductor material shaped like a bead, with a thicker skin depth and shorter length to minimize standing-wave effects. These physical properties assure correspondence between lengthwise low-frequency and rf power distribution to provide the necessary inherent accuracy of the bolometer.

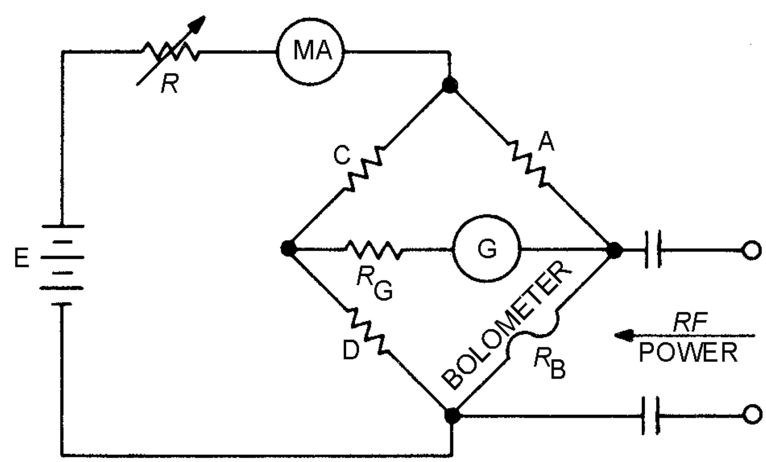


Figure 3-16.—Basic bolometer bridge circuit.

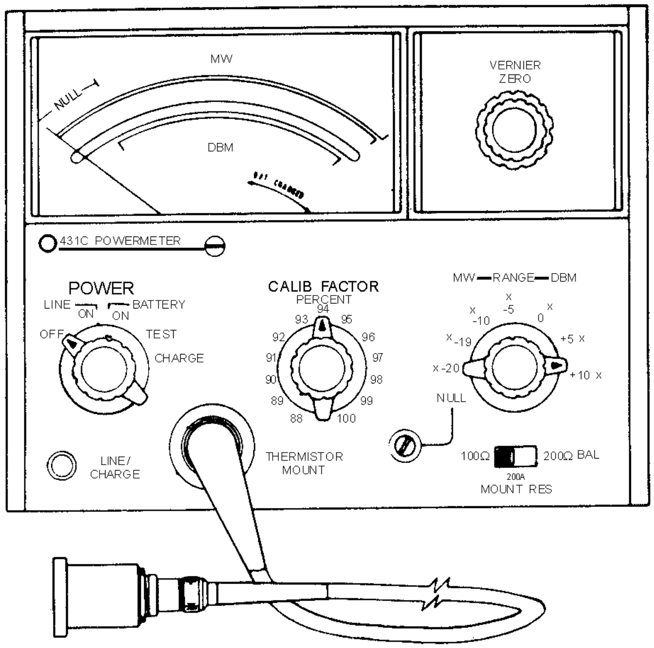
An air-mounted bolometer provides a power sensitivity 100 or more times greater than that provided by static calorimetric devices. Additional sensitivity may be obtained by mounting the element within an evacuated envelope to eliminate convective heat loss. The small size of bolometer elements is associated with small thermal mass and short thermal time constants. The thermal time constant varies directly with the volume-to-area ratio of the element for a particular shape and composition. Typical time is up to 0.1

second for thermistor beads. The thermistor type of bolometer element is usually composed of a ceramic-like mixture of metallic oxides having a large negative temperature coefficient of resistance. Two fine platinum-alloy wires are embedded in the bead, after which the bead is heated and coated with a glass film. Typical dimensions of a thermistor bead used for microwave measurements are 0.015 inch along its major axis and 0.010 inch along its minor axis. The thermistor bead may be operated at high temperatures; it is rugged, both electrically and mechanically; it has high resistance-power sensitivity; and it has a good temperature-power sensitivity. In addition, it can endure large pulse energies; it has a sluggish thermal response; and it has negligible pulsed-power measurement errors. The more sensitive thermistor requires thermal shielding or heat compensation for best operation.

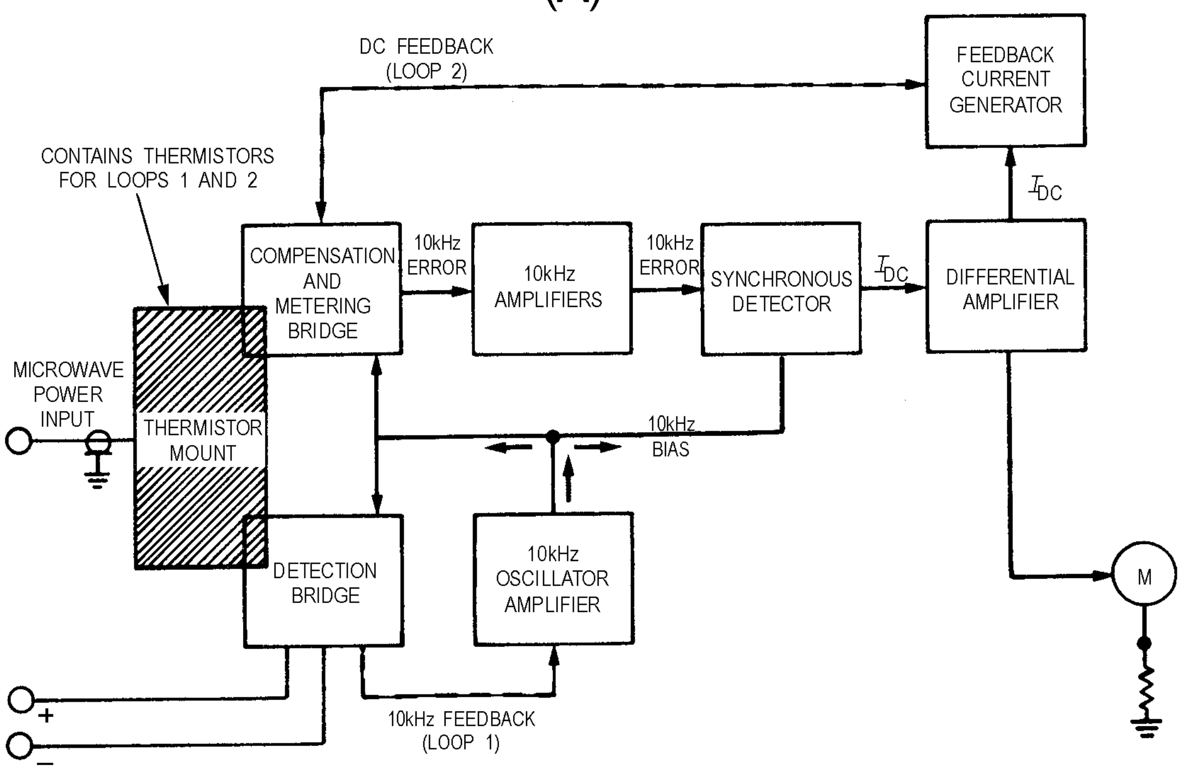
Q-14. What type of material is used in the construction of bolometers and thermistors?

Bolometer Power Meter

The standard power meter used in the Navy (Hewlett-Packard 431 C) is an automatic self-balancing instrument employing dual-bridge circuits. It is designed to operate with temperature-compensated thermistor mounts that enable you to measure power in a 50-ohm coaxial system from 10 MHz to 18 GHz and in a waveguide system from 2.6 GHz to 40 GHz. This power meter can be operated from either an ac or a dc primary power source. The ac source can be either 115 or 230 volts at 50 to 400 hertz. The dc source is a 24-volt rechargeable battery. A seven-position range switch allows full-scale power measurements of 10 microwatts to 10 milliwatts or of −20 dBm to +10 dBm. These ranges can be further extended with the aid of attenuators. The thermistor mount (as shown in fig. 3-17) contains two thermistors: one in the detection bridge, which absorbs the microwave power to be measured, and the other in the compensation and metering bridge, which supplies temperature compensation and converts the measured rf power to a meter indication. Each bridge includes its respective thermistor element as a bridge arm.



(A)



(B)

Figure 3-17.—Power meter.

Basically, the power meter circuit consists of two bridges; each bridge includes one of the thermistor elements as a bridge arm. The bridges are made self-balancing through the use of feedback loops. Positive or regenerative feedback is used in feedback loop 1; degenerative (negative) feedback is used in feedback loop 2. Both bridges are excited by a common 10-kHz source. The 10-kHz amplifier-oscillator supplies 10-kHz power to bias the thermistor in feedback loop 1 to produce the resistance required to balance the rf bridge. An equal amount of 10-kHz power is supplied by the same oscillator to the second thermistor in feedback loop 2 through two series-connected transformers. Feedback loop 2 balances the meter bridge. When rf is applied to the thermistor in the detection bridge (but not to the compensation and metering bridge), an amount of 10-kHz power is present, equal to the rf power being removed from the detection bridge by the self-balancing action of the bridge. Since the rf power replaced the 10-kHz power, the detection bridge is in balance; however, the metering bridge must be balanced by its separate feedback loop. Sufficient dc power to equal the 10-kHz power lost by the metering bridge is automatically replaced, balancing this loop. Hence the dc power applied to the metering bridge thermistor is equal to the microwave power applied to the detection bridge. The meter circuit senses the magnitude of the feedback current. The resultant meter current passes through a differential amplifier to the indicating meter. The two thermistors are matched with respect to their temperature characteristics; therefore, there is only a very small amount of drift of the zero point with ambient temperature changes. When there is a change in temperature, there is a change in the electrical power needed by the thermistors to maintain constant operating resistances. This change is automatically performed by feedback loop 1, which changes the amount of 10-kHz power for both thermistors by the proper amount. The dc power in feedback loop 2 is not changed; and since it is this dc power that is metered, the temperature change has not affected the meter indication.

CALORIMETERS

The calorimeters are the most accurate of all instruments for measuring high power. Calorimeters depend on the complete conversion of the input electromagnetic energy into heat. Direct heating requires the measurement of the heating effect on the medium, or load, terminating the line. Indirect heating requires the measurement of the heating effect on a medium or body other than the original power-absorbing material. Power measurement with true calorimeter methods is based solely on temperature, mass, and time. *Substitution* methods use a known, low-frequency power to produce the same physical effect as an unknown rf power being measured. Calorimeters are classified as STATIC (nonflow) types and CIRCULATING (flow) types.

Q-15. Power measurements performed with calorimeters are based on what three variables?

Static Calorimeters

The static calorimeter uses a thermally shielded body. Since an isolated body loses little heat to a surrounding medium, the temperature increase of the body is in direct proportion to the *time* of applied power. The product of the rate of temperature rise in the calorimetric body and its heat capacity equals applied power. Figure 3-18 illustrates a static-type calorimeter.

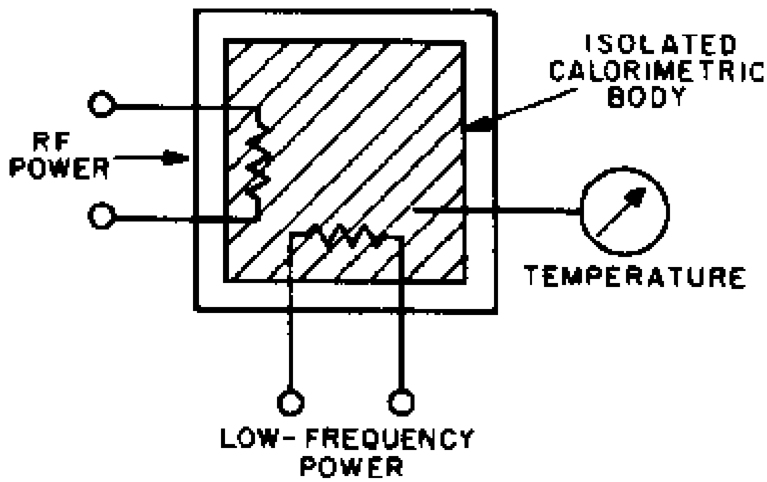


Figure 3-18.—Static calorimeter using low-frequency power substitution.

The most common type of static calorimeter is the ADIABATIC calorimeter. In the adiabatic meter, power is applied directly to a thermally isolated body; and the rate of temperature rise is determined from a temperature change measurement during a sufficiently long, known time interval. Figure 3-19 illustrates an adiabatic calorimeter using water as the body contained in a covered Dewar flask. A tapered-wall, open-ended waveguide contains a sealed, inclined glass partition to create a wedge-shaped water load of low-reflection coefficient. Thorough mixing of the water is accomplished with a stirrer, and a sensitive thermometer measures the temperature rise. A heating coil is wound around the waveguide inside the calorimeter and is used for calibrating purposes when low-frequency power is applied. This type of meter can be used for accurate measurement of several hundred watts of average power and can withstand 50 kilowatts of peak power.

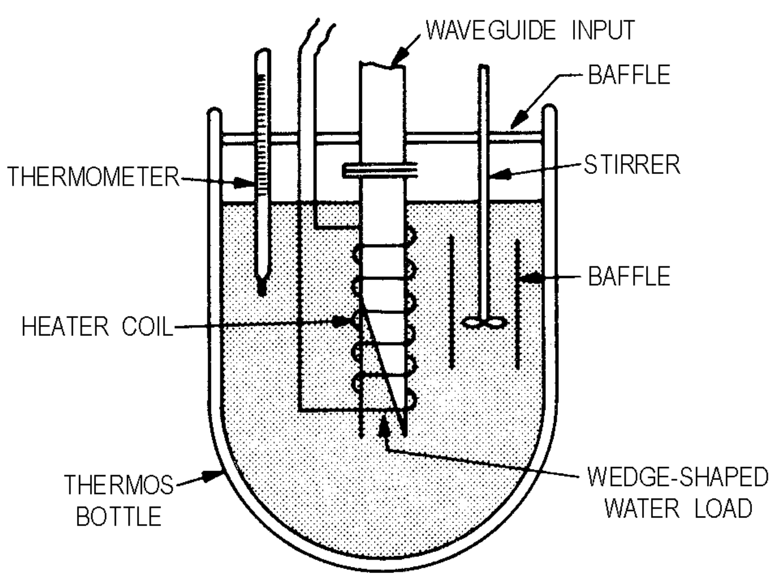


Figure 3-19.—Adiabatic calorimeter.

The NONADIABATIC calorimeter uses an rf termination with a resistive film strip or LOSSY dielectric materials (solids or liquids that are designed to efficiently dissipate the applied power) as a load. Temperature indication can be accomplished with thermocouples, thermopiles, thermistors, thermometers, bimetallic strips, and manometers. Calibration is against a power standard or known low-frequency power.

Based on the above principle, a coaxial calorimeter of good sensitivity with a short, 50-ohm resistive film on a lava (dielectric) center conductor, enclosed within a tapered, thin-walled outer conductor, is used for frequencies between 0 and 1.2 GHz. The rf termination is electrically connected to, but thermally isolated from, a massive mounting plate by a short section of silvered-lava coaxial line with a high thermal resistance. The steady-state temperature rise of the outer casing of the load with respect to the mounting plate is measured by a differential platinum-resistance thermometer in a Wheatstone bridge. Low-frequency power applied to the termination provides a method of calibration. Power in the range of 0 to 2.5 watts may be measured. A 70-second time constant and steady-state temperatures are attained in about 6 minutes. The small physical size of termination (to keep convective and radiative heat losses low) provides high sensitivity. Calibration with lower frequency power is extremely accurate, because the termination is broadband and should exhibit the same power distribution from dc to 10 gigahertz.

A twin calorimeter provides a method of using two calorimetric bodies thermally shielded against ambient temperature variations and improves sensitivity. Figure 3-20 illustrates this type of calorimetric device. The power to be measured is applied to one calorimetric body; the other calorimetric body acts as a temperature reference. The steady-state temperature difference between the two calorimeters is used as a measure of rf power. Calibration is performed by applying low-frequency power. A differential-air, thermometer-type temperature difference indicator, shown in figure 3-21, is used with a twin calorimeter to measure microwave power in the 0.1-mW range. This instrument consists of two similar glass cells connected by a capillary tube containing a liquid pellet. Each glass cell contains a tapered, carbon-coated strip; and the entire assembly is mounted in a rectangular waveguide. Balancing dc power heats one strip; the other strip is heated by rf power. The liquid pellet, which indicates the differential expansion of the air within the two cells, is viewed through an aperture in the waveguide wall, preferably with a microscopy for highest sensitivity. This procedure permits a 2% accuracy at 10 mW.

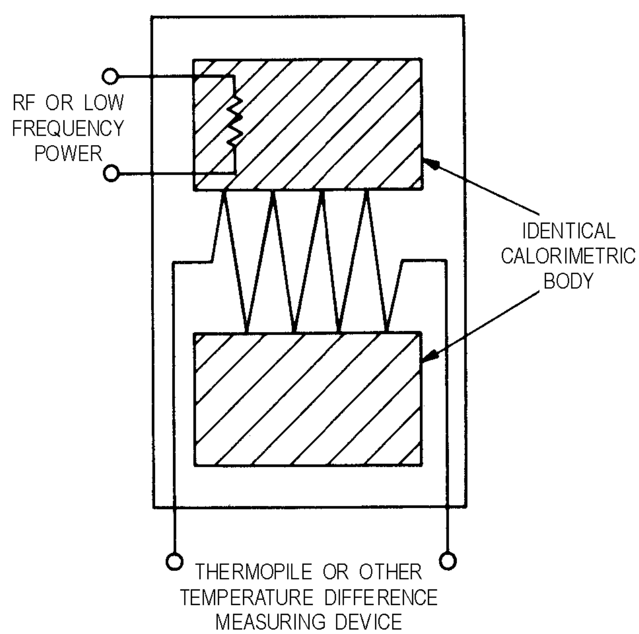


Figure 3-20.—Twin calorimetric system.

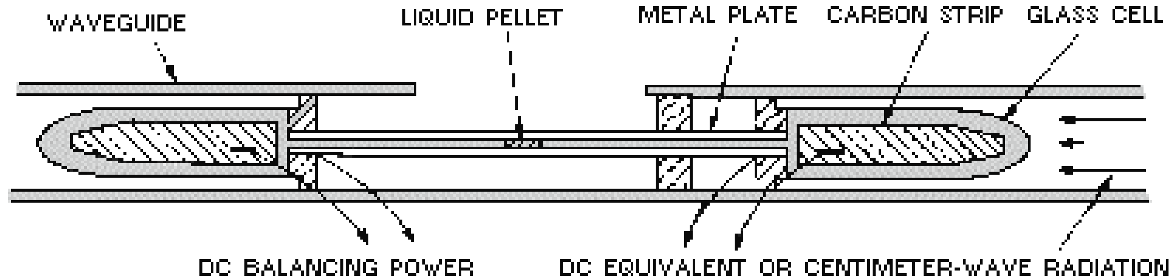


Figure 3-21.—Differential-air, thermometer-type calorimeter.

Flow Calorimeters

Flow calorimeters are classified by the type of circulating method used (open or closed), the type of heating used (direct or indirect), and the type of measurement performed (true calorimetric or substitution). Water or other calorimetric fluid is used only once in an open system. An overflow system is used to maintain a constant rate of flow. Closed systems recirculate the fluid continuously by means of a pump, and a cooling system restores the fluid to ambient temperatures prior to its return to the calorimeter. Closed systems are more elaborate and permit the use of fluids other than water.

Flow calorimeters provide the primary standards for the measurement of high power levels; and, in conjunction with calibrated directional couplers, attenuators, power dividers, or other similar devices serve to standardize medium- and low-power measuring instruments. The measurement time depends on the required time for the entering fluid to reach the outlet, where the rise in temperature is measured. The circulating fluid may serve in a dual capacity as the dissipative medium and coolant, using the direct heating method, or solely as a coolant, using the indirect heating method. Because of its excellent thermal properties and high dielectric losses at 1 GHz or higher, water is normally used in both heating methods. Water is rarely used as the fluid at frequencies lower than 100 MHz, because of insufficient dielectric losses. The indirect heating method offers a wider frequency and power-range coverage and can be used in substitution-type measurements.

True calorimetric measurements contain appreciable error, because of nonuniformity of flow rate, air bubbles, flow-rate measurement inaccuracies, and temperature rise. Flow regulators, bubble traps, and good thermal insulation are required to eliminate the majority of these errors. Substitution methods do not involve direct heat dissipation measurement of moving fluid. Greater accuracy is obtained because known low-frequency power is substituted for the unknown rf power, with all other measurement parameters remaining constant. The accuracy depends on the exactness of the low-frequency power determination and the degree to which factors remain fixed during the substitution of one type of power with another.

Figure 3-22 illustrates a flow calorimeter using low-frequency power substitution. Two different measurement techniques are possible with this type of meter: the calibration technique and the balance technique. The CALIBRATION TECHNIQUE uses an adjustable known power to exactly reproduce the same temperature indication originally obtained by the unknown rf power measurement. The BALANCE TECHNIQUE uses an initial low-frequency power (P_1) to provide a steady-state temperature rise in the calorimetric fluid. When unknown rf power is applied, the original power (P_1) is reduced to a new power (P_2) to maintain the same temperature indication. Therefore, the actual power equals P_1 minus P_2. Figure 3-23 illustrates a widely used method of power measurement using a balanced-flow calorimeter. Temperature-sensitive resistors are bridge-connected as the thermometric elements and are balanced at ambient temperature prior to the application of power. Low-frequency balancing power and the unknown rf power are applied to maintain the bridge at null. This occurs when the temperature rise caused by the unknown rf power equals the temperature rise caused by the known low-frequency power.

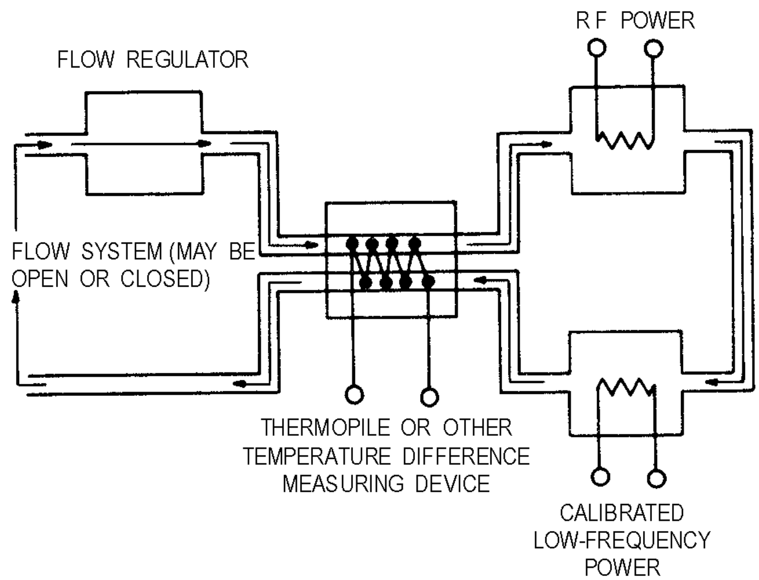


Figure 3-22.—Flow calorimetric system using substitution at low-frequency power.

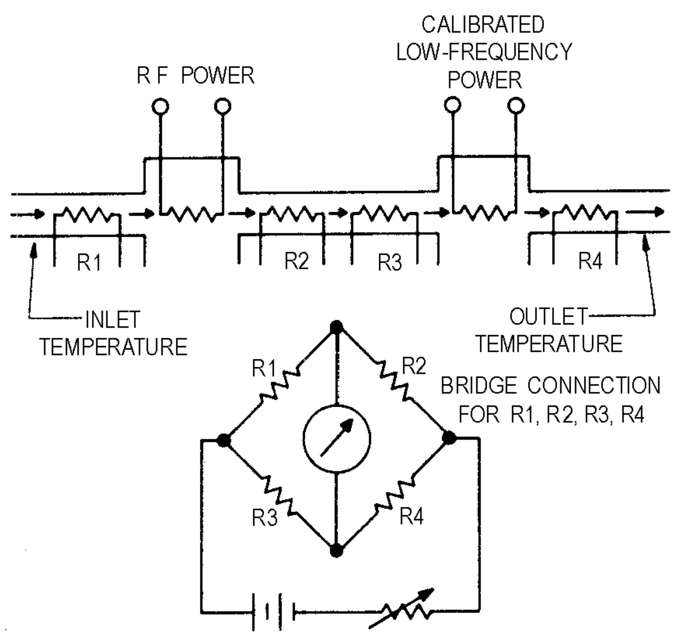


Figure 3-23.—Balanced-flow calorimeter.

Q-16. What is the result of applying power to a calorimeter?

FREQUENCY MEASUREMENTS

Frequency measurements are an essential part of preventive and corrective maintenance for communication and electronic equipment. Rotation frequencies of some mechanical devices must be determined; the output frequency of electric power generators is checked when the engine is started and during preventive maintenance routines; carrier equipment that operates in the audio-frequency range must be adjusted to operate at the correct frequencies; and radio transmitters must be accurately tuned to the assigned frequencies to provide reliable communications and to avoid interfering with radio circuits operating on other frequencies. These are only a few of the applications for making frequency measurements.

FREQUENCY-MEASUREMENT METHODS

Frequency-measuring equipment and devices, particularly those used to determine radio frequencies, constitute a distinct class of test equipment, because of the important and critical nature of such measurements. The requirement of precise calibration is extremely important in all frequency-measuring work. To provide accurate measurements, every type of frequency-measuring device must be calibrated against some frequency standard.

FREQUENCY STANDARDS

Of considerable importance in the measurements of frequency or wavelength are the standards against which frequency-measuring devices are compared and calibrated. Frequency standards belong to two general categories: primary and secondary standards. The PRIMARY FREQUENCY STANDARD maintained by the U.S. National Bureau of Standards has long-term stability and an accuracy of 1 part in 10^{12}, using an atomic clock. A SECONDARY FREQUENCY STANDARD is a highly stable and accurate standard that has been calibrated against the primary standard. Secondary standards are maintained by calibration laboratories that service your test equipment.

The National Bureau of Standards provides time and frequency standards from station WWV at Fort Collins, Colorado, and from station WWVH at Kekaha, Kauai, Hawaii. The following technical radio services are given continuously by these stations:

- Standard radio frequencies
- Standard audio frequencies
- Standard time intervals
- Standard musical pitch
- Time signals
- Radio propagation notices (WWV only)
- Geophysical alerts
- Universal Time Coordinated (UTC)
- + UT^1 Corrections

The UTC scale uses the ATOMIC SECOND as a time interval. UT^1 is based on the earth's uniform rate of rotation. Since the earth's rotation is not precisely uniform, UT^1 is an adjustable interval.

To ensure reliable coverage of the United States and extensive coverage of other parts of the world, radio stations WWV and WWVH provide the primary standard radio frequencies listed in table 3-1. The transmission of WWV and WWVH are interrupted for 5 minutes of each hour. The silent period begins at 15 minutes past the hour for station WWVH and 45 minutes past the hour for station WWV. These silent periods are provided to eliminate errors caused by interference.

Table 3-1.—NBS Frequency Standards and Time Transmission

TRANSMISSION	WWV	WWVH
RF Signal Frequency MHz	5, 10, and 15	5, 10, and 15
Frequency Stability	1 part in 1011	1 part in 1011
Frequency Deviation	1 part in 1012 per day	1 part in 1012 per day
Seconds Frequency and Duration	5 cycles of 1000Hz for .005 seconds	6 cycles of 1200Hz for .005 seconds
Audio Tones	600Hz and 500Hz with 440Hz to mark the hour	600Hz and 500Hz with 440Hz to mark the hour
Frequency Accuracy	1 part in 1012	1 part in 1012
Propagation Forecast	14 min. past the hour (in voice)	None

Two primary standard audio-frequency tones (440 Hz and 600 Hz) are broadcast on all WWV and WWVH carrier frequencies. In the absence of a message, a 500-Hz tone is broadcast during the message interval. The 440-Hz signal that denotes the 1-hour mark is the standard musical pitch, A above middle C. The 600-Hz tone provides a frequency standard for checking the 60-Hz power-line frequency.

The standard time pulse marking interval of 1 second consists of five cycles of a 1,000-Hz tone at WWV and six cycles of a 1,200-Hz tone at WWVH. These marker pulses are heard as clock ticks. Intervals of 1 minute are marked by a 0.8-second, 100-Hz tone for WWV and a 0.8-second, 1,200-Hz tone for WWVH. Each hour is marked by a 0.8-second, 1,500-Hz tone on both stations. Universal Time Coordinated (UTC) is announced on WWVH between the 45 and 52.5 seconds of each minute and on WWV between the 52.5 and 60 seconds of each minute.

An announcement of radio propagation conditions (geophysical alert) for the North Atlantic area is broadcast by station WWV in voice at 18 minutes after each hour. For example, these short-term announcements might state, "The radio propagation quality forecast at ... (normal, unsettled, disturbed)." The propagation format is repeated phonetically and in numerical code to ensure clarity. The letter designations N, U, and W, signifying "normal," unsettled," and "disturbed," respectively, classify the radio propagation conditions at the time of the broadcast. The digits from 1 to 9 indicate the expected radio propagation conditions during the next 6 hours; refer to table 3-2 for code interpretations. The National Bureau of Standards forecasts are based on information obtained from a worldwide network of geophysical and solar observations.

Table 3-2.—NBS Radio Propagation Coding

PHONETIC	PROPAGATION CONDITION
Whiskey	Disturbed
Uniform	Unsettled
Normal	Normal
NUMERAL	
1	Useless
2	Very poor
3	Poor
4	Poor to fair
5	Fair
6	Fair to good
7	Good
8	Very good
9	Excellent

Q-17. What government agency is responsible for monitoring our primary frequency standards?

MECHANICAL ROTATION AND VIBRATION METHODS

There are many instances when you are very much concerned with the question of rotational or vibratory speeds. Knowledge of rotational speeds is necessary where the output of a direct current generator has fallen below a minimum desired output or where the speed of a motor (such as the motor in a teletypewriter or radar antenna) must be maintained at a constant value. There are many instruments that you can use for this purpose, such as tuning forks, stroboscopes, vibrating-reed meters, and electromechanical counters. The oscilloscope and the frequency counter are two of the other devices which may be used, but their use may require the employment of accessory equipment.

Tuning Fork Methods

A tuning fork is generally used in conjunction with the measurement of the rotational speed of a teletypewriter or facsimile motor but is not limited to this application. However, you must remember that the tuning fork can be used at only one frequency, the frequency of vibration for which it was manufactured, and therefore cannot be used on variable-speed motors. To use the tuning fork, you direct a source of light upon the point to be observed. In the case of a teletypewriter, a black-and-white segmented target is painted on the outer circumference of the motor governor. Radial spokes in a flywheel could be used equally well. Permit the motor to reach operational speed under normal load conditions; otherwise, the motor will slow down considerably when the normal load is applied. Strike the tuning fork against the side of your hand to set it into vibration. Then observe the target through the slots in the plates attached to the tines of the fork. The correct speed is obtained when the segments of the target appear to be stationary. If the segments seem to move backward, apparently against the known motor rotational direction, the speed is too low. If the segments seem to move forward, the speed is too high. There is also the possibility that the target segments will appear to jump back and forth or to disappear suddenly. Such erratic action is often because of governor malfunctioning. The correct speed adjustment is reached when the targets appear to be stationary.

Q-18. What is the primary measurement application for tuning forks?

Stroboscope Methods

When using a stroboscope to measure the speed of rotating or reciprocating mechanisms, hold the instrument so that the light from the stroboscope lamp falls directly on the part to be observed. If the part is uniform, or symmetrical, place an identification mark with chalk or a grease pencil on the portion to be observed. This method provides a positive means of identification, because if only one reference mark is observed during measurement, you can be sure that either the fundamental synchronization or a submultiple thereof has been obtained. If the approximate speed of rotation is known, the stroboscope controls may be set to the appropriate positions prior to actual measurement. The main frequency control that determines the rate of the flashing light is then varied until the reference mark on the moving part appears to be standing still. The calibrated scale of the stroboscope will then show the speed directly in revolutions per minute (rpm).

If you have no idea of the speed of the moving part, it is best to start the measurement procedure at the highest frequency that the stroboscope can deliver. The flashing rate of the stroboscope can then be gradually reduced until a single stationary image of the reference mark is obtained. This is the point of fundamental synchronism that corresponds to the speed of the moving part. Do not continue to reduce the flashing rate of the instrument beyond this point without a valid reason for doing so. If you do continue the reduction, a stationary image will still be observed, but the stroboscope will indicate a submultiple of the true rotational speed; thus, a measurement error will be introduced.

Stroboscopes generally have a high- and low-range switch. The typical low range is from 600 to 3,600 rpm, and the upper range is from 3,600 to 15,000 rpm; there is a slight overlap in ranges to ensure reliable frequency coverage. In view of the limitation imposed by flasher tube life, the stroboscope should always be operated at a flashing rate that is as low as possible, consistent with the rotational speed of the observed part. If you should be required to operate this instrument over a long period of time, use a submultiple of the fundamental synchronous speed. The pattern will remain just as stationary, and the tube life will be greatly extended. In addition, the quality of the light is better at the lower ranges than at the upper end of the scale. Sometimes you will encounter a rotating or vibrating device that is moving faster (or slower) than the measuring range of the stroboscope will accommodate. Although such speeds can still be measured, you must use the multiple or submultiple synchronism points.

There are two methods of measuring high speeds. The first method is to obtain a single stationary image of the rotating object at a subharmonic speed relationship and to record that value as *A*. Then obtain a second single stationary image at the next lower subharmonic speed relationship, and record this value as *B*. The unknown speed may then be computed from the following formula:

$$\text{unknown speed} = \frac{A \times B}{A - B}$$

For example, assume reading *A* was 4,000 rpm and reading *B* was 3,500 rpm. The computation would be as follows:

$$\text{unknown speed} = \frac{(4 \times 10^3)(3.5 \times 10^3)}{(4 \times 10^3) - (3.5 \times 10^3)}$$

$$\text{unknown speed} = \frac{14 \times 10^6}{5 \times 10^2}$$

$$\text{unknown speed} = 28,000 \text{ rpm}$$

The second method is used where the value of $A \times B$ becomes progressively smaller. The A reading is obtained as in the previous example (for the sake of easier computation, suppose that the A reading is still 4,000 rpm). Then obtain another submultiple reading for B, keeping in mind the number of times a stationary single image was observed. If a stationary single image was observed seven different times and the final B reading was 2,000 rpm, the calculation would become as follows:

$$\text{unknown speed} = (N_x)\frac{AB}{A - B}$$

$$\text{unknown speed} = (7_x)\frac{(4 \times 10^2)(2 \times 10^2)}{(4 \times 10^2) - (2 \times 10^2)}$$

$$\text{unknown speed} = (7_x)\frac{8 \times 10^6}{2 \times 10^3}$$

$$\text{unknown speed} = 28,000 \text{ rpm}$$

At speeds lower than the lowest range of the stroboscope, multiple images will be observed. For example, assume a dial reading of 900 rpm was obtained when two stationary images were observed. Then dividing the rpm by the number of images will give the unknown shaft speed, as shown below:

$$\text{unknown speed} = \frac{900}{2} = 450 \text{ rpm}$$

WARNING

Exercise caution in using a stroboscope. The *illusion* of stopped motion is very convincing. Do not attempt to touch the moving equipment.

Q-19. If you are required to monitor the speed of a device with a stroboscope over an extended period of time, what step should be taken to prolong the life of the flasher tube?

Frequency Counter Methods

Various frequency counters have found application as an ELECTRONIC TACHOMETER to obtain accurate measurements of high-speed rotating machinery. A tachometer pickup may be used to produce signals that are fed directly to the frequency counter. If the tachometer pickup is designed to generate 1 signal per revolution, the counter will indicate directly in revolutions per second; if the pickup is designed to produce 60 signals per revolution, the counter will indicate directly in revolutions per minute.

AUDIO-FREQUENCY MEASUREMENTS

Audio-frequencies can be measured with a variety of nonelectronic and electronic devices.

Examples of nonelectronic measuring devices are the vibrating-reed meter and the moving-disk frequency meter. (Both of these devices were discussed in NEETS, module 3.) They are used primarily to measure the frequency of ac power, 60 Hz. However, such instruments do not have a wide frequency range. The most common instruments available for the measurement of audio frequencies are oscilloscopes and frequency counters.

OSCILLOSCOPE METHOD

The frequency of a waveform can readily be determined by using an oscilloscope. The most common oscilloscope method of measuring a frequency is accomplished by first measuring the time duration of the waveform. Frequency is the reciprocal of time

$$f = \frac{1}{t}$$

and may be easily computed, as shown in figure 3-24.

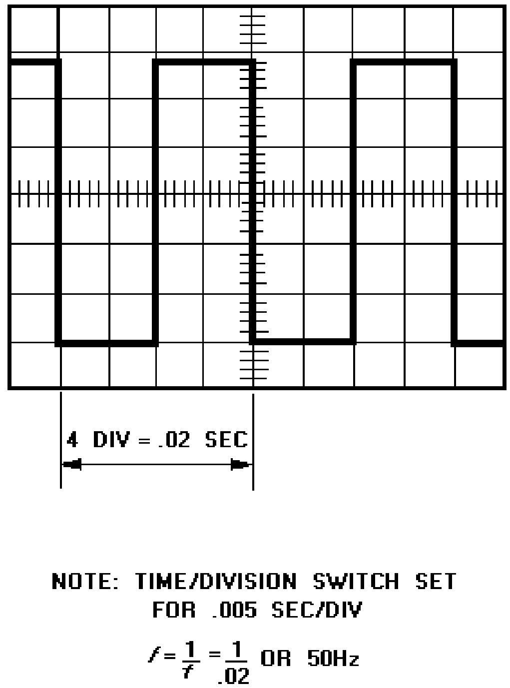


Figure 3-24.—Oscilloscope method of determining frequency.

Another common method of determining the frequency of a waveform is by using Lissajous patterns. This method was discussed in NEETS, module 19.

FREQUENCY COUNTER METHOD

While oscilloscopes can be used to compare rectangular waveforms for the purpose of measuring the frequency of a signal, frequency counters, as shown in figure 3-25, are much more useful for this purpose. The fundamental measurement of frequency is accomplished by totaling the number of cycles into the counter for a precise period of time. The result is then displayed as an exact digital readout. The audio-frequency signal must be of sufficient amplitude to trigger the counter. The AUTO-MANUAL switch provides two methods of frequency counter operation. One method is to initiate the count simultaneously with the initiation of the signal to be measured. With this method, the AUTO-MANUAL switch should be set to the MANUAL position. The second method assumes that the signal to be measured has been operating over some indefinite period of time and that it will continue to do so after a measurement has been taken (hence, only that segment of the signal required to make the frequency measurement is important). With this method, the AUTO-MANUAL switch is to be set to the AUTO position.

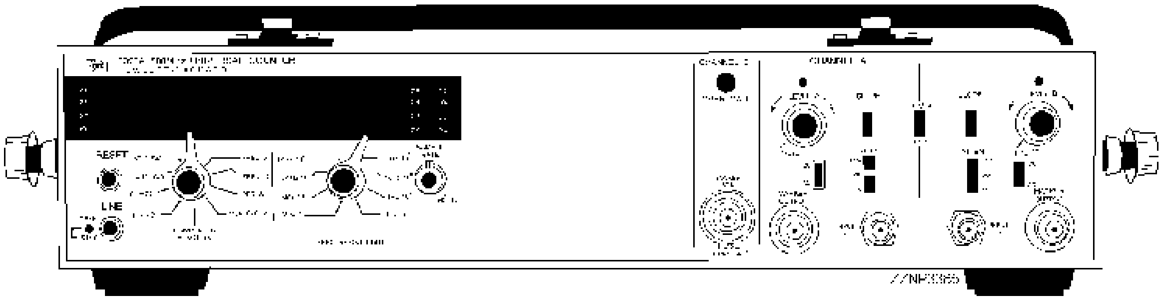


Figure 3-25.—Frequency counter.

RADIO-FREQUENCY (RF) MEASUREMENTS

Radio-frequency measurements are primarily made with frequency counters. Most oscilloscopes are limited in use to approximately 100 MHz. Frequency meters, such as the Hewlett-Packard 530 series, are widely used but lack the accuracy of frequency counters.

Frequency Meters

Prior to the invention of the frequency counter, most frequency measurements above the af range were made primarily with frequency meters.

This process involved heterodyning the frequency to be measured against the calibrated output of the frequency meter to obtain a zero beat from which the measured frequency was then read. This method proved inaccurate because of reading errors.

Frequency meters as we know them today are entirely different from their predecessors. Today's frequency meters (fig. 3-26) contain waveguide or coaxial lines coupled to quarter-wavelength resonant cavities. The meter is adjusted until the cavity is tuned to the resonant frequency of the signal being measured. At resonance, power is absorbed by the cavity and produces a dip in the output-power level, as measured at the frequency meter's output connector. The resonant frequency is read directly from the frequency meter dial and is accurate, in most cases, to approximately ±0.2%. Frequency meters are capable of measuring frequencies in the range of 1 to 40 gigahertz, far exceeding the frequency limitations of the average frequency counter.

Q-20. What happens when a frequency meter is adjusted to the frequency of the signal being measured?

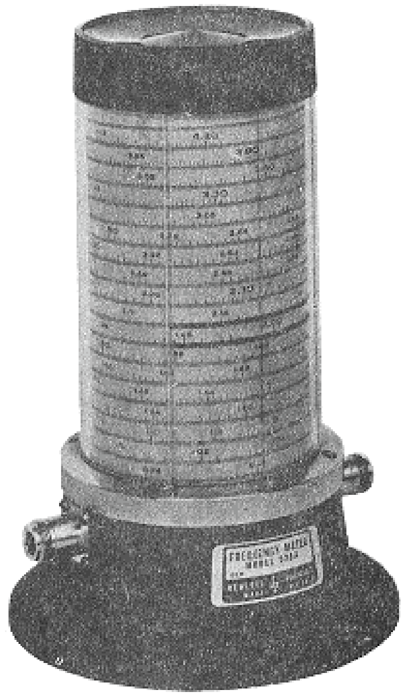


Figure 3-26.—Frequency counter.

Frequency Counters

In the early 1950s, the frequency counter was developed. The device could measure and accurately indicate frequencies up to 10 MHz. Present-day frequency counters can accurately read frequencies as high as 40 GHz. In addition to direct frequency measurement indication, some types of frequency counters can measure the WAVE PERIOD, which is the inverse of frequency; RATIO, which compares one frequency against another; and TIME INTERVAL, the time between two events or the time between two functions of an event. In addition, frequency counters can totalize event indications. This is similar to measuring the frequency except that a manual or an electronic start-stop gate controls the time over which the measurement is taken. Frequency counters can also provide scaling in the form of a digital output signal from the frequency counter that represents a frequency-related division of the input frequency.

All of the above functions have useful applications. For pulse timing, the period function is used; totalizing is used in digital applications; and ratio is used in comparing harmonic-related signals. Scaling is used for triggering other test equipment used in conjunction with the frequency counter; and time-interval capability is used in measuring the interval between two pulses or between two sets of pulses. Because of the wide variety of frequency counters in use, the technical manual for a specific frequency counter should be consulted to determine the instrument's full capabilities.

Frequency Counter Accuracy

All frequency counter measurements are measured with 1 part in 10^8 of accuracy. However, frequency counters have provisions for input from external frequency standards. This extends the accuracy of the frequency to that of the standard. A frequency self-check capability is provided to determine if the counting and lighting circuits are operating properly.

Wavemeters

Wavemeters are calibrated resonant circuits used to measure frequency. Although the accuracy of wavemeters is not as high as that of heterodyne frequency meters, they have the advantage of being comparatively simple and can be easily carried about.

Any type of resonant circuit may be used in wavemeter applications. The exact kind of circuit employed depends on the frequency range for which the meter is intended. Resonant circuits consisting of coils and capacitors are used for low-frequency wavemeters. Butterfly circuits, adjustable transmission line sections, and resonant cavities are used in vhf and microwave instruments.

There are three basic kinds of wavemeters: the absorption, the reaction, and the transmission types. Absorption wavemeters are composed of the basic resonant circuit, a rectifier, and a meter for indicating the amount of current induced into the wavemeter. In use, this type of wavemeter is loosely coupled to the circuit to be measured. The resonant circuit of the wavemeter is then adjusted until the current meter shows a maximum deflection. The frequency of the circuit under test is then determined from the calibrated dial of the wavemeter.

The reaction type derives its name from the fact that it is adjusted until a marked reaction occurs in the circuit being measured. For example, the wavemeter is loosely coupled to an oscillator, and the resonant circuit of the meter is adjusted until it is in resonance with the oscillator frequency. The setting of the wavemeter dial is made by observing the output current of the oscillator. At resonance, the wavemeter circuit takes energy from the oscillator, causing the current to dip sharply. The frequency of the oscillator is then determined from the calibrated dial of the wavemeter.

The transmission wavemeter is an adjustable coupling link. When it is inserted between a source of rf energy and an indicator, energy is transmitted to the indicator only when the wavemeter is tuned to the frequency of the source. Transmission wavemeters are widely used in measuring microwave frequencies.

In figure 3-27, a typical cavity wavemeter is illustrated.

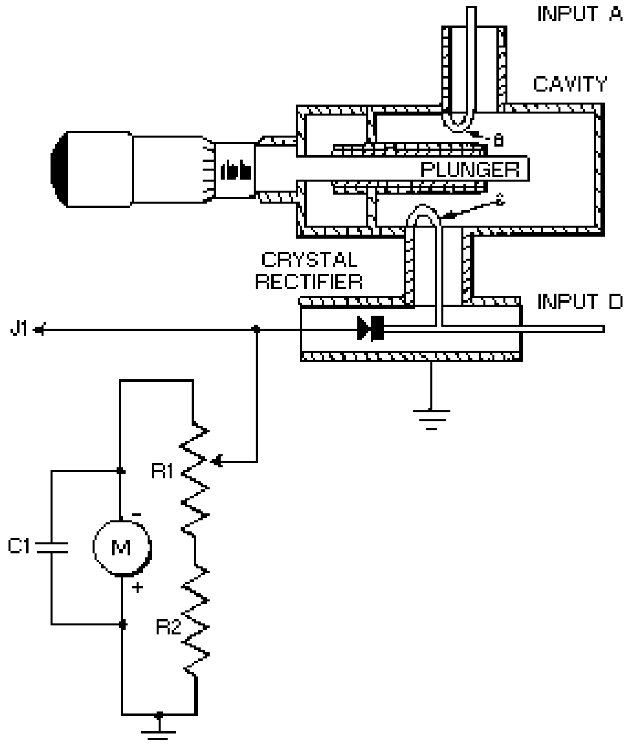


Figure 3-27.—Typical cavity wavemeter.

The wavemeter illustrated is of the type commonly used for the measurement of microwave frequencies. The device employs a resonant cavity, which effectively acts as a high-Q LC tank circuit. The resonant frequency of the cavity is varied by means of a plunger that is mechanically connected to a micrometer mechanism. Movement of the plunger into the cavity reduces the cavity size and increases the resonant frequency. Conversely, an increase in the size of the cavity (made by withdrawing the plunger) lowers the resonant frequency. The microwave energy from the equipment under test is fed into the wavemeter through one of two inputs, A or D. A crystal rectifier then detects or rectifies the signal, and the rectified current is indicated on the current meter, M.

The instrument can be used as either a transmission type or an absorption type of wavemeter. When used as a transmission wavemeter, the unknown signal is coupled into the circuit by means of input A. When the cavity is tuned to the resonant frequency of the signal, energy is coupled through coupling loop B into the cavity and out through loop C to the crystal rectifier where it is rectified and indicated on the meter. At frequencies off resonance little or no current flows in the detector and the meter reading is small. Therefore, the micrometer and attached plunger are varied until a maximum meter reading is obtained. The micrometer setting is then compared with a calibration chart supplied with the wavemeter to determine the unknown frequency.

When the unknown signal is relatively weak, such as the signal from a klystron oscillator, the wavemeter is usually used as an absorption type of device. Connection is made to the instrument at input D. Rf loop C then acts as an injection loop to the cavity. When the cavity is tuned to the resonant frequency of the klystron, maximum energy is absorbed by the cavity, and the current indicated on the meter dips. When the cavity is not tuned to the frequency of the klystron, high current is indicated on the current meter. Therefore, the cavity is tuned for a minimum reading, or dip, in the meter; and the resonant frequency is determined from the micrometer setting and the calibration chart.

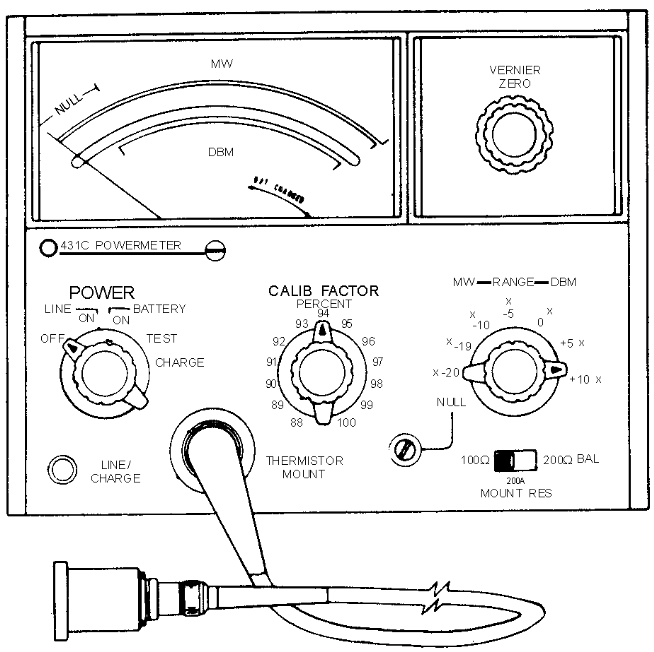
The potentiometer, R1, is used to adjust the sensitivity of the meter from the front panel of the instrument. J1 is a video jack and is provided for observing video waveforms with a test oscilloscope.

SUMMARY

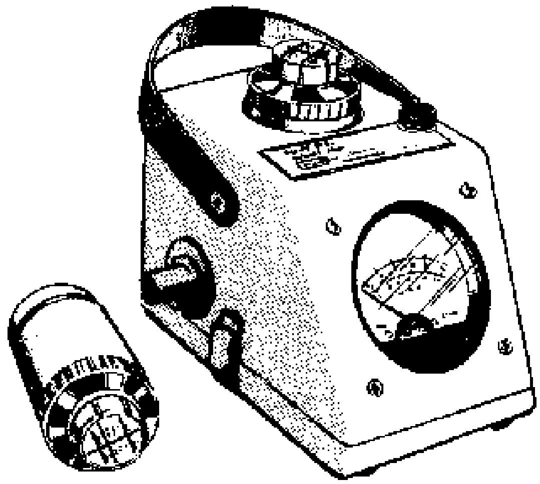
The following is a brief summary of the important points of this chapter.

All **IMPEDANCE BRIDGES** have several things in common. Each type of bridge has a comparing circuit and a measuring circuit. They measure an unknown impedance by comparing the characteristics of the device under test with the characteristics of components within the test set.

POWER METERS that are designed to measure af power can be separated into two distinct groups. Power meters that are designed for measuring sine waves are basically electronic voltmeters calibrated in dB or dBm. VU METERS are designed to measure the average value of complex waveforms, such as a voice.

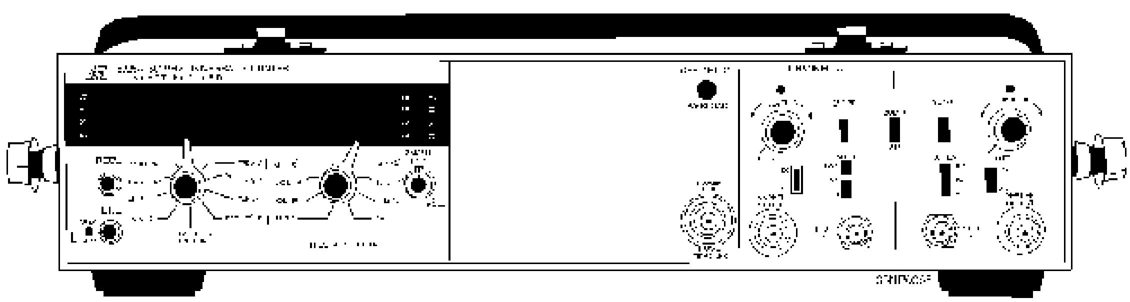


The most common type of test equipment used to measure rf power is the **ABSORPTION POWER** METER. Absorption power meters are designed to absorb all or part of the signal being measured. Examples of absorption power meters are output power meters, in-line wattmeters, and meters employing bolometers.



CALORIMETERS are the most accurate type of test equipment used for measuring high power. As power is applied to a calorimeter, its medium (either liquid or solid) is heated. The heat that is produced is directly proportional to the amount of applied power. The amount of applied power is determined by measuring the change in temperature of the medium.

Today's **ELECTRONIC FREQUENCY COUNTERS** are capable of measuring frequencies from dc to 40 GHz. Most have added features that enable period averaging, time-interval measurements, and scaling. Frequency counter accuracy can be extended by using an external frequency standard in lieu of its internal frequency standard.



REFERENCES

Coaxial Frequency Meter*, NAVSHIPS 0969-092-3010, Naval Ships Systems Command, Washington, D.C., 1978.*

EIMB, Test Methods and Practices*, NAVSEA 08967-LP-000-0130, Naval Sea Systems Command, Washington, D.C., 1980.*

Fire Control Technician (M) 3 & 2*, NAVEDTRA 10209-A, Naval Education Training and Program Development Center, Pensacola, Fla., 1974.*

NEETS, Module 16, Introduction to Test Equipment*, NAVEDTRA 172-16-00-84, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1984.*

ANSWERS TO QUESTIONS Q1 THROUGH Q20.

A-1. *A bridge circuit is balanced when the opposite legs of the comparing and measuring circuits exhibit the same voltage drop.*

A-2. *The capacitive and inductive characteristics of the test leads.*

A-3. *As the supply voltage increases, bridge components may heat up and become less accurate.*

A-4. *Small values of resistances.*

A-5. *A standard capacitor.*

A-6. *Both measure phase angle and magnitude in determining impedance.*

A-7. *High vswr, which equates to poor reception or a loss of power output.*

A-8. *DB, dBm, and vu.*

A-9. *600-ohm load.*

A-10. *DB meters are used for measuring sine waves. Vu meters are used to measure the average value of complex waveforms.*

A-11. *Current transformers.*

A-12. *Electronic wattmeters are capable of measuring high-fequency signals.*

A-13. *Most in-line wattmeters are capable of measuring both forward and reflected power.*

A-14. *Temperature-sensitive material that exhibits a large negative temperature coefficient.*

A-15. *Temperature, mass, and time.*

A-16. *As power is applied, the medium heats up in proportion to the applied power.*

A-17. *The National Bureau of Standards.*

A-18. *They are used to monitor fixed motor speeds.*

A-19. *Monitor a submultiple frequency to prolong the flasher-tube life.*

A-20. *Power is absorbed by the frequency meter cavity; and a pronounced dip in power, at the output, will be observed.*

CHAPTER 4

QUALITATIVE MEASUREMENTS

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to do the following:

1. Identify the various methods of measuring standing-wave ratios.
2. Identify the various methods of determining electrical losses caused by deterioration of transmission lines.
3. Identify the methods of measuring intermodulation distortion.

INTRODUCTION TO QUALITATIVE MEASUREMENTS

As a technician, you are responsible for repairing and maintaining complex electronic systems. The basic ability to repair a specific piece of equipment is only the first step in becoming a qualified technician. Your ultimate goal should be to become proficient at *systems* fault isolation — in other words, to know the entire system like the back of your hand. To reach this goal you will need to be familiar with all parts of the system and know how they are interconnected and interact with each other. There are numerous shortcuts or tricks of the trade that can only be learned through experience on any system, but the most practical thing for you to remember is to approach all problems in a logical manner.

Various combinations of electronic equipment are interconnected to form a system capable of performing specific functions. You must be able to apply general test methods and practices to installation, tuning, maintenance, and repair of the system. This requires you to have a thorough knowledge of many types of electronic equipment. When radar, communication, and digital computers are interconnected, they require different maintenance procedures than when they are operated separately. Revised test procedures may be necessary. Detrimental interactions between equipment or facilities must be corrected and effective preventive maintenance procedures must be planned for all equipment within the system. System quality figures, such as sensitivity and coverage, must be determined and measured during equipment preventive maintenance checks to assure efficient operation. System monitoring at specific test points is often used to help localize a problem.

System testing and monitoring are frequently accomplished by using an external piece of electronic equipment, which is designed specifically for testing a particular system. Some computers and computer systems build in their own monitoring and testing devices and will inform the operator when and where failure has occurred. You must realize that any equipment designed to test, monitor, or repair another system is itself subject to malfunction and will require periodic checks and preventive maintenance. This chapter will cover some of the basic test methods and practices associated with system-level troubleshooting.

STANDING-WAVE RATIO (SWR) MEASUREMENTS

Standing-wave ratio (swr) is the ratio of the maximum voltage or current to the minimum voltage or current at any point along a transmission line. Swr measurements are used to determine the matching quality of the termination of the line.

A variety of methods and test equipments may be used to measure the voltage or current distribution along a transmission line. An open transmission line is accessible for coupling to many types of voltage-measuring devices, such as a wavemeter or a grid-dip meter. However, at higher frequencies where coaxial cables or waveguides are used to minimize skin effect losses, (discussed in NEETS, module 10) access is more complicated. Access to the interior of the waveguide or center conductor of the coaxial cable must be gained by using a unidirectional or bidirectional coupler, which is inserted into the transmission line. The coupler contains a slot into which an rf probe is inserted and positioned with respect to directivity.

The conditions that produce standing waves and their adverse effects are discussed in detail in NEETS, module 10. The different methods of detecting and measuring standing waves are discussed in the following paragraphs.

PROBES

A magnetic or electric probe can be used to observe the standing wave on a short-circuited, terminated line. The wavelength is obtained by measuring the distance between alternate maximum or minimum current points along the line. A typical setup operating at 300 megahertz might use two 10-foot lengths of number 14 phosphor-bronze wires, which are spaced 1 inch apart and supported parallel to a set of probe guide rails. The line should be partially matched to the source generator by means of a parallel-wire shorting stub connected in parallel with the transmission line and the oscillator output line. Figure 4-1, view A and view B, illustrates the types of probes required for this method of measurement.

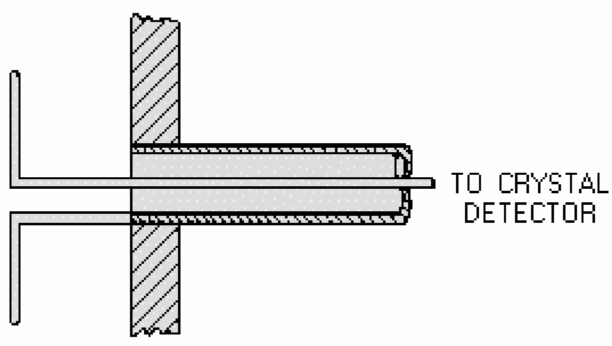


Figure 4-1A.—Typical electromagnetic probe.

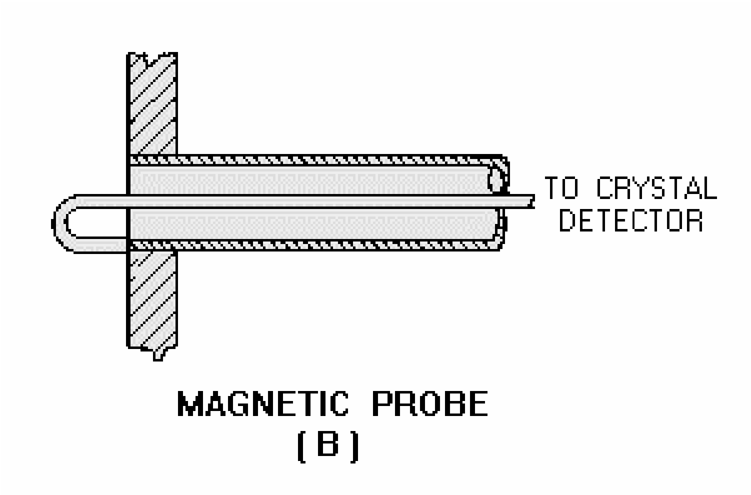


Figure 4-1B.—Typical electromagnetic probe.

NEON-LAMP AND MILLIAMMETER METHODS

In this method of measurement, a neon bulb or milliammeter is moved along the two-wire parallel transmission line. Points of maximum voltage (standing-wave voltage peaks) with the lamp or points of maximum current (standing-wave current peaks) with the indicator will have maximum brilliance or indication, respectively.

Q-1. At what points along a transmission line will a neon lamp glow the brightest?

BRIDGE METHODS

The bridge method permits measurement of the standing-wave ratio without actually measuring the standing waves. The bridge method is applicable because the input impedance of a line terminated in its characteristic impedance is a pure resistance equal to the characteristic impedance. A line terminated in this way can be used as the unknown resistance in a bridge circuit and a null can be obtained in the indicating device when the other resistance arms of the bridge are properly adjusted.

Many types of bridges can be used. For example, an ac bridge that is independent of the applied frequency can be used. The bridge will become unbalanced when the line is no longer properly terminated. Improper termination will produce a reactive component as well as a resistive component in the input impedance of the line and result in a standing wave. The reading of the indicating device depends on the degree of imbalance, which becomes more severe as the mismatch caused by the termination becomes worse. The indicating device can be calibrated directly to indicate the standing-wave ratio. The most common indicator consists of a crystal rectifier, a filtering circuit, and a sensitive dc meter movement in series with a high resistance.

RESISTANCE-CAPACITANCE BRIDGE

A resistance-capacitance bridge circuit is shown in view A of figure 4-2. The bridge is theoretically independent of the applied frequency.

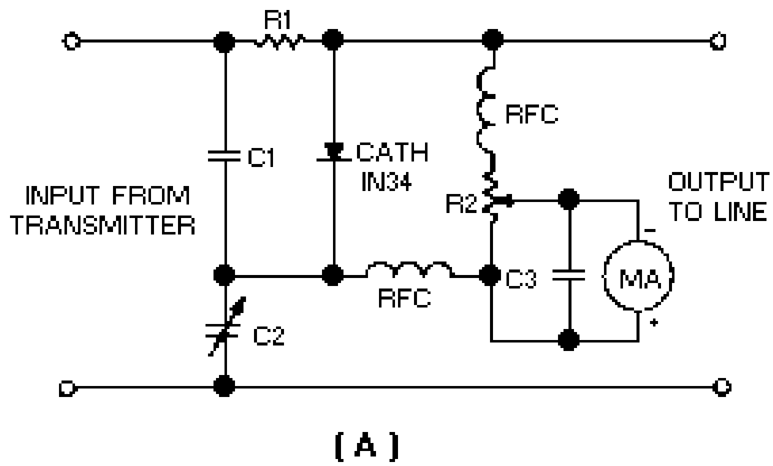


Figure 4-2A.—Resistance-capacitance bridge circuit for measuring standing-wave ratio.

However, the applied frequency must be low enough to avoid skin effect, stray inductance, capacitance, and coupling between circuit elements and wiring. The leads must be kept short to eliminate stray reactance, which causes bridge imbalance. The rectifier circuit wiring must be isolated from other bridge component fields so that induced voltages do not cause an erroneous indication. You should only use resistors having negligible capacitance and inductance effects.

Before you calibrate a newly constructed bridge, the following procedure must be followed if residual readings caused by stray effects are to be held to a minimum:

1. Connect a noninductive resistor (R_L in view B) that is equal to the characteristic impedance of the line to the output terminals of the bridge.

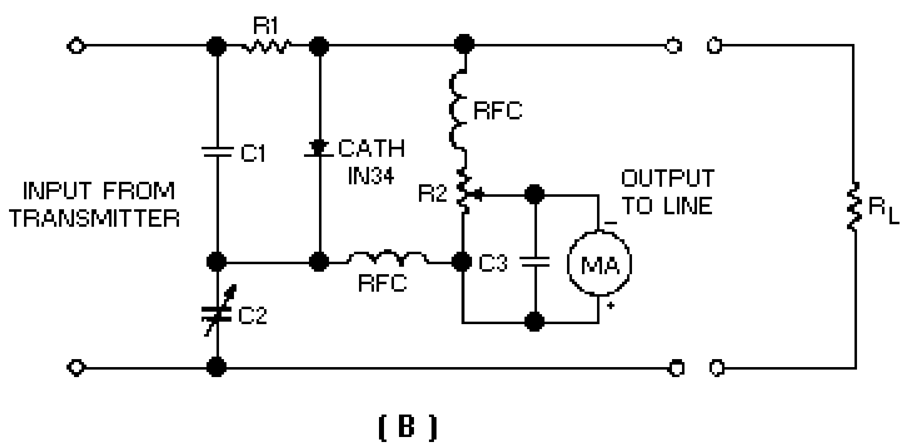


Figure 4-2B.—Resistance-capacitance bridge circuit for measuring standing-wave ratio.

2. Apply an rf voltage to the input terminals and adjust the variable capacitor for a minimum reading on the meter.

3. Reconnect the resistor (R_L) to the input terminals and connect the rf power source to the output terminals.

4. Adjust the rf voltage amplitude applied to the bridge until a full-scale meter reading is obtained.
5. Reconnect the bridge in the normal manner (resistor R_L to the output terminals, etc.). If the meter reading is now more than 1% or 2% of the full-scale reading, different arrangements (lead dress) of the internal wiring must be tried until the null is reduced to 0 or as close as possible to the 0 point.

The bridge can be calibrated after completion of the preceding check. Connect the transmission line under investigation to the output terminals of the bridge and connect a succession of noninductive resistors (R_O in view C) to the load end of the transmission line until the bridge is balanced. Assuming that the bridge was originally balanced for the characteristic impedance of the line, the standing-wave ratio can be computed from the following equation:

$$SWR = \frac{R_L}{R_0} \text{ or } \frac{R_0}{R_L}$$

Where:

R_0 = line impedance

R_L = load resistance

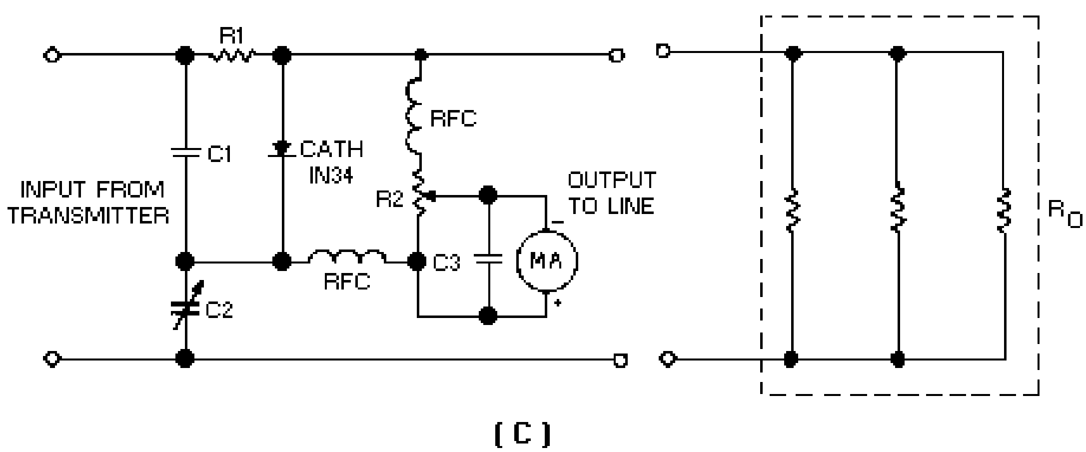


Figure 4-2C.—Resistance-capacitance bridge circuit for measuring standing-wave ratio.

Select the formula that yields a ratio greater than unity. The swr calibration can be recorded on the meter scale directly, recorded on a chart in terms of the meter deflection, or plotted on a graph against the meter deflection. The variable capacitor, in turn, can be calibrated for various characteristic impedances. This is accomplished by applying suitable resistors (R_O) across the output terminals and noting the capacitor settings at the respective balance points. A range of 50 to 300 ohms should prove attainable.

ACCURACY OF BRIDGE MEASUREMENTS

To assure accurate measurements, the rf signal applied to the bridge must be properly adjusted each time a calibrated instrument is used. Essentially, this adjustment is a repetition of the previously described reversed-bridge procedure. The following steps are to be performed:

1. Connect the line to the input terminals of the bridge and connect the transmitter to the output terminals.

2. Adjust the transmitter coupling until full-scale deflection is obtained. From this point on, the coupling must be left untouched.

3. Reconnect the bridge in the usual way and proceed with the measurement.

POWER OUTPUT VERSUS IMPEDANCE MATCHING

For maximum transfer of the power out of an rf source, with minimum heating from reflected power, the total output impedance sensed by the rf source must be equal to the internal impedance of that source. A perfect impedance match between transmitter and load would exist if the swr were 1 to 1. As discussed in NEETS, module 10, test equipment designed to measure the instantaneous voltage of a standing wave will give you a voltage standing-wave ratio (vswr). Test equipment designed to measure the instantaneous current component of a standing wave will give you the current standing-wave ratio (iswr). Regardless of the type of test equipment selected, both ratios will be the same.

Q-2. What vswr is a perfect match between a transmitter and its load?

SWR METERS

The Hewlett-Packard Model 415E swr meter, shown in figure 4-3, is a commonly used swr meter. It is extremely accurate, sensitive, lightweight, easy to use, and portable. It is essentially a high-gain, tuned audio amplifier with a square-law meter that is calibrated to read swr directly. The meter is designed to be operated at a mean center frequency of 1,000 hertz.

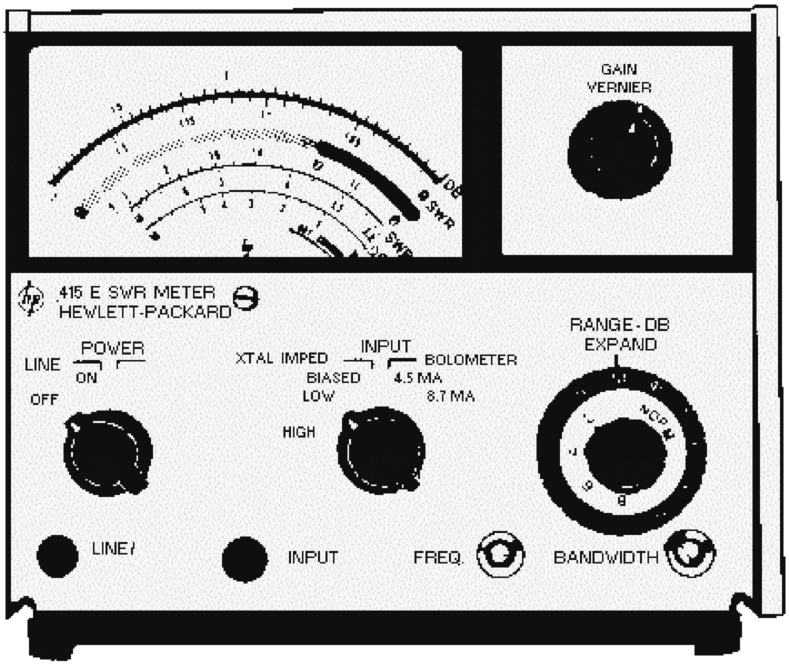


Figure 4-3.—Typical swr meter.

Figure 4-4 shows a typical swr measurement setup using the swr meter. The signal source is usually a sinusoidal wave that is square-wave modulated at 1,000 hertz.

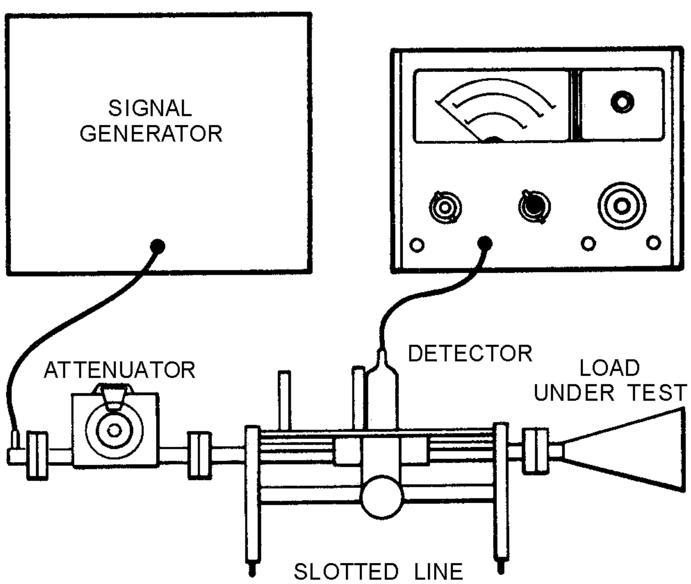


Figure 4-4.—Typical setup for measuring swr.

The swr meter usually gets its input from a detector, either a barretter or a crystal diode. This detector must be a square-law device (its output voltage is proportional to the applied rf power) to ensure the accuracy of the meter. The input is amplified and applied directly to the meter. To perform the measurement as shown in figure 4-4, you move the detector along the slotted line so that its probe is at a voltage maximum and adjust the gain of the meter with the RANGE-DB, GAIN, and VERNIER controls (EXPAND switch to NORM) for full-scale deflection (1.0 on the 1.0 to 4 SWR scale). Then move the probe toward a minimum. If the meter drops below 3.2, rotate the RANGE-DB switch one position clockwise and read on the 3.2 to 10 SWR scale. If the pointer drops below this scale, rotate the RANGE-DB switch one more position clockwise and read on the 1.0 to 4 scale and multiply by 10. This pattern continues for still higher swr readings.

The dB scales can be used for a standing-wave-ratio measurement by setting the meter to full scale at a voltage maximum, then turning the RANGE-DB switch clockwise for an on-scale reading at a voltage minimum and noting the difference in dB reading at the maximum and minimum. A dB reading is obtained by adding the RANGE-DB switch setting and meter indication. The swr meter may also be used for high resolution insertion loss measurements.

The setup for performing insertion loss or attenuation measurements is shown in figure 4-5. It requires that you initially establish a convenient reference on the DB scale of the meter. This is accomplished by connecting the signal source directly to the detector and using the GAIN and VERNIER controls to adjust the meter pointer to a convenient reference. Then you can insert the device to be measured between the signal source and the detector and note the change in dB, as shown on the meter.

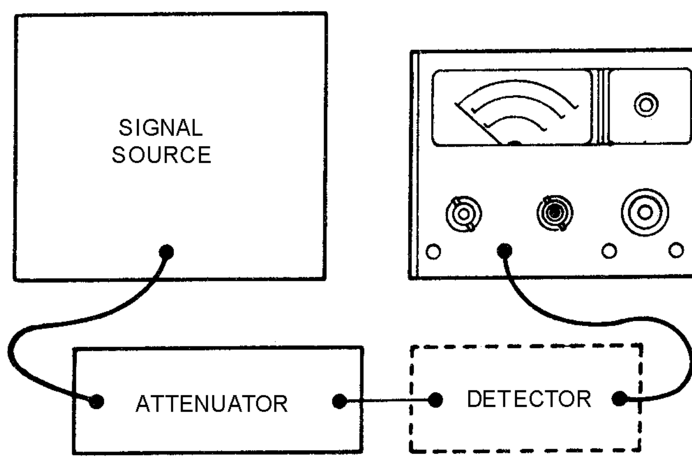


Figure 4-5.—Typical setup for measuring attenuation or insertion loss.

ATTENUATION AND INSERTION LOSS MEASUREMENTS OF TRANSMISSION LINES

Transmission lines are sometimes subjected to extremes of weather and the corrosive effects of salt water. You should be aware of the adverse effects of this environment on transmission lines and how to determine electrical losses caused by transmission-line deterioration.

Q-3. What are the two common causes of transmission-line deterioration?

LOSS MEASUREMENT

Insertion loss measurement of transmission lines requires the use of a good signal generator and an accurate power meter. The method is identical to the insertion loss measurements used on most couplers. When a known frequency, at a predetermined level of power, is inserted into one end of a transmission line, then the same frequency and the same level of power should be transmitted to the other end of a transmission line. Because all transmission lines contain some degree of resistance, some loss of power will occur during the test.

Exposure to the elements over a period of time causes transmission-line deterioration. To determine the accuracy of this test, you should use the power meter to measure the output of the signal generator at the end of the test cable to be attached to the transmission line. Any power loss associated with the test cables should be recorded and subtracted from the measurement taken with the transmission line connected.

You should note that transmission lines, like all other electronic components, are designed to operate over a specific range of frequencies. It is not uncommon for a transmission line to operate improperly at one frequency, yet operate properly over the remainder of its frequency spectrum. You should check transmission-line losses over their entire frequency range. Insertion loss measurements are normally taken when a system is first installed or the transmission line is replaced. Periodic measurements should be performed to enable you to determine if system performance is being degraded by transmission-line deterioration.

Q-4. Is it possible for a transmission line to operate improperly at certain frequencies and properly at others?

TRANSMISSION-LINE FORMULAS

Transmission lines are engineered and manufactured to meet certain specifications. The most important of these specifications relates to frequency, power-handling capabilities, and characteristic impedance. The dielectric constant (K) of the insulating material is probably the manufacturer's most important consideration and is the primary factor that affects the size of the coaxial cable. The formulas in the following sections discuss some aspects of coaxial transmission-line engineering.

A cross section of a coaxial line is shown in figure 4-6. The characteristic impedance of a coaxial line can be determined by the following formula:

$$Z_O = \frac{138}{\sqrt{K}} \left(\log \frac{D}{d}\right)$$

Where:

D = the inside diameter of the outer conductor

d = the outside diameter of the inner conductor and must be expressed in the same units as D

K = the dielectric constant of the insulating material (See table 4-1.)

Z_O = the characteristic impedance

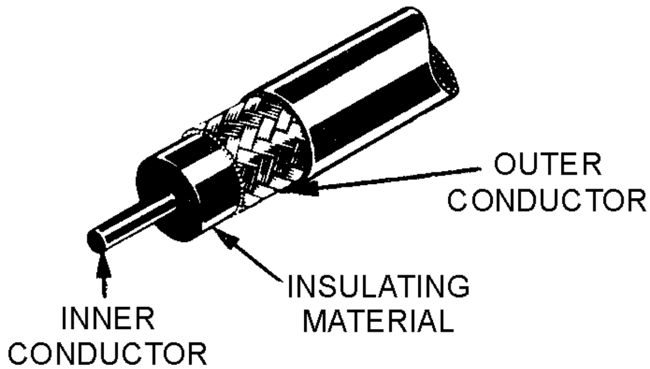


Figure 4-6.—Coaxial line.

Table 4-1.—Dielectric Constants of Materials

Material	Dielectric constant (Approx.)	Material	Dielectric constant (Approx.)
Air	1.0	Lucite	2.5
Amber	2.6-2.7	Mica (electrical)	4.0-9.0
Asbestos Fiber	3.1-4.8	Mica (clear India)	7.5
Bakelite (asbestos base)	5.0-22	Mica (filled phenolic)	4.2-5.2
Bakelite (mica filled)	4.5-4.8	Micaglass (titanium dioxide)	9.0-9.3
Barium Titanate	100-1250	Micarta	3.2-5.5
Beeswax	2.4-2.8	Mycalex	7.3-9.3
Cambric (varnished)	4.0	Neoprene	4.0-6.7
Carbon Tetrachloride	2.17	Nylon	3.4-22.4
Celluloid	4.0	Paper (dry)	1.5-3.0
Cellulose Acetate	2.9-4.5	Paper (coated)	2.5-4.0
Durite	4.7-5.1	Paraffin (solid)	2.0-3.0
Ebonite	2.7	Plexiglas	2.6-3.5
Epoxy Resin	3.4-3.7	Polycarbonate	2.9-3.2
Ethyl Alcohol (absolute)	6.5-25	Polyethylene	2.5
Fiber	5.0	Polyimide	3.4-3.5
Formica	3.6-6.0	Polystyrene	2.4-3.0
Glass (electrical)	3.8-14.5	Porcelain (dry process)	5.0-6.5
Glass (photographic)	7.5	Porcelain (wet process)	5.8-6.5
Glass (Pyrex)	4.6-5.0	Quartz	5.0
Glass (window)	7.6	Quartz (fused)	3.78
Gutta Percha	2.4-2.6	Rubber (hard)	2.0-4.0
Isolantite	6.1	Ruby Mica	5.4
Selenium (amorphous)	6.0	Styrofoam	1.03
Shellac (natural)	2.9-3.9	Teflon	2.1
Silicone (glass) (molding)	3.2-4.7	Titanium Dioxide	100
Silicone (glass) (laminate)	3.7-4.3	Vaseline	2.16
Slate	7.0	Vinylite	2.7-7.5
Soil (dry)	2.4-2.9	Water (distilled)	34-78
Steatite (ceramic)	5.2-6.3	Waxes, mineral	2.2-2.3
Stearite (low loss)	4.4	Wood (dry)	1.4-2.9

Attenuation in a coaxial line in terms of decibels per foot can be determined by the following formula:

$$a = \frac{4.6\sqrt{f}\,(D + d)}{D \times d\,(\log\frac{D}{d})}$$

Where:

D = the inside diameter of the outer conductor (in inches)

d = the outside diameter of the inner conductor (in inches)

f = the frequency (in megahertz)

a = the attenuation (in decibels per foot of line)

As a technician, you need not be concerned with designing coaxial transmission lines. It is, however, our feeling that you should be familiar with the parameters that go into making a transmission line. It can readily be seen by the above formulas that transmission lines cannot be randomly selected without consideration of system requirements. NAVSHIPS 0967-000-0140, EIMB, *Reference Data*, section 3, lists the characteristics of most common transmission lines.

Q-5. What factor has the greatest effect on the physical size of a coaxial cable?

Q-6. Is the attenuation of a coaxial cable independent of frequency?

INTERMODULATION DISTORTION MEASUREMENTS

Intermodulation distortion occurs when two or more frequencies become mixed across a nonlinear device. The resultants are the difference frequency and the sum frequency, both components of the originals. Undesirable frequencies can be generated by a mixing of two discrete frequencies. Spurious radiation, arising from close spacing of transmitter and receiver, is a prime source of an undesirable frequency that can cause intermodulation distortion in an electronic circuit. This is particularly the case when antenna couplers are employed. Cross modulation and parasitic generation (described in the next section) are two other sources of undesirable frequencies that may cause intermodulation distortion.

Q-7. What is the main cause of intermodulation distortion?

CROSS MODULATION AND PARASITIC GENERATION

CROSS MODULATION occurs when a signal from an adjacent channel crosses over into a second channel and modulates the frequency of the second channel. PARASITIC GENERATION occurs when regenerative feedback is sufficient to cause a circuit to oscillate, even though it is not designed to oscillate. Both types of distortion are common to systems that are misaligned.

INTERMODULATION DISTORTION DETECTION

The presence of intermodulation distortion is determined by a two-tone test method. Two sinusoidal frequencies of equal amplitude are introduced into the system under test. The two frequencies are spaced close together with reference to the unit under test. The output of the system under test (an amplifier, receiver, or transmitter) is monitored on a spectrum analyzer that is comparable in characteristics to the suspect system. The resultant display should be an exact reproduction of the input frequencies. If not, some form of intermodulation distortion is present. To determine if external sources are causing the intermodulation distortion, you can use a single-frequency signal. If the display on the spectrum analyzer does not show the single frequency, then intermodulation distortion is present.

Intermodulation distortion cannot be entirely suppressed, but it can be minimized by shielding components and circuitry, parasitic suppression circuitry, and antenna spacing. These factors are incorporated in the design of the system and are tested during production. Any shields or parasitic suppressors that are removed by the technician must be replaced before troubleshooting and/or repair can be effective. Antenna locations also pose a consideration when installing a new system. Ship alteration specifications must be observed when new antenna systems are being installed.

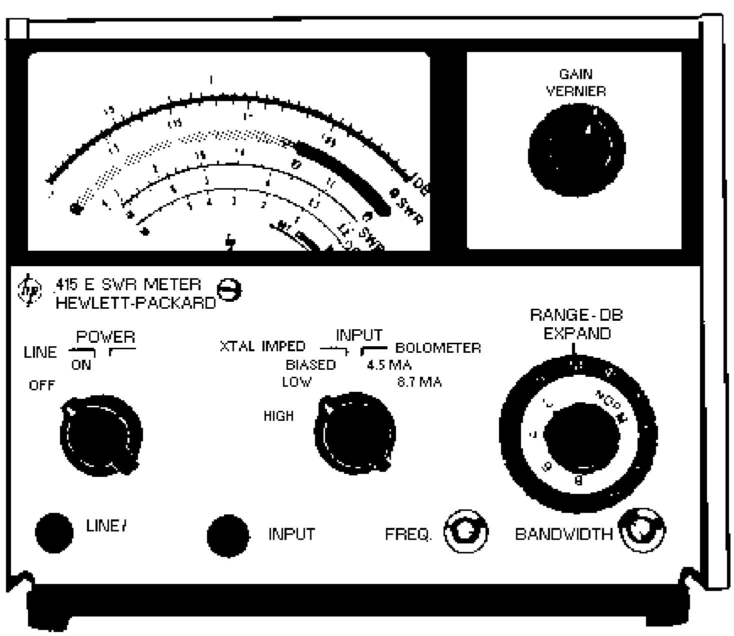
Q-8. When you are testing a piece of equipment for intermodulation distortion, what should the output of the equipment look like?

SUMMARY

The important points of this chapter are summarized in the following paragraphs:

STANDING WAVES are the result of an impedance mismatch between a transmission line and its load. If a transmission line is not properly terminated, it will cause a percentage of the transmitter power to be reflected back to the source. The reflected wave or standing wave will increase in magnitude as the mismatch becomes greater.

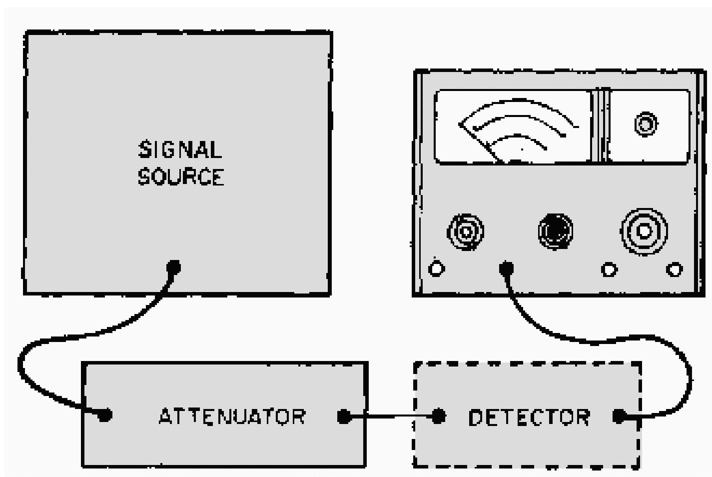
VSWR refers to the voltage ratio of the incident wave (that which is transmitted to the load) and the reflected wave (that which is reflected by the load back to the transmitter). An ideal vswr is considered to be 1 to 1.



Standing waves that are present on a transmission line can be used to determine the **TRANSMITTER FREQUENCY**. Voltage or current peaks are present at half-wavelength intervals. By measuring the distance between peaks, you can compute frequency mathematically.

TWO-WIRE, PARALLEL TRANSMISSION LINES are usually tested for standing waves with test devices that are inductively coupled to the line. These test devices vary greatly in their complexity, ranging from bridge circuits to simple neon lamps.

INSERTION LOSS MEASUREMENTS are performed by injecting a signal of a known amplitude into a transmission line and then monitoring the signal at the far end of the cable with a power meter. Loss measurements must be taken at various frequencies to determine if the transmission line is good across its frequency range.



The most common cause of **INTERMODULATION DISTORTION** is improper spacing of transmitters and receivers. CROSS MODULATION is common to equipment that is misaligned. Intermodulation distortion can be tested by injecting two signals (different frequencies) into a piece of equipment and then monitoring its output for distortion using a spectrum analyzer. Intermodulation distortion is usually caused by improper antenna spacing or by poorly shielded components or circuits.

REFERENCES

EIMB, *Test Methods and Practices Handbook*, NAVSEA 0967-LP-000-0130, Naval Sea Systems Command, Washington, D.C., 1980.

NEETS, Module 10, *Wave Propagation*, Transmission Lines, and Antennas, NAVEDTRA 172-10-00-83, Naval Education Training and Program Development Center, Pensacola, Fla., 1983.

SWR Meter 415E, NAVSHIPS 0969-139-2010, Hewlett-Packard Co., Palo Alto, Calif. 1968.

ANSWERS TO QUESTIONS Q1. THROUGH Q8.

A-1. At standing-wave voltage peaks.

A-2. 1 to 1.

A-3. Corrosive effects of salt water and weather extremes.

A-4. Yes, it is quite common.

A-5. The dielectric constant of the insulating material.

A-6. No.

A-7. Close spacing of transmitters and receivers.

A-8. An exact reproduction of the input.

CHAPTER 5

INTRODUCTION TO WAVEFORM INTERPRETATION

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to do the following:

1. Explain the use of waveform interpretation in testing applications.
2. Identify the different types of modulation and methods of measuring modulation.
3. Explain the various uses of spectrum analyzers.
4. Explain the various uses of time-domain reflectometers.
5. Identify the various tests that can be performed with the swept-frequency technique.

INTRODUCTION TO WAVEFORM INTERPRETATION

Measurements performed with oscilloscopes, time-domain reflectometers, and spectrum analyzers enable you to view the signal produced by the equipment or circuit under test. However, a visual display is of no value unless you are able to interpret the signal characteristics.

A displayed waveform is a representation of a varying signal related to time. You can graphically plot an unknown waveform by using a system of coordinates in which the amplitude of the unknown signal is plotted linearly against time. An analysis of the resultant waveform provides you with valuable information in determining the characteristics of many electronic (and some mechanical) devices. For example, the waveform of a signal may indicate the presence of harmonics or parasitic oscillations, or it may indicate how closely a device is following a desired cycle of operation. As the parts in an amplifier begin to shift in value or deteriorate, waveform distortion often occurs and indicates abnormal operation of a circuit and often precedes circuit breakdown. Malfunctioning of electrical or electronic circuits within equipment can usually be traced, by waveform inspection, to a specific part or parts of the circuit responsible for the distorted signal. On the basis of these facts, it is apparent that there is an important need for test equipment that can provide a visual presentation of a waveform at the instant of its occurrence in a circuit.

DISTORTION is a term used by technicians and engineers alike that generally signifies dissatisfaction with the shape of the wave processed by an amplifier. Distortion of a waveform is the undesired change or deviation in the shape of the observed signal with respect to a reference waveform. Classifying any waveform as a *distorted* wave without reference to the electronic circuitry involved is meaningless. A waveform that can be validly termed distorted with respect to a specific amplifier circuit may be the normal waveform to be expected from another amplifier circuit. One of the most important steps in waveform analysis, the one that usually proves the most difficult for the maintenance personnel, is the interpretation of patterns viewed on the test equipment.

This chapter will cover some of the basic test methods and practices associated with waveform interpretation.

MODULATION MEASUREMENTS

Modulation measurements are sometimes required during tuning procedures to adjust transmitting equipment for the proper amount of modulation. During maintenance tests of modulated transmitter equipment, you should determine the amount of distortion in the output signal and the modulation level or index. The modulation level in multiplexing equipment is usually set at the factory or during corrective maintenance procedures. Proper adjustment of the input signal level and automatic signal-level regulation circuits provides the correct amount of modulation. Defects in modulation circuits of a transmitter can be detected by measurements of the quality of the received signals at the receiver. Corrective maintenance analysis of multiplex equipment modulation circuits can usually be made by signal-level measurements.

Some radio transmitters, when operating in the AM mode, must be adjusted for correct modulation during normal tuning procedures. If the modulation level is low, the transmitter is not operating at its maximum efficiency. On the other hand, modulation in excess of 100% produces serious distortion. Since neither of these conditions is desirable, amplitude modulation should be maintained between 60% and 95% when possible. The modulation level or index of AM and fm radio transmitters that operate in the vhf range is initially adjusted by the manufacturer or during corrective maintenance. The amplifier gain of the modulator can be initially adjusted by reference to the modulation meter provided on the front panel of the equipment.

Pulse modulation of radar and radio beacon signals can be measured by waveform displays presented on a standard oscilloscope. The amount of usable energy in a pulsed waveform, as measured by a spectrum analyzer, is also an indication of the pulse modulation quality.

Attaining 100% amplitude modulation of an rf carrier with a sine wave requires a modulating power equal to one-half of the rf carrier power. Under this condition, the average power of the modulated carrier is equal to 1.5 times the average unmodulated carrier power. The added power is divided equally between the upper and lower sidebands. During the peaks of 100% modulation, the amplitude of the carrier is doubled. This will cause the instantaneous peak power to be four times the instantaneous unmodulated peak power $P = E^2/R$. When voice modulation is employed, only the highest amplitude peaks can be allowed to modulate the carrier 100%. Since many speech components do not modulate the carrier 100%, the average power required for voice modulation is less than that required for modulation with a sine wave. Voice peaks usually modulate a carrier 100% when the modulation increases the average carrier output power 25% over its normal value.

Q-1. What is the result of overmodulating an AM signal?

Q-2. For AM transmissions, the carrier is normally modulated within what range?

AMPLITUDE-MODULATION MEASUREMENTS

An increase in the power output of an AM transmitter is indicated by an increase in antenna current. The increase can be taken as a measure of the degree of modulation and can be expressed as a percentage, as shown in figure 5-1. The graph for this figure was developed from the relationship existing between the carrier power and the increased power resulting from the added modulation power. The formula for calculating the PERCENTAGE of MODULATION is as follows:

$$\text{Percentage of modulation} = \frac{100\ (E_{max} - E_0}{E_0}$$

Where:

E_{max} = the highest peak voltage

E_0 = the unmodulated carrier voltage

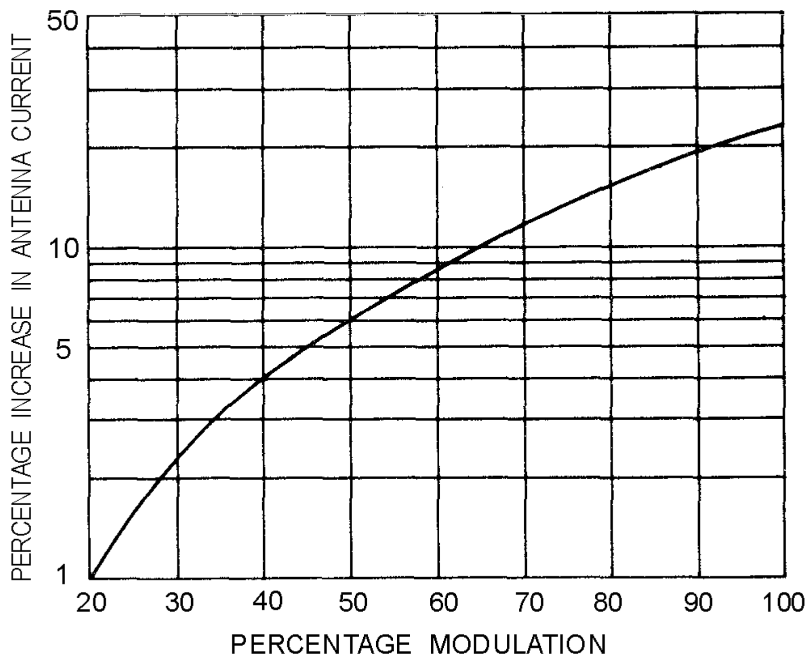


Figure 5-1.—Antenna current increase with amplitude modulation.

The use of this formula is based on the assumption that the modulating voltage is a pure sine wave. Normal broadcasting, however, is characterized by complex envelope patterns, as illustrated in figure 5-2. In this light, the previous formula is not so clear. Consequently, the preceding formula should be viewed more correctly as the PERCENTAGE OF POSITIVE PEAK MODULATION. When the minimum voltage (E_{min}) rather than the peak voltage (E_{max}) is used to compute percentage of modulation, the computed percentage (shown below) is the PERCENTAGE OF NEGATIVE PEAK MODULATION:

$$\text{negative peak percentage of modulation} = \frac{100(E_0 - E_{min})}{E_0}$$

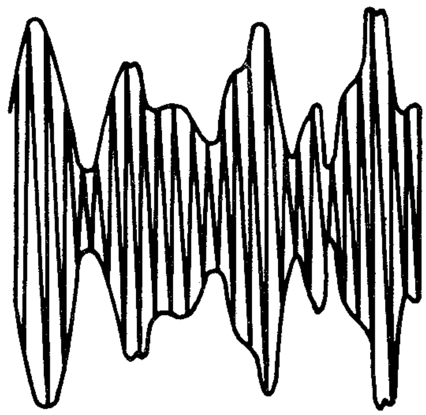


Figure 5-2.—Rf carrier amplitude-modulated by a complex wave envelope.

Since the preceding two modulation percentages often differ, you should define the AVERAGE PERCENTAGE OF MODULATION, as shown below (refer to fig. 5-3):

$$\text{average percentage of modulation} = \frac{100\ E_{max} - E_{min}}{2 \times E_o}$$

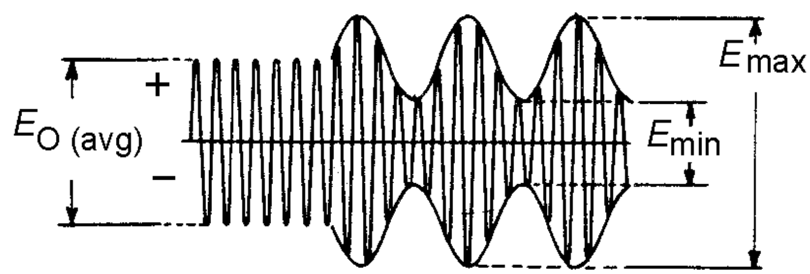


Figure 5-3.—Rf amplitude percentage modulation wave envelope.

From the preceding definitions of percentage of modulation, you should note that methods of *measuring* all three types of modulation percentages must be devised. When differing values are obtained, however, the cause may not necessarily be directly related to unequal positive and negative peaks of a complex modulation wave. Another possibility is distortion caused by carrier shift. Distortion may also be produced by effects other than the modulation process — for example, parasitic oscillation, nonlinear radio-frequency amplification of modulated signals, and distortion present in the audio amplifiers.

Unfortunately, continuous variations in the percentage of modulation create a number of additional problems. For example, damping is necessary so that a meter can provide an average reading despite fluctuations. An average reading, on the other hand, will not disclose the presence of transient overmodulation. This shortcoming is serious because of the large number of sideband frequencies produced in addition to the normal ones whenever overmodulation occurs. Not only do these extra frequencies interfere drastically with other transmissions, but they also may significantly distort the modulation signal. These considerations account for the importance of using a meter that responds to modulation peak; specifically, both positive-peak and negative-peak overmodulation must be indicated. Positive-peak overmodulation occurs when the *positive* modulation exceeds 100%; negative-peak overmodulation occurs when the *negative* modulation exceeds 100%.

Oscilloscope Measurement Methods

The oscilloscope is widely used as an amplitude-modulation monitor and measuring instrument. Since it is capable of presenting visual indications of the modulated output of AM transmitters, the oscilloscope is reliable for detecting overmodulation and determining the percentage of modulation. For example, the relative error of most measurements taken with a 5-inch crt is about 10%. Although such accuracy is adequate for many maintenance checks, the oscilloscope is usually considered more valuable as a monitor of general modulation conditions. It is also used to monitor the amplitude-modulated output of a radio transmitter when photographic records are desired.

Types of Modulation Display

Two types of modulation patterns are provided by the oscilloscope, depending upon the hookup used. These patterns are the WAVE-ENVELOPE PATTERNS, as shown in figures 5-2 and 5-3, and the TRAPEZOIDAL PATTERN, as shown in figure 5-4.

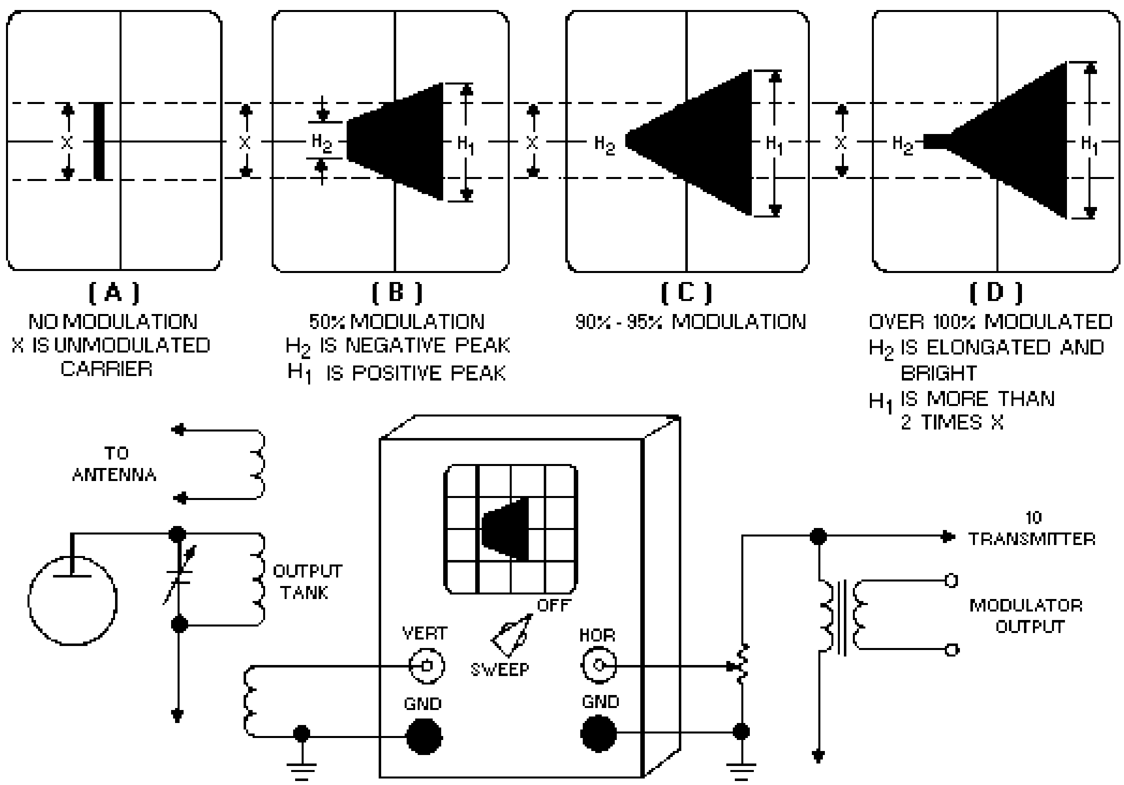


Figure 5-4.—Trapezoidal modulation patterns.

Figure 5-2 shows an oscilloscope presentation of an rf carrier that is amplitude-modulated by a complex wave, such as that of speech. Figures 5-4 and 5-5 show the effects of over 100% modulation on the carrier wave. The carrier wave envelope pattern (as shown in fig. 5-3) is obtained by applying the rf-modulated wave to the vertical input of the oscilloscope. The trapezoidal pattern is obtained in a similar manner except that the modulation signal from the transmitter is used to horizontally sweep the oscilloscope (instead of having the sweep signal generated internally by the oscilloscope). Both methods are limited by the frequency response of the oscilloscope; therefore, these methods find greater applicability in the lf to hf ranges.

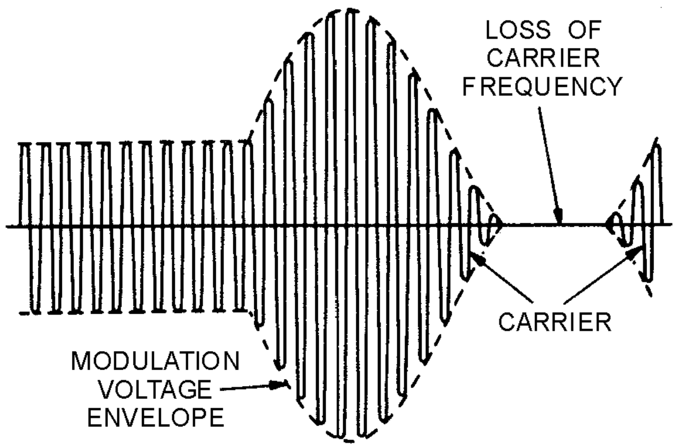


Figure 5-5.—Overmodulated rf carrier.

VHF AND UHF MEASUREMENTS

In the vhf and uhf ranges, modulation is normally measured by applying a specific-level, 1-kilohertz tone to the input of the modulator. This, in turn, produces a significant drop in the plate voltage of the final output stage of the modulator. The correct setting of output plate voltage ensures that overmodulation will not occur.

SINGLE-SIDEBAND MEASUREMENTS

Single-sideband modulation is a form of amplitude modulation in which only one sideband is transmitted with a suppressed carrier. Since balanced modulators are used to provide carrier cancellation, the exact balancing of the carriers to provide cancellation requires a null adjustment. The null can be observed and adjusted by using either a detector and an indicator, such as a voltmeter, or an oscilloscope for observation of the output while tuning the transmitter.

Measurements peculiar to sideband technology also include special modulation-amplitude and modulation-distortion checks. If the sideband modulator is overdriven or mistuned or the associated linear amplifiers are improperly loaded or overdriven, spurious output frequencies are produced. These are harmonically related to the driving signals and can cause splatter over a large range of frequencies, thus causing interference to other transmitting stations.

To determine the proper amplitude so that the modulation will not cause distortion or splatter, you use the audio two-tone modulation test. The resulting signals are shown in views A, B, and C of figure 5-6. The two-tone test is used for initial adjustment and for precise checking because it will indicate distortion. The two-tone test corresponds to the wave envelope method of AM modulation checking. Two signals of equal amplitude but of slightly different frequencies beating together are applied to the sideband modulator input to produce a single tone of approximately 1,000 hertz. On an oscilloscope, the output appears as a series of fully modulated sine waves and is similar to a 100-percent-amplitude-modulated waveform, as shown in view A. A spectrum analyzer presentation is shown in view B.

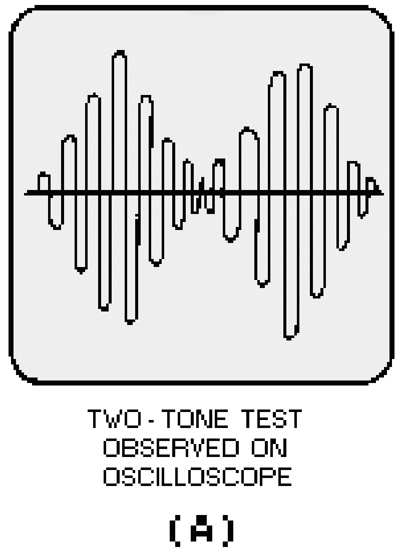


Figure 5-6A.—Examples of ideal two-tone test waveforms.

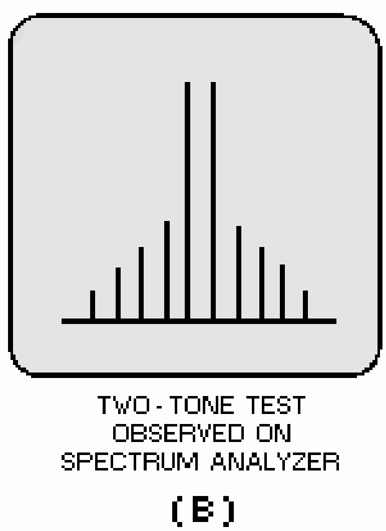


Figure 5-6B.—Examples of ideal two-tone test waveforms.

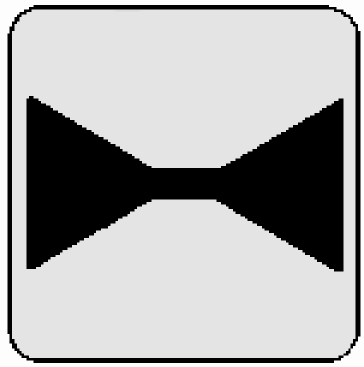


Figure 5-6C.—Examples of ideal two-tone test waveforms.

When the trapezoidal method is used, two opposed triangles appear on the oscilloscope, as shown in figure 5-6, view C. When equally balanced modulators are used, the triangles are mirror images. Elliptical or straight-line patterns appear when the phase-distortion check is used.

It is also possible to make a rough operating adjustment by varying the audio drive from the microphone so that on peak swings a definite value of final plate current is not exceeded. This check depends upon the initial accuracy of calibration and response characteristics of the ammeter in the final stage, as well as other factors.

FREQUENCY MODULATION

In frequency modulation, the carrier amplitude remains constant, and the output frequency of the transmitter is varied about the carrier (or mean) frequency at a rate corresponding to the audio frequencies. The extent to which the frequency changes in one direction from the unmodulated (carrier) frequency is called the FREQUENCY DEVIATION.

Deviation in frequency is usually expressed in kilohertz. It is equal to the difference between the carrier frequency and either the highest or lowest frequency reached by the carrier in its excursions with modulation. There is no modulation percentage in the usual sense. With suitable circuit design, the frequency deviation may be made as large as desired without encountering any adverse effects that are equivalent to the overmodulation in amplitude-modulation transmissions. However, the maximum permissible frequency deviation is determined by the width of the band assigned for station operation.

In frequency modulation, the equivalent of 100% modulation occurs when the frequency deviation is equal to a predetermined maximum value. There are several methods of measuring the modulation in frequency-modulated transmissions.

The frequency-deviation measurement of a frequency-modulated signal is normally performed with either a spectrum analyzer or with a modulation analyzer. The modulation analyzer method is more commonly used because of its accuracy. Typical accuracies for a modulation analyzer are within ±1%. Figure 5-7 shows a typical modulation analyzer.

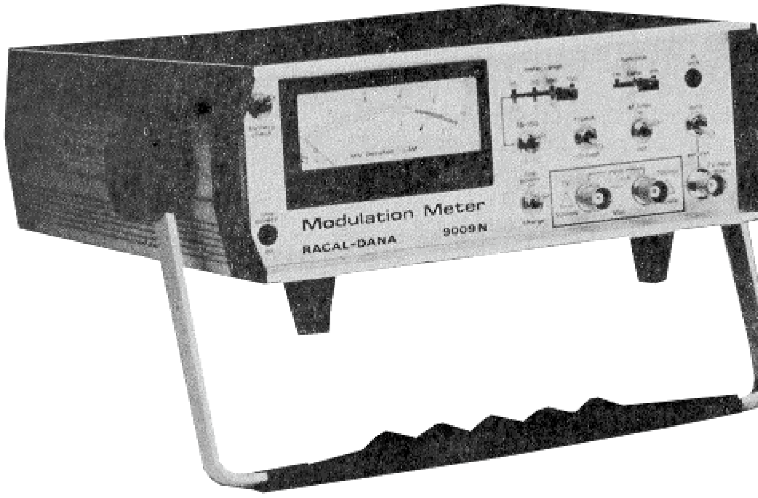


Figure 5-7.—Typical modulation analyzer.

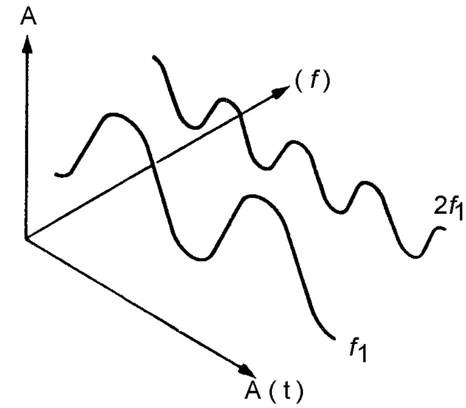
Q-3. What is meant by frequency deviation?

SPECTRUM WAVEFORM ANALYSIS AND MEASUREMENTS

An analysis of a complex waveform, prepared in terms of a graphic plot of the amplitude versus frequency, is known as SPECTRUM ANALYSIS. Spectrum analysis recognizes the fact that waveforms are composed of the summation of a group of sinusoidal waves, each of an exact frequency and all existing together simultaneously.

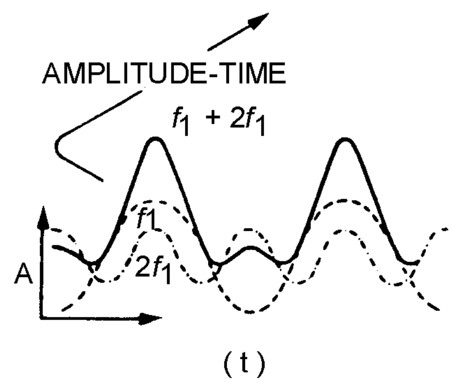
Three axes of degree (amplitude, time, and frequency) are important when considering varying frequency. The time-domain (amplitude versus time) plot is used to consider phase relationships and basic timing of the signal and is normally observed with an oscilloscope. The frequency-domain (amplitude versus frequency) plot is used to observe frequency response - the spectrum analyzer is used for this purpose. Figure 5-8 illustrates the differences between frequency- and time-domain plots. View A illustrates a three-dimensional coordinate of a fundamental frequency (f_1) and its second harmonic ($2f_1$) with respect to time, frequency, and amplitude. View B shows the time-domain display as it would be seen on an oscilloscope. The solid line, $f_1 + 2f_1$ is the actual display. The dashed lines, f and $2f_1$ are drawn to illustrate the fundamental and second harmonic frequency relationship used to formulate the composite signal $f_1 + 2f_1$. View C is the frequency-domain display as it would be seen on a spectrum analyzer. Note in view C that the components of the composite signal are clearly seen.

Q-4. A spectrum analyzer is designed to display what signal characteristic?



(A)

Figure 5-8A.—Time versus frequencies.



(B)

Figure 5-8B.—Time versus frequencies.

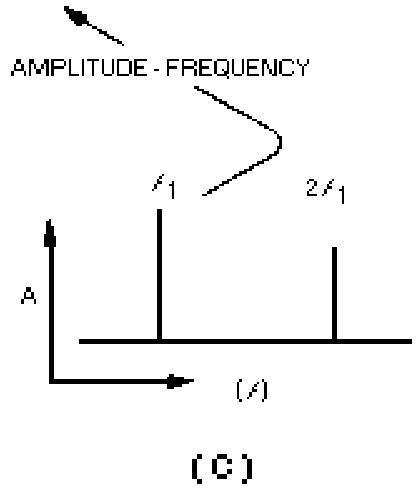


Figure 5-8C.—Time versus frequencies.

FREQUENCY-DOMAIN DISPLAY CAPABILITIES

The frequency domain contains information not found in the time domain. The spectrum analyzer can display signals composed of more than one frequency (complex signals). It can also discriminate between the components of the signal and measure the power level at each one. It is more sensitive to low-level distortion than an oscilloscope. Its sensitivity and wide, dynamic range are also useful for measuring low-level modulation, as illustrated in views A and B of figure 5-9. The spectrum analyzer is useful in the measurement of long- and short-term stability such as noise sidebands of an oscillator, residual fm of a signal generator, or frequency drift of a device during warm-up, as shown in views A, B, and C of figure 5-10.

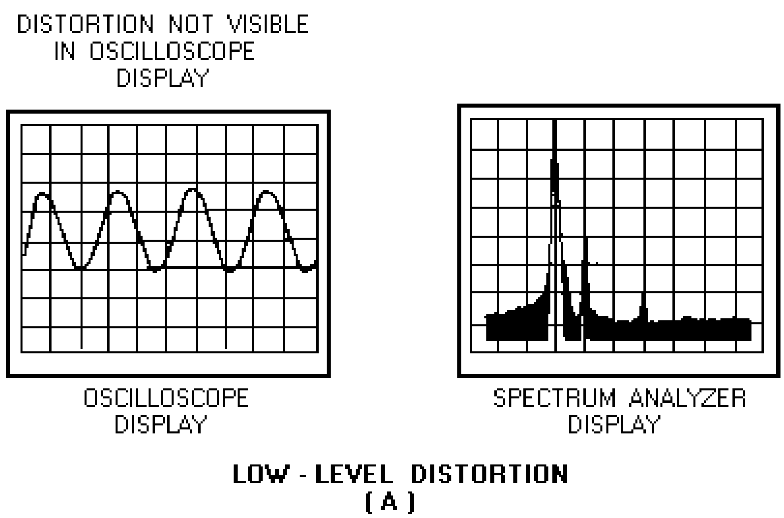


Figure 5-9A.—Examples of time-domain (left) and frequency-domain (right) low-level signals.

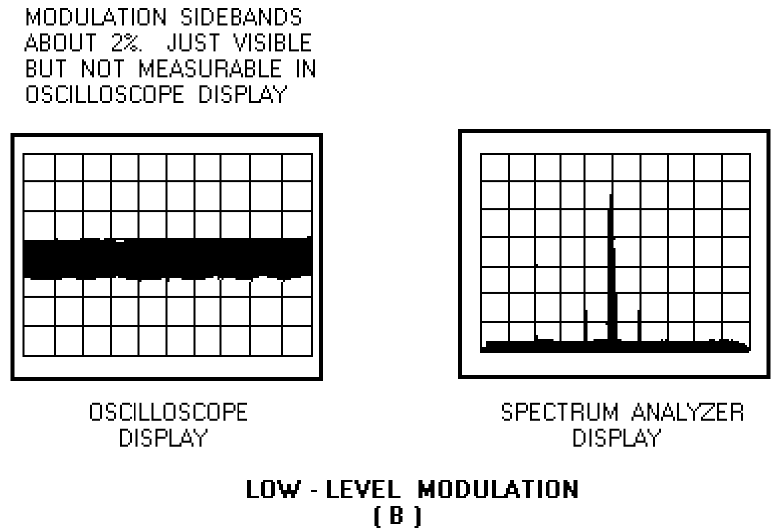


Figure 5-9B.—Examples of time-domain (left) and frequency-domain (right) low-level signals.

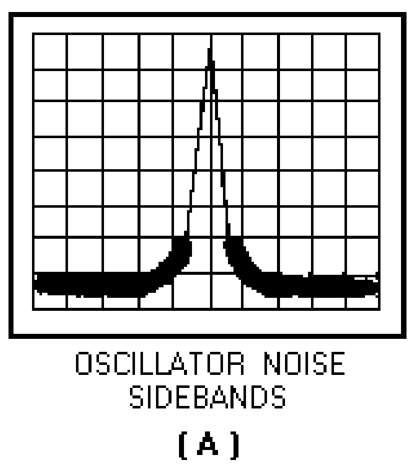


Figure 5-10A.—Spectrum analyzer stability measurements.

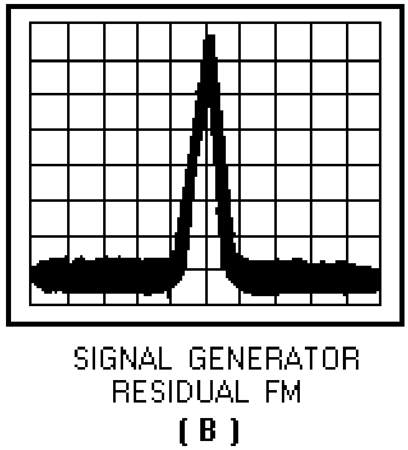


Figure 5-10B.—Spectrum analyzer stability measurements.

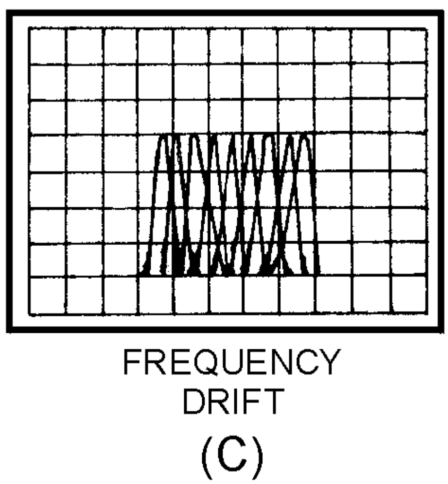


Figure 5-10C.—Spectrum analyzer stability measurements.

The swept-frequency response of a filter or amplifier and the swept-distortion measurement of a tuned oscillator are also measurable with the aid of a spectrum analyzer. However, in the course of these measurements, a variable persistence display or an X-Y recorder should be used to simplify readability. Examples of tuned-oscillator harmonics and filter response are illustrated in figure 5-11. Frequency-conversion devices such as mixers and harmonic generators are easily characterized by such parameters as conversion loss, isolation, and distortion. These parameters can be displayed, as shown in figure 5-12, with the aid of a spectrum analyzer.

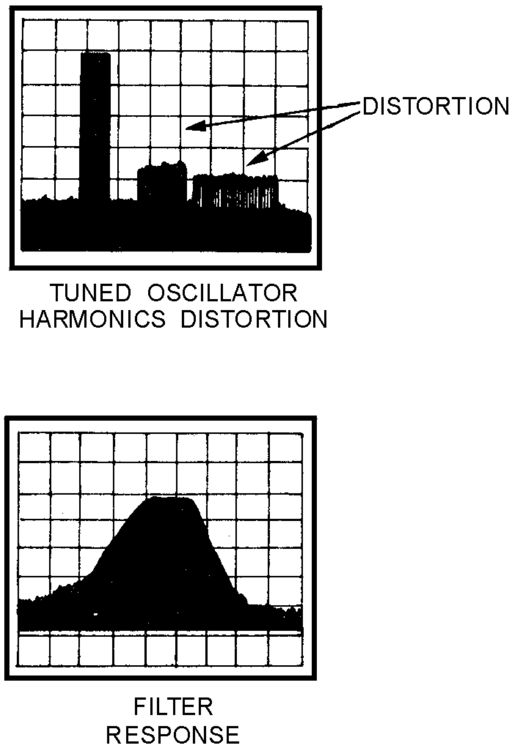


Figure 5-11.—Swept-distortion and response characteristics.

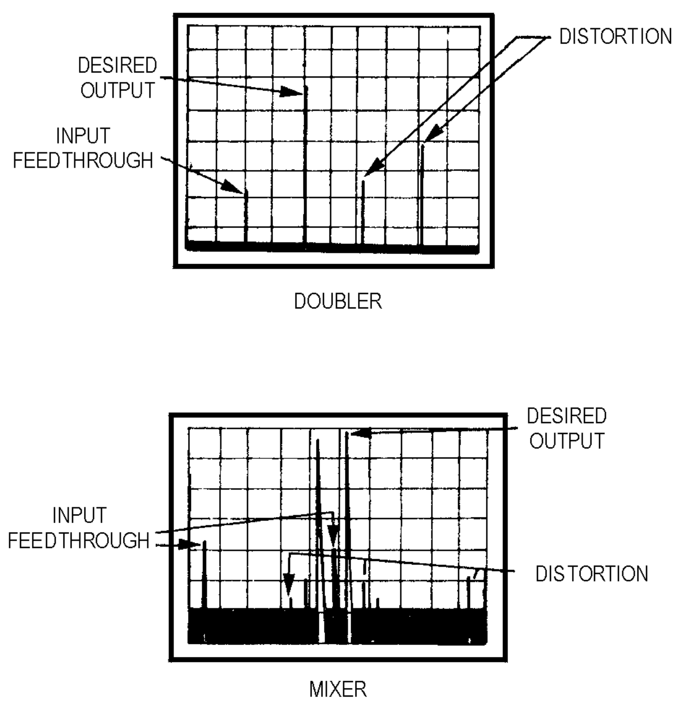


Figure 5-12.—Frequency-conversion characteristics.

Present-day spectrum analyzers can measure segments of the frequency spectra from 0 hertz to as high as 300 gigahertz when used with waveguide mixers.

SPECTRUM ANALYZER APPLICATIONS

Figure 5-13 shows a typical spectrum analyzer. The previously mentioned measurement capabilities can be seen with a spectrum analyzer. However, you will find that the spectrum analyzer generally is used to measure spectral purity of multiplex signals, percentage of modulation of AM signals, and modulation characteristics of fm and pulse-modulated signals. The spectrum analyzer is also used to interpret the displayed spectra of pulsed rf emitted from a radar transmitter.

33NP0180

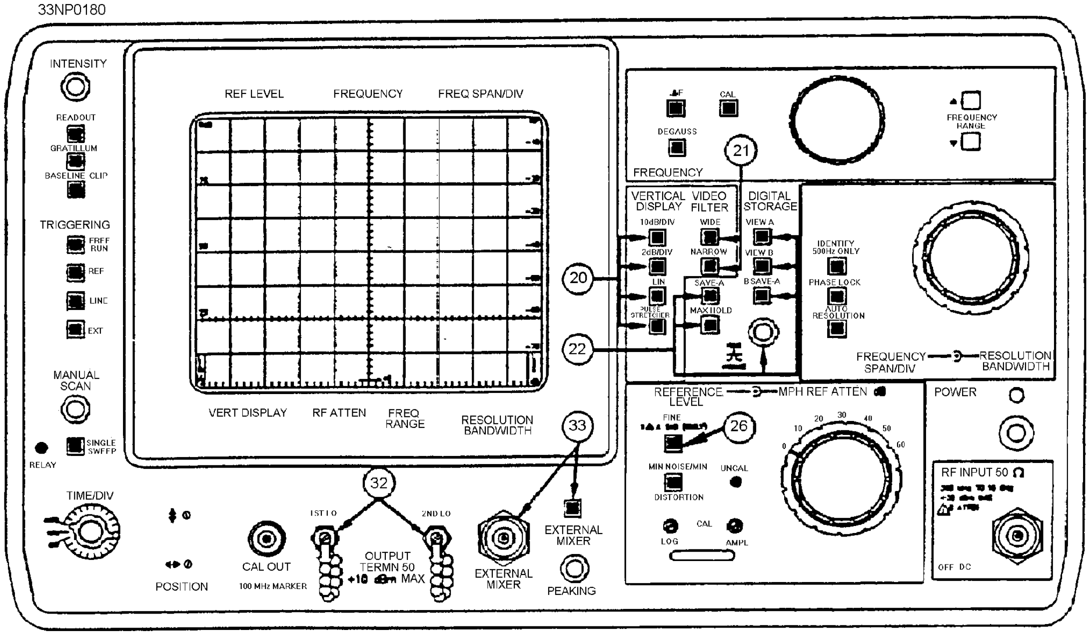


Figure 5-13.—Typical spectrum analyzer.

COMPLEX WAVEFORMS

Complex waveforms are divided into two groups, PERIODIC WAVES and NONPERIODIC WAVES. Periodic waves contain the fundamental frequency and its related harmonics. Nonperiodic waves contain a continuous band of frequencies resulting from the repetition period of the fundamental frequency approaching infinity and thereby creating a continuous frequency spectrum.

MODULATION MEASUREMENTS

In all types of modulation, the carrier is varied in proportion to the instantaneous variations of the modulating waveform. The two basic properties of the carrier available for modulation are the AMPLITUDE CHARACTERISTIC and ANGULAR (frequency or phase) CHARACTERISTIC.

Amplitude Modulation

The modulation energy in an amplitude-modulated wave is contained entirely within the sidebands. Amplitude modulation of a sinusoidal carrier by another sine wave would be displayed as shown in figure 5-14. For 100% modulation, the total sideband power would be one-half of the carrier power; therefore, each sideband would be 6 dB less than the carrier, or one-fourth of the power of the carrier. Since the carrier component is not changed with AM transmission, the total power in the 100-percent-modulated wave is 50% higher than in the unmodulated carrier. The primary advantage of the log display that is provided by the spectrum analyzer over the linear display provided by the oscilloscopes for percentage of modulation measurements is that the high dynamic range of the spectrum analyzer (up to 70 dB) allows accurate measurements of values as low as 0.06%. It also allows the measurements of low-level distortion of AM signals. Both capabilities are illustrated in figure 5-15, view A, view B, and view C. The chart in figure 5-16 provides an easy conversion of dB down from carrier into percentage of modulation.

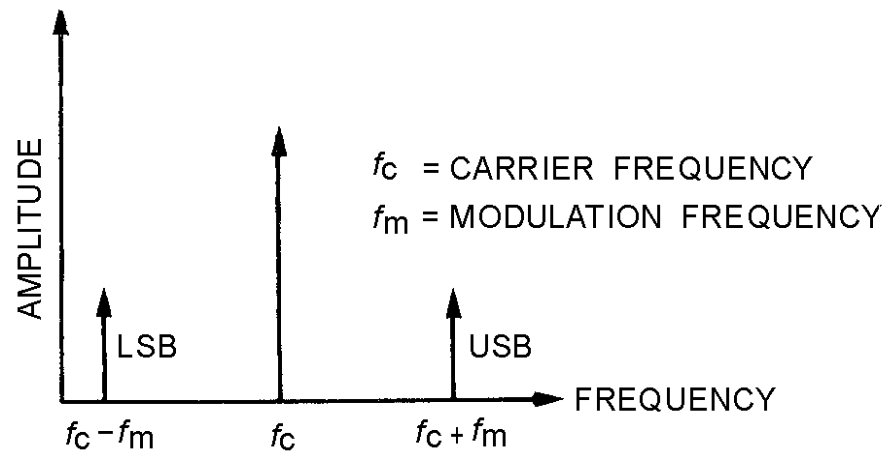
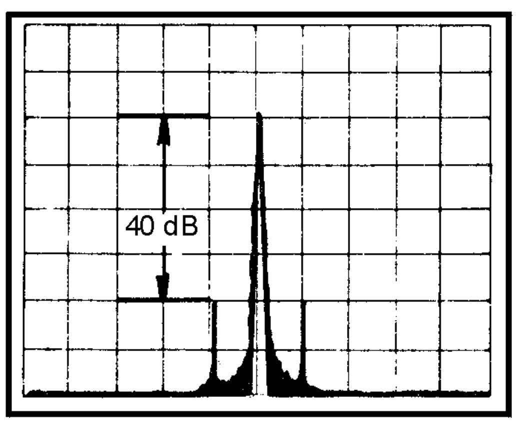


Figure 5-14.—Spectrum analyzer display of an AM signal.



2% MODULATION
SIDEBANDS 40 dB DOWN

(A)

Figure 5-15A.—Spectrum analyzer displays of AM signals.

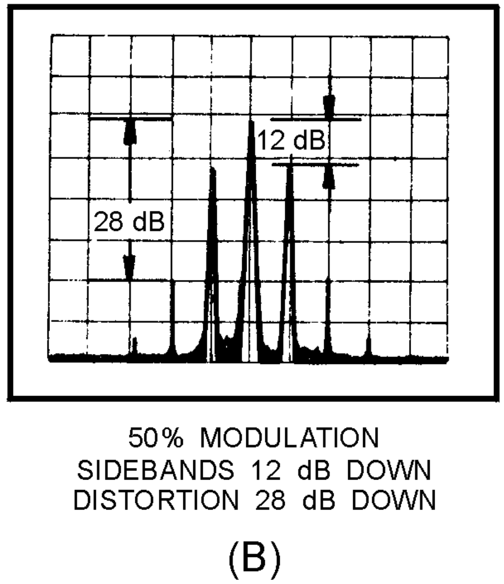


Figure 5-15B.—Spectrum analyzer displays of AM signals.

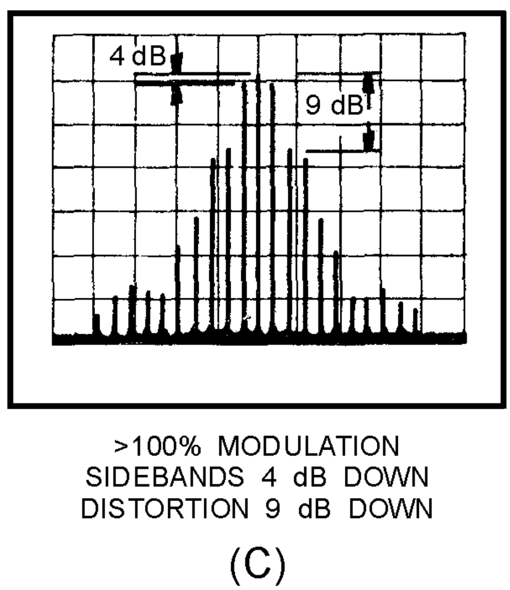


Figure 5-15C.—Spectrum analyzer displays of AM signals.

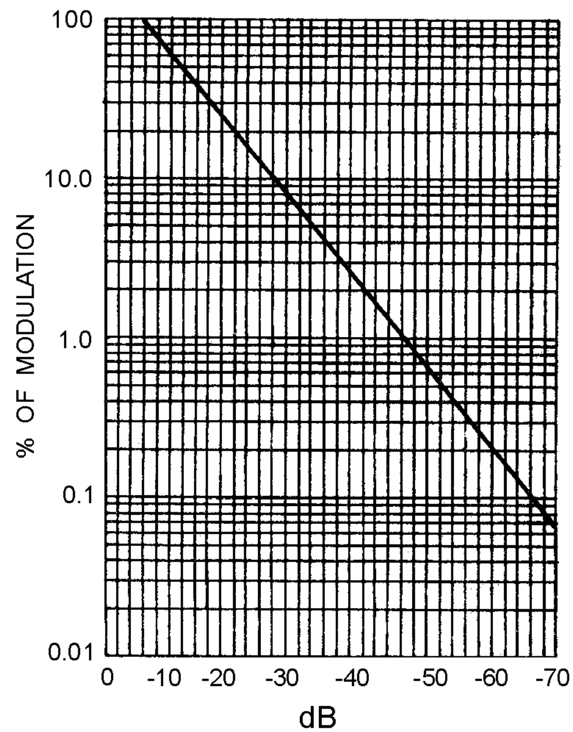
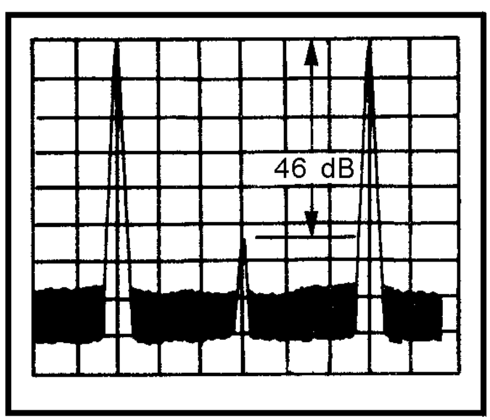


Figure 5-16.—Modulation percentage versus sideband levels.

NOTE: Anything greater than -6 dB exceeds 100% modulation and produces distortion, as shown in figure 5-16.

In modern, long-range hf communications, the most important form of amplitude modulation is ssb (single-sideband). In ssb either the upper or lower sideband is transmitted, and the carrier is suppressed. Ssb requires only one-sixth of the output power required by AM to transmit an equal amount of intelligence power and less than half the bandwidth. Figure 5-17 shows the effects of balancing the carrier of an AM signal. The most common distortion experienced in ssb is intermodulation distortion, which is caused by nonlinear mixing of intelligence signals. The two-tone test is used to determine if any intermodulation distortion exists. Figure 5-18 illustrates the spectrum analyzer display of the two-tone test with the modulation applied to the upper sideband input.



CARRIER BALANCE 46 dB

Figure 5-17.—Double sideband carrier suppressed.

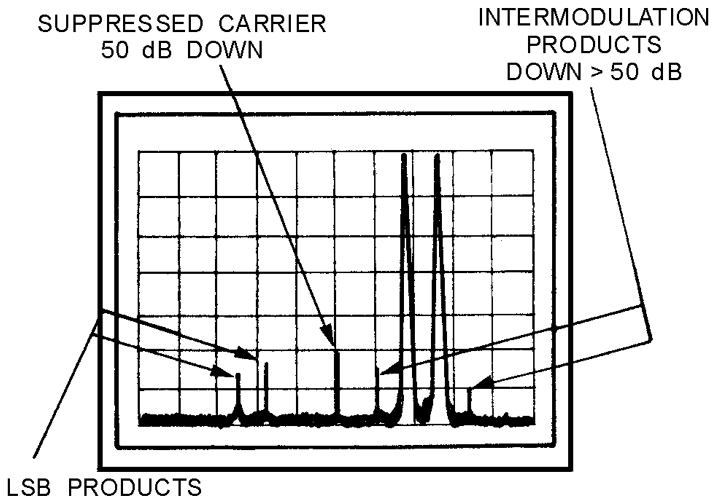


Figure 5-18.—Two-tone test.

Q-5. What is the advantage of single-sideband (ssb) transmission over AM transmission?

Frequency Modulation

In frequency modulation, the instantaneous frequency of the radio-frequency wave varies with the modulation signal. As mentioned in NEETS, module 12, the amplitude is kept constant. The number of times per second that the instantaneous frequency varies from the average (carrier frequency) is controlled by the frequency of the modulating signal. The amount by which the frequency departs from the average is controlled by the amplitude of the modulating signal. This variation is referred to as the FREQUENCY DEVIATION of the frequency-modulated wave. We can now establish two clear-cut rules for frequency deviation rate and amplitude in frequency modulation:

- Amount of frequency shift is proportional to the amplitude of the modulating signal.

(This rule simply means that if a 10-volt signal causes a frequency shift of 20 kilohertz, then a 20-volt signal will cause a frequency shift of 40 kilohertz.)

- Rate of frequency shift is proportional to the frequency of the modulating signal.

(This second rule means that if the carrier is modulated with a 1-kilohertz tone, then the carrier is changing frequency 1,000 times each second.)

The amplitude and frequency of the signal used to modulate the carrier will determine both the number of significant sidebands (shown in fig. 5-19) and the amplitude of the sidebands (shown in fig. 5-20). Both the number of significant sidebands and the bandwidth increase as the frequency of the modulating signal increases.

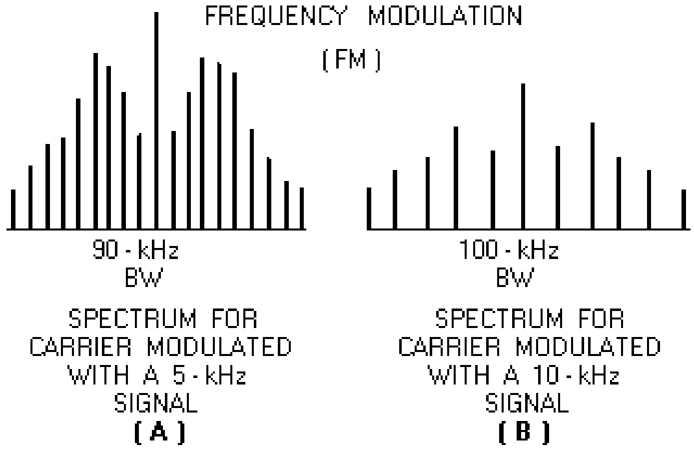


Figure 5-19.—Distribution of sidebands.

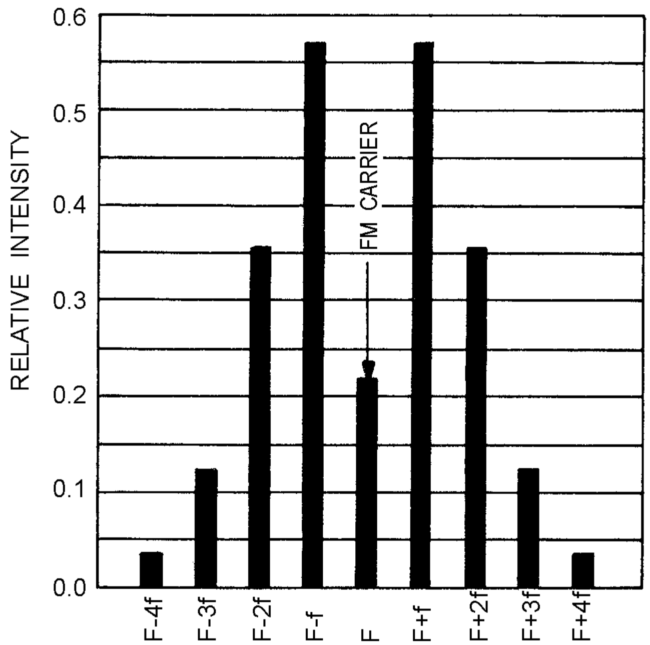


Figure 5-20.—Spectrum distribution for a modulation index of 2.

NEETS, module 12, should be consulted for an in-depth discussion of frequency-modulation principles.

Q-6. What happens to an fm signal as you increase the frequency of the modulating signal?

PULSED WAVES

An ideal pulsed radar signal is made up of a train of rf pulses with a constant repetition rate, constant pulse width and shape, and constant amplitude. To receive the energy reflected from a target, the radar receiver requires almost ideal pulse radar emission characteristics. By observing the spectra of a pulsed radar signal, you can easily and accurately measure such characteristics as pulse width, duty cycle, and peak and average power. The principles of radar are covered in NEETS, Module 18, *Radar Principles*, which can be consulted for an explanation of pulsed waves.

Rectangular Pulse

A rectangular wave is used to pulse-modulate the constant frequency rf carrier to produce the pulse radar output. The rectangular wave is made up of a fundamental frequency and its combined odd and even harmonics. Figure 5-21 shows the development of a rectangular wave.

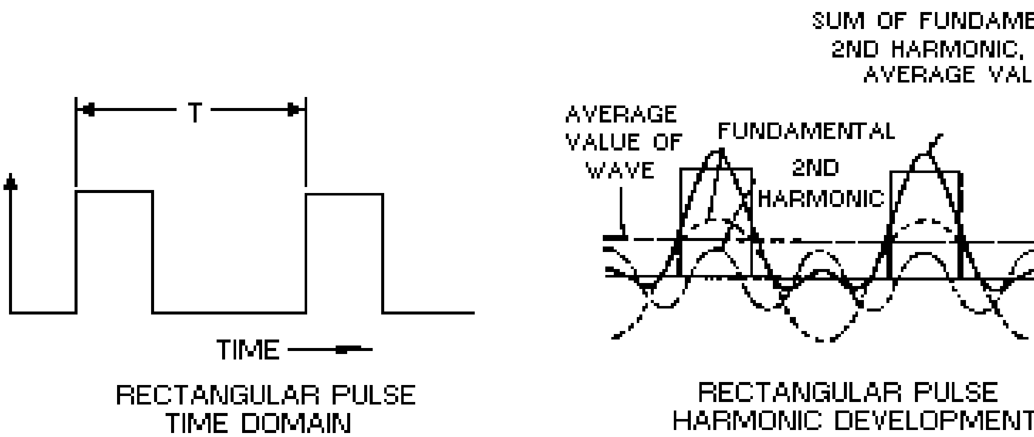


Figure 5-21.—Rectangular pulse.

Pulsed Wave Analysis

In amplitude modulation, sidebands are produced above and below the carrier frequency. A pulse is also produced above and below the carrier frequency, but the pulse is made up of many tones. These tones produce multiple sidebands that are commonly referred to as SPECTRAL LINES, or RAILS, on the spectrum analyzer display. Twice as many rails will be in the pulse-modulated output of the radar as there are harmonics contained in the modulating pulse (upper and lower sidebands), as shown in figure 5-22. In the figure, the pulse repetition frequency (prf) is equal to the pulse interval of 1/T. The actual spectrum analyzer display would show the lower lobes (shown below the reference line in the figure) on top because the spectrum analyzer does not retain any polarity information. Changing the pulse interval, or pulse width, of the modulation signal will change the amount of rails (prf), or number of lobe minima, as illustrated in figure 5-23.

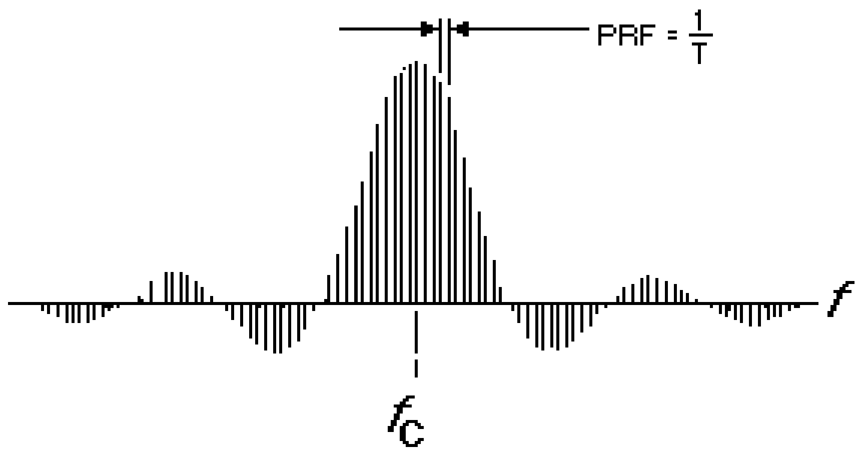


Figure 5-22.—Pulsed radar output.

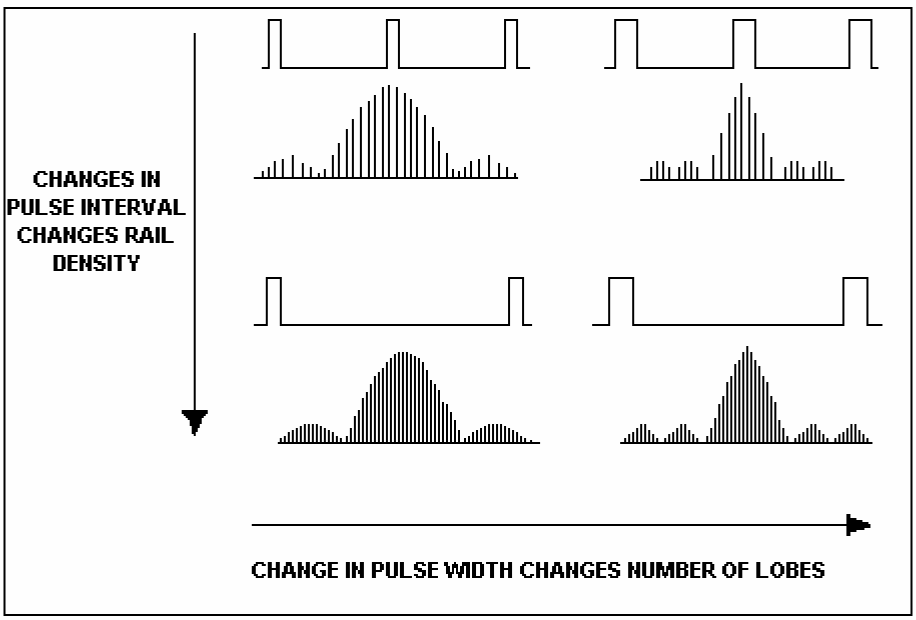


Figure 5-23.—Pulsed radar changes caused by modulating signal changes.

ANALYZING THE SPECTRUM PATTERN

The leading and trailing edges of the radiated pulse-modulated signal must have a sharp rise time and decay time and a constant amplitude between them. Incorrect pulse shape will cause frequency spread and pulling, which results in less available energy at the frequency to which the receiver is tuned. The primary reason for analyzing the spectrum is to determine the exact amount of amplitude and frequency modulation present. The amount of amplitude modulation determines the increase in the number of sidebands within the applied pulse spectrum; an increase in frequency modulation increases the amplitude of the side-lobe frequencies. In either case, the energy available to the main spectrum lobe is decreased.

SPECTRUM ANALYZER OPERATION

The information desired from the spectra to be analyzed determines the SPECTRUM ANALYZER requirements. Real-time analysis is used if a particular point in the frequency spectrum is to be analyzed, such as a line spectra display. Continuous- or swept-frequency analysis, which is the most common mode of observation, is used to display a wider portion of the frequency spectrum or (in some cases) the entire range of the spectrum analyzer in use. Changing the spectrum analyzer setting from one mode to another is accomplished by varying the scan time and the bandwidth of the spectrum analyzer or a combination of the two. Most real-time spectrum analyzers, however, are preceded by mechanical filters, which limit the input bandwidth of the spectrum analyzer to the desired spectra to be analyzed. Tunable- or swept-spectrum analyzers function basically the same as heterodyne receivers, the difference being that the local oscillator is not used but is replaced by a voltage-controlled oscillator (vco). The vco is swept electronically by a ramp input from a sawtooth generator. The output of the receiver is applied to a crt, which has its horizontal sweep in synchronization with the vco. The lower frequency appears at the left of the crt display. As the trace sweeps to the right, the oscillator increases in frequency. Figure 5-24 is a block diagram of a heterodyne spectrum analyzer.

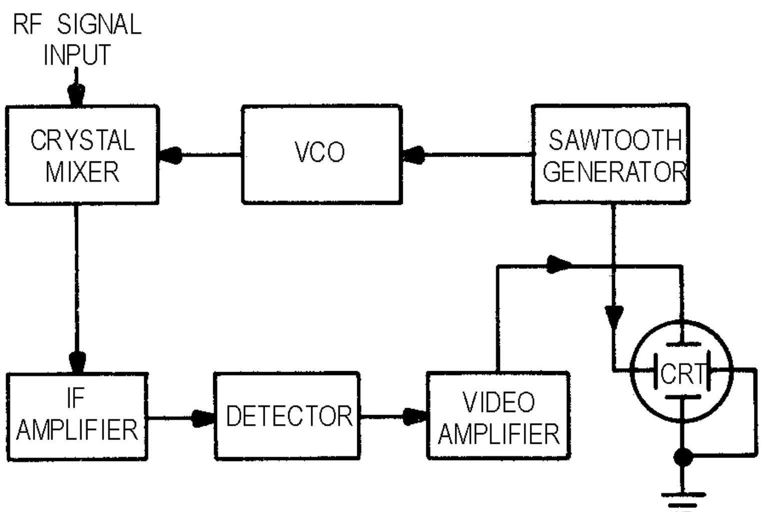


Figure 5-24.—Block diagram of a heterodyne spectrum analyzer.

Before the frequency of a signal can be measured on a spectrum analyzer, it must be RESOLVED. Resolving a signal means distinguishing it from other signals near it. Resolution is limited by the narrowest bandwidth of the spectrum analyzer because the analyzer traces out its own IF bandwidth shape as it sweeps through a signal. If the narrowest bandwidth is 1 kilohertz, the nearest any two signals can be, and still be resolved, is 1 kilohertz. Reducing the IF bandwidth indefinitely would obtain infinite resolution except that the usable IF bandwidth is limited by the stability of the spectrum analyzer. The smaller the IF bandwidth, the greater the capability of the analyzer to resolve closely spaced signals of unequal amplitudes. Modern spectrum analyzers have been refined to the degree that IF bandwidths are less than 1 hertz.

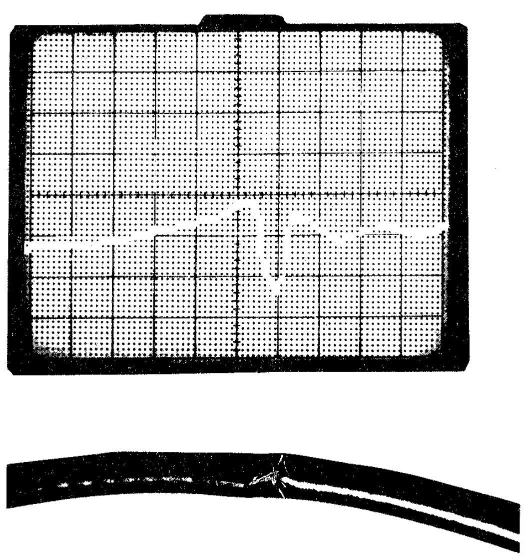
It is important that the spectrum analyzer be more stable in frequency than the signals being measured. The stability of the analyzer depends on the frequency stability of its vco. Scan time of the spectrum analyzer must be long enough, with respect to the amplitude of the signal to be measured, to allow the IF circuitry of the spectrum analyzer to charge and recover. This will prevent amplitude and frequency distortion.

Q-7. When referring to spectrum analyzers, what is meant by the term resolving signals?

TIME-DOMAIN REFLECTOMETRY

TIME-DOMAIN REFLECTOMETRY is a testing and measurement technique that has found increasing usefulness in testing transmission lines (both metallic and fiber-optic), cables, strip lines, connectors, and other wideband systems or components. Basically, time-domain reflectometry is an extension of an earlier technique in which reflections from an electrical pulse were monitored to locate faults and to determine the characteristics of power transmission lines. You can compare time-domain reflectometry to a closed-loop radar system in which the transmitted signal, a very fast step pulse, is fed into the system and the reflections resulting from discontinuities or impedance deviations in the system are monitored on a crt.

The technique used in time-domain reflectometry consists of feeding an impulse of energy into the system and then observing that energy as it is reflected by the system at the point of insertion. When the fast-rise input pulse meets with a discontinuity or impedance mismatch, the resultant reflections appearing at the feed point are compared in phase, time, and amplitude with the original pulse. By analyzing the magnitude, deviation, and shape of the reflected waveform, you can determine the nature of the impedance variation in the transmission system. Also, since distance is related to time and the amplitude of the reflected step is directly related to impedance, the comparison indicates the distance to the fault as well as the nature of the fault. Figure 5-25, view A, view B, view C, and view D, illustrates typical transmission line problems that can easily be identified by using a time-domain reflectometer (tdr). In addition to showing both the distance to and the nature (resistive, inductive, or capacitive) of each line discontinuity, time-domain reflectometry also reveals the characteristic impedance of the line and indicates whether losses are shunt or series. They are also used to locate and analyze connectors and splices.



CRIMPED CABLE
(A)

Figure 5-25A.—Time-domain reflectometer display of transmission line problems.

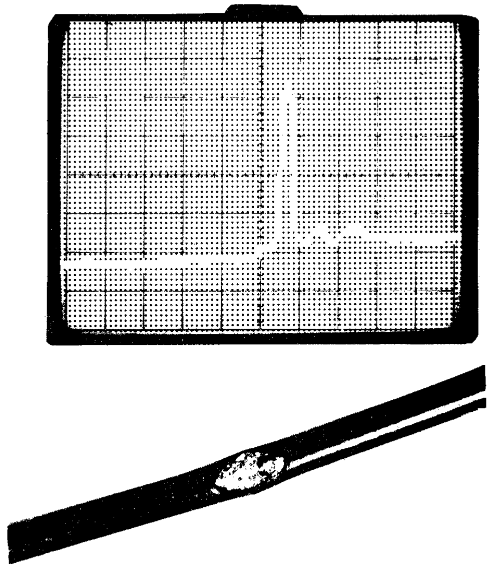

FRAYED CABLE
(B)

Figure 5-25B.—Time-domain reflectometer display of transmission line problems.

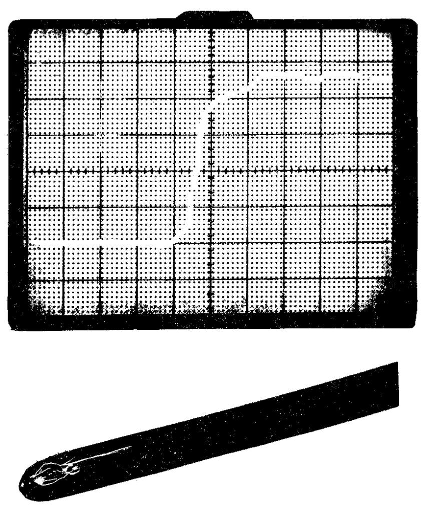
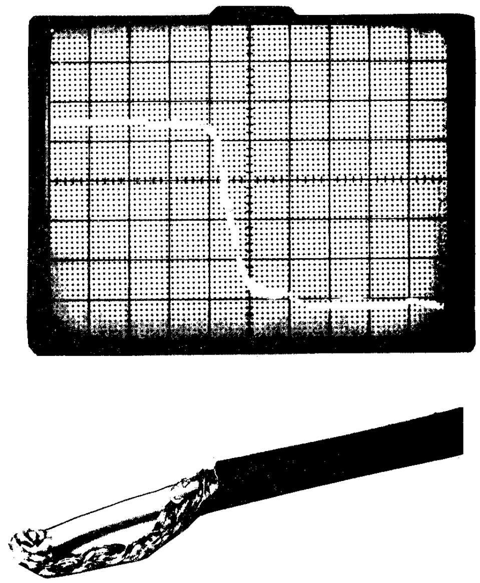

OPEN CABLE
(C)

Figure 5-25C.—Time-domain reflectometer display of transmission line problems.



SHORTED CABLE
(D)

Figure 5-25D.—Time-domain reflectometer display of transmission line problems.

A conventional method of evaluating high-frequency transmission systems and components has been through the use of standing wave ratio (swr) measurements to obtain an overall indication of transmission line performance. This method involves feeding a sine-wave signal into the system and measuring the maximum and minimum amplitudes of the standing waves that result from system discontinuities or load mismatches. The ratio between the minimum and maximum swr values is then taken as the system FIGURE OF MERIT. The swr measurement, however, does not isolate individual discontinuities or mismatches when multiple reflections are present; it only indicates their total effect. Time-domain reflectometry measurements, on the other hand, isolate the line characteristics in time (location). As a result, multiple reflections resulting from more than one discontinuity or impedance variation that are separated in distance on the line are also separated in time at the monitoring point and can be individually analyzed.

Prior to the advent of time-domain reflectometers, time-domain reflectometry was performed with the aid of sampling oscilloscopes and pulse generators with very fast rise times.

Figure 5-26 shows the earlier type of test setup, which is still an option. However, today's time-domain reflectometers have several advantages over the old pulse-generator and oscilloscope methods. Modern time-domain reflectometers are compact, lightweight, are often supplied with battery pack options for field use, and provide a direct readout of distances instead of time. Some equipments provide a paper-tape recording for a permanent record. Figure 5-27 shows a typical time-domain reflectometer.

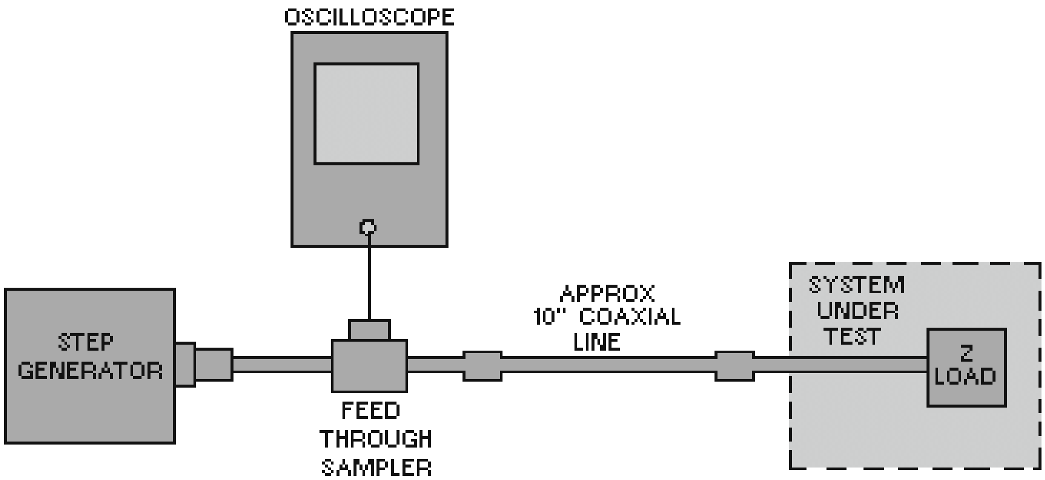


Figure 5-26.—Time-domain reflectometry, basic equipment setup.

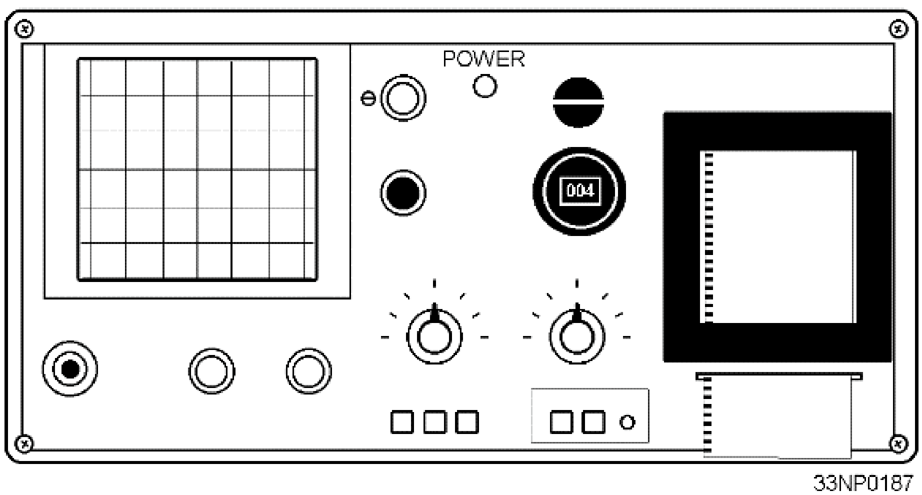


Figure 5-27.—Typical time-domain reflectometer.

Q-8. Why are time-domain reflectometers often compared to a radar system?

Q-9. What is the main advantage of using a time-domain reflectometer (tdr) to test a transmission line?

SWEPT-FREQUENCY TESTING EQUIPMENT

SWEPT-FREQUENCY testing is used to determine the bandwidth, alignment, frequency response, impedance matching, and attenuation in various circuits, systems, and components. Swept-frequency testing can be used to quickly determine the broadband response of a device that otherwise would require a number of separate measurements and manual plotting of the response curve. Swept-frequency techniques are applicable over the entire electronic spectrum from vlf to ehf and are generally limited only by your resourcefulness and the basic limitation of the equipment employed. The basic swept-frequency arrangement is shown in figure 5-28.

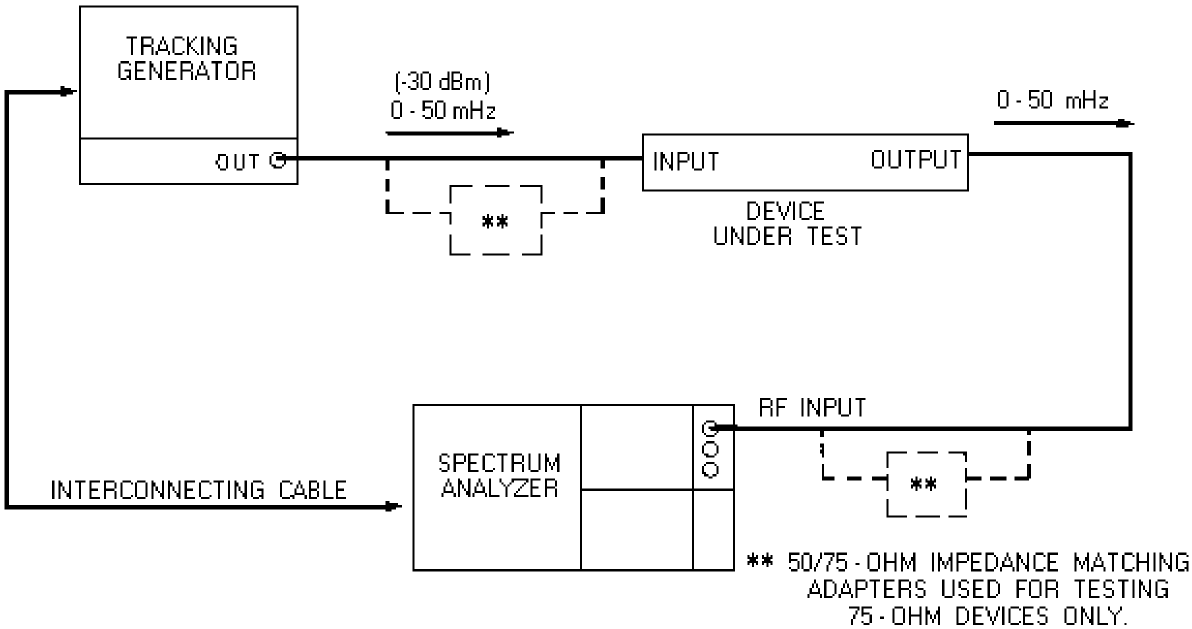


Figure 5-28.—Frequency-response test.

The swept-frequency technique can effectively determine the frequency response of an amplifier or filter and is useful in the alignment or bandwidth determination of an IF or rf stage. The test equipment permits direct visual readout on the crt of the spectrum analyzer. The spectrum analyzer can also be connected to an X-Y chart recorder if a permanent record or print is desired. Figure 5-29 shows a spectrum analyzer crt display of the frequency response of a multicoupler. The tracking generator used must be capable of sweeping the desired frequency range of the device under test.

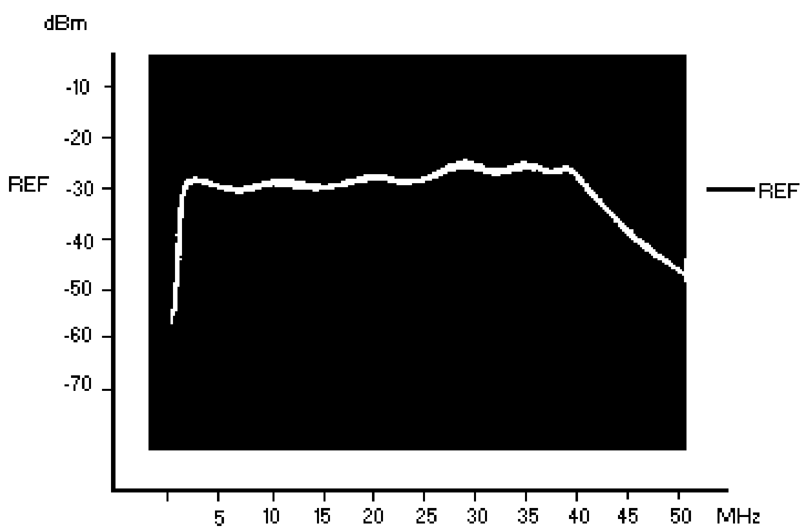


Figure 5-29.—Typical spectrum analyzer frequency-response display.

Q-10. What is the purpose of swept-frequency testing?

TRACKING GENERATOR

Figure 5-30 shows a typical tracking generator used with the Hewlett-Packard 141 T spectrum analyzer. A TRACKING GENERATOR is basically a sweep generator in which the sweep rate is matched to that of the spectrum analyzer. The output circuitry of the tracking generator contains a network that ensures a constant output amplitude over the entire range being swept. When the fm signal produced by the tracking generator is applied to a device or circuit under test, the instantaneous output amplitude is always proportional to the response of the circuit to the frequency at that instant. Thus, the original fm input signal is changed in passing through the circuit under test. The output signal, therefore, would consist of an fm signal that is also amplitude-modulated. For equal deviations, the positive and negative portions of this envelope are symmetrical, making it necessary to observe only one side of the envelope. After the detection stage in the spectrum analyzer, only the modulation remains to appear on the face of the crt. This presentation will appear as a continuous curve because of the persistence of vision and the phosphor characteristic of the crt. The polarity of the detector determines whether a positive or a negative output is displayed. The frequency at any point on the crt display can be analyzed by arresting the scan of the spectrum analyzer either electronically or manually at the point of interest. For greater accuracy in frequency determination, a frequency counter may be attached to the output of the tracking generator at the point of the arrested scan.

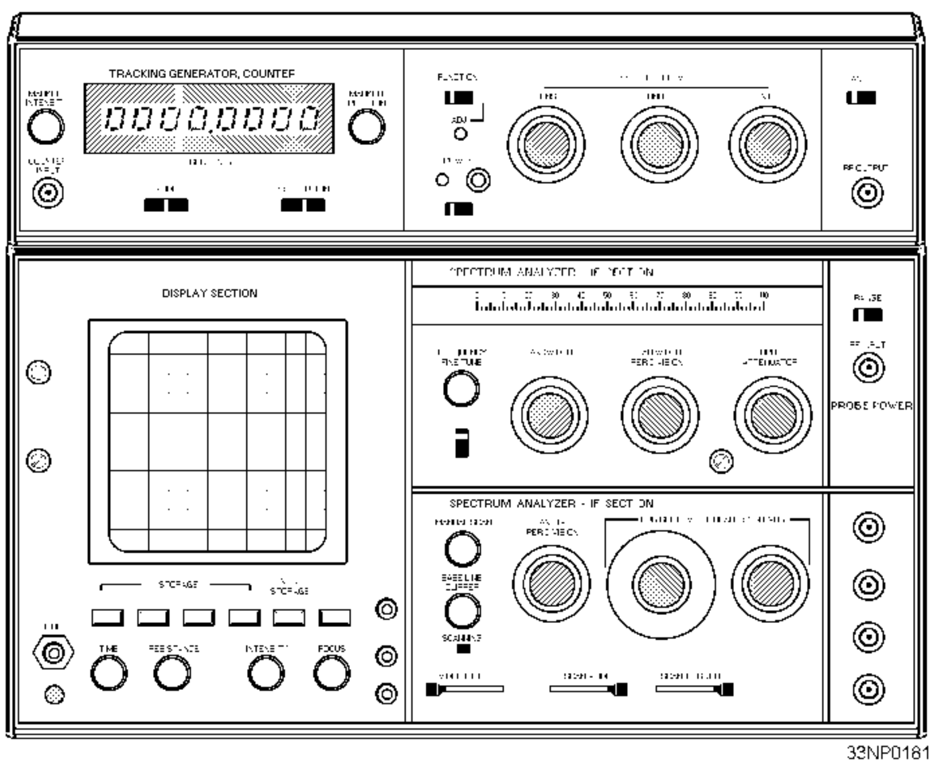


Figure 5-30.—Tracking generator used with a spectrum analyzer.

IMPEDANCE MATCHING

Conventional tuners cannot be used successfully to cancel source or load reflections in swept-frequency measurements. This is because the tuning is effective only at single frequencies; therefore, pads or isolators are required. However, by the use of automatic-level control, the power output of the sweep generator can be maintained relatively constant at the point of measurement. The source impedance may thus be maintained very close to the nominal value. With this arrangement, any impedance variation in the connecting cables, connectors, and adapters is effectively cancelled since these components are within the leveling loop. The attenuation of a device under test will be displayed on the associated crt as a continuous response curve as it is scanned. This will result in an attenuation versus frequency plot of the device under test only.

IMPEDANCE

Circuit impedance is measured conveniently by using the reflectometer principle. The individual values of the incident and reflected signals (swr) in a transmission line feeding an unknown impedance are measured. The ratio between these signals indicates how closely the load impedance matches that of the transmission line.

Another method is the use of an auto-mechanical load control to hold the forward power at a constant level while the return load of a specific load is measured. A short is then placed in the circuit, and 100% reflected power is measured. The loss detected is then calculated to obtain swr figures.

NOISE FIGURE

By using a frequency-sweeping receiver and an automatic noise-figure meter, you can make noise-figure measurements on broadband microwave devices, such as a traveling-wave-tube amplifier. To conduct such a test properly, you must first check the receiver noise figure.

Q-11. In swept-frequency testing the impedance of a transmission line, what electrical characteristic is actually being measured?

SWEEPING ANTENNAS

Antenna system testing is one of the more common and useful applications for using the swept-frequency technique. The main parameters that an antenna system is tested for are vswr, frequency response, and impedance. Figure 5-31 shows a typical test setup for testing a transmitting antenna for vswr. Remember that any transmitting antenna can also act as a receiving antenna and send induced power from adjacent antennas back to the test equipment. You should make an initial power check on the antenna to prevent damage to your test equipment. Figure 5-32 shows a typical hf transmitting antenna vswr display as measured using the swept-frequency technique. The setup for testing a receiving antenna vswr, shown in figure 5-33, is similar, with the exception of the attenuators. The measured vswr (within the operating frequency range) of any broadband antenna should not exceed a vswr of 2.5 to 1. The vswr for any single-tuned antenna should not exceed 1.5 to 1 at the tuned frequency.

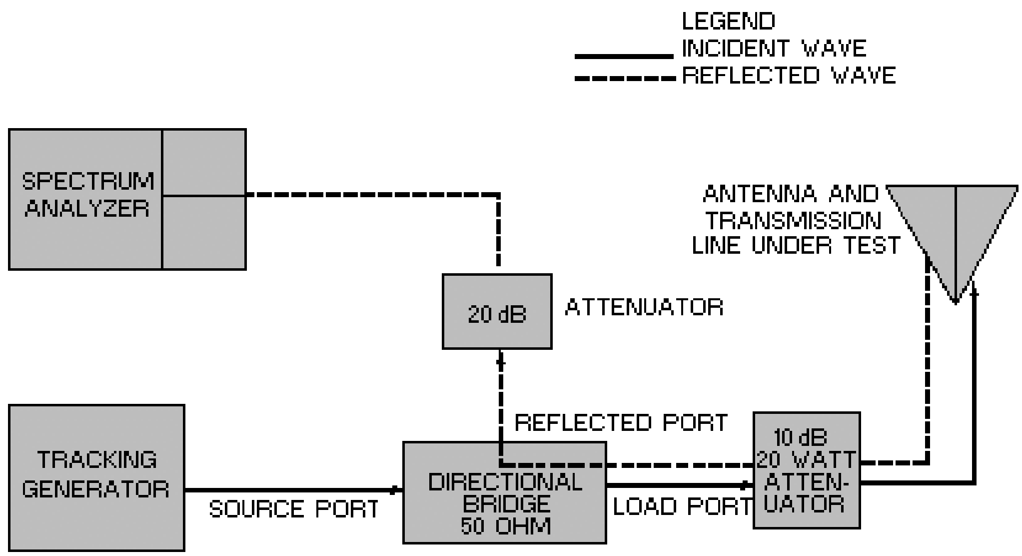


Figure 5-31.—Vswr test for transmitting antennas.

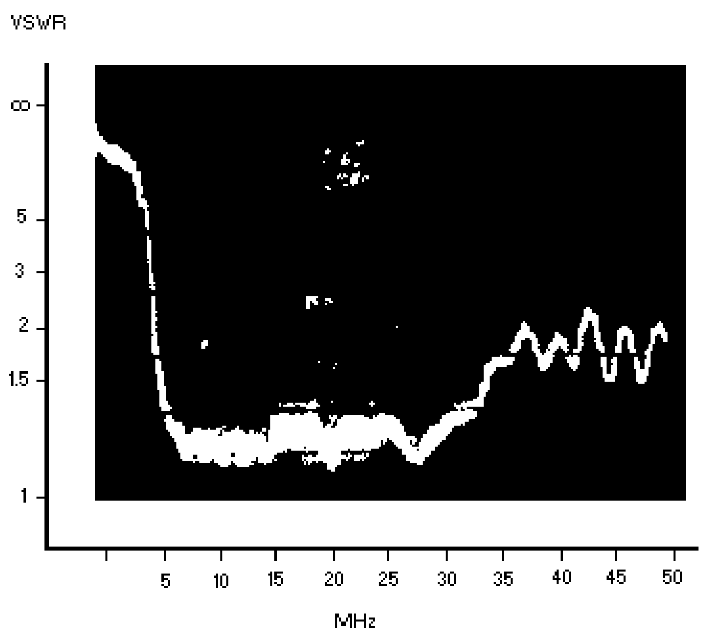


Figure 5-32.—Typical spectrum analyzer vswr display.

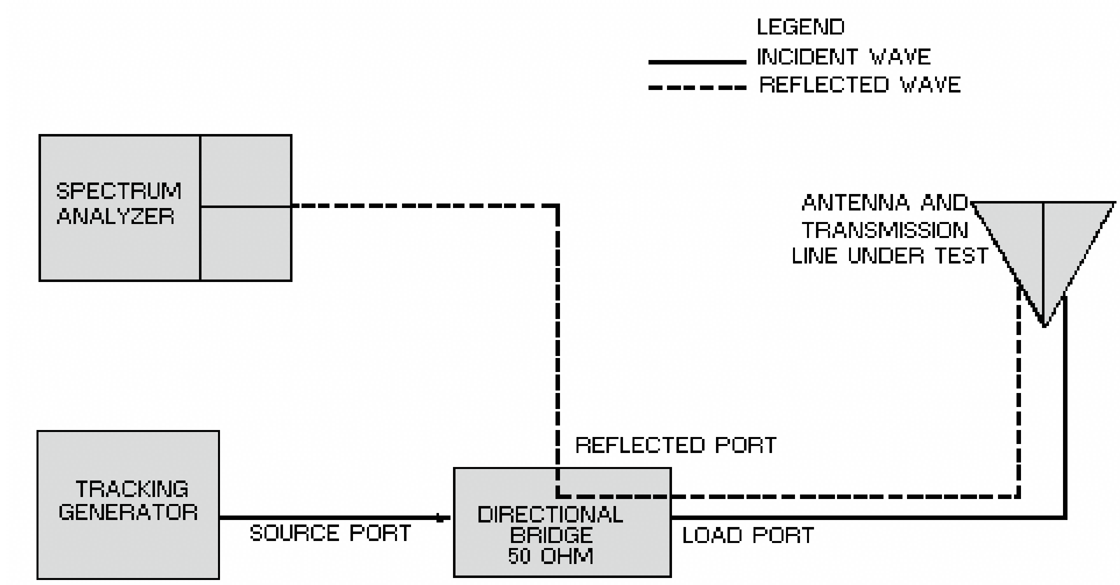


Figure 5-33.—Vswr test for receiving antennas.

Q-12. What precautions must be taken when sweeping a transmitting antenna?

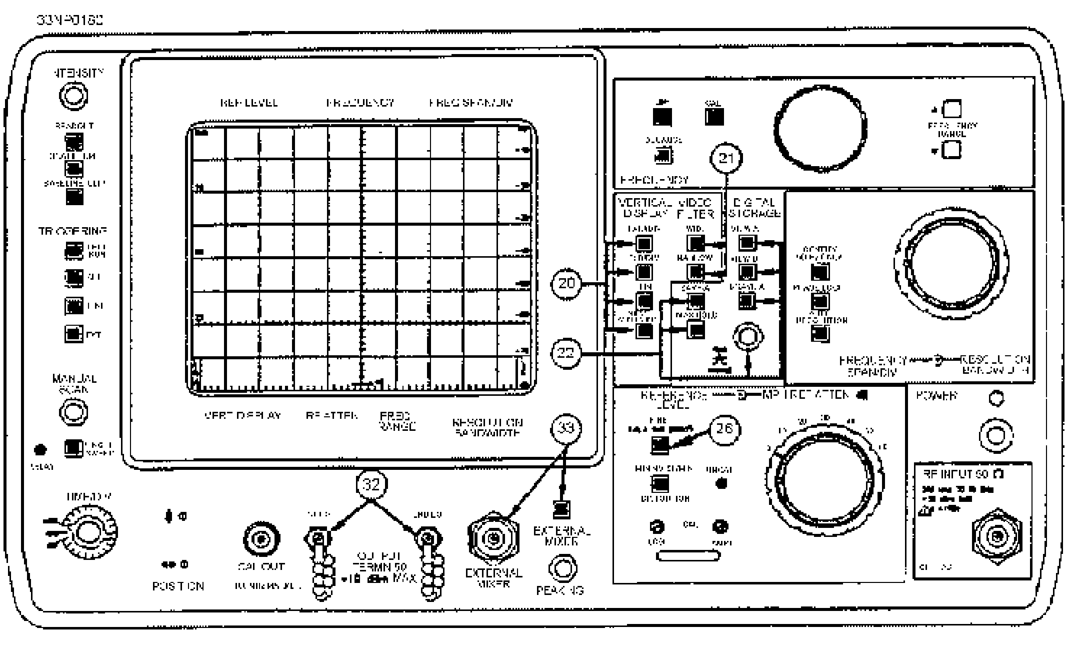
SUMMARY

This chapter has presented information on waveform interpretation. The information that follows summarizes the important points of this chapter.

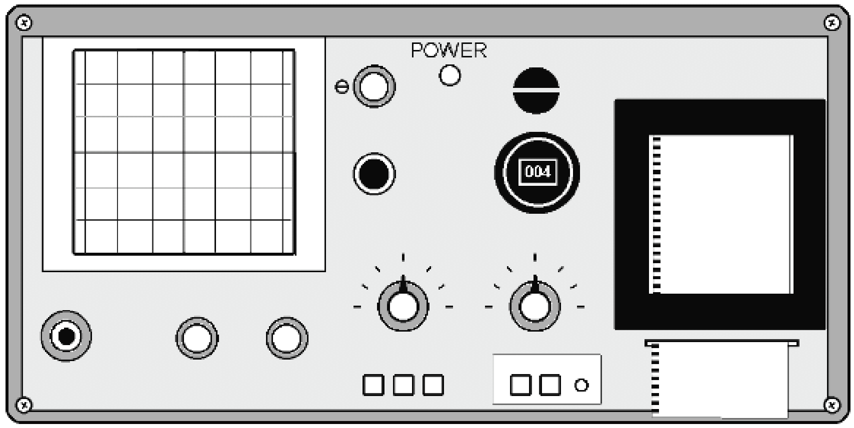
Interpretation of a waveform is best accomplished with test equipment that gives you a visual indication of the waveform. The most common devices used in systems applications are **OSCILLOSCOPES** and **SPECTRUM ANALYZERS**.

An amplitude-modulated signal can be tested with either an oscilloscope or a spectrum analyzer to determine its percentage of modulation, sideband characteristics, and carrier frequency. Frequency-modulated signals are normally tested with a spectrum analyzer or a modulation analyzer.

Oscilloscopes are designed to view a time-domain waveform (amplitude versus time). Spectrum analyzers are designed to view a frequency-domain waveform (amplitude versus frequency). One advantage of using a spectrum analyzer is its ability to graphically display the composition of **COMPLEX WAVEFORMS**.

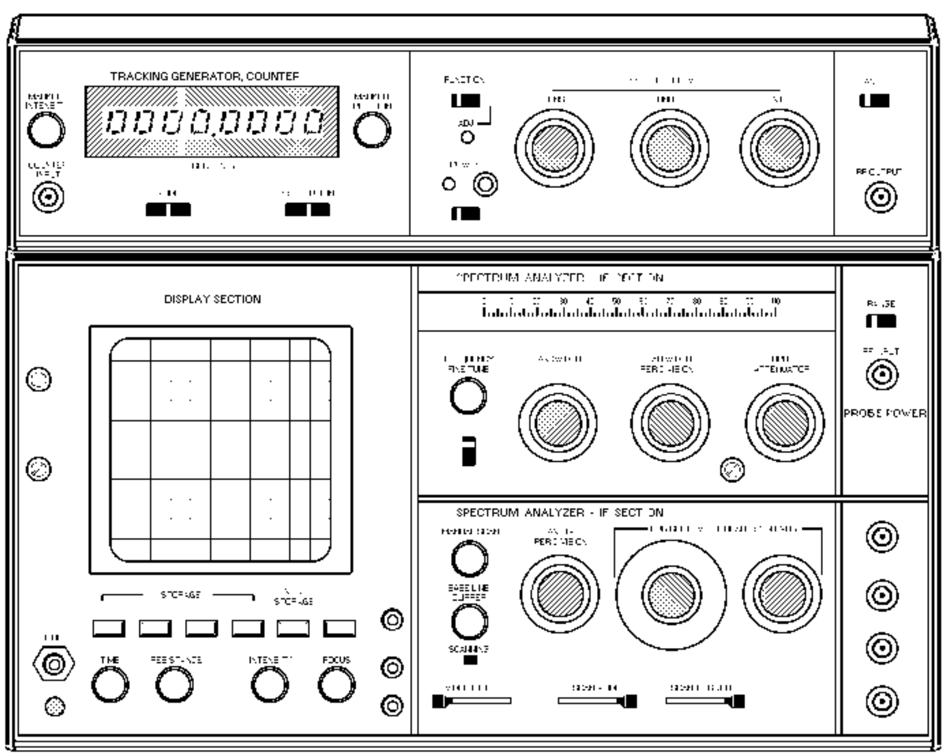


TIME-DOMAIN REFLECTOMETERS work on a principle similar to that used in radar. A precise signal is produced by the time-domain reflectometer and injected into the device under test (usually a transmission line); and the resulting reflections are displayed to discover such things as impedance mismatches, opens, and shorts. The display sections of time-domain reflectometers are calibrated to give you a graphical display of amplitude versus distance.



33NP0187

SWEPT-FREQUENCY TESTING is performed by using a **TRACKING GENERATOR** to inject a signal into a device and then monitoring the output of the device with a spectrum analyzer. The tracking generator is designed to sweep or scan through the entire frequency range of the device being tested. Its sweep rate must be matched with the sweep rate of the spectrum analyzer.



33NP0181

REFERENCES

Communications Systems, *NAVTELCOMINST 2313.1, Naval Telecommunications Command, Washington, D.C., 1984.*

EIMB, Test Methods and Practices, *NAVSEA 0967-LP-000-0130, Naval Sea Systems Command, Washington, D.C., 1980.*

Modulation Principles, *NAVEDTRA 172-12-00-83. Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1983.*

ANSWERS TO QUESTIONS Q1. THROUGH Q12.

A-1. Distortion.

A-2. 60% to 95%.

A-3. The difference between the carrier frequency of an fm signal and its maximum frequency excursion when modulated.

A-4. Amplitude versus frequency (the frequency domain of the signals).

A-5. The same amount of intelligence can be transmitted with one-sixth of the output power with less than one-half the bandwidth.

A-6. Both the bandwidth and the number of significant sidebands increase.

A-7. The ability of the analyzer to discriminate between display signals of slightly different frequencies.

A-8. Both transmit a pulse and analyze the signal reflection.

A-9. A Tdr will indicate the nature of and the distance to or location of any faults.

A-10. To determine various characteristics of a component, piece of equipment, or system over its operational frequency range.

A-11. Swr on the transmission line.

A-12. You must ensure that power induced from any adjacent transmitting antennas does not damage your test equipment.

APPENDIX I

GLOSSARY

ABSORPTION WAVEMETER—A device used for measuring frequency, consisting of a tuned circuit or cavity that is loosely coupled to the frequency being measured. Maximum energy is absorbed at the resonant frequency.

BOLOMETER—An rf detector that converts rf power to heat, which causes a change in the resistance of the material used in the detector. This change in resistance varies in proportion with the amount of applied power and is used to measure the amount of applied power.

CALORIMETER—A device that measures rf power by measuring the heat the rf power generates.

CAVITY WAVEMETER—An instrument used to measure microwave frequencies. The resonant frequency of the cavity is determined by its inside dimensions.

COAXIAL-LINE WAVEMETERS—A shorted section of a coaxial line used to measure rf frequencies. It is calibrated in either wavelength or frequency.

CROSS MODULATION—An intermodulation condition that occurs when a carrier is modulated by an undesired signal.

CURRENT PROBE—An inductive device used for measuring the current in a conductor. Probes are designed to be clamped around the insulated conductor.

CURRENT TRACER—An inductively coupled device used for tracing current paths to determine the cause of low-impedance faults on a printed-circuit board.

DECADE RESISTOR (DECADE RESISTANCE BOX)—It typically has two or more sections, each containing 10 precision resistors wired to selector switches. A piece of test equipment that provides a ready source of various resistances for engineering and measurement applications.

DECIBEL (dB)—A standard unit for expressing *relative* power levels as the ratio of power out to power in.

dBm—A unit used to express power levels above or below a l-milliwatt reference level at a designated load impedance (usually 600 ohms).

DIFFERENTIAL VOLTMETER—A precision voltmeter that measures an unknown voltage by comparing it to a precision internal-reference voltage supply.

ELECTROSTATIC-DISCHARGE SENSITIVE (ESDS) DEVICE—Electronic components that are susceptible to damage from static charges.

FIBER OPTICS—Conductors that are usually constructed of plastic or glass fibers that readily pass light. Used primarily for transmission of high-speed data over relatively short distances.

FREQUENCY DEVIATION—Refers to the difference between the carrier frequency of an fm signal and the instantaneous frequency of its modulated wave.

FREQUENCY DOMAIN—A plot of frequency versus amplitude as shown by a spectrum analyzer display.

FREQUENCY RESPONSE—(I) The ability of a component or device to operate over a portion of the frequency spectrum. (2) In reference to test equipment, that portion of the frequency spectrum that the test equipment is capable of sensing and measuring accurately.

GALVANOMETER—A meter used to measure small values of current by electromagnetic or electrodynamic means.

IMPEDANCE ANGLE METER—A device that measures circuit impedance by comparing the phase angle between voltage and current.

INSERTION LOSS—The difference between the amount of power applied to a load before and after the insertion of a device in the line.

INTEGRATED CIRCUIT (IC)—(1) A circuit in which many elements are fabricated and interconnected by a single process (into a single chip), as opposed to a nonintegrated circuit in which the transistors, diodes, resistors, and other components are fabricated separately and then assembled. (2) Elements inseparably associated and formed on or within a single substrate.

INTERMODULATION DISTORTION—Nonlinear distortion characterized by the appearance (at the system output) of frequencies equal to the sums and differences of two or more frequencies present at the input.

LOAD—(I) A device through which an electric current flows that changes electrical energy into another form. (2) Power consumed by a device or circuit in performing its function.

LOGIC CLIPS—A device that can be clipped onto an in-circuit, dual-in-line package (DIP) logic IC to determine the logic state of each pin of the IC.

LOGIC PROBE—A hand-held probe used to determine the logic state (high or low) of test points in a logic circuit. A logic high is represented by a lit indicator light on the probe.

LOGIC PULSER—A hand-held probe used to pulse, or change the logic state, of in-circuit logic ICs.

METROLOGY CALIBRATION (METCAL) PROGRAM—A Navy calibration program designed to ensure the accuracy of test equipment through comparisons with calibration laboratory standards of known accuracy.

MICROPHONICS—Electrical noise caused by the mechanical motion of the internal parts of a device. The term is usually associated with vacuum tubes.

MODULATION INDEX—When a sine wave is used to modulate an fm signal, the ratio of the frequency deviation to the frequency of the modulating wave.

NATIONAL BUREAU OF STANDARDS (NBS)—A bureau of the United States government that is responsible for maintaining the nation's electrical and physical standards. The accuracy of all *calibrated* test equipment is traceable to NBS through the Navy's METCAL program.

OPTICAL TIME-DOMAIN REFLECTOMETER (OTDR)—A piece of test equipment used to test a fiber-optic cable for such things as attenuation, localized losses, and defects. It transmits an optical pulse (usually a laser) into the fiber-optic cable and analyzes the resulting reflections in terms of amplitude versus distance.

PARALLAX ERROR—The error in meter readings that results when you look at a meter from some position other than directly in line with the pointer and meter face. A mirror mounted on the meter face aids in eliminating parallax error.

PARASITIC—In electronics, an undesirable frequency in an electronic circuit. Usually associated with vacuum-tube amplifiers and oscillators.

PERCENT OF MODULATION—In AM signals, the ratio of half the difference between the maximum and minimum amplitudes of a modulated wave to its average amplitude.

RF IMPEDANCE BRIDGE—A piece of test equipment used for measuring the combined resistance and reactance of a component, piece of equipment, or system at rf frequencies.

SENSITIVITY—In reference to test equipment, the ratio of the response of the test equipment to the magnitude of the measured quantity. Sometimes expressed indirectly by stating the property by which sensitivity is computed (e.g., ohms per volt).

STANDARD—An exact value of an electrical quantity (established by international agreement), which serves as a model for measurement of that quantity.

STANDING WAVE—The distribution of voltage and current along a transmission line formed by the incident and reflected waves, which has minimum and maximum points on a resultant wave that appears to stand still.

STROBOSCOPE—An instrument that allows viewing of rotating or reciprocating objects by producing the optical effect of a slowing or stopping motion.

SWEPT-FREQUENCY TESTING—Testing the frequency response of a component or system by applying an rf signal, in which the frequency is varied back and forth through a set frequency range at a steady rate, to the input of a device. The output is then monitored to determine the amplitude of the output with respect to frequency.

THERMISTOR—(1) A semiconductor device in which the resistance varies with temperature. (2) A type of bolometer characterized by a decrease in resistance as the dissipated power increases.

TIME-DOMAIN REFLECTOMETER—A piece of test equipment used to test a transmission line for defects, such as shorts and opens. It transmits an electrical pulse into the transmission line and analyzes the resulting reflections in terms of emplidute versus distance.

TRAIC—A three-terminal device that is similar to two SCRs back-to-back with a common gate and common terminals. Although similar in construction and operation to the SCR, the Triac controls and conducts current flow during both alternations of an ac cycle.

TUNING FORK—A two-pronged mechanical device that is designed to vibrate only at its natural frequency. In electronics, it is used primarily to determine the correct speed of a motor.

UNIJUNCTION TRANSISTOR (UJT)—A three-terminal, semiconductor device with a negative-resistance characteristic that is used in switching circuits, oscillators, and wave-shaping circuits.

VOLUME UNIT (VU)—Unit of measurement of a complex audio signal such as voice or music. A 0 level is referenced to 1 milliwatt of power into a 600-ohm load.

WAVEMETERS—(1) Calibrated resonant circuits that are used to measure frequency. (2) An instrument for measuring the wavelength of an rf wave.

ZENER DIODE—A pn-junction diode designed to operate in the reverse-bias breakdown region.

MODULE 21 INDEX

Component testing—Continued

D

R

S

T

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Basic Measurements," pages 1-1 through 1-26. Chapter 2, "Component Testing," pages 2-1 through 2-8.

1-1. What is the purpose of the Navy's Metrology Calibration Program?

1. To provide the fleet with new types of test equipment
2. To provide quality control for your test equipment
3. To improve the efficiency of sophisticated electronic systems
4. To establish test equipment pools from which technicians can borrow

1-2. At each higher echelon METCAL calibration laboratory, the accuracy of the test equipment increases by a factor of

1. 10
2. 2
3. 100
4. 4

1-3. Most equipment technical manuals contain voltage charts. For which of the following purposes are they used?

1. To list the equipment's power supplies
2. To list the input power requirements of the equipment
3. To provide handy reference guides for calculating voltage drops across fixed impedances
4. To list correct voltages at major test points

THIS SPACE LEFT BLANK INTENTIONALLY.

1-4. Which, if any, of the following statements correctly describes the effect input impedance of test equipment can have on readings taken?

1. The greater the input impedance of your test equipment, the less accurate the readings
2. The lower the input impedance of your test equipment, the more accurate the readings
3. A piece of test equipment with an infinite input impedance will absorb no energy and readings will be more accurate
4. None of the above

1-5. A piece of test equipment with a low input impedance can cause readings taken to be inaccurate. To eliminate this problem, the input impedance of your test equipment should exceed the impedance of the circuit under test by what minimum ratio?

1. 1 to 1
2. 2 to 1
3. 10 to 1
4. 100 to 1

THIS SPACE LEFT BLANK INTENTIONALLY.

USE THE FOLLOWING INFORMATION TO ANSWER QUESTION 1-6: YOU NEED TO TAKE A CRITICAL VOLTAGE READING, BUT YOU DO NOT HAVE A HIGH IMPEDANCE METER AVAILABLE. INSTEAD, YOU CONNECT TWO LOWER IMPEDANCE METERS IN SERIES AND PLACE THEM ACROSS THE COMPONENT IN QUESTION. YOU ADD THE READINGS SHOWN ON THE TWO METERS TO GET YOUR MEASUREMENT.

1-6. Compared to using just one of the lower impedance meters, what is the advantage of using two meters connected in series?

1. Input impedance increases and voltage-measuring accuracy increases
2. Frequency response of the test setup doubles
3. Accuracy of current measurements decreases
4. Input impedance decreases and voltage-measuring accuracy increases

1-7. On an analog multimeter, where on the scale are the most accurate readings taken?

1. At the highest end of the scale
2. At the lowest end of the scale
3. Midscale
4. It makes no difference if the meter is properly calibrated

1-8. What can you do to reduce the problem of meter-reading errors caused by parallax?

1. Close one eye when reading the meter
2. Use short meter leads
3. Use a meter that has a mirror built into the scale
4. View the meter face from either the left or right side, but not directly in front

1-9. For what primary reason are oscilloscopes used in circuit testing?

1. They provide a visual presentation of the signal under test
2. They present a low input impedance to the circuit under test
3. They provide numerical readouts of signals under test
4. They measure voltages more accurately than other pieces of test equipment

1-10. THIS QUESTION HAS BEEN DELETED.

1-11. Digital multimeters effectively eliminate which of the following disadvantages of analog meters?

1. Parallax
2. Low impedance
3. Poor accuracy
4. All of the above

1-12. Which of the following pieces of test equipment is most accurate for measuring dc voltages?

1. Vtvm
2. Oscilloscope
3. Digital voltmeter
4. Differential voltmeter

1-13. If you exceed the frequency limitations of your voltmeter, which of the following results is likely?

1. The meter will be destroyed
2. The circuit under test will be damaged
3. The measurement will be inaccurate
4. The meter will indicate average voltage

1-14. When performing measurements with an oscilloscope, you should ensure that the trace extends across what minimum portion of the vertical viewing area?

1. 15%
2. 25%
3. 45%
4. 60%

1-15. When using an oscilloscope to measure a high voltage, you should use which of the following procedures?

1. Use the logic probe instead of the normal probe
2. Use the high voltage probe instead of the normal probe
3. Use two oscilloscopes connected in series
4. Place a 10-ohm shunt across the vertical input of the oscilloscope

1-16. Oscilloscopes are normally calibrated to display which of the following types of voltages?

1. Peak
2. Average
3. Peak-to-peak
4. Both 2 and 3 above

1-17. When using an oscilloscope to observe a sine wave, what, if anything, must you do to determine the rms voltage?

1. Divide the observed peak-to-peak voltage by 3.65
2. Multiply the observed peak-to-peak voltage by 2; then divide by 1.414
3. Divide the observed peak-to-peak voltage by 2; then multiply by 0.707
4. Nothing

1-18. The frequency-measuring capabilities of a digital multimeter can be extended by using which of the following devices?

1. An rf probe
2. A frequency doubler
3. A high-voltage probe
4. A frequency divider network

1-19. When performing ac voltage measurements, you should use which of the following pieces of equipment to obtain the most accurate reading?

1. A differential voltmeter
2. An oscilloscope
3. A Simpson 260
4. A wattmeter

1-20. For which of the following purposes would you connect two ammeters in parallel?

1. To perform voltage measurements
2. To increase frequency-measuring capabilities
3. To decrease input impedance
4. To increase input impedance

1-21. When taking measurements with two ammeters connected in parallel, how do you determine the resulting readings?

1. The current equals the sum of both meter readings
2. The current equals the difference of the two meter readings
3. The current equals the product of the two readings divided by their sum
4. Read either meter directly; the same current flows through both meters

1-22. Current tracers indicate the presence of a current in which of the following ways?

1. By the lighting of an indicator lamp
2. By a clicking noise
3. Both 1 and 2 above
4. By the movement of a meter

1-23. Which of the following is an advantage of using a current probe?

1. It is the most accurate method of measuring current
2. It senses current by induction without being connected directly into the circuit
3. It is battery operated
4. It is capable of measuring current at frequencies above 40 GHz

1-24. When troubleshooting a specific piece of equipment, you can find an accurate listing of resistance readings for specific test points in which of the following documents?

1. In equipment PMS cards
2. In test equipment manuals
3. In equipment technical manuals
4. In Naval Ships Technical Manuals

1-25. An ohmmeter that is used for field work should meet which of the following criteria?

1. It should be extremely accurate
2. It should be portable
3. It should be simple to operate
4. Both 2 and 3 above

1-26. When you use an analog multimeter to measure resistance, which of the following actions should you take first?

1. Make sure the meter is zeroed
2. Set the meter for dc voltage
3. Set the meter for ac voltage
4. Make sure the meter leads do not exceed 36 inches

1-27. Digital multimeters are used to test semiconductors for which of the following reasons?

1. They produce voltage sufficient to gate all Zener diodes
2. Their LED displays are easier to read than analog displays
3. They typically limit the current flow through the semiconductor to less than 1 milliamp
4. They produce in excess of the 500 milliamps normally required to gate a PN junction

1-28. Compensation for the resistance in test leads of digital multimeters used to perform resistance measurements is accomplished by which, if any, of the following methods?

1. Short the leads, note the lead resistance displayed, and add this value to subsequent resistance measurements
2. Short the leads, note the lead resistance displayed, and subtract the value from subsequent resistance measurements
3. Add 10% to the reading
4. None of the above

1-29. Which of the following is a typical use for a megger?

1. Testing MOSFETs
2. Testing filter capacitors
3. Testing thermistor mounts
4. Testing an ac power cord for insulation breakdown

1-30. When large capacitors are stored as spare parts, why should their terminals be shorted with a piece of wire?

1. It prevents dielectric leakage
2. It prevents deterioration of the plates
3. It prevents the capacitors from becoming charged when in close proximity to an rf field
4. It prevents electrolytic capacitors from changing value during periods of storage

1-31. Capacitance meters can be grouped into which of the following basic categories?

1. Wheatstone type and Kelvin Varley type
2. Bridge-type and reactance-type
3. Depletion-type and enhancement-type
4. Resistive-type and reactive-type

1-32. Which of the following statements correctly describes the accuracy and use of a reactance-type capacitance meter?

1. It gives approximate values and is usually portable
2. It gives approximate values and is used in calibration laboratories only
3. It is very accurate and is usually portable
4. It is very accurate and is used to measure capacitors that have a high power factor

1-33. Which of the following types of inductor core materials produces the greatest inductance?

1. Mica
2. Magnetic metal
3. Polyparoloxylene
4. Nonmagnetic metal

1-34. As frequency increases, the inherent resistance of the inductor causes which of the following types of losses to become more critical?

1. Hysteresis
2. Skin effect
3. Eddy currents
4. Standing waves

1-35. Most capacitance test sets are capable of testing capacitors and what other type of component?

1. TRIACS
2. Inductors
3. Resistors
4. Barretters

1-36. When using reactance-type test equipment to measure inductance, what relationship exists between the inductor and the voltage drop across the reactance of the inductor?

1. The voltage drop is directly proportional to the value of inductance
2. The voltage drop is inversely proportional to the value of inductance
3. The voltage drop is proportional to the dielectric constant (K) of the inductor
4. The voltage drop is inversely proportional to the frequency of the applied voltage

.

1-37. Aboard ship you should be able to troubleshoot equipment failures to the component level for which of the following reasons?

1. Ships must be self-sustaining units when deployed
2. Storage space on board ships limits the number of bulky items or electronic modules that can be stored
3. Individual components may be easier to obtain than modules or larger equipment pieces
4. All of the above

1-38. What is the most common cause of electron tube failures?

1. Vibration damage
2. Open filaments
3. Shorted elements
4. Power supply voltage surges

1-39. The simplest way to test a tube is by which of the following methods?

1. Using a tube tester
2. Measuring tube element voltages
3. Feeling for signs of overheating
4. Substituting tubes

1-40. Test conditions for the electron tube tester described in the text are set by which of the following methods?

1. By a technician setting switches
2. By using a magnetic tape program
3. By using a prepunched card program
4. By inserting the tube into the appropriate socket

1-41. The electron tube tester can be used to test common low-power tubes for which of the following conditions?

1. Gas
2. Quality
3. Leakage
4. Each of the above

1-42. Pushbuttons on the electron tube tester are used to test for which of the following conditions?

1. Emission
2. Transconductance
3. Other quality tests
4. Each of the above

1-43. Which of the following tests is automatically performed when the electron tube tester card switch is first actuated?

1. Gas
2. Shorts
3. Opens
4. Quality

1-44. Which of the following methods is normally used to test high-power amplifier tubes?

1. Using tube testers
2. Performing interelectrode resistance checks
3. Making gain measurements with an oscilloscope
4. Observing built-in meters that measure grid and plate current and power output

1-45. Which, if any, of the following problems occur when klystrons are left in storage or not used for more than 6 months?

1. They become gassy
2. The elements become tarnished and ruin the tube
3. All external metallic parts become tarnished and must be cleaned prior to use
4. None

1-46. Which of the following actions should you take to restore operation if the klystron is gassy?

1. Replace the klystron with a new one
2. Return it to the nearest depot for intermediate maintenance
3. Evacuate the gas by igniting the tube's getter
4. Operate it at reduced beam voltage for approximately 8 hours

1-47. Traveling-wave tubes (twt) should be replaced if they deviate from design specifications by what minimum percentage?

1. 1%
2. 10%
3. 25%
4. 33%

1-48. If a twt used as an oscillator fails, which of the following indications should you observe?

1. The twt will become noisy
2. Equipment line fuses will blow
3. The twt will have reduced output power
4. The twt will fail to break into oscillation when all other conditions are normal

1-49. Which of the following is an appropriate reason to use transistors instead of electron tubes?

1. Transistors are more rugged than electron tubes
2. Transistors are not as heat sensitive as electron tubes
3. Transistors are not as sensitive as electron tubes to voltage overloads
4. Transistors are capable of handling greater amounts of power than electron tubes

1-50. When using an ohmmeter to test a transistor's base-to-emitter or base-to-collector junction, what minimum back-to-forward resistance ratio should you expect to read?

1. 5 to 1
2. 10 to 1
3. 50 to 1
4. 100 to 1

1-51. When taking forward and reverse resistance readings between a transistor's emitter and collector, what type of reading should you get?

1. Both the forward and reverse readings should be nearly the same
2. A short in both the forward and reverse directions
3. Less than 15 ohms when measuring in the forward direction and infinite in the reverse direction
4. Less than 15 ohms in the reverse direction and infinite in the forward direction

1-52. When using an ohmmeter to test transistors, you should avoid using R×1 range for which of the following reasons?

1. The R×1 range is not as accurate as the other ranges
2. Most ohmmeters do not produce sufficient voltage on the R×1 range to properly bias a transistor junction
3. Some ohmmeters produce in excess of 100 milliamps of current on the R×1 range and could possibly damage the transistor
4. The R×1 scale is not capable of measuring the high resistances that are typical of a PN junction when forward biased

1-53. When using a soldering iron to replace transistors, you must be sure there is no current leakage between the power source and the tip of the iron. Which of the following actions should you take if current leakage is detected?

1. Reduce the wattage of the heating element
2. Use an isolation transformer to power the soldering iron
3. Use a soldering gun instead of a soldering iron
4. Isolate the soldering iron from ground by disconnecting the soldering iron safety ground wire

1-54. Which of the following is a description of ESDS devices?

1. Components that are sensitive to electrostatic discharge
2. Components that are sensitive to the electromagnetic pulse produced by a nuclear detonation
3. State-of-the-art devices used to detect electronic emissions
4. Devices designed to withstand any type of electromagnetic or electrostatic interference

1-55. MOS and CMOS devices without input diode protection circuitry belong in which, if any, of the following ESDS device categories?

1. Sensitive devices
2. Very sensitive devices
3. Moderately sensitive devices
4. None of the above

1-56. Wearing a grounded wrist strap when repairing electronic circuit boards serves which of the following purposes?

1. It identifies you as being 2M qualified
2. It protects ESDS devices from damage
3. It protects the technician from electrical shock
4. It protects you from rf burns when working near radar antennas

1-57. What, if any, precaution should you take before you open a package that contains an ESDS device?

1. Rub the package against a dissimilar material
2. Discharge any static electricity by connecting a grounded lead to the package
3. Measure the static charge on the package with an oscilloscope to ensure that it is within tolerance
4. None

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Component Testing," pages 2-8 through 2-48. Chapter 3, "Quantitative Measurements," pages 3-1 through 3-15.

2-1. Which of the following servicing techniques applies to semiconductors?

1. Substituting a semiconductor with a known good semiconductor is a simple way to test them
2. Voltage and resistance measurements are taken prior to substituting semiconductors
3. Substituting semiconductors is cumbersome if more than one is bad or if they are soldered into the circuit
4. All of the above

2-2. What minimum ratio of back-to-forward resistance should you expect when testing a diode?

1. 1 to 1
2. 10 to 1
3. 50 to 1
4. 100 to 1

2-3. Which of the following characteristics of a diode cannot be determined by using a multimeter?

1. How the diode reacts to various voltages
2. How the diode reacts to various frequencies
3. Both 1 and 2 above
4. How the diode reacts to forward and reverse dc biasing

2-4. How are SCRs normally used in the Navy?

1. As rectifiers
2. As power control devices
3. As voltage regulators
4. As switching diodes in digital applications

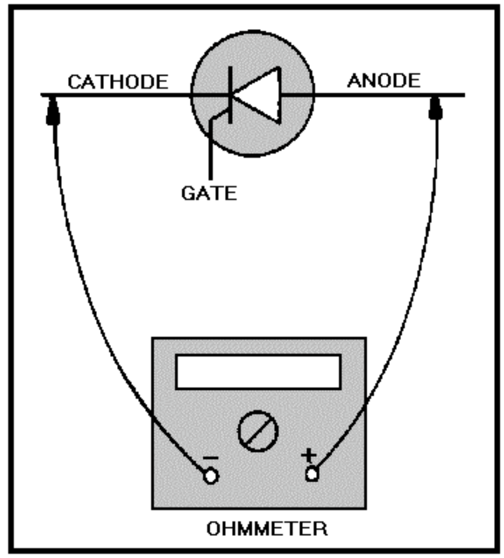


Figure 2A. —Testing an SCR with an ohmmeter.

IN ANSWERING QUESTIONS 2-5 AND 2-6, REFER TO FIGURE 2A. NOTE THAT THE CONNECTIONS OF THE OHMMETER ARE ALREADY MADE.

2-5. To forward bias an SCR, which elements should you short together?

1. The gate and anode
2. The cathode and anode
3. The cathode and gate
4. All three elements

2-6. What, if anything, will be the result of removing the short after it has been made?

1. Current flow from the cathode to the anode will stop
2. Current flow from the anode to the cathode will stop
3. Current can flow in either direction between the anode and cathode
4. Nothing

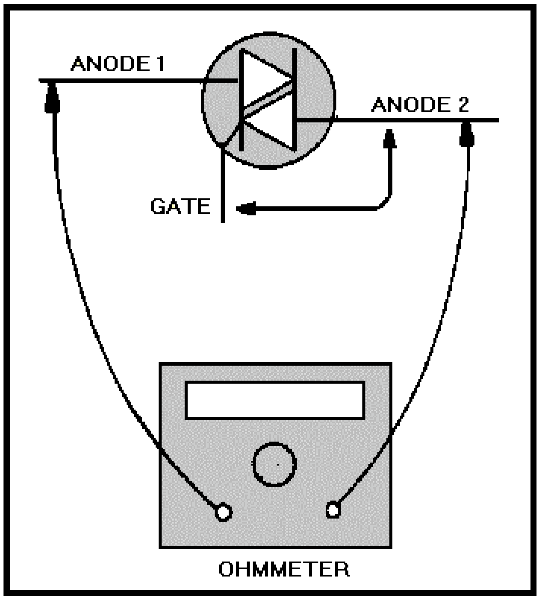


Figure 2B. —Testing a TRIAC with an ohmmeter.

IN ANSWERING QUESTION 2-7, REFER TO FIGURE 2B.

2-7. With a momentary short connected between the gate and anode 2, the TRIAC will be forward biased and allow current to flow between what elements?

1. From the gate to anode 1
2. From the gate to anode 2
3. From anode 1 to anode 2 only
4. In either direction between the two anodes

THIS SPACE LEFT BLANK INTENTIONALLY.

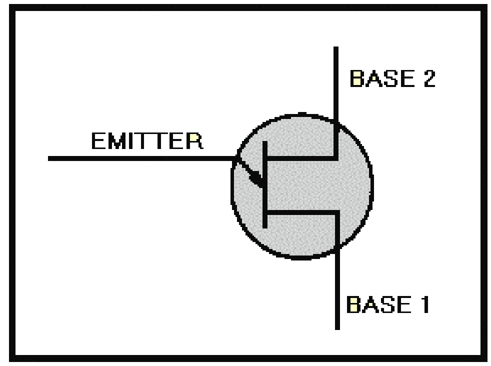


Figure 2C. —Unijunction transistor.

IN ANSWERING QUESTIONS 2-8 AND 2-9, REFER TO FIGURE 2C.

2-8. What readings should you expect to find when you measure the resistance between bases 1 and 2 of a UJT?

1. A short regardless of the polarity of the meter leads
2. A high resistance value regardless of the polarity of the meter leads
3. Approximately 15 ohms between base 1 and base 2 with the negative meter lead connected to base 1
4. Approximately 15 ohms between the two bases with the negative meter lead connected to base 2

2-9. For which, if any, of the following reasons do JFETs have circuit applications similar to those of vacuum tubes?

1. JFETs have a high input impedance and are voltage-responsive
2. JFETs have a low input impedance and a frequency response comparable to that of vacuum tubes
3. JFETs have a high input impedance and are current-responsive
4. None of the above

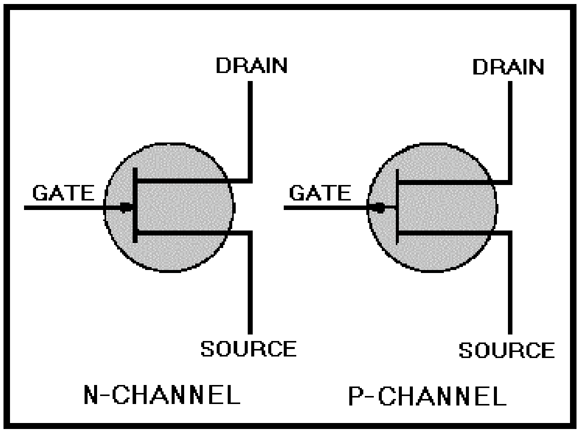


Figure 2D. —Junction FET.

IN ANSWERING QUESTION 2-10, REFER TO FIGURE 2E.

2-10. With the negative lead of an ohmmeter attached to the gate and the positive lead attached to the source, which of the JFETs in figure 2D would be good if the meter shows infinity?

1. P-channel
2. N-channel
3. Both 1 and 2 above
4. Neither 1 or 2

THIS SPACE LEFT BLANK INTENTIONALLY.

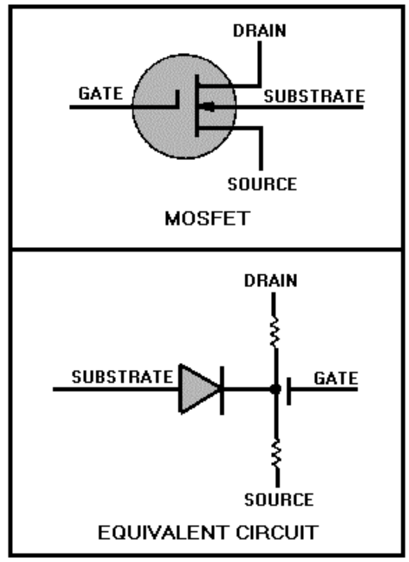


Figure 2E. —MOSFET (depletion/enhancement type) and equivalent circuit.

IN ANSWERING QUESTIONS 2-11 AND 2-12, REFER TO FIGURE 2E.

2-11. When measuring resistance between the drain and source of a depletion/enhancement type of MOSFET, what readings should you expect?

1. 15 ohms in one direction and infinity in the other direction
2. The same value of resistance in both directions
3. A short in one direction and infinity in the other
4. Infinite reading regardless of meter lead polarities

2-12. When measuring resistance between the gate, source, and drain of a depletion/enhancement type of MOSFET with the negative lead attached to the gate, what readings should you expect?

1. Both readings should be infinity
2. Both readings should be between 15 ohms and 100 ohms
3. Both readings should be approximately 1,000 ohms
4. Both readings should be less than 10 ohms

2-13. When unsoldering a MOSFET from a printed circuit board, you should avoid using a vacuum plunger solder sucker for which of the following reasons?

1. Solder suckers can generate high electrostatic charges that can damage MOSFETs
2. Solder suckers create a vacuum that can physically damage MOSFETs
3. Solder suckers are not authorized for any type of equipment repair
4. Solder suckers require the use of a high wattage soldering iron that may damage MOSFETs

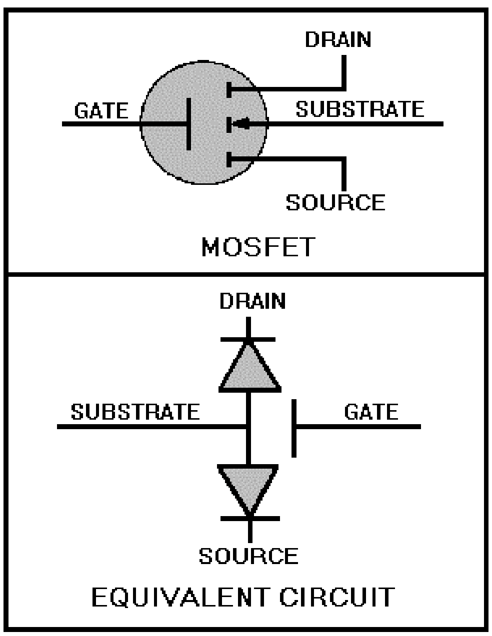


Figure 2F. —MOSFET (enhancement type) and equivalent circuit.

IN ANSWERING QUESTION 2-14, REFER TO FIGURE 2F.

2-14. When comparing resistance readings of an enhancement type of MOSFET to those of a depletion/enhancement type of MOSFET, which, if any, of the following differences should you notice?

1. The resistance between the substrate and the gate of the enhancement type should be less than 15 ohms
2. The measurement between the drain and source of the enhancement type should read infinite regardless of meter lead polarity
3. The resistances between the gate and the drain and between the gate and the source of the enhancement type should be noticeably higher
4. None of the above

2-15. Which of the following is/are (an) advantage(s) of integrated circuits when compared to circuits made up of separate components and interconnections?

1. Lower power consumption
2. Smaller size of the equipment
3. Lower equipment cost
4. All of the above

2-16. Which of the following is a characteristic of linear ICs?

1. They do not require regulated power supplies
2. They are typically sensitive to their supply voltages
3. They are never classed as electrostatic discharge sensitive devices
4. They are comparable in size to their equivalent transistor circuits

2-17. For which of the following reasons would you classify an IC as a "black box" device?

1. Because ICs are always black in color
2. Because all you can check are the inputs and outputs, not the internal operation of ICs
3. Because printed circuit boards that contain ICs cannot be repaired
4. Because ICs are designed to be repairable components

2-18. Test equipment used to detect the logic state of a digital IC should have which of the following characteristics?

1. A capability of measuring rms voltages
2. A frequency response in excess of 40 GHz
3. A high input impedance
4. A low input impedance

2-19. Which of the following statements describes) the use of logic clips?

1. Logic clips are designed to monitor the input and output of an IC simultaneously
2. Logic clips can only be used to test an IC that is out of the circuit
3. Logic clips can only be used with flat pack ICs
4. All of the above

2-20. For which of the following purposes are logic comparators used?

1. To test linear ICs
2. To compare different types of ICs
3. To inject pulse trains into digital ICs
4. For in-circuit testing of digital ICs by comparing them with reference ICs

2-21. Which of the following is an advantage of using a logic probe instead of an oscilloscope to test a digital IC?

1. Logic probes are usually larger than an oscilloscope but much lighter
2. Logic probes have a low input impedance
3. Logic probes are battery powered and do not react to line voltage variations
4. Logic probes are capable of detecting short-duration pulses that most oscilloscopes cannot display

2-22. For which of the following purposes are logic pulsers used?

1. To detect the logic state of digital ICs
2. To detect the logic state of linear ICs
3. To inject a pulse or pulse train into a logic circuit
4. To inject a 1-kHz sine wave into a circuit for signal tracing

2-23. Which of the following is a typical application for a logic analyzer?

1. To program EPROMs
2. To test individual logic ICs
3. To analyze the spectral purity at the output of a logic IC
4. To perform timing analysis by monitoring and comparing more than one timing signal simultaneously

2-24. Which of the following instruments is used to test the specific gravity of a lead-acid battery's electrolyte?

1. Hydrometer
2. Hygrometer
3. Electrometer
4. Gravitometer

2-25. Smoking is prohibited in the vicinity of lead-acid storage batteries for which of the following reasons?

1. Cigarette smoke neutralizes the electrolyte
2. Lead-acid batteries produce explosive hydrogen when they are being charged
3. Fumes produced by a lead-acid battery mixed with cigarette smoke produce a toxic by-product
4. All of the above

2-26. When testing dry cell batteries, which of the following procedures should you follow?

1. The battery should be tested under load conditions
2. The battery should not be tested under load conditions
3. The battery should be tested at various temperatures
4. Both 2 and 3 above

2-27. Which of the following dry cell batteries is rechargeable?

1. NICAD
2. Alkaline
3. Carbon-zinc
4. Mercury cells

2-28. Which of the following is the correct maximum charge rate for a NICAD battery rated at 300 milliampere hours?

1. 300 milliamperes for 15 hours
2. 60 milliamperes for 15 hours
3. 30 milliamperes for 15 hours
4. 600 milliamperes for 15 hours

2-29. Which of the following is a characteristic of fixed rf attenuators?

1. They are used to match impedances
2. They are designed to handle small amounts of rf power
3. They are usually built into the equipment in which they are used
4. They are capable of handling several kilowatts of power

2-30. Which of the following is an easy method of performing an operational test on a decade resistor?

1. Use an swr meter
2. Use the resistance substitution method
3. Read the resistance with an ohmmeter
4. Apply an rf voltage across the decade, measuring the voltage drop and computing the resistance

2-31. Which of the following is/are the disadvantages of glass-core, fiber-optic cables?

1. They are smaller in diameter than plastic-core fibers
2. They are extremely susceptible to mechanical damage
3. They exhibit signal losses as high as 25 dB/km
4. Both 2 and 3 above

2-32. Which of the following types of test equipment should you use to measure the losses in a fiber optic cable if only one end of the cable is accessible?

1. An optical ohmmeter
2. A Wheatstone bridge
3. An optical power meter
4. An optical time-domain reflectometer

2-33. When using the AN/USM-465 portable service processor, which of the following procedures is possible?

1. Identifying faulty components on digital printed circuit boards
2. Troubleshooting its own printed circuit boards
3. When using the guided probe, it will tell you if you have placed the probe on the wrong test point
4. All of the above

2-34. Which of the following measurements add resistance and inductive and capacitive reactance?

1. Q
2. Resonance
3. Impedance
4. Figure of merit

2-35. Bridge circuits are used in the measurement of impedance for which of the following reasons?

1. Bridges are one of the most accurate devices for measuring impedance
2. Bridges are only slightly less accurate than vtvm's when measuring impedance
3. Bridges are useful in measuring frequency
4. Both 2 and 3 above

2-36. Bridge circuits typically contain which of the following sections?

1. A measuring circuit and comparing circuit
2. A detector circuit
3. A power circuit
4. All of the above

2-37. When approximate values for resistance, capacitance, or inductance to be measured by a bridge are unknown, which, if any, of the following actions should you take?

1. Connect two bridges in parallel to make the measurement
2. Assign a temporary value to the component and set up the bridge accordingly
3. Place an adjustable shunt across the meter terminals
4. None of the above

2-38. The most serious errors affecting the accuracy of bridge measurements can be attributed to which of the following problems?

1. The capacitive and inductive characteristics of the connecting leads
2. The resistance of the test leads
3. D'Arsonval meter movements used as detectors
4. Improper selection of meter shunts

2-39. Which of the following considerations should be given when applying external excitation to a bridge circuit?

1. The voltage applied should equal the maximum voltage rating of the component under test
2. The higher the voltage, the more accurate the measurement
3. Apply only enough voltage to obtain a reliable indicator deflection
4. External excitation should be limited to 115 v 60 Hz

2-40. It is difficult to measure resistances less than 1 ohm with a bridge because of which of the following factors?

1. Contact resistance is present between the resistor being measured and the binding posts of the bridge
2. Excessive supply voltage is required to excite the galvanometer
3. The frequency of the excitation source creates excessive skin currents in the resistor under test
4. The excitation voltage causes low-value resistors to heat excessively

2-41. What type of bridge is recommended for measuring resistances less than 1 ohm?

1. Wheatstone bridge
2. Schering bridge
3. Maxwell bridge
4. Kelvin bridge

2-42. When using resistance-ratio bridges to measure capacitance, inductance, or resistance, you should compare the unknown component with which of the following components?

1. A capacitor
2. An inductor
3. A resistor
4. A similar standard

2-43. A Hay bridge measures unknown inductances by comparing them with which, if any, of the following components?

1. A standard inductor
2. A standard resistor
3. A standard capacitor
4. None of the above

2-44. What is the advantage of using a Maxwell bridge over a Hay bridge?

1. The Maxwell bridge can measure greater range of inductances
2. The Maxwell bridge can measure much smaller resistances
3. The Maxwell bridge can provide a greater accuracy over a smaller range
4. The Maxwell bridge can measure inductances having a high Q

2-45. Which of the following pieces of test equipment measure(s) the magnitude and phase angle of an unknown impedance?

1. The vector bridge
2. The impedance-angle meter
3. Both 1 and 2 above
4. The Hay bridge

2-46. Maximum transfer of rf energy between transmitter/receiver and antenna will occur under which of the following circumstances?

1. When the transmitter or receiver is properly matched to the antenna
2. When the receiver is tuned one sideband above the transmitter
3. When the transmitter is tuned one sideband above the receiver
4. Both 2 and 3 above

2-47. Rf impedance bridge measurements require the use of which of the following pieces of equipment?

1. An ac power source, a detector, and a Wheatstone bridge
2. An rf signal generator, an oscilloscope, and a power supply
3. An rf signal generator, an rf bridge, and a detector
4. A Schering bridge, a detector, and an rf power supply

2-48. What unit of measure is used to express the power level of a complex voice signal?

1. Vu
2. dB
3. dBm
4. Watt

2-49. The function of a dB meter is described in which of the following descriptions?

1. A current-measuring device
2. A user-calibrated constant current device
3. An electronic voltmeter calibrated in terms of dB
4. A frequency selective voltmeter calibrated in terms of true power

2-50. Electrodynamic wattmeters are used to measure which of the following types of power?

1. Ac power
2. Dc power
3. Both 1 and 2 above
4. Shf power in the 2-32 GHz frequency range

2-51. An electrodynamic wattmeter can be converted into an instrument for measuring reactive power by which of the following methods?

1. Installing a capacitor in series with the input
2. Shunting the meter movement with a 0. 1 ufd capacitor
3. Replacing the resistor which is normally in series with the voltage coil with a large inductance
4. Shunting the input terminals with an LC network adjusted to the resonant frequency of the signal being measured

2-52. What precaution(s), if any, must be taken when checking components with the Huntron Tracker 2000?

1. Voltages must not exceed 5 V dc
2. Voltages must not exceed 5 V ac
3. Device to be tested must have all power turned off and capacitors discharged
4. None of the above

2-53. When you are testing components by comparison, what is the most common mode used on the Huntron Tracker?

1. Automatic
2. Pulse generator
3. Single sweep
4. Alternate

2-54. Why is it necessary to electrically isolate a component while testing individual components with the Huntron Tracker 2000?

1. A resistor in series may give you an inaccurate signature
2. A diode in series may give you an inaccurate signature
3. A resistor in parallel may give you an inaccurate signature
4. All of the above

ASSIGNMENT 3

Textbook assignment: Chapter 3, “Quantitative Measurement,” pages 3-26 through 3-39. Chapter 4, “Qualitative Measurements,” pages 4-1 through 4-14. Chapter 5, “Introduction to Waveform Interpretation,” pages 5-1 through 5-35.

3-1. The AN/URM-120 in-line wattmeter is capable of measuring which of the following values?

1. Rf levels up to 500 watts between 30 MHz and 1,000 MHz
2. Rf levels up to 1 kW between 2 MHz and 30 MHz
3. Both 1 and 2 above
4. Af levels up to 500 watts between 1 kHz and 15 kHz

3-2. When rf power is applied to a bolometer, the heat generated by the semiconductor bead results in which of the following characteristic changes?

1. A capacitive change
2. An inductive change
3. A resistive change
4. A frequency change

3-3. The Hewlett-Packard 431 C power meter is capable of measuring power within which of the following frequency ranges in a coaxial system?

1. 1 MHz to 9 MHz
2. 10 MHz to 18 GHz
3. 41 GHz to 100 GHz
4. 101 GHz to 1,000 GHz

3-4. Calorimeters measure power by converting the input electromagnetic energy into which of the following forms?

1. Heat
2. Dc power
3. Pulsed rf energy
4. Electrodynamic energy

3-5. Which of the following relationships exist(s) between the temperature increase of the calorimetric body of a static calorimeter and the applied power?

1. The temperature increase is proportional to the frequency of the applied power
2. The temperature increase is inversely proportional to the amount of applied power
3. The temperature increase is directly proportional to the time of the applied power
4. Both 2 and 3 above

3-6. Which of the following statements describe(s) the method of using a twin calorimeter?

1. Rf power is applied to one calorimetric body and the other body acts as a temperature reference
2. The steady-state temperature difference between the two calorimetric bodies is used as a measure of rf power
3. Both 1 and 2 above
4. Power is applied to both calorimetric bodies through a directional coupler

3-7. Flow calorimeters are classified by the type of measurement performed, the type of heating used, and what other characteristic?

1. Number of calorimetric bodies
2. Type of circulation method used
3. Type of rf loads that they employ
4. Number of calorimetric fluids used

3-8. When performing measurements above 1 GHz in a flow calorimeter, which of the following dielectrics do you normally use?

1. Water
2. Oil
3. MEK
4. H_2SO_4

3-9. Which of the following government agencies is/are responsible for maintaining our primary?

1. U. S. National Bureau of Standards
2. U. S. Naval Observatory
3. Department of Weights and Measures
4. All of the above

3-10. When using a stroboscope to measure an unknown frequency, which, if any, of the following steps should you take?

1. Start the measurement at the lowest frequency that the stroboscope can deliver and increase the flashing rate until a single image is obtained
2. Start the measurement at the highest frequency that the stroboscope can deliver and reduce the flashing rate until a single stationary image is obtained
3. Start the measurement at the midscale range of the stroboscope and adjust the flashing rate, in either direction, until a harmonic of the primary frequency is obtained
4. None of the above

THIS SPACE LEFT BLANK INTENTIONALLY.

3-11. If you anticipate using a stroboscope over an extended period of time, which of the following actions can you take to extend flasher-tube life?

1. Operate the stroboscope at a submultiple of the fundamental synchronous speed
2. Lower the plate voltage of the flasher tube
3. Lower the filament voltage of the flasher tube
4. Operate the stroboscope at a multiple of the fundamental synchronous speed

3-12. Vibrating reed meters and moving disk meters are primarily used to measure which of the following values?

1. The frequency of 60-Hz ac power
2. The rotational speed of synchronous motors
3. Frequencies between 1 kHz and 10 MHz
4. The frequencies of multiplexed signals

3-13. When using an oscilloscope to measure frequencies, which of the following formulas should you use?

1. $f = \frac{1}{t}$
2. $f = t$
3. $f = \frac{t}{1}$
4. $f = (N_x) \cdot \frac{AB}{A - B}$

3-14. Most oscilloscopes are limited in their frequency-measuring capability to which of the following upper frequency limits?

1. 50 kHz
2. 100 kHz
3. 500 kHz
4. 100 MHz

3-15. Which of the following indications should you observe when a frequency meter is adjusted to the resonant frequency of the signal under test?

1. An audible beat-frequency signal
2. A pronounced dip in output at resonance
3. A pronounced increase in output power
4. A bright glow of the frequency meter glow lamp

3-16. What is the purpose of the time interval measurement of a frequency counter?

1. It indicates the wave period
2. It indicates the time between two events
3. It indicates the time between two functions of an event
4. Both 2 and 3 above

3-17. What are the three basic categories of wavemeters?

1. Resonant, active, and passive
2. Absorption, active, and passive
3. Reaction, resonant, and absorption
4. Absorption, reaction, and transmission

3-18. In becoming a qualified technician, which of the following goals should you seek to achieve?

1. To be able to repair a specific piece of equipment
2. To be able to isolate faults in an entire system
3. To demonstrate a basic knowledge of system interconnections
4. To demonstrate minimum maintenance ability on a piece of equipment

3-19. When attempting to correct a technical problem, which of the following procedures should you follow?

1. Use short cuts
2. Do random testing
3. Use a logical approach
4. Do a self-test of the equipment

3-20. Efficient operation of equipment is assured by which of the following actions?

1. Using tricks of the trade
2. Making quick repairs when problems occur
3. Observing system quality figures during preventive maintenance
4. Monitoring all system test points continuously

3-21. The standing-wave ratio (swr) in a transmission line is figured by using which of the following ratios?

1. Maximum voltage to maximum current
2. Maximum voltage to minimum voltage
3. Maximum current to maximum voltage
4. Minimum voltage to minimum current

3-22. Swr measurements are taken for which of the following purposes?

1. To determine the output frequency of the system under test
2. To determine the matching quality of the transmission line termination
3. To determine the coupling quality of the transmission line
4. To determine system output power

3-23. Couplers containing slots are used with rf probes to provide access to which of the following components?

1. Wavemeters
2. Unidirectional couplers
3. Open transmission lines
4. Waveguides

3-24. The wavelength of a standing wave is measured on a short-circuited, terminated line using a magnetic or electric probe in which of the following ways?

1. By multiplying the average current by the peak current
2. By dividing the average voltage by the peak voltage
3. By measuring the distance between a maximum voltage point and a maximum current point
4. By measuring the distance between alternate maximum or minimum current points along the line

3-25. A neon lamp moved parallel to a two-wire parallel transmission line will glow at its brightest at which of the following points?

1. Maximum current points
2. Maximum voltage points
3. Maximum voltage and current points
4. Maximum and minimum current points

3-26. A milliammeter moved parallel to a two-wire transmission line will show its highest indication at which of the following points?

1. Maximum voltage points
2. Maximum current points
3. Maximum current and voltage points
4. Maximum and minimum voltage points

3-27. Which of the following devices may be used to measure swr without measuring the standing wave?

1. Bridge
2. Rf probe
3. Neon lamps
4. Milliammeter

3-28. When using an RC bridge to measure swr, which of the following factor(s) must you consider as the applied frequency increases?

1. Skin effect
2. Stray inductance
3. Stray capacitance
4. All of the above

3-29. Before a newly constructed bridge can be calibrated, adjustments must be made for which of the following reasons?

1. To determine the frequency range of the bridge
2. To keep stray effects at a minimum
3. To adjust the rf voltage amplitude
4. To determine the characteristic impedance of the circuit

3-30. Which of the following formulas apply(ies) when measuring swr with a bridge?

1. $Swr = \dfrac{R_L}{R_O}$
2. $Swr = \dfrac{R_O}{R_L}$
3. Both 1 and 2 above--use the one that yields an swr ratio greater than 1 to 1
4. $Swr = \dfrac{R_O \times R_L}{R_O + R_L}$

3-31. The ideal impedance match between transmitter and load is

1. 1 to 1
2. 2 to 1
3. 3 to 1
4. 4 to 1

3-32. When comparing vswr and iswr, which, if any, of the following is the correct ratio?

1. Vswr will exceed iswr by a minimum of 100%
2. Vswr will exceed iswr by a minimum of 50%
3. Vswr and iswr ratios will be equal
4. None of the above

3-33. Electrical losses caused by transmission line deterioration are best measured using which of the following pieces of equipment?

1. A signal generator and a power meter
2. A signal generator and a frequency counter
3. An swr meter and an oscilloscope
4. A frequency counter and a power meter

3-34. If a 9.5 GHz, 20 watt signal is inserted into a transmission line, approximately what signal should be measured at the other end of the transmission line?

1. 5 GHz, 10 watts
2. 5 GHz, 20 watts
3. 0 GHz, 10 watts
4. 0 GHz, 20 watts

3-35. To accurately determine transmission line losses, you should perform insertion losses at which of the following frequencies?

1. Midrange of the transmission line's frequency spectrum
2. At the upper and lower entrances of the transmission line's frequency spectrum
3. Across the entire frequency spectrum of the transmission line
4. Midrange of the transmission line's frequency spectrum, plus and minus 10 kHz

3-36. Which of the following transmission line specifications is/are considered important?

1. Frequency
2. Characteristic impedance
3. Power-handling capabilities
4. All of the above

3-37. Mixing two or more frequencies across a nonlinear device produces which of the following signals?

1. Crosstalk
2. Intermodulation distortion
3. Single sideband (ssb) transmission
4. Undesirable carrier frequency deviation

3-38. Which of the following statements describes cross modulation?

1. Degenerative feedback that causes a circuit to oscillate
2. Overmodulation that produces an echo
3. The signal from one channel that modulates the signal on an adjacent channel
4. Oscillation that is caused by system misalignment

3-39. Distortion caused by excessive regenerative feedback is called

1. echo
2. crosstalk
3. detected distortion
4. parasitic generation

3-40. When using a two-tone test to detect intermodulation distortion, what is the ideal indication you should see on a spectrum analyzer?

1. An exact reproduction of the input frequencies
2. The sum and difference of the input frequencies
3. A single frequency with the amplitude equal to the sum of the input frequencies
4. The beat frequency of the two inputs

3-41. Which of the following actions minimizes the effects of intermodulation distortion?

1. Using proper antenna spacing
2. Shielding components and circuitry
3. Using parasitic suppression circuits
4. All of the above

3-42. At what point does an amplitude-modulated signal begin to produce distortion?

1. Below 50% modulation
2. At 65% modulation
3. At 95% modulation
4. Above 100% modulation

3-43. To obtain 100% amplitude modulation of an rf carrier with a sine wave, the modulating power must equal what minimum percent of the rf carrier power?

1. 10%
2. 15%
3. 25%
4. 50%

3-44. The damping of a meter movement that is being used to measure modulation has which of the following disadvantages?

1. The frequency response of the meter is reduced
2. The accuracy of the meter movement is reduced
3. An average reading does not disclose transient overmodulation
4. The amount of current required to drive the meter is reduced

3-45. Which of the following modulation patterns can be observed on an oscilloscope?

1. Wave-envelope and trapezoidal
2. Lissajous and wave-envelope
3. Time division and frequency division
4. Lissajous and trapezoidal

3-46. The frequency response of most oscilloscopes limits the capability of measuring percentage of modulation to which of the following frequency bands?

1. Lf and hf
2. Slf and shf
3. Uhf
4. Vhf

3-47. When using the two-tone test (trapezoidal method) to check a transmitter, you should see what pattern on the oscilloscope?

1. A series of fully modulated sine waves
2. A 100% amplitude-modulated signal
3. Two pulses of equal amplitude and duration
4. Two opposing triangles that are mirror images of each other

3-48. Frequency deviation of an fm signal is usually expressed in which of the following units of measurements?

1. Kilohertz
2. dB
3. dBm
4. Volts

3-49. What limits an fm transmitter's maximum frequency deviation?

1. The width of the band assigned for station operation
2. The maximum power output rating of the transmitter
3. The distortion that occurs at 100% modulation
4. The transmitting antenna height

3-50. Spectrum analysis is a graphic plot of

1. amplitude versus time
2. time versus frequency
3. amplitude versus frequency
4. amplitude versus power

3-51. Time-domain plots are used by technicians to graphically view which of the following waveform parameters?

1. Amplitude versus time
2. Frequency versus time
3. Frequency versus distance
4. Amplitude versus power

3-52. Frequency-domain plots are used by technicians to graphically view which of the following waveform parameters?

1. Amplitude versus time
2. Amplitude versus frequency
3. Frequency versus distance
4. Amplitude versus power

3-53. Which of the following pieces of test equipment should you use to determine what signals make up a complex signal?

1. Oscilloscope
2. Sweep oscillator
3. Spectrum analyzer
4. Time-domain reflectometer

3-54. At 100% amplitude modulation, the total power in the sidebands equals what percentage of the carrier power?

1. 6%
2. 50%
3. 66%
4. 100%

3-55. When viewing a 100% amplitude-modulated signal with a spectrum analyzer, what type of display should you observe?

1. A center frequency and both the upper and lower sidebands 6 dB down from the center frequency
2. A center frequency and both the upper and lower sidebands of equal amplitude
3. A center frequency that is -6 dB down from both the upper and lower sidebands
4. A suppressed carrier with both the upper and lower sidebands of equal amplitude

3-56. Which of the following is/are an advantage of ssb transmission?

1. The voice quality of ssb transmissions is superior to both AM and fm transmissions
2. Ssb transmissions are not susceptible to interference caused by sun spots and atmospherics
3. Ssb requires one-sixth of the output power and less than half the bandwidth required by AM to transmit the same amount of intelligence power
4. All of the above

3-57. In fm, the AMOUNT of frequency deviation (shift) is proportional to

1. the frequency of the carrier
2. the amplitude of the modulating signal
3. the frequency of the modulating signal
4. the plate current of the transmitter's linear amplifier

3-58. In fm, the RATE of frequency deviation (shift) is proportional to

1. the impedance of the antenna
2. the power output of the transmitter
3. the amplitude of the modulating signal
4. the frequency of the modulating signal

3-59. When analyzing the composition of a rectangular wave with a spectrum analyzer, which of the following types of displays will you see?

1. A fundamental frequency and its odd harmonics only
2. A fundamental frequency and its even harmonics only
3. A fundamental frequency and its combined even and odd harmonics
4. An infinite number of fundamental frequencies

3-60. The ability of a spectrum analyzer to resolve signals refers to its ability to

1. distinguish one signal from other signals
2. shape signals through the use of filters
3. determine a receiver's minimum discernible signal
4. measure the frequency of a signal

3-61. The ability of a spectrum analyzer to resolve signals is limited by which of the following factors?

1. The amplitude of the signal under test
2. The narrowest bandwith of the spectrum analyzer
3. The upper frequency limits of the spectrum analyzer
4. The lower frequency limits of the spectrum analyzer

3-62. Which of the following characteristics of a transmission line fault can be observed using time-domain reflectometry?

1. Nature of the fault
2. Distance to the fault
3. Both 1 and 2 above
4. Figure of merit of the fault

3-63. What is the primary application of swept-frequency testing?

1. To determine the broadband frequency response of a device
2. To determine the characteristics of a device at a specific frequency
3. To determine the impedance of a transmission line
4. To determine the swr of a transmission line

3-64. You should perform an initial power check on a transmitting antenna before sweeping the antenna for which of the following reasons?

1. To prevent damage to the test equipment
2. To ensure the transmitter is deenergized
3. To ensure the transmitter is energized
4. To ensure the transmitter is keyed

NONRESIDENT
TRAINING
COURSE
SEPTEMBER 1998

Navy Electricity and Electronics Training Series

Module 22—Introduction to Digital Computers

NAVEDTRA 14194

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Although the words “he,” “him,” and “his” are used sparingly in this course to enhance communication, they are not intended to be gender driven or to affront or discriminate against anyone.

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: To introduce the student to the subject of Digital Computers who needs such a background in accomplishing daily work and/or in preparing for further study.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and either the occupational or naval standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards,* NAVPERS 18068.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
FCCM(SW) Robert A. Gray

NAVSUP Logistics Tracking Number
0504-LP-026-8450

Sailor's Creed

"I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country's Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all."

TABLE OF CONTENTS

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 4 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

Student Comments

Course Title: *NEETS Module 22*
Introduction to Digital Computers

NAVEDTRA: 14194 **Date**: ____________________

We need some information about you:

Rate/Rank and Name: ____________________ SSN: ____________ Command/Unit ____________

Street Address: ______________________ City: ____________ State/FPO: ________ Zip ________

Your comments, suggestions, etc.:

Privacy Act Statement: Under authority of Title 5, USC 301, information regarding your military status is requested in processing your comments and in preparing a reply. This information will not be divulged without written authorization to anyone other than those within DOD for official use in determining performance.

NETPDTC 1550/41 (Rev 4-00)

CHAPTER 1

OPERATIONAL CONCEPTS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions placed throughout the chapters are based on the objectives. By successfully completing the Nonresident Training Course (NRTC), you indicate that you have met the objectives and have learned the information. The learning objectives for this chapter are listed below.

Upon completion of this chapter, you will be able to do the following:

1. Describe the history of computers.
2. Describe how computers are classified.
3. Explain how digital computers have changed during each generation.
4. Describe the practical applications of digital computers in the Navy.
5. Describe the initial steps needed to use a microcomputer.
6. Explain storage media handling, backups, and the threats to storage media.

INTRODUCTION

Digital computers are used in many facets of today's Navy. It would be impossible for one *NEETS* module to cover all the ways they are used in any depth. A few of these ways are covered later in this chapter.

The purpose of this module is to acquaint you, the trainee, with the basic principles, techniques, and procedures associated with digital computers. We will use a desktop (personal) computer for most of the examples. Personal computers should be more familiar to you than the large mainframes, and the operating principles of personal computers relate directly to the operating principles of mainframe computers. You will learn the basic terminology used in the digital-computer world. When you have completed these chapters satisfactorily, you will have a better understanding of how computers are able to perform the demanding tasks assigned to them.

If we were to define the word computer, we would say a computer is an instrument for performing mathematical operations, such as addition, multiplication, division, subtraction, integration, vector resolution, coordinate conversion, and special function generation at very high speeds. But the usage of computers goes well beyond the mathematical-operations level.

Computers have made possible military, scientific, and commercial advances that before were considered impossible. The mathematics involved in orbiting a satellite around the earth, for example, would require several teams of mathematicians for a lifetime. Now, with the aid of electronic digital computers, the conquest of space has become reality.

Computers are employed when repetitious calculations or the processing of large amounts of data are necessary. The most frequent applications are found in the military, scientific, and commercial fields. They are used in many varied projects, ranging from mail sorting, through engineering design, to the identification and destruction of enemy targets. The advantages of digital computers include speed, accuracy, reliability, and man-power savings. Frequently computers are able to take over routine jobs, releasing people for more important work; work that cannot be handled by a computer.

HISTORY OF COMPUTERS

The ever increasing need for faster and more efficient computers has created technological advances that can be considered amazing. Ever since humans discovered that it was necessary to count objects, we have been looking for easier ways to do it. Contrary to popular belief, digital computers are not a new idea. The abacus is a manually operated digital computer used in ancient civilizations and used to this day in the Orient (see fig. 1-1). For those who consider the abacus outdated, in a contest between a person using a modern calculator and a person using an abacus, the person using the abacus won.

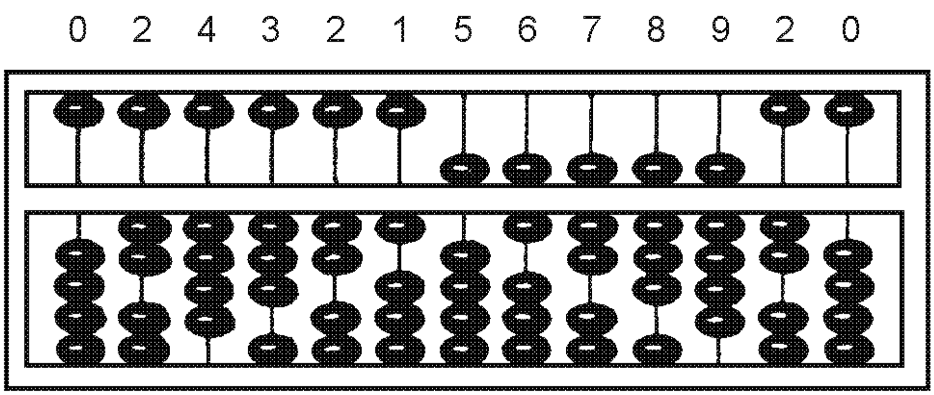


Figure 1-1.—Abacus.

The first mechanical adding machine (calculator) was invented by Blaise Pascal (French) in 1642. Twenty years later, an Englishman, Sir Samuel Morland, developed a more compact device that could multiply, add, and subtract. In 1682, Wilhelm Liebnitz (German) perfected a machine that could perform all the basic operations (addition, subtraction, division, and multiplication), as well as extract the square root. Liebnitz's principles are still in use today in our modern electronic digital computers.

As early as 1919, electronics entered the scene. An article by W. H. Eccles and F. W. Jordan described an electronic "trigger circuit" that could be used for automatic counting. It was the ECCLES-JORDAN multivibrator which was a little ahead of its time because a trigger circuit is one of many components required to make an electronic digital computer. Modern digital computers use these circuits, known as flip-flops, to store information, perform arithmetic operations, and control the timing sequences within the computer.

Under the pressure of military needs in World War II, the science of electronic data processing made giant strides forward. In 1944, Harvard University developed a computing system known as the Automatic Sequence Controlled Calculator. After the initial design and construction, several improved models were built.

Meanwhile, at the University of Pennsylvania, a second system was being developed. This system, completed in 1946, was named ENIAC (Electronic Numerical Integrator and Computer). ENIAC employed 18,000 vacuum tubes in its circuitry; and in spite of these bulky, hot tubes, it worked quite successfully. The first problem assigned to ENIAC was a calculation in nuclear physics that would have taken 100 human-years to solve by conventional methods. The ENIAC solved the problem in 2 weeks, only 2 hours of which were actually spent on the calculation. The remainder of the time was spent checking the results and operational details. All modern computers have their basics in these two early developments conducted at Harvard University and University of Pennsylvania.

In 1950, the UNIVAC I was developed. This machine was usually regarded as the most successful electronic data processor of its day. An outstanding feature of the UNIVAC I was that it checked its own results in each step of a problem; thus eliminating the need to run the problems more than once to ensure accuracy.

During the first outbreak of publicity about computers (especially when the UNIVAC predicted the outcome of the 1952 presidential election), the term "giant brain" caused much confusion and uneasiness. Many people assumed that science had created a thinking device superior to the human mind. Currently most people know better. By human standards the giant brain is nothing more than a talented idiot that is wholly dependent upon human instructions to perform even the simplest job. A computer is only a machine and definitely cannot think for itself. The field of artificial intelligence, however, is developing computer systems that can "think"; that is, mimic human thought in a specific area and improve performance with experience and operation. The field of digital computers is still in the growing stages. New types of circuitry and new ways of accomplishing things are continuing to be developed at a rapid rate.

In the military field, the accomplishments of digital computers are many and varied. One outstanding example is in weapons systems. Most of the controlling is done by digital computers.

CLASSIFICATIONS OF COMPUTERS

Computers can be classified in many different ways. They can be classified by the type of technology they use (mechanical, electromechanical, or electronic), the purpose for which they were designed (general purpose or special purpose), by the type of data they can handle (digital or analog), by the amount they cost (from $50 to $10 million and up), and even by their physical size (handheld to room size). We will briefly explain mechanical, electromechanical, and electronic computers; special-purpose and general-purpose computers; and analog and digital computers.

MECHANICAL COMPUTERS

Mechanical or analog computers are devices used for the computation of mathematical problems. They are made up of components, such as integrators, sliding racks, cams, gears, springs, and driveshafts. Figure 1-2 shows a typical mechanical computer used by the Navy. These computers are analog in nature, and their physical size depends on the number of functions the computer has to perform. In an analog computer, a continuing input will give a constantly updated output. This being perfect for target information, the Navy uses these analog computers primarily for gun fire control. As systems for naval weapons became more and more complex, the need for a different computer was apparent. The functions that had to be performed had increased the size of the computer to an unreasonable scale.

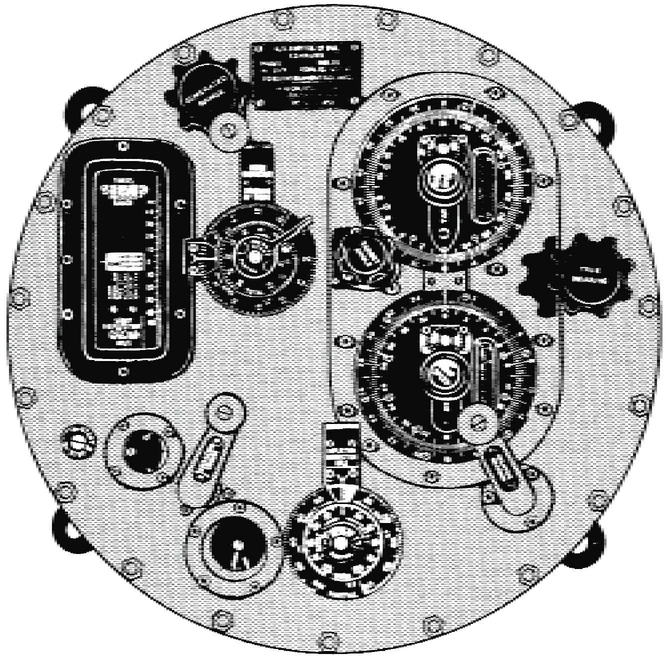


Figure 1-2.—Bulkhead-type mechanical computer

ELECTROMECHANICAL COMPUTERS

Electromechanical computers came next and differ from mechanical computers in that they use electrical components to perform some of the calculations and to increase the accuracy. Because the electrical components are smaller than their mechanical counterparts, the size of the computer was reduced, even though it performs more functions. The components used to perform the calculations are devices such as synchros, servos, resolvers, amplifiers, servo amplifiers, summing networks, potentiometers, and linear potentiometers. Figure 1-3 shows one of the Navy's electromechanical computers. These computers are used in gun fire control and missile fire control. Even though they are better than the mechanical computer, they still have their drawbacks. Of prime importance is that they are special-purpose computers. This means they can only be used for one job, dependent on their design characteristics. By today's Navy standards they are still too large, and the maintenance time on them is excessive. The need for a more accurate, reliable, versatile, and smaller computer was recognized.

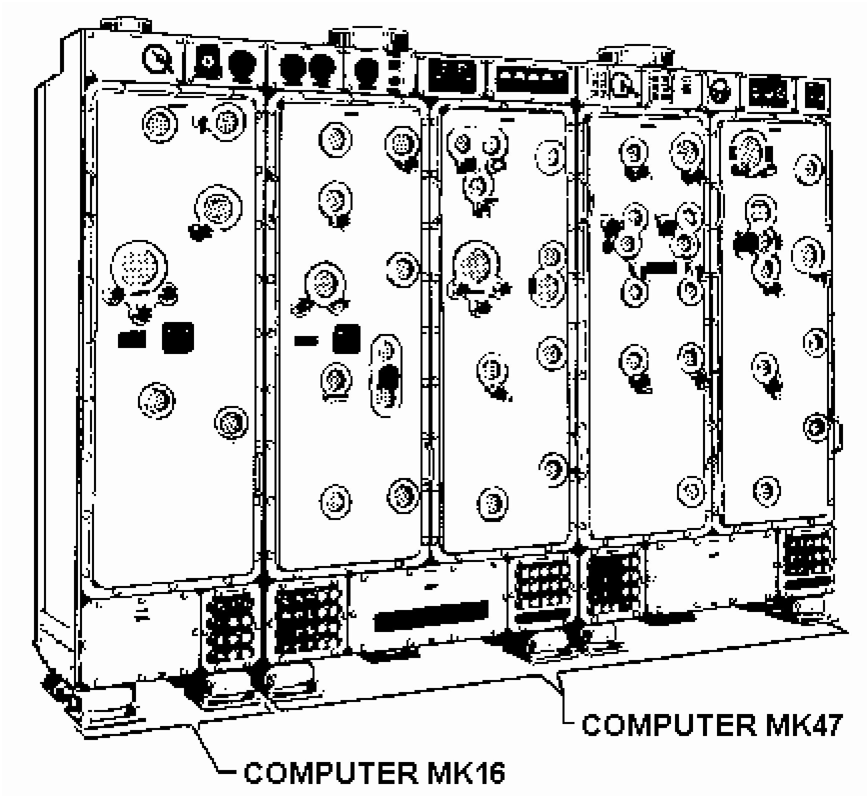


Figure 1-3.—Electromechanical computer.

ELECTRONIC COMPUTERS

Next came electronic computers. The early electronic computers' mathematical processes were solved by using electrical voltages only, applied to elements such as amplifiers, summing networks, differentiating, and integrating circuits. The weak link in this type of electrical computation was the vacuum tube. To correct this, transistors which consume less power and last longer than vacuum tubes were used in the amplifiers. Through technological research and development, we have progressed from tubes, to transistors, to miniaturized circuits, to integrated circuitry. These advances have made it possible to reduce the size and weight of our computers. Figure 1-4 is an example of one of our modern electronic digital computers.

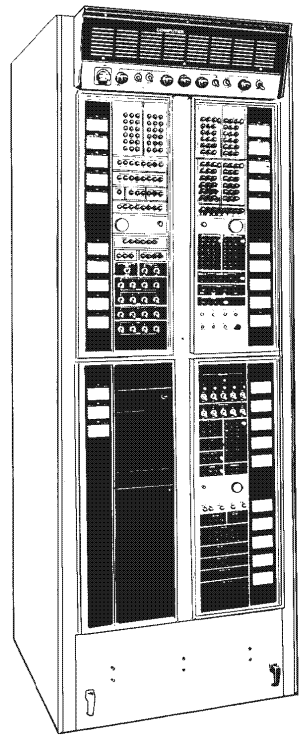


Figure 1-4.—Electronic digital computer.

SPECIAL-PURPOSE COMPUTERS

A special-purpose computer, as the name implies, is designed to perform a specific operation and usually satisfies the needs of a particular type of problem. Such a computer system would be useful in weather predictions, satellite tracking, or oil exploration. While a special-purpose computer may have many of the same features found in a general-purpose computer, its applicability to a particular problem is a function of its design rather than to a stored program. The instructions that control it are built directly into the computer, which makes for a more efficient and effective operation. A drawback of this specialization, however, is the computer's lack of versatility. It cannot be used to perform other operations.

GENERAL-PURPOSE COMPUTERS

General-purpose computers are designed to perform a wide variety of functions and operations. You will probably use this type of computer. A general-purpose computer is able to perform a wide variety of operations because it can store and execute different programs in its internal storage. Unfortunately, having this ability is often achieved at the expense of speed and efficiency. In most situations, however, you will find that having this flexibility makes this compromise a most acceptable one.

ANALOG COMPUTERS

All analog computers are special-purpose computers. They are designed to measure continuous electrical or physical conditions, such as current, voltage, flow, temperature, length, or pressure. They then convert these measurements into related mechanical or electrical quantities. The early analog computers were strictly mechanical or electromechanical devices. They did not operate on digits (in binary notation, either of the characters, 0 and 1). If digits were involved at all, they were obtained indirectly. Your wrist watch (if nondigital); your car's speedometer; and oil pressure, temperature, and fuel gauges are also considered analog computers. The output of an analog computer is often an adjustment to the control of a machine; such as, an adjustment to a valve that controls the flow of steam to a turbine generator or a temperature setting to control the ovens in the ship's galley for baking. Analog computers are also used for controlling processes. To do so, they must convert analog data to digital form, process it, and then convert the digital results back to analog form.

You should know that a digital computer can process data with greater accuracy than an analog computer, but an analog computer can process data faster than a digital computer, in some systems. Some computers combine the functions of both analog and digital computers. They are called hybrid computers.

DIGITAL COMPUTERS

Digital computers perform arithmetic and logic functions on separate discrete data, like numbers, or combinations of discrete data, such as name, rate, and division. This makes them different from analog computers that operate on continuous data, like measuring temperature changes. We generally use digital computers for business and scientific data processing. The following are examples:

Accounting—Computers are ideal for keeping payroll records, printing paychecks, billing customers, preparing tax returns, and taking care of many of the other accounting tasks in an organization.

Recordkeeping—Computers can record information like inventories and personnel files. They can also keep track of books checked out of a library. Airline ticket counters are much more efficient than they used to be, thanks to centralized reservation computers that can be reached over the telephone lines.

Industrial Uses—Industrial computers save considerable time and reduce waste by efficiently performing hundreds of industrial tasks, ranging from filling sales orders and routing parts to various locations on an assembly line, to designing earthquake-resistant structures, and controlling an entire oil refinery.

Science—The research and development applications are the most numerous. Digital computers are being used to do lengthy and complicated mathematical calculations millions of times faster than human beings. They are also used to collect, store, and evaluate data from experiments, analyze weather patterns, forecast crop statistics, and, believe it or not, design other computers.

Word Processing—Remember, these words were typed into a desktop computer! Word processing is among the most common applications for personal computers. If you have not discovered the advantages of computer writing, it's time to visit a computer dealer for a personalized demonstration.

None of this work could be performed by a computer without first instructing the computer how to do it by means of a list of instructions called a program. The instructions in the program must be written in one of the languages the computer understands. The most popular generic term for computer programs is software (this is covered in chapter 3). Hardware (covered in chapter 2), of course, refers to the computer and related equipment. It is easy to say that both computer hardware and software are interdependent because neither can perform useful work without the other. Digital computers may be either special or general purpose.

ACCURACY OF COMPUTERS

The fundamental difference between analog and digital computers is that digital computers deal with discrete quantities such as beads on an abacus, notches on a toothed wheel, or electrical pulses, while analog computers deal with continuous physical variables such as electrical voltages or mechanical shaft rotations. Computation with analog computers depends on the relation of information to a measurement of some physical quantity. For example, you can determine the number of boards in a picket fence by either a digital or an analog system as follows. In the digital method (fig. 1-5, view A), you use an adding machine and count the boards one by one. In the analog method (fig. 1-5, view B), you draw a string (marked off in inches for the width of each board including the gap) over the length of the fence, then measure the length of the string. The number of boards may then be determined by dividing the length of string by the number of inches per board.

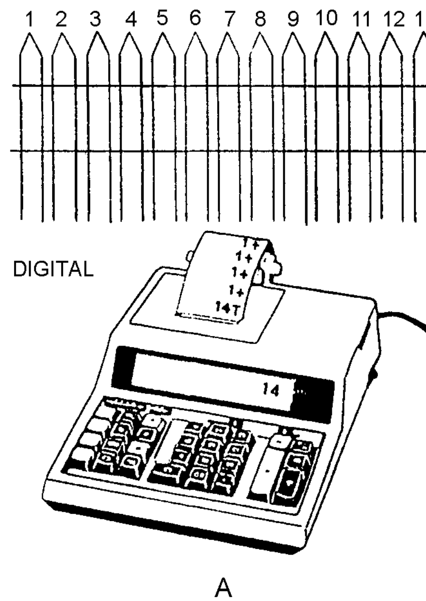


Figure 1-5A.—Digital computation.

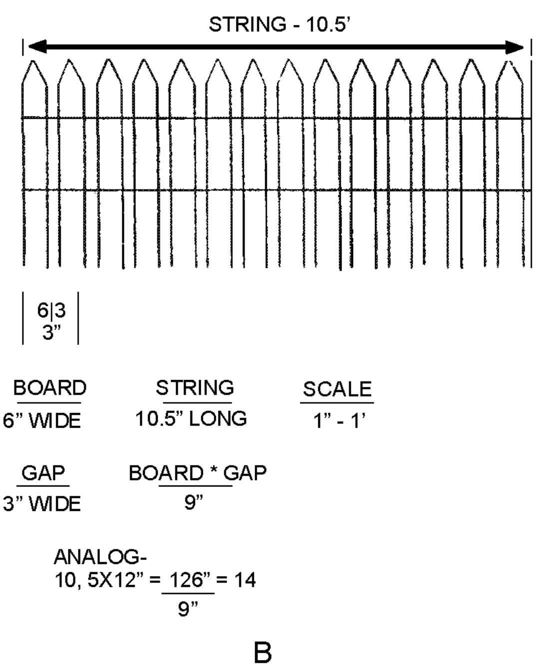


Figure 1-5B.—Analog computation.

The accuracy of an analog computer is restricted to the accuracy with which physical quantities can be sensed and displayed. This, in turn, is related to the quality of the components used in constructing the computer; for example, the tolerance of electrical resistors or mechanical shafts and the quality of the output equipment. In an analog computer, for example, if the constant is represented by a voltage, it probably could be read only to the third decimal place.

On the other hand, the accuracy of a digital computer is governed by the number of significant figures carried in the computations. This, in turn, is determined by the computer's design. In a digital computer, the number of decimal places in the constant could be many, depending on the design of the computer processing unit. The digital computer is, therefore, capable of higher precision and accuracy. However, a computer, regardless of its accuracy, would do you no good if the wrong one were chosen for a given task.

Most of the computer systems you will work with will be general-purpose digital computers. The remainder of this module will be about general-purpose digital computers.

Q-1. How are computers classified?

Q-2. Mechanical computers are considered to be of what type?

Q-3. The Navy uses analog computers primarily for what purpose?

Q-4. How do electromechanical computers differ from the mechanical computers?

Q-5. In electronic computers, vacuum tubes were replaced by transistors and transistors have been replaced by what device?

Q-6. A computer that is designed to perform a specific operation and usually satisfies the needs of a particular type of problem, is said to be what type of computer?

Q-7. Rather than using a stored program, a special-purpose computer's applicability to a particular problem is a function of what?

Q-8. What is a drawback to the special-purpose computer?

Q-9. A general-purpose computer is designed for what purpose?

Q-10. How is a general-purpose computer able to perform different operations?

Q-11. In a general-purpose computer, the ability to perform a wide variety of operations is achieved at the expense of what capabilities?

Q-12. All analog computers are what type of computers?

Q-13. What are analog computers designed to measure?

Q-14. Early analog computers were what type of devices?

Q-15. What are computers called that combine the functions of both analog and digital computers?

Q-16. Digital computers are generally used for what purposes?

Q-17. What is the fundamental difference between analog and digital computers?

Q-18. How is the accuracy of an analog computer restricted?

Q-19. A constant represented by a voltage can be read to what decimal place?

Q-20. The accuracy of a digital computer is governed by what factor?

Q-21. In a digital computer, what does the number of decimal places in the constant depend on?

Q-22. You will most likely be working with what type of computer?

DIGITAL COMPUTER GENERATIONS

In the electronic computer world, we measure technological advancement by generations. A specific system is said to belong to a specific "generation." Each generation indicates a significant change in computer design. The UNIVAC I represents the first generation. Currently we are moving toward the fourth generation.

FIRST GENERATION

The computers of the first generation (1951-1958) were physically very large machines characterized by the vacuum tube (fig. 1-6). Because they used vacuum tubes, they were very unreliable, required a lot of power to run, and produced so much heat that adequate air conditioning was critical to protect the computer parts. Compared to today's computers, they had slow input and output devices, were slow in processing, and had small storage capacities. Many of the internal processing functions were measured in thousandths of a second (millisecond). The software (computer program) used on first generation computers was unsophisticated and machine oriented. This meant that the programmers had to code all computer instructions and data in actual machine language. They also had to keep track of where instructions and data were stored in memory. Using such a machine language (see chapter 3) was efficient for the computer but difficult for the programmer.

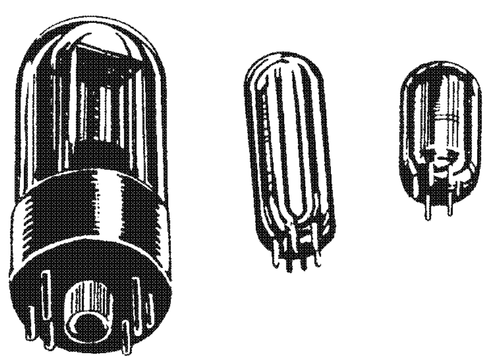


Figure 1-6.—First generation computers used vacuum tubes.

SECOND GENERATION

The computers of the second generation (1959-1963), were characterized by transistors (fig. 1-7) instead of vacuum tubes. Transistors were smaller, less expensive, generated almost no heat, and required very little power. Thus second generation computers were smaller, required less power, and produced a lot less heat. The use of small, long lasting transistors also increased processing speeds and reliability. Cost performance also improved. The storage capacity was greatly increased with the introduction of magnetic disk storage and the use of magnetic cores for main storage. High speed card readers, printers, and magnetic tape units were also introduced. Internal processing speeds increased. Functions were measured in millionths of a second (microseconds). Like the first generation, a particular computer of the second generation was designed to process either scientific or business oriented problems but not both. The software was also improved. Symbolic machine languages or assembly languages were used instead of actual machine languages. This allowed the programmer to use mnemonic operation codes for instruction operations and symbolic names for storage locations or stored variables. Compiler languages were also developed for the second generation computers (see chapter 3).

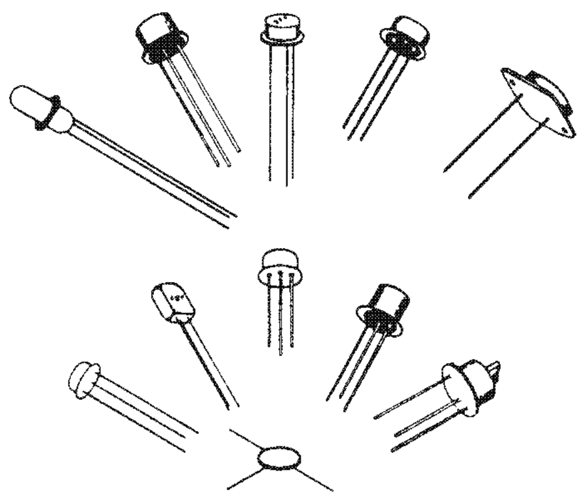


Figure 1-7.—Second generation computers used transistors.

THIRD GENERATION

The computers of this generation (1964-1970), many of which are still in use, are characterized by miniaturized circuits. This reduces the physical size of computers even more and increases their durability and internal processing speeds. One design employs solid-state logic microcircuits (fig. 1-8) for which conductors, resistors, diodes, and transistors have been miniaturized and combined on half-inch ceramic squares. Another smaller design uses silicon wafers on which the circuit and its components are etched. The smaller circuits allow for faster internal processing speeds resulting in faster execution of instructions. Internal processing speeds are measured in billionths of a second (nanoseconds). The faster computers make it possible to run jobs that were considered impractical or impossible on first or second generation equipment. Because the miniature components are more reliable, maintenance is reduced. New mass storage, such as the data cell, was introduced during this generation, giving a storage capacity of over 100 million characters. Drum and disk capacities and speed have been increased, the portable disk pack has been developed, and faster, higher density magnetic tapes have come into use. Considerable improvements were made to card readers and printers, while the overall cost has been greatly reduced. Applications using online processing, real-time processing, time sharing, multiprogramming, multiprocessing, and teleprocessing have become widely accepted. More on this in later chapters.

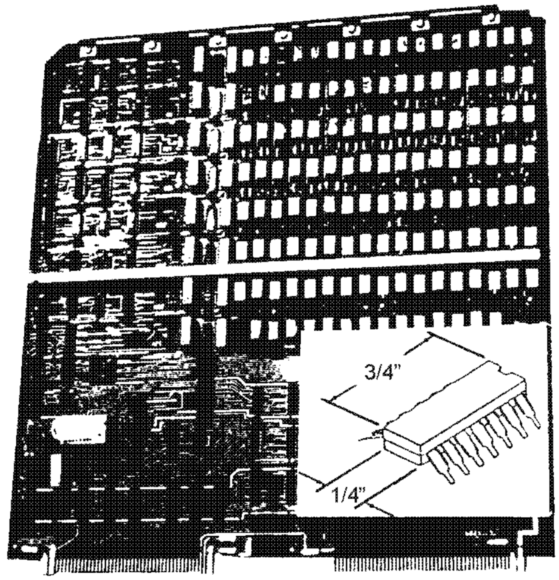


Figure 1-8.—Third generation computers used microcircuits.

Manufacturers of third generation computers are producing a series of similar and compatible computers. This allows programs written for one computer model to run on most larger models of the same series. Most third generation systems are designed to handle both scientific and business data processing applications. Improved program and operating software has been designed to provide better control, resulting in faster processing. These enhancements are of significant importance to the computer operator. They simplify system initialization (booting) and minimize the need for inputs to the program from a keyboard (console intervention) by the operator.

FOURTH GENERATION AND BEYOND

The computers of the fourth generation are not easily distinguished from earlier generations, yet there are some striking and important differences. The manufacturing of integrated circuits has advanced to the point where thousands of circuits (active components) can be placed on a silicon wafer only a fraction of an inch in size (the computer on a chip). This has led to what is called large scale integration (LSI) and very large scale integration (VLSI). As a result of this technology, computers are significantly smaller in physical size and lower in cost. Yet they have retained large memory capacities and are ultra fast. Large mainframe computers are increasingly complex. Medium sized computers can perform the same tasks as large third generation computers. An entirely new breed of computers called microcomputers (fig. 1-9) and minicomputers are small and inexpensive, and yet they provide a large amount of computing power.

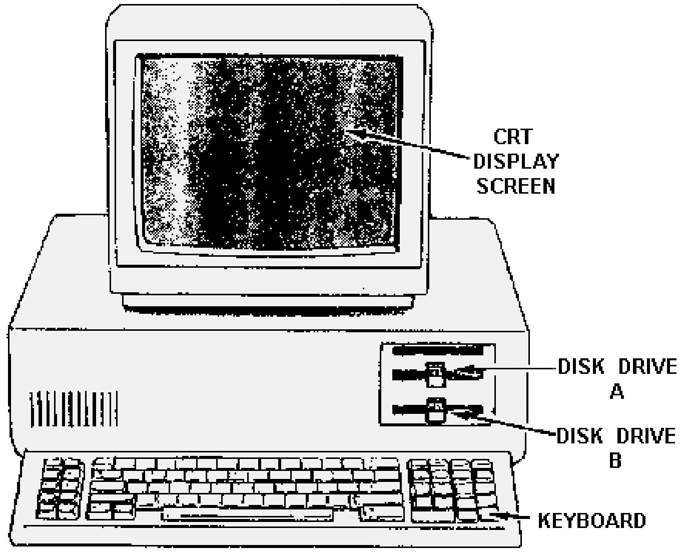


Figure 1-9.—Fourth generation desktop (personal) computer.

What is in store for the future? The computer industry still has a long way to go in the field of miniaturization. You can expect to see the power of large mainframe computers on a single super chip. Massive data bases, such as the Navy's supply system, may be written into read-only memory (ROM) on a piece of equipment no bigger than a desktop calculator (more about ROM in chapter 2). The future challenge will not be in increasing the storage or increasing the computer's power, but rather in properly and effectively using the computing power available. This is where software (programs such as assemblers, report generators, subroutine libraries, compilers, operating systems, and applications programs) will come into play (see chapter 3). Some believe developments in software and in learning how to use these extraordinary, powerful machines we already possess will be far more important than further developments in hardware over the next 10 to 20 years. As a result, the next 20 years (during your career) may be even more interesting and surprising than the last 20 years.

Q-23. Technological advancement is measured by what, in the electronic computer world?

Q-24. What does each generation of computer systems indicate?

Q-25. What were computers of the first generation characterized by?

Q-26. How did vacuum tubes cause a problem for first generation computers?

Q-27. In first generation computers, internal processing functions were measured by what division of time?

Q-28. The software (computer program) used on first generation computers was what type?

Q-29. How were processing speed and reliability increased in second generation computers?

Q-30. In second generation computers, how was the storage capacity greatly increased?

Q-31. With improvements in software, what kind of computer languages could be used on second generation computers?

Q-32. What do the smaller circuits in third generation computers allow for?

Q-33. On third generation computers, what results are gained by faster internal processing speeds?

Q-34. The data cell had a storage capacity of how many characters?

Q-35. What type of applications were most third generation computer systems designed to accomplish?

Q-36. What type of computers are small and inexpensive yet provide a lot of computing power?

Q-37. What does the acronym ROM stand for?

Q-38. What will be one of the future challenges involving computer power?

Q-39. What term is used for programs such as assemblers, compilers, and operating systems?

USES OF A DIGITAL COMPUTER

In the modern computer world of today, the uses of the digital computer are almost as limitless as a person's imagination. New and better programs are being written everyday for easier and greater uses. Consider how many mathematicians it would take to put an astronaut in orbit around the moon, but it only takes one computer. Think back to the days without word processing when a document had to be retyped entirely when any changes were needed. Think back to the days of using an adding machine to prepare and revise budgets and accounting reports. Let's look at three of the primary uses of general-purpose digital computers in the Navy: word processing, accounting/recordkeeping, and work center uses.

WORD PROCESSING

One of the more widespread uses of the computer is word processing. The word processor can be considered a typewriter with a display screen. To the hundreds of thousands of word processor users, the computer is nothing more than a typewriter. Both have keyboards, and both have a mechanism for making the image of the character you strike on the keyboard appear on some type of visual medium. When using an electric typewriter, the process is strictly mechanical. When you press the key, it causes the type face to strike the paper, and in so doing, it leaves an impression. In the computer, the process is more indirect. A program stored in the computer's memory causes a visual representation to appear on a crt (cathode-ray tube) or at a printer. However, from the view point of the user, the result is the same, a printed document.

The great advantage of computers over typewriters is in correcting errors. In the past, correcting a document with a typewriter has meant typing it all over again. Since computers allow the movement of information from one part of memory to another, it is possible to make many changes on a document, and print the result. If the document is still not correct, only the changes need to be entered. The use of computers in this particular way came to be known as word processing.

A further breakthrough came with the development of word-processing application programs for microcomputers. These programs cost a fraction of their office machine counterparts, and could be run on general-purpose microcomputers. This was unique because general-purpose microcomputers could be used for functions such as spreadsheets, data base management systems, and programming in common computer languages.

The Navy saw the obvious uses to which microcomputers using the word processing programs could be put. Some of these are manuscript writing, memorandum writing, identification-card application filing, and recordkeeping.

ACCOUNTING AND RECORDKEEPING

There are virtually unlimited applications for the computer in today's modern business world, from basic accounting functions to controlling the manufacture of products, and of course, keeping records of these actions. Six standard systems dealing with accounting applications are widely accepted. These systems are (1) order entry; (2) inventory control; (3) accounts receivable; (4) accounts payable; (5) general ledger; and (6) payroll. (Figure 1-10 shows a simplified flowchart of payroll.) The area of recordkeeping has two requirements, legal and audit. The Navy has included similar functions in its Shipboard Non-Tactical ADP Program for work center use.

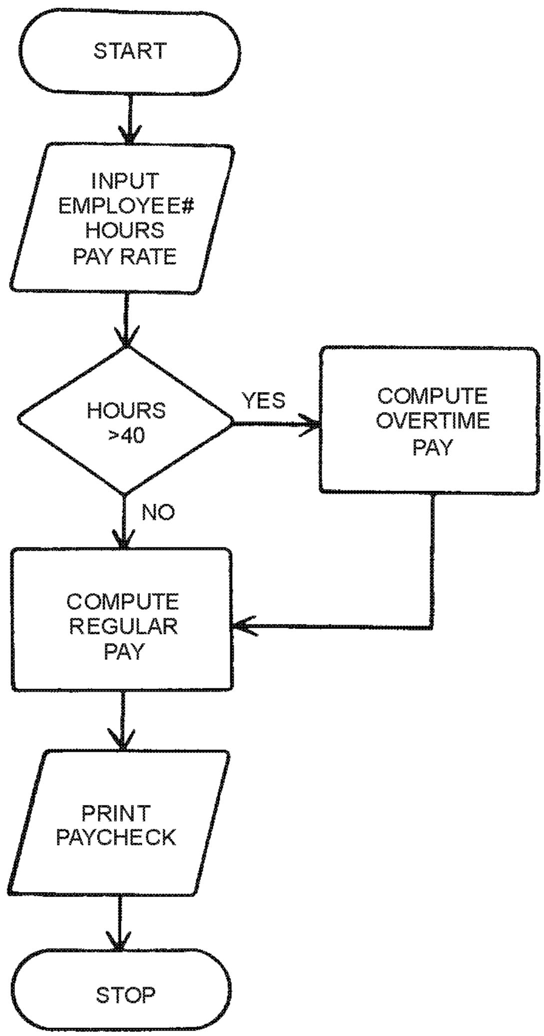


Figure 1-10.—Programming flowchart used to build a payroll program.

WORK CENTER USES (SNAP II)

Every Navy rating has the responsibility for some element of ship's maintenance. And for every rate, recordkeeping has been a "tough nut to turn," an administrative chore that goes along with the work to be done, but takes a "back burner" position to the physical maintenance of the ship and equipment. Today, aboard some ships and soon aboard most, much of that hassle will be done with a SNAP.

The Navy has looked at the paperwork blizzard of recordkeeping responsibility of the essential records and reports that must be generated, and has offered relief to the fleet. This is in the form of S-N-A-P, which stands for Shipboard Non-Tactical ADP Program.

SNAP II is a modern shipboard computer system designed to support shipboard and intermediate-level maintenance, supply, financial, and administrative functions. If this sounds confusing, it really isn't, for the systems are designed to be user-friendly; that is, operating instructions are written in everyday English. Figure 1-11 shows the AN/UYK-62 (V) Data Processing Set. This is the SNAP II computer and its associated hardware.

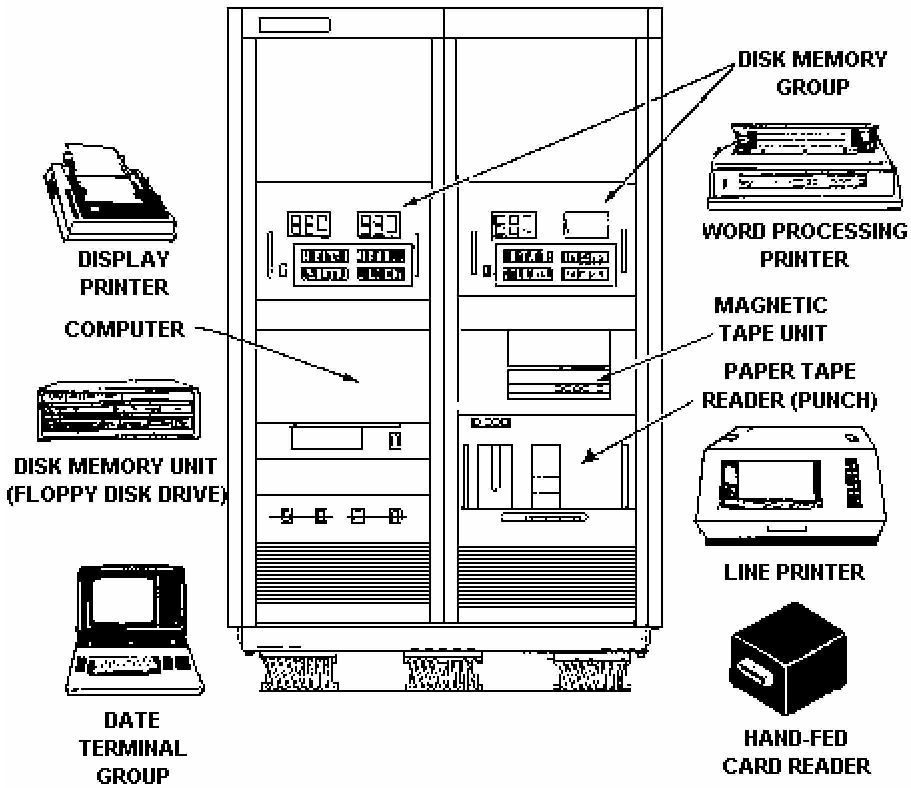


Figure 1-11.—AN/UYK-62 (V) Data Processing Set (SNAP II).

Over the next 3 years, new functions will be added to SNAP to support more of the ship's administrative workload. Pay, personnel, food service, ship's store, PMS, training, medical and dental data are all to be added to SNAP systems.

The SNAP concept is to take the power of the modern computer, the ability to process information, and put that power in the hands of the work center sailors. The sailors can use the system to reduce the labor associated with the paperwork function. User terminals are placed in the different work centers for use by the work center supervisor. Each work center has a different access code. This access code or password prevents unauthorized entry into the main computer's program. Different levels of entry are also defined. The levels depend on a work center's need.

Information stored in the computer for a typical work center normally has the following items that can be updated by the work center supervisor. COSAL (coordinated onboard ship/shore allowance list) is a listing of the repair parts that are allowed to be kept onboard ship, at all times. APL (allowance parts list) is the reference for stock numbers, part numbers, and quantity allowed onboard for a specific system. EIC (equipment identification code) identifies a system, sub-system, or equipment. SHIP'S FORCE WORK LIST is a listing of all work to be performed by a certain work center during a given time period. CSMP (current ship's maintenance projects) provides shipboard maintenance managers with a consolidated listing of deferred maintenance to manage and control its accomplishment. These are but a few of the uses of SNAP II that can be updated by the work center supervisor.

Although the information is usually viewed on a display screen (cathode-ray tube), printed (hard) copies can be obtained. Today, hard copy output from SNAP can be sent to higher authorities in lieu of written reports. In the future, these hard copy transmittals may be replaced by disks or tapes containing the same data. In some cases, the shipboard computers will have an extra telephone wire to the pier or tender, and information can be exchanged electronically.

And there are other important benefits. In practice, the system expedites the storage and retrieval of information the Navy has about its ships. In turn, information that is more accessible means a more timely supply of parts, an improved aid to planners on when and how long to schedule ships' overhauls, and updated information for making decisions whether to place additional or remove unnecessary shipboard equipment. These decisions are now made by laboriously using stacks of printed files. SNAP can sort through these files electronically so Navy planners can make more effective and timely decisions.

SNAP II is a system for unclassified use only at present. This cuts the costs of the installation and many of the physical and electronic security requirements.

Q-40. What is one of the more widespread uses of the computer?

Q-41. What is the great advantage of computers over typewriters?

Q-42. How are word processing programs used by the Navy?

Q-43. How many systems dealing with accounting applications have been widely accepted?

Q-44. What does the acronym S-N-A-P stand for?

Q-45. For what purposes is the SNAP II system designed?

Q-46. What does user friendly mean in computer terms?

Q-47. What does a password prevent?

Q-48. In the SNAP II system, how are the different levels of entry defined?

Q-49. The work center supervisor can update what items from a user terminal?

Q-50. At present what type of classified use is allowed for SNAP II?

USING A DESKTOP COMPUTER

To use a desktop (personal) computer effectively, you'll need to learn about the hardware (the equipment) and the software (the programs). You will also need to know how to handle disks and how to back up programs and data files. So let's assume you have a desktop computer system to use. Its hardware consists of a display screen, a keyboard, a computer, two floppy disk drives (A & B), and a printer. Look at the example in figure 1-12. You need software (computer programs) to make the computer operate. The first program you need is the operating system. The operating system manages the computer and allows you to run application programs like word processing or recordkeeping programs. So let's begin with the operating system.

Figure 1-12.—Typical microcomputer system with display, keyboard, floppy disk drives, and printer.

OPERATING SYSTEM

An operating system is simply a set of programs and routines that lets you and other programs use the computer. A digital computer uses one central set of programs called the operating system to manage execution of other programs and to perform common functions like read, write, or print. Other programs, or you the user, can order the operating system to perform these common functions. These orders are called system calls when other programs use them, or simply commands when you put them through the keyboard.

First, you must load the operating system into the computer so we, and our programs, can use the computer. Remember, in our example, we have a desktop computer with two floppy disk drives, named A and B.

Booting the System

Each desktop computer has a built-in program called "bootstrap loader." When you turn the computer on, this program tries to load, or "boot," an external operating system from disk, usually from drive A, into the computer's internal memory. Disk drive B is usually used for data file disks. The term boot comes from the idea of pulling yourself up by your bootstraps. The computer loads a little program from the disk that tells it how to load a second, bigger program (the operating system). The operating system then tells it how to load another program (an applications program or utility program) to perform a specific job or function. The first thing you need to learn about using a computer is that computers and their programs are very particular. They require complete accuracy and attention to detail on your part. They are not good at guessing what you meant. You'll quickly learn there are a few things that can go wrong at this point, in which case the computer will give you an error message on the display screen similar to this:

Device Error

This means the computer is not reading anything in A drive. Check for:

1. No floppy disk in drive A

2. Floppy disk inserted incorrectly in drive

3. Lock handle on drive A not lowered

Another error message you might receive at this time is:

No System

This means the computer is reading a properly inserted floppy disk, but the disk does not have an operating system on it. Replace the disk with one that does contain the operating system.

Once the operating system is properly booted (loaded), you will see a display similar to this:

A>

You now have what is called a prompt. At this point you can tell the computer what to do next, such as run an application program for example: word processing, accounting, or recordkeeping.

Running an Application Program

To load an application program into the computer from drive A, you put the disk with the application program in disk drive A. Next you type the name of the program following the operating system prompt (A>).

A>**WORDPROC**

This tells the system what program to load and run; in this case Word Processing. The computer then does what the application program tells it. If the application is word processing, the system is ready for you to type a new document, correct an existing document, print a document, and so on. You'll learn more about both the operating system and application programs in chapter 3.

Each application program will have its own set of instructions to follow. In addition to printed documentation, many will include online HELP screens you can display while you are working. These will tell you how to perform a given function or operation.

Another area that needs your constant attention relates to handling floppy disks and making backup copies to be sure your work is not lost.

STORAGE MEDIA HANDLING AND BACKUP

Floppy disks (fig. 1-13) are one means by which you will store data (files that you create) either directly or in backing up the data you store on hard (or fixed) disk. For this reason, and because floppy disks are extremely fragile, you should follow certain guidelines to ensure their proper care and handling. This includes properly labeling and backing up disks.

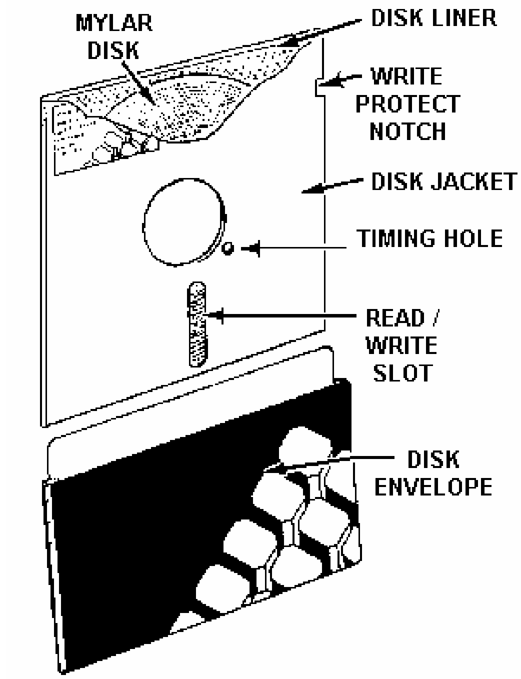


Figure 1-13.—Floppy disk.

Handling

Never touch the exposed surface of a disk. As you know (or will learn), most of the surface of the actual disk is protected most of the time; however, there are areas that are exposed. These areas are the timing hole and the read/write slots. Touching an exposed area can ruin that particular area. If you are familiar with Murphy's Law, you will realize the area you ruin will invariably contain the most important data on that disk.

Storage

Never bend, fold, or otherwise distort the shape of a disk. Never place heavy objects such as books on top of disks. Store disks in the box they came in, or in filing containers that are specifically designed for storing disks. Try to store disks vertically, but if you do store disks horizontally, do not stack more than 10 disks.

Exposure

Disks are subject to exposure from magnetic fields, smoke, heat, and sunlight. X-rays may also have a negative effect.

MAGNETIC FIELDS.—Disks should never be exposed to anything that could be the source of a magnetic field. Exposure of a disk to a magnetic field could cause the destruction of some or all of the data contained on that disk. Some common sources of magnetic energy are crt's, disk drives, and perhaps the most common, the telephone.

SMOKE.—Smoke can cause buildup on disks and on disk drives. DO NOT SMOKE while you work at a terminal or computer.

HEAT AND SUNLIGHT.—Never expose disks to excessive heat or direct sunlight. Either can cause the disks to become warped or distorted so they cannot be used. Disks are made of a plastic material, and if you have ever seen a phonograph record that has been exposed to heat or sunlight, you have some idea of the damage that can result from exposure. Typically, disks will operate only between 10 and 50 degrees Celsius (50 to 120 degrees Fahrenheit). They will accept a relative humidity of 10% to 80%.

X-RAYS.—There is some question about the effect that airport x-ray machines have on disks. It has been the normal experience that the walk-through x-ray machines at airports have no effect on floppy disks; however, this is not to say there will be no effect. It is up to you because these disks contain the data you work with and need. You may not want to take the chance the disks will be affected.

Labeling

When labeling the outside of a floppy disk, write the label before attaching it to the disk. Never use a pencil or ballpoint pen to write on a label once that label has been attached to a disk. When you use an instrument with a sharp point to write on the label, you can actually etch into the surface of the disk underneath the protective sheath, thereby destroying that disk. If you must write on a label once it has been attached to a disk, use a felt-tip marker.

Data Backup

In virtually all computer systems, the possibility exists for errors to occur that accidentally alter or destroy the data stored in the data bases or files. This may occur because of natural disasters, such as fire, flood, or power outages. It may occur through operator error. It may occur through equipment malfunction.

It is essential, therefore, to provide a means to ensure that any data lost can be recovered. The most common method is backup files. A backup file is merely a copy of a file. If for some reason the file or data base is destroyed or becomes unusable, the backup file can be used to recreate the file or data base. Two media are commonly used for backup: disk or tape.

Disk—The most common method of creating a backup for a microcomputer is to use a floppy disk and the diskcopy procedure. This is accomplished by using the original data base or file and copying the information onto a blank floppy disk. The instructions for this procedure will be provided with the particular computer and program you are using.

Tape—Another method of creating a backup is to use magnetic tape. The information contained on your disk, whether it is a data base or file, can be copied onto a tape. The instructions for this procedure will also be provided with the particular computer and program you are using.

Q-51. What is a central set of programs called that manages the execution of other programs and performs common functions like read, write, and print?

Q-52. What is the function of a built-in program called a bootstrap loader?

Q-53. When you see the error message NO SYSTEM, what does it mean?

Q-54. When an operating system prompt (A>) is displayed on the screen, what do you enter from the keyboard to load an application program?

Q-55. If disks are stored horizontally, how many can be stacked?

Q-56. What can exposure to a magnetic field do to the data on a disk?

Q-57. What is the temperature range within which a disk will operate?

Q-58. What is the most common method to ensure that any stored data lost can be recovered?

Q-59. The most common method of creating a backup for a microcomputer is what?

Q-60. Other than disk, what is another media used for backup files?

SUMMARY

This chapter has presented information on the history and classification of computers. It introduced you to electronic digital computers, their uses and operation. The information that follows summarizes the important points of this chapter.

Early computers were **MECHANICAL** or **ELECTROMECHANICAL**. **ELECTRONIC COMPUTERS** came into use in the 1940s.

ANALOG COMPUTERS are special-purpose computers designed to measure continuous electrical or physical conditions.

DIGITAL COMPUTERS are special- or general-purpose computers designed to perform arithmetic and logic functions on separate discrete data. They are generally used for business and scientific data processing.

Digital computers have evolved through four generations: vacuum tubes, transistors, miniaturized circuits, and integrated circuits.

WORD PROCESSING is one of the most widespread uses of desktop computers.

ACCOUNTING AND RECORDKEEPING are also major uses of computers. Included are order entry, inventory control, accounts receivable, accounts payable, general ledger, and payroll.

The Navy's **SHIPBOARD NON-TACTICAL ADP PROGRAM (SNAP)** consists of computers used by work center supervisors for logistic and administrative support. This system expedites the storage and retrieval of information the Navy has about its ships.

A **DESKTOP (PERSONAL) COMPUTER** is a microcomputer with at least a display screen, keyboard, floppy disk drive, and printer. It may also have additional devices such as a second floppy or a hard disk drive.

An **OPERATING SYSTEM** is loaded into the computer to let you and other programs use the computer. It also provides common functions like read, write, and print.

You can direct the computer to run an **APPLICATION PROGRAM** by telling the operating system the name of program to run. Common application programs are word processing, accounting, and recordkeeping.

You will probably be using **FLOPPY DISKS** for data storage and backup. To ensure you don't damage a disk, use care in handling, labeling and storing the disks.

ANSWERS TO QUESTIONS Q1. THROUGH Q60.

A-1. *Technology (mechanical, electromechanical, electronic), purpose (special or general), type of data they handle (analog or digital), cost, physical size (handheld to room size).*

A-2. *Analog.*

A-3. *Gun fire control.*

A-4. *Electromechanical computers use electrical components to perform some of the calculations.*

A-5. *Integrated circuits.*

A-6. *Special-purpose.*

A-7. *Its design.*

A-8. *Lack of versatility.*

A-9. *To perform a wide variety of functions and operations.*

A-10. *By storing different programs in its internal storage.*

A-11. *Speed and efficiency.*

A-12. *Special-purpose.*

A-13. *Continuous electrical or physical conditions.*

A-14. *Mechanical or electromechanical.*

A-15. *Hybrid computers.*

A-16. *Business and scientific data processing.*

A-17. *Digital computers deal with discrete quantities, while analog computers deal with continuous physical variables.*

A-18. *By the accuracy with which physical quantities can be sensed and displayed.*

A-19. *Third.*

A-20. *The number of significant figures carried in the computations.*

A-21. *Design of the computer processing unit.*

A-22. *General-purpose digital computer.*

A-23. *Generations.*

A-24. *Significant change in computer design.*

A-25. *The vacuum tube.*

A-26. *They were unreliable, required a lot of power to run, and produced so much heat that air conditioning was needed to protect computer parts.*

A-27. Thousandths of a second (millisecond).

A-28. Unsophisticated and machine oriented.

A-29. By the use of small, long lasting transistors.

A-30. With the introduction of magnetic disk storage and the use of core for main storage.

A-31. Symbolic machine languages or assembly languages.

A-32. Faster internal processing speeds.

A-33. Faster execution of instructions.

A-34. Over 100 million.

A-35. Both scientific and business data processing applications.

A-36. Microcomputers and minicomputers.

A-37. Read-only memory.

A-38. How to properly and effectively use the computing power available.

A-39. Software.

A-40. Word processing.

A-41. Correcting errors.

A-42. For manuscript writing, memorandum writing, identification-card application filing, and recordkeeping.

A-43. Six.

A-44. Shipboard Non-tactical ADP Program.

A-45. To support shipboard and intermediate level maintenance, supply, financial, and administrative functions.

A-46. Operating instructions are written in everyday English.

A-47. Unauthorized entry into the main computer's program.

A-48. Dependent on a work center's need.

A-49. COSAL, APL, EIC, SHIP'S FORCE WORK LIST, and CSMP.

A-50. Unclassified.

A-51. Operating system.

A-52. To load an external operating system into the computer's internal memory.

A-53. The computer is reading a properly inserted floppy disk, but it does not have an operating system on it.

A-54. The program name.

A-55. No more than ten.

A-56. Destroy some or all of it.

A-57. 10 to 50 degrees Celsius or 50 to 120 degrees Fahrenheit.

A-58. Backup files.

A-59. Use a floppy disk and the diskcopy procedure.

A-60. Magnetic tape.

CHAPTER 2

HARDWARE

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to do the following:

1. Explain the cpu and describe the functions of the different sections.
2. Categorize the types of storage and their functions.
3. Describe how storage is classified.
4. Analyze and compare the input/output devices and explain their functions.

INTRODUCTION

Components or tools of a computer system are grouped into one of two categories, hardware or software. We refer to the machines that compose a computer system as hardware. This hardware includes all the mechanical, electrical, electronic, and magnetic devices within the computer itself (the central processing unit) and all related peripheral devices (printers, magnetic tape units, magnetic disk drive units, and so on). These devices will be covered in this chapter to show you how they function and how they relate to one another. Take a few minutes to study figure 2-1. It shows the functional units of a computer system: the inputs, the central processing unit (cpu), and the outputs. The inputs can be on any storage medium from punched cards, paper tape, or magnetic ink to magnetic tape, disk, or drum; or they can be entries from a console keyboard or a cathode-ray tube (crt) terminal. The data from one or more of these inputs will be processed by the central processing unit to produce output. The output may be in punched cards or paper tape, on magnetic tape, disk, or drum, or it may be printed reports or information displayed on a console typewriter or crt terminal. The figure also shows the data flow, instruction flow, and flow of control. We'll start our hardware discussion with the cpu and then move into storage media (disk, tape, and drum). We'll end the chapter with a discussion of input/output devices and how they work.

CENTRAL PROCESSING UNIT (CPU)

The brain of a computer system is the central processing unit, which we generally refer to as the cpu or mainframe. The central processing unit IS THE COMPUTER. It is the cpu that processes the data transferred to it from one of the various input devices, and then transfers either the intermediate or final results of the processing to one of many output devices. A central control section and work areas are required to perform calculations or manipulate data. The cpu is the computing center of the system. It consists of a control section, internal storage section (main or primary memory), and arithmetic-logic section (fig. 2-1). Each of the sections within the cpu serves a specific function and has a particular relationship to the other sections within the cpu.

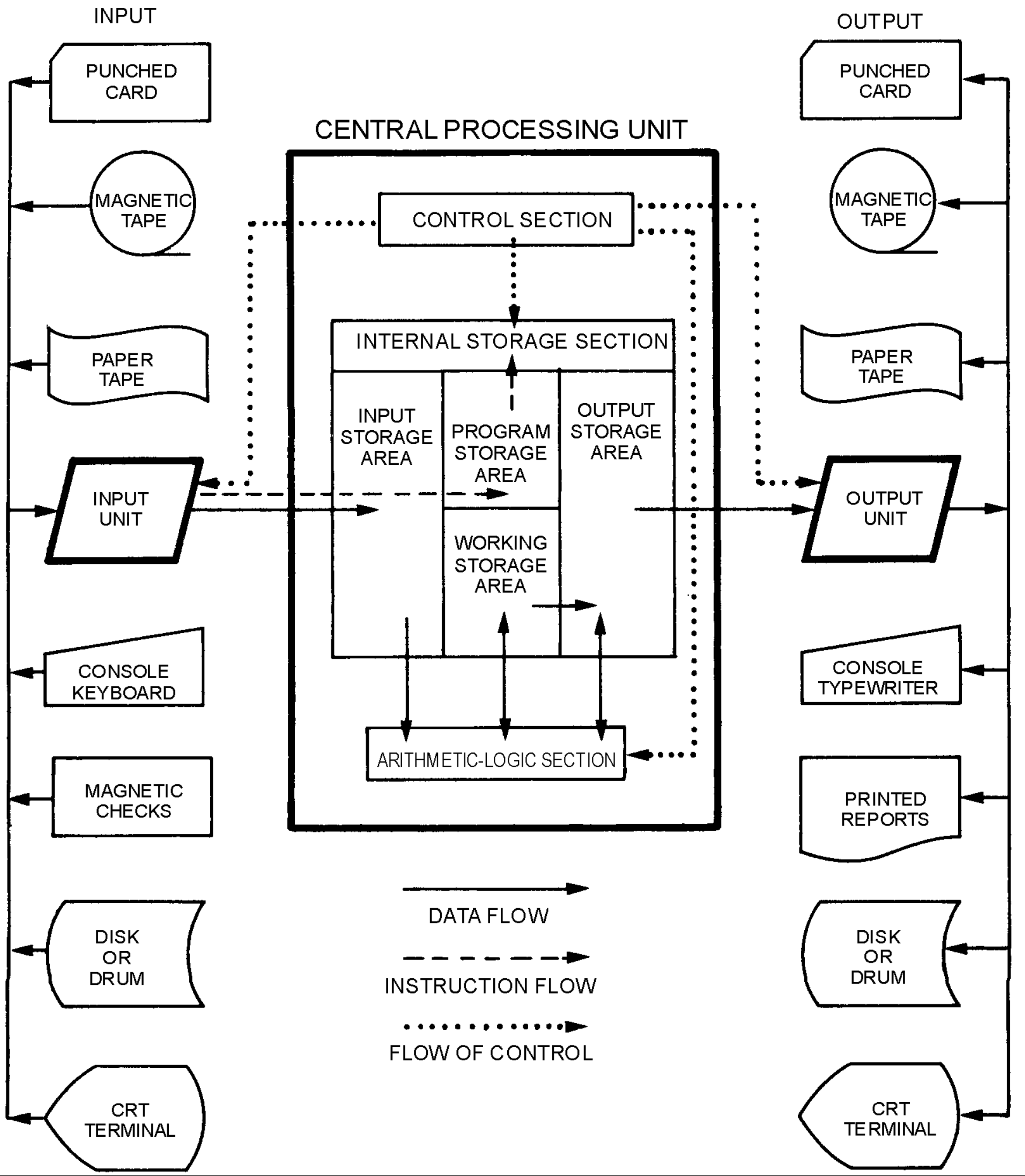


Figure 2-1.—Functional units of a computer system.

CONTROL SECTION

The control section may be compared to a telephone exchange because it uses the instructions contained in the program in much the same manner as the telephone exchange uses telephone numbers. When a telephone number is dialed, it causes the telephone exchange to energize certain switches and control lines to connect the dialing phone with the phone having the number dialed. In a similar manner, each programmed instruction, when executed, causes the control section to energize certain control lines, enabling the computer to perform the function or operation indicated by the instruction.

The program may be stored in the internal circuits of the computer (computer memory), or it may be read instruction-by-instruction from external media. The internally stored program type of computer, generally referred to only as a stored-program computer, is the most practical type to use when speed and fully automatic operation are desired.

Computer programs may be so complex that the number of instructions plus the parameters necessary for program execution will exceed the memory capacity of a stored-program computer. When this occurs, the program may be sectionalized; that is, broken down into modules. One or more modules are then stored in computer memory and the rest in an easily accessible auxiliary memory. Then as each module is executed producing the desired results, it is swapped out of internal memory and the next succeeding module read in.

In addition to the commands that tell the computer what to do, the control unit also dictates how and when each specific operation is to be performed. It is also active in initiating circuits that locate any information stored within the computer or in an auxiliary storage device and in moving this information to the point where the actual manipulation or modification is to be accomplished.

The four major types of instructions are (1) transfer, (2) arithmetic, (3) logic, and (4) control. Transfer instructions are those whose basic function is to transfer (move) data from one location to another. Arithmetic instructions are those that combine two pieces of data to form a single piece of data using one of the arithmetic operations.

Logic instructions transform the digital computer into a system that is more than a high-speed adding machine. Using logic instructions, the programmer may construct a program with any number of alternate sequences. For example, through the use of logic instructions, a computer being used for maintenance inventory will have one sequence to follow if the number of a given item on hand is greater than the order amount and another sequence if it is smaller. The choice of which sequence to use will be made by the control section under the influence of the logic instruction. Logic instructions, thereby, provide the computer with the ability to make decisions based on the results of previously generated data. That is, the logic instructions permit the computer to select the proper program sequence to be executed from among the alternatives provided by the programmer.

Control instructions are used to send commands to devices not under direct command of the control section, such as input/output units or devices.

ARITHMETIC-LOGIC SECTION

The arithmetic-logic section performs all arithmetic operations-adding, subtracting, multiplying, and dividing. Through its logic capability, it tests various conditions encountered during processing and takes action based on the result. As indicated by the solid arrows in figure 2-1, data flows between the arithmetic-logic section and the internal storage section during processing. Specifically, data is transferred as needed from the internal storage section to the arithmetic-logic section, processed, and returned to the internal storage section. At no time does processing take place in the storage section. Data may be transferred back and forth between these two sections several times before processing is completed. The results are then transferred from internal storage to an output unit, as indicated by the solid arrow (fig. 2-1).

MEMORY (INTERNAL STORAGE) SECTION

All memory (internal storage) sections must contain facilities to store computer data or instructions (that are intelligible to the computer) until these instructions or data are needed in the performance of the computer calculations. Before the stored-program computer can begin to process input data, it is first necessary to store in its memory a sequence of instructions, and tables of constants and other data it will use in its computations. The process by which these instructions and data are read into the computer is called loading.

Actually, the first step in loading instructions and data into a computer is to manually place enough instructions into memory using the keyboard or electronically using an operating system (discussed in chapter 1), so that these instructions can be used to bring in more instructions as desired. In this manner a few instructions are used to bootstrap more instructions. Some computers make use of an auxiliary (wired) memory that permanently stores the bootstrap program, thereby making manual loading unnecessary.

The memory (internal storage) section of a computer is essentially an electronically operated file cabinet. It has a large number (usually several hundred thousand) of storage locations; each referred to as a storage address or register. Every item of data and program instruction read into the computer during the loading process is stored or filed in a specific storage address and is almost instantly accessible.

Q-1. What is the brain of a computer system?

Q-2. How many sections make up the central processing unit?

Q-3. What are the names of the sections that make up the cpu?

Q-4. The control section can be compared to what?

Q-5. What are the four major types of instructions in the control section?

Q-6. What capability allows the arithmetic/logic section to test various conditions encountered during processing and take action based on the result?

Q-7. In the arithmetic/logic section, data is returned to what section after processing?

Q-8. What is the process by which instructions and data are read into a computer?

TYPES OF INTERNAL STORAGE

You already know that the internal storage section is the holding area in which instructions and data are kept. For the control section to control and coordinate all processing activity, it must be able to locate each instruction and data item in storage. About now, you are probably wondering how the control section is able to find these instructions and data items. To understand this, let's look at storage as nothing more than a collection of mailboxes. Each mailbox has a unique address and represents a location in memory as shown in figure 2-2. Like the mail in your mailbox, the contents of a storage location can change, but the number on your mailbox or memory address always remains the same. In this manner, a particular program instruction or data item that is held in storage can be located by knowing its address. Some computers can address each character of data in memory directly. Others address computer words which contain a group of characters at a single address. Each computer word contains a group of characters at a single address. Some of the more common types of internal storage media used in today's computers are as follows: magnetic core, semiconductor, and bubble.

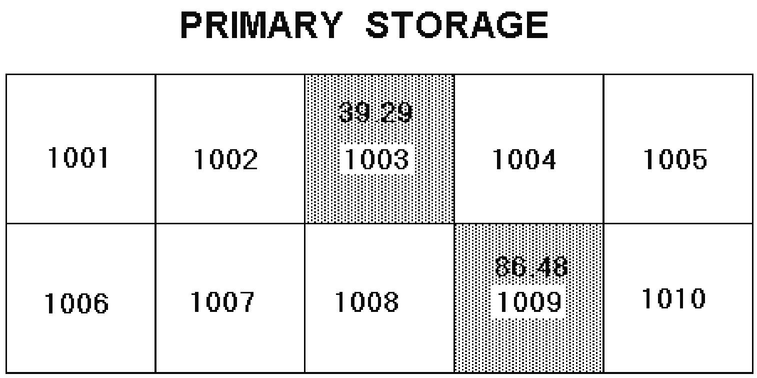


Figure 2-2.—Memory locations.

MAGNETIC CORE STORAGE

Although magnetic core storage is no longer as popular as it once was, we will cover it in some detail because its concepts are easily understood and apply generally to the more integrated semiconductor and bubble-type memories. Magnetic core storage is made up of tiny doughnut-shaped rings made of ferrite (iron), that are strung on a grid of very thin wires (fig. 2-3). Since data in computers is stored in binary form (refer to *NEETS*, module 13), a two-state device is needed to represent the two binary digits (bits), 0 for off and 1 for on. In core storage, each ferrite ring can represent a 0 or 1 bit, depending on its magnetic state. If magnetized in one direction, it represents a 1 bit, and if magnetized in the opposite direction, it represents a 0 bit. These cores are magnetized by sending an electric current through the wires on which the core is strung. It is this direction of current that determines the state of each core.

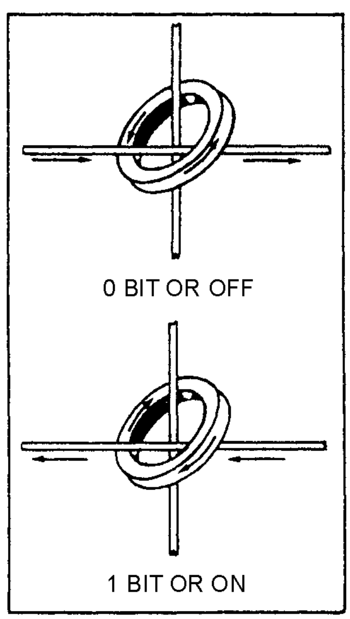


Figure 2-3.—Two-state principle of magnetic storage.

SEMICONDUCTOR STORAGE (THE SILICON CHIP)

Semiconductor memory consists of hundreds of thousands of tiny electronic circuits etched on a silicon chip (fig. 2-4). Each of these electronic circuits is called a bit cell and can be in either an off or on state to represent a 0 or 1 bit, depending on whether or not current is flowing in that cell. Another name you will hear used for semiconductor memory chips is integrated circuits (ICs). Developments in technology have led to large scale integration (LSI), which means that more and more circuits can be squeezed onto the same silicon chip. Companies are even manufacturing very large scale integrated circuits (VLSI), which means even further miniaturization.

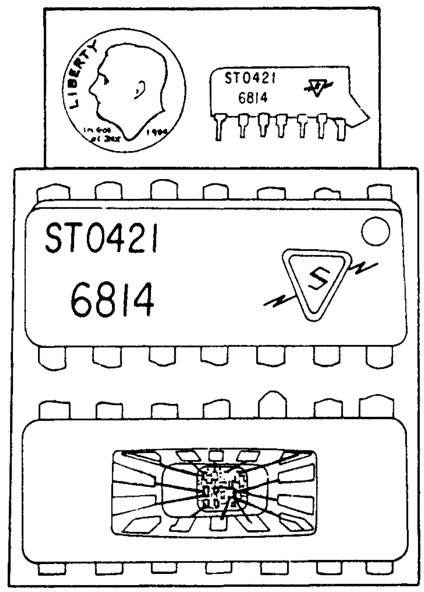


Figure 2-4.—A semiconductor memory chip (integrated circuit).

Some of the advantages of semiconductor storage are fast internal processing speeds, high reliability, low power consumption, high density (many circuits), and low cost. However, there is a drawback to this type of storage. It is volatile, which means all data in memory is lost when the power supply is removed. Should the power on your computer fail and you have no backup power supply, all the stored data is lost. This is not the case with magnetic core storage. Core storage is nonvolatile. This means the data is retained even if there is a power failure or breakdown, since the cores store data in the form of magnetic charges rather than electric current.

BUBBLE STORAGE

One of the latest technological developments in storage media is the introduction of bubble memory. Bubble memory consists of a very thin crystal made of semiconductor material. The molecules of this special crystal act as tiny magnets (fig. 2-5). The polarity of these molecules or "magnetic domains" can be switched in an opposite direction by passing a current through a control circuit imprinted on top of the crystal. In this manner, data can be stored by changing the polarity of the magnetic domains. Since the principle is the same as for magnetic core storage, bubble memory is considered nonvolatile. The data is retained even if there is a power failure. Furthermore, the process of reading from bubble memory is nondestructive, meaning that the data is still present after being read. This is not the case with core storage, which must be regenerated after being read. If we were to view these magnetic domains under a microscope, they would look like tiny bubbles; hence the name, bubble memory.

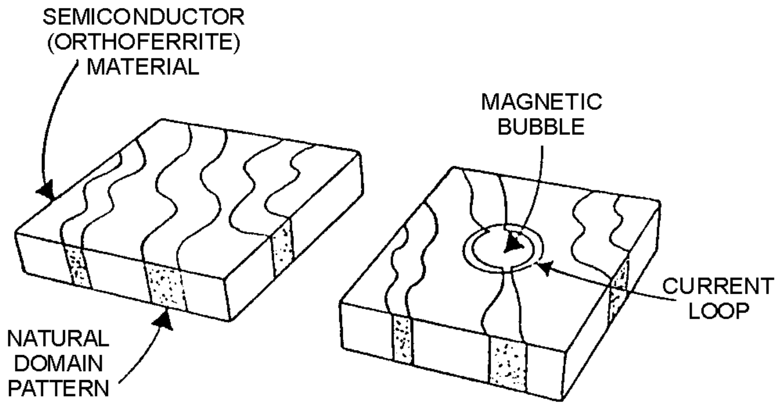


Figure 2-5.—Bubble memory.

Q-9. Magnetic core storage is made up of what?

Q-10. A semiconductor memory consists of what?

Q-11. What is another name for semiconductor memory chips?

Q-12. In computer storage, what does volatile mean?

Q-13. What type of storage can retain its data even if there is a power failure or breakdown?

Q-14. Bubble memory consists of what?

Q-15. How are the magnetic domains of a bubble memory switched?

Q-16. What do we mean when we say that reading from bubble memory is nondestructive?

CLASSIFICATIONS OF INTERNAL STORAGE

Up to this point, you have learned some of the general functions of the cpu, the physical characteristics of memory, and how data is stored in the internal storage section. Now, we will explain yet another way to classify internal (primary or main) storage. This is by the different kinds of memories used within the cpu: read-only memory, random-access memory, programmable read-only memory, and erasable programmable read-only memory.

READ-ONLY MEMORY (ROM)

In most computers, it is useful to have often used instructions, such as those used to bootstrap (initial system load) the computer or other specialized programs, permanently stored inside the computer. Memory that enables us to do this without the programs and data being lost (even when the computer is powered down) is called read-only memory. Only the computer manufacturer can provide these programs in ROM and once done, they cannot be changed. Consequently, you cannot put any of your own data or programs in ROM. Many complex functions such as routines to extract square roots, translators for programming languages, and operating systems can be placed in ROM memory. Since these instructions are hard wired (permanent), they can be performed quickly and accurately. Another advantage of ROM is that your computer facility can order programs tailored for its needs and have them permanently installed in ROM by the manufacturer. Such programs are called microprograms or firmware.

RANDOM-ACCESS MEMORY (RAM)

Another kind of memory used inside computers is called random-access memory (RAM) or read/write memory. RAM memory is rather like a blackboard on which you can scribble down notes, read them, and rub them out when you are finished with them. In the computer, RAM is the working memory. Data can be read (retrieved) from or written (stored) into RAM just by giving the computer the address of the location where the data is stored or is to be stored. When the data is no longer needed, you can simply write over it. This allows you to use the storage again for something else. Core, semiconductor, and bubble storage all have random access capabilities.

PROGRAMMABLE READ-ONLY MEMORY (PROM)

An alternative to ROM is programmable read only memory (PROM) that can be purchased already programmed by the manufacturer or in a blank state. By using a blank PROM, you can enter any program into the memory. However, once the PROM has been written into, it can never be altered or changed. Thus you have the advantage of ROM with the additional flexibility to program the memory to meet a unique need. The main disadvantage of PROM is that if a mistake is made and entered into PROM, it cannot be corrected or erased. Also, a special device is needed to "burn" the program into PROM.

ERASABLE PROGRAMMABLE READ-ONLY MEMORY (EPROM)

The erasable programmable read-only memory (EPROM) was developed to overcome the drawback of PROM. EPROMs can also be purchased blank from the manufacturer and programmed locally at your command/activity. Again, this requires special equipment. The big difference with EPROM is that it can be erased if and when the need arises. Data and programs can be retrieved over and over again without destroying the contents of the EPROM. They will stay there quite safely until you want to reprogram it by first erasing the EPROM with a burst of ultra-violet light. This is to your advantage, because if a mistake is made while programming the EPROM, it is not considered fatal. The EPROM can be erased and corrected. Also, it allows you the flexibility to change programs to include improvements or modifications in the future.

Q-17. In what type of memory are often used instructions and programs permanently stored inside the computer?

Q-18. Who provides the programs stored in ROM?

Q-19. Can programs in ROM be changed?

Q-20. What is another name for random-access memory (RAM)?

Q-21. How is data read from or written into RAM?

Q-22. In what two states can programmable read-only memory (PROM) be purchased?

Q-23. What is the main disadvantage of PROM?

Q-24. What does EPROM stand for?

Q-25. How is EPROM erased?

SECONDARY STORAGE

The last kind of memory we will briefly introduce here is called secondary storage or auxiliary storage. This is memory outside the main body of the computer (cpu) where we store programs and data for future use. When the computer is ready to use these programs and data, they are read into internal storage. Secondary (auxiliary) storage media extends the storage capabilities of the computer system. We need it for two reasons. First, because the computer's internal storage is limited in size, it cannot always hold all the data we need. Second, in secondary storage, data and programs do not disappear when power is turned off. Secondary storage is nonvolatile. This means information is lost only if you, the user, intentionally erase it. The three types of secondary storage we most commonly use are magnetic disk, tape, and drum.

MAGNETIC DISK

The popularity of disk storage devices is largely because of their direct-access capabilities. Most every system (micro, mini, and mainframe) will have disk capability. Magnetic disks resemble phonograph records (round platters), coated with a magnetizable recording material (iron oxide), but their similarities end there. Magnetic disks come in many different sizes and storage capacities. They range from 3 inches to 4 feet in diameter and can store from 2.5 million to 600 million characters (bytes) of data. They can be portable in that they are removable, or they can be permanently mounted in the storage devices called disk drive units or disk drives. They can be made of rigid metal (hard disks) or flexible plastic (floppy disks or diskettes) as shown in figure 2-6.

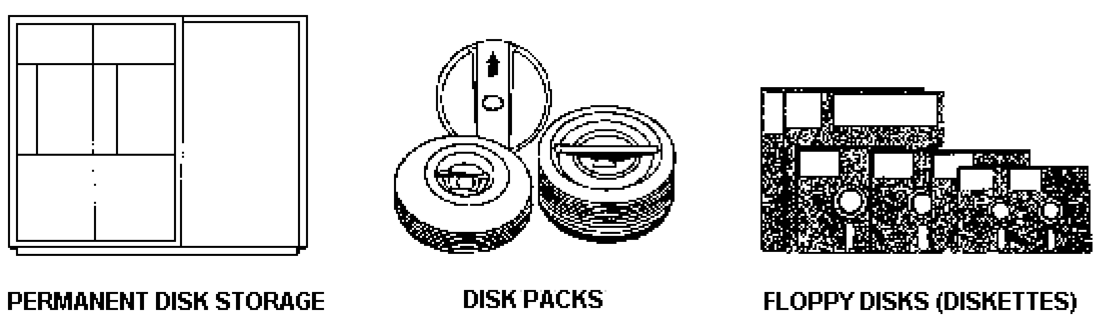


Figure 2-6.—Various types and sizes of magnetic disk storage.

Music is stored on a phonograph record in a continuous groove that spirals into the center of the record. But there are no grooves on a magnetic disk. Instead, data is stored on all disks in a number of invisible concentric circles called tracks. Each track has a designated number beginning with track 000 at the outer edge of the disk. The numbering continues sequentially toward the center to track 199, 800, or whatever the highest track number is. No track ever touches another (fig. 2-7). The number of tracks can vary from 35 to 77 on a floppy disk surface and from 200 to over 800 on hard disk surfaces.

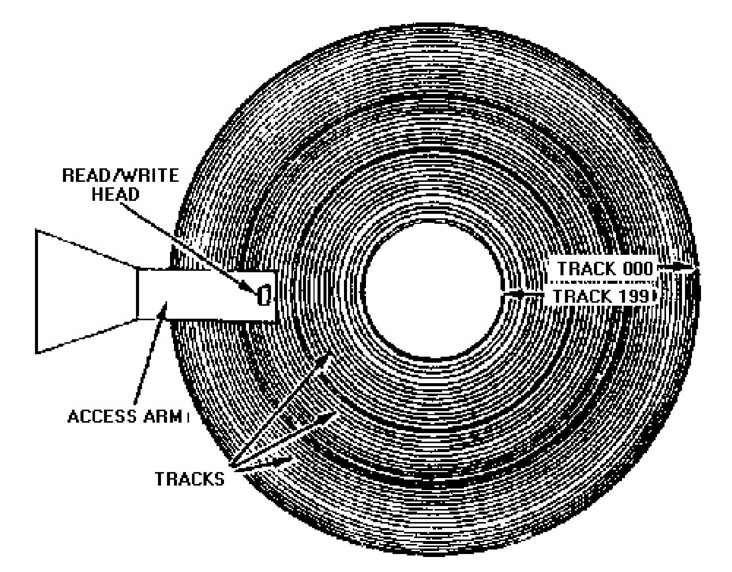


Figure 2-7.—Location of tracks on the disk's recording surface.

Data is written as tiny magnetic bits (or spots) on the disk surface. Eight-bit codes are generally used to represent data. Each code represents a different number, letter, or special character. In chapter 4, you'll learn how the codes are formed. When data is read from the disk, the data on the disk remains unchanged. When data is written on the disk, it replaces any data previously stored on the same area of the disk.

Characters are stored on a single track as strings of magnetized bits (0's and 1's) as shown in figure 2-8. The 1 bits indicate magnetized spots or ON bits. The 0 bits represent unmagnetized portions of the track or OFF bits. Although the tracks get smaller as they get closer to the center of the disk platter, each track can hold the same amount of data because the data density is greater on tracks near the center.

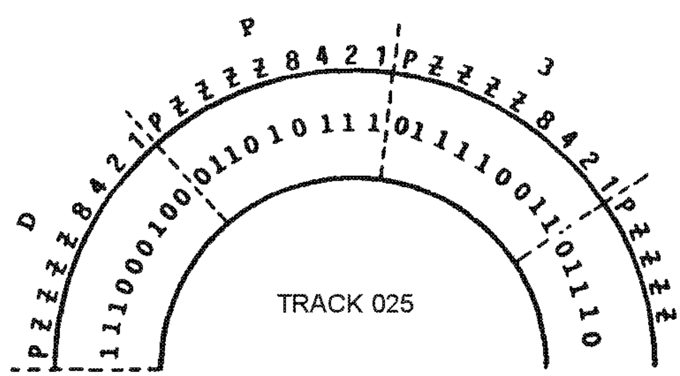


Figure 2-8.—A string of bits written to disk on a single track.

A track can hold one or more records. A record is a set of related data treated as a unit. The records on a track are separated by gaps in which no data is recorded, and each of the records is preceded by a disk address. This address indicates the unique position of the record on the track and is used to directly access the record. Figure 2-9 shows a track on which five records have been recorded. Because of the gaps and addresses, the amount of data we can store on a track is reduced as the number of records per track is increased. Records on disk can be blocked (grouped together). Only one disk address is needed per block, and as a result, fewer gaps occur. We can use the blocking technique to increase the amount of data we can store on one track.

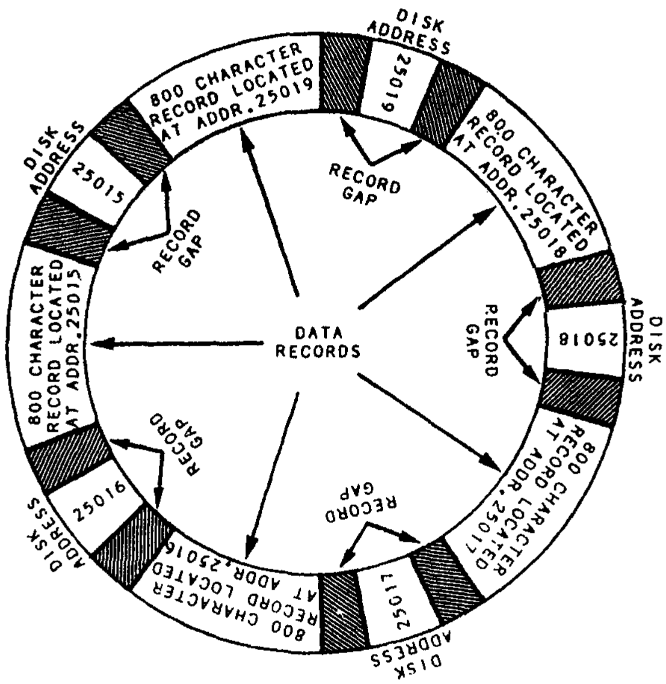


Figure 2-9.—Data records as they are written to disk on a single track.

The storage capacity of a disk depends on the bits per inch of track and the tracks per inch of surface. Using Winchester technology, the designers of disk drive units were able to increase the data density of a disk by increasing the number of tracks. Winchester was the code name used by IBM during the development of this technology. The designers originally planned to use dual disk drives to introduce the new concept. Each drive was to have a storage capacity of 30 million characters, and thus was expected to be a "30-30." Since that was the caliber of a famous rifle, the new product was nicknamed "Winchester." The designers found that data density could be improved and storage capacity increased by reducing the flying height, the distance of the read/write heads over the disk surfaces when reading and writing. By doing this, smaller magnetized spots could be precisely written and then read. The read/write heads were moved so close to the disk that a human hair looked like a mountain in the path of the flying head. Winchester technology reduces this potential problem by sealing the disks in a contamination-free container. This eliminates foreign objects from coming in contact with the read/write heads.

Data can be physically organized in one of two ways on a disk pack, depending on the manufacturer and the model of disk drive you are using. One way uses the cylinder method, and the other uses the sector method. On diskettes, data is organized using the sector method.

The cylinder method uses a cylinder as the basic reference point. When you look at figure 2-10, view A, you will see a disk pack containing six disk platters with 10 recording surfaces. Imagine you are looking down through the disk pack from the top. All the tracks with the same number line up vertically. Together they are called a cylinder. These 10 tracks, one on each recording surface, can be referenced by the 10 read/write heads on the five access arms at each discrete location where the access arms can be positioned.

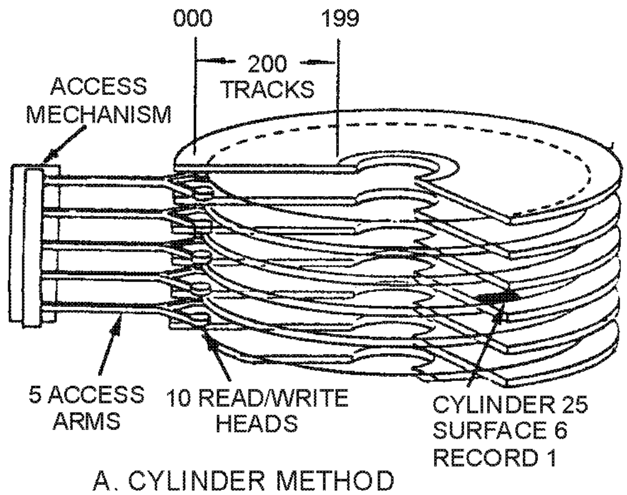


Figure 2-10A.—Physical organization of data on a disk. CYLINDER METHOD.

Therefore, to physically reference a record stored using the cylinder method, a computer program must specify the cylinder number, the recording surface number, and the record number as shown in figure 2-10, view A. Here, the record is stored in cylinder 25 of recording surface 6 and is the first record on that track. Special data stored on each track specifies the beginning of the track so that the first record, second record, third record, and so on, can be identified.

Another way to physically organize data on the disk pack (and on diskettes) is to use the sector method. This requires that each of the tracks be divided into individual storage areas called sectors (shown in figure 2-10, view B). The number of sectors varies with the disk system used; however, there are usually eight or more. Each sector holds a specific number of characters. Before a record can be accessed, a computer program must again give the disk drive the record's address specifying the track number, the surface number, and the sector number of the record. One or more read/write heads are then moved to the proper track, the head over the specified surface is activated, and the data is read from or written to the designated sector as it spins under the head.

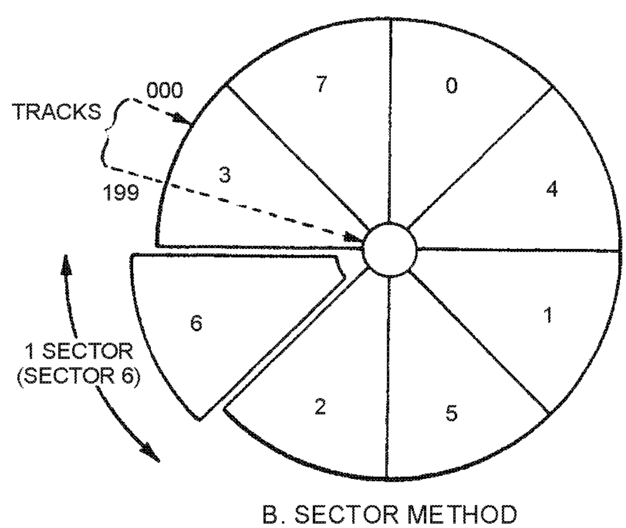


Figure 2-10B.—Physical organization of data on a disk. SECTOR METHOD.

MAGNETIC TAPE

Another type of storage device is magnetic tape which is similar to the tape used with commercial tape recorders. It is used mainly for secondary storage. It differs from commercial tape in that it is usually wider (ranging from one-half inch to an inch), and it is manufactured to more rigid quality specifications. It is made of a MYLAR® base coated with a magnetic oxide that can be magnetized to store data. Magnetic tape comes in a variety of lengths (from 600 to 3,000 feet), and is packaged in one of three ways: open reel, cartridge, or cassette, as shown in figure 2-11. Large computers use standard open reels, 1/2-inch wide tape, 2,400 feet in length. Magnetic tape units are categorized by the type of packaging used for the tape. The tape unit (or drive) shown in figure 2-12 uses open reels, while cartridge tape units use tape cartridges and cassette units use tape cassettes. Cartridge tape units are often used on personal computers to provide backup for hard disk.

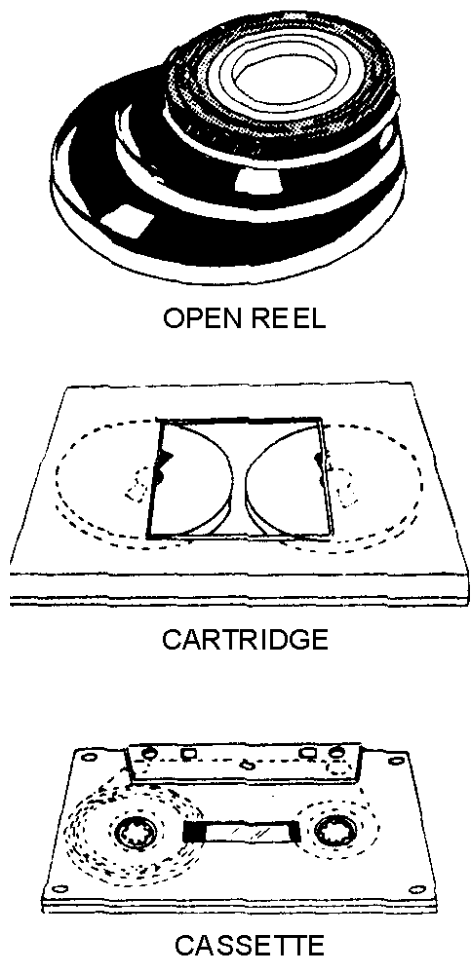


Figure 2-11.—Various types of magnetic tape storage.

Figure 2-12.—Mounting a magnetic tape

A standard 1/2-inch tape may have either seven (fig. 2-13, view A) or nine tracks (fig. 2-13, view B) of data stored on it, depending upon the particular read/write heads installed in the tape unit. Read/write heads are usually designed to read (or write) data (in the form of bits) concurrently across the width of the tape.

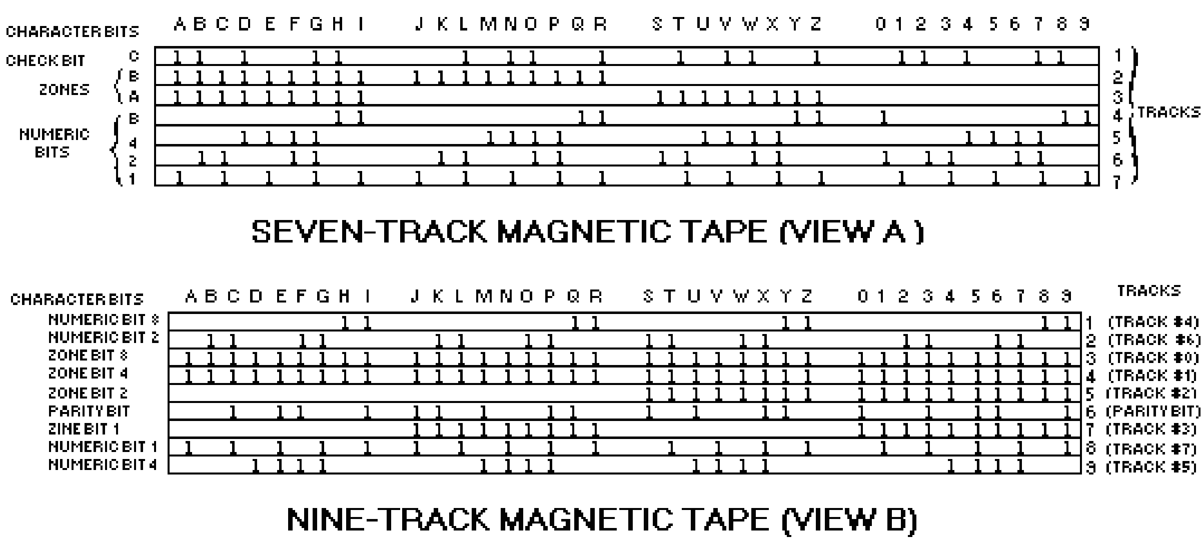


Figure 2-13.—Multi-track magnetic tape.

The amount of data or the number of binary digits (0 and 1 bits) that can be written (stored) on a linear inch of tape is known as the tape's recording density. Common recording densities for multitrack tapes range from 200 to 6,250 bits/bytes per inch (BPI). Also note that sometimes the density of a tape is referred to as the number of frames per inch (FPI) or characters per inch (CPI) rather than BPI. Regardless of which term is used, a frame or byte is a group of related bits that make up a single character written across the width of the tape. Most magnetic tape units are capable of reading and writing in several different densities.

Magnetic tapes have many common features and data recording formats. Each tape is physically marked in some manner to indicate where reading and writing on tape is to begin (known as the beginning-of-tape [BOT]), and where it ends (known as the end-of-tape [EOT]). The length of tape between the BOT and EOT is referred to as the usable recording (reading/writing) surface or usable storage area. BOT/EOT markers are usually made of short silver strips of reflective tape (1/4-inch wide by 1/2-inch long) as shown in figure 2-14. The BOT marker is normally placed toward the front edge of the tape (the side nearest you when the tape is mounted on the tape unit). The EOT marker is placed toward the back edge (the side farthest from you when the tape is mounted on the tape unit). They are placed approximately 15 to 20 feet in from each end on the shiny side of the tape. Sometimes, holes or clear plastic inserts are used as markers in place of reflective strips. Regardless of the method used, the BOT/EOT markers are sensed by an arrangement of lamps and/or photodiode sensors to indicate where reading and writing is to begin and end.

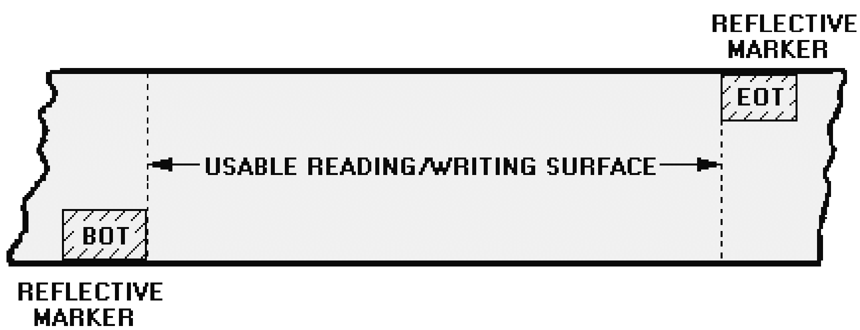


Figure 2-14.—Beginning-of-tape (BOT) and end-of-tape (EOT) markers.

We can make records on magnetic tape any size we need to hold the data. We are restricted only by the length of the tape or the capacity of internal storage. For example, a record can be one character, several characters, or thousands of characters in length. The collection of records is called a file. A file containing payroll records is called a payroll file; a file containing supply inventory records is called a supply inventory file.

Records can be placed on tape either separately as single records (unblocked) as shown in figure 2-15, view A, or multiple records can be grouped together (blocked) as shown in figure 2-15, view B, to form a record block. The number of records stored in a record block is the blocking factor. In this example, the blocking factor is five.

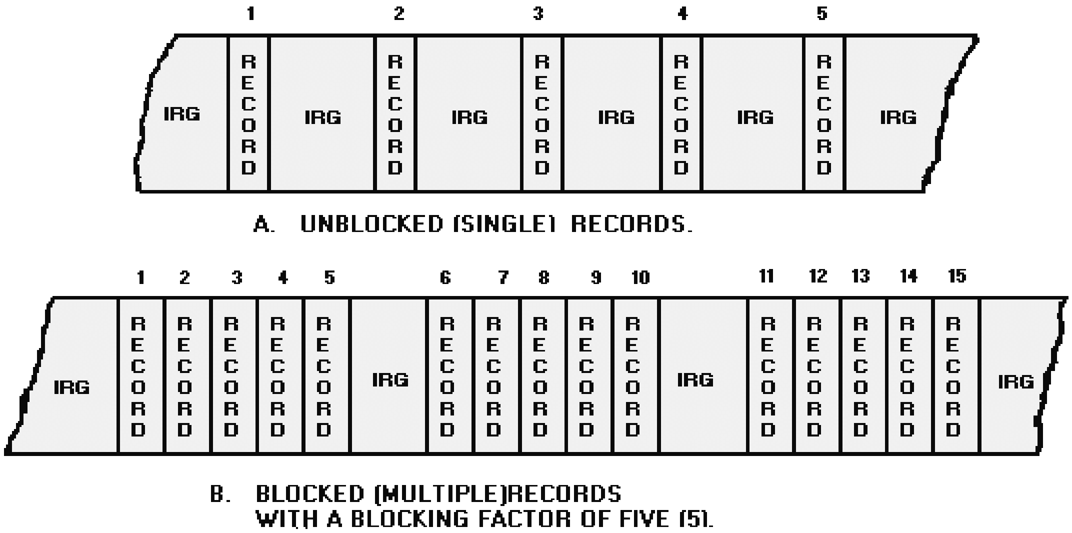


Figure 2-15.—Record formats on magnetic tape.

All magnetic tape must be moving at a predetermined speed for data to be read from or written on the tape. Data cannot be read or written while the tape is coming up to speed, slowing down, or stopped. During this time delay, the tape moves a short distance creating a blank spot on the tape. This interrecord gap or interblock gap separates each single record or block of records on the tape. The length of the gap varies, depending upon the particular system and method of recording, but is approximately 2/5 to 3/4 inch in length. If single records are stored on the tape, the interrecord gap may be longer than the portion of tape used to store the record. Therefore, much of the tape's recording surface is wasted.

To overcome the inefficiency of storing single data records, we normally block records. In figure 2-15, view B, you will notice the tape is used more efficiently than the tape in figure 2-15, view A. Blocking allows more data to be stored on a reel of tape.

During reading, the record begins with the first character sensed following an interrecord or interblock gap and continues until the next gap is reached. All input records read are internally stored in accordance with the amount of storage area set aside by the applications program.

Magnetic tape, as a storage media, offers several useful features. We can store large amounts of data in a variety of convenient package sizes (open reels, cartridges, or cassettes). Magnetic tapes are easily interchangeable between similar tape units of different computer systems, and tapes are less prone to damage than other types of storage media.

MAGNETIC DRUM

Like the magnetic disk, the magnetic drum is another example of a direct-access storage device. Although the magnetic drum was once used as main (or primary) storage, it is now used as secondary (or auxiliary) storage. Unlike some disk packs, the magnetic drum cannot be physically removed. The drum is permanently mounted in the device.

Magnetic drum storage devices consist of either a hollow cylinder (thus, the name drum) or a solid cylinder that rotates at a constant velocity (from 600 to 6,000 rpm). The outer surface is coated with an iron-oxide material capable of being magnetized.

A magnetic drum differs from a magnetic disk in that the tracks in which the data is stored are assigned to channels located around the circumference of the drum as shown in figure 2-16. That is, the channels form circular bands around the drum. The coded representation of data in figure 2-16 is similar to that used on 9-track magnetic tape, 8-bit code. The basic functions of the read/write heads are to place magnetized spots (those little binary 0's and 1's) on the drum during a writing operation and to sense these spots during a reading operation. The read/write heads of a drum perform in a manner similar to the read/write heads of a magnetic tape unit or disk drive unit.

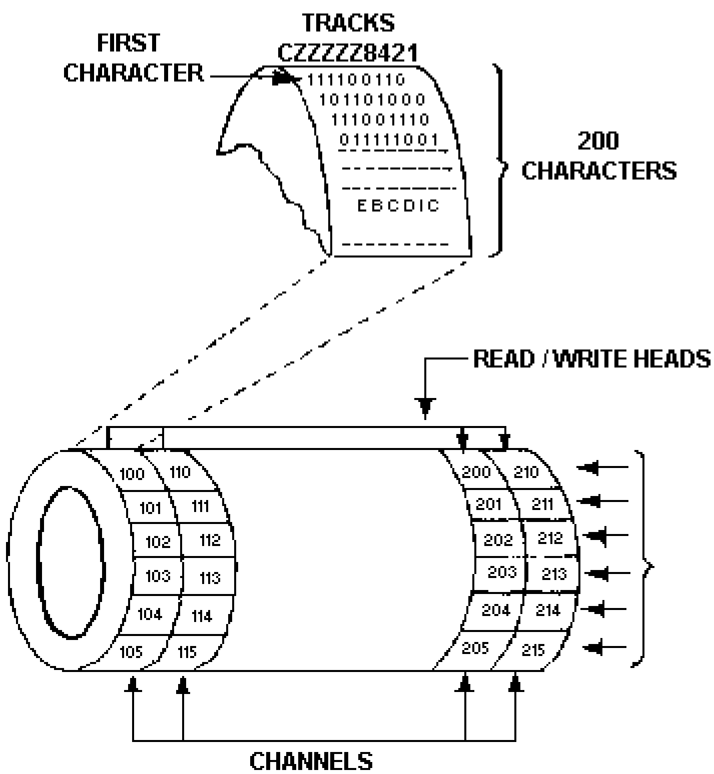


Figure 2-16.—Magnetic drum.

The tracks on each channel are grouped into sectors as illustrated in figure 2-16. Does this sound familiar to you? It sounds almost like the format used on disk packs when referring to tracks (or cylinders) and sectors. As the drum rotates, the reading or writing occurs when the specified sector of a given channel passes under the read/write head for that channel.

Some drums are mounted in a horizontal position, such as the one shown in figure 2-16, while others are mounted in a vertical position. Another major difference in the design is the number of read/write heads. Some drums use only one read/write head, which services all channels on the drum. In this case, the head moves back and forth (or up and down) over the surface of the drum as required. Other drums, using multiple read/write heads, have one principal advantage over drums with the single-head type. Since one read/write head is assigned to each channel, no read/write head movement is required. That is, the time required for head positioning is zero. The only significant time required when reading or writing is the rotational delay that occurs in reaching a desired record location.

To give you some idea of speed and storage capacities, some high-speed drums are capable of transferring over one million characters of data per second, which is roughly equivalent to reading a stack of punched cards 8 feet high in one second. The storage capacities of magnetic drums range from 20 million to more than 150,000 million characters (or bytes) of data.

Q-26. Why are disk storage devices popular?

Q-27. How is data stored on all disks?

Q-28. What precedes each record on a disk?

Q-29. How is the storage capacity of a disk determined?

Q-30. What two ways can data be physically organized on a disk pack?

Q-31. The amount of data that can be stored on a linear inch of tape is known by what term?

Q-32. The length of tape between BOT and EOT is referred to by what term?

Q-33. How does a magnetic drum differ from a magnetic disk?

Q-34. Tracks on each channel of a magnetic drum are grouped into what?

INPUT/OUTPUT DEVICES (EXTERNAL)

Input and output devices are similar in operation but perform opposite functions. It is through the use of these devices that the computer is able to communicate with the outside world. Input data may be in any one of three forms:

1. Manual inputs from a keyboard or console
2. Analog inputs from instruments or sensors
3. Inputs from a source on or in which data has previously been stored in a form intelligible to the computer

Computers can process hundreds of thousands of computer words or characters per second. Thus, a study of the first method (manual input) reflects the inability of human-operated keyboards or keypunches to supply data at a speed that matches the speed of digital computers. A high average speed for keyboard operation is two or three characters per second, that, when coded to form computer words, would reduce the data input rate to the computer to less than a computer word per second. Since mainframe computers are capable of reading several thousand times this amount of information per second, it is clear that manual inputs should be minimized to make more efficient use of computer time. However, as a rule, the keyboard is the normal input media for microcomputers.

Input data that has previously been recorded on paper tapes, magnetic tapes, magnetic disks, or floppy disks in a form understood by the program may also be entered into the computer. These are much faster methods than entering data manually from a keyboard. The most commonly used input devices in this category are magnetic tape units, magnetic disk drive units, and floppy disk drive units.

Output information is also made available in three forms:

1. Displayed information: codes, numbers, words, or symbols presented on a display device like a cathode-ray screen

2. Control signals: information that operates a control device, such as a lever, aileron, or actuator

3. Recordings: information that is stored in a machine language or human language on tapes, disks, or printed media

Devices that display, store, or read information include magnetic tape units, magnetic disk drive units, floppy disk drive units, printers, and display devices.

MAGNETIC TAPE UNITS (INPUT/OUTPUT)

The purpose of any magnetic tape unit (drive or device) is to write data on or read data from a magnetic tape (fig. 2-17). Tape stores data in a sequential manner. In sequential processing, the computer must begin searching at the beginning and check each record until the desired data is found. Like a tape cassette with recorded music, to play the fifth song recorded, you must play or fast forward the tape past the first four songs before you can play the fifth.

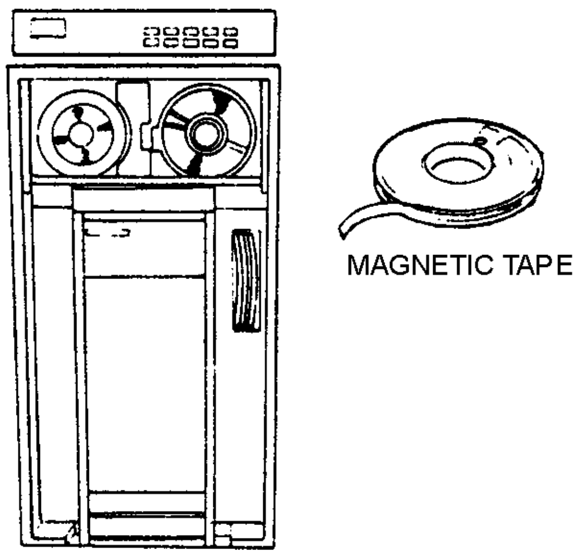


Figure 2-17.—Magnetic tape unit.

Two reels are used, tape moves from a supply reel to a take-up reel (both are mounted on hubs). Figure 2-18 shows the basic tape drive mechanism. The magnetic oxide coated side of the tape passes directly over the read/write head assembly, making contact with the heads. The magnetic tape unit reads and writes data in parallel channels or tracks along the length of the tape as shown in figure 2-19, view A. Each channel or track is used by a read/write head (one for each channel), as the tape moves across the magnetic gap of the head. Read/write heads may be either one gap or two gap as shown in figure 2-19, views B and C. The one-gap head has only one magnetic gap at which both reading and writing occur. The two-gap head has one gap for reading and another for writing. Although the one gap is satisfactory, the two-gap head gives increased speed by checking while writing. For example, a tape being written on passes over the write gap where the data is recorded, and then the data is read as it passes over the read gap to make a comparison. With this method, errors are detected almost instantly. When you look closely at figure 2-19, view B (top view), you will notice that there is one read/write coil in the write head for each channel (or track). In this particular case, there are seven. It is the electrical current flowing through these coils that magnetizes the iron-oxide coating on the surface of the tape.

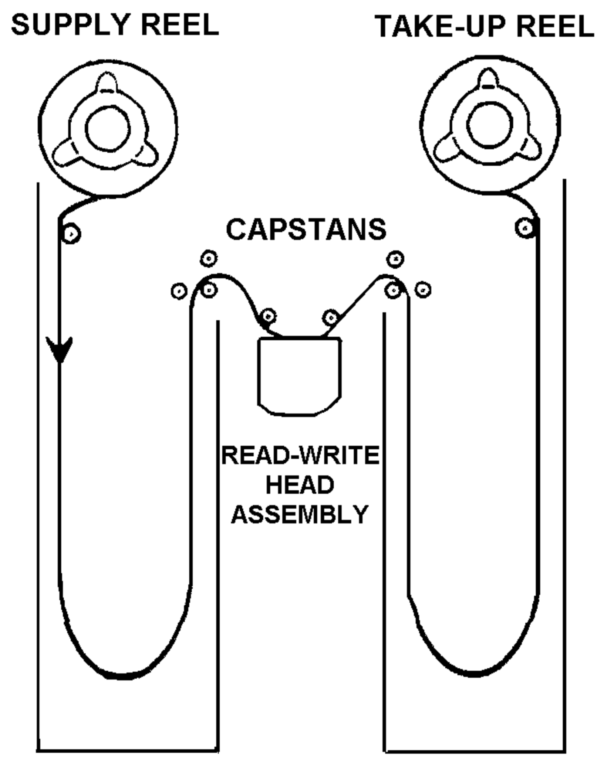


Figure 2-18.—A basic tape drive mechanism.

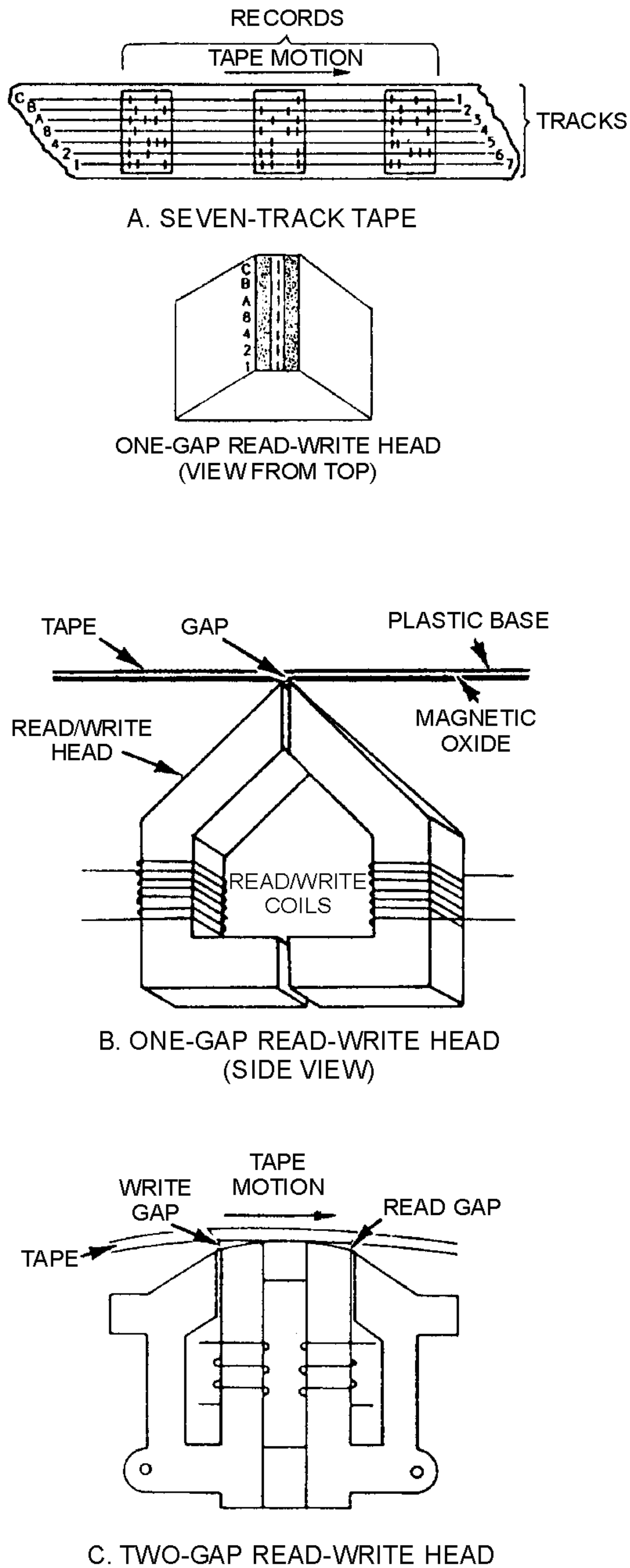


Figure 2-19.—Read/write head assemblies.

The major differences between magnetic tape units are the speed at which the tape is moved past the read/write head and the density of the recorded information. You know that density describes the number of binary digits, bytes, or frames we can record on an inch of tape. The most common tape densities are 800 and 1,600 BPI (or FPI). Tape speed (or tape movement) varies to a great extent, from less than 50 inches per second to more than 100 inches per second. How fast a tape unit reads and writes is specified as the character transfer rate which is calculated by multiplying the speed of the magnetic tape unit by the character density.

MAGNETIC DISK DRIVE UNITS (INPUT/OUTPUT)

Magnetic disk drive units are storage devices that read and write information on the magnetized surfaces of rotating disks (fig. 2-20). The disks are made of thin metal, coated on each side so that data can be recorded in the form of magnetized spots. As the disks spin around like music records, characters can be stored on them or retrieved in a direct manner. This direct accessing of data has a big advantage over the sequential accessing of data. It gives us fast, immediate access to specific data without having to examine each and every record from the beginning. You can direct the disk drive to begin reading at any point. This is like the phonograph record, you can place the needle at any point and begin playing at any point.

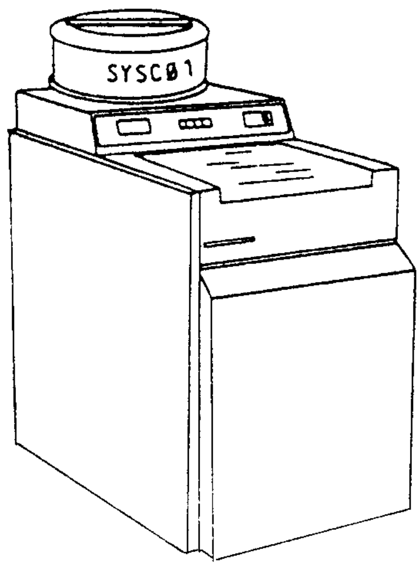


Figure 2-20.—Magnetic disk drive unit.

Located within each disk drive unit is a drive motor that rotates the disk at a constant speed, normally 3,600 revolutions per minute (rpm); or, if you prefer, 60 revolutions per second. The rotational speed for floppy disks is usually between 300 and 400 rpm because of their plastic base. Data is written on the tracks of a spinning disk surface and read from the surface by one or more (multiple) read/write heads. When reading from and writing to hard disks (rigid disks), the read/write heads float on a cushion of air and do not actually touch the surface of the disk. The distance between the head and the surface varies from a millionth of an inch to one-half millionth of an inch. This distance is called the flying height. When multiple disks (platters) are packaged together as a unit in a disk pack, a number of access arms and read/write heads are used to access both surfaces of each platter (fig. 2-21). The disk pack shown consists of six metal disks mounted on a central spindle. Data can be recorded on all surfaces except the top surface of the top disk, and the bottom surface of the bottom disk. These two surfaces are intentionally left blank for protection.

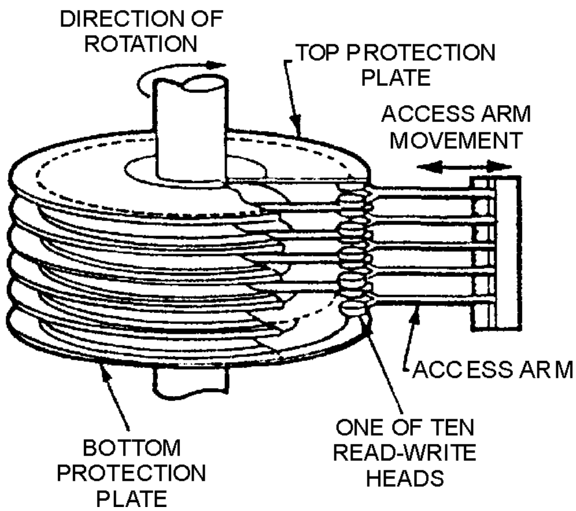


Figure 2-21.—Multiple access arms, read/write heads used with disk packs.

FLOPPY DISK DRIVE UNITS (INPUT/OUTPUT)

Floppy disk drive units are physically smaller than magnetic disk drive units and are typically used with personal (desktop) computers (fig. 2-22). The unit consists of a disk drive in which the disk rotates and a controller containing the electronic circuitry that feeds signals onto and from the disk. The disk (diskette) is a thin, flexible platter (floppy disk) coated with magnetic material so characters can be recorded on the surface in the form of magnetized spots. Floppy disks come in several sizes from 3 to 8 inches in diameter. The most common are the 8-inch disk, the 5 1/4-inch disk, and the 3 1/2-inch disk.

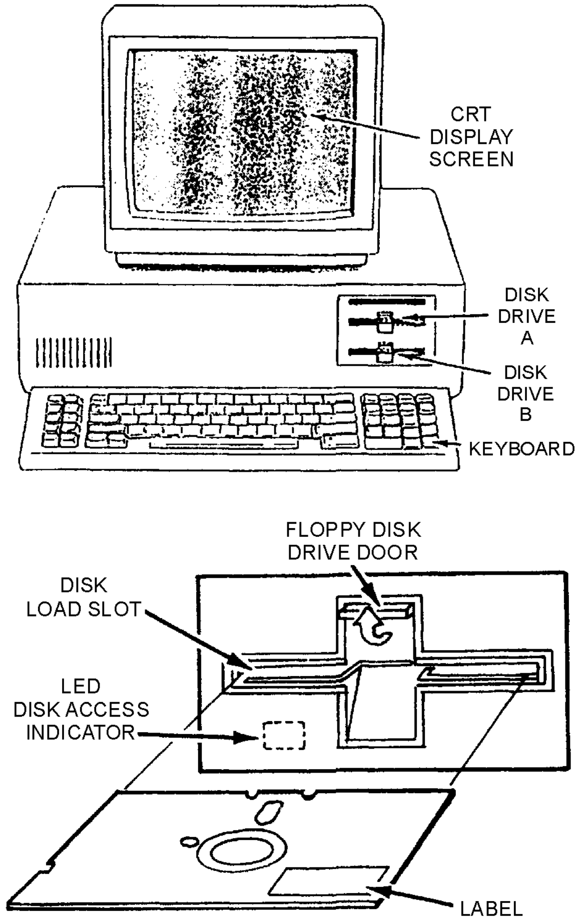


Figure 2-22.—Floppy disk drive unit.

PRINTERS (OUTPUT)

Printers are widely used output devices that express coded characters on hard (paper document) copy (fig. 2-23). They print out computer program results as numbers, letters, words, symbols, graphics, or drawings. Printers range from electronic typewriters to high-speed printers. High-speed printers are usually used on mainframes and minis to prepare supply requisitions, pay checks, inventory, or financial reports at 10 lines per second and faster. The types of printers we'll discuss are daisy-wheel, dot matrix, ink jet, and laser. These are the ones commonly used with personal computers.

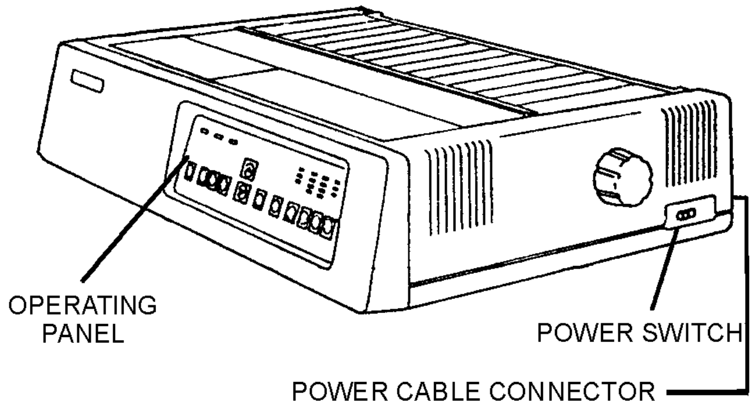


Figure 2-23.—Printer.

Daisy-Wheel Printers

Daisy-wheel printers have the most professional-looking, pleasing-to-the-eye print of all the printers in the character-at-a-time impact printer class. Daisy-wheel printers are often used in an office or word processing environment, where crisp, sharp, high-quality (letter quality) characters are a must. The daisy-wheel printer uses a round disk, with embossed characters located at the end of each petal-like projection (one character per petal), similar to the petals of a daisy, as shown in figure 2-24. A drive motor spins the wheel at a high rate of speed. When the desired character spins to the correct position, the print hammer strikes that character causing it to be printed on the paper. Once printed, the daisy wheel continues to move, searching out the next character to be printed, until the line is completed. The speeds of daisy-wheel printers range from 30 to 60 characters per second (cps).

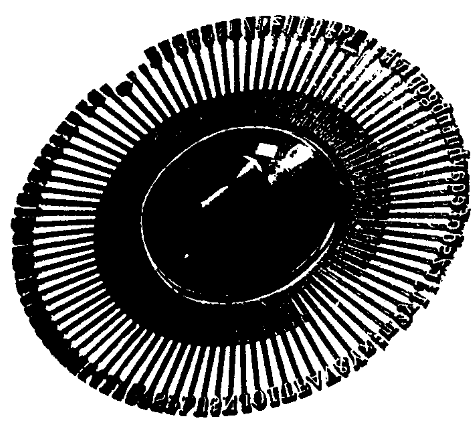


Figure 2-24.—A daisy-wheel print wheel.

Dot-Matrix Printers

Dot-matrix printers, (also known as the wire matrix printers) create characters in much the same way you see numbers on the scoreboard at a baseball or football game. In contrast to the daisy-wheel printers, dot-matrix printers use an arrangement of tiny pins or hammers, called a dot matrix, to generate characters a dot-at-a-time. A dot-matrix print head builds characters out of the dots created by the pins in the matrix. Figure 2-25, view A, shows what dot matrix characters look like when printed.

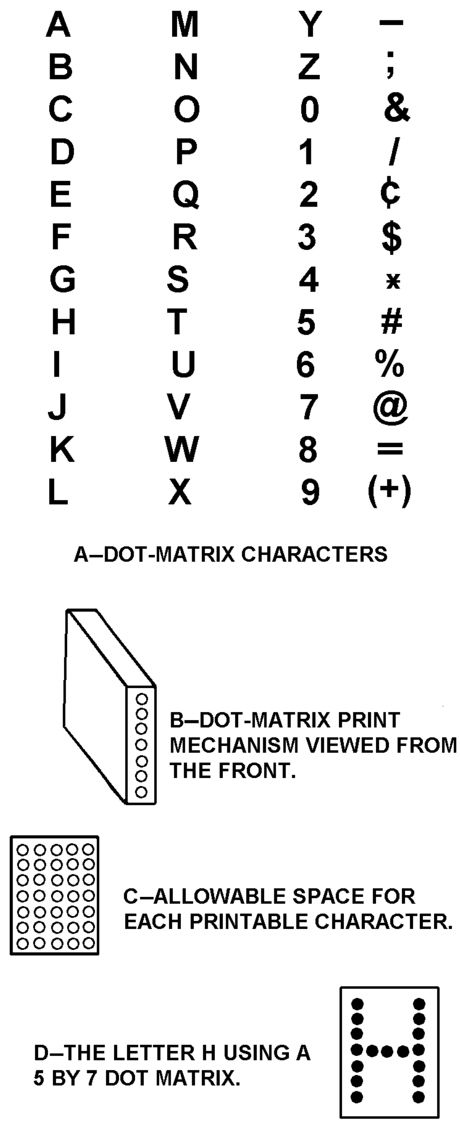


Figure 2-25.—Dot-matrix printing.

The dot matrix is defined in terms of rows and columns of dots. A 5 by 7 matrix uses up to five vertical columns of seven dots to create a character. An example of a 5 by 7 matrix printing the letter H is shown in figure 2-25, view D. The size of dot matrixes varies from a 5 by 7 matrix to as large as a 58 by 18 matrix. A number of dot-matrix printers use a single vertical column of pins to print characters, as shown in figure 2-25, view B. The characters are printed by moving (stepping) the print head a small amount and printing the vertical columns one at a time until the character is printed as shown in figure 2-25, views C and D.

The size of the matrix determines the quality of the printed character. In other words, the more dots used to print a character, the better the character is filled in and the higher its print quality. Dot-matrix printers are faster than the daisy-wheel printers with speeds ranging from 60 to 350 cps, but their print quality is not as good.

Ink Jet Printers

Ink jet printers employ a technique very similar to the way we use a can of spray paint and a stencil. A spray of electrically charged ink is shot (under pressure) toward the paper. Before reaching the paper, the ink is passed through an electrical field which forms the letters in a matrix form. The print resulting from this process consists of easy to read, high-quality characters. Some manufacturers use large droplets of ink for faster printing, while others use small droplets for better clarity but with slightly reduced printing speeds. These printers can print up to 300 cps (characters per second).

Laser Printers

Laser printers direct a beam of light through a rotating disk containing the full range of print characters. The appropriate character image is directed onto photographic paper, which is then put through a toner, developed, and used to make additional copies. The print resulting from this process consists of sharp, clean images that are easy on the eyes. These printers can print up to 20,000 plus lines per minute, or 26,666 cps (characters per second).

KEYBOARDS (INPUT)

A keyboard is nothing more than an array of switches called keyswitches. Keyboards are designed to input a code to the computer when a keyswitch is depressed. Each keyswitch, or key, on the keyboard is assigned a particular code value; and it is usually imprinted with a legend to identify its function. Figure 2-26 shows a keyboard combined with a crt on a microcomputer.

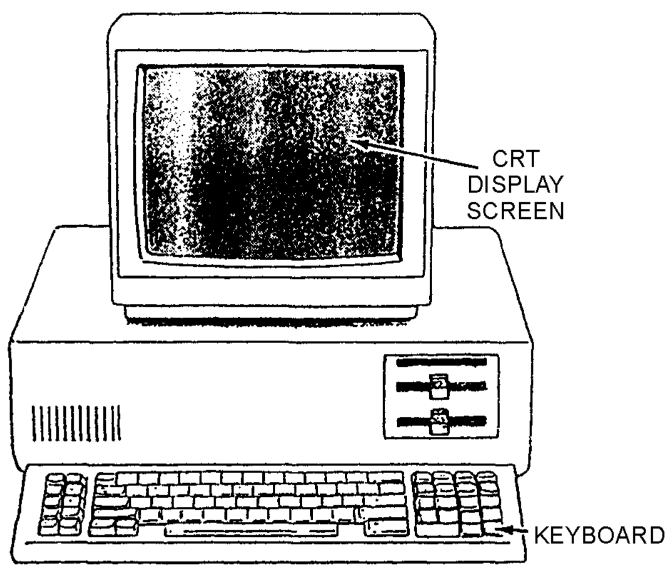


Figure 2-26.—Keyboard combined with a crt and microcomputer.

The primary purpose of a keyboard is to enter or input alphanumeric (numbers, letters, and special characters) character codes. The major grouping of keyswitches on a keyboard will be in one of the two styles of a typewriter keyboard arrangement (QWERTY or DVORAK). The typewriter keyswitches are arranged in 4 rows of 10 or more switches. The keyboard arrangement shown in figure 2-27 is QWERTY. The rows are usually offset to the row above to make it easier to reach all the keys when typing. The tops of the individual keyswitches are sculptured to conform to the shape of the human finger.

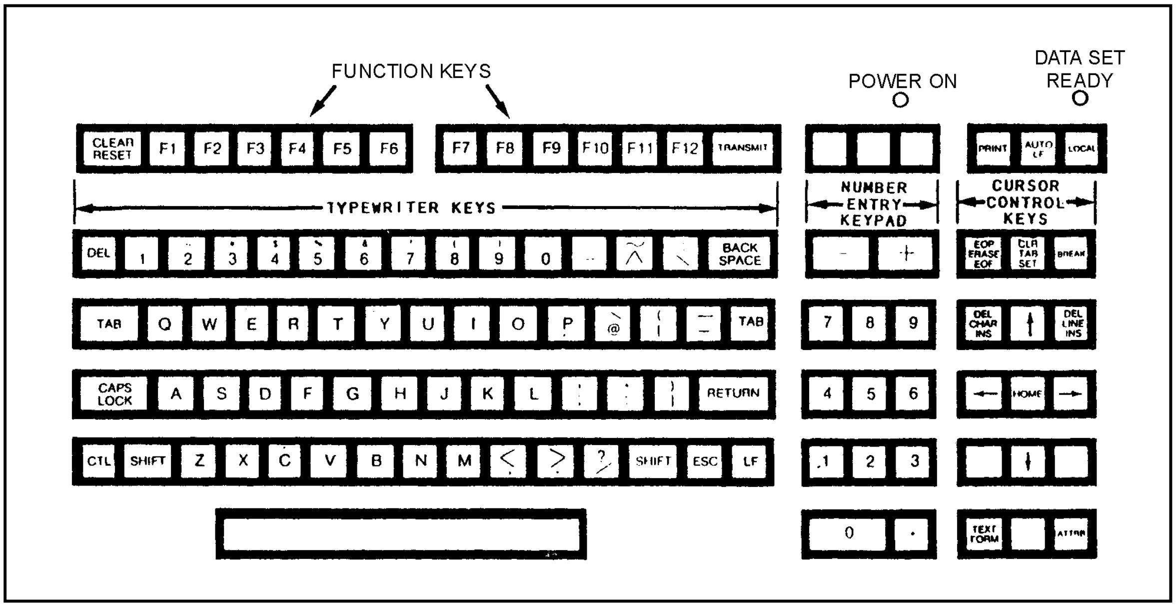


Figure 2-27.—Keyboard layout.

Other groupings of keyswitches are used for special purposes, such as number entry (calculator) keypads, special function switches (F1-F12), and cursor control keys. The special function switches allow an operator to use the special functions designed in the software. For example, in a word processing program, you can use them to spell check a document, search for a particular portion of text, move text from one place to another, and to print hard copies of a document. These are but a few of the functions allowed; however, as you become more familiar with computers you will learn them all. The cursor control key allows you to move to different locations on the screen.

The design of keyboards varies from device to device and is dependent on the requirements of the system in which the keyboards are installed.

Keyboards are generally used with nontactical computer systems. However, the newer tactical display system consoles have optional keyboards for data entry. A keyboard may be built into the display device, or it may be a separate component connected only by a communication cable.

DISPLAY DEVICES

Display devices are the crts and other displays that are part of computer terminals, computer consoles, and microcomputers. They are designed to project, show, exhibit, or display softcopy information (alphanumerics or graphic symbology).

The information displayed on a display device screen is not permanent. That is where the term soft-copy comes from. The information is available for viewing only as long as it is on the display screen. Two types of display devices used with personal/microcomputers are the raster scan crt's and the flat panel displays.

Raster Scan Crts

Raster scan crts (tv scan video monitors or display monitors) are used extensively in the display of alphanumeric data and graphics. They are used primarily in nontactical display applications such as SNAP II user terminals and desktop computers.

The raster is a series of horizontal lines crossing the face of the crt screen (fig. 2-28). Each horizontal line is made up of one trace of the electron beam from left to right. The raster starts at the top left corner of the crt screen. As each horizontal line is completed, the blanked electron beam is rapidly returned or retraced to the left of the screen.

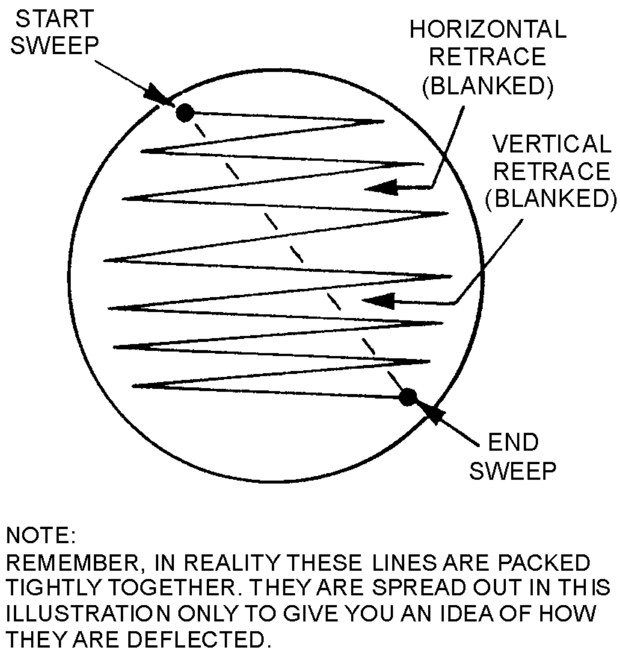


Figure 2-28.—Raster or TV scan.

Vertical deflection moves the beam down, and the horizontal sweep repeats. When the vertical sweep reaches the bottom line of the raster, a vertical blanked retrace returns the sweep to the starting position of the raster, and the process is repeated.

Each completed raster scan is referred to as a field; two fields make up a frame. The display rate of fields and frames determines the amount of flicker in the display that is perceived by the human eye. Each field is made up of approximately 525 horizontal lines. The actual number of horizontal lines varies from device to device. A frame consists of the interlaced lines of two fields. The horizontal lines of the two fields are interlaced to smooth out the display. A display rate of 30 frames per second produces a smooth, flicker-free raster and corresponding display on the screen.

PICTURE ELEMENTS.—The actual display of data results from the use of picture elements. A picture element is a variable dot of light derived from video signals input to the display monitor. The picture elements, often called pixels or pels, are contained in the horizontal scan lines crossing the face of the crt screen. The horizontal and vertical sweeps are continuous and repetitive in nature.

Pictures with alphanumeric characters and graphics can be created and displayed by varying the intensity or brightness of the picture element dots. This is done in conjunction with the phosphor coating on the face of the crt.

The number of picture elements in each horizontal line varies from device to device. The actual number of picture elements is dependent on the frequency bandwidth of the video monitor, the number of characters to be displayed on a line, and the physical size of the screen.

Each picture element is addressable by a row and column address. Picture elements are numbered from left to right on each horizontal line (column number). Each horizontal line has a row number. Picture elements, at a minimum, will have off (blanked) or on (full intensity) states. Many display devices have the capability to display picture elements at varying degrees of intensity for the display of graphics.

Characters are assembled on the screen in much the same way as a dot-matrix print head prints a character. It takes several horizontal lines and picture elements on each line to create a character. Figure 2-29 shows the generation of the character A, 7 picture elements wide and 9 horizontal lines high. The character is built using what is, in effect, a 7 by 9 dot matrix. The picture elements used to build the character would be at full intensity; the remaining picture elements in the matrix would be blanked. If dark characters on a lighted screen were desired, then the character picture elements would be blanked and the remainder displayed at full intensity.

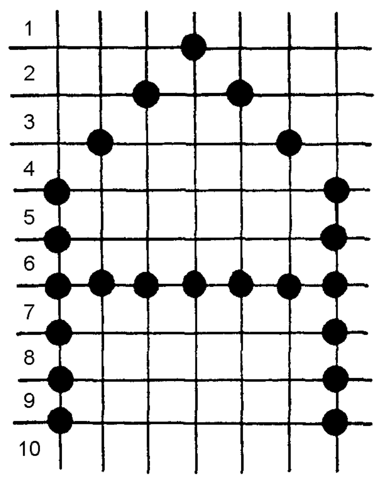


Figure 2-29.—A 7 by 9 picture element character.

Approximately 640 picture elements per horizontal line are required for the display of an 80 character line. Therefore, you can expect 140,000 picture elements on a raster scan display screen (80 alphanumeric characters per line and 25 lines).

HORIZONTAL AND VERTICAL RESOLUTION.—Horizontal resolution is defined in terms of the number of picture elements that can be displayed on the horizontal line without overlapping or running into each other. It is often stated in terms of lines of resolution. In other words, a monitor with a horizontal resolution of 1,000 lines can display 1,000 vertical lines using 1,000 picture elements per line.

Vertical resolution depends on the number of horizontal scan lines used by the particular display raster. Generally, the greater the number of scan lines, the easier it is to resolve a horizontal line of display. This characteristic remains true up to a point, called the merge point, where the variation between the lines cannot be detected by the human eye.

DISPLAYING DATA ON RASTER SCAN SCREENS.—Raster scan displays are repetitive in nature. The raster frame is displayed approximately 30 times a second.

The basic video monitor does nothing more than display the video signals it receives. If no video signals are received, then all the picture elements remain blanked, and the screen is blank in each frame. For data to be displayed accurately, each and every frame must blank and unblank the same picture elements.

The digital logic that drives video monitors is designed to take advantage of the repetitive nature of frames. There can only be a fixed number of picture elements on the screen of a display; therefore, the contents of the display screen are organized into a data unit called a page.

The page contains the status of every picture element on the display screen. The page is usually stored in some form of random-access memory, RAM chips being the most common. The contents of page memory, or, as it is sometimes called, video memory, are continually scanned by the video generation logic and used to develop the video signals for the picture element display. The picture element locations in page memory are read in time to develop the video signals for the picture element display on the horizontal lines.

If the display is to be changed, the contents of page memory must be changed. The display on the screen changes as new data is stored in page memory. Two addressing methods are used with page memory.

Unformatted Displays.—Displays that reference page memory by picture element address are called unformatted or fully populated displays. These displays are more commonly used for graphics rather than alphanumeric characters.

Formatted Displays.—Often displays are organized by character position and line number. These displays are known as formatted displays. This display method is used with devices displaying alphanumeric characters only or those with an alternate graphic capability.

The video generation logic of these types of displays scans the entire page memory, as before, to generate the display picture elements. The difference is in the way the new data is written into the page memory. Individual picture element addresses are not used. Character addresses are used to reference page memory.

The screen is organized into character lines. Each line is made up of a fixed number of character positions or columns. A fixed number of character lines can be displayed. A common arrangement found on display screens is twenty five 80-character lines, or 2,000 characters.

The character set that can be displayed on a device's formatted screen is stored in ROMs or PROMS. That is, the dot-matrix (picture element) patterns for each individual character to be displayed are stored. Different character sets may be displayed by simply replacing the appropriate ROM or PROM chips with new chips containing different character patterns.

Upon receipt of a character code and a row and column address, the device logic reads the picture element pattern (dot matrix) from the ROM and writes the pattern into the appropriate character position in the page memory. The desired character is then displayed at the correct position. Other display devices store the codes in page memory and convert the codes to picture element dots when scanning memory to refresh or redisplay the characters on the screen. The use of formatted displays greatly simplifies the programming requirements for the display of alphanumeric data.

Flat Panel Displays

A number of display methods are in use that are designed to reduce the depth of the crt display caused by the length of the tube. These devices are collectively known as flat panel displays. Three types of flat panel displays commonly in use with computer systems are liquid crystal displays (LCDs), gas plasma displays (GPDs), and electroluminescent displays (ELDs).

The screens of these flat panel displays are made up of pairs of electrodes. Each pair of electrodes is used to generate one picture element.

The liquid crystal display differs from the gas plasma and electroluminescent displays in that it does not generate its own light for the picture elements. The LCD requires an external light source, often called a backlight, for computer applications. The liquid crystal material between the charged electrodes becomes translucent when voltage is applied and allows the backlight to shine through as a picture element.

In the gas plasma and electroluminescent displays, the picture element light is generated by ionizing a gas (neon or neon argon) between the charged electrodes (gas plasma display) or by stimulating a luminescent material in the same manner (electroluminescent display). In either case, the picture element only emits light when the electrodes have voltage applied to them.

One of the advantages of flat panel displays is that smaller voltages are required for their operation than for a crt. Gas plasma displays use approximately 200 volts to charge the electrodes, and electroluminescent displays require only 20 volts.

The picture elements in these displays are addressed by the row and column method. Displays with as many as 737,280 picture elements (960 rows by 768 columns) have been developed.

The picture elements on flat panel displays are not lighted continually. This would require a large amount of power and generate excessive heat. A sequential scan similar to a crt raster is used. Once again a page memory is required. The picture element electrodes are on and off as the scan sequentially addresses page memory.

Those picture elements that are to display a dot are momentarily turned on and off starting with the first picture element in the top row, or line, and ending with the last picture element on the bottom row. The picture elements are turned on and off at a high enough frequency that the human eye cannot detect the flicker of the off-on-off cycle.

The sequential scan used to light the picture elements is continuous and repetitive. Once again, the page memory must be changed to change the display. Flat panel displays may be formatted or unformatted in the same manner as crt displays.

Q-35. What is the purpose of any magnetic tape unit?

Q-36. What are the major differences between magnetic tape units?

Q-37. Why is direct accessing of data a big advantage over the sequential accessing of data?

Q-38. What is a floppy disk?

Q-39. What are the three most common sizes of floppy disks?

Q-40. What output device expresses coded characters as hard copy (paper documents)?

Q-41. What four types of printers are commonly used with personal computers?

Q-42. What is the primary purpose of a keyboard?

Q-43. Raster scan or tv scan video monitors are used extensively for what purpose?

Q-44. How many fields make up a frame?

Q-45. A field is approximately how many horizontal lines?

Q-46. What are picture elements often called?

Q-47. Vertical resolution depends on what?

Q-48. Flat panel displays are designed to reduce what problem of a crt display?

Q-49. What does the liquid crystal display require for computer applications?

SUMMARY

Now that you have finished chapter 2, you should be feeling more at ease with digital computers. You should realize by now that they are not so hard to understand, once you have the terminology down. The information that follows summarizes the important points of this chapter.

The **CENTRAL PROCESSING UNIT** is the brain of the computer. We generally refer to it as the cpu or mainframe.

The **CONTROL SECTION** directs the flow of traffic (operations) and data, and maintains order within the computer.

The **ARITHMETIC-LOGIC SECTION** performs all arithmetic operations-adding, subtracting, multiplying, and dividing. It also tests various conditions during processing and takes action based on the result.

INTERNAL STORAGE is sometimes referred to as primary storage, main storage, or main memory (because its functions are similar to our own human memory). It stores the programs and data.

MAGNETIC CORE STORAGE is made up of tiny doughnut-shaped rings made of ferrite (iron) that are strung on a grid of very thin wires.

SEMICONDUCTOR STORAGE consists of hundreds of thousands of tiny electronic circuits etched on a silicon chip.

BUBBLE STORAGE is made of semiconductor material in the form of a very thin crystal.

READ-ONLY MEMORY (ROM) allows us to permanently store programs that will not be lost even when the computer is powered down.

RANDOM-ACCESS MEMORY (RAM) is read/ write memory. It is the working memory, rather like a blackboard, that you can scribble down notes, read them, and rub them out when you are finished with them.

SECONDARY STORAGE is the memory outside the main body of the computer (cpu) where we store programs and data for future use.

MAGNETIC TAPE is a sequential access storage device.

MAGNETIC DISK is a direct access storage device.

INPUT/OUTPUT DEVICES are the means by which the computer communicates with the outside world. These include magnetic tape units, magnetic disk drive units, floppy disk drive units, printers (daisy-wheel, dot-matrix, ink jet, and laser), and display devices (raster scan crt and flat panel).

ANSWERS TO QUESTIONS Q1. AND Q49.

A-1. The central processing unit.

A-2. Three.

A-3. Control section, internal storage section, and arithmetic-logic section.

A-4. A telephone exchange.

A-5. Transfer, arithmetic, logic, and control.

A-6. Logic.

A-7. Internal storage.

A-8. Loading.

A-9. Tiny doughnut-shaped rings made of ferrite iron.

A-10. Hundreds of thousands of tiny electronic circuits etched on a silicon chip.

A-11. Integrated circuits.

A-12. All data in memory is lost when the power source is removed.

A-13. Nonvolatile (magnetic core storage and bubble memory are examples).

A-14. A very thin crystal made of semiconductor material.

A-15. By passing a current through a control circuit imprinted on top of the crystal.

A-16. The data is still present after being read.

A-17. Read-only memory (ROM).

A-18. Only the manufacturer.

A-19. No.

A-20. Read/write memory.

A-21. By giving the computer the address of the location where the data is stored or is to be stored.

A-22. Already programmed by the manufacturer or in a blank state.

A-23. If a mistake is made and entered, it cannot be corrected or erased.

A-24. Erasable programmable read-only memory.

A-25. With a burst of ultra-violet light.

A-26. Largely because of their direct access capabilities.

A-27. In a number of invisible concentric circles called tracks.

A-28. A disk address.

A-29. By the bits per inch of track and the tracks per inch of surface.

A-30. By cylinder or sector.

A-31. Recording density.

A-32. The usable recording (reading/writing) surface or usable storage area.

A-33. The tracks in which the data is stored are assigned to channels that form circular bands around the drum.

A-34. Sectors.

A-35. To write data on or read data from a magnetic tape.

A-36. The speed at which the tape is moved past the read/write head and the density of the recorded information.

A-37. It gives us fast, immediate access to specific data without having to examine each and every record from the beginning.

A-38. A thin, flexible platter coated with magnetic material so characters can be recorded.

A-39. 8 inch, 5 1/4 inch, and 3 1/2 inch.

A-40. Printers.

A-41. Daisy-wheel, dot-matrix, ink jet, and laser.

A-42. To enter or input alphanumeric character codes.

A-43. The display of alphanumeric data and graphics.

A-44. Two.

A-45. 525.

A-46. Pixels or pels.

A-47. The number of horizontal scan lines used.

A-48. Reduce the depth of the crt display caused by the length of the tube.

A-49. An external light source, called a backlight.

CHAPTER 3
SOFTWARE

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to do the following:

1. Recognize and compare the different types and functions of operating systems.
2. Identify the types of utilities and explain their functions.
3. Describe the different types and functions of programming languages.
4. Explain the steps necessary to develop a program and describe the tools used.
5. Compare and describe the types and functions of applications packages.

INTRODUCTION

Up to now we have been discussing computer OPERATIONAL CONCEPTS and HARDWARE (the computer and its peripheral devices), and how these devices work and communicate with each other. What about this thing called SOFTWARE? Do we really need it? We most certainly do! Software plays a major role in computer data processing. For example, without software, the computer could not perform simple addition. It's the software that makes everything happen. Or putting it another way, software brings the computer to life.

You already know it takes a program to make the computer function. You load an operating system into the computer to manage the computer's resources and operations. You give job information to the operating system to tell it what you want the computer to do. You may tell it to assemble or compile a COBOL program. You may tell it to run the payroll or print inventory reports. You may tell it to copy a tape using a utility program. You may tell it to print the data from a disk file, also using a utility program. You may tell it to test a program. This job information may be entered through the console or read into the computer from tape or disk. It also may be entered by the programmer or user from a remote computer terminal. The operating system receives and processes the job information and executes the programs according to that job information.

Software can be defined as all the stored programs and routines (operating aids) needed to fully use the capabilities of a computer. Generally speaking, we say, "If it is not hardware then it must be software."

OPERATING SYSTEMS

The operating system is the heart of any computer system. Through it, everything else is done. Basically, operating systems are designed to provide the operator with the most efficient way of executing many user programs.

An operating system is a collection of many programs used by the computer to manage its own resources and operations. These programs control the execution of other programs. They schedule, assign resources, monitor, and control the work of the computer. There are several types.

TYPES OF OPERATING SYSTEMS

Operating systems are designed to provide various operating modes. Some systems can only do one task at a time, while others can perform several at a time. Some systems allow only one person to use the system, and others allow multiple users. Single user/single tasking operating systems are the simplest and most common on microcomputers. CP/M®-80[1], CP/M-86®[1], and MS-DOS®[2] are examples. Single user/multitasking operating systems allow you to do more than one task as long as the tasks don't use the same type of resources. For example, you can print one job while you run another, as long as the second job does not require the printer. Examples are Concurrent CP/M® -86 [3], Concurrent® DOS [3], and MS-DOS®; (3.0 and above). Multiuser/multitasking operating systems let more than one user access the same resources at the same time. This is especially useful for sharing common data. These are only feasible on processors (the functional unit in a computer that interprets and executes instructions) of 16 bits or more and with large memories. UNIX [4] is an example. There are also multiprocessor systems, shared resource systems. This means each user (or operator) has a dedicated microprocessor (cpu), which shares common resources (disks, printers, etc.).

1. CP/M and CP/M-86 are registered trademarks of Digital Research Inc.

2. MS-DOS is a registered trademark of Microsoft Corporation.

3. Concurrent CP/M and Concurrent DOS are trademarks of Digital Research Inc.

4. UNIX is a trademark of AT&T.

COMPATIBILITY WITH APPLICATIONS SOFTWARE

To use an applications program, it must be compatible with the operating system. Therefore, the availability of application software for a particular operating system is critical. Because of this, several operating systems have become the most popular. For 8-bit microcomputers, CP/M (Control Program for Microprocessors) is widely used because many hardware manufacturers have adopted it. MS-DOS (MicroSoft Disk Operating System) designed from CP/M dominates in lower performance 16-bit systems. UNIX, an operating system for larger computers, is being used on the more powerful 16-bit and 32-bit microcomputers. Other operating systems are offered by microcomputer manufacturers.

To overcome the applications software compatibility problem, some software comes in several versions so it can be run under several different operating systems. The point to remember is that not all applications software will run on all systems. You have to check to see that compatibility exists. You need the right version.

OPERATING SYSTEM FUNCTIONS

To give you a better idea of what you can expect to see on your microcomputer display screen, we will show a few fundamental disk operating system commands and messages. Again, the functions of each operating system are about the same, but each may use a different command to do about the same thing. For example, try not to get confused because CP/M uses the command PIP (peripheral interchange program) to copy a file, while MS-DOS uses the command COPY.

Remember, the first thing you need to do is boot (initial program load) the system. There are many ways this can be done. Here is an example. When you turn on the power, a prompt may appear on the screen. You then insert the operating system floppy disk into the drive A. Type a B (for boot) and press the RETURN key. The operating system will load from the disk. If you are using a system set up for automatic booting, you won't have to type the B. The system automatically loads the operating system when you insert the disk that contains it. Some systems will then ask for date and time. Enter these. You will next see a prompt, usually A> (or A:). The system is ready and drive A is assigned as your primary drive. One thing you might want to do is to display the disk directory to see what is on the disk. To do this, enter DIR following the A>. This will list your files.

COMMAND	.COM
CONFIGUR	.COM
DATDBL	.BAK
DATDBL	.DOC
FINANCE	.BAS
MASTER	.DOC

It may also give you file size and the date and time of the file. Let's take an example. Let's say you are to copy the file "MASTER.DOC" from the floppy disk in drive A to the floppy disk in drive B and then delete the file on the floppy disk in drive A. You have just displayed the directory of the floppy disk in drive A. Check to see that the file you want is on the floppy disk in drive A. It is. You then insert the floppy disk on which you want the copy into drive B. Be sure it is formatted with the track and sector information so it is ready to receive data. Also, be sure the disk is not write-protected. On a 5-1/4 inch floppy disk that means the write protect notch is uncovered. Following the A> type

COPY MASTER.DOC B:

and press RETURN. The system will copy the file and give it the same name. Next you might want to display the directory on drive B to see that the file was copied. You can do this by entering DIR B: following the A> prompt. To delete the file on the floppy disk in drive A, type

DEL MASTER.DOC

following the A> prompt on the screen and press RETURN.

You probably noticed each entry in the directory is followed by three characters. These are called extensions, and we use them to tell us the type of file we are working with. For example,

.BAK	Means backup file.
.BAS	Means BASIC source program.
.TMP	Means temporary file.
.DOC	Means ASCII document file.
.BIN	Means binary file, and so on.

Other typical built-in operating system commands you can use might include:

RENAME	to change the name of a file
DISKCOPY	to copy a whole floppy disk
FORMAT	to initialize a floppy disk, get it ready to receive data and programs from the system
TIME	to display or set the time

You will learn to use these and many other system commands as you operate a specific computer.

We won't go into any more detail here. You will have documentation and reference manuals for the specific version of the operating system you will be using.

Q-1. What is the heart of any computer system?

Q-2. Which types of operating systems are the simplest and most common on microcomputers?

Q-3. What types of operating systems let more than one user access the same resources at the same time?

Q-4. Why is the availability of applications software for a particular operating system critical?

Q-5. How is the applications software compatibility problem overcome?

UTILITY PROGRAMS

Now that you have learned about operating systems, let's go into another type of program, utilities. In addition to the utility commands (like diskcopy and rename), which are built into the operating system, you will probably have some independent utility programs. These are standard programs that run under control of the operating system just like your applications programs. They are called utilities because they perform general types of functions that have little relationship to the content of the data. Utility programs eliminate the need for programmers to write new programs when all they want to do is copy, print, or sort a data file. Although a new program is not needed, we do have to tell the program what we want it to do. We do this by providing information about files, data fields, and the process to be used. For example, a sort program arranges data records in a specified order. You will have to tell the sort program what fields to sort on and whether to sort in ascending or descending sequence.

Let's examine two types of utility programs to give you some idea of how a utility program works. The first will be sort-merge and the second the report program generator (RPG).

SORT-MERGE PROGRAMS

Sorting is the term given to arranging data records in a predefined sequence or order. Merging is the combining of two or more ordered files into one file. For example, we normally think of putting a list of people's names in alphabetical order arranging them in sequence by last name. We arrange those with the same last name in order by first name.

If we do this ourselves, we know the alphabetic sequence —B comes after A, C after B, and so on, and it is easy to arrange the list, even if it is a time consuming job. On a computer, the sequence of characters is also defined. It is called the collating sequence. Every coding system has a collating sequence. The capability of a computer to compare two values and determine which is greater (B is greater than A, C is greater than B, and so on) makes sorting possible. What about numbers and special characters? They are also part of the collating sequence. In EBCDIC, (EBCDIC is a computer code that is discussed in detail in chapter 4) special characters, such as #, $, &, and *, come in front of the alphabetic characters, and numbers follow. When you sort records in the defined sequence, they are in ascending sequence. Most sort programs also allow you to sort in reverse order. This is called descending sequence. In EBCDIC, it is 9-0, Z-A, then special characters.

To sort a data file, you must tell the sort program what data field or fields to sort on. These fields are called sort (or sorting) keys. In our example, the last name is the major sort key and the first name is the minor sort key.

Sorting is needed in many applications. For example, for mailing we need addresses in ZIP Code order; personnel records may be kept in service number order; and inventory records may be kept in stock number order. We could go on and on. Because many of our files are large, sorting is very time consuming, and it is one of the processes most used on computer systems. As a user, you will become very familiar with this process.

Sort-merge programs usually have phases. First they initialize: read the parameters, produce the program code for the sort, allocate the memory space, and set up other functions. The sort-merge program then reads in as many input data records as the memory space allocated can hold, arranges (sorts) them in sequence, and writes them out to an intermediate sort-work file. It continues reading input, sorting and writing intermediate sort-work files until all the input is processed. It then merges (combines) the ordered intermediate sort-work files to produce one output file in the sequence specified. The merging process can be accomplished with less memory than the sort process since the intermediate sort work files are all in the same sequence. Records from each work file can be read, the sort keys compared based on the collating sequence and sort parameters, and records written to the output file maintaining the specified sequence.

REPORT PROGRAM GENERATORS

Report program generators (RPG) are used to generate programs to print detail and summary reports of data files. Figure 3-1 is an example of a printed report. RPGs were designed to save programming time. Rather than writing procedural steps in a language like BASIC or COBOL, the RPG programmer writes the printed report requirements on specially designed forms.

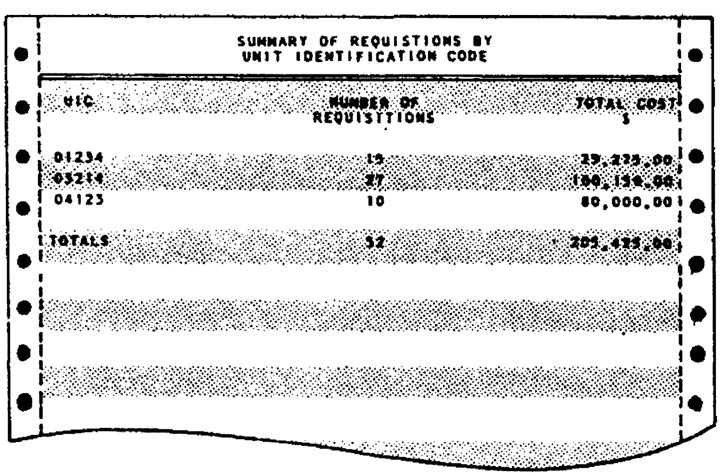


Figure 3-1.—Printed report using a report program generator (RPG) program.

Included in the requirements are an input file description, the report heading information lines, the input data record fields, the calculations to be done, and the data fields to be printed and summarized. The RPG program takes this information and generates a program for the specific problem. You then run that program with the specified input data file to produce the printed report. The input data file must be in the sequence in which you want the report to summarize the data.

In our example (fig. 3-1), we summarized requisitions based on unit identification codes (UIC). We first sorted the input data file on the field that contains the UIC. We provided specifications to the RPG program to tell it to accumulate totals from the detail (individual) data records until the UIC changed. We then told it to print the total number of requisitions and total cost for that UIC. We did not have it print each detail record, although we could have. The UIC is called the control field. Each time the control field changes, there is a control break. Each time there is a control break, the program prints the summary information. After all records are read and processed, it prints a summary line (TOTALS) for all UICs. You can also use RPGs to generate a program to update data files.

Q-6. What programs eliminate the need for programmers to write new programs when all they want to do is copy, print, or sort a data file?

Q-7. How do we tell a utility program what we want it to do?

Q-8. What is the term given to arranging data records in a predefined sequence or order?

Q-9. To sort a data file, what must you tell the sort program?

Q-10. What are report program generators used for?

PROGRAMMING LANGUAGES

Programmers must use a language that can be understood by the computer. Several methods can achieve human-computer communication. For example, let us assume the computer only understands French and the programmer speaks English. The question arises: How are we to communicate with the computer? One approach is for the programmer to code the instructions with the help of a translating dictionary before giving them to the processor. This would be fine so far as the computer is concerned; however, it would be very awkward for the programmer.

Another approach is a compromise between the programmer and computer. The programmer first writes instructions in a code that is easier to relate to English. This code is not the computer's language; therefore, the computer does not understand the orders. The programmer solves this problem by giving the computer another program, one that enables it to translate the instruction codes into its own language. This translation program, for example, would be equivalent to an English-to-French dictionary, leaving the translating job to be done by the computer.

The third and most desirable approach from an individual's standpoint is for the computer to accept and interpret instructions written in everyday English terms. Each of these approaches has its place in the evolution of programming languages and is used in computers today.

MACHINE LANGUAGES

With early computers, the programmer had to translate instructions into the machine language form that the computers understood. This language was a string of numbers that represented the instruction code and operand address(es).

In addition to remembering dozens of code numbers for the instructions in the computer's instruction set, the programmer also had to keep track of the storage locations of data and instructions. This process was very time consuming, quite expensive, and often resulted in errors. Correcting errors or making modifications to these programs was a very tedious process.

SYMBOLIC LANGUAGES

In the early 1950s, mnemonic instruction codes and symbolic addresses were developed. This improved the program preparation process by substituting letter symbols (mnemonic codes) for basic machine language instruction codes. Each computer has mnemonic codes, although the symbols vary among the different makes and models of computers. The computer still uses machine language in actual processing, but it translates the symbolic language into machine language equivalent. Symbolic languages have many advantages over machine language coding. Less time is required to write a program. Detail is reduced. Fewer errors are made. Errors which are made are easier to find, and programs are easier to modify.

PROCEDURE-ORIENTED LANGUAGES

The development of mnemonic techniques and macroinstructions led to the development of procedure-oriented languages. Macroinstructions allow the programmer to write a single instruction that is equivalent to a specified sequence of machine instructions. These procedure-oriented languages are oriented toward a specific class of processing problems. A class of similar problems is isolated, and a language is developed to process these types of applications. Several languages have been designed to process problems of a scientific-mathematical nature and others that emphasize file processing.

Procedure-oriented languages were developed to allow a programmer to work in a language that is close to English or mathematical notation. This improves overall efficiency and simplifies the communications process between the programmer and the computer. These languages have allowed us to be more concerned with the problems to be solved rather than with the details of computer operation. For example:

COBOL (COmmon Business Oriented Language) was developed for business applications. It uses statements of everyday English and is good for handling large data files.

FORTRAN (FORmula TRANslator) was developed for mathematical work. It is used by engineers, scientists, statisticians, and others where mathematical operations are most important.

BASIC (Beginner's All-Purpose Symbolic Instruction Code) was designed as a teaching language to help beginning programmers write programs. Therefore, it is a general-purpose, introductory language that is fairly easy to learn and to use. With the increase in the use of microcomputers, BASIC has regained popularity and is available on most microcomputer systems.

Other languages gaining in popularity are PASCAL and Ada. PASCAL is being used by many colleges and universities to teach programming because it is fairly easy to learn; yet is a more powerful language than BASIC. Although PASCAL is not yet a standardized language, it is still used rather extensively on microcomputers. It has greater programming capabilities on small computers than are possible with BASIC.

Ada's development was initiated by the U.S. Department of Defense (DOD). Ada is a modern general-purpose language designed with the professional programmer in mind and has many unique features to aid in the implementation of large scale applications and real-time systems. Because Ada is so strongly supported by the DOD and other advocates, it will become an important language like those previously mentioned. Its primary disadvantage relates to its size and complexity, which will require considerable adjustment on the part of most programmers.

The most familiar of the procedure-oriented languages are BASIC and FORTRAN for scientific or mathematical problems, and COBOL for file processing.

Programs written in procedure-oriented languages, unlike those in symbolic languages, may be used with a number of different computer makes and models. This feature greatly reduces reprogramming expenses when changing from one computer system to another. Other advantages to procedure-oriented languages are (1) they are easier to learn than symbolic languages; (2) they require less time to write; (3) they provide better documentation; and (4) they are easier to maintain. However, there are some disadvantages of procedure-oriented languages. They require more space in memory, and they process data at a slower rate than symbolic languages.

Q-11. With early computers, the programmer had to translate instructions into what type of language form?

Q-12. When were mnemonic instruction codes and symbolic addresses developed?

Q-13. What led to the development of procedure oriented languages?

Q-14. What computer language was developed for mathematical work?

Q-15. What are two disadvantages of procedure oriented languages?

PROGRAMMING

Programming is, simply, the process of planning the computer solution to a problem. Thus, by writing:

1. Take the reciprocal of the resistance of all resistors (expressed in ohms);
2. Sum the values obtained in step 1;
3. Take the reciprocal of the sum derived in step 2.

A generalized process or program for finding the total resistance of a parallel resistance circuit has now been derived.

To progress from this example to preparing a program for a computer is not difficult. However, one basic characteristic of the computer must be kept in mind. It cannot think. It can only follow certain commands, and these commands must be correctly expressed and must cover all possibilities. Thus, if a program is to be useful in a computer, it must be broken down into specifically defined operations or steps. Then the instructions, along with other data necessary for performing these operations or steps, must be communicated to the computer in the form of a language or code that is acceptable to the machine. In broad terms, the computer follows certain steps in executing a program. It must first read the instructions (sequentially unless otherwise programmed), and then in accordance with these instructions, it executes the following procedures:

1. Locates the parameters (constants) and such other data as may be necessary for problem solution
2. Transfers the parameters and data to the point of manipulation

3. Manipulates the parameters and data in accordance with certain rules of logic

4. Stores the results of such manipulations in a specific location

5. Provides the operator (user) with a useful output

Even in a program of elementary character such as the one above, this would involve breaking each of the steps down into a series of machine operations. Then these instructions, parameters, and the data necessary for problem solution must be translated into a language or code that the computer can accept.

Next, we'll provide an introduction to the problem solving concepts and flow charting necessary to develop a program.

OVERVIEW OF PROGRAMMING

Before learning to program in any language, it is helpful to establish some context for the productive part of the entire programming effort. This context comprises the understanding and agreement that there are four fundamental and discrete steps involved in solving a problem on a computer.

The four steps are as follows:

1. State, analyze, and define the problem.

2. Develop the program logic and prepare a program flowchart or decision table.

3. Code the program, prepare the code in machine readable form, prepare test data, and perform debug and test runs.

4. Complete the documentation and prepare operator procedures for implementation and production.

Figure 3-2 depicts the evolution of a program. Programming can be complicated, and advance preparation is required before you can actually start to write or code the program. The first two steps, problem understanding/definition and flowcharting, fall into the advance planning phase of programming. It is important at this point to develop correct habits and procedures, since this will prevent later difficulties in program preparation.

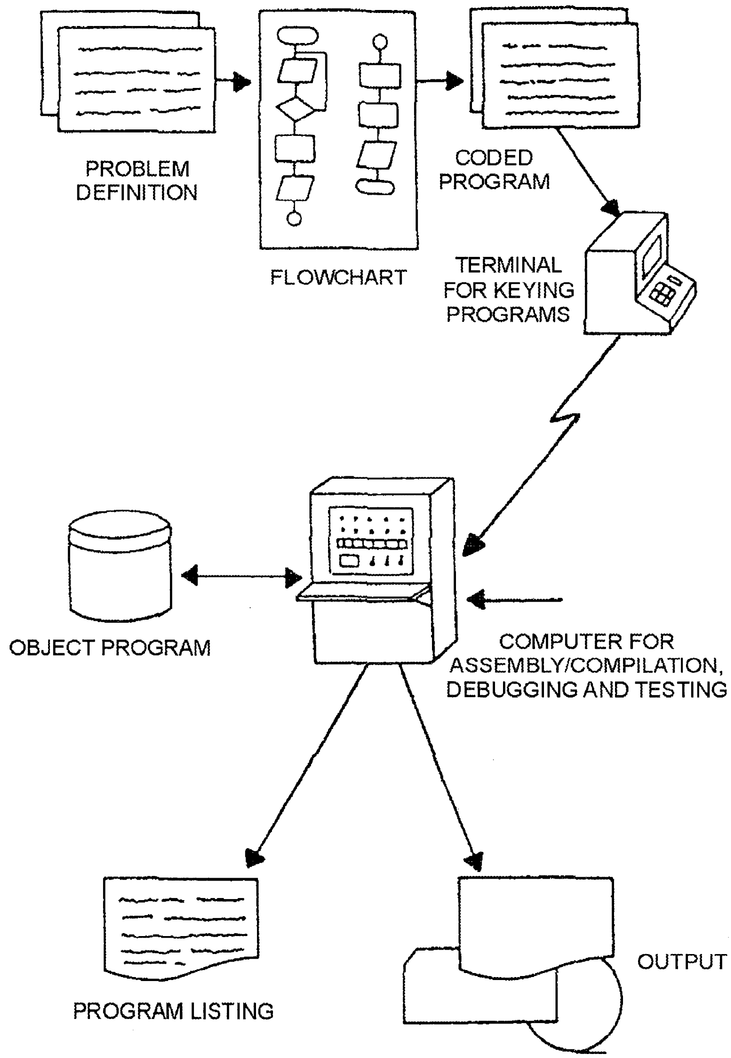


Figure 3-2.—Evolution of a program.

Whether you are working with a systems analyst, a customer, or solving a problem of your own, it is extremely important that you have a thorough understanding of the problem.

Every aspect of the problem must be defined:

- What is the problem?
- What information (or data) is needed?
- Where and how will the information be obtained?
- What is the desired output?

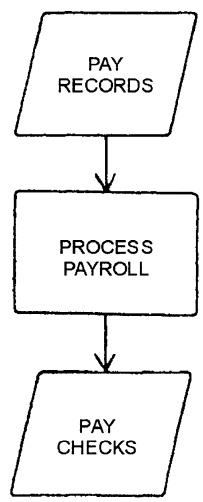
Starting with only a portion of the information, or an incomplete definition, will result in having to constantly alter what has been done to accommodate the additional facts as they become available. It is easier and more efficient to begin programming after all of the necessary information is understood. Once you have a thorough understanding of the problem, the next step is flowcharting.

FLOWCHARTING

Flowcharting is one method of pictorially representing a procedural (step-by-step) solution to a problem before you actually start to write the computer instructions required to produce the desired results. Flowcharts use different shaped symbols connected by one-way arrows to represent operations, data flow, equipment, and so forth.

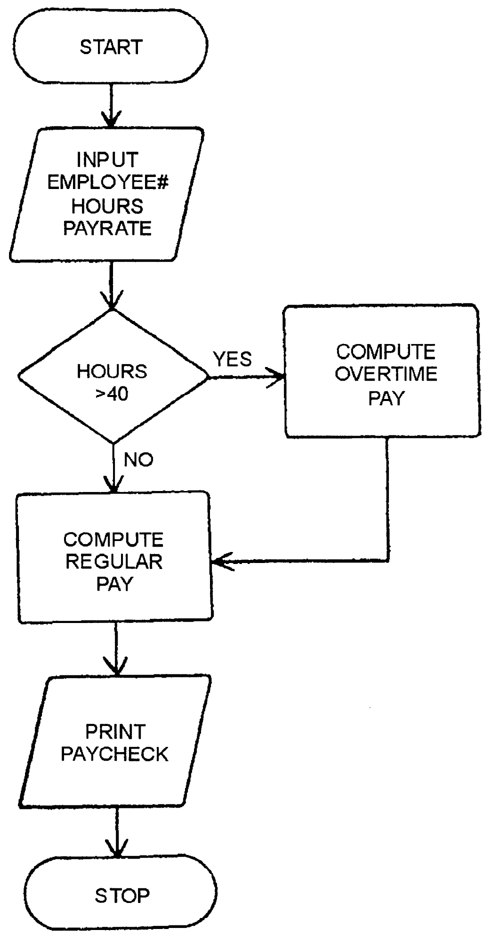
There are two types of flowcharts, system (data) flowcharts and programming flowcharts. A system (data) flowchart defines the major phases of the processing, as well as the various data media used. It shows the relationship of numerous jobs that make up an entire system. In the system (data) flowchart, an entire program run or phase is always represented by a single processing symbol, together with the input/output symbols showing the path of data through a problem solution. For example:

SYSTEM FLOWCHART



The second type of flowchart, and the one we will talk about in this section is the programming flowchart. It is constructed by the programmer to represent the sequence of operations the computer is to perform to solve a specific problem. It graphically describes what is to take place in the program. It displays specific operations and decisions, and their sequence within the program. For example:

PROGRAMMING FLOWCHART



Tools of Flowcharting

Next we will take a look at the tools used in flowcharting. These tools are the fundamental symbols, graphic symbols, flowcharting template, and the flowcharting worksheet.

FUNDAMENTAL SYMBOLS.—To construct a flowchart, you must know the symbols and their related meanings. They are standard for the military, as directed by *Department of the Navy Automated Data Systems Documentation Standards*, SECNAVINST 5233.

Symbols are used to represent functions. These fundamental functions are processing, decision, input/output, terminal, flow lines, and connector symbol. All flowcharts may be initially constructed using only these fundamental symbols as a rough outline to work from. Each symbol corresponds to one of the functions of a computer and specifies the instruction(s) to be performed by the computer. The contents of these symbols are called statements. Samples of these fundamental symbols, definitions, examples, and explanations of their uses are shown in figure 3-3.

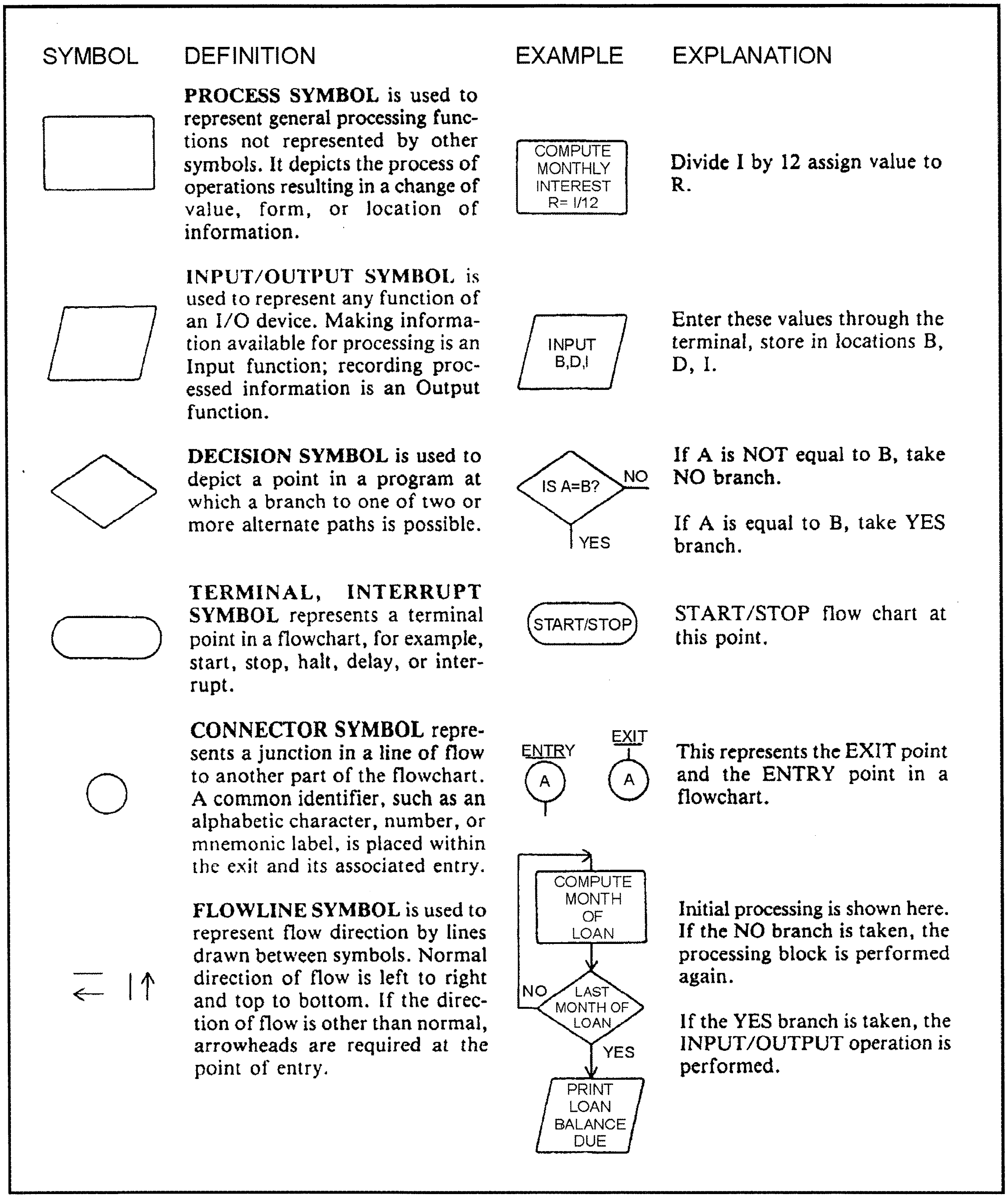


Figure 3-3.—Fundamental flowcharting symbols.

GRAPHIC SYMBOLS.—Within a flowchart, graphic symbols are used to specify arithmetic operations and relational conditions. The following are commonly-used arithmetic and relational symbols.

+	plus, add
-	minus, subtract
*	multiply
/	divide
±	plus or minus
=	equal to
>	greater than
<	less than
≥	greater than or equal to
≤	less than or equal to
≠	not equal
YES or Y	Yes
NO or N	No
TRUE or T	True
FALSE or F	False

FLOWCHARTING TEMPLATE.—To aid in drawing the flowcharting symbols, you may use a flowcharting template. Figure 3-4 shows a template containing the standard symbol cutouts. A template is usually made of plastic with the symbols cut out to allow tracing the outline.

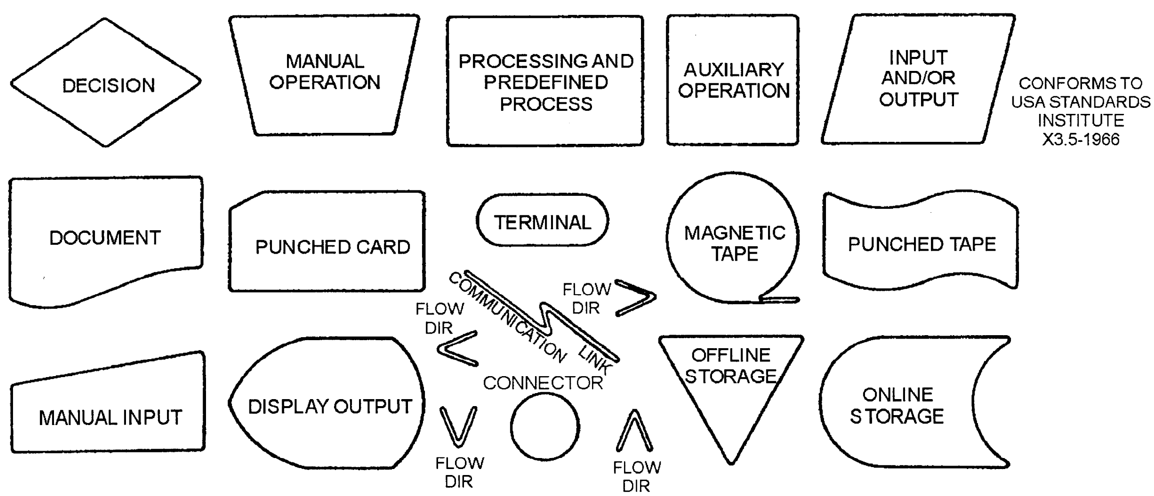


Figure 3-4.—Flowchart template.

FLOWCHART WORKSHEET.—The flowchart worksheet is a means of standardizing documentation. It provides space for drawing programming flowcharts and contains an area for identification of the job, including application, procedure, date, and page numbers (fig. 3-5). You may find it helpful when you develop flowcharts. If you don't have this form available, a plain piece of paper will do.

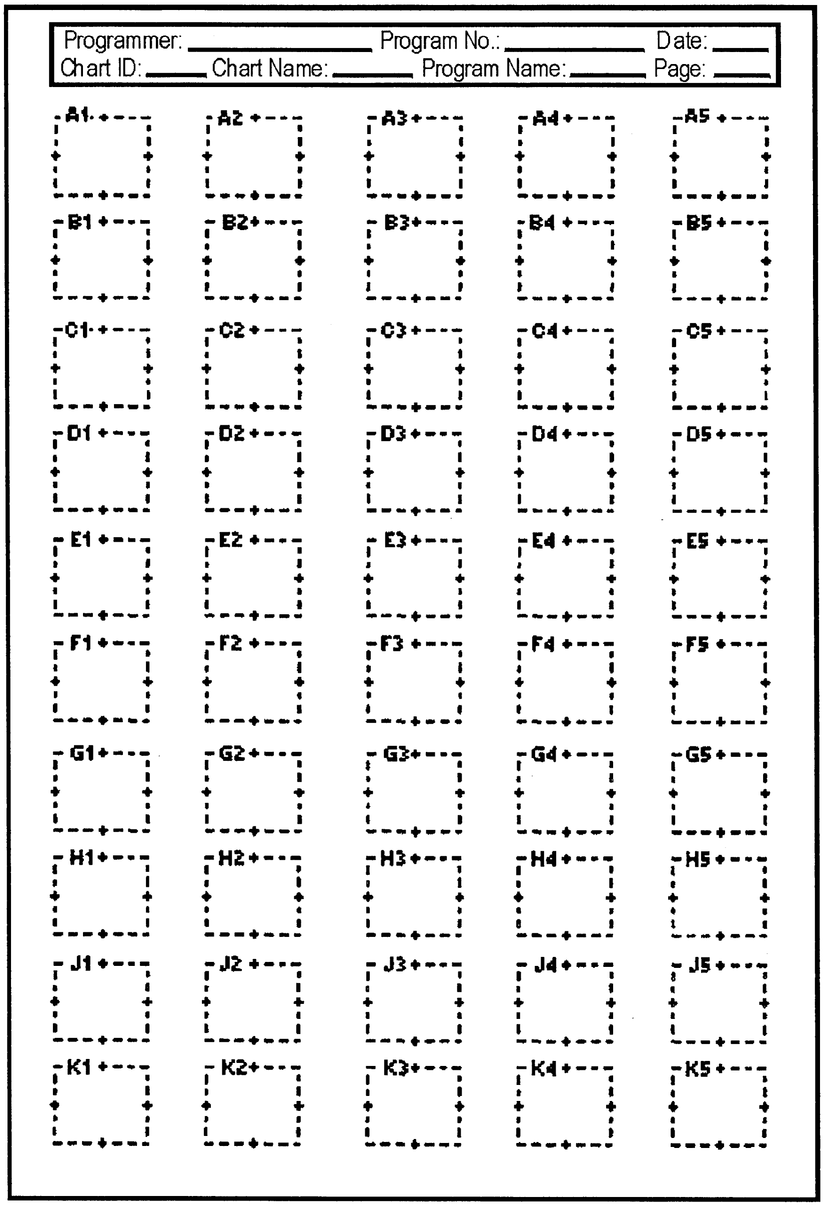


Figure 3-5.—Flowchart worksheet.

Constructing a Flowchart

There is no "best way" to construct a flowchart. There is no way to standardize problem solution. Flowcharting and programming techniques are often unique and conform to the individual's own methods or direction of problem solution.

This section will show an example of developing a programming flowchart. It is not the intent to say this is the best way; rather, it is one way to do it.

By following this text example you should grasp the idea of solving problems through flowchart construction. As you gain experience and familiarity with a computer system, these ideas will serve as a foundation.

To develop a flowchart, you must first know what problem you are to solve. It is then your job to study the problem definition and develop a flowchart to show the logic, steps, and sequence of steps the computer is to execute to solve the problem.

As an example, suppose you have taken a short-term second mortgage on a new home, and you want to determine what your real costs will be, the amount of interest, the amount to be applied to principal, and the final payment at the end of the 3-year loan period.

The first step is to be sure you understand the problem completely—What are the inputs and the outputs and what steps are needed to answer the questions? Even when you are specifying a problem of your own, you will find we don't usually think in small, detailed sequential steps. However, that is exactly how a computer operates, one step after another in a specified order. Therefore, it is necessary for you to think the problem solution through step by step. You might clarify the problem as shown by the Problem Definition in figure 3-6.

PROBLEM DEFINITION

MORTGAGE AMORTIZATION- This program is to determine the monthly amount of interest (A) and amount applied to the principal (P) of the mortgage giving the balance (B) at the end of a thirty-six month period.

INPUT: The monthly payment is to be entered as variable D, the beginning balance of the mortgage is to be entered as variable B, and the annual interest rate is to be entered as variable I. This input is to be entered into the system via the terminal.

OUTPUT: The end result is to be a listing displaying the amount applied to principal and interest and the current loan balance each month, with one final entry showing the final payment on the mortgage.

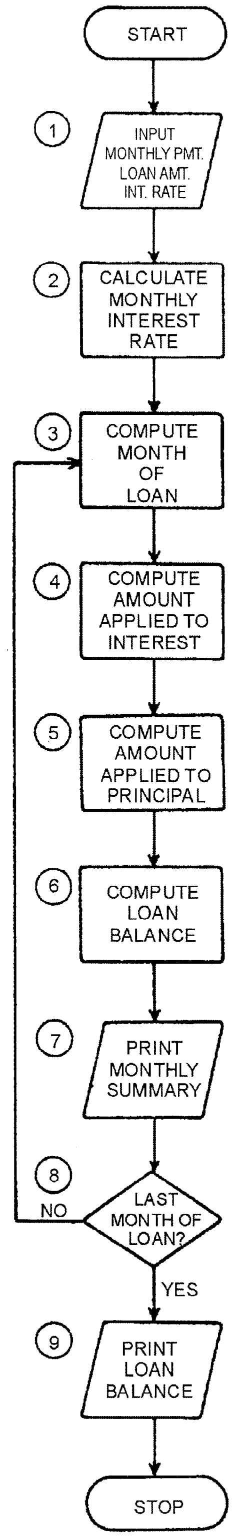


Figure 3-6.—Problem definition and programming flowchart.

After you have this level of narrative problem definition, you are ready to develop a flowchart showing the logic, steps, and sequence of steps you want the computer to execute to solve the problem. A programming flowchart of this problem is also shown in figure 3-6. Study both the problem definition and the flowchart to see their relationship and content.

You now have a plan of what you want the computer to do. The next step is to code a program that can be translated by a computer into a set of instructions it can execute. This step is called program coding.

PROGRAM CODING

It is important to remember program coding is not the first step of programming. Too often we have a tendency to start coding too soon. As we just discussed, a great deal of planning and preparation must be done before sitting down to code the computer instructions to solve a problem. For the example amortization problem (fig. 3-6), we have analyzed the specifications in terms of (1) the output desired; (2) the operations and procedures required to produce the output; and (3) the input data needed. In conjunction with this analysis, we have developed a programming flowchart that outlines the procedures for taking the input data and processing it into usable output. You are now ready to code the instructions that will control the computer during processing. This requires that you know a programming language.

All programming languages, FORTRAN, COBOL, BASIC and so on, are composed of instructions that enable the computer to process a particular application, or perform a particular function.

Instructions

The instruction is the fundamental element in program preparation. Like a sentence, an instruction consists of a subject and a predicate. However, the subject is usually not specifically mentioned; rather it is some implied part of the computer system directed to execute the command that is given. For example, the chief tells a sailor to "dump the trash." The sailor will interpret this instruction correctly even though the subject "you" is omitted. Similarly, if the computer is told to " ADD 1234," the control section may interpret this to mean that the arithmetic-logic section is to add the contents of address 1234 to the contents of the accumulator (a register in which the result of an operation is formed).

In addition to an implied subject, every computer instruction has an explicit predicate consisting of at least two parts. The first part is referred to as the command, or operation; it answers the question "what?" It tells the computer what operation it is to perform; i.e., read, print, input. Each machine has a limited number of built-in operations that it is capable of executing. An operation code is used to communicate the programmer's intent to the computer.

The second specific part of the predicate, known as the operand, names the object of the operation. In general, the operand answers the question "where?" Operands may indicate the following:

1. The location where data to be processed is found.

2. The location where the result of processing is to be stored.

3. The location where the next instruction to be executed is found. (When this type of operand is not specified, the instructions are executed in sequence.)

The number of operands and the structure or format of the instructions vary from one computer to another. However, the operation always comes first in the instruction and is followed by the operand(s). The programmer must prepare instructions according to the format required by the language and the computer to be used.

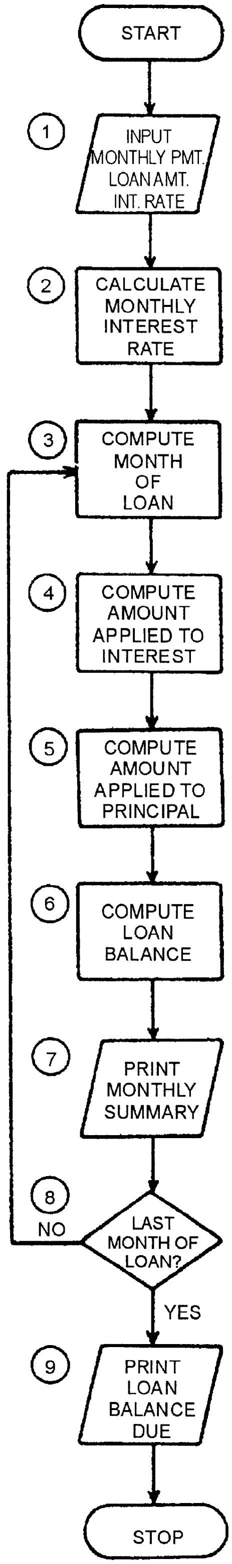
Instruction Set

The number of instructions in a computer's instruction set may range from less than 30 to more than 100. These instructions may be classified into categories by the action they perform such as input/output (I/O), data movement, arithmetic, logic, and transfer of control. Input/output instructions are used to communicate between I/O devices and the central processor. Data movement instructions are used for copying data from one storage location to another and for rearranging and changing of data elements in some prescribed manner.

Arithmetic instructions permit addition, subtraction, multiplication, and division. They are common in all digital computers. Logic instructions allow comparison between variables, or between variables and constants. Transfer of control instructions are of two types, conditional and unconditional. Conditional transfer of control instructions are used to branch or change the sequence of program control, depending on the outcome of the comparison. If the outcome of a comparison is true, control is transferred to a specific statement number. If it proves false, processing continues sequentially through the program. Unconditional transfer of control instructions are used to change the sequence of program control to a specified program statement regardless of any condition.

Coding a Program

Regardless of the language used, there are strict rules the programmer must adhere to with regard to punctuation and statement structure when coding any program. Using the programming flowchart introduced earlier, we have now added a program coded in BASIC to show the relationship of the flowchart to the actual coded instructions (fig. 3-7). Don't worry about complete understanding, just look at the instructions with the flowchart to get an idea of what coded instructions look like.




```
10 PRINT "ENTER MONTHLY PAYMENT, ";
20 PRINT "LOAN AMOUNT, INTEREST RATE"
30 INPUT D,B,I
40 LET R= I/12
50 FOR M = 1 TO 36
60 LET A = B*R
70 LET P = D-A
80 LET B = B-P
90 PRINT M;D;B,P,A
100 NEXT M
110 PRINT "WITH ONE FINAL PAYMENT OF" ,B
120 END
```

Figure 3-7.—Programming flowchart and coded program.

You will have to have specific information about the computer you are to use. It will tell you how the language is implemented on that particular computer, in order to code a program. The computer manufacturers or software designer will provide these specifics in their user's manual. Get a copy of the user's manual and study it before you begin to code. The differences may seem minor to you, but they may prevent your program from running.

Once coding is completed, the program must be debugged and tested before implementation.

Debugging

Errors caused by faulty logic and coding mistakes are referred to as "bugs." Finding and correcting these mistakes and errors that prevent the program from running and producing correct output is called "debugging."

Rarely do complex programs run to completion on the first attempt. Often, time spent debugging and testing equals or exceeds the time spent in program coding. This is particularly true if insufficient time was spent on program definition and logic development. Some common mistakes which cause program bugs are mistakes in coding punctuation, incorrect operation codes, transposed characters, keying errors, and failure to provide a sequence of instructions (a program path) needed to process certain conditions.

To reduce the number of errors, you will want to carefully check the coding sheets before they are keyed into the computer. This process is known as "desk-checking" and should include an examination for program completeness. Typical input data should be manually traced through the program processing paths to identify possible errors. In effect, you will be attempting to play the role of the computer. After you have desk checked the program for accuracy, the program is ready to be assembled or compiled. Assembly and compiler programs prepare your program (source program) to be executed by the computer, and they have error diagnostic features which detect certain types of mistakes in your program. These mistakes must be corrected. Even when an error-free pass of the program through the assembly or compiler program is accomplished, this does not mean your program is perfected. However, it usually means the program is ready for testing.

Testing

Once a program reaches the testing stage, generally, it has proved it will run and produce output. The purpose of testing is to determine that all data can be processed correctly and that the output is correct. The testing process involves processing input test data that will produce known results. The test data should include: (1) typical data, which will test the commonly used program paths; (2) unusual but valid data, which will test the program paths used to process exceptions; (3) incorrect, incomplete, or inappropriate data, which will test the program's error routines. If the program does not pass these tests, more testing is required. You will have to examine the errors and review the coding to make the coding corrections needed. When the program passes these tests, it is ready for computer implementation. Before computer implementation takes place, documentation must be completed.

Documentation

Documentation is a continuous process, beginning with the problem definition. Documentation involves collecting, organizing, storing, and otherwise maintaining a complete record of the programs and other documents associated with the data processing system.

The Navy has established documentation standards to ensure completeness and uniformity for computer system information between commands and between civilian and Navy organizations. SECNAVINST 5233.1 establishes minimum documentation requirements.

A documentation package should include:

1. A definition of the problem. Why was the program written? What were the objectives? Who requested the program, and who approved it? These are the types of questions that should be answered.

2. A description of the system. The system environment (hardware, software, and organization) in which the program functions should be described (including systems flowcharts). General systems specifications outlining the scope of the problem, the form and type of input data to be used, and the form and type of output required should be clearly defined.

3. A description of the program. Programming flowcharts, program listings, program controls, test data, test results, and storage dumps — these and other documents that describe the program and give a historical record of problems and/or changes should be included.

4. Operator instructions. Items that should be included are computer switch settings, loading and unloading procedures, and starting, running, and termination procedures.

Implementation

After the documentation is complete, and the test output is correct, the program is ready for use. If a program is to replace a program in an existing system, it is generally wise to have a period of parallel processing; that is, the job application is processed both by the old program and by the new program. The purpose of this period is to verify processing accuracy and completeness.

Q-16. What is programming?

Q-17. In programming, how many steps are involved in solving a problem on a computer?

Q-18. What is required before you can actually start to write or code a program?

Q-19. In flowcharting, what method is used to represent different operations, data flow, equipment, and so forth?

Q-20. What type of flowchart is constructed by the programmer to represent the sequence of operations the computer is to perform to solve a specific problem?

Q-21. How many tools are used in flowcharting?

Q-22. Is there a "best way" to construct a flowchart?

Q-23. What controls the computer during processing?

Q-24. What is the fundamental element in program preparation?

Q-25. What type of instructions permit addition, subtraction, multiplication, and division?

Q-26. Where is specific information about the computer you are to use contained?

Q-27. How do we refer to errors caused by faulty logic and coding mistakes?

Q-28. What is the purpose of testing a program?

PACKAGED SOFTWARE

Fortunately you don't have to write a program for every problem to be solved. Instead, you can use packaged or off-the-shelf programs that are designed for specific classes of applications. Everyday more and more packaged software (software written by the manufacturer, a software house, or central design agency [CDA]) becomes available for general use. It may be up to you to set up and process a job within the specifications of a packaged program. Let's look at four classes of packaged software you may work with: word processing, data management, spreadsheets, and graphics.

WORD PROCESSING

You can use word processing software for any function that involves text: letters, memos, forms, reports, and so on. At a minimum, it includes routines for creating, editing, storing, retrieving, and printing text. Under the word processing software control, you generally enter the text on the keyboard and it is printed on a display screen as shown in figure 3-8. At that point, you may store it on disk or tape, print it on a printer, or change (edit) it. Using the edit functions you can add or delete words, characters, lines, sentences, or paragraphs. You can rearrange text; for example, move a paragraph or block of information to another place in the same document or even move it to a different document. Word processing is particularly useful for text documents that are repetitive or that require a lot of revisions. It saves a lot of rekeying.

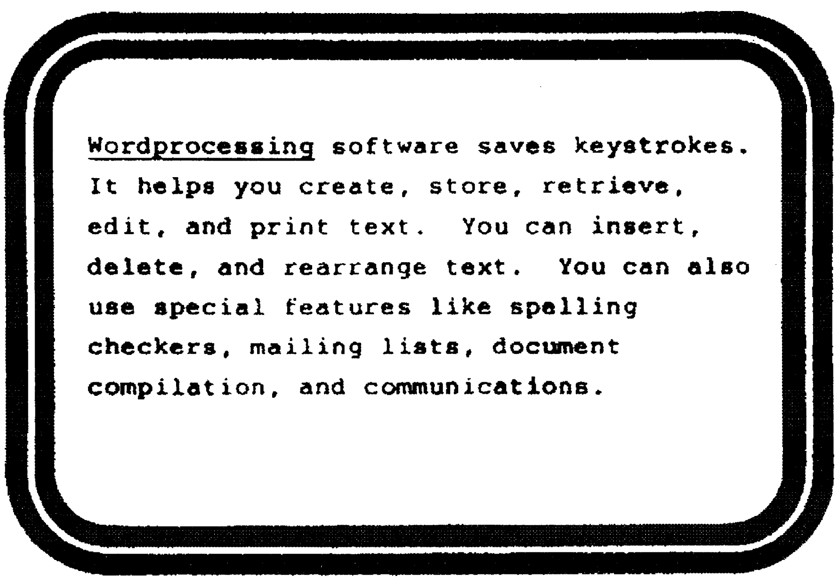


Figure 3-8.—Word processing example.

Other features and software often available with a word processing software package include: spelling checkers, mailing list programs, document compilation programs, and communications programs.

Spelling checker software helps find misspelled words but not misused words. It scans the text matching each word against a dictionary of words. If the word is not found in the dictionary, the system flags the word. You check it. If it is misspelled, you can correct it. You will still have to proofread the document to see that everything was keyed and that the words are used correctly.

Mailing list programs are for maintaining name and address files. They often include a capability to individualize letters and reports by inserting names, words, or phrases to personalize them.

Document compilation programs are useful when you have standard paragraphs of information that you need to combine in different ways for various purposes. For example, you may be answering inquiries or putting together contracts or proposals. Once you select the standard paragraphs you want, you add variable information. This saves both keying and proofreading time.

Communications software and hardware enable you to transmit and receive text on your microcomputer. Many organizations use this capability for electronic mail. In a matter of minutes you can enter and transmit a memo to other commands or to personnel in other locations. You can transmit monthly reports, notices, or any documents prepared on the microcomputer.

DATA MANAGEMENT

Data management software allows you to enter data and then retrieve it in a variety of ways. You define your data fields and set up a display screen with prompts. You enter the data records according to the prompts. Figure 3-9, view A, shows an example. The system writes the records on a disk or tape. Once you have a file keyed and stored, you can retrieve records by a field or several fields or by searching the records for specific data. For example, if you wanted a list of all personnel who reported aboard before January 1988, you could tell the system to search the file and print selected fields of all records that meet that condition. You tell the system what fields to print (that is name, rate, SSN, date reported) and where (what print positions) to print them. At the same time, you can specify in what order you want the records printed. For example, figure 3-9, view B, shows the records printed in alphabetical order by last name. The software also provides routines so you can easily add, delete, and change records.

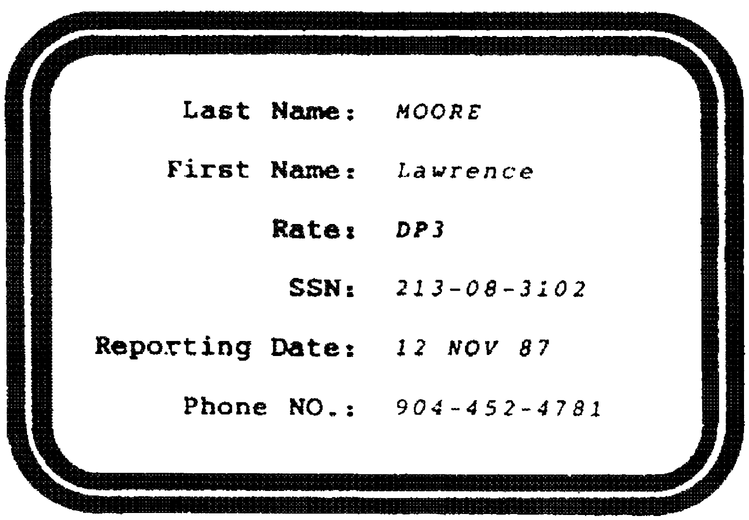


Figure 3-9A.—Data management example. PROMPTS (IN BOLD) AND DATA (IN ITALICS).

RATE	NAME LAST	FIRST	SSN	REPORTING DATE
DS2	FREDERICK	Jan	712-36-4279	5 SEP 87
DP2	HOWARD	Al	123-07-0736	15 DEC 87
DP3	MOORE	Lawrence	213-08-3102	12 NOV 87
:				
ET2	RAY	William	219-47-3261	5 NOV 87
:				

Figure 3-9B.—Data management example. SAMPLE PRINTED REPORT (SORTED BY LAST NAME).

You can also generate reports by specifying what records to use, what fields to print, where to print the fields, and which data fields, if any, need to be combined. For example, your supply officer wants to know the value of the inventory. You can specify that the extended price is to be calculated by multiplying the item quantity by the unit price, and that the extended prices are to be totaled.

INVENTORY VALUE			
ITEM	QUANTITY	UNIT PRICE	EXTENDED PRICE
Swabs	47	1.65	77.55
Brooms	62	2.25	139.50
Foxtails	36	1.85	66.60
TOTAL			283.65

You can also specify the information to be used in report and column headings.

While the data management programs on micros are not as sophisticated as the data base management systems on mainframes and minis, they do provide an extremely useful capability in offices or aboard ship.

SPREADSHEETS

Spreadsheets are tables of rows and columns of numbers. Figure 3-10 shows an example. Spreadsheet processors allow you to set up a table of rows and columns and specify what calculations to perform on the columns. You enter values for the basic information into the appropriate rows and columns. Then the processor performs the calculations. In our example (fig. 3-10), we used a spreadsheet to project magnetic media costs. You enter the item descriptions, column headings, report title, and data for columns 1, 2, and 4, and the software calculates column 3 by adding columns 1 and 2. Then it multiplies column 3 times column 4 and puts the result in column 5. It also subtotals and totals the columns you specify; in this case, columns 1 through 3 and column 5.

MAGNETIC MEDIA REQUIREMENTS
SPREAD SHEET

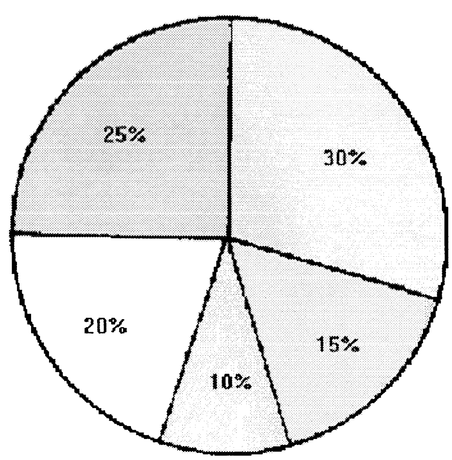
ITEM	NUMBER TO BE REPLACED (1)	NUMBER FOR EXPANSION (2)	TOTAL NEEDED (3)	COST PER ITEM (4)	TOTAL COST (5)
TAPES	*15*	*30*	**45**	*27.50*	**1237.50**
DISKS	*4*	*5*	**9**	*150.00*	**3150.00**
DISKETTES					
5"	*10*	*30*	**40**	*1.25*	**70.00**
3 1/4"	*10*	*50*	**60**	*1.50*	**74.00**
SUBTOTAL	**20**	**80**	**100**		**148.00**
TOTAL	**39**	**115**	**154**		**4535.50**

ITALICS -- DATA YOU ENTER
BOLD -- INFORMATION CALCULATED BY THE PROGRAM

Figure 3-10.—Spreadsheet example.

GRAPHICS

Graphics capability is available on many microcomputers. One use is to produce data displays, like bar charts, pie charts, and graphs. See figure 3-11, view A and view B. On some micros, you can do line drawings; on others you can create sophisticated engineering drawings. High resolution color graphics are also available for specialized applications.



A. PIE CHART

Figure 3-11A.—Graphics examples. PIE CHART.

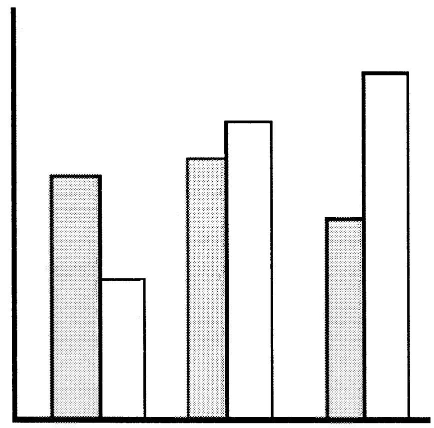


Figure 3-11B.—Graphics examples. BAR CHART.

You cannot use all printers for graphics output. They must be capable of producing graphics and also be compatible with the software. Some character printers can be used for limited graphics. Dot-matrix printers and plotters work well for graphics output. Laser and ink jet printers are also good for both text and graphics.

Q-29. What is packaged software?

Q-30. What are some of the other features and software available with a word processing software package?

Q-31. What software allows you to enter data and then retrieve it in a variety of ways?

Q-32. What are spreadsheets?

Q-33. Are all printers capable of handling graphics output?

SUMMARY

Congratulations you have just finished chapter 3. By now you should be convinced that anyone, with a little study, can understand digital computers. You probably thought when you first started this *NEETS* MODULE that it would get more difficult as your studies progressed. Our objective was to show you that the more you learn, the easier the material is to understand.

OPERATING SYSTEMS are a collection of many programs used by the computer to manage its own resources and operations and to perform commonly used functions like copy, print, and so on.

UTILITY PROGRAMS perform such tasks as sorting, merging, and transferring (copying) data from one input/output device to another: card to tape, tape to tape, tape to disk, and so on.

SORT-MERGE PROGRAMS arrange data records in a predefined sequence or order and are capable of combining two or more ordered files into one file.

REPORT PROGRAM GENERATORS are used to generate programs to print detail and summary reports of data files.

PROGRAMMING LANGUAGES are the means by which human-computer communication is achieved. They are used to write the instructions to tell the computer what to do to solve a given problem.

A **MACHINE LANGUAGE** uses a string of numbers that represent the instruction codes and operand addresses to tell the computer what to do.

SYMBOLIC LANGUAGES improved the program preparation process by substituting letter symbols (mnemonic codes) for basic machine language instruction codes.

A **PROCEDURE ORIENTED LANGUAGE** is a programming language oriented toward a specific class of processing problems. Examples are BASIC, COBOL, and FORTRAN.

PROGRAMMING is the process of planning and coding the computer instructions to solve a problem.

FLOWCHARTING is one method of pictorially representing a procedural (step-by-step) solution to a problem before you actually start to write the computer instructions required to produce the desired results.

PACKAGED SOFTWARE is designed for specific classes of applications. Examples are word processing, spreadsheets, data management, and graphics. These off-the-shelf programs are written by the manufacturer, a software house, or a central design agency.

ANSWERS TO QUESTIONS Q1. THROUGH Q33.

A-1. *The operating system.*

A-2. *Single user/single tasking.*

A-3. *Multiuser/multitasking.*

A-4. *Because, to use applications software, it must be compatible with the operating system.*

A-5. *Some software comes in several versions so it can run under several different operating systems.*

A-6. *Utility programs.*

A-7. *By providing information about files, data fields, and the process to be used.*

A-8. *Sorting.*

A-9. *What data field or fields to sort on.*

A-10. *To generate programs to print detail and summary reports of data files.*

A-11. *Machine.*

A-12. *In the early 1950's.*

A-13. *The development of mnemonic techniques and macroinstructions.*

A-14. *FORTRAN.*

A-15. They require more space in memory and they process data at a slower rate than symbolic languages.

A-16. The process of planning the solution to a problem.

A-17. Four.

A-18. Advance preparation.

A-19. Different shaped symbols.

A-20. A programming flowchart.

A-21. Four.

A-22. No, there isn't a way to standardize problem solution.

A-23. Coded instructions.

A-24. The instruction.

A-25. Arithmetic.

A-26. In the computer manufacturers or software designers user's manual.

A-27. Bugs.

A-28. To determine that all data can be processed correctly and that the output is correct.

A-29. Off-the-shelf programs designed for specific classes of applications.

A-30. Spelling checkers, mailing list programs, document compilation programs, and communications programs.

A-31. Data management.

A-32. They are tables of rows and columns of numbers.

A-33. No.

CHAPTER 4

DATA REPRESENTATION AND COMMUNICATIONS

LEARNING OBJECTIVES

Upon completion of this chapter, you will be able to do the following:

1. Explain data and how it is represented.
2. Explain computer coding systems.
3. Define a parity bit and what it is used for.
4. Explain data storage concepts.
5. Describe three storage access methods.
6. Describe networks and data communications.

INTRODUCTION

One of the major problems we face in using a digital computer is communicating with it. We must have one or more ways of getting data into the computer to be processed. You learned in chapter 2 that there are several types of input devices that read data into a computer. But how does one prepare the data to be used as input? How do we convert human-readable documents into a computer-readable form, and what type of input media do we use? If the data is to be used by another computer some distance away, how do we transmit it? Well, as you probably suspect, there are several ways to perform this conversion and transmission process, and that is the chapter of our discussion.

DATA

Data is a general term used to describe raw facts. To put it simply, data is nothing more than a collection of related elements or items, that when properly coded into some type of input medium, can be processed by a computer. Data items might include your service number, your name, your paygrade, or any other fact. Until some meaning has been given to the data, nothing can really be determined about it; therefore, it remains data. When this data has been processed together with other facts, it then has meaning and it becomes information we can understand and properly use.

DATA REPRESENTATION

Data is represented by symbols. Symbols convey meaning only when understood. The symbol itself is not the information, but merely a representation of it. Symbol meaning is one of convention (fig. 4-1). Symbols may convey one meaning to you and me, another meaning to others, and no meaning at all to those that do not know their significance. Data must be reduced to a set of symbols that the computer can read and interpret before there can be any communication with the computer. The first computers were designed to manipulate numbers to solve arithmetic problems. But as you can see in figure 4-1, we create, use, and manipulate many other symbols to represent facts in the world in which we live. We are fortunate that early computer experts soon realized the need to manipulate nonnumerical symbols as well. Manipulating these symbols is possible if an identifying code or coded number is assigned to the symbol to be stored and processed. Thus, the letters in a name such as ALBERT or CAROL can be represented by different codes, as can all special characters, such as #, (,), &, $, @, and yes, even the comma. The data to be represented is called source data.

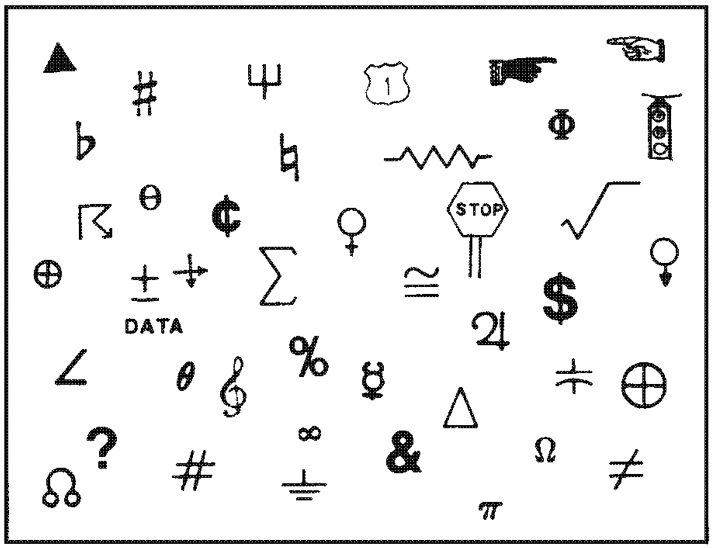


Figure 4-1.—Communications symbols.

SOURCE DATA

Source data or raw data is typically written on some type of paper document, which we refer to as a source document. The data contained on the source document must be converted into a machine-readable form for processing either by direct or indirect means. The data may be entered directly into the computer in its original form; namely right from the source document on which it is recorded by way of magnetic ink characters, optically recognizable characters, or bar code recognition. Or the data on the documents may be entered indirectly on input media, such as punched cards, paper tape, magnetic tape, or magnetic disk. It may also be keyed directly into a computer from a keyboard.

If you look at figure 4-2, you see a list of SERVMART items that have been typed on a preprinted form. To most people this is just another piece of paper; however, to the Storekeeper (SK) it is a source document to be used to provide input data to the computer. In this example, the SERVMART form deals with requisitioning supplies. The form could be sent to the data-entry department to be used as a source document. There the data-entry operator can key the data into or on whatever computer medium is to be used, according to a prescribed format. The data elements are numbered in the order they are to be keyed: (1) document identification, (2) stock number, (3) unit of issue, (4) quantity, and so on. You'll notice we need more than numbers, and that is where coding systems come into play.

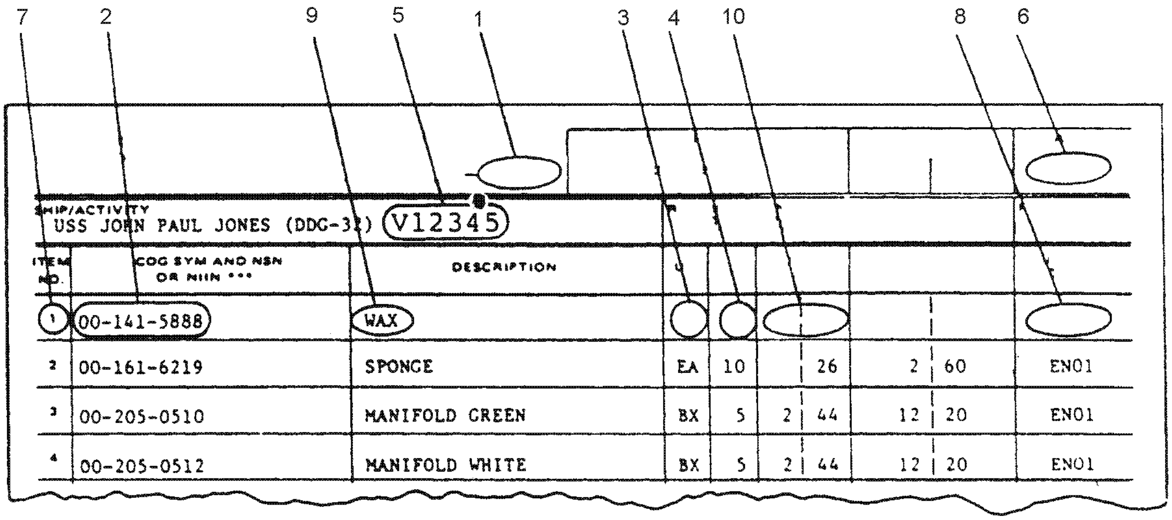


Figure 4-2.—SERVMART shopping list (source document).

Q-1. What is a general term used to describe raw facts?

Q-2. How is data represented?

Q-3. What were the first computers designed to manipulate in order to solve arithmetic problems?

Q-4. By what two means can the data contained on a source document be converted into a machine-readable form for processing?

Q-5. What are some of the types of input media on which data may be indirectly entered?

COMPUTER CODING SYSTEMS

To represent numeric, alphabetic, and special characters in a computer's internal storage and on magnetic media, we must use some sort of coding system. In computers, the code is made up of fixed size groups of binary positions. Each binary position in a group is assigned a specific value; for example 8, 4, 2, or 1. In this way, every character can be represented by a combination of bits that is different from any other combination.

In this section you will learn how the selected coding systems are used to represent data. The coding systems included are Extended Binary Coded Decimal Interchange Code (EBCDIC), and American Standard Code for Information Interchange (ASCII).

EXTENDED BINARY CODED DECIMAL INTERCHANGE CODE (EBCDIC)

Using an 8-bit code, it is possible to represent 256 different characters or bit combinations. This provides a unique code for each decimal value 0 through 9 (for a total of 10), each uppercase and lowercase letter (for a total of 52), and for a variety of special characters. In addition to four numeric bits, four zone bit positions are used in 8-bit code as illustrated in figure 4-3. Each group of the eight bits makes up one alphabetic, numeric, or special character and is called a byte.

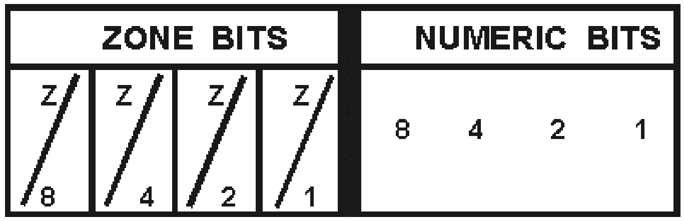


Figure 4-3.—Format for EBCDIC and ASCII codes.

When you look at figure 4-3, you will notice that the four rightmost bits in EBCDIC are assigned values of 8, 4, 2, and 1. The next four bits to the left are called the zone bits. The EBCDIC coding chart for uppercase and lowercase alphabetic characters and for the numeric digits 0 through 9 is shown in figure 4-4, with their hexadecimal equivalents. Hexadecimal is a number system used with some computer systems. It has a base of 16 (0-9 and A-F). A represents 10; B represents 11; C represents 12; D represents 13; E represents 14; and F represents 15. In EBCDIC, the bit pattern 1100 is the zone combination used for the alphabetic characters A through I, 1101 is used for the characters J through R, and 1110 is the zone combination used for characters S through Z. The bit pattern 1111 is the zone combination used when representing decimal digits. For example, the code 11000001 is equivalent to the letter A; the code 11110001 is equivalent to the decimal digit 1. Other zone combinations are used when forming special characters. Not all of the 256 combinations of 8-bit code have been assigned characters. Figure 4-5 illustrates how the characters DP-3 are represented using EBCDIC.

ALPHABETIC CHARACTERS					
UPPERCASE			LOWERCASE		
PRINTS AS	EBCDIC IN BINARY	EBCDIC IN HEXA-DECIMAL	PRINTS AS	EBCDIC IN BINARY	EBCDIC IN-HEXA DECIMAL
	ZZZZ 8421			ZZZZ 8421	
A	1100 0001	C 1	a	1000 0001	8 1
B	1100 0010	C 2	b	1000 0010	8 2
C	1100 0011	C 3	c	1000 0011	8 3
D	1100 0100	C 4	d	1000 0100	8 4
E	1100 0101	C 5	e	1000 0101	8 5
F	1100 0110	C 6	f	1000 0110	8 6
G	1100 0111	C 7	g	1000 0111	8 7
H	1100 1000	C 8	h	1000 1000	8 8
I	1100 1001	C 9	i	1000 1001	8 9
J	1101 0001	D 1	j	1001 0001	9 1
K	1101 0010	D 2	k	1001 0010	9 2
L	1101 0011	D 3	l	1001 0011	9 3
M	1101 0100	D 4	m	1001 0100	9 4
N	1101 0101	D 5	n	1001 0101	9 5
O	1101 0110	D 6	o	1001 0110	9 6
P	1101 0111	D 7	p	1001 0111	9 7
Q	1101 1000	D 8	q	1001 1000	9 8
R	1101 1001	D 9	r	1001 1001	9 9
S	1110 0010	E 2	s	1010 0010	A 2
T	1110 0011	E 3	t	1010 0011	A 3
U	1110 0100	E 4	u	1010 0100	A 4
V	1110 0101	E 5	v	1010 0101	A 5
W	1110 0110	E 6	w	1010 0110	A 6
X	1110 0111	E 7	x	1010 0111	A 7
Y	1110 1000	E 8	y	1010 1000	A 8
Z	1110 1001	E 9	z	1010 1001	A 9
NUMERIC CHARACTERS					
0	1111 0000	F 0	5	1111 0101	F 5
1	1111 0001	F 1	6	1111 0110	F 6
2	1111 0010	F 2	7	1111 0111	F 7
3	1111 0011	F 3	8	1111 1000	F 8
4	1111 0100	F 4	9	1111 1001	F 9

Figure 4-4.—Eight-bit EBCDIC coding chart (including hexadecimal equivalents).

D	P	—	3
1100 0100	1101 0111	0110 0000	1111 0011

Figure 4-5.—DP-3 represented using 8-bit EBCDIC code.

Since one numeric character can be represented and stored using only four bits (8-4-2-1), using an 8-bit code allows the representation of two numeric characters (decimal digits) as illustrated in figure 4-6. Representing two numeric characters in one byte (eight bits) is referred to as packing or packed data. By packing data (numeric characters only) in this way, it allows us to conserve the amount of storage space required, and at the same time, increases processing speed.

DECIMAL VALUE	9	2	7	3
EBCDIC CODE	1001	0010	0111	0011
BIT PLACE VALUES	8421	8421	8421	8421
	BYTE 1		BYTE 1	

Figure 4-6.—Packed data.

AMERICAN STANDARD CODE FOR INFORMATION INTERCHANGE (ASCII)

Another 8-bit code, known as the American Standard Code for Information Interchange (ASCII) (pronounced ASS-KEY), was originally designed as a 7-bit code. Several computer manufacturers cooperated to develop this code for transmitting and processing data. The purpose was to standardize a binary code to give the computer user the capability of using several machines to process data regardless of the manufacturer: IBM, HONEYWELL, UNIVAC, BURROUGHS, and so on. However, since most computers are designed to handle (store and manipulate) 8-bit code, an 8-bit version of ASCII was developed. ASCII is commonly used in the transmission of data through data communications and is used almost exclusively to represent data internally in microcomputers.

The concepts and advantages of ASCII are identical to those of EBCDIC. The important difference between the two coding systems lies in the 8-bit combinations assigned to represent the various alphabetic, numeric, and special characters. When using ASCII 8-bit code, you will notice the selection of bit patterns used in the positions differs from those used in EBCDIC. For example, let's look at the characters DP3 in both EBCDIC and ASCII to see how they compare.

Character	D	P	3
EBCDIC	1100 0100	1101 0111	1111 0011
ASCII	0100 0100	0101 0000	0011 0011

In ASCII, rather than breaking letters into three groups, uppercase letters are assigned codes beginning with hexadecimal value 41 and continuing sequentially through hexadecimal value 5A. Similarly, lowercase letters are assigned hexadecimal values of 61 through 7A. The decimal values 1 through 9 are assigned the zone code 0011 in ASCII rather that 1111 as in EBCDIC. Figure 4-7 is the ASCII coding chart showing uppercase and lowercase alphabetic characters and numeric digits 0 through 9.

ALPHABETIC CHARACTERS					
UPPERCASE			LOWERCASE		
PRINTS AS	ASCII CODE IN BINARY	ASCII CODE IN HEXA-DECIMAL	PRINTS AS	ASCII CODE IN BINARY	ASCII CODE IN HEXA-DECIMAL
	8421 8421			8421 8421	
A	0100 0001	4 1	a	0110 0001	6 1
B	0100 0010	4 2	b	0110 0010	6 2
C	0100 0011	4 3	c	0110 0011	6 3
D	0100 0100	4 4	d	0110 0100	6 4
E	0100 0101	4 5	e	0110 0101	6 5
F	0100 0110	4 6	f	0110 0110	6 6
G	0100 0111	4 7	g	0110 0111	6 7
H	0100 1000	4 8	h	0110 1000	6 8
I	0100 1001	4 9	i	0110 1001	6 9
J	0100 1010	4 A	j	0110 1010	6 A
K	0100 1011	4 B	k	0110 1011	6 B
L	0100 1100	4 C	l	0110 1100	6 C
M	0100 1101	4 D	m	0110 1101	6 D
N	0100 1110	4 E	n	0110 1110	6 E
O	0100 1111	4 F	o	0110 1111	6 F
P	0101 0000	5 0	p	0111 0000	7 0
Q	0101 0001	5 1	q	0111 0001	7 1
R	0101 0010	5 2	r	0111 0010	7 2
S	0101 0011	5 3	s	0111 0011	7 3
T	0101 0100	5 4	t	0111 0100	7 4
U	0101 0101	5 5	u	0111 0101	7 5
V	0101 0110	5 6	v	0111 0110	7 6
W	0101 0111	5 7	w	0111 0111	7 7
X	0101 1000	5 8	x	0111 1000	7 8
Y	0101 1001	5 9	y	0111 1001	7 9
Z	0101 1010	5 A	z	0111 1010	7 A

NUMERIC CHARACTERS

0	0011 0000	3 0	5	0011 0101	3 5
1	0011 0001	3 1	6	0011 0110	3 6
2	0011 0010	3 2	7	0011 0111	3 7
3	0011 0011	3 3	8	0011 1000	3 8
4	0011 0100	3 4	9	0011 1001	3 9

Figure 4-7.—Eight-bit ASCII coding chart (including hexadecimal equivalents).

At this point you should understand how coding systems are used to represent data in both EBCDIC and ASCII. Regardless of what coding system is used, each character will have an additional bit called a check bit or parity bit.

PARITY BIT

This additional check or parity bit in each storage location is used to detect errors in the circuitry. Therefore, a computer that uses an 8-bit code, such as EBCDIC or ASCII, will have a ninth bit for parity checking.

The parity bit (also called a check bit, the C position in a code) provides an internal means for checking the validity, the correctness, of code construction. That is, the total number of bits in a character, including the parity bit, must always be odd or always be even, depending upon whether the particular computer system or device you are using is odd or even parity. Therefore, the coding is said to be in either odd or even parity code, and the test for bit count is called a parity check.

Now, let's talk about bits and bytes, primary storage, and storage capacities; or, to put it another way, the capacity of a storage location. Sit back, keep your memory cycling, and we will explain the ways data may be stored and retrieved inside the computer.

Q-6. What does the acronym EBCDIC stand for?

Q-7. By using an 8-bit code, how many characters or bit combinations can be represented?

Q-8. What is the base of a hexadecimal number system?

Q-9. What term is used for the representation of two numeric characters stored in eight bits?

Q-10. What does the acronym ASCII mean?

Q-11. What was the purpose of several computer manufacturers cooperating to develop ASCII code for processing and transmitting data?

Q-12. Are there any differences in the concepts and advantages of ASCII and EBCDIC?

Q-13. How is the parity bit in each storage location used?

Q-14. A computer or device that uses 8-bit ASCII or EBCDIC will use how many bits to store each character?

DATA STORAGE CONCEPTS

You learned in chapter 2 that a computer's primary storage area is divided into four areas, each serving a specific purpose. The input storage area accepts and holds input data to be processed. The working storage area holds intermediate processing results. The output storage area holds the final processing results. The program storage area holds the processing instructions (the program). You also learned that these separate areas do not have built-in physical boundaries, rather the boundaries are determined by the individual programs being used.

You also may recall in chapter 2, we talked about the different types of primary storage used in computers and how they differ from one another. Some were magnetic in nature, such as magnetic core storage; others were electronic, such as semiconductor and bubble storage. For purposes of simplicity, we have selected magnetic core storage to show you how data is represented and stored in the computer's primary memory.

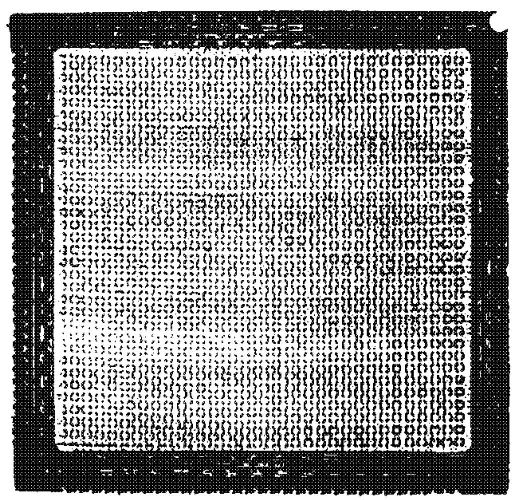
BITS AND BYTES

A bit is a single binary digit. It represents the smallest unit of data just like the good old American penny). However, computers usually do not operate on single bits, rather they store and manipulate a fixed number of bits. Most often, the smallest unit or number of bits a computer works with is eight bits. These eight bits make up a byte. You just learned that both EBCDIC and ASCII codes use eight bits (excluding the parity bit), and that eight bits represent a single character, such as the letter A or the number 7. Thus, the computer can store and manipulate an individual byte (a single character) or a group of bytes (several characters, a word) at a time. These individual bytes, or groups of bytes, form the basic unit of memory.

Primary storage capacities are usually specified in number of bytes. The symbol "K" is used whenever we refer to the size of memory, especially when the memory is quite large. The symbol K is equal to 1,024 units or positions of storage. Therefore, if a computer has 512K bytes (not bits) of primary storage, then it can hold 512 × 1,024 or 524,288 characters (bytes) of data in its memory.

MAGNETIC CORE STORAGE

In primary storage, many magnetic cores are strung together on a screen of wire to form what is called a core plane (fig. 4-8, view A). As you may know, each core can store one binary bit (0 or 1) of data. A core is magnetized by current flowing through the wires on which the core is strung. Hence, a core magnetized in one direction represents a binary 0, and when magnetized in the opposite direction, a binary 1. It is the direction that the core is magnetized that determines whether it contains a binary 0 or a binary 1 (refer to fig. 4-8, view B). These core planes look very much like small window screens and are arranged vertically to represent data as shown in figure 4-8, view C. In looking at this figure, you will notice that nine planes are needed to code in 8-bit EBCDIC. The ninth plane provides for a parity (check) bit. Figure 4-8, view C, shows DP-3 in EBCDIC code, even parity.



A. ONE OF MANY CORE PLANES THAT MAKE UP PRIMARY STORAGE.

Figure 4-8A.—Core storage with DP-3 represented using 8-bit EBCD1C code.

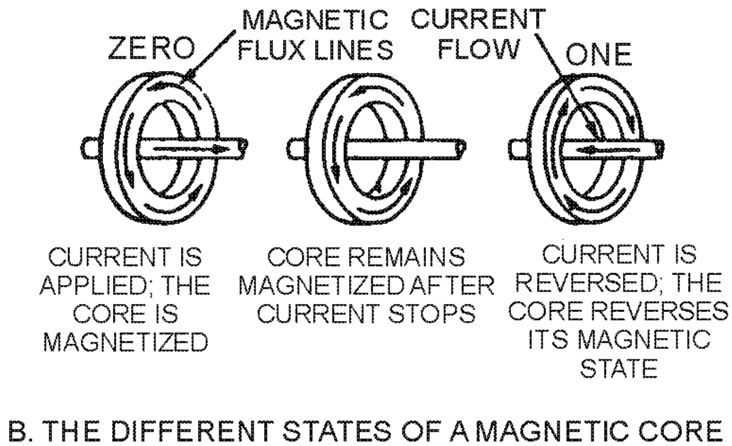


Figure 4-8B.—Core storage with DP-3 represented using 8-bit EBCD1C code.

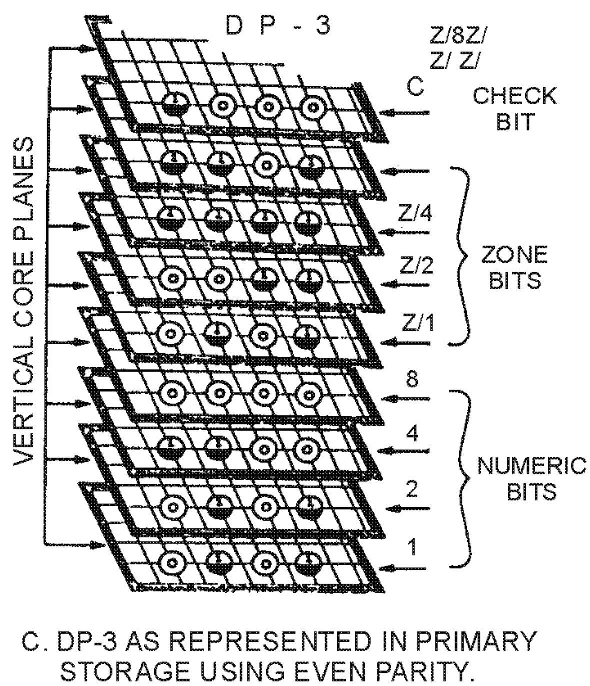


Figure 4-8C.—Core storage with DP-3 represented using 8-bit EBCD1C code.

STORAGE CAPACITY AND ADDRESSES

The storage capacity of an address is designed and built into the computer by the manufacturer. Over the years several different design approaches to partition primary storage have been used. With this in mind, let's take a look at some of the ways primary storage is partitioned into addresses.

One way to design or organize the primary storage section is to store a fixed number of characters (bytes) at each address location. We can then reference these characters as a single entity called a word, as illustrated in figure 4-9, view A. The name CHARLIE (address location 400) or the amount he is paid, in this case $69.00 (address location 401), are each treated as a single word. Computers that are built to retrieve, manipulate, and store a fixed number of characters in each address are said to be word-oriented, word-addressable machines, or fixed-word-length computers.

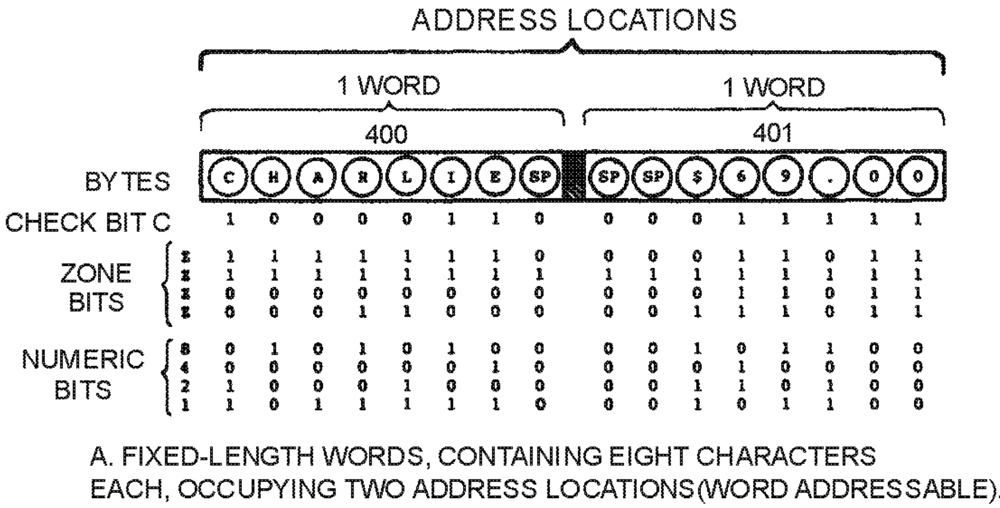


Figure 4-9A.—Fixed-word-length vs variable-word-length storage. FIXED-LENGTH WORDS, CONTAINING EIGHT CHARACTERS EACH, OCCUPYING TWO ADDRESS LOCATIONS (WORD ADDRESSABLE).

Another way to design the primary storage section is to store a single character, such as the letter L or the number 8, in each address location. An address is assigned to each location in storage. Computers designed in this way are said to be character-oriented or character addressable. We also call them variable-word-length computers. Therefore, the name CHARLIE (fig. 4-9, view B) now requires seven address locations (300 through 306), while amount paid ($69.00) occupies six address locations (307 through 312).

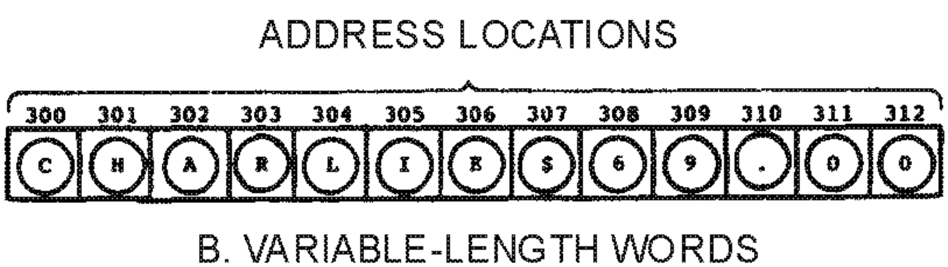


Figure 4-9B.—Fixed-word-length vs variable-word-length storage. VARIABLE-LENGTH WORDS (CHARACTER ADDRESSABLE).

Whether a computer addresses a group of bytes as a word or addresses each byte individually is a function of the circuitry. Both designs have advantages and disadvantages. Variable-word-length computers make the most efficient use of available storage space, since a character can be placed in every storage location. In a fixed-word-length computer, storage space may be wasted. For example, if the storage capacity in each address of a fixed-word-length computer is eight bytes, and some of the data elements to be stored contain only three or four characters, then many of the storage positions in each word are not being used.

Fixed-word-length computers have faster calculating speeds. They can add two data words in a single operation. This is not so with character-addressable computers. Here, only one digit (byte) in each number can be added during a single machine operation. Thus, eight steps are required to complete the calculation.

The larger mainframe computers (super-computers like the CRAY-1 and CYBER 205) use only fixed-word-length storage. Most microcomputers use the variable-word-length approach allowing them to operate on one character at a time. Somewhere in between these two extremes are the dozens of existing minicomputer and mainframe models that have what is called built-in flexibility.

These flexible computers are byte-oriented but can operate in either a fixed- or variable-word-length mode through the use of proper program instructions. Let's take a look at how these flexible computers operate in a variable- and fixed-word-length environment.

Working in a variable-word-length environment, each address holds one alphanumeric character as shown in figure 4-9, view B. Since a byte usually represents a single alphanumeric character, unless you are using packed decimal, a flexible computer is often said to be byte-addressable. Don't become confused; the terms character-addressable, character-oriented, and byte-addressable all have the same meaning. By using the appropriate program instructions, a programmer can retrieve a stored data element by identifying the address of the first character (say position 300 as in fig. 4-9, view B) and specifying the number of address locations to be included in the word. In this case there are seven, positions 300 through 306.

When a flexible computer is working in a fixed-word-length environment, each address identifies a group of bytes that can be operated on as a unit. This processing method helps to achieve faster calculating speeds. A programmer can use program instructions to cause the computer to automatically retrieve, manipulate, and store, as a unit, a fixed word of say, two, four, or eight bytes of data in one machine operation by identifying the address of the first character of data. At the same time all remaining bytes are acted upon as a unit moving from left to right. Figure 4-10 illustrates the different word lengths possible with many byte-addressable computers. They are half-word (2 bytes), full-word (4 bytes), and double-word (8 bytes).

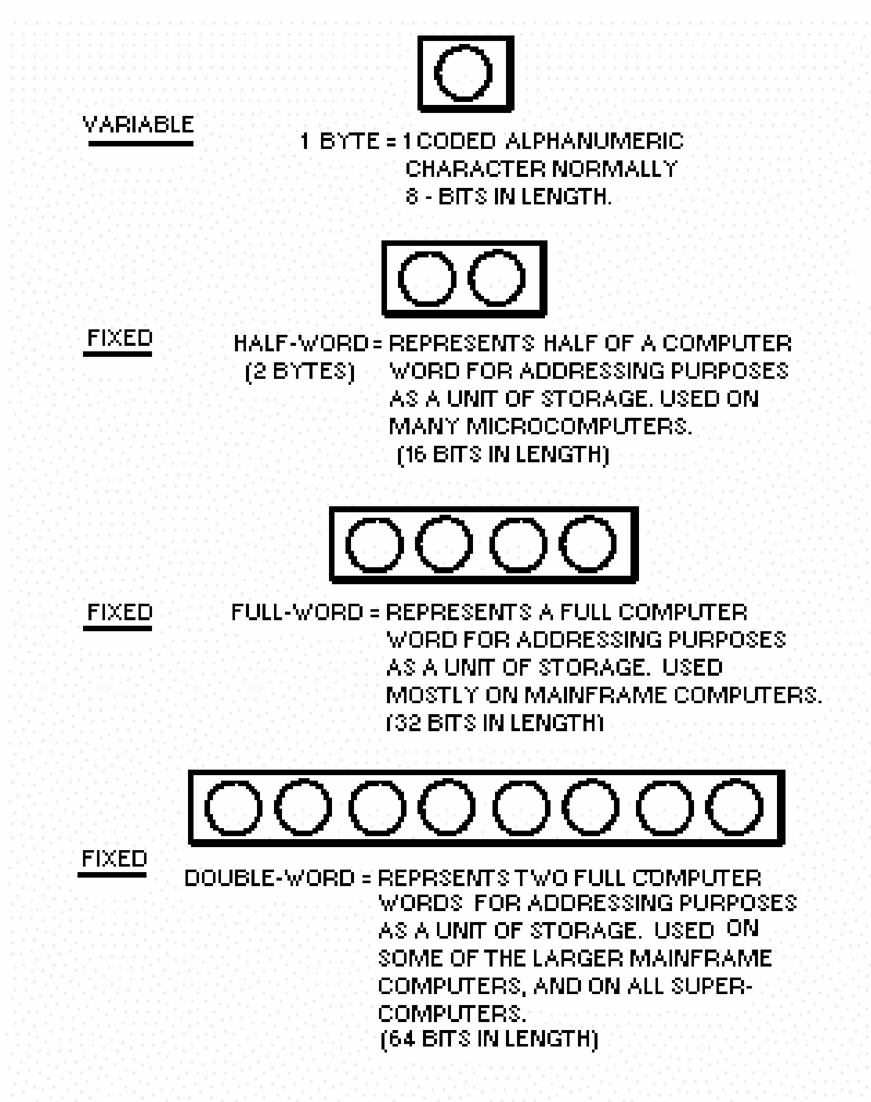


Figure 4-10.—Word lengths used on flexible byte-addressable computers.

By now, you should have a good idea of how primary storage locations are identified by their storage addresses, how these addressable storage locations are used, and how the storage capacity at an address can vary depending on the design of the computer.

Now, let's go one step further, to see how these bits and bytes are represented (coded) on some of the more common secondary storage media.

SECONDARY STORAGE DATA ORGANIZATION

Remember, secondary storage devices (also called auxiliary or mass storage devices) are those devices which are not part of the central processing unit (cpu). They include: external core; semiconductor, thin film, and bubble memories; punched cards; paper tape; and several different types of mass storage, such as magnetic tape, disk, and drum.

You already know it takes a certain number of bits to make one byte (normally eight), and when bytes are grouped together at a single address they make up a word in the computer's memory. When data is recorded on some type of magnetic storage medium, such as disk or tape, it is normally organized by bits, characters (bytes), fields, records, and files (fig. 4-11). The following definitions should help you understand the relationship between bits, characters, bytes, words, fields, records, and files.

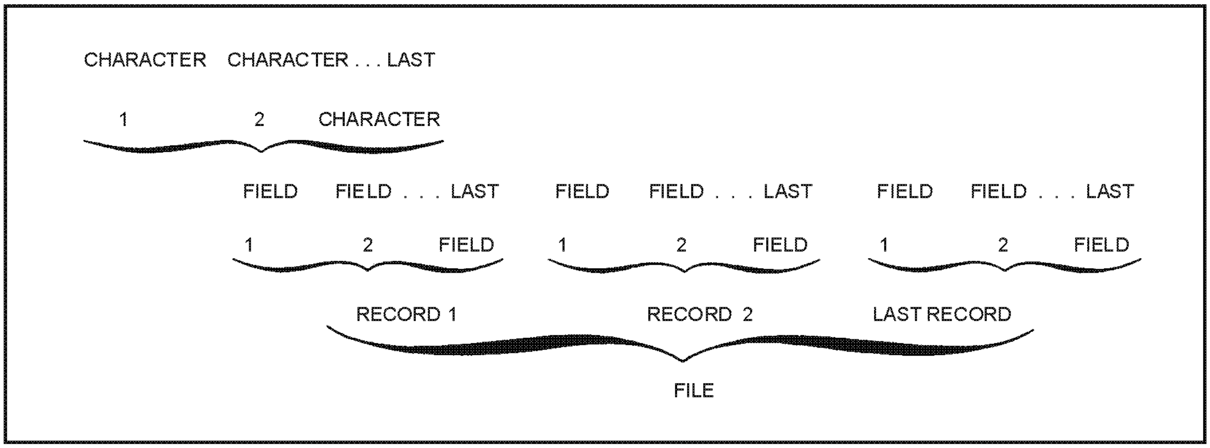


Figure 4-11.—Data organization.

BIT—The smallest unit of data; it represents one binary digit (0 or 1).

CHARACTER (BYTE)—A group of related bits (usually eight) that make up a single character-letter, number, or special character.

WORD—A group of related bytes that are treated as a single addressable unit or entity in memory.

FIELD—One or more related characters that are treated as a unit of information. A field (also referred to as a data item) may be alphabetic, numeric, or alphanumeric, and may be either fixed or variable in length. For example, your social security number (SSN) is of a fixed length; that is, it's always 9 positions in length. Whereas, names are variable length because they may be from 2 to 25 positions in length.

RECORD—A group of related fields, all pertaining to the same subject; a person, a thing, or an event. For example, your payroll record (LES statement) might include fields for your name, amount paid, taxes withheld, earned leave, and any allotments you might have. On the other hand, a supply inventory record might consist of fields containing stock number, the name of the item, its unit price, the quantity on hand, and its bin location.

FILE—A collection of related records, such as the payroll or supply inventory records. Normally, all records within the file are in the same format.

When processing data, we think in terms of data files. For example, to process a parts inventory, you would need the master parts inventory file and the file that contains up-to-date information on each part that has been issued. The master parts inventory file would have a record for every part in the inventory. The update file, parts issued file, would have a record for each part issued. You would use a program to read the records on the parts issued file and update the matching records on the master parts inventory file. Depending on whether the data is stored on magnetic tape or disk or in internal storage, the program would use different methods to access storage to obtain the data. In the next section you'll learn about storage access methods.

Q-15. What area in the computer's primary storage area holds the processing instructions (the program)?

Q-16. How are the boundaries determined for the separate areas of the computer's primary storage area?

Q-17. What is a bit?

Q-18. How many bits make up a byte?

Q-19. Primary storage capacities are usually specified in what unit of measure?

Q-20. How are core planes formed?

Q-21. Where are core planes used?

Q-22. Who designs and builds the storage capacity of an address into a computer?

Q-23. What is another name for computers designed to be character-oriented or character-addressable?

Q-24. Which computer has the faster calculating speeds, the variable-word-length or the fixed-word-length?

Q-25. What is the normal organization of data recorded on magnetic storage media?

Q-26. What is a file?

STORAGE ACCESS METHODS

How data files are stored in secondary storage varies with the types of media and devices you are using. Data files may be stored on or in sequential-access storage, direct-access storage, or random-access storage.

SEQUENTIAL-ACCESS STORAGE

Punched cards, paper tape, and magnetic tape are examples of sequential-access storage media. When operating in a sequential environment, a particular record can be read only by first reading all the records that come before it in the file. When you store a file on tape, the 125th record cannot be read until the 124 records in front of it are read. The records are read in sequence. You cannot read just any record at random. This is also true when reading punched cards or paper tape.

DIRECT-ACCESS STORAGE

Direct-access storage allows you to access the 125th record without first having to read the 124 records in front of it. Magnetic disks and drums are examples of direct-access storage media. Data can be obtained quickly from anywhere on the media. However, the amount of time it takes to access a record is dependent to some extent on the mechanical process involved. It is usually necessary to scan some (but not all) of the preceding data.

RANDOM-ACCESS STORAGE

Random-access storage media refers to magnetic core, semiconductor, thin film, and bubble storage. Here, a given item of data can be selected from anywhere in storage without having to scan any preceding items. And, the access time is independent of the storage location.

Q-27. Punched cards, paper tape, and magnetic tape use what storage access method?

Q-28. What kind of storage allows you to access the 125th record without having to read the 124 records in front of it?

Q-29. Random-access storage media refers to what types of storage?

NETWORKS

A network can be defined as any system composed of one or more computers and terminals; however, most are composed of multiple terminals and computers. In this section you will learn how this allows dissimilar computers to work together as a team.

LOCAL AREA NETWORKS (LANs)

In local area networks (LANs), various machines are linked together within a building or adjacent buildings. Figure 4-12 shows an example of a LAN. A LAN allows dissimilar machines to exchange information within one universal system. With the ability to communicate, the dissimilar machines act as a team. The information that exists in one system can be reused without being reentered via keyboard or disk into another separate system. A universal system for the integration and exchange of information is connected to all input devices. The entire system is usually housed within the same building or the same geographic area. A local area network is made up of a communications facility (for example, a coaxial cable, such as that used for cable television) and interface units creating a link for the computers and terminals to the communications facility. Two designs can be used: broadband or baseband.

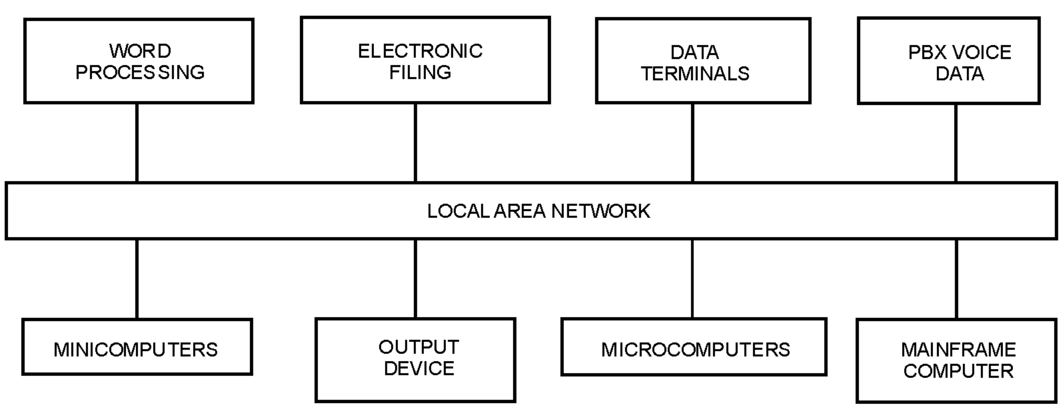


Figure 4-12.—Local area network system.

A baseband communications channel uses the basic frequency band of radio waves and a coaxial cable. This coaxial cable has one channel, which is like a party line. Only two machines can use this cable at one time, even though many have the channel available, but there is no central switching unit to route traffic over the network.

A more expensive channel, called a broadband communications channel, can handle more advanced applications. This includes transmission of voice as well as data and text. Because of the use of a controller to route traffic for a large number of simultaneous users, the users are able to share one of the many individual channels of the system.

WIDE AREA NETWORKS

Wide area networks provide for global connections and are sometimes referred to as global networks. Organizations are able to send information from city to city, across the nation, and to other countries throughout the world, through the expansion of local area networks into larger network configurations. Combinations of telephone lines, microwave radio links, and satellites are used by these larger telecommunications networks to send information. In 1965, the first successful communications satellite for business applications was launched. It was not the only try, it was preceded by many more primitive satellites. With the launching of larger and more complex satellites, the size and complexity of earth stations have been shrinking. Since satellite services' costs have been steadily decreasing, it is becoming more cost effective to employ them for business-type uses.

MODEMS

Since both signals and data can be transmitted and received through cables (communications lines), we refer to them as input/output channels. And when we transmit data directly to a computer over long distances, it becomes necessary to add two other devices, one at each end of the communications line. These devices are called modems (fig. 4-13). The word modem is an acronym for modulator/ demodulator (combines first syllable of each word). A modem converts the digital signal produced by your terminal or the computer to an audio signal suitable for transmission over the communications line. The modem at the other end of the line converts the audio signal back to a digital signal before it is supplied to the computers or your terminal. If this conversion were not carried out, the digital signal would degenerate during transmission and become garbled.

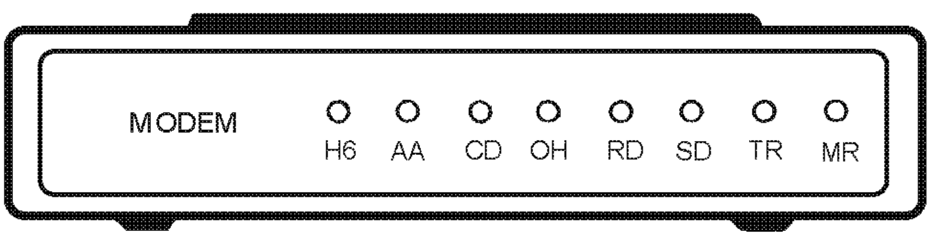


Figure 4-13.—Modem.

The physical link or medium that is used to carry (or transmit) data from one location to another is a communications channel. It allows remotely located input/output devices to communicate directly with the computer's central processing unit (cpu). Telephone lines (often referred to as land lines) are a frequently used type of communications channel.

In a simple data communications system, terminals and other remote I/O devices are linked directly to one or more cpu's to allow users to enter data and programs and receive output information. Interface elements (those devices that serve to interconnect), such as modems are used to bridge and control the different data communications environments. Modems are used to permit the system to switch back and forth from computer digital data to analog signals that can be transmitted over communications lines.

A modem never knows exactly when to expect data; therefore, it must be given some type of signal warning that data is about to be transmitted. This gives the modem time to get itself aligned and in synchronization with the incoming signal. Special characters, known as message characters, provide this warning and are placed in front of and behind the data to mark the beginning and ending of the message. Two methods are used: asynchronous and synchronous.

With asynchronous transmission, each character of data must be surrounded by message characters. As a result, more total bits must be transmitted (transferred) than would be necessary if the synchronous method were used.

With synchronous transmission, only a single set of start and stop message characters is needed per block of data, thus allowing more characters to be transmitted per second. As you can see, synchronous transmission is more efficient and faster. However, it has the disadvantage of requiring a more complex and expensive modem than does asynchronous transmission.

You should be aware that whenever data is transferred between devices, it also involves an exchange of prearranged signals. This is known as handshaking. These signals, in combination with a prearranged pattern of message characters, define the rules for exchanging data over a communications line. The exact rules depend upon each individual computer manufacturer, the telephone company, and the related devices (the modems) that make up the computer system. Protocol is the term used for the specific set of rules that govern handshaking and message characters.

In the system illustrated in figure 4-14, data to be sent to the main computer's cpu is entered through a remote online user terminal (far left). As the data is keyed, it is keyed in digital form and sent to a nearby modem to be converted into an analog signal suitable for transmission. This converted data is then transmitted over the telephone (or land) lines to another modem, that is located near the main computer system's cpu. The data, now in digital form, can be sent directly to the cpu for processing. The same route is followed when information is sent from the cpu back to the remote terminal.

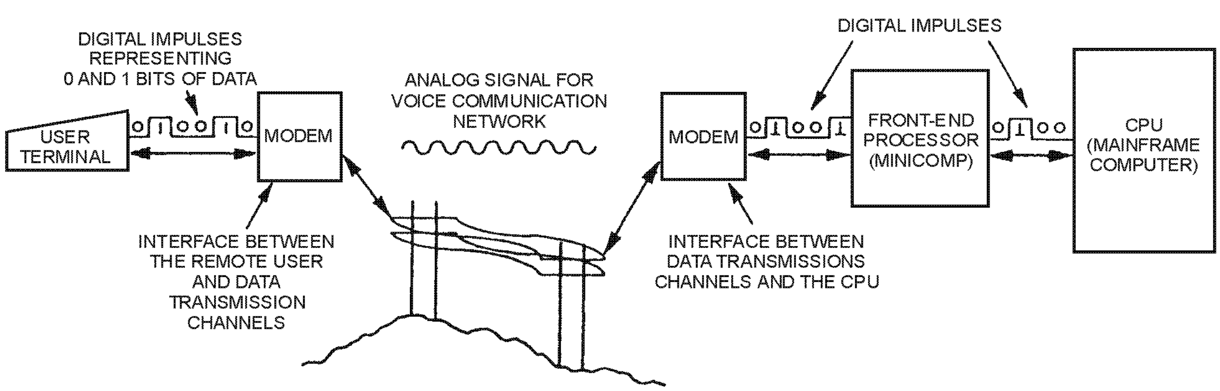


Figure 4-14.—Modems used in network system.

Data communications and networks expand our use of computer technology by providing a means for computers and other machines to talk to each other.

Q-30. Any system composed of one or more computers and terminals can be defined as what?

Q-31. A network system allows dissimilar machines to do what within one universal system?

Q-32. What does the make-up of a local area network consist of?

Q-33. How many designs of local area networks are there that can be used?

Q-34. What are the different designs of local area networks called?

Q-35. What is a baseband communication channel like?

Q-36. What do wide area networks provide for?

Q-37. Where does the word modem come from?

Q-38. What are interface elements?

Q-39. How does a modem know when to expect data?

Q-40. Whenever data is transferred between devices, it involves the exchange of prearranged signals; what is this process called?

SUMMARY

Congratulations! You have just finished the last chapter in *Introduction to Digital Computers*. In this chapter you learned about many things that were mentioned in other chapters, without a detailed explanation. This was done intentionally, as some of the subjects would have been too difficult and hard to understand without background knowledge. Through your study of chapters 1, 2, and 3, you gained enough knowledge to understand chapter 4. This chapter should have answered a lot of questions for you and made certain subjects more clear.

DATA is a general term used to describe raw facts like your service number, name, and paygrade.

SOURCE DATA is raw data typically written on some type of paper document.

DATA REPRESENTATION is accomplished by the use of symbols. The symbol itself is not the information, but merely a representation of it. Symbols convey meaning only when understood. In computers, symbols are represented by CODES.

COMPUTER CODING SYSTEMS are used to represent numeric, alphabetic, and special characters in computer storage and on magnetic media.

EXTENDED BINARY CODED DECIMAL INTERCHANGE CODE (EBCDIC) is an 8-bit code used in computers to represent numbers, letters, and special characters.

AMERICAN STANDARD CODE FOR INFORMATION INTERCHANGE (ASCII) is another 8-bit code developed to standardize a binary code to give the computer user the capability of using several machines to process data regardless of the manufacturer.

A **PARITY (CHECK) BIT** is used to detect errors in computer circuitry.

MAGNETIC CORE STORAGE is used as primary storage in some computers.

PRIMARY STORAGE CAPACITY AND ADDRESSES are designed and built into the computer by the manufacturer.

Computers may be **WORD-ADDRESSABLE, CHARACTER-ADDRESSABLE**, or **FLEXIBLE**.

Data in **SECONDARY STORAGE** like disk or tape is normally organized by bits, characters (bytes), fields, records, and files.

STORAGE ACCESS METHODS vary with the types of media and devices you are using.

SEQUENTIAL-ACCESS STORAGE is associated with punched cards, paper tape, and magnetic tape.

DIRECT-ACCESS STORAGE is obtained by using magnetic disks and drums.

RANDOM-ACCESS STORAGE refers to magnetic core, semiconductor, thin film, and bubble storage.

A **NETWORK** is any system composed of one or more computers and terminals; however, most are composed of multiple terminals and computers.

LOCAL AREA NETWORKS (LANs) allow dissimilar machines to exchange information within one universal system within a building or small geographic area.

WIDE AREA NETWORKS provide for global connections and are sometimes referred to as global networks.

A **MODEM** converts the digital signal produced by your terminal or the computer to an audio signal suitable for transmission over a communications line. It also converts the audio signal back to a digital signal before it is supplied to your terminal or computer.

ANSWERS TO QUESTIONS Q1. THROUGH Q40.

A-1. Data.

A-2. By symbols.

A-3. Numbers.

A-4. By either direct or indirect means.

A-5. Punched cards, paper tape, magnetic tape, or magnetic disk.

A-6. Extended Binary Coded Decimal Interchange Code.

A-7. 256.

A-8. 16.

A-9. Packing or packed data.

A-10. American Standard Code for Information Interchange.

A-11. To standardize a binary code to give the computer user the capability of using several machines to process data regardless of the manufacturer.

A-12. No, they are identical.

A-13. To detect errors in the circuitry.

A-14. Nine.

A-15. Program storage area.

A-16. By the individual programs being used.

A-17. A single binary digit.

A-18. Eight.

A-19. Number of bytes.

A-20. Magnetic cores are strung together on a screen of wire.

A-21. In primary storage.

A-22. The manufacturer.

A-23. Variable-word-length or byte-addressable.

A-24. Fixed-word-length.

A-25. By bits, characters (bytes), fields, records, and files.

A-26. A collection of related records.

A-27. Sequential-access.

A-28. Direct-access storage.

A-29. Magnetic core, semiconductor, thin film, and bubble.

A-30. A network.

A-31. Exchange information.

A-32. A communications facility and interface units.

A-33. Two.

A-34. Broadband and baseband.

A-35. A party line.

A-36. Global connections.

A-37. It is an acronym for modulator/demodulator.

A-38. Those devices that serve to interconnect.

A-39. It is given a signal warning that data is about to be transmitted.

A-40. Handshaking.

APPENDIX I

GLOSSARY

ACCESS TIME—The amount of time between the time a request for data from a storage device is made and the time the data is delivered.

ADA—A high-level programming language designed by the Department of Defense.

ALPHAMERIC (ALPHANUMERIC) CHARACTER SET—The set of characters that includes letters, numbers, and special characters.

ANALOG COMPUTER—A computer that solves problems using continuous data from physical quantities like voltage or temperature.

APPLICATION (PROGRAMS) SOFTWARE—Programs written to solve user problems.

ARITHMETIC-LOGIC UNIT—The part of the cpu that contains the logic capability and performs all the arithmetic functions (addition, subtraction, multiplication, and division).

ARTIFICIAL INTELLIGENCE—The capability of a machine to perform human-like intelligence functions, such as learning, adapting, reasoning, and self-correction.

ASCII (American Standard Code for Information Interchange)—A standardized 8-bit code (originally a 7-bit code) designed for transmitting and processing data.

ASSEMBLER—A computer program that translates source programs written in assembly language into machine language (object) programs.

ASSEMBLY LANGUAGE—A low-level, machine-oriented programming language in which each instruction (written as a mnemonic) translates into a single machine language (computer) instruction.

AUTOMATIC DATA PROCESSING (ADP)—A general term used to define a system for automatically performing a series of data processing functions by means of machines using mechanical, electromechanical, and electronic circuitry.

AUXILIARY EQUIPMENT—The peripheral equipment or devices that may or may not be in direct communication with the central processing unit of a computer.

AUXILIARY STORAGE—See storage, secondary.

BACKUP FILE—A copy of a program or data file to be used in the event something happens to the original.

BASIC (Beginners All Purpose Symbolic Instruction Code)—A high-level, general-purpose programming language primarily used on microcomputers. See NAVEDTRA 10079, *Introduction to Programming in BASIC*.

BAUD—A unit for measuring data transmission speed. For practical purposes, it is now used interchangeably with bits per second as the unit of measure of data flow.

BINARY—Two values (0 or 1) or states (ON or OFF); the number system used in computers.

BIT—An abbreviation for binary digit; the smallest unit of data, either a 0 or 1.

BLOCKED RECORDS—One or more logical records grouped and treated as a unit (physical record or block) for input/output processing.

BLOCKING FACTOR—The number of records stored in a record block.

BOOT OR BOOTSTRAP—(1) A set of instructions that causes additional instructions to be loaded until the complete computer program is in storage. (2) A technique or device designed to bring itself into a desired state by means of its own action; e.g., a machine routine whose first few instructions are sufficient to bring the rest of itself into the computer from an input device. (3) That part of a computer program used to establish another version of the computer program.

BUG—A mistake in a program.

BYTE—A group of bits next to each other that is considered a unit; for example an 8-bit byte.

CENTRAL PROCESSING UNIT (cpu)—The part of the computer hardware that directs the sequence of operations, interprets the coded instructions, performs arithmetic and logical operations, and initiates the proper commands to the computer circuits for execution. It controls the computer operation as directed by the program it is executing.

CHARACTER—One symbol; for example, A, Z, a, z, 0, 1, 9, !, ".

CHIP—A small piece of silicon impregnated with impurities in such a way as to form transistors, diodes, and resistors. Electrical paths are formed on the silicon by depositing thin layers of aluminum or gold.

COBOL (COmmon Business Oriented Language)—A high-level programming language designed for business-type applications.

COMPATIBLE SOFTWARE—Programs that can be run on more than one type of computer. These programs come in several different versions so they can be run under several different operating systems.

COMPILER—A program that translates source programs written in a high-level programming language (for example COBOL or FORTRAN) into machine language.

COMPUTER—A programmable electronic device that can store, retrieve, and process data.

COMPUTER OPERATOR—The person who sets up and operates the computer system.

COMPUTER PROGRAMMER—A person who designs, writes, tests, debugs, and documents programs.

COMPUTER SYSTEM—The cpu (mainframe) with its console, input, and output devices, and secondary (auxiliary) storage devices.

COMPUTER USERS—See users.

COMPUTER WORD—See word, computer.

CONSOLE—The unit of the computer used by the computer operator to communicate with and control the computer system.

CONTROL SECTION—The part of the cpu that directs the flow of operations and data, maintains order in the computer, and initiates execution of the instructions.

CRT (Cathode-Ray Tube) TERMINAL—A computer terminal that displays its output on a television-like screen that may be black and white or color.

CURSOR—A pointer (a dot of light) on a crt screen to let you know the next position in which data will be entered. By depressing cursor control keys, the operator can move the cursor from line to line and from character to character.

CUSTOM SOFTWARE—Programs designed and written to the specifications of a user or an organization.

DATA—Facts represented by numbers, letters, or symbols to which meaning is or can be assigned.

DATA BASE—A structured collection of data that can be extracted, organized, and manipulated by a program.

DATA COMMUNICATIONS—The means by which data is transmitted electronically from one location to another over a communications channel.

DATA ELEMENT—One item of information; the smallest unit of data that can be referenced.

DATA FLOWCHART—See flowchart.

DATA, REFERENCE—The source document identification.

DATA REPRESENTATION—The symbols and codes used by computers to represent letters, numbers, and special characters.

DEBUGGING—The process of finding errors (bugs) in a program or system and correcting them so that the program or system runs correctly.

DENSITY, RECORDING—The number of bits, bytes, characters, or frames per linear inch on a recording medium, like tape or disk.

DIGITAL COMPUTER—A computer that solves problems on discrete data using 0's and 1's (OFF and ON states) to represent data and operations.

DIRECT ACCESS—A storage method that allows the computer to locate and read a particular record without having to search through an entire file. The computer is able to access data independent of its location. Magnetic disks, diskettes, and drums are considered direct access devices.

DISK DRIVE—A direct-access storage device for recording and retrieving data on hard (rigid) disk or floppy disks (diskettes).

DISK PACK—A mass storage device in which information is stored on one or both sides of a rigid disk that can be magnetized. The disk is rotated by a disk drive and information is stored and retrieved by one or more magnetically sensitive read/write heads.

DISKETTE—A mass storage device in which information is stored on one or both sides of a flexible disk that can be magnetized. The diskette is rotated by a diskette drive and information is stored and retrieved by one or more magnetically sensitive read/write heads. Diskettes are also called floppy disks because the disk bends easily.

DOWNTIME—The length of time the computer is not operating, either because of preventive maintenance (scheduled downtime) or a malfunction (nonscheduled downtime).

EBCDIC (Extended Binary Coded Decimal Interchange Code)—An 8-bit coding system for representing uppercase and lowercase letters, numbers, and special characters.

EPROM—The acronym for erasable programmable read-only memory.

FIELD, DATA—An item of information in a data record. One or more related characters that are treated as a unit of information.

FILE—A collection of related records; for example, a payroll file. Any collection of records holding similar data or transactions that are stored together to permit systematic access and modification.

FILE, MASTER—The file that contains all the data records in up-to-date form. It is a main reference file of relatively more permanent information, which is usually updated periodically.

FIRMWARE—A set of program instructions, a microprogram, permanently stored in read-only memory.

FLAT PANEL DISPLAY—A display device that consists of a grid of electrodes in a flat, gas-filled panel. The image can persist for a long period of time without refresh.

FLOPPY DISK—See diskette.

FLOWCHART—A graphic representation of the processing steps (logic) of a program (a program flowchart) or the inputs, outputs, and processing steps of a system (a systems [data] flowchart). The graphic representation uses symbols to represent operations and directional lines to indicate sequence and direction of flow.

FORMAT—The arrangement or layout of data in or on a data medium.

FORTRAN (FORmula TRANslator)—A high-level programming language for scientific and mathematical applications.

FULL-DUPLEX CHANNEL—A channel that provides for simultaneous transmission in both directions, such as the telephone.

GENERAL-PURPOSE COMPUTER—A computer designed to operate on a program of instructions for the purpose of solving many different types of processing problems.

GENERATIONS OF COMPUTERS—Historically, the distinctive types of computers from the 1940s to the present; the first generation was based on vacuum tubes, the second on transistors, the third and current features integrated circuits. Recent developments involve the use of VLSI (very large scale integration) and semiconductor memories.

GRAPHICS—The use of pictorial means to present data in the form of plotted curves, graphs (bar, pie, line, and so on), or diagrams. These may be displayed on a crt or printed.

HANDSHAKING—The process through which the rules for exchanging data over a communications line are defined for the two devices involved.

HARD COPY—The term given to humanly readable printed output from a computer.

HARDWARE—The visible, physical equipment of a system, including the computer (cpu) and related peripheral equipment; as distinguished from software.

HEAD POSITIONING—Placing a read/write head over a specified track on a disk or drum.

HEXADECIMAL—The number system with base 16 (0-9 and A-F). A represents 10; B represents 11; C represents 12; D represents 13; E represents 14; and F represents 15. Used in some computer systems.

HIGH-LEVEL LANGUAGES—Programming languages that allow the programmer to write programs in English-like terms and symbols and mathematical notation, rather than the 0's and 1's used by the computer. These high-level programs must be translated into machine language before the computer can execute them. FORTRAN, Ada, COBOL, and BASIC are examples.

HOST COMPUTER—The main or controlling computer in a distributed data processing network (ddp).

HYBRID COMPUTER—A computer that combines the functions of both analog and digital computers.

I/O—Input/Output.

INPUT—The data entered into a computer system for processing.

INPUT DEVICES—Devices for reading data and programs into the computer system for processing.

INPUT/OUTPUT DEVICES—Secondary storage devices for writing and reading data. Magnetic tape drives, magnetic disk drives, and drums are examples.

INTEGRATED CIRCUIT—A miniaturized chip in which semiconductor components and other such technology combine the functions of a number of conventional components (such as transistors, resistors, capacitors, and diodes).

INTERNAL STORAGE (MEMORY)—See storage, primary.

INTERRECORD/INTERBLOCK GAP—A blank section of recording surface separating each record or block of records on a magnetic data medium.

K—An abbreviation for the value 1,024 which is 210. Often used to express the memory capacity of a computer. For example, a 512K computer has 524,288 bytes of memory.

KEY-TO-DISK—A process, similar to key-to-tape, in which data is transmitted from a keyboard to magnetic disk.

KEY-TO-TAPE—An operation in which data is transmitted from a keyboard to magnetic tape.

LANGUAGE TRANSLATOR—A program that reads a source program and converts it into an object (machine language) program. Assemblers and compilers are examples.

LOCAL-AREA NETWORK—A network that normally operates within a well-defined and generally self-enclosed area. The communication stations or terminals are usually linked by cable and are within 1,000 feet of each other.

LOGICAL RECORD—A record that includes all the data that belongs together as a unit regardless of the physical size or storage location of the data.

M—A unit of measurement approximately equal to one million and used to express the capacity of a computer memory. 1M is about 1,000,000 units. Memory size is usually measured in words or bytes.

MACHINE LANGUAGE—Machine instructions in binary bit patterns that the central processing unit can execute directly without additional interpretation or translation.

MACHINE (TAKE-UP) REEL—A reel that remains on the tape drive and on which magnetic tape is wound during the processing of a tape.

MAGNETIC MEDIA—Magnetic cards, tapes, disks, drums, cartridges, and cassettes used to record data or information.

MAGNETIC TAPE—A mass storage device in which information is stored on a plastic tape coated with a magnetic film. The tape is wound on reels that are rotated by tape drives. Information is stored and retrieved sequentially by magnetically sensitive read/write heads.

MAINFRAME COMPUTERS—This term is usually used to designate large-scale computer systems, although the precise definition of mainframe is the cpu and the control elements of any computer system.

MAIN STORAGE (MEMORY)—See storage, primary.

MASS STORAGE—Any external storage medium (magnetic tape, disk, drum, and so on) that supports and can be linked to the cpu's main memory in the computer. When the power is turned off, information in the mass storage is retained (not lost).

MEDIUM—The material on which data and instructions are recorded, such as punched cards, paper tape, and all forms of magnetic media (tape, disk, drum, and so on).

MEMORY—A device or section of the computer in which computer instructions and data can be stored for retrieval (synonymous with primary or internal storage).

MICROCOMPUTERS—The smallest category of computers, usually with the entire central processing unit on a single chip. Unlike large-scale and minicomputer systems, they are designed to be used by one person at a time (hence the term, personal computer [PC]).

MICROPROCESSOR—The semiconductor central processing unit (cpu) of a microcomputer that fits on a small silicon chip. The microprocessor is the central chip containing the control units of the computer.

MICROSECOND—One millionth of a second.

MILLISECOND—One thousandth of a second.

MINICOMPUTERS—Midsize computers that are smaller than large-scale systems but with the same components. They are less expensive and have less strict environmental requirements.

MODEM—Acronym for MOdulator-DEModulator. A device that converts data from digital-to-analog format for transmission on analog transmission lines and converts data in analog format to digital format for computer processing.

MULTIPROCESSING—A computer processing mode that provides for simultaneous processing of two or more programs or routines by use of multiple cpu's.

MULTIPROGRAMMING—A computer processing mode that provides for overlapping or interleaving the execution of two or more programs at the same time by a single processor.

NANOSECOND—A billionth of a second.

NETWORK—Computers and terminals linked together through a communications system to allow users at different locations to share data files, devices, and programs.

ONLINE PROCESSING—Processing from terminals under the direct control of a computer.

OPERATING SYSTEM—Software that controls the execution of programs. An operating system may provide services such as input/output control and data management. It may also provide job scheduling, memory allocation, and other general functions. It is usually loaded by a bootstrap program.

OUTPUT—The results of computer processing. It may be data transferred to tape, disk, paper, and so on.

PACKAGED SOFTWARE—Programs already written (and tested) to solve specific types of problems; usually designed by a central design agency (CDA) or purchased from a software firm or computer manufacturer.

PACKED DECIMAL—In ASCII and EBCDIC, the representation of two digits stored in one eight-bit byte.

PAPER TAPE—See punched tape.

PARALLEL TRANSMISSION—A method of data transmission in which all bits of a particular character are transmitted simultaneously.

PARITY BIT—A check bit; an extra bit added to a group of bits for use in detecting errors during data transfer.

PARITY CHECK—An internal error checking method in which the binary digits in a character or word are added and the sum is checked against a single previously computed parity digit. The check tests whether the number of one bits in a character or word are odd or even, depending on the parity of the computer.

PASSWORD—A protected word or string of characters that identifies or authenticates a user for access to a specific resource, such as a file or record.

PERIPHERAL EQUIPMENT—Equipment used for data entry, storage, or retrieval, but which is not a part of the central processing unit. Peripherals include crt displays, terminals, printers, and mass storage (tape, disk, and drum) devices.

PERSONAL COMPUTER (PC)—A computer, usually a microcomputer, that is more affordable than minicomputers or mainframes and is used by one person at a time.

PICTURE ELEMENT—Synonym for pixel. See pixel.

PIXEL—In computer graphics, the smallest element of a display surface that can be independently assigned color or intensity.

PRIMARY STORAGE—See storage, primary.

PRINTER—A device used with a computer to produce hard copy, printed output.

PROGRAM—(1) Verb—The act of writing instructions for computer execution. (2) Noun—The set of instructions that tells the computer the steps to execute to automatically solve a problem.

PROGRAM FLOWCHART—See flowchart.

PROM—Acronym for programmable read-only memory.

PUNCHED CARD—A card punched with hole patterns that represent data or program instructions. Punched cards can be read by an input device (card reader) to a computer.

PUNCHED TAPE—A tape punched with hole patterns that represent data or program instructions. Tape can be read by an input device to a computer.

RAM—Acronym for random-access memory.

RANDOM ACCESS—A method of accessing data (or instructions) without having to scan any preceding information. Magnetic core, semiconductor, and bubble memories are considered random access storage devices.

REAL-TIME PROCESSING—A computer processing method in which data about a particular event is entered directly into the computer as the event occurs and is immediately processed so it can influence future processing.

RECORD—A group of related fields, all pertaining to the same subject.

RECORD BLOCK—Several records blocked together.

RECORD LENGTH—The number of characters in a record.

REMOTE TERMINAL—A display terminal, such as a crt or other piece of equipment, which is not located with the computer but is connected by a communications line. In a typical online, real-time communications system, the remote device is usually a teletypewriter or a crt visual display unit.

ROM—Acronym for read-only memory.

ROTATIONAL DELAY—The time required for the read/write head to find a specified record on a disk, diskette, or drum once head positioning has occurred.

SECONDARY STORAGE—See storage, secondary.

SECTORS—The pie-shaped segments of a disk's recording surface.

SEQUENTIAL ACCESS—A storage technique in which the stored items of information become available only in a one after the other sequence, whether or not all the information or only some of it is desired. Magnetic tape is an example.

SOFT COPY—Output of a computer displayed on a display terminal or monitor (crt). It is nonpermanent.

SOFTWARE—Programs, routines, codes, and other written information used to direct the operation of a computer; as distinguished from hardware.

SORT—The process of arranging data records in a predefined sequence by use of sort keys; for example, to sequence personnel records by social security number (the sort key).

SOURCE DATA—The data in its initial state to be processed by a computer system.

SOURCE DOCUMENT—The document that contains the initial (raw) data for computer processing.

SOURCE PROGRAM—A computer program written in a language like COBOL, FORTRAN, or assembly language. It must be translated into an object program before it can be executed by a computer.

SPECIAL-PURPOSE COMPUTER—A computer designed to perform one specific function such as a weather computer.

STORAGE, PRIMARY (MAIN, INTERNAL)—The section of the cpu in which instructions and data are held. Also called main memory.

STORAGE, SECONDARY (AUXILIARY, EXTERNAL)—Storage outside the cpu where programs and data are stored for future computer processing; for example tapes, disks, and punched cards.

STORED PROGRAM—The set of instructions stored in computer memory for execution.

SUBROUTINE LIBRARY—A set of standard and proven computer routines that are kept on file for use at any time.

TELECOMMUNICATIONS—The transmission of data between computer systems and/or terminals in different locations.

TELEPROCESSING—A method of data processing in which communication devices are used.

TERMINAL—A device linked to the central processor for entering or receiving data and programs.

TIME SHARING—A processing mode in which many users share the computer systems' resources through online terminals. Each user gets a slice of computer time.

TRACK—(1) One of seven or nine, horizontal rows stretching the entire length of a magnetic tape and on which data can be recorded. (2) One of a series of concentric circles on the surface of a disk. (3) One of a series of circular bands on a drum.

UNBLOCKED—Having a blocking factor of one logical record per block.

USER PROGRAMS—Programs written to solve specific user problems, called applications software.

USERS—(1) The people who use the output from computer processing. (2) The people who operate a computer for their own purposes.

UTILITY PROGRAMS (UTILITIES)—Programs designed to perform often needed general functions.

WIDE AREA NETWORK—A network that usually covers large geographical areas. Communications between stations or terminals usually occur using standard telephone lines or microwave relays.

WORD, COMPUTER—A group of related bytes treated as a single addressable unit or entity in computer memory.

APPENDIX II

REFERENCE LIST

Casady, Mona J., and Sandburg, Dorothy C., *Word/Information Processing,* South-Western Publishing Co., Cincinnati, Ohio, 1985.

Davis, Gordon B., *Introduction to Electronic Computers*, McGraw-Hill Book Co., New York, N.Y., 1971.

Delco Electronics, *Fundamentals of Digital Computers Course,* General Motors Corp., Milwaukee, Wis., 1972.

Fuori, William M., D'Arco, Anthony M.S., and Ovilia, Lawrence, *Introduction To Computer Operations,* 2nd edition, Prentice-Hall Inc., Englewood Cliffs, N.J., 1981.

Jansen, John T., *How Microprocessors Function*, Dun-Donnelley Publishing Corp., New York, N.Y., 1983.

Lavin, Albert J. III DSCS(SW), *Data Systems Technician 3 & 2,* Volume 2, NAVEDTRA 10231, NETPDTC, Pensacola, Fla., 1987.

McGee, Fred H. III DPC, *Introduction to Programming in BASIC,* NAVEDTRA 10079-2, NETPDTC, Pensacola, Fla., 1983.

Mims, Forrest M. III, *Understanding Digital Computers,* Tandy Corp. Co., Fort Worth, Tex., 2nd edition, 1987.

Moore, Alvin W., et. al., *Microprocessor Applications Manual,* McGraw-Hill Book Co., New York, N.Y., 1975.

Rau Bernie, ZENITH 100 *Series Microcomputer Operations,* NAVEDTRA 40801, Naval Oceanography Command Facility, Bay St. Louis, Miss., 1985.

Shelly, Gary B., and Thomas J. Cashman, *Introduction To Computers and Data Processing,* Anaheim Publishing Co., Brea, Calif., 1980.

U S Army Signal School, *Digital Computers FM 11-72*, Fort Gordon, Ga., 1977.

Van Overberghe, Albert G., Jr. DPCS, *Data Processing Technician 3*, NAVEDTRA 10263, NETPDTC, Pensacola, Fla., 1987.

Vermillion Associates, *Computer Concepts For Small Business*, Heath Company, Benton Harbor, Mich., 1978.

MODULE 22 INDEX

I

K

L

M

N

O

P

R

S

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, “Operational Concepts,” pages 1-1 through 1-25.

1-1. When was the first mechanical adding machine invented?

1. 1264
2. 1426
3. 1462
4. 1642

1-2. What year did electronics enter the computer scene?

1. 1918
2. 1919
3. 1920
4. 1921

1-3. In modern digital computers, circuits that store information, perform arithmetic operations, and control the timing sequences are known as what?

1. Flip-flops
2. Amplifiers
3. Oscillators
4. Multipliers

1-4. When was the UNIVAC I developed?

1. 1944
2. 1946
3. 1950
4. 1951

1-5. The field of research that is developing computer systems which mimic human thought in a specific area and improve performance with experience and operation is what field of research?

1. Human intelligence
2. Artificial intelligence
3. Animal intelligence
4. Computer intelligence

1-6. Mechanical computers are what type of devices?

1. Digital
2. Electrical
3. Analog
4. Electromechanical

1-7. What determines the size of an analog computer?

1. Where it will be installed
2. Number of operators using it
3. Cost
4. Number of functions it has to perform

1-8. What is the primary use of analog computers in the Navy?

1. Gun fire control
2. Data processing
3. Ships steering
4. Missile fire control

1-9. Compared to mechanical computers, electromechanical computers are different in which of the following ways?

1. They cost more
2. They are bigger
3. They are less accurate
4. They use electrical components to perform some of the calculations and to increase the accuracy

1-10. In early electronic computers, what was the weak link in electrical computations?

1. Transistors
2. Resistors
3. Vacuum tubes
4. Capacitors

1-11. A computer that is designed to perform a specific operation is what kind of computer?

1. All-purpose
2. General-purpose
3. Special-purpose
4. Single-purpose

1-12. How are the instructions that control a computer applied to a special-purpose computer?

1. From a stored program
2. From a keyboard
3. From an input device
4. From built-in instructions

1-13. What is a drawback to the specialization of a special-purpose computer?

1. Low speed
2. Lack of versatility
3. Large size
4. High cost

1-14. What gives a general-purpose computer the ability to perform a wide variety of operations?

1. It can store and execute different programs in its internal storage
2. It is a much larger computer
3. It has a huge built-in program
4. It can operate faster than other computers

1-15. All analog computers are what type?

1. Mechanical
2. Electromechanical
3. Special-purpose
4. General-purpose

1-16. What are computers called that combine the functions of both analog and digital?

1. Analog-digital computers
2. Mixed computers
3. Duplexed computers
4. Hybrid computers

1-17. A digital computer knows how to do its work by what means?

1. By a list of instructions called a program
2. By a list of instructions called a job sequence
3. By its hardware
4. By its peripheral equipment

1-18. What is the most popular generic term for computer programs?

1. Hardware
2. Software
3. Wordprocessing
4. Graphics

1-19. First generation computers were characterized by what technology?

1. Transistors
2. Resistors
3. Vacuum tubes
4. Printed circuits

1-20. What type of computer language was used with first generation computers?

1. Machine
2. COBOL
3. BASIC
4. Fortran

1-21. Computers of the second generation were characterized by what technology?

1. Vacuum tubes
2. Capacitors
3. Transistors
4. Resistors

1-22. The small, long lasting transistors used in second generation computers had which of the following effects?

1. They increased processing speeds and reliability
2. They decreased processing speeds and increased reliability
3. They increased processing speeds and decreased reliability
4. They decreased processing speeds and reliability

1-23. Internal processing speeds of second generation computers were measured at what speed?

1. Hundredths of a second
2. Thousandths of a second
3. Millionths of a second
4. Billionths of a second

1-24. Third generation computers are characterized by what technology?

1. Capacitors
2. Transistors
3. Resistors
4. Miniaturized circuits

1-25. A circuit and its components can be etched onto which of the following materials?

1. Plastic
2. Silicon
3. Gold
4. Pressed fiber

1-26. The internal processing speeds of third generation computers are measured at what speed?

1. Hundredths of a second
2. Thousandths of a second
3. Millionths of a second
4. Billionths of a second

1-27. Fourth generation technology has which of the following results for the computer industry?

1. Computers that are significantly smaller and lower in cost
2. Computers that are significantly larger and lower in cost
3. Computers that are significantly smaller and higher in cost
4. Computers that are significantly larger and higher in cost

1-28. What does the acronym ROM stand for?

1. Run-on manual
2. Read-only minutes
3. Read-only memory
4. Read-only manual

1-29. Which of the following will be one of the future challenges involving computer power?

1. How to properly and effectively use the computing power available
2. How to increase computer storage capacity
3. How to increase computer power
4. How to properly install the computers available

1-30. What term is used for programs such as assemblers, compilers, operating systems, and applications programs?

1. Hardware
2. Peripheral devices
3. Software
4. Sub-systems

1-31. Which of the following is one of the more widespread uses of the computer in the Navy?

1. Research
2. Word processing
3. Manufacturing
4. Games

1-32. Computers have an advantage over typewriters in what area?

1. Cost
2. Speed
3. Reliability
4. Correcting errors

1-33. What does the acronym S-N-A-P stand for?

1. Shipboard navigational aid package
2. Shipboard Navy applied program
3. Shipboard non-tactical ADP program
4. Shipboard nuclear active program

1-34. Which computer is used with the SNAP II system?

1. UYK-4
2. UYK-7
3. UYK-20
4. AN/UYK-62 (V)

1-35. Where are the user terminals for SNAP II placed on board ship?

1. Engineering spaces
2. Work centers
3. Supply spaces
4. Electronics spaces

1-36. The work center supervisor can update which of the following items from a user terminal?

1. COSAL, APL, EIC, and CSMP only
2. APL, EIC, SHIP'S FORCE WORK LIST, and CSMP only
3. COSAL, APL, SHIP'S FORCE WORK LIST, and CSMP only
4. COSAL, APL, EIC, SHIP'S FORCE WORK LIST, and CSMP

1-37. What type of classified use does SNAP II allow?

1. Unclassified
2. Confidential
3. Secret
4. Top Secret

1-38. What is a central set of programs called that manages execution of other programs and performs common functions like read, write, and print?

1. Managing system
2. Execution system
3. Operating system
4. Word processing system

1-39. What is the function of a built-in program called a bootstrap loader?

1. To load a word processor into the computer's internal memory
2. To load an external operating system into the computer's internal memory
3. To load a graphics program into the computer's internal memory
4. To load a bootstrap program into the computer's internal memory

1-40. When an error message such as device error is shown on the display screen, which of the following problems could be the cause?

1. No floppy disk in drive A
2. Floppy disk inserted incorrectly in drive
3. Lock handle on drive A not lowered
4. Each of the above

1-41. A display similar to this A> means what in computer terminology?

1. A device error
2. No system
3. A prompt
4. Run again

1-42. What does it mean when the computer displays a prompt on the screen?

1. The computer has made an error
2. There is no system in the computer
3. You need to stop putting information into the computer
4. You can tell the computer what to do next

1-43. To tell the operating system what program to run, you should take which of the following actions following the operating system prompt A>?

1. Type help
2. Reboot the computer
3. Press the execute key
4. Type the program name

1-44. Online HELP screens serve what purpose?

1. Display the contents of memory
2. Display the operating system directory
3. Tell the operator how to perform a given function
4. Stop computer processing so the operator can read the instruction manual

1-45. Floppy disks provide which of the following functions?

1. Store data
2. Perform arithmetic operations
3. Provide alternate power to the computer
4. Check the accuracy of computer operations

1-46. Touching the exposed area seen through the timing hole and the read/write slots on a floppy disk can do what, if anything, to the data in that area?

1. Ruin it
2. Add to it
3. Move it
4. Nothing

1-47. What maximum number of disks should be stacked horizontally?

1. 5
2. 10
3. 15
4. 20

1-48. What is perhaps the most common source of a magnetic field that can affect a floppy disk?

1. Crt's
2. Printer
3. Telephone
4. Disk drives

1-49. In which of the following ways does smoke affect a computer?

1. It damages the electronics
2. It causes the monitor to fail
3. It coats the keyboard
4. It causes buildup on disks and disk drives

1-50. What, if anything, can happen to a floppy disk when it is exposed to direct sunlight or excessive heat?

1. It can become warped or distorted so it cannot be used
2. It can become sticky, which stops the drive
3. It can lose part of the data recorded on it
4. Nothing, it is not affected

1-51. Typically, floppy disks will operate only in what temperature range?

1. 40 to 120 degrees Fahrenheit
2. 50 to 120 degrees Fahrenheit
3. 60 to 120 degrees Fahrenheit
4. 70 to 120 degrees Fahrenheit

1-52. A floppy disk will accept what relative humidity range?

1. 5% to 60%
2. 10% to 70%
3. 10% to 80%
4. 10% to 90%

1-53. When a pencil or ballpoint pen is used to write on the label after it is attached to the disk, what, if anything, can happen to a disk?

1. Some of the data written on the label can be added to the disk
2. All of the data can be lost, but the disk can be used again
3. The disk can be destroyed
4. Nothing; there can be no effect

1-54. In the computer world, what method provides a means to ensure that any data lost can be recovered?

1. Records
2. Backup files
3. Tracks
4. Blocks

1-55. What two media are commonly used for backup?

1. Paper tape and punched cards
2. Magnetic tape and punched cards
3. Disk and magnetic tape
4. Disk and drum

1-56. What is the most common method of creating a backup for a microcomputer?

1. Copying the disk onto a magnetic tape
2. Copying the disk onto a paper tape
3. Copying the disk onto a punched card
4. Copying the disk onto another disk

ASSIGNMENT 2

Textbook assignment: Chapter 2, "Hardware," pages 2-1 through 2-35.

2-1. The components or tools of a computer system can be grouped into what two categories?

1. Hardware and software
2. Hardware and firmware
3. Firmware and software
4. Software and programs

2-2. What section/unit is the brain of a computer system?

1. Control section
2. Arithmetic-logic section
3. Central processing unit
4. Input unit

2-3. What section/unit is the computing center of a computer system?

1. Arithmetic-logic section
2. Central processing unit
3. Control section
4. Output unit

2-4. The central processing unit is made up of which of the following sections?

1. Control and internal storage only
2. Central and arithmetic-logic only
3. Arithmetic-logic and internal storage only
4. Control, internal storage, and arithmetic-logic

2-5. When a program is so large and complex that it exceeds the memory capacity of a stored-program computer, where is the overflow stored?

1. Input storage area
2. Output storage area
3. Primary memory
4. Auxiliary memory

2-6. What part of the computer dictates how and when each specific operation is to be performed?

1. Control section
2. Arithmetic-logic section
3. Input storage area
4. Output storage area

2-7. Of the four major types of instructions, which one has the basic function of moving data from one location to another?

1. Control
2. Logic
3. Arithmetic
4. Transfer

2-8. To send commands to devices not under direct command of the control section, what type of instructions are used?

1. Control
2. Logic
3. Arithmetic
4. Transfer

2-9. Operations like adding and multiplying are performed by what section?

1. Control-logic
2. Storage-logic
3. Arithmetic-logic
4. Transfer-logic

2-10. When processing is taking place, data is transferred back and forth between what two sections?

1. Control and internal storage
2. Internal storage and arithmetic-logic
3. Control and arithmetic
4. Arithmetic and output

2-11. The process by which instructions and data are read into a computer is called what?

1. Moving
2. Storing
3. Inputting
4. Loading

2-12. An auxiliary (wired) memory is used in some computers to permanently store a small program that makes manual loading unnecessary. What is this program called?

1. Operating system
2. Bootstrap
3. Word processing
4. Graphics

2-13. The tiny doughnut-shaped rings used to make up magnetic core storage are made of what material?

1. Aluminum
2. Steel
3. Tin
4. Ferrite

2-14. Data is stored in computers in what form?

1. Binary
2. Decimal
3. Octal
4. Hexadecimal

2-15. The state of each core in magnetic core storage is changed by what?

1. The amount of magnetism
2. The amount of current
3. The direction of magnetism
4. The direction of current

2-16. Electronic circuits are placed on a silicon chip by what method?

1. Wired
2. Drawn
3. Etched
4. Printed

2-17. Each of the individual electronic circuits on a silicon chip is called what?

1. A memory cell
2. A bit cell
3. A byte cell
4. A holding cell

2-18. Semiconductor storage has which of the following drawbacks?

1. It is too slow
2. It is expensive
3. It is volatile
4. It is unreliable

2-19. Using a very thin crystal made of semiconductor material, what type of memory can be created?

1. Bubble
2. Magnetic core
3. Semiconductor
4. Capacitive

2-20. In bubble memory, where is the control circuit imprinted on the crystal of semiconductor material?

1. The side
2. The bottom
3. The middle
4. The top

2-21. Who installs the programs in read-only memory?

1. The programmer
2. The manufacturer
3. The operator
4. The dealer

2-22. Programs that are tailored to certain needs and permanently installed in ROM by the manufacturer are called what?

1. Firmware
2. Software
3. Hardware
4. Diskware

2-23. What kind of memory used inside computers has a read/write capability without any additional special equipment?

1. ROM
2. RAM
3. EPROM
4. PROM

2-24. A special device is needed to burn the program into what type of memory?

1. ROM
2. PROM
3. ERAM
4. RAM

2-25. EPROM can be erased by what method?

1. With a current charge
2. With a voltage change
3. With a burst of ultra-violet light
4. With a special program

2-26. To coat magnetic disks, what magnetizable recording material is used?

1. Plastic
2. Mylar®
3. Aluminum oxide
4. Iron oxide

2-27. What is the size range of the diameters of magnetic disks?

1. 3 inches to 4 feet
2. 4 inches to 3 feet
3. 5 inches to 6 feet
4. 6 inches to 5 feet

2-28. Data is stored on all disks in a number of invisible concentric circles called what?

1. Cracks
2. Grooves
3. Paths
4. Tracks

2-29. A floppy disk surface has what maximum number of tracks?

1. 66
2. 77
3. 88
4. 99

2-30. When data is written on a disk in the same area where data is already stored, the old data is affected in which of the following ways, if at all?

1. It is moved to a new area
2. It is mixed with the new data
3. It is replaced
4. It is not affected

2-31. How are records on a track separated?

1. By a gap in which no data is recorded
2. By a gap in which the name of the record is recorded
3. By a gap in which the record is numbered
4. By a gap in which the operator's name is placed

2-32. To increase the amount of data we can store on one track, what technique can be used?

1. Records
2. Files
3. Disk address
4. Blocking

2-33. Designers were able to increase the data density of a disk by increasing the number of tracks. What code name was given to this technology?

1. Computer
2. Winchester
3. Solid state
4. Colt

2-34. During reading and writing, which of the following changes are achieved by reducing the distance of the read/write heads over the disk surface?

1. Data density can be improved and storage capacity decreased
2. Data density is lessened and storage capacity increased
3. Data density can be improved and storage capacity increased
4. Data density is lessened and storage capacity decreased

2-35. To physically organize data on diskettes, what method is used?

1. Records
2. Cylinder
3. Files
4. Sector

2-36. The lengths of magnetic tapes used with computers have what range?

1. From 400 to 1,000 feet
2. From 500 to 2,000 feet
3. From 600 to 3,000 feet
4. From 700 to 4,000 feet

2-37. Magnetic tapes can be packaged in which of the following ways?

1. Open reel only
2. Cartridge and cassette only
3. Open reel and cartridge only
4. Open reel, cartridge, and cassette

2-38. By which of the following methods are magnetic tape units categorized?

1. Type of packaging used for tape
2. Size of tape
3. Speed of tape
4. Cost of tape

2-39. What determines if a standard 1/2-inch tape will have either seven or nine tracks of data?

1. The brand of tape
2. The read/write heads installed in the tape unit
3. The type of computer used
4. The speed at which the tape unit is run

2-40. For multitrack tapes, what is the range of common recording densities in bits/bytes per inch (bpi)?

1. From 200 to 6,250 bpi
2. From 300 to 6,275 bpi
3. From 400 to 6,300 bpi
4. From 500 to 6,350 bpi

2-41. On magnetic tape, the size of a record that holds the data is restricted in what two ways?

1. By the thickness of the tape and the capacity of internal storage
2. By the length of the tape and the speed of internal storage
3. By the width of the tape and the speed of internal storage
4. By the length of the tape and the capacity of internal storage

2-42. In computer terminology, what is called a file?

1. A collection of tapes
2. A collection of disks
3. A collection of records
4. A collection of characters

2-43. In order for data to be read from or written on a magnetic tape, the tape must do what?

1. Speed up
2. Move at a predetermined speed
3. Slow down
4. Stop

2-44. Storing single records on a magnetic tape has which of the following disadvantages?

1. It takes too long to record the data
2. It takes too long to recover the data
3. Too much of the recording surface is wasted
4. Too much of the recording surface is used

2-45. The magnetic drum is another example of what type of access storage device?

1. Random
2. Direct
3. Multiple
4. Single

2-46. What is the speed range of a magnetic drum?

1. 300 to 3,000 rpm
2. 400 to 4,000 rpm
3. 500 to 5,000 rpm
4. 600 to 6,000 rpm

2-47. When using a magnetic drum, what is rotational delay?

1. Time that occurs in coming up to speed
2. Time that occurs in slowing down
3. Time that occurs in reaching a desired record location
4. Time that occurs in changing a drum

2-48. What is the storage capacity range of magnetic drums in characters or bytes of data?

1. From 20 million to more than 150,000 million
2. From 30 million to more than 150,000 million
3. From 40 million to more than 200,000 million
4. From 50 million to more than 200,000 million

2-49. Input data may be in any one of how many forms?

1. Five
2. Two
3. Three
4. Four

2-50. When data is input from a keyboard, a high average speed is how many characters per second?

1. One to two
2. Two to three
3. Three to four
4. Four to five

2-51. Output information is made available in how many forms?

1. One
2. Two
3. Three
4. Four

2-52. Magnetic tape stores data in what manner?

1. Sequential
2. Non-sequential
3. Direct
4. Random

2-53. The magnetic tape unit reads and writes data in channels or tracks along the length of the tape. How are these tracks referenced to each other?

1. Perpendicular
2. Parallel
3. Vertical
4. Random

2-54. How does a two gap head allow for increased speed?

1. By checking before writing
2. By using two gaps to write
3. By checking while writing
4. By using two gaps to check

2-55. What are the most common tape densities in bits/bytes per inch?

1. 500 and 1,000 bpi
2. 600 and 1,200 bpi
3. 700 and 1,500 bpi
4. 800 and 1,600 bpi

2-56. The drive motor of a disk drive unit rotates the disk at a constant speed, normally how many revolutions per minute?

1. 2,000 rpm
2. 2,500 rpm
3. 3,000 rpm
4. 3,600 rpm

2-57. The usual range of rotational speed for floppy disks is what?

1. 100 to 200 rpm
2. 200 to 300 rpm
3. 300 to 400 rpm
4. 400 to 500 rpm

2-58. The distance between the read/write head and the surface of a hard disk is called what?

1. The flying height
2. The disk height
3. The head height
4. The recording height

2-59. Floppy disks come in several sizes with diameters of what size range?

1. 2 to 6 inches
2. 3 to 8 inches
3. 4 to 9 inches
4. 5 to 10 inches

2-60. In the character-at-a-time impact printer class, which printer has the most professional-looking, pleasing-to-the-eye print?

1. Dot-matrix
2. Ink jet
3. Daisy-wheel
4. Laser

2-61. What is another name for the dot-matrix printer?

1. Hammer-matrix
2. Pin-matrix
3. Ink-matrix
4. Wire-matrix

2-62. Dot-matrix printers have which of the following ranges of speeds in characters per second?

1. 50 to 200 cps
2. 60 to 350 cps
3. 70 to 400 cps
4. 80 to 450 cps

2-63. Ink jet printers have what maximum speed in characters per seconds?

1. 300 cps
2. 400 cps
3. 500 cps
4. 600 cps

2-64. Laser printers can print up to approximately what total number of characters per second?

1. 20,666 cps
2. 22,666 cps
3. 24,666 cps
4. 26,666 cps

2-65. What are the two styles of typewriter keyboard arrangements used with a computer?

1. QWERYT or DVORAK
2. QWERTY or DVORAK
3. ABCDEF or DVORAK
4. ABCDEF or DVOARK

2-66. When working with display devices, what does the term soft-copy mean?

1. The information displayed is not permanent
2. The information displayed is permanent
3. The information displayed has a soft glow
4. The information displayed has no glow

2-67. On a raster scan crt, a raster is a series of what type of lines across the face of a crt?

1. Diagonal
2. Vertical
3. Horizontal
4. Wavy

2-68. Each field of a raster scan crt is made up of approximately how many lines?

1. 525
2. 550
3. 575
4. 600

2-69. In a video monitor, what do the frequency bandwidth, the number of characters to be displayed on a line, and the physical size of the screen determine?

1. Actual level of brightness
2. Actual number of picture elements
3. Actual speed of scan rate
4. Actual number of vertical lines that can be displayed

2-70. A monitor that uses 1,000 picture elements per line with a horizontal resolution of 1,000 can display what total number of vertical lines?

1. 10
2. 100
3. 1,000
4. 10,000

2-71. A raster frame is displayed approximately how many times a second?

1. 5
2. 10
3. 20
4. 30

2-72. To reduce the depth of the crt caused by the length of the tube, what type of displays were designed?

1. Wide panel
2. Flat panel
3. Narrow panel
4. Short panel

2-73. Compared to the gas plasma and electroluminescent displays, a liquid crystal display differs in which of the following ways?

1. It does not use as many picture elements
2. It does not use a light for the picture elements
3. It does not generate its own light for the picture elements
4. It does not have a backlight

2-74. The operation of an electroluminescent display requires what total number of volts?

1. 5
2. 10
3. 15
4. 20

ASSIGNMENT 3

Textbook assignment: Chapter 3, “Software,” pages 3-1 through 3-29.

3-1. What must you load into a computer to manage its resources and operations?

1. Bootstrap program
2. Word processor
3. Graphics program
4. Operating system

3-2. What program controls the execution of other programs according to job information?

1. An operating system
2. A bootstrap program
3. A word processor
4. A utility program

3-3. The simplest and most commonly used operating systems on microcomputers are which of the following types?

1. Multiuser/single tasking
2. Single user/single tasking
3. Single user/multitasking
4. Multiuser/multitasking

3-4. Which of the following programs must be compatible with the operating system in use?

1. CP/M-86
2. UNIX
3. Applications
4. MS-DOS

THIS SPACE LEFT BLANK INTENTIONALLY.

3-5. To overcome the applications software compatibility problem, which of the following is done so the application can be run under several different operating systems?

1. Some software comes in several versions
2. Computers are designed to accept all applications software
3. Software comes in a universal version
4. Operating systems are changed to be compatible

3-6. What is another term for "initial program load" the system?

1. Start
2. Boot
3. Kick
4. Run

3-7. When the symbol A> is on the screen of a computer crt, it tells the operator/user which of the following information?

1. The system is not ready, and drive A is busy
2. The system is ready, and drive A is assigned as your secondary drive
3. The system is ready, and drive A is assigned as your primary drive
4. The system is activating, and no drive is available

3-8. The three characters following each directory entry are called what?

1. Files
2. Records
3. Locators
4. Extensions

3-9. Commands built into the operating system that control actions, like diskcopy and rename, are what type of commands?

1. Independent
2. Copy
3. Spread
4. Utility

3-10. To eliminate the need for programmers to write new programs when all they want to do is copy, print, or sort a data file, which of the following types of programs can be used?

1. Word processor
2. Graphics
3. Utility
4. Spreadsheet

3-11. What is the term given to arranging records in a predefined sequence or order?

1. Sorting
2. Merging
3. Writing
4. Shifting

3-12. On a computer, what is the sequence of characters called?

1. Numerical sequence
2. Collating sequence
3. Random sequence
4. Alphabetic sequence

3-13. To sort a data file, what must you tell the sort program?

1. How many characters are in the file
2. How many records are in the file
3. The length of the data file
4. The data field or fields to sort on

3-14. Sort-merge programs usually have which of the following characteristics?

1. Specific file length
2. Specific run time
3. Phases
4. Names

3-15. What personnel or methods are used to generate programs to print detail and summary reports of data files?

1. Programmers
2. Operating systems
3. Sort-merge programs
4. Report program generators

3-16. What are report program generators designed to save?

1. Run time
2. Programming time
3. Operator time
4. Printer time

3-17. Each time there is a control break, what does the program developed by the report program generator print?

1. Input information
2. Output information
3. Summary information
4. Programming information

3-18. A computer language that is a string of numbers which represents the instruction code and operand addresses is what type?

1. Machine
2. Printed
3. Symbolic
4. Procedure-oriented

3-19. Mnemonic instruction codes and symbolic addresses were developed early in what decade?

1. 1940s
2. 1950s
3. 1960s
4. 1970s

3-20. Compared to machine language coding, symbolic languages have which of the following advantages?

1. Detail is reduced
2. Fewer errors are made
3. Less time is required to write a program
4. All of the above

3-21. An instruction that allows the programmer to write a single instruction which is equivalent to a specified sequence of machine instructions is what type of instruction?

1. Machine language instruction
2. Graphic language instruction
3. Macroinstruction
4. Scientific instruction

3-22. What does the acronym COBOL stand for?

1. Computer ordered byte oriented language
2. Computer ordered business oriented language
3. Common business oriented language
4. Common business ordered language

3-23. PASCAL is being used by many colleges and universities to teach programming for which of the following reasons?

1. It is fairly easy to learn and more powerful than BASIC
2. It is hard to learn and weaker than BASIC
3. It is easy to learn and cheaper than BASIC
4. It is a shorter course and produces better programmers

3-24. The development of Ada was initiated by what organization?

1. U. S. Navy
2. U. S. Army
3. U. S. Department of Defense
4. U. S. Department of Transportation

3-25. What are the two most familiar of the procedure-oriented languages used for scientific or mathematical problems?

1. PASCAL and FORTRAN
2. PASCAL and COBOL
3. COBOL and FORTRAN
4. BASIC and FORTRAN

3-26. Compared with programs written in symbolic languages, programs written in procedure-oriented languages differ in which of the following ways?

1. They can only be used with small computers
2. They can only be used with large computers
3. They can only be used with the computer for which the program was written
4. They can be used with a number of different computer makes and models

3-27. Compared with symbolic languages, procedure-oriented languages have which of the following disadvantages?

1. They require more space in memory, and they process data at a slower rate
2. They require more space in memory, and they process data too fast for some printers
3. They require a special memory, and they process data at a slower rate
4. They require a special memory, and they process data too fast for some printers

3-28. Which of the following is a simple definition of programming?

1. The process of planning which computer system to use
2. The process of planning the computer solution to a problem
3. The process of planning the mathematical solution to a problem
4. The process of planning which computer program to use

3-29. Which of the following is NOT a basic characteristic of a computer?

1. It needs commands
2. It needs specifically defined operations
3. It can think
4. It can understand instructions only in an acceptable form

3-30. How many fundamental and discrete steps are involved in solving a problem on a computer?

1. Five
2. Two
3. Three
4. Four

3-31. In the advance planning phase of programming, what are the first two steps?

1. Program coding and machine readable coding preparation
2. Problem understanding/ definition and flowcharting
3. Test data preparation and test run performance
4. Documentation completion and operator procedures preparation

3-32. Which of the following is NOT part of defining every aspect of a problem?

1. What information (or data) is needed
2. Where and how will the information be obtained
3. What is the desired output
4. What is the computation time

3-33. Once you have a thorough understanding of the problem, what is the next step in programming?

1. Gathering information
2. Coding the program
3. Flowcharting
4. Debugging

3-34. The method of pictorially representing a step-by-step solution to a problem before computer instructions are written to produce the desired results is called what?

1. Flowcharting
2. Constructing
3. Documenting
4. Debugging

3-35. What two types of flowcharts are there?

1. System and programming
2. System and data
3. Processing and programming
4. Processing and data

3-36. What are the four basic tools used in flowcharting?

1. Advanced symbols, graphic symbols, flowcharting template, and flowcharting worksheet
2. Fundamental symbols, graphic symbols, flowcharting template, and flowcharting worksheet
3. Fundamental symbols, mathematical symbols, flowcharting symbols, and flowcharting worksheet
4. Fundamental symbols, advanced symbols, flowcharting template, and flowcharting worksheet

QUESTION 3-37 IS TO BE JUDGED TRUE OR FALSE.

3-37. Fundamental symbols are standard for the military, as directed by Department of the Navy Automated Data Systems Documentation Standards, SECNAVINST 5233.1.

1. True
2. False

3-38. Within a flowchart, what type of symbols are used to specify arithmetic operations and relational conditions?

1. Fundamental symbols
2. Graphic symbols
3. Arithmetic symbols
4. Arabic symbols

3-39. What is the graphic symbol for less than or equal to?

1. >
2. <
3. ≤
4. ≥

QUESTION 3-40 IS TO BE JUDGED TRUE OR FALSE.

3-40. The flowchart worksheet is a means of standardizing documentation.

1. True
2. False

3-41. To develop a flowchart, which of the following must you know first?

1. What type of computer is to be used
2. What problem you are to solve
3. What code you are going to use
4. What logic the computer will use to solve a problem

3-42. In solving a problem, which of the following ways does a computer operate?

1. Two steps at a time in random order
2. It processes the problem as a whole
3. One step after another in specified order
4. One step after another in random order

3-43. What is the step called in which you code a program that can be translated by a computer into a set of instructions it can execute?

1. Program booting
2. Program execution
3. Program logic
4. Program coding

QUESTION 3-44 IS TO BE JUDGED TRUE OR FALSE.

3-44. It is important to remember program coding is the first step of programming.

1. True
2. False

3-45. Before sitting down to code the computer instructions to solve a problem, you should complete which of the following activities?

1. A course in computer operation
2. A course in mathematics
3. Planning and coding
4. Planning and preparation

3-46. What is the fundamental element in program preparation?

1. Subject
2. Predicate
3. Computer
4. Instruction

3-47. The first part of a computer instruction, which answers the question what, is known by which of the following terms?

1. Operation only
2. Command only
3. Command or operation
4. Operand

3-48. The second specific part of the predicate in a computer instruction, known as the operand, in general answers what question?

1. Who
2. What
3. When
4. Where

3-49. What part of the program must the programmer prepare according to the format required by the language and the computer to be used?

1. Documentation
2. Implementation
3. Instructions
4. Length

3-50. To copy data from one storage location to another and to rearrange and change data elements in some prescribed manner, what type of instructions are used?

1. Input/output
2. Data movement
3. Transfer of control
4. Conditional logic

3-51. Transfer of control instructions are classified as which of the following types?

1. Conditional only
2. Unconditional only
3. Conditional and unconditional
4. Conditional and distributed

3-52. Errors caused by faulty logic and coding mistakes are referred to as what?

1. Mistakes
2. Errors
3. Faults
4. Bugs

3-53. The process of carefully checking the coding sheets before they are keyed into the computer is known as what?

1. Desk-checking
2. Code-checking
3. Program-checking
4. Computer-checking

3-54. A definition of the problem, a description of the system, a description of the program, and operator instructions make up what package?

1. Training
2. Security
3. Orientation
4. Documentation

3-55. Which of the following is another name for packaged software?

1. Rented programs
2. Manufactured programs
3. Off-the-shelf programs
4. On-the-shelf programs

3-56. Under the word processing software control, you generally enter the text using what method?

1. Tape
2. Disk
3. Drum
4. Keyboard

QUESTION 3-57 IS TO BE JUDGED TRUE OR FALSE.

3-57. Spelling checker software helps find misspelled words not misused words.

1. True
2. False

3-58. What type of software allows you to enter data and then retrieve it in a variety of ways?

1. Communications
2. Data retrieval
3. Data management
4. Document compilation

QUESTIONS 3-59 AND 3-60 ARE TO BE JUDGED TRUE OR FALSE.

3-59. Spreadsheets are tables of rows and columns of text.

1. True
2. False

3-60. You can use all printers for graphics output.

1. True
2. False

Assignment 4

Assignment: Topic 4, "Data Representation and Communications"
Pages: 4-1 through 4-17

4-1. In using a digital computer, which of the following is one of the major problems we face?

1. Finding disks to fit the drives
2. Locating a stable power source
3. Communicating with the computer
4. Arranging the proper environment

4-2. In computer terminology, what is a general term to describe raw facts?

1. Characters
2. Bits
3. Bytes
4. Data

4-3. In computer terminology, when data has been processed with other facts and has meaning, it is described as which of the following?

1. Information only the computer can understand and properly use
2. Information we can understand and properly use
3. Information the input device can understand and properly use
4. Information the output device can understand and properly use

4-4. Data is represented by which of the following means?

1. By symbols
2. By electricity
3. By magnetics
4. By mechanics

QUESTION 4-5 IS TO BE JUDGED TRUE OR FALSE.

4-5. The first computers were designed to manipulate numbers to solve arithmetic problems.

1. True
2. False

4-6. What is data to be represented called?

1. Numeric data
2. Alphanumeric data
3. Information data
4. Source data

4-7. Raw data is typically written on some type of paper document referred to as what type of document?

1. End document
2. Source document
3. Classified document
4. Unclassified document

4-8. Numeric, alphabetic, and special characters are represented in a computer's internal storage and on magnetic media through the use of what kind of system?

1. Coding
2. Reading
3. Writing
4. Labeling

4-9. It is possible to represent a maximum of 256 different characters or bit combinations by using which of the following codes?

1. 8-bit
2. 16-bit
3. 32-bit
4. 64-bit

4-10. In addition to four numeric bits, there are four other bit positions used in an a-bit code, what are they called?

1. Area bits
2. Zone bits
3. Region bits
4. District bits

4-11. Which of the following numbering systems has a base of 16?

1. Octal
2. Binary
3. Decimal
4. Hexadecimal

4-12. Representing two numeric characters in one byte (eight bits) is referred to as what?

1. Packing
2. Stacking
3. Doubling
4. Crowding

4-13. By packing data within an 8-bit code, which of the following results are achieved?

1. Storage space required increases and processing speed increases
2. Storage space required increases and processing speed decreases
3. Storage space required decreases and processing speed increases
4. Storage space required decreases and processing speed decreases

4-14. Through the cooperation of several manufacturers, what a-bit code was developed for transmitting and processing data?

1. EBCDIC
2. EBCDTC
3. ASCIT
4. ASCII

QUESTION 4-15 IS TO BE JUDGED TRUE OR FALSE.

4-15. The concepts and advantages of ASCII are identical to those of EBCDIC.

1. True
2. False

4-16. The letter D is represented by what coding in (a) EBCDIC and (b) ASCII?

1. (a) 1111 0011
   (b) 0011 0011
2. (a) 0011 0011
   (b) 1111 0011
3. (a) 0100 0100
   (b) 1100 0100
4. (a) 1100 0100
   (b) 0100 0100

4-17. An additional bit in each storage location called a parity bit is used for what purpose?

1. To stop errors
2. To erase errors
3. To detect errors
4. To reroute errors

4-18. The parity bit is also called what?

1. Odd bit
2. Even bit
3. Code bit
4. Check bit

4-19. The test for bit count is called what?

1. Odd check
2. Even check
3. Stop check
4. Parity check

4-20. What storage area accepts and holds input data to be processed?

1. Input
2. Output
3. Working
4. Program

4-21. A single binary digit is called what?

1. Bit
2. Byte
3. Word
4. Record

4-22. What symbol is used when we refer to the size of computer memory?

1. M
2. m
3. K
4. k

4-23. When many magnetic cores are strung together on a screen of wire, what type of plane is formed?

1. A wire plane
2. A core plane
3. A screen plane
4. A magnetic plane

4-24. When a core is magnetized, what characteristic of magnetism determines whether it contains a binary 0 or a binary 1?

1. Amount
2. Duration
3. Direction
4. Saturation

4-25. The storage capacity of an address is designed and built into the computer by which of the following people or organizations?

1. Operator
2. Installer
3. Programmer
4. Manufacturer

4-26. Computers that are built to retrieve, manipulate, and store a fixed number of characters in each address are said to be which of the following types of computers?

1. Fixed-bit-length
2. Fixed-word-length
3. Fixed-file-length
4. Fixed-number-length

4-27. Computers that store a single character in each address location are said to be which of the following types of computers?

1. Variable-addressable
2. Data-addressable
3. Fixed-addressable
4. Character-addressable

QUESTION 4-28 IS TO BE JUDGED TRUE OR FALSE.

4-28. Fixed-word-length computers have slower calculating speeds than character-addressable computers.

1. True
2. False

4-29. Flexible computers that are byte oriented can operate in either a fixed- or variable-word-length mode by which of the following techniques?

1. Proper program length
2. Proper program density
3. Proper program instructions
4. Proper program flexibility

4-30. Different word lengths, such as half-word, full-word, and double-word, are possible with what type of computer?

1. Bit-addressable
2. Byte-addressable
3. File-addressable
4. Record-addressable

4-31. When a flexible computer is working in a fixed-word-length environment, each address identifies what group of elements that can be operated on as a unit?

1. Bits
2. Bytes
3. Files
4. Records

4-32. To automatically retrieve, manipulate, and store a fixed word of data as a unit on a flexible computer, what means can a programmer use?

1. Parity bit
2. Program length
3. Storage capacity
4. Program instructions

4-33. In computer terminology, a group of related bits is known by which of the following terms?

1. Word
2. File
3. Record
4. Character

4-34. What group of related items form a record?

1. Bits
2. Bytes
3. Fields
4. Characters

4-35. The variations in how data files are stored in secondary storage is determined by what?

1. Types of media and devices used
2. Cost of the installation
3. Size of the installation
4. Type of power available

4-36. When you store a file on tape, the 125th record cannot be read until the 124 records in front of it are read. This is called what type of storage?

1. Input-access
2. Direct-access
3. Random-access
4. Sequential-access

4-37. When data can be obtained quickly from anywhere on the media without having to read the records in front of it, which of the following types of storage is being used?

1. Reading-access
2. Direct-access
3. Sequential-access
4. Processing-access

4-38. Which of the following is an example of random-access storage?

1. Disk
2. Thin film
3. Paper tape
4. Magnetic tape

4-39. Any system composed of one or more computers and terminals is the definition for what?

1. Network
2. ADP system
3. Computer system
4. Supply system

4-40. A network that consists of various machines linked together within a building or adjacent buildings is what type?

1. Wide area
2. Linked area
3. Local area
4. Narrow area

4-41. When dissimilar machines have the ability to communicate, they act in which of the following ways?

1. Human
2. As a team
3. Individually
4. Against each other

4-42. For local area networks, what two designs are used?

1. Broadband and bandpass
2. Broadpass and bandpass
3. Broadbase and baseband
4. Broadband and baseband

4-43. The communications channel that uses the basic frequency band of radio waves and a coaxial cable is what type?

1. Broadband
2. Broadpass
3. Baseband
4. Bandpass

4-44. The transmission of voice as well as data and text can be handled by what type of communications channels?

1. Broadband
2. Broadpass
3. Baseband
4. Bandpass

4-45. Wide area networks are sometimes referred to as which of the following networks?

1. Local
2. Global
3. Satellite
4. Telephone

4-46. The first successful communications satellite for business applications was launched in what year?

1. 1955
2. 1959
3. 1962
4. 1965

4-47. When we transmit data directly to a computer over long distances, it becomes necessary to add two other devices, one at each end of the communications line. These devices are called what?

1. Modems
2. Printers
3. Converters
4. Input/output buffers

4-48. A modem converts the digital signal produced by your terminal or the computer to what type of signal suitable for transmission over the communications line?

1. Video
2. Audio
3. Carrier
4. Hybrid

4-49. If conversion of the digital signal to be transmitted were not carried out, it would degenerate and become what?

1. Lost
2. Strong
3. Doubled
4. Garbled

4-50. Telephone lines are a frequently used type of communications channel. They are often referred to by which of the following terms?

1. Land lines
2. Microwave link
3. High frequency link
4. Communications lines

4-51. In communications, what name is given to those devices that serve to interconnect?

1. Connectors
2. System controllers
3. Interface elements
4. Impedance matchers

4-52. When using a modem, what are the two methods of data transmission?

1. Digital and analog
2. Mechanical and light
3. Continuous and non-continuous
4. Asynchronous and synchronous

4-53. The transmission method that uses a single set of start and stop message characters per block of data is which of the following types?

1. Synchronous
2. Asynchronous
3. Microwave link
4. Frequency modulated

4-54. Whenever data is transferred between devices, it also involves an exchange of prearranged signals known as what?

1. Nodding
2. Spacing
3. Handshaking
4. Coordinating

4-55. The specific set of rules used to govern handshaking and message characters is called what?

1. Sending
2. Protocol
3. Modulator
4. Transmitting

NONRESIDENT
TRAINING
COURSE
SEPTEMBER 1998

Navy Electricity and Electronics Training Series

Module 23—Magnetic Recording

NAVEDTRA 14195

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Although the words "he," "him," and "his" are used sparingly in this course to enhance communication, they are not intended to be gender driven or to affront or discriminate against anyone.

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: To introduce the student to the subject of Magnetic Recording who needs such a background in accomplishing daily work and/or in preparing for further study.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and either the occupational or naval standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards,* NAVPERS 18068.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
CTMC(SS) Milton Charles Georgo

NAVSUP Logistics Tracking Number
0504-LP-026-8460

Sailor’s Creed

“I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country’s Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all.”

TABLE OF CONTENTS

CHAPTER **PAGE**

CREDITS

The figures listed below and included in this edition of NEETS, Module 23, *Magnetic Recording*, were provided by Datatape, Inc. Permission to use these illustrations is gratefully acknowledged.

SOURCE	FIGURE
Datatape, Inc.	3-2, 3-3, 3-4, 5-3, 6-3, 6-4 A&B, 7-1, 7-4, 7-5, 7-6

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7, *Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 2 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

THIS PAGE LEFT BLANK INTENTIONALLY.

Student Comments

Course Title: *NEETS Module 23*
Magnetic Recording

NAVEDTRA: 14195 **Date**: ______________________

We need some information about you:

Rate/Rank and Name: ______________________ SSN: ____________ Command/Unit ______________

Street Address: ______________________ City: ____________ State/FPO: ________ Zip ________

Your comments, suggestions, etc.:

Privacy Act Statement: Under authority of Title 5, USC 301, information regarding your military status is requested in processing your comments and in preparing a reply. This information will not be divulged without written authorization to anyone other than those within DOD for official use in determining performance.

NETPDTC 1550/41 (Rev 4-00)

CHAPTER 1

INTRODUCTION TO MAGNETIC RECORDING

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. They serve as a preview of the information you are expected to learn in the chapter. The comprehension check questions placed within the text are based on the objectives. By successfully completing those questions and the associated NRTC, you show that you have met the objectives and have learned the information. The learning objectives for this chapter are listed below.

After completing this chapter, you will be able to do the following:

1. Describe the history and purpose of magnetic recording.

2. State the prerequisites for magnetic recording.

3. Describe a magnetic recording head, how it's constructed, and how it operates.

INTRODUCTION

Have you ever wondered how a whole album of your favorite music got onto one of those little cassette tapes? Or, what about computer floppy disks; have you ever wondered how they can hold 180 or more pages of typed text? The answer to both of these questions is *magnetic recording*.

Magnetic recording devices seldom get much attention until they fail to work. But without magnetic recording, recording your favorite television show on a video cassette recorder would be impossible, portable tape players wouldn't exist, and you wouldn't be able to get money from an automated bank teller machine at two o'clock in the morning.

Now what about the Navy? Could it operate without magnetic recording? The answer is definitely no. Without it:

- Computer programs and data would have to be stored on either paper cards or on rolls of paper tape. Both of these methods need a lot of storage space, and they take much longer to load into and out of the computer.

- There wouldn't be any movies to show or music to play on the ship's entertainment system when the ship is at sea and is out of range for television and radio reception.

- Intelligence-collection missions would be impossible since you couldn't store the collected signals for later analysis.

As you can see, magnetic recording plays a very important part both in our Navy life and in our civilian life.

HISTORY OF MAGNETIC RECORDERS

In 1888, Oberlin Smith originated the idea of using permanent magnetic impressions to record sounds. Then in 1900, Vladeniar Poulsen brought Mr. Smith's dream to reality. At the Paris Exposition, he demonstrated a Telegraphone. It was a device that recorded sounds onto a steel wire. Although everyone thought it was a great idea, they didn't think it would succeed since you had to use an earphone to hear what was recorded. It wasn't until 1925, when electronic amplifiers were developed, that magnetic recording started to receive the attention it deserved.

The best magnetic recording is the one that produces an output signal identical to the input signal. It didn't take long to realize that the magnetism generated during the recording process didn't vary directly to the current which caused it. This is because there's a *step* in the magnetism curve where it crosses the zero point and changes polarity. This step causes the output signal to be distorted when compared with the input signal. Figure 1-1 shows this step.

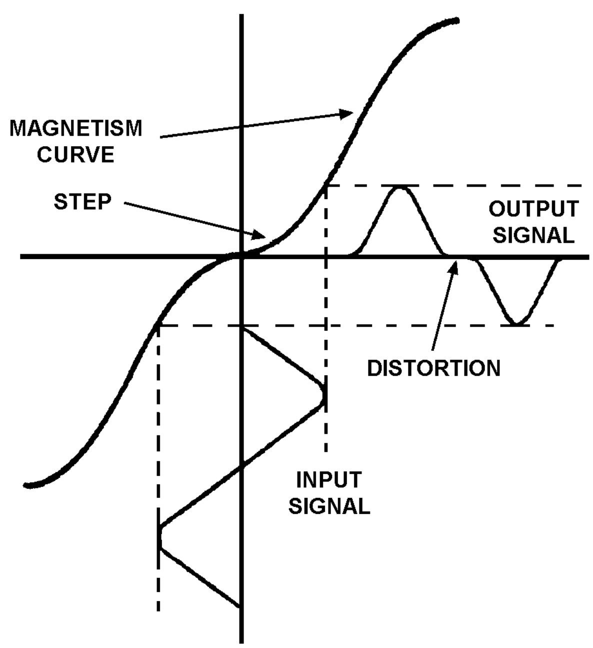


Figure 1-1.—Magnetic recording without bias voltage.

In 1907, Mr. Poulsen discovered a solution to this problem. He discovered *dc* bias. He found that if a fixed dc voltage were added to the input signal, it moved the input signal away from the *step* in the magnetism curve. This prevented the input signal from crossing the zero-point of the magnetism curve. The result is an output signal exactly like the input signal. Figure 1-2 shows this process.

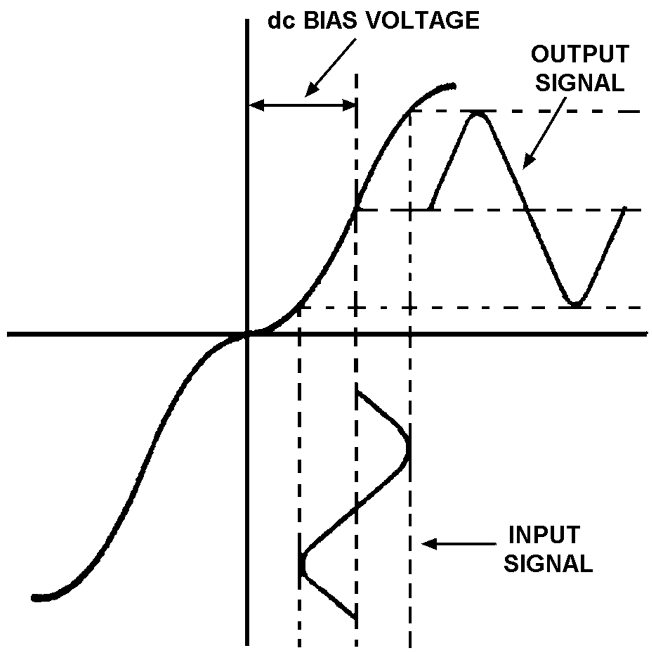


Figure 1-2.—Magnetic recording with dc bias voltage.

Unfortunately, dc bias had its problems. Since only a small portion of the magnetism curve was straight enough to use, the output signal was weak compared with the natural hiss of the unmagnetized tape passing the playback head. This is commonly called poor signal-to-noise ratio (SNR). We'll explain SNR in more detail later.

From the beginning, the U.S. Naval Research Laboratories (NRL) saw great potential in magnetic recording. They were especially interested in using it to transmit telegraph signals at high speed. After electronic amplifiers were invented around 1925, W.L. Carlson and G.W. Carpenter at the NRL made the next important magnetic recording discovery. They found that adding an ac bias voltage to the input signal instead of a fixed dc bias voltage would

- reproduce a stronger output signal
- greatly improve the signal-to-noise ratio
- greatly reduce the natural tape hiss that was so common with dc bias

To make ac bias work, they used an ac frequency for the bias voltage that was well above what could be heard, and a level that placed the original input signal away from both *steps* in the magnetism curve. This resulted in two undistorted output signals that could be combined into one strong output. See figure 1-3.

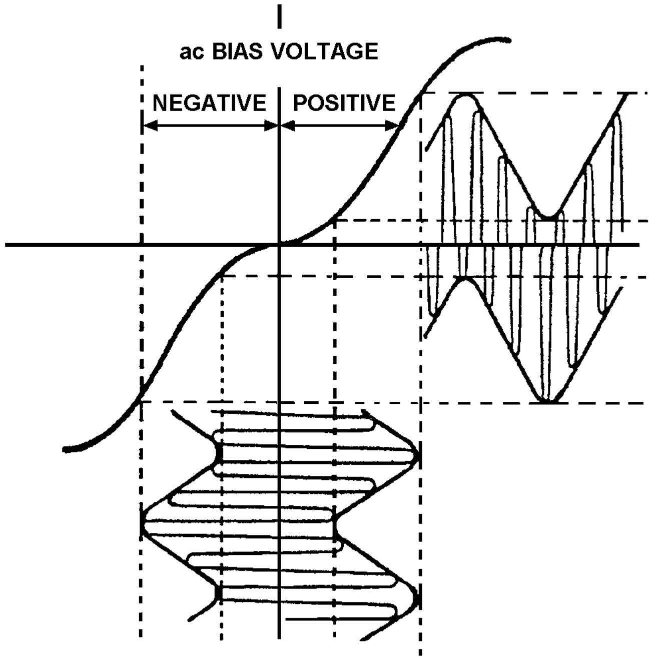


Figure 1-3.—Magnetic recording with ac bias voltage.

Until 1935, all magnetic recording was on steel wire. Then, at the 1935 German Annual Radio Exposition in Berlin, Fritz Pfleumer demonstrated his Magnetophone. It used a cellulose acetate tape coated with soft iron powder. The Magnetophone and its "paper" tapes were used until 1947 when the 3M Company introduced the first plastic-based magnetic tape.

In 1956, IBM introduced the next major contribution to magnetic recording—the hard disk drive. The disk was a 24-inch solid metal platter and stored 4.4 megabytes of information. Later, in 1963, IBM reduced the platter size and introduced a 14-inch hard disk drive.

Until 1966, all hard disk drives were "fixed" drives. Their platters couldn't be removed. Then in 1966, IBM introduced the first removable-pack hard disk drive. It also used a 14-inch solid metal platter.

In 1971, magnetic tape became popular again when the 3M Company introduced the first 1/4-inch magnetic tape cartridge and tape drive. In that same year, IBM invented the 8-inch floppy disk and disk drive. It used a flexible 8-inch platter of the same material as magnetic tape. Its main goal was to replace punched cards as a program-loading device.

The next contribution to magnetic recording literally started the personal computer (PC) revolution. In 1980, a little-known company named Seagate Technology invented the 5-1/4-inch floppy disk drive. Without it, PCs as we know them today would not exist.

From then on, it was all downhill. Magnetic tape became more sophisticated. Floppy disks and disk drives became smaller, while their capacities grew bigger. And hard disk capacities just went through the roof. All of the major hurdles affecting magnetic recording had been successfully cleared, and it was just a matter of refining both its methods and materials.

Q-1. Why did the early inventors of magnetic recording find it necessary to add a fixed dc bias to the input signal?

Q-2. How does dc bias added to the input signal correct the distortion in the output signal?

Q-3. Why does adding dc vice ac bias voltage to the input signal result in a poor signal-to-noise ratio (SNR)?

Q-4. What are three advantages of adding an ac bias voltage to the input signal instead of adding a fixed dc bias voltage?

Q-5. Why does using ac vice dc bias voltage result in a stronger output signal?

PREREQUISITES FOR MAGNETIC RECORDING

To perform magnetic recording, you need three things:

1. An input signal you wish to record.
2. A recording medium. (This is a recording surface that will hold the signal you wish to record.)
3. A magnetic head to convert the input signal into a magnetic field so it can be recorded.

If any one of these are missing, magnetic recording cannot take place.

Input Signal

An input signal can come from a microphone, a radio receiver, or any other source that's capable of producing a recordable signal. Some input signals can be recorded immediately, but some must be *processed* first. This processing is needed when an input signal is weak, or is out of the frequency response range of the recorder.

Recording Medium

A recording medium is any material that has the ability to become magnetized, in varying amounts, in small sections along its entire length. Some examples of this are magnetic tape and magnetic disks. These are thoroughly discussed in chapter 2 of this module.

Magnetic Heads

Magnetic heads are the heart of the magnetic recording process. They are the transducers that convert the electrical variations of your input signal into the magnetic variations that are stored on a recording medium. Without them, magnetic recording isn't possible.

Magnetic heads actually do three different things. They transfer, or *record*, the signal information onto the recording medium. They recover, or *reproduce*, the signal information from the recording medium. And they remove, or erase, the signal information from the recording medium.

MAGNETIC HEAD CONSTRUCTION.—A magnetic head is a magnetic core wrapped with a coil of very thin wire (see figure 1-4). The core material is usually shaped like the letter C, and is made from either iron or ceramic-ferrite material. The number of turns of wire placed on the core depends on the purpose of that specific head. The gap in the core is called a *head gap*. It's here that magnetic recording actually takes place. We'll go into more detail of magnetic head construction in chapter 3.

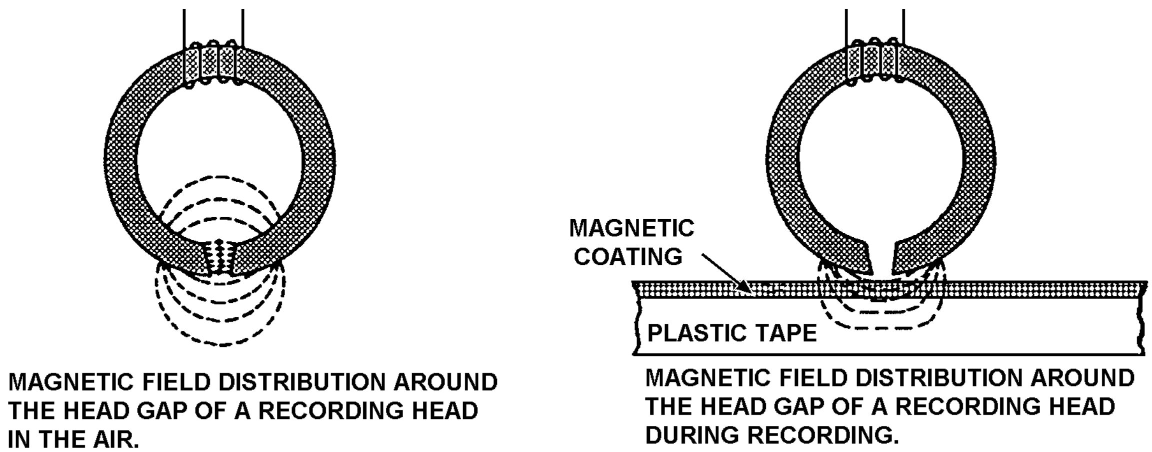


Figure 1-4.—Magnetic field distribution around the head gap.

MAGNETIC HEAD OPERATION.—Whether you're recording on magnetic tapes or disks, all magnetic heads operate the same way. When an electric current passes through the coil of a magnetic head, magnetic field lines associated with the electric current follow paths through the core material. When the magnetic fields get to the head gap, some of them spread outside the core to form a *fringing field*. When a recording medium is passed through this fringing field, it is magnetized in relation to the electric current. This is called magnetic recording. Figure 1-4 illustrates this process.

Q-6. What three things are required to perform magnetic recording?

Q-7. What is the meaning of the term recording medium as it pertains to magnetic recording?

Q-8. What are the three functions of the magnetic heads on a magnetic recording device?

SUMMARY

This chapter briefly covered the historical development of magnetic recording principles and devices. The following is a summary of important points in the chapter.

The **BEST MAGNETIC RECORDING** is one that produces an output signal that is identical to the input signal. However, a *step* in the magnetic curve causes the output signal to be distorted.

In 1907, **DC BIAS** was added to the input signal to remove the distortion in the output signal. But the dc bias caused a weak output signal with a poor SNR. Around 1925, the NRL used AC BIAS to reproduce a stronger output signal and greatly improve the SNR.

To perform magnetic recording, you need (1) an **INPUT SIGNAL**, (2) a **RECORDING MEDIUM**, and (3) a **MAGNETIC HEAD**.

A **RECORDING MEDIUM** is any material that can become magnetized in varying amounts (such as magnetic tape and disks).

MAGNETIC HEADS are used to (1) *record* the signal onto the recording medium, (2) *reproduce* the signal from the recording medium, and (3) *erase* the signal from the recording medium.

ANSWERS TO QUESTIONS Q1. THROUGH Q8.

A-1. Because a step in the magnetism curve where it crosses the zero point and changes polarity causes the output signal to be distorted. See figure 1-1.

A-2. The dc bias moves the input signal away from the step in the magnetism curve. This prevents the input signal from crossing the zero-point of the magnetism curve. See figure 1-2.

A-3. With dc bias, the SNR is poor because only a small portion of the magnetism curve is straight enough to use, thus the output signal is weak compared with the natural tape hiss.

A-4.

- *a. Reproduces a stronger output signal.*
- *b. Greatly improves the SNR.*
- *c. Greatly reduces the natural tape hiss.*

A-5. Because an ac bias voltage of the proper frequency and level places the input signal away from both steps in the magnetism curve. The result is two undistorted output signals that are combined into one strong output.

A-6.

- *a. An input signal.*
- *b. A recording medium.*
- *c. A magnetic head.*

A-7. A recording medium is any material that has the ability to become magnetized, in varying amounts, in small sections along its entire length.

A-8.

- *a. Record the signal onto the recording medium.*
- *b. Reproduce the signal from the recording medium.*
- c. *Erase the signal from the recording medium.*

CHAPTER 2

MAGNETIC TAPE

LEARNING OBJECTIVES

After completing this chapter, you'll be able to do the following:

1. Describe the physical properties of magnetic tape in terms of:

   a. The Three Basic Materials Used To Make Magnetic Tape.

2. The function of the magnetic tape's *base material*, *oxide coating*, and *binder glue*.

3. Describe the two types of magnetic recording tape.

4. Describe the following types of tape errors and their effects on magnetic tape recording: *signal dropout*, *noise*, *skew*, and *level*.

5. Describe the following causes of magnetic tape failure: *normal wear*, *accidental damage*, *environmental damage*, and *winding errors*.

6. Describe the purpose and makeup of tape reels and tape cartridges.

7. Describe the two methods for erasing magnetic tape, the characteristics of automatic and manual tape degaussers, and the procedures for degaussing magnetic tape.

8. Describe the proper procedures for handling, storing, and packaging magnetic tape, tape reels, and tape cartridges.

PHYSICAL PROPERTIES OF MAGNETIC TAPE

The three basic materials used to make magnetic tape are (1) the base material, (2) the coating of magnetic oxide particles, and (3) the glue to bind the oxide particles onto the base material. See figure 2-1.

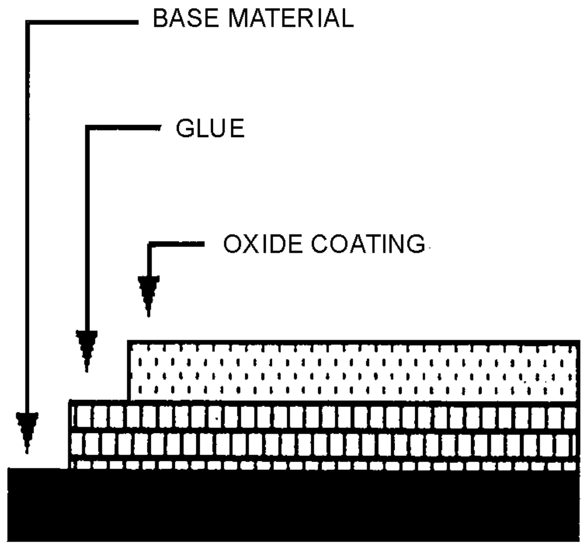


Figure 2-1.—Magnetic tape construction.

BASE MATERIAL

The base material for magnetic tape is made of either plastic or metal. Plastic tape is used more than metal tape because it's very flexible, it resists mildew and fungus, and it's very stable at high temperatures and humidity.

OXIDE COATING

Oxide particles that can be magnetized are coated onto the base material. The most common magnetic particles used are either gamma ferric oxide or chromium dioxide. It's very important that these magnetic particles are uniform in size. If they're not, the tape's surface will be abrasive and will reduce the life of the recorder's magnetic heads.

An ideal magnetic particle is needle-shaped. It's actual size depends on the frequency of the signal to be recorded. Generally, long particles are used to record long wavelength signals (low-frequency signals), and short particles are used to record short wavelength signals (high-frequency signals).

GLUE

The glue used to bond the oxide particles to the base material is usually an organic resin. It must be strong enough to hold the oxide particles to the base material, yet be flexible enough not to peel or crack.

TYPES OF MAGNETIC RECORDING TAPE

There are two basic types of magnetic recording tape in common use: *analog* and *digital*. Analog magnetic tape is used to record, store, and reproduce audio and instrumentation type signals. These signals are usually in a frequency band from very-low frequency (VLF) to 2.5 MHz. Digital magnetic tape is used to record, store, and reproduce computer programs and data. It's base material thickness is about 50 percent thicker than analog magnetic tape. This allows the digital tape to withstand the more strenuous starts and stops associated with digital magnetic recorder search, read, and write functions.

Digital magnetic tape is also held to much stricter quality control standards. It's important not to have any blemishes or coating flaws on the tape's surface. Because, if you lost one digital data bit, your computer program or data would be bad. In contrast, losing one microsecond of an analog signal is not nearly as critical.

Q-1. Magnetic tape is made of what three basic materials?

Q-2. Why is plastic magnetic tape used more than metal tape?

Q-3. Which of the two types of magnetic tape is used to record audio and instrumentation type signals in the VLF to 2.5MHz frequency range?

Q-4. What type of magnetic tape is used to record computer programs and data, and what are the additional thickness and quality standards for this type of tape?

TAPE ERRORS AND THEIR EFFECTS

Four types of tape errors that will degrade the performance of a magnetic recording system are signal dropout, noise, skew, and level (signal amplitude changes).

DROPOUT ERRORS

Signal dropout is the most common and the most serious type of tape error. It's a temporary, sharp drop (50% or more) in signal strength caused by either contaminates on the magnetic tape or by missing oxide coating on part of the tape.

During recording and playback, the oxide particles on the tape can flake off and stick to the recorder's guides, rollers, and heads. After collecting for awhile, the oxide deposits (now oxide lumps) break loose and stick to the magnetic tape. As the tape with the lumps passes over the head, the lumps get between the tape and the head and lift the tape away from the head. This causes the signal dropouts. Although oxide lumps cause most signal dropouts, remember that any contaminate (such as dust, lint or oil) that gets between the tape and the head can cause signal dropouts.

NOISE ERRORS

Noise errors are unwanted signals that appear when no signal should appear. They're usually caused by a cut or a scratch on the magnetic tape. It's the lack of oxide particles at the cut or the scratch that causes the noise error.

SKEW ERRORS

Skew errors only occur on multi-track magnetic tape recorders. The term skew describes the time differences that occur between individual tracks of a single magnetic head when the multi-track tape isn't properly aligned with the magnetic head.

There are two types of skew errors: *fixed* and *dynamic*. Fixed skew happens when properly aligned magnetic tape passes an improperly aligned magnetic head. Dynamic skew happens when misaligned tape passes a properly aligned head. This type of skew is usually caused by one or more of the following:

- A misaligned or worn-out tape transport system.
- A stretched or warped magnetic tape.

- A magnetic tape that is improperly wound on a reel.

LEVEL ERRORS

Magnetic tape is manufactured to have a specified output signal level (plus or minus some degree of error). Level errors happen when the actual output signal level either drops or rises to a level outside the expected range. For example, if a magnetic tape is rated for 10 volts (+/–10%), any output signal level below 9 volts or above 11 volts is a level error. Level errors are caused by an uneven oxide coating on the magnetic tape. This can come from either the original manufacturing process or from normal wear and tear.

Some causes of level errors are permanent and cannot be removed by any means. For example, a crease in the tape, a hole in the oxide, or a damaged edge. Other causes of level errors are removable and may be cleaned off the tape. For example, oxide flakes or clumps, metallic particles, or dirt are removable.

Q-5. What are four types of tape errors that can degrade a magnetic recording system's performance?

Q-6. What are signal dropouts, and what are two tape defects that can cause signal dropouts?

Q-7. What is the most common and most serious type of signal dropout?

Q-8. You see a build-up of dust and lint on the take-up reel of a tape recorder. This can cause which of the four types of tape errors?

Q-9. What type of tape error causes noise to appear on the tape when no signal should appear? What causes this type of tape error?

Q-10. The multi-track tape recorder in your computer system has a fixed skew error. What does this mean and what is the probable cause?

Q-11. Some tapes you are using may have level errors. What does this mean and what is the cause?

CAUSES OF MAGNETIC TAPE FAILURE

Tape failure happens when a magnetic tape's performance degrades to a point where it's no longer usable. The *exact* point where failure occurs will vary, depending on the type of tape and how it is used.

There are four main causes for tape failure:

1. Normal wear (natural causes)

2. Accidental damage

3. Environmental damage

4. Winding errors

NORMAL WEAR

Normal wear occurs because the tape must come in contact with fixed surfaces, such as a recorder's magnetic heads, rollers, and guides. Over time, this repeated contact with the fixed surfaces causes excessive dropout errors and makes the tape unusable.

ACCIDENTAL DAMAGE

Accidental tape damage that causes tape failure is any damage that wouldn't normally occur under ideal operating and handling conditions. It can be caused by either a human operator or the tape recorder itself. Accidental tape damage caused by human operators can range from accidentally dropping a reel of magnetic tape to improperly threading a magnetic tape recorder. Accidental tape damage caused by recording equipment can occur if the recorder is poorly designed or if the tape transport mechanism is adjusted improperly.

ENVIRONMENTAL DAMAGE

The negative effect of environmental extremes on tape can also cause tape failure. Magnetic tape is very flexible and can be used in a wide range of environmental conditions. It's designed for use in a temperature range of about 2 to 130 degrees Fahrenheit (−20 to 55 degrees Celsius), and in a relative humidity range of about 10 to 95%. Of course, these numbers are the *extreme*. Ideally, magnetic tape should be used and stored at a temperature of about 60 to 80º F (room temperature), and in a relative humidity of about 40 to 60%.

Large changes from the ideal relative humidity cause tape to expand or contract and thus affect the uniformity of a tape's oxide coating. High relative humidity causes the tape to stretch and increases the tape's friction. The increased friction causes increased head wear, head clog by oxide particles, and head-to-tape sticking. Low relative humidity encourages oxide shedding and increases static build-up on tape surfaces, causing the tape to collect airborne contaminants.

The effects of exceeding the ideal temperature and humidity ranges described above can cause the following environmental damage to magnetic tape: *tape deformation, oxide shedding, head-to-tape sticking, layer-to-layer sticking, dirt build-up, and excessive tape and head wear.*

Tape Deformation

Magnetic tapes are wound onto tape reels with tension applied. This tension causes great layer-to-layer pressure within the reel pack. Changes in temperature and humidity can cause the backing material to expand or contract, creating even more pressure. All of this pressure causes the tape to become deformed or warped.

Oxide Shedding

At temperatures above 130º F, a tape's oxide coating tends to become soft. At temperatures below 2º F, the oxide coating tends to be brittle. In both cases, the oxide coating will shed, flake off, or otherwise become separated from the base material. These free pieces of oxide will then stick to parts of the tape transport, to the magnetic heads, or back onto the tape and cause dropout or level errors.

Head-to-Tape Sticking

At higher temperatures, the tape binder glue can soften to the point where it will stick to the recorder's magnetic head. This head-to-tape sticking causes jerky tape motion.

Layer-to-Layer Adhesion

When reels of magnetic tape are stored at higher temperatures, the tape's binder glue may melt and cause the layers of tape to stick to one another. In very severe cases, layer-to-layer adhesion can separate the oxide coating from the base material and completely destroy a tape.

Dirt Build-up

Dirt build-up happens when the relative humidity level is less than 10%. The low humidity causes static electricity that attracts dirt and dust which builds up on the magnetic tape and other parts of the magnetic tape recorder.

Excessive Tape and Head Wear

When the relative humidity is more than 95%, the high humidity causes increased friction as the tape passes over the heads. This, in turn, causes excessive tape and head wear.

Q-12. What is tape failure?

Q-13. What are four main causes of tape failure?

Q-14. How does normal wear cause tape failure?

Q-15. Accidental damage to magnetic tape is normally caused by the tape recorder itself or by human operators of the recorder. What are three frequent causes of such accidental damage?

Q-16. Environmental damage to magnetic tape can occur when the tape is stored in an area that exceeds what <u>ideal</u> temperature and humidity ranges?

Q-17. What six types of environmental damage can occur to tapes in storage when the ideal temperature and humidity ranges are exceeded?

Q-18. After using a tape that was stored in an area where temperatures exceeded 130º F you notice pieces of oxide sticking to the recorder's tape-transport mechanism, to its magnetic heads, and onto the tape. What is the probable cause of these symptoms?

Q-19. Your activity stores its magnetic tape in an area where the temperature is 100º F. What two types of environmental damage could occur that would make these tapes unusable?

Q-20. When the relative humidity is below 10%, what happens to magnetic tape and parts of a tape recorder that could cause environmental damage?

Q-21. How does relative humidity over 95% cause excessive tape and head wear?

WINDING ERRORS

Winding errors are another cause of tape failure. They happen when improper winding practices create an excessive or uneven force as the tape is being wound onto a tape reel. The form taken by the tape after it is wound onto the reel is called the *tape pack*. Winding errors can cause a deformed tape pack that will prevent good head-to-tape contact.

In most cases, a deformed tape pack can be fixed simply by rewinding it onto another reel at the proper tension and at the right temperature and humidity. The four most common types of tape pack deformation are:

1. Cinching
2. Pack-slip
3. Spoking

4. Windowing

Cinching

Cinching happens when a tape reel is stopped too quickly. The sudden stop causes the outer layers of magnetic tape to continue to spin after the inner layers have stopped. This causes any loosely wound tape within the pack to unwind and pile up. Figure 2-2 shows an example of a cinched tape pack (note the complete foldover of one tape strand).

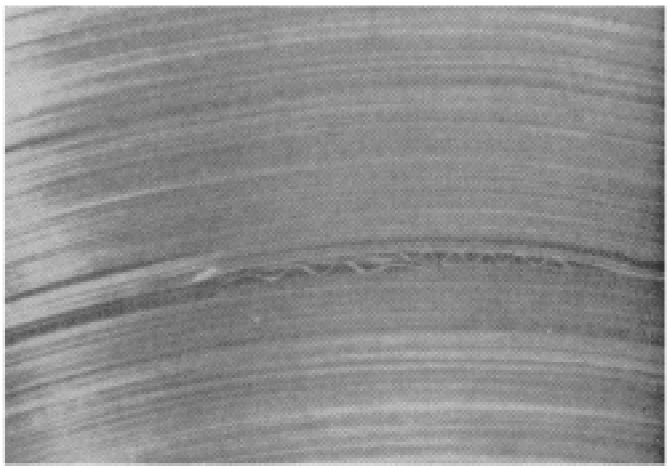


Figure 2-2.—Example of cinched tape pack.

Pack Slip

Pack slip happens when the tape is loosely wound on the reel and is exposed to excessive vibration or too much heat. This causes the tape to shift (side-to-side), causing *steps* in the tape pack. When a tape reel with pack slip is used, the magnetic tape will unwind unevenly and rub against the sides of the tape reel or the recorder's tape guides. This can damage the magnetic tape and cause oxide shedding. Figure 2-3 shows an example of pack slip.

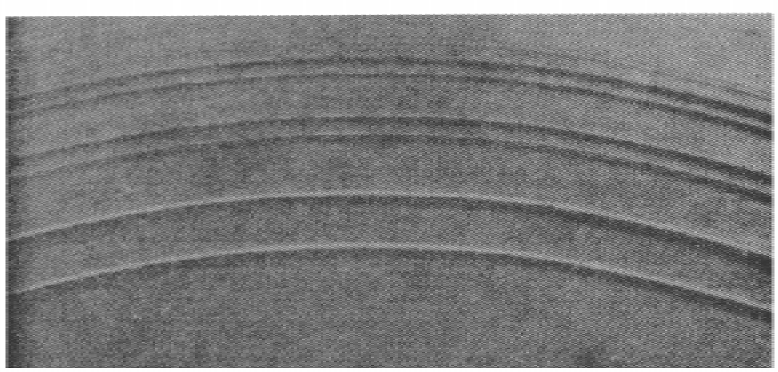


Figure 2-3.—Example of pack slip.

Spoking

Spoking happens when magnetic tape is wound onto the tape reel with the tension increasing toward the end of the winding. The higher tension on the outside of the tape pack causes the inner pack to buckle and deform. Spoking is also caused by the uneven pressures created when a tape is wound on a reel that has a distorted hub, or when the tape is wound over a small particle that is deposited on the hub. Figure 2-4 shows a spoked tape pack.

Figure 2-4.—Example of spoked tape pack.

Windowing

Windows are voids or see-through air gaps in the tape winding. They happen when magnetic tape is loosely wound onto a tape reel, and especially when the loosely wound reel is later exposed to extreme heat or humidity. Figure 2-5 shows a windowed tape pack.

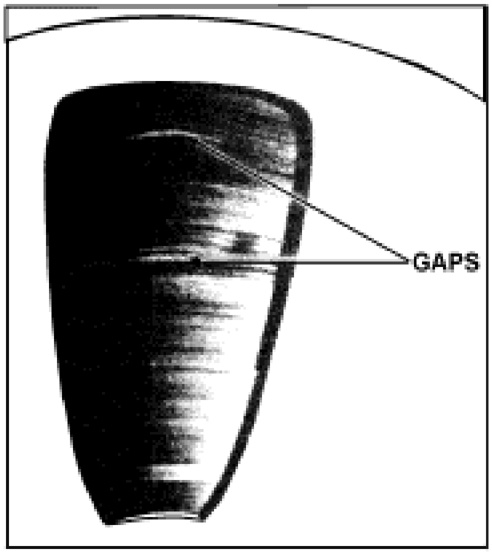


Figure 2-5.—Example of windowed tape pack.

Q-22. Tape winding errors can cause a deformed tape pack. What are four common types of tape pack deformation?

Q-23. After rewinding a tape onto its supply reel, you examine the tape pack and notice pile-ups of tape resembling the example in figure 2-2. What causes this condition?

Q-24. You notice steps in the tape pack such as those in figure 2-3. What causes this and how does it damage the magnetic tape?

Q-25. A tape pack is buckled and deformed as shown in figure 2-4. What are three possible causes for this condition?

Q-26. A tape pack has gaps in the tape winding as shown in figure 2-5. What causes this condition?

TAPE REELS AND TAPE CARTRIDGES

There are two types of magnetic tape carriers: tape *reels* and *tape* cartridges. Both types can be used for either analog or digital recording. Tape cartridges are normally used only for digital recording.

TAPE REELS

Tape reels are used on magnetic recorders that use a manually loaded tape supply reel and a separate take-up reel. A reel's purpose is to protect the magnetic tape from damage and contamination. It can be made of plastic, metal, or glass. A reel has two parts, the hub and the flanges.

A tape reel is designed to hold magnetic tape on its hub without letting the magnetic tape touch the sides of the flanges. Contrary to popular belief, the flanges <u>are not</u> designed to *guide* the magnetic tape onto the tape reel.

TAPE CARTRIDGES

Tape cartridges hold a spool of magnetic tape in the same way as tape reels, except that the inside of the cartridge contains both the supply reel and the take-up reel. Unlike tape reels which must be manually loaded into a recorder, when you insert a tape cartridge into a recorder, it's automatically loaded and ready to use. Figure 2-6 shows two typical tape cartridges.

Figure 2-6.—Typical tape cartridges.

Q-27. When winding a tape onto a plastic or metal reel, should the tape ever touch the reel's flanges?

TAPE ERASING AND DEGAUSSING

One advantage of magnetic tape is that you can erase what you've previously recorded, and record on the same tape again and again. The erasing is done by demagnetizing the magnetic tape. You demagnetize a magnetic tape by exposing it to a gradually decreasing ac (alternating current) magnetic field. There are two ways to do this: (1) with an *erase head* that's mounted on the magnetic recorder, or (2) with a separate *tape degausser*.

ERASE HEADS

A magnetic recorder's erase head erases magnetic tape by saturating it with an ac signal that's higher in frequency than the frequency range of the recorder itself. This method of erasing a tape works well in some cases, but it's not the best way because:

- It's slow; the tape must be run through the recorder to be erased.

- If the erase head is not completely demagnetized, it may not do a complete erasure.
- Some recorders do not have erase heads installed.

MAGNETIC TAPE DEGAUSSERS

By far, the best way to erase a magnetic tape is to use a separate magnetic tape degausser. There are two types of degaussers: *automatic* and *manual*.

Automatic Tape Degausser

Automatic degaussers erase magnetic tape by automatically moving the whole tape reel or cartridge slowly and steadily in and out of an intense ac magnetic field. This type of degausser erases a tape very well. Some automatic degaussers are made specifically for tape reels, and some are made for both tape reels and tape cartridges. Figure 2-7 shows a typical automatic degausser.

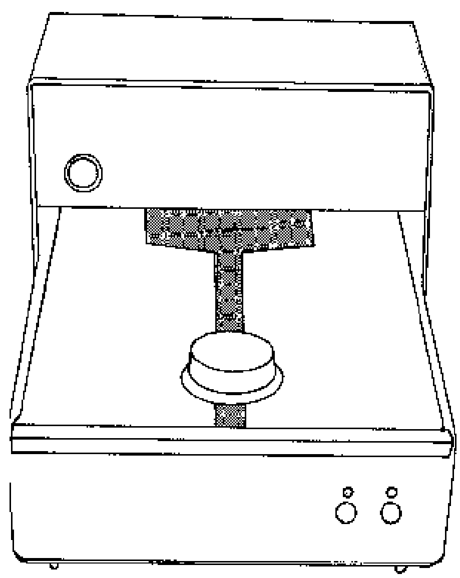


Figure 2-7.—Automatic tape degausser.

Manual Tape Degausser

Both manual and automatic tape degaussers use the same electronic principles for erasing magnetic tape. However, the manual version is much more portable. It's small, hand-held, and much less expensive. Figure 2-8 shows a typical manual degausser.

Figure 2-8.—Manual tape degausser.

To erase tapes with a manual degausser:

1. Place the tape reel or cartridge to be erased on a flat surface.

2. Hold the degausser very close to the magnetic tape and turn it on.

3. Slowly rotate the degausser in circles around the tape reel or cartridge for a few seconds.

4. Then slowly move it away until you're about 12 to 14 inches away from the tape reel or tape cartridge.

5. Turn off the degausser.

Q-28. *What are two disadvantages of using a recorder's erase head to erase data recorded on a magnetic tape?*

Q-29. *What method for erasing magnetic tape is much more effective and reliable than using a recorder's erase head?*

HANDLING, STORING, AND PACKAGING MAGNETIC TAPE

Today's magnetic tape coatings can store recorded signals for years. The data recorded is a permanent record that won't fade or weaken with age. And, it'll remain unchanged until it's altered by another magnetic field or until the tape coating deteriorates.

When magnetic tape recordings are ruined, the cause is usually poor handling, improper storage, or shipping damage. If you want your tape recordings to last a long time, you need to know how to properly handle, store, and ship magnetic tape.

HANDLING MAGNETIC TAPE

A magnetic tape reel or cartridge should always be in one of two places, either mounted on a tape recorder or in its storage container. When you handle magnetic tape, follow these rules:

- DO use extreme care when handling magnetic tape. Careless handling can damage magnetic tape, tape reels, and tape cartridges. Always hold a tape reel by the hub, NEVER by the flanges, and NEVER handle or touch the working tape surface.

- DO NOT let the magnetic tape trail on the floor. Even though the end of the tape may not have data stored on it, it can pick up dirt and dust that ends up on the recorder.

- DO clean your hands before handling magnetic tape. You can contaminate magnetic tape with dirt and oils from dirty hands.

- DO mount tape reels and cartridges properly. Improperly seated tape reels can cause unnecessary wear and tear on the magnetic tape.

- DO replace any warped take-up reels, as they can damage magnetic tape.

- DO keep the magnetic recorder and its take-up reel clean. Magnetic tape can pick up dirt and dust from the recorder itself.

- DO NOT use the top of a magnetic recorder as a work area. This can expose the magnetic tape to dirt, excessive heat, and stray magnetic fields.

- DO NOT allow eating, drinking, or smoking in areas where magnetic tape or devices are exposed.

STORING MAGNETIC TAPE

Most magnetic tape reels and cartridges spend a lot of time in storage. It's very important that you protect the stored tape from physical damage and the damaging effects of contamination and temperature and humidity extremes. If you don't, damage to the tape pack such as oxide shedding, layer-to-layer sticking, and tape deformation can happen. To protect magnetic tape from damage during storage, follow these rules:

- DO make sure that magnetic tape is wound properly on the reel hub and at the proper tension.

- DO always store tape reels vertically. DO NOT lay them on their side.

- DO maintain a proper environment. Keep the storage area clean, and at a 60 to 80F degree temperature and a 40 to 60% relative humidity.

- DO NOT store magnetic tapes near any equipment that generates stray magnetic fields.

- DO handle <u>all</u> tape reels and cartridges as gently as possible.

- DO NOT eat, drink, or smoke in a magnetic tape storage area.

PACKAGING MAGNETIC TAPE FOR SHIPPING

There may be times when you are asked to package magnetic tape reels or tape cartridges for shipment. If you want the tape to arrive in good condition, you must pack it properly to protect it from damage. The packaging you use must protect the tape reels or cartridges from impact, vibration, and temperature and humidity changes. Here are some simple rules to follow:

- DO always package tape reels so that they're supported by their hub. This prevents any pressure on the reel's flanges that might flex the flanges against the tape pack. Figure 2-9 shows a shipping box that supports the tape reel by the hub.

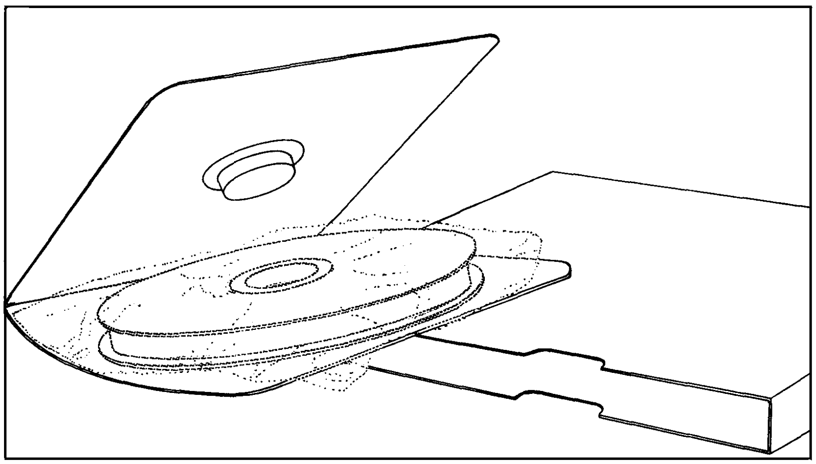


Figure 2-9.—Reel box that supports reel by the hub.

- DO always use reel bands where available. Reel bands are for placement around the outside edges of the reel flanges to help prevent the flanges from flexing and damaging the tape.

- DO always ship magnetic reels in a container designed so its normal positioning is with the reels in a vertical position. This will prevent the tape pack from shifting and damaging the edges of the magnetic tape.

- DO always package tape cartridges in their shipping cases. Tape cartridges are more durable than tape reels, but they still need to be protected during shipment.

Q-30. When magnetic tapes are ruined, what three factors are normally the cause?

Q-31. What is the correct way to hold a magnetic tape reel?

Q-32. The take-up reel on your recorder is warped. What should you do to/with the reel?

Q-33. If magnetic tape is stored in areas with temperature and humidity extremes, what are three types of tape damage that may occur?

Q-34. List four rules you should follow when storing magnetic tape to protect it from damage.

Q-35. When packaging tape reels or cartridges for shipping, what are four rules you should follow to protect the tape reels from impact and vibration?

SUMMARY

Now that you've finished chapter 2, you should be able to (1) describe the physical properties of magnetic tape, (2) recognize the four most common magnetic tape errors, (3) recognize the four causes of tape failure, (4) describe the two methods for erasing magnetic tape, and (5) use the proper procedures for handling, storing, and packaging magnetic tape, tape reels, and tape cartridges. The following is a summary of the important points in this chapter.

The three **BASIC MATERIALS** used to make magnetic tape are the (1) base material, (2) the oxide particles, and (3) the binder glue.

ANALOG and **DIGITAL** are the two basic types of magnetic tape in common use.

BLEMISHES OR **COATING FLAWS ON DIGITAL TAPE** can easily ruin the data or the computer program stored on the tape.

SIGNAL DROPOUT, NOISE, SKEW, AND **LEVEL** are four types of tape errors. Dropout errors are the most common.

OXIDE LUMPS accumulated on the tape cause most dropout errors. Other causes are dust or lint on the tape, or missing oxide coating on part of the tape.

MAGNETIC TAPE FAILURE has four main causes: (1) normal wear, (2) accidental damage, (3) environmental damage, and (4) winding errors.

IDEAL TEMPERATURE AND HUMIDITY RANGES for using and storing magnetic tape are 60 to 80º F and 40 to 60% relative humidity.

ENVIRONMENTAL TAPE DAMAGE caused by excessive temperature or humidity includes the following: (1) tape deformation, (2) oxide shedding, (3) head-to-tape sticking, (4) layer-to-layer sticking, (5) dirt buildup, and (6) excessive tape and head wear.

WINDING ERRORS can cause tape pack deformation. The four most common types are: (1) cinching, (2) pack slip, (3) spoking, and (4) windowing.

The **TWO PARTS OF A TAPE REEL** are the hub and the flanges. The tape should be wound on the hub. No part of the tape should be touching the flange sides.

ERASE HEADS AND **TAPE DEGAUSSERS** are two methods for erasing tape. Degaussers are the fastest and the most reliable.

Rules for **HANDLING MAGNETIC TAPE** are (1) always hold the reel by the hub, not the flanges, (2) never touch the working tape surface, (3) replace warped or damaged reels, and (4) mount reels and cartridges properly.

Rules for **STORING MAGNETIC TAPE** are (1) wind tape properly on the reel hub, (2) store tapes vertically, (3) keep storage area clean and at proper temperature and humidity levels, and (4) store tapes away from equipment that generates stray magnetic fields.

Rules for **PACKAGING TAPE FOR SHIPPING** are (1) support reels by their hubs, (2) use reel bands, (3) pack reels in containers vertically, and (4) keep tape cartridges in their shipping cases.

ANSWERS TO QUESTIONS Q1. THROUGH Q35.

A-1.

a. Base material.

b. Coating of magnetic oxide particles.

c. Glue that bonds the particles to the base.

A-2. Plastic tape is used more than metal because it's more flexible, resists mildew and fungus, and is very stable at high temperatures and humidity.

A-3. Analog magnetic tape.

A-4. Digital magnetic tape is for computer programs and data. Its base material is about 50% thicker. The tape's surface must not have blemishes or coating flaws because losing even one digital data bit could ruin the recorded computer program or data.

A-5. Signal dropout, noise, skew, and level. Dropout is the most common.

A-6. Dropouts are temporary, sharp drops (50% or more) in signal strength. They're caused by contaminates that lift the tape away from the magnetic head, or when magnetic oxide coating is missing on part of the tape.

A-7. Oxide particles that get onto the magnetic tape.

A-8. Signal dropout errors and level errors. The dust and lint on the reel will eventually get onto the tape where it can get between the tape and the recorder's heads.

A-9. Noise error is usually caused by a cut or a scratch on the magnetic tape.

A-10. Skew means there are time differences between the individual tracks of a multi-track recorder's magnetic head. It happens when the tape isn't properly aligned with the head. Fixed skew happens when the tape passes over an improperly aligned magnetic head.

A-11. The actual output signal level of the tape exceeds the manufacturer's specified range for the output signal level (+ / – 10%). It's caused by an uneven oxide coating on the tape due to worn tape or defective manufacture.

A-12. Tape's performance degrades to a point where it's no longer usable.

A-13. Normal wear, accidental damage, environmental damage, and winding errors.

A-14. Repeated contact with a recorder's fixed surfaces such as magnetic heads, tape rollers, and tape guides.

A-15.

a. Improperly adjusted tape transport mechanism.

b. Dropping a reel of tape.

c. Improperly threading tape.

A-16. Ideally, use and store tape at 60 to 80º F and at 40 to 60% relative humidity.

A-17. Tape deformation, oxide shedding, head-to-tape sticking, layer-to-layer sticking, dirt build-up, and excessive tape and head wear.

A-18. Oxide shedding. At temperatures above 130º F, oxide coating becomes soft and sheds.

A-19. Head-to-tape sticking and layer-to-layer adhesion.

A-20. Dirt build-up caused by static electricity.

A-21. High humidity causes increased friction as the tape passes over the heads.

A-22. Cinching, pack slip, spoking, and windowing.

A-23. The tape is stopped too quickly when winding or rewinding.

A-24. Pack slip. It's caused by loosely wound tape on a reel that is exposed to excessive vibration or heat. The vibration or heat causes the tape to shift, causing steps in the tape pack. The uneven tape will then rub against the reel's sides and the recorder's tape guides.

A-25.

- *a. Reel has a distorted hub,*
- *b. tape wound over small particle deposited on hub, and*
- *c. tape wound on reel with tension increasing toward end of winding.*

A-26. Tape is loosely wound on reel.

A-27. No. The reel is designed to hold the tape on its hub without letting the tape touch the sides of the flanges.

A-28. Using an erase head is slow, and it may not completely erase the tape.

A-29. Using a magnetic tape degausser.

A-30. Poor handling, improper storage, or shipping damage.

A-31. Always hold reel by the hub, never by the flanges. Never touch the working tape surface.

A-32. Always replace a warped reel.

A-33. Oxide shedding, layer-to-layer sticking, and tape deformation.

A-34.

- *a. Make sure the tape is wound properly on the reel hub,*
- *b. store tapes vertically,*
- *c. keep storage area at right temperature and humidity,*
- *d. store away from equipment that generates stray magnetic fields.*

A-35.

a. *Package reels so they're supported by their hub,*

b. *use reel bands,*

c. *package reels in vertical position,*

d. *package tape cartridges in their shipping cases.*

CHAPTER 3

MAGNETIC TAPE RECORDER HEADS

LEARNING OBJECTIVES

After completing this chapter, you'll be able to do the following:

1. Describe the construction, function, and placement of magnetic tape recorder record, reproduce, and erase heads.
2. Describe the preventive maintenance requirements for magnetic tape recorder heads.

MAGNETIC TAPE RECORDER HEADS

Magnetic tape recorder heads are the *heart* of magnetic tape recording, because it's the magnetic heads (as we'll call them in this chapter) that actually:

1. Record signal or data information onto magnetic tape
2. Reproduce (play back) signal or data information from magnetic tape
3. Erase any signal or data off of magnetic tape

To do these things, a magnetic tape recorder can have up to three different heads installed: one head for recording, one for reproducing, and one for erasing. Some magnetic tape recorders will use the same head for both recording and reproducing. Figure 3-1 shows a typical multitrack magnetic head.

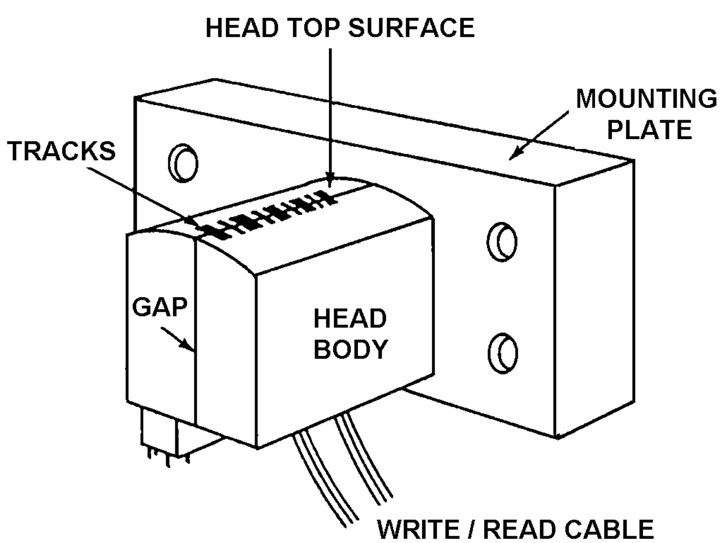


Figure 3-1.—Typical multitrack magnetic tape recorder head.

MAGNETIC HEAD CONSTRUCTION

Magnetic head construction is *basically* the same for all magnetic heads. They're all made up of a magnetic core wrapped with a coil of very thin wire. But, there's where the similarity ends. From here on, each magnetic head is built to perform a specific job. Will the head be used on a single track recorder? Will it be used on a multitrack recorder? Will it be a record head or a reproduce head? Or, will it be an erase head? What frequency will it be recording and/or reproducing? The answers to these questions will determine the final construction of a magnetic head.

Figure 3-2 shows the construction of a typical multitrack magnetic head. Magnetic cores are wound with very thin wire, cemented together, and placed inside a half-bracket. A tip piece is then placed on top of the ferrite core, and the two half-brackets are assembled together. It's during this final assembly process that the headgap and the resulting frequency response of the magnetic head are determined. After some final contouring to give the magnetic head its curved face, it's ready for use.

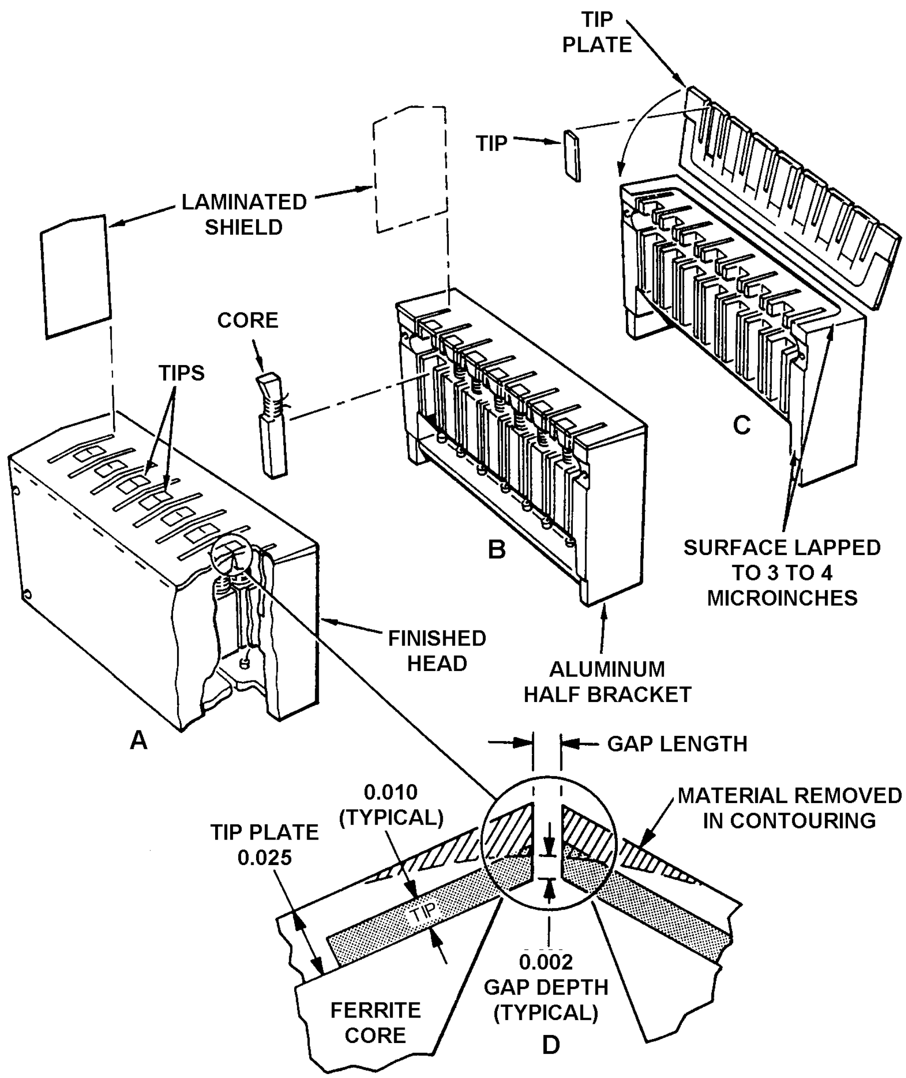


Figure 3-2.—Multitrack tape recorder head construction.

Record and Reproduce Heads

Record and reproduce heads convert and transfer electrical signals onto and off of magnetic tape. The maximum frequency these heads can transfer depends on the size of the headgap and the speed of the magnetic tape (we'll discuss speed in the next chapter). Most record and reproduce heads are in one of these three general bandwidth categories:

1. Narrowband—100 Hz to 100 kHz
2. Intermediate band—100 Hz to 500 kHz
3. Wideband—400 Hz to 2 mHz

The only physical difference between a record head and a reproduce head is in the number of turns of wire on the core. A reproduce head will have more turns than a record head. This is because reproduce heads must be able to recover low-level signals from magnetic tape. The extra turns of wire allow the reproduce head to output the highest level possible and at a good signal-to-noise level.

Record heads are always placed before reproduce heads on magnetic tape recorders. This allows you to monitor signals that you're recording. Figure 3-3 shows the placement of record and reproduce heads. Figure 3-4 shows some of the typical track arrangements used.

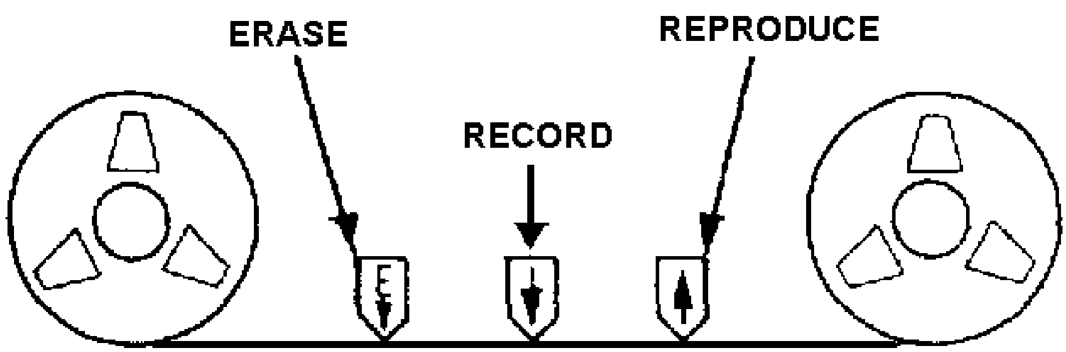
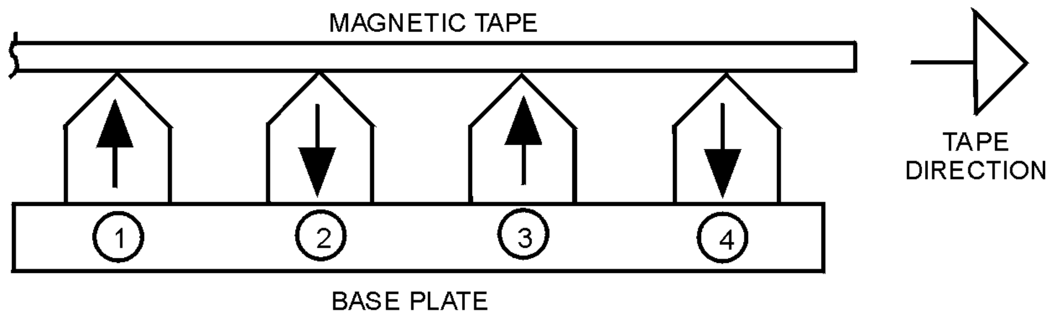


Figure 3-3.—Placement of magnetic tape recorder heads.



1 - ODD RECORD STACK
2 - ODD REPRODUCE STACK
3 - EVEN RECORD STACK
4 - EVEN REPRODUCE STACK

TRACK NUMBERING

HEAD STACK	14 TRACK	28 TRACK	42 TRACK
1	1-13 RECORD	1-27 RECORD	1-41 RECORD
2	1-13 REPRODUCE	1-27 REPRODUCE	1-41 REPRODUCE
3	2-14 RECORD	2-28 RECORD	2-42 RECORD
4	2-14 REPRODUCE	2-28 REPRODUCE	2-42 REPRODUCE

Figure 3-4.—Magnetic head track placement.

Erase Heads

Erase heads transfer a signal to the magnetic tape that causes the magnetic particles to assume a *neutralized* or *erased* state. To do this, a high current signal that is 3 to 5 times higher in frequency than the maximum frequency response of the record and reproduce heads is used. In some audio recorders, a simple direct current (dc) voltage is used.

Erase heads are always placed before the record and the reproduce heads on tape recorders. This allows you to erase the magnetic tape before it's recorded on. Figure 3-3 shows the placement of erase heads.

Q-1. Magnetic tape recorders can have up to three different heads installed. What are the three functions performed by a recorder's heads?

Q-2. The way a magnetic head will be used determines how it is constructed. Name three factors that determine the final construction of a magnetic head.

Q-3. What two specifications determine the maximum frequency that a recorder's record and reproduce heads will be able to transfer?

Q-4. Most record and reproduce heads are in one of what three bandwidth categories?

Q-5. Why are record heads always placed before reproduce heads on recorders?

Q-6. A recorder's erase head is always placed in what sequence on the record/reproduce track?

MAGNETIC HEAD MAINTENANCE

It's very important to *regularly* maintain magnetic heads. If you do, you'll greatly reduce the chance of getting a poor recording or playback. Regular preventive maintenance will also increase the life of the magnetic heads. There are two things you must do to properly maintain magnetic heads: (1) keep them clean, and (2) keep them demagnetized.

Cleaning Magnetic Heads

Through use, magnetic heads pick up dirt, dust, lint, and oxide particles from the magnetic tape. These particles collect on the magnetic head and, if left unchecked, could cause signal dropout errors that degrade the quality of recording and playback. To keep magnetic heads clean, regularly clean them with a cotton-tipped applicator soaked in either isopropyl alcohol or in a magnetic head cleaner recommended by the recorder's manufacturer. A good rule of thumb is to clean the heads each time you change a tape reel or cartridge.

Demagnetizing Magnetic Heads

Magnetic heads can become magnetized from many sources. It could happen

- during ac power losses,
- during testing or alignment,
- because of stray magnetic fields,
- from normal use.

No matter the cause, magnetized magnetic heads degrade the quality of the magnetic recording or playback.

To demagnetize magnetic heads, you'll use a hand-held degausser. It could be like the one shown in figure 3-5, or like the manual degausser shown in the previous chapter. No matter how they look, they all generate an ac magnetic field that demagnetizes the metal parts of a magnetic head.

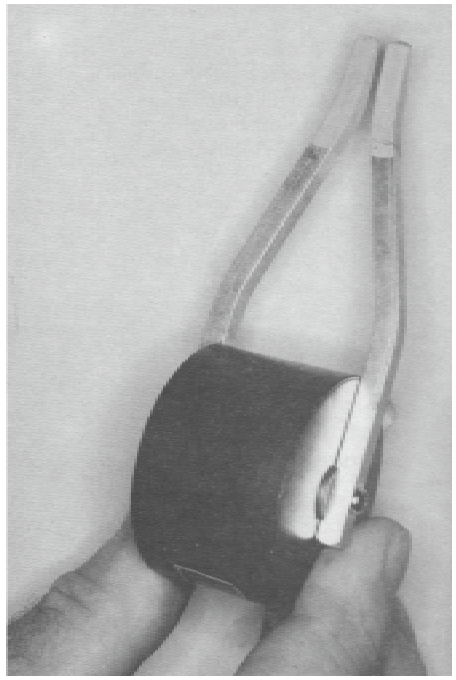


Figure 3-5.—Hand-held head degausser.

The procedure for demagnetizing a magnetic head is similar to the procedure for degaussing a magnetic tape. Here are the basic steps:

1. Remove the tape (reel or cartridge) from the magnetic recorder.

2. Holding the degausser an arm's length away from the magnetic head, energize the degausser.

3. *Slowly* bring the degausser closer and closer to the magnetic head. **Don't touch the head with the degausser**.

4. Move the degausser back and forth across the head for 15 to 30 seconds. Figure 3-6 shows how this looks.

5. *Slowly* move the degausser away from the magnetic head. When the degausser is an arm's length away, de-energize it.

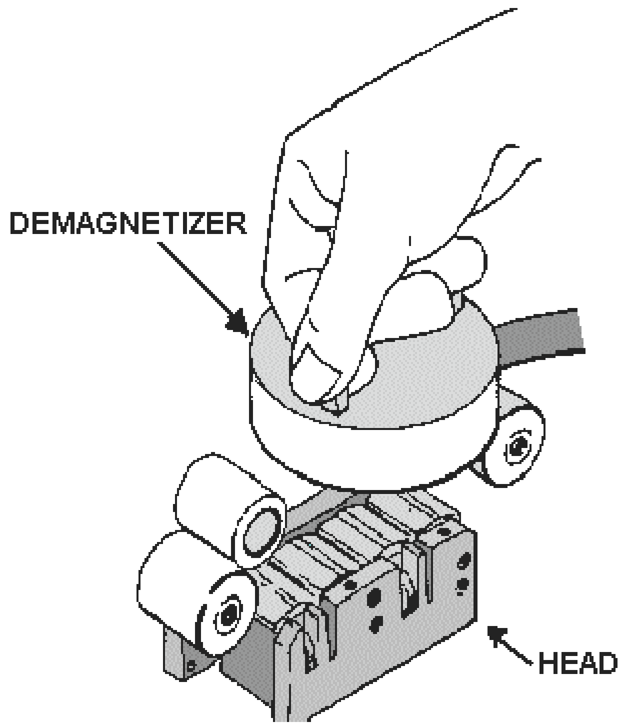


Figure 3-6.—Demagnetizing magnetic heads with a degausser.

That's all there is to it. It's hard to determine *exactly* how often magnetic heads should be de-magnetized. Manufacturer's recommendations vary from every 8 to 25 hours of operation. To be safe, check the equipment's technical manual.

Q-7. What two preventive maintenance actions must you do regularly to increase magnetic head life and to ensure good tape recording and playback?

Q-8. How should you clean your recorder's magnetic heads?

Q-9. What are four sources that can cause magnetic heads to become magnetized?

Q-10. What type of equipment should you use to demagnetize your recorder's magnetic heads?

Q-11. How often should you demagnetize a recorder's magnetic heads?

SUMMARY

Now that you've finished chapter 3, you should be able to (1) describe the construction of magnetic tape recorder heads; (2) describe the purpose and placement of record, reproduce, and erase heads; and (3) describe the preventive maintenance requirements for tape recorder heads. The following is a summary of important points in this chapter:

Magnetic tape recorders have up to **THREE MAGNETIC HEADS** to perform the erase, record, or reproduce function.

Three factors that determine the **CONSTRUCTION OF A MAGNETIC HEAD** are the (1) type of head, (2) frequencies it will record, reproduce, or erase, and (3) use on a single or multitrack recorder.

Most tape recorder heads are designed for **ONE OF THREE BANDWIDTHS**: (1) narrowband, (2) intermediate band, or (3) wideband.

A recorder's magnetic heads are in the following **SEQUENCE** on its record/reproduce track: (1) erase, (2) record, and (3) reproduce.

Two important **PREVENTIVE MAINTENANCE** requirements for magnetic heads are cleaning and demagnetizing.

ANSWER TO QUESTIONS Q1. THROUGH Q11.

A-1. Record, reproduce, and erase.

A-2.

a. Type of head (record, reproduce, or erase).

b. Frequencies it will record or reproduce.

c. Whether it will be used on a single or multitrack recorder.

A-3.

a. Size of the headgap.

b. Speed of the magnetic tape.

A-4.

a. Narrowband—100 Hz to 100 kHz.

b. Intermediate band—100 Hz to 500 kHz.

c. Wideband—400 Hz to 2 mHz.

A-5. Allow you to monitor the signals you're recording.

A-6. First, before the record and reproduce heads.

A-7.

a. Keep the heads clean.

b. Keep the heads demagnetized.

A-8. With a cotton-tipped applicator soaked in either isopropyl alcohol or a head cleaner recommended by the recorder's manufacturer.

A-9.

a. During ac power losses.

b. During testing.

c. Because of stray magnetic fields.

d. From normal use.

A-10. A hand-held degausser like the manual degaussers used for degaussing magnetic tape.

A-11. Every 8 to 25 hours depending on the manufacturer's recommendations.

CHAPTER 4

MAGNETIC TAPE RECORDER TRANSPORTS

LEARNING OBJECTIVES

After completing this chapter, you'll be able to do the following:

1. Describe the function and components of a basic magnetic tape transport system.
2. Describe the operating characteristics and parts of the three most common tape reeling systems.
3. Describe the physical characteristics of the two basic tape reeling configurations, *co-planar* and *co-axial*.
4. Describe the characteristics of *open-loop drive* and *closed-loop drive* tape transport configurations and the three most common *closed-loop* designs.
5. Describe the capstan speed control function of a tape transport system and the relationship of the six basic parts of a typical capstan speed control unit.
6. Explain why, and describe how, magnetic tape transports must be cleaned and degaussed.

INTRODUCTION

Magnetic tape recorder transports are precisely built assemblies that move the magnetic tape across the magnetic heads and hold and protect the tape. Figure 4-1 shows a basic tape transport assembly. Tape transports have four basic parts:

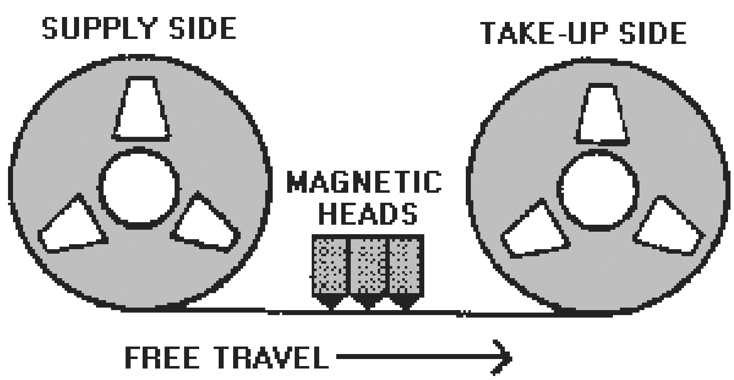


Figure 4-1.—Basic tape transport assembly.

1. A **tape reeling system** that, with the aid of tape guides, physically moves the tape across the magnetic heads.

2. A **tape speed control system** that monitors and controls the movement of the magnetic tape.

3. An **electronic subsystem** that activates the reeling device to move the magnetic tape.

4. A **basic enclosure** that holds and protects the reels or cartridges of magnetic tape.

This chapter describes these basic parts, tells how they work, and shows diagrams of the more common ones.

TAPE REELING SYSTEMS

A basic magnetic recorder tape reeling system (figure 4-1) has one supply reel and one take-up reel. Its job is to move the magnetic tape from one reel to the other. When this happens, four things occur:

1. The supply reel feeds out magnetic tape at a constant tension.

2. The tape passes the magnetic heads in a straight line.

3. The take-up reel accepts the magnetic tape at a constant tension.

4. Both the supply and take-up reels start and stop smoothly while maintaining the proper tape tension.

These four things must happen, or the magnetic tape could be damaged. Three of the most commonly used tape reeling systems are (1) take-up control, (2) two-motor reeling, and (3) tape buffering.

TAKE-UP CONTROL REELING SYSTEMS

This system uses a motorized take-up reel which pulls the magnetic tape off of a *free-spooling* supply reel. It maintains tape tension by using mechanical drag on the supply reel. As you might guess, this method has its disadvantages. It only works in one direction, and the tape tension doesn't remain constant throughout the reel. As the supply reel gives out tape, the tape tension varies. Uneven tape tension can cause stretched tape, poorly wound tape reels, and tape damage during starts and stops.

TWO-MOTOR REELING SYSTEMS

To overcome the problems of take-up control reeling systems, designers added a motor to the supply reel. By using two motors, the magnetic tape direction can be forward or reverse.

Two-motor reeling configurations usually use dc (direct current) motors, instead of ac (alternating current) motors, because dc motors run smoother and are easier to control. To help control tape tension, a small hold-back voltage is added to the motor for the supply reel.

Unfortunately, two-motor reeling systems do not properly control tape tension during starts and stops. Something called *tape buffering* must be added.

TAPE BUFFERING REELING SYSTEMS

Controlling a recorder's tape tension during starts and stops is a big problem. Tape buffering overcomes this problem by regulating the tape reel speed and by protecting against changes in tape tension.

Every manufacturer of high-quality, high performance magnetic tape recorders uses some sort of tape buffering. It's especially important in magnetic recorders that operate at many different speeds, where precise tape tension must be maintained.

Figure 4-2 shows the relationship between the tape reeling system and the tape buffering system. As you can see, the speed at which a tape reel will give up or take up magnetic tape is controlled by its respective speed control servo. Feedback from the supply and take-up buffers tells the servo to speed up or slow down.

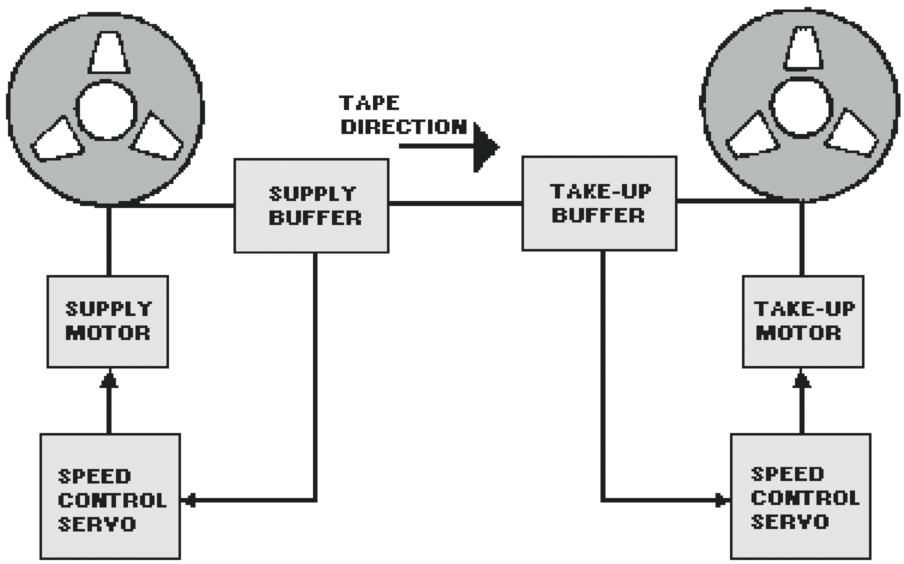


Figure 4-2.—Tape buffering arrangement.

There are two basic types of reeling system buffers: (1) spring-tension, and (2) vacuum-column.

1. **Spring-tension** buffering systems use an electro-mechanical device to sense changes in tape tension. These changes are *feedback* that the speed control servo needs to adjust the speed of the tape reels. Figures 4-3 and 4-4 show two of the more common arrangements for spring-tension buffers.

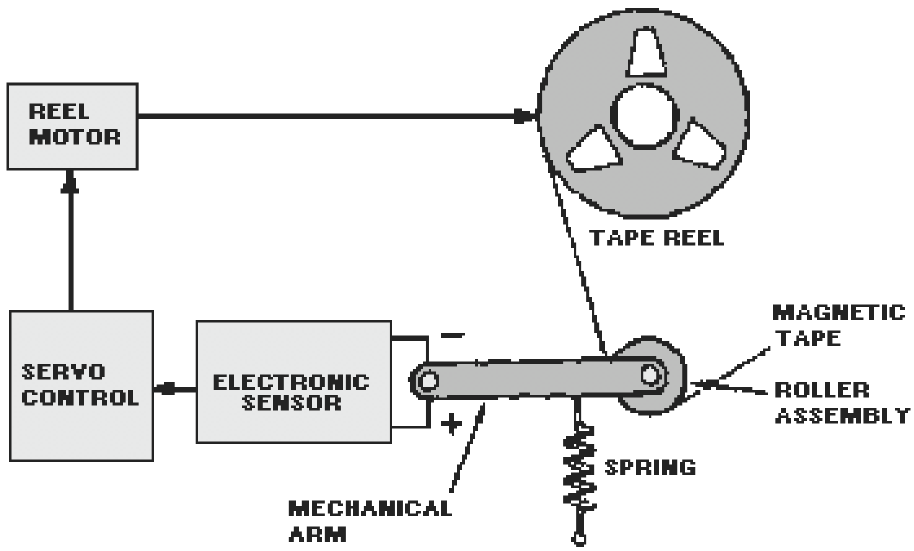


Figure 4-3.—Mechanical arm spring-tension tape buffering.

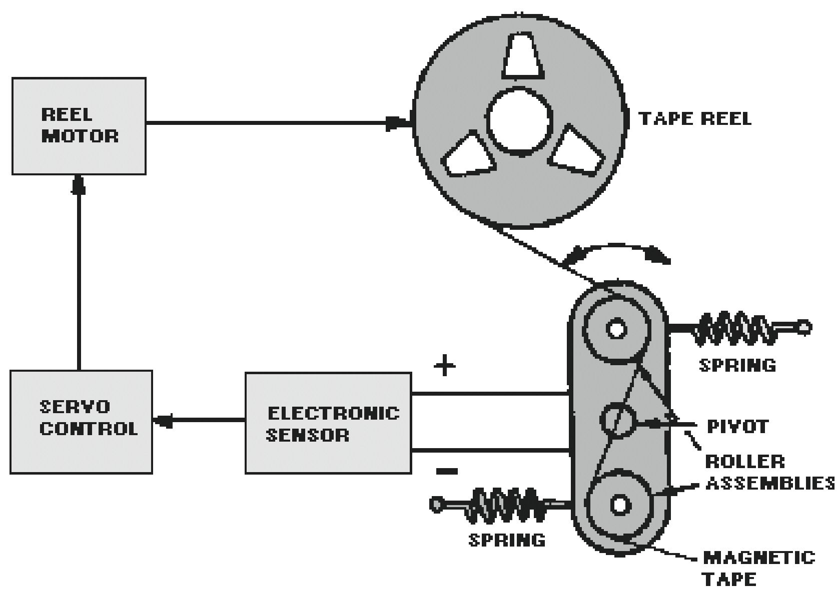


Figure 4-4.—Mechanical arm spring-tension tape buffering.

2. **Vacuum-column** buffering systems operate like the spring-tension systems. They also regulate the speed control servos that control tape reel speed. But, as shown in figure 4-5, the vacuum-column buffer system uses a vacuum chamber instead of a spring to hold a length of magnetic tape as *slack* during tape recorder starts and stops. An electronic sensor in the vacuum chamber helps to control how much tape is in the buffer.

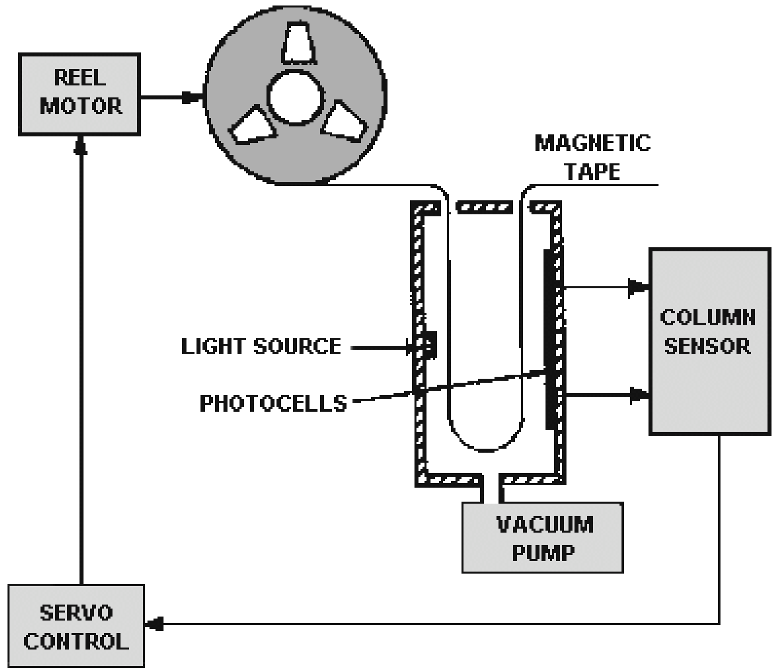


Figure 4-5.—Vacuum-column tape buffering system.

TAPE GUIDES

Another job of a tape reeling system is to make sure the magnetic tape is protected from damage during operation. To do this, tape reeling systems use tape guides. Tape guides come in two designs, *fixed* and *rotary*. Both of these are shown in figure 4-6. Each type of tape guide has its drawbacks. Fixed tape guides produce a lot more friction, and rotary tape guides are more likely to cause errors because of their moving parts.

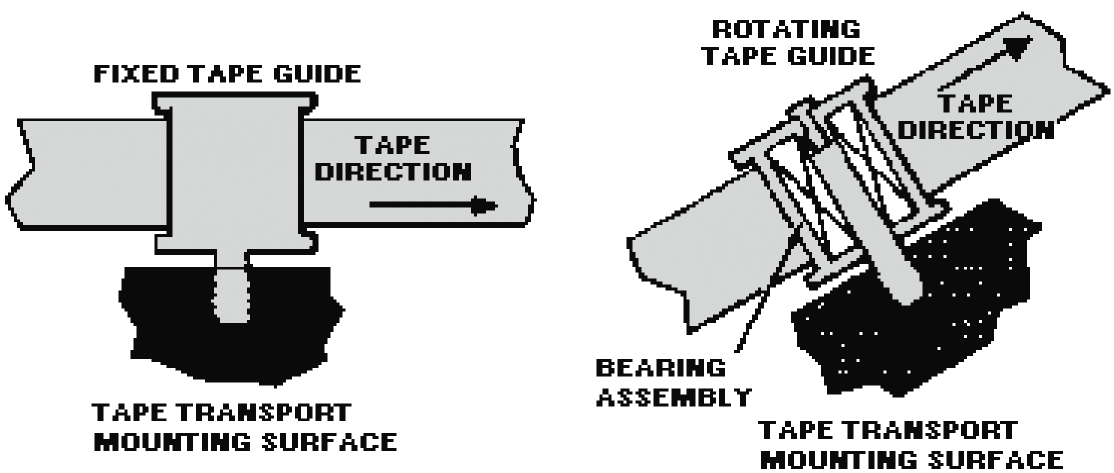


Figure 4-6.—Typical fixed and rotary tape guides.

Tape guides are strategically placed in a tape reeling system to make sure the magnetic tape is kept straight with respect to the supply and take-up reels and the magnetic heads. Some magnetic recorders use only fixed tape guides, some use rotary tape guides, and some use a combination of the two.

TAPE REELING CONFIGURATIONS

There are two basic tape reeling configurations: (1) co-planar, and (2) co-axial. Both of these describe the physical relationship between the supply reel and the take-up reel. The co-planar, which is used more often than the co-axial, has the supply reel and the take-up reel side by side. Figure 4-7 shows this configuration.

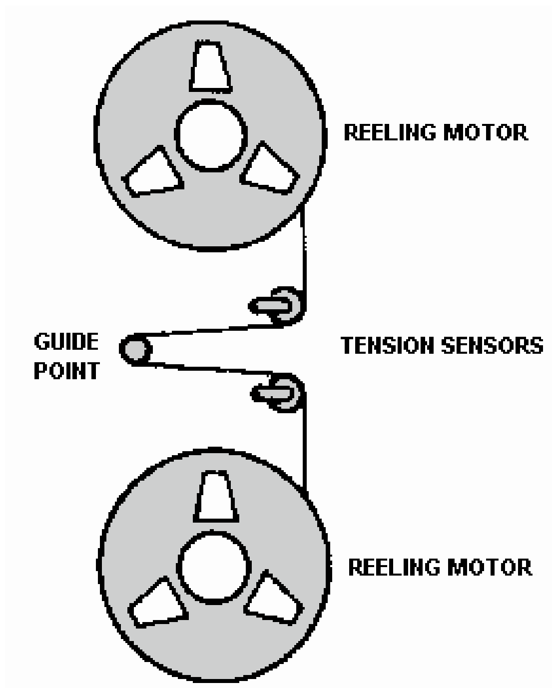


Figure 4-7.—Co-planar tape reeling configuration.

The co-axial configuration is used when physical space is limited. It places the supply and take-up reels on top of each other. Figure 4-8 shows this configuration.

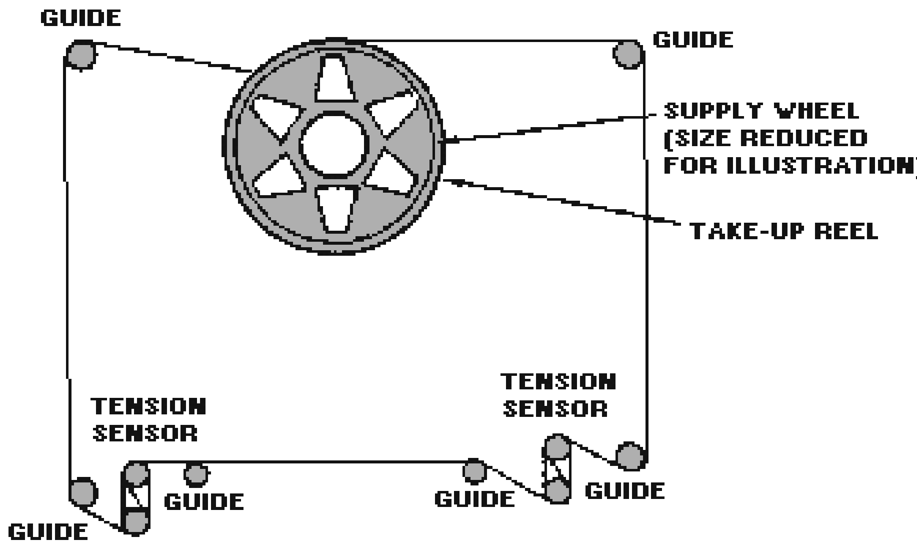


Figure 4-8.—Co-axial tape reeling configuration.

Q-1. What are the four basic parts of a magnetic tape recorder's tape transport system?

Q-2. What are the three most commonly used tape reeling systems?

Q-3. What are two disadvantages of the take-up control reeling system?

Q-4. What are two advantages of a two-motor reeling system over a take-up control reeling system?

Q-5. What type of reeling system best controls a tape recorder's tape tension during starts and stops?

Q-6. What are the two basic types of tape buffering reeling systems?

Q-7. How do the tape guides on a tape reeling system protect the tape from damage during operation?

TAPE TRANSPORT CONFIGURATIONS

There are two types of tape transport configurations: (1) *open-loop capstan drive*, and (2) *closed-loop capstan drive*. The following paragraphs describe each of these.

OPEN-LOOP CAPSTAN DRIVE

This is probably the simplest tape transport configuration. Figure 4-9 shows how the magnetic tape is pulled off of the supply reel, taken across the magnetic heads, and wound onto the take-up reel. The tape is *pulled* by sandwiching it between a single capstan and a pinch roller. As the capstan turns, the friction between it and the pinch roller pulls the tape across the magnetic heads. The magnetic tape is held against the magnetic heads by using tape guides.

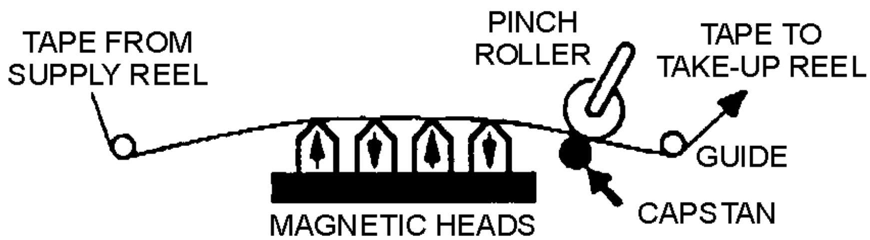


Figure 4-9.—Open-loop capstan drive tape transport.

The open-loop drive transport configuration has two major drawbacks:

1. It can only work in one direction. It can *pull* the tape, but it can't *push* it across the magnetic heads.
2. Tape tension and head-to-tape contact can vary. If the capstan motor hesitates or speeds up, the tape tension will vary.

CLOSED-LOOP CAPSTAN DRIVE

Closed-loop capstan drive tape transports were designed to overcome the drawbacks of the open-loop drive design. They use more than one capstan and/or pinch roller to *clamp* the magnetic tape in the area around the magnetic heads. This keeps tape tension constant and improves the quality of the recording or the playback. Figure 4-10 shows the basic arrangement of the closed-loop capstan drive.

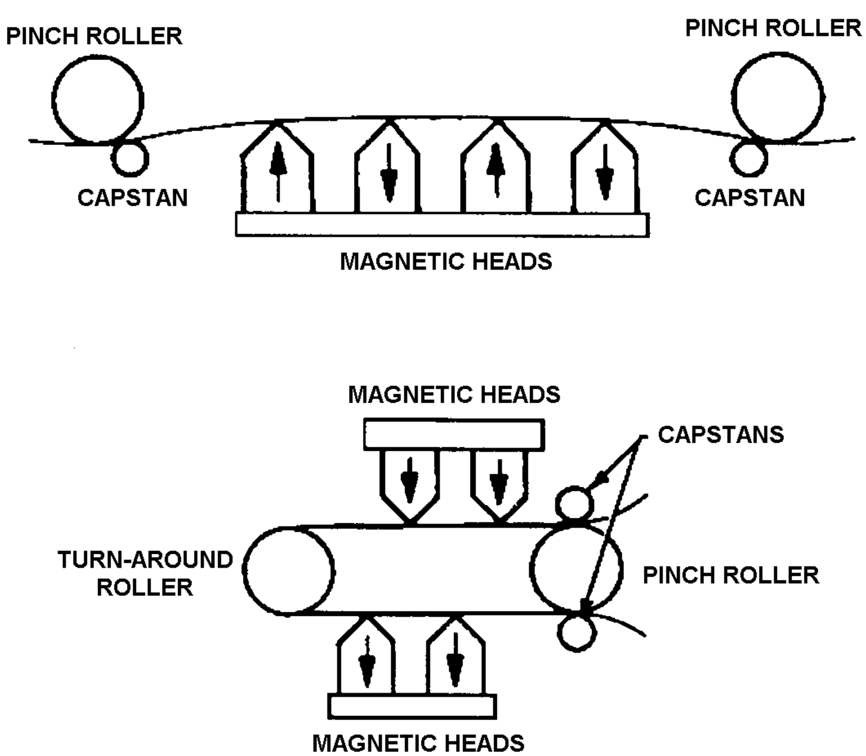


Figure 4-10.—Closed loop capstan drive tape transport.

The three most common closed-loop capstan drive designs are (1) *differential velocity capstans*, (2) *dual-motors dual capstans*, and (3) *peripheral drive capstans*.

Differential Velocity Capstans

Figure 4-11 shows a differential velocity capstan. In this design, the take-up capstan is made a little larger than the supply capstan. This causes the take-up capstan to pull the tape away from the heads slightly faster than the supply capstan feeds the tape to the heads. The result is a constant tape tension in the area around the magnetic heads.

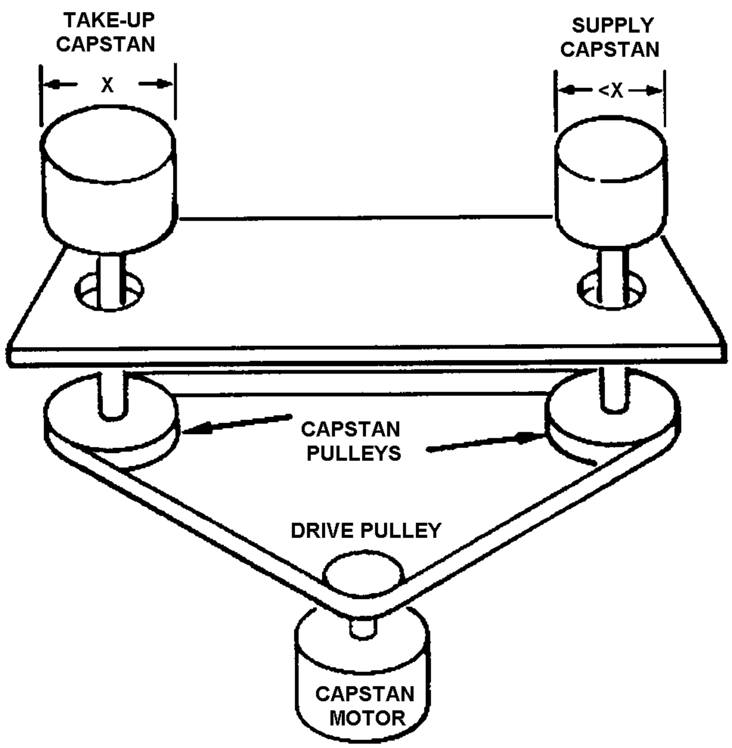


Figure 4-11.—Differential velocity capstan drive.

Both capstans are turned by a single motor which is coupled to the capstan pulleys by a belt. This arrangement is very efficient in one direction, but, unfortunately, differential velocity capstans don't work in reverse. If you reversed the tape direction, a negative tension would occur, and the tape would bunch up in the area around the magnetic heads.

Dual-Motors Dual Capstans

Figure 4-12 shows a dual-motor dual capstan drive. In this design, each capstan is driven by its own motor. Tape tension is maintained by slowing down one of the motors. When reverse tape motion is needed, the opposite motor is slowed down.

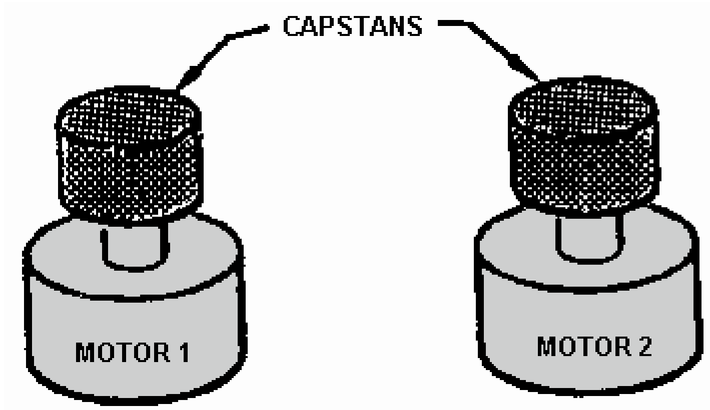


Figure 4-12.—Dual-motors dual capstans drive system.

Peripheral Drive Capstans

In this design, the magnetic tape is moved by a capstan placed directly against the tape reel or tape pack. Figure 4-13 shows two different peripheral drive capstan arrangements.

The first arrangement, figure 4-13A, shows a single capstan design. In this method, two tightly wound tape reels, without flanges, are pushed against the capstan. As the capstan turns, it forces the tape reels to turn in the appropriate direction. Magnetic tape tension is maintained by using either spring loading or servo control.

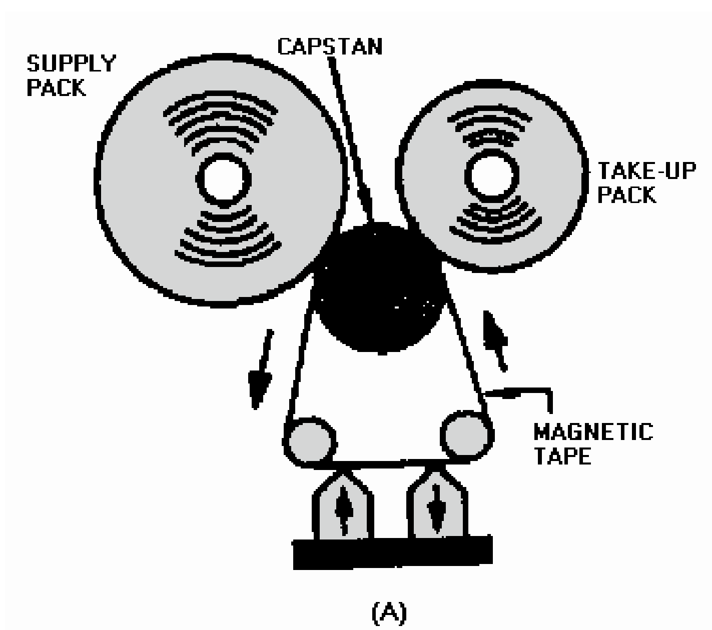


Figure 4-13A.—Peripheral drive capstans.

The second arrangement, figure 4-13B, uses two capstans. In this method, the two tightly wound tape reels, without flanges, are pressed directly against the capstans. Tension in the magnetic head area is maintained by controlling the speed of the individual capstans.

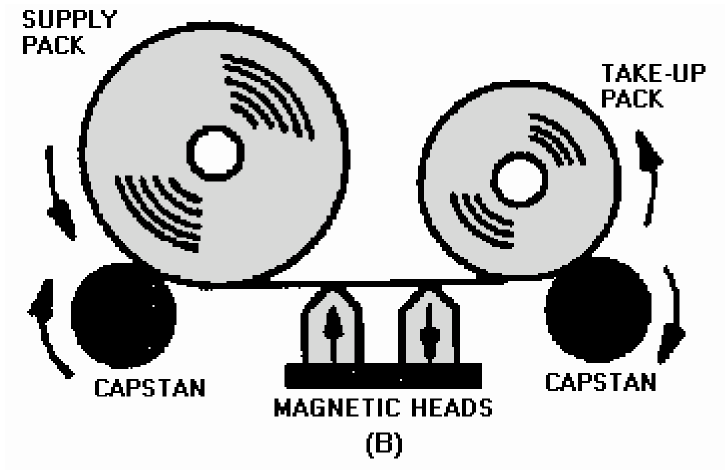


Figure 4-13B.—Peripheral drive capstans.

CAPSTAN SPEED CONTROL

Capstan speed control is an important part of the magnetic tape transport system. It makes sure the capstan is turning (1) at the right speed and (2) at a constant speed. This is important because errors in speed control can cause poor recordings and playbacks.

Capstans are turned either by a motor only, or by a motor, belt, and pulley arrangement. In either case, it's the motor that the capstan speed control function acts upon to do its job. A capstan speed control function typically consists of six basic parts. Figure 4-14 shows these six parts and how they're related. Each of the parts is described below.

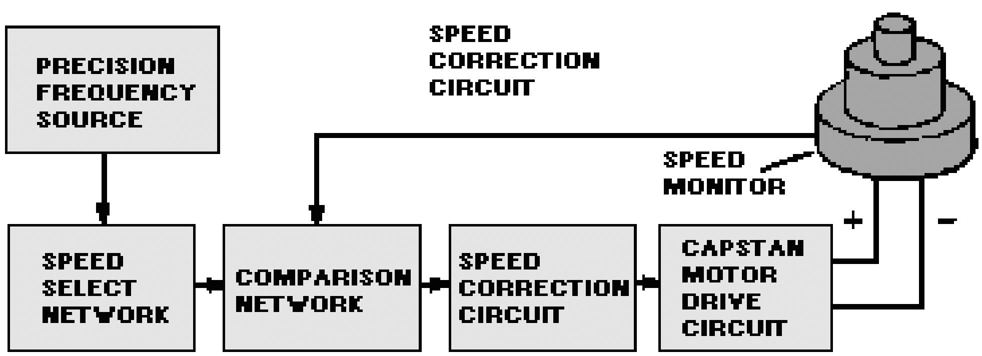


Figure 4-14.—Six parts of the capstan speed control function.

PRECISION FREQUENCY SOURCE

This part of the capstan speed control provides a reference frequency that the *speed select network* and the *comparison network* use to drive the capstan motor. The precision frequency source is usually a very-high-frequency crystal with an accuracy of at least .001 percent.

SPEED SELECT NETWORK

This network selects the desired operating tape speed. It takes the reference frequency from the *precision frequency source* and (depending on the desired operating tape speed) generates another specific reference signal that the *comparison network* uses to control the speed of the capstan. Table 4-1 is a list of the speed control reference signal frequencies for the various operating tape speeds.

Table 4-1.—Typical speed control reference signal frequencies

Operating Tape Speed (inches per second)	Speed Control Frequency (kilohertz)
15/16	1.5625
1 7/8	3.125
3 3/4	6.25
7 1/2	12.5
15	25
30	50
60	100
120	200
240	400

CAPSTAN SPEED MONITOR

This circuit monitors the true capstan motor speed. It sends the true speed to the *comparison network* circuit. Most capstan speed monitor circuits are made using a photo-optical tachometer that's directly attached to the shaft of the capstan motor. Figure 4-15 shows this.

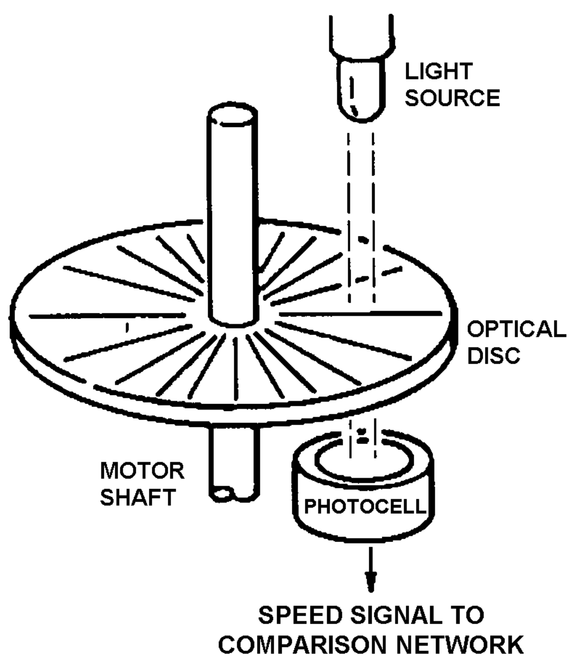


Figure 4-15.—Capstan speed monitor using a photo-optical tachometer.

COMPARISON NETWORK

This network takes the input signals from the *speed select network* and the *capstan speed monitor*, compares the two signals, and decides if the capstan is at the right speed. If not, it tells the speed *correction circuit*.

Sometimes, a third input signal, which comes from the magnetic tape itself, is supplied to the *comparison network.* It's called a *servo control from tape signal*. Tape recordings made on a specific recorder are sometimes shipped off for further analysis and played back on a different recorder. To help compensate for speed errors in the tape transport systems of the two recorders, the *precision reference frequency* of the originating recorder is recorded onto a track of the magnetic tape. During playback, this reference signal is also fed to the recorder's *comparison network* and is used to correct speed errors.

SPEED CORRECTION CIRCUIT

This circuit takes speed correction signals from the *comparison network* and tells the *capstan motor drive circuit* to either speed up or slow down the capstan motor.

CAPSTAN MOTOR DRIVE CIRCUIT

This circuit takes the speed-up or slow-down signals from the speed correction circuit and actually speeds up or slows down the capstan motor.

MAGNETIC TAPE TRANSPORT MAINTENANCE

If you want good recordings and playbacks, you must keep magnetic tape transports clean and demagnetized. The following paragraphs describe preventive maintenance procedures for magnetic tape transport systems.

MAGNETIC TAPE TRANSPORT CLEANING

You can clean most magnetic tape transports with isopropyl alcohol, cotton swabs, and lint-free cloths. (Caution: Cotton swabs <u>are not</u> lint free, so use them only in places you can't get to with the lint-free cloths.) Figure 4-16 shows a technician cleaning a capstan. Here are some other points to remember:

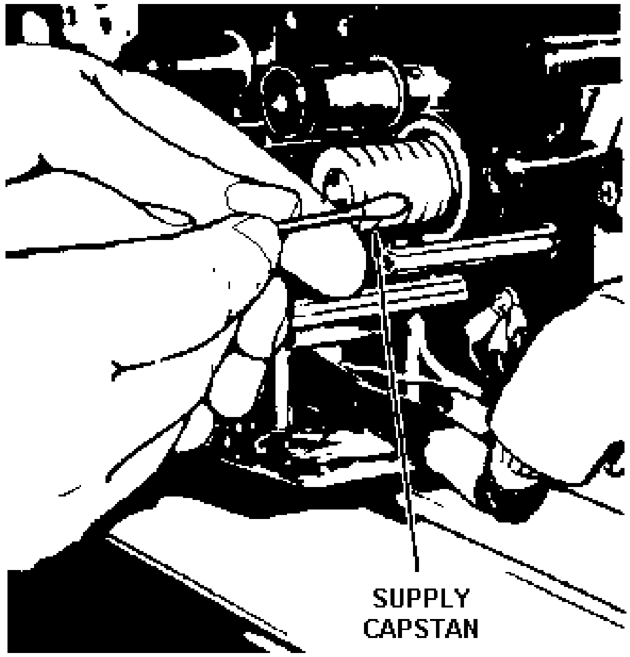


Figure 4-16.—Cleaning the capstan on a magnetic tape transport system.

- DO always remove the magnetic tape from the transport before cleaning it.
- DO apply the cleaner onto the lint free cloth or cotton swab; DON'T apply it directly onto the tape transport.
- DO pay extra attention to the flanged parts of tape guides. It's here that oxide particles collect the most.
- DON'T use the same lint-free cloth or cotton swab to clean many parts of the tape transport. Switch cloths and swabs often. If you don't, you may transfer dirt and oxide particles from one part of the tape transport to another.

MAGNETIC TAPE TRANSPORT DEMAGNETIZING

With use, tape transport parts become magnetized. It's hard to say *exactly* what will happen if the magnetic tape passes a magnetized part of the tape transport before the tape is recorded on. The effects can range from just a little more noise on the tape to a complete tape saturation. Either way, magnetized tape transport parts can ruin magnetic recordings.

To prevent this, you must periodically demagnetize the tape transport. The procedures for doing this are identical to those listed in chapter 2 for demagnetizing magnetic heads. You'll even use the same manual hand-held degausser you saw in figure 2-8 of chapter 2. Figure 4-17 shows a technician demagnetizing a tape guide.

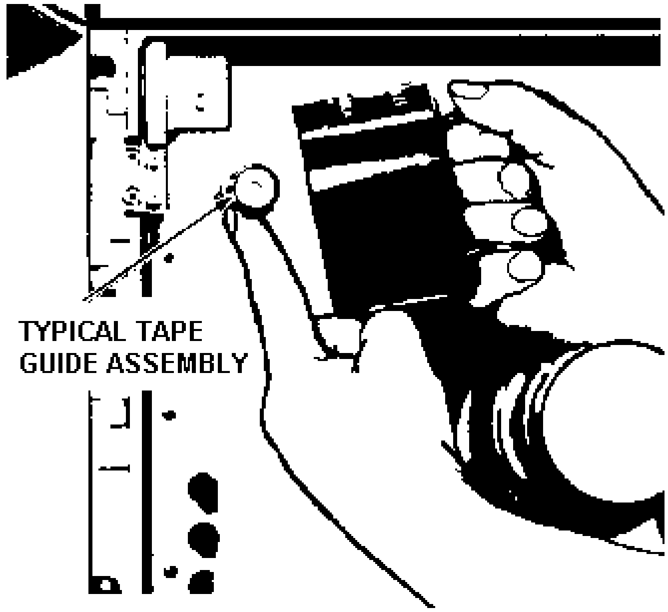


Figure 4-17.—Demagnetizing a tape guide with a hand-held degausser.

Q-8. There are two types of tape transport configurations, open-loop capstan drive and closed-loop capstan drive. What are two major disadvantages of open-loop capstan drive tape transports?

Q-9. How do closed-loop capstan drive tape transports overcome the disadvantages of the open-loop drive design?

Q-10. What are the three most common closed-loop capstan drive designs?

Q-11. How do tape transports with differential velocity capstans maintain a constant tape tension in the area around the magnetic heads?

Q-12. How do dual-motor dual capstan drives maintain a constant tape tension while operating in either a forward or reverse direction?

Q-13. What are the two critical functions of the capstan speed control part of a magnetic tape transport system?

Q-14. Which part of the capstan speed control function monitors the true capstan motor speed?

Q-15. Sometimes it's necessary, but why should you avoid using cotton swabs when cleaning a magnetic tape transport?

Q-16. When cleaning the parts of a tape transport, why should you switch lint-free cloths and swabs often?

Q-17. What equipment should you use to de-magnetize a magnetic tape transport?

SUMMARY

Now that you've finished chapter 4, you should be able to describe magnetic tape transport systems in terms of their operating characteristics, parts, and preventive maintenance requirements. The following is a summary of the important points in this chapter.

A **MAGNETIC TAPE RECORDER TRANSPORT** has four basic parts: (1) tape reeling system, (2) tape speed control system, (3) electronic subsystem, and (4) basic enclosure.

The **TAPE REELING SYSTEM** must move the tape in a straight line at a constant tension, and it must start and stop smoothly while maintaining the proper tension.

Three of the **MOST COMMON REELING SYSTEMS** are (1) take-up control, (2) two-motor reeling, and (3) tape buffering.

The two types of **TAPE TRANSPORT CONFIGURATIONS** are (1) *open-loop capstan drive* and (2) *closed-loop capstan drive*. The open-loop type works in only one direction, and the tape tension can vary. The closed-loop type keeps the tape tension constant.

Three types of **CLOSED-LOOP CAPSTAN DRIVES** are (1) differential velocity capstans, (2) dual-motors dual capstans, and (3) peripheral drive capstans.

The **CAPSTAN SPEED CONTROL** component of a tape transport keeps the capstan turning at the correct operating speed and at a constant speed. It has these six parts: (1) precision frequency source, (2) speed select network, (3) capstan speed motor, (4) comparison network, (5) speed correction circuit, and (6) capstan motor drive circuit.

You should **CLEAN** magnetic tape transports with isopropyl alcohol, cotton swabs, and lint free cloths and **DEMAGNETIZE** them using a hand-held degausser.

ANSWERS TO QUESTIONS Q1. THROUGH Q17.

A1.

- *a. Tape reeling system.*
- *b. Tape speed control system.*
- *c. Electronic subsystem.*
- *d. Basic enclosure.*

A2.

- *a. Take-up control.*
- *b. Two-motor reeling.*
- *c. Tape buffering.*

A3.

a. It only works in one direction.

b. The tape tension varies as the supply reel unwinds, which can cause damage during starts and stops.

A4. The two-motor configuration runs in both directions and a holdback voltage helps control tape tension, but it does not properly control tape tension during starts and stops.

A5. A tape buffering reeling system.

A6.

a. Spring-tension ***buffering systems****.*

b. Vacuum-column ***buffering systems****.*

A7. They keep the tape straight with respect to both the supply and take-up reels and the magnetic heads.

A8.

a. Only operates in one direction.

b. The tape tension and head-to-tape contact can vary.

A9. Closed loop capstan drive transports use more than one capstan to clamp the tape in the area around the magnetic head.

A10.

a. Differential velocity capstans.

b. Dual motors dual capstans.

c. Peripheral drive capstans.

A11. The supply capstan is slightly larger than the take-up capstan. This causes the take-up capstan to pull the tape slightly faster than the supply capstan feeds the tape.

A12. Each capstan is driven by its own motor. It maintains tape tension by slowing down one of the motors. When the tape motion is reversed, the opposite motor is slowed down.

A13. Makes sure the capstan turns at the right speed and at a constant speed.

A14. Capstan speed monitor.

A15. Cotton swabs are not lint free.

A16. You may transfer dirt or oxide particles from one part of the tape transport to another.

A17. A hand-held degausser.

CHAPTER 5

MAGNETIC TAPE RECORDER RECORD AND REPRODUCE ELECTRONICS

LEARNING OBJECTIVES

After completing this chapter, you'll be able to do the following:

1. State the two types of record and reproduce electronics used on magnetic tape recorders.
2. Describe the purpose and function of *direct record* electronics and the four main parts of a recorder's direct record component.
3. Describe the purpose and function of *direct reproduce* electronics and the three main parts of a recorder's direct reproduce component.
4. Describe the purpose and function of *frequency modulation* (FM) record electronics and the three main parts of a recorder's FM record component.
5. Describe the purpose and function of FM reproduce electronics and the four main parts of a recorder's FM record component.

RECORD AND REPRODUCE ELECTRONICS

There are two ways to record and reproduce analog signals. The first way is *direct record*. It's also called *amplitude modulation* (AM) electronics. The second way is *frequency modulation* (FM). Even though direct record and reproduce circuits are much different from FM record and reproduce electronics, they both share the same two very important jobs. They both must:

1. Take an input signal, process it as needed, and then send it to the record magnetic head for reproduction.
2. Take the reproduced signal from the reproduce magnetic head, process it as needed, and output it for listening or analysis.

DIRECT RECORD ELECTRONICS

Direct record electronics record input signals onto magnetic media just as the signals appeared at the recorder's input. The only processing an input signal receives is the adding of a bias signal. The added bias signal makes sure the signal stays away from the *steps* of the magnetism curve. Figure 5-1 shows a basic block diagram of a recorder's direct record electronics.

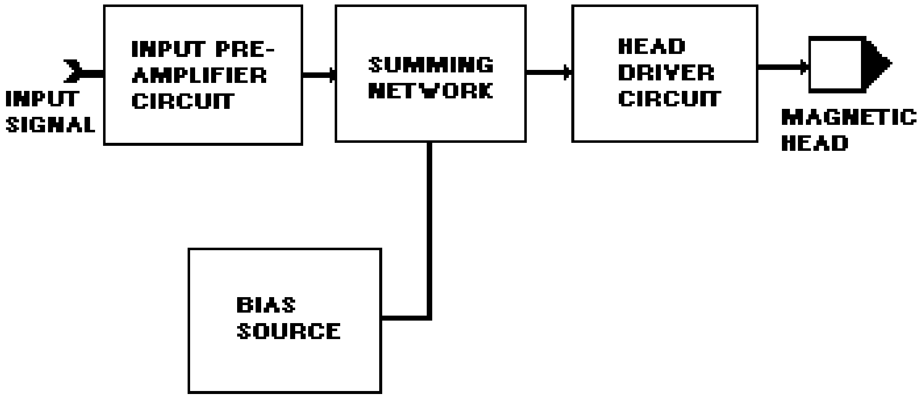


Figure 5-1.—Direct record electronics.

Direct record electronics has four main parts:

1. Input pre-amplifier circuit. This circuit takes the input signal, amplifies it, and sends it to the summing network. It also matches the impedance between the source of the input signal and the magnetic tape recorder.

2. Bias source. This circuit generates the high-frequency bias signal and sends it to the summing network. Normally, the frequency of the bias signal will be five to ten times higher than the highest frequency the tape recorder can record.

3. Summing network. This network takes the input signal and the bias signal, mixes them, and sends the resulting signal to the head driver circuit.

4. Head driver circuit. This circuit takes the signal from the summing network, amplifies it, and sends it to the record head for recording.

DIRECT REPRODUCE ELECTRONICS

Direct reproduce electronics amplify the *very* weak input signals from the reproduce head, and send them out for listening or analysis, as needed. Figure 5-2 shows a basic block diagram of direct reproduce electronics.

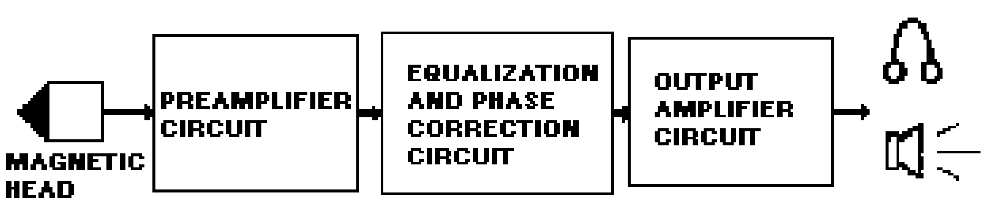


Figure 5-2.—Direct reproduce electronics.

Direct reproduce electronics consists of three main parts:

1. Pre-amplifier circuit. This circuit takes the very weak reproduced signal from the reproduce head and (a) amplifies the signal, (b) removes any bias signal that was used during the recording process, and (c) sends the signal to the equalization and phase correction circuit.

2. Equalization and phase correction circuit. This circuit takes the pre-amplified signal and fixes any frequency response problems that the reproduce magnetic head may have caused. To better understand this, look at the voltage *versus* frequency response graph in figure 5-3. The top of the graph shows the input signal that comes from the pre-amplifier and the bottom shows the *equalization* signal generated by the equalization circuit. In the top part of the graph, note how the output voltage level varies as the frequency of the signal varies. This isn't good. A good output voltage level is one that remains constant as the frequency changes. The equalization signal corrects this problem. Notice that when the input signal and the equalization signal are combined they cancel each other out. This allows a nice flat (voltage *versus* frequency) output signal to go to the output amplifier circuit.

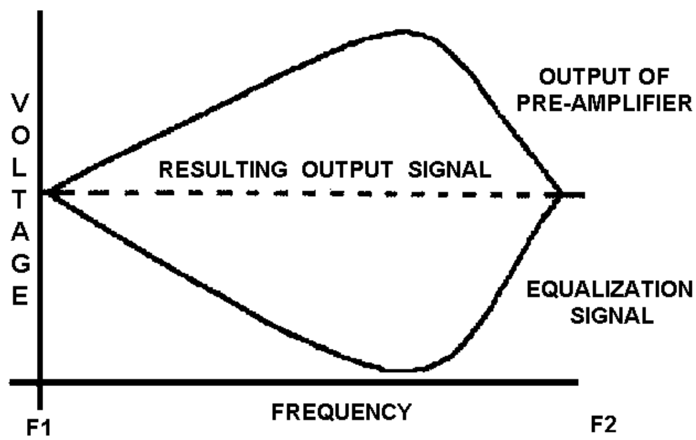


Figure 5-3.—Equalization process.

3. Output amplifier circuit. This circuit takes the signal from the equalization and phase correction circuit and amplifies it for output. It also matches the magnetic recorder's impedance to the output device that is used for listening or recording.

FM RECORD ELECTRONICS

FM record electronics process signals to be recorded differently than direct record electronics. Instead of recording the input signal just as it appears at the recorder's input, FM record electronics use the input signal to vary (modulate) the carrier frequency of a *record oscillator*. The frequency modulated output signal of the *record oscillator* then becomes the signal that's actually recorded onto the magnetic media. Figure 5-4 shows a block diagram of the FM record electronics.

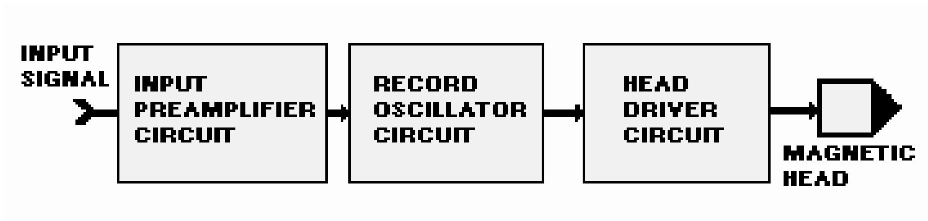


Figure 5-4.—FM record electronics.

FM record electronics consist of three main parts:

1. Input pre-amplifier circuit. This circuit does two things: (a) it serves as an impedance matcher between the signal source and the magnetic recorder, and (b) it pre-amplifies the input signal.

2. Record oscillator circuit. This circuit generates a carrier signal onto which the input signal will be modulated. The input signal is used to vary (frequency modulate) the carrier signal. This is how the input signal gets frequency modulated onto the carrier signal. The output of this circuit is the frequency-modulated carrier signal. The center frequency of the carrier depends on two things: (a) the bandwidth of the signal you're recording, and (b) the media onto which you're recording. For magnetic tape, the carrier frequency can be as low as 1.688 kHz for an operating tape speed of 1-7/8 inches per second, and as high as 900 kHz for 120 inches per second.

3. Head driver circuit. This circuit takes the frequency-modulated output from the record oscillator circuit, amplifies it, and sends it to the magnetic head for recording. The output level of this circuit is set to be *just below* the magnetic saturation point of the magnetic media.

FM REPRODUCE ELECTRONICS

The FM reproduce electronics work just like direct reproduce electronics, with one exception. FM reproduce electronics must first demodulate the original input signal from the carrier frequency before the intelligence can be sent to the output device for listening or analysis. Figure 5-5 shows a block diagram of the FM reproduce electronics.

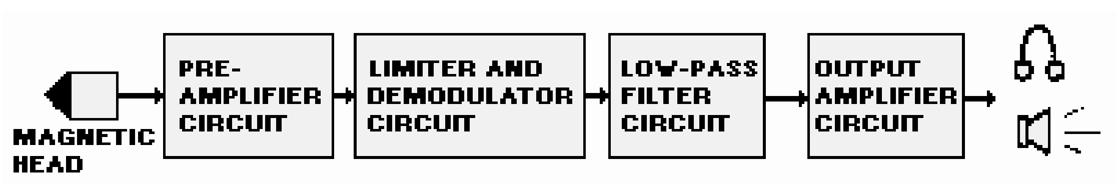


Figure 5-5.—FM reproduce electronics.

FM reproduce electronics consist of four main parts:

1. Pre-amplifier circuit. This circuit takes the frequency modulated carrier frequency from the reproduce head and amplifies it.

2. Limiter/demodulator circuit. This circuit takes the output of the preamplifier, stabilizes the amplitude level, and demodulates the signal intelligence from the carrier frequency.

3. Low-pass filter circuit. This circuit takes the signal intelligence from the limiter/demodulator circuit and cleans up any noise or left over carrier signal.

4. Output amplifier circuit. This circuit takes the output from the low-pass filter and amplifies it for output. It also matches the impedance of the magnetic recorder to the output device.

Q-1. What two types of record and reproduce electronics are used by magnetic tape recorders?

Q-2. The head driver circuit in a tape recorder's direct record electronics component (figure 5-1) performs what function?

Q-3. The equalization and phase correction circuit in a tape recorder's direct reproduce electronics (figure 5-2) performs what function?

Q-4. How do FM record electronics differ from AM (direct record) electronics?

Q-5. The head driver circuit of a tape recorder's FM record electronics (figure 5-4) performs what function?

Q-6. What is the major difference between direct reproduce electronics and FM reproduce electronics?

SUMMARY

Now that you've finished chapter 5, you should be able to (1) state the two types of record and reproduce electronics used on magnetic tape recorders and (2) describe the function and main parts of *direct record and reproduce electronics* and *FM record and reproduce electronics*. The following is a summary of important points in this chapter:

DIRECT RECORD (AM) and **FREQUENCY MODULATION (FM)** are the two types of record and reproduce electronics used by magnetic tape recorders.

The four main parts of **DIRECT RECORD ELECTRONICS** are the (1) input pre-amplifier circuit, (2) bias source, (3) summing network, and (4) head driver circuit.

The three main parts of **DIRECT REPRODUCE ELECTRONICS** are the (1) pre-amplifier circuit, (2) equalization and phase correction circuit, and (3) output amplifier circuit.

FM RECORD ELECTRONICS record a frequency modulated signal onto the magnetic tape. It has three main parts: (1) input pre-amplifier circuit, (2) record oscillator circuit, and (3) head driver circuit.

FM REPRODUCE ELECTRONICS must demodulate the original input signal from the carrier signal. It has four main parts: (1) preamplifier circuit, (2) limiter and demodulator circuit, (3) low-pass filter circuit, and (4) output amplifier circuit.

ANSWERS TO QUESTIONS Q1. THROUGH Q6.

A1.

a. Direct record (AM).

b. Frequency modulation (FM).

A2. It takes the signal from the summing network, amplifies it, and sends it to the record head for recording.

A3. It generates an equalization signal which corrects any frequency response problems in the input signal from the pre-amplifier circuit.

A4. Instead of recording the signal just as it appears at the recorder's input, FM record electronics records a frequency-modulated carrier signal from a record oscillator (figure 5-4) onto the magnetic tape.

A5. It amplifies the frequency-modulated output from the record oscillator and sends it to the record head.

A6. FM record electronics must use a limiter and demodulator circuit (figure 5-5) to demodulate the signal intelligence from the carrier frequency.

CHAPTER 6

MAGNETIC TAPE RECORDING SPECIFICATIONS

LEARNING OBJECTIVES

After completing this chapter, you'll be able to do the following:

1. Define the seven most common magnetic tape recording specifications.
2. Describe a magnetic tape recorder's *signal-to-noise ratio* (SNR) *specification*, how it's measured, and why a high SNR is important.
3. Describe a tape recorder/reproducer's *frequency-response* specification, how it's measured, and the three factors that can limit or degrade a recorder's frequency response.
4. Describe a tape recorder's *harmonic-distortion* specification, how it's measured, and how a recorder produces harmonic distortion.
5. Describe a recorder's *phase-response* specification, how it's measured, and why good phase response is important.
6. Describe a recorder's *flutter* specification, how it's measured, and why minimal flutter is important.
7. Describe a recorder's *time-base error* (TBE) specification, how it's measured, and why minimal TBE is important.
8. Describe a multi-track magnetic tape recorder's *skew* specification, how it's measured, and why minimal skew is important.

INTRODUCTION

Have you ever gone to a store to buy a magnetic tape recorder? Were you able to decide which of the displayed models was the *good* one to buy? Or, did you instead end up confused when the salesperson started spouting words like *SNR*, *flutter*, and *bandwidth.* If so, you weren't alone.

This chapter (1) defines the seven most common magnetic tape recording specifications, (2) describes their effect on the magnetic recording process, and (3) tells how to measure each specification. The remaining paragraphs in this chapter describe each of the following magnetic tape recorder specifications:

1. Signal-to-noise ratio
2. Frequency response
3. Harmonic distortion
4. Phase response

5. Flutter
6. Time-base error
7. Skew

SIGNAL-TO-NOISE RATIO

Signal-to-noise ratio (SNR) is the first magnetic tape recorder specification we'll describe. It's one of the most important specifications of a magnetic tape recorder.

SIGNAL-TO-NOISE RATIO DEFINITION

The SNR is *the ratio of the normal signal level to the magnetic tape recorder's own noise level.* It's measured in decibels (dB). In other words, the higher the SNR of a magnetic tape recorder, the wider the range of input signals it can properly record and reproduce.

The *noise* part of the signal-to-noise ratio is generated in the magnetic tape recorder itself. Although noise can be generated by almost any part of the magnetic tape recorder, it's usually generated by either the magnetic heads or the magnetic tape.

SIGNAL-TO-NOISE RATIO MEASUREMENT

You can measure the SNR with a vacuum tube voltmeter (VTVM) and a signal generator. The equipment set up for measuring the SNR is shown in figure 6-1. After equipment setup, measure the SNR as follows:

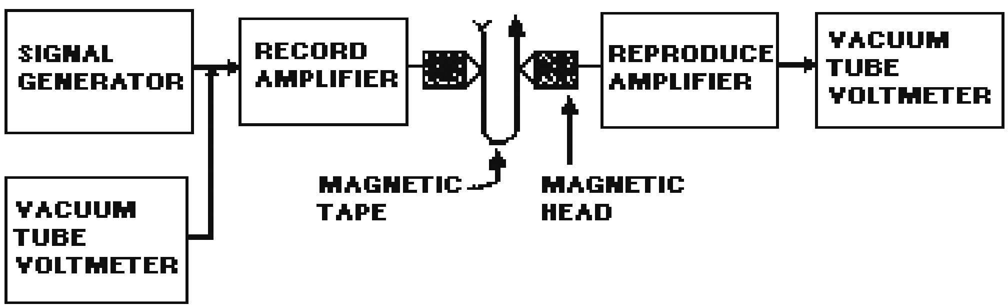


Figure 6-1.—Test equipment setup for measuring signal-to-noise ratio.

1. Set the signal generator to inject a test signal into the tape recorder. The technical manual for the tape recorder you're testing will tell you how to set up the signal generator.

2. While recording and reproducing, set the output level of the tape recorder's reproduce electronics to a level that displays 0-dB reference on the VTVM.

3. Disconnect the signal generator. The voltage displayed on the VTVM will drop from 0-dB to some *negative* dB level. This level is the magnetic tape recorder's SNR.

There are two things you should know when reading SNR specifications in technical manuals, equipment brochures, etc.

First, the SNR is stated in three ways. You'll see it as (1) *root-mean-square (RMS) signal-to-RMS noise*, (2) *peak-to-peak signal-to-RMS noise*, or (3) *peak signal-to-RMS noise*. If the SNR specification doesn't state which way it was measured, you could be mislead. For example, a 25-dB RMS SNR is equal to a 34-dB peak-to-peak signal-to-RMS noise ratio, or a 28-dB peak signal-to-RMS noise ratio.

Second, all SNR specifications should include the record level that was used. Since the SNR varies directly to the record level, you could be mislead by a SNR that doesn't include the record level of the test signal used when the SNR was measured.

FREQUENCY RESPONSE

The frequency-response specification of a magnetic tape recorder is sometimes called the *bandwidth*. A typical frequency-response specification might read *within* + / – 3 *db from 100 Hz to 100 kHz at 60 ips*. This means the magnetic tape recorder is capable of recording all frequencies between 100 Hz and 100 kHz at 60 inches per second (ips) without varying the output amplitude more than 3 dB.

FREQUENCY-RESPONSE DEFINITION

Frequency response is *the amplitude variation with frequency over a specified bandwidth*. Let's convert this to plain English. The frequency-response specification of a magnetic tape recorder tells you the range of frequencies the recorder can *effectively* record and reproduce. What exactly does the word *effectively* mean? That's hard to say because frequency response varies from recorder to recorder, and from manufacturer to manufacturer. But a good rule of thumb is that *an effective frequency-response specification tells the lowest and highest frequencies that the recorder can record and reproduce with no more than + / – 3-dB difference in output amplitude*.

FREQUENCY-RESPONSE MEASUREMENT

The equipment setup for measuring the frequency response of a magnetic tape recorder is the same as for measuring the signal-to-noise ratio. It's shown in figure 6-1. After equipment setup, measure a recorder's frequency response as follows:

1. Set the signal generator to output a test signal. The technical manual for the tape recorder will tell you how.

2. Set the recorder's reproduce electronics output level to a 0-dB reference on the VTVM.

3. While recording at a set speed, vary the frequency of the signal generator from the lowest to highest frequency you're checking. Make sure that the output level of the signal generator stays the same.

4. As you sweep through the frequencies, look at the VTVM. You'll see the amplitude rise and fall as you vary the output frequency of the signal generator. As you approach the lowest and the highest frequencies that the magnetic tape recorder can *effectively* record, you'll see the VTVM drop to less than – 3 dB. This determines the lower and upper limits of the frequency-response specification for that magnetic tape recorder.

FREQUENCY-RESPONSE LIMITING FACTORS

Four factors that can limit or degrade the frequency response of magnetic tape recorders are:

1. A too-high or too-low bias signal level setting for the record head.
2. An improper reproduce head.
3. An improper tape transport speed.
4. A poor magnetic tape-to-head contact.

The magnetic record head transforms the electrical signal into a magnetic field for recording onto magnetic tape. If the bias signal level is set to high, you might erase the higher frequencies. If it's too low, you'll get excessive tape distortion.

The reproduce head transforms the magnetic field from the magnetic tape back into an electrical signal. As explained in chapters 3 and 5, the head gap of a recorder's reproduce head and the operating speed of the magnetic tape transport determine the wavelength of the reproduce head. The wavelength determines the *center* frequency of a recorder's frequency-response specification. Once you pass this center frequency, both below and above, the output voltage level of the recorder's reproduce head will decrease. Figure 6-2 shows this. This is why the equalization circuits described in chapter 5, figure 5-3, are used.

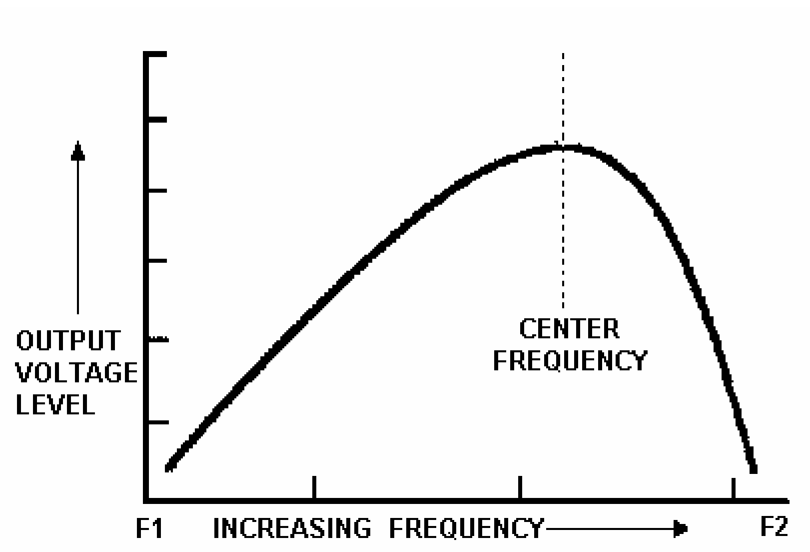


Figure 6-2.—Frequency response of a reproduce head.

Poor tape-to-head contact can seriously degrade the record and reproduce process. Magnetic heads are designed to reduce tape-to-head gap as much as possible. A tape-to-head gap is *extremely* degrading at the higher frequencies. Figure 6-3 shows this. Note how a .1-mil gap causes only a small loss at 10 kHz. But, at 1 MHz, it causes a 46-dB loss! <u>You must maintain tape-to-head contact.</u> Keeping the magnetic tape recorder heads and tape transport clean is the best way to do this.

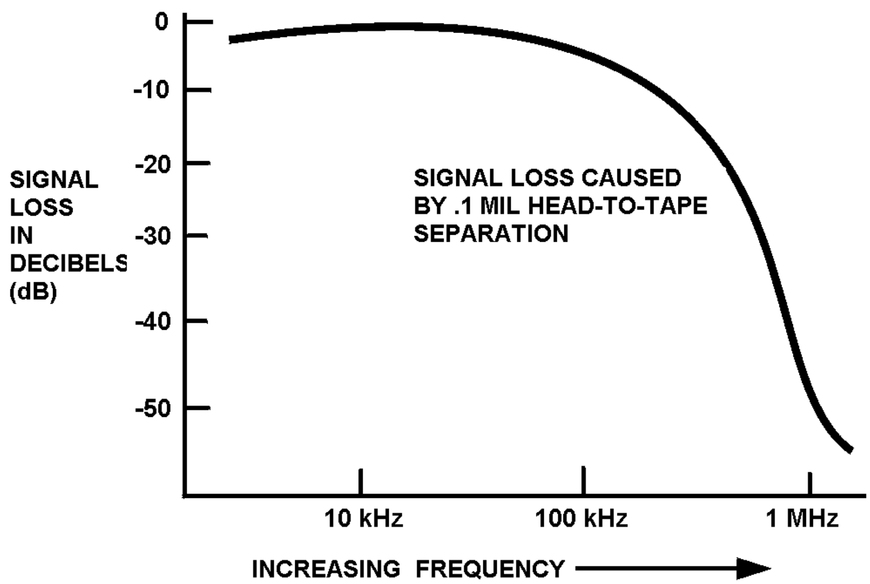


Figure 6-3.—Effects of poor tape-to-head contact.

Q-1. Two tape recorders have signal-to-noise ratios (SNRs) of 25-dB RMS and 35-dB RMS respectively. Which of the SNRs can record and reproduce the widest range of input signals and why?

Q-2. You plan to measure your tape recorder's SNR. What test equipment will you need?

Q-3. Technical manuals for tape recorders can state the SNR in what three different ways?

Q-4. The frequency-response specification of your tape recorder reads within +/– 3 dB from 150 Hz to 150 kHz at 60 ips. What does this mean?

Q-5. While measuring frequency response, as the signal generator approaches the lowest and highest frequency the recorder can effectively record, the VTVM reading drops to less than – 3 dB. What does this indicate?

Q-6. List four factors that can degrade the frequency response of magnetic tape recorders.

HARMONIC DISTORTION

A magnetic tape recorder's harmonic-distortion specification is very important. It usually determines where the record level of a recorder's electronics should be set. The record level is also used to determine the signal-to-noise ratio and frequency-response specifications. A typical harmonic-distortion specification might read "1% third harmonic of a 100-kHz signal at 60 ips." This means that the magnetic tape recorder has 1% third-harmonic distortion of a 100-kHz signal at 60 ips.

HARMONIC-DISTORTION DEFINITION

Harmonic distortion is *the production of harmonic frequencies by an electronic system when a signal is applied at the input.* When an input signal goes through nonlinear electronic circuitry, the output signal will include some harmonic distortion (or unwanted frequencies). If you analyzed this distortion, you'd see that a pattern exists. A pattern, whereby the frequency of each unwanted frequency is a multiple ($\times 1$, $\times 2$, $\times 3$, etc.) of the center frequency of the input signal.

There are two types of harmonic distortion: even-order and odd-order. If the frequencies of the distortion are 2, 4, 6, etc., times the center frequency, it's even-order harmonics. If the frequencies of the distortion are 3, 5, 7, etc., times the center frequency, it's odd-order harmonics.

Odd-order harmonics are normally caused by the magnetic tape itself. Even-order harmonics are normally caused by (1) permanently magnetized magnetic heads, (2) faulty circuits, or (3) asymmetrical or unbalanced bias signals. As you might guess, even-order harmonics can be reduced by doing the right maintenance and periodic performance tests.

The primary harmonic distortion in magnetic tape recorder systems is third-order harmonics. If the level of third-order harmonics in a recorder increases, the level of distortion will also increase (figures 6-4A and B show this relationship). Two things that determine the level of third-order harmonics in a recorder are (1) the signal bias level, and (2) the record level. Figure 6-4A shows how third-order harmonic distortion decreases as the signal bias level increases. Figure 6-4B shows how the third harmonic increases gradually at first and then abruptly as the record level increases. That's why the third harmonic is used to determine the normal record level.

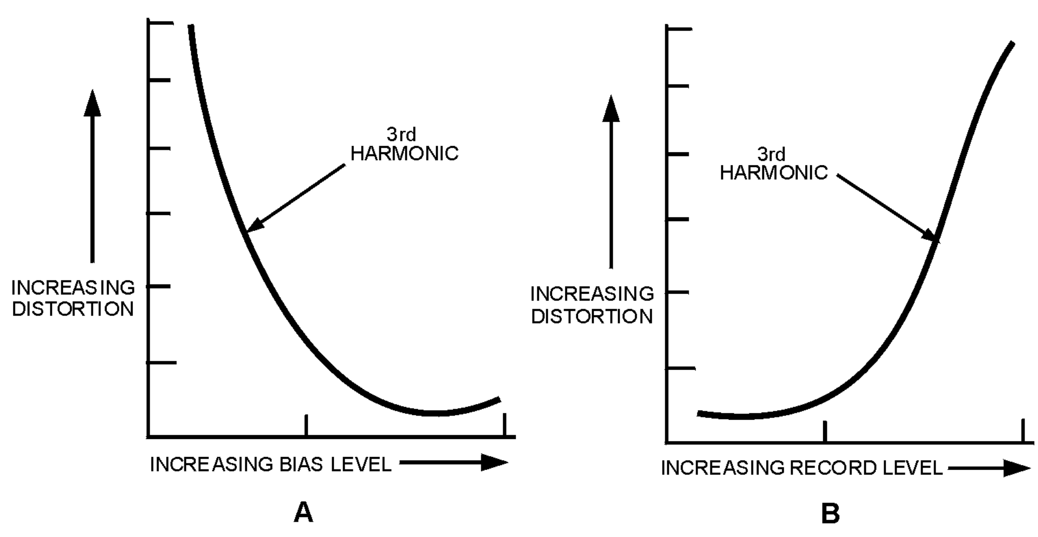


Figure 6-4 A & B.—Effect of signal bias level and record level on harmonic-distortion level.

HARMONIC-DISTORTION MEASUREMENT

Figure 6-5 shows a typical test equipment setup for measuring harmonic distortion. With this setup, the test signal from the signal generator is recorded and reproduced by the magnetic tape recorder at a normal record level. The amount of harmonic distortion is measured at the recorder's output on the wave analyzer.

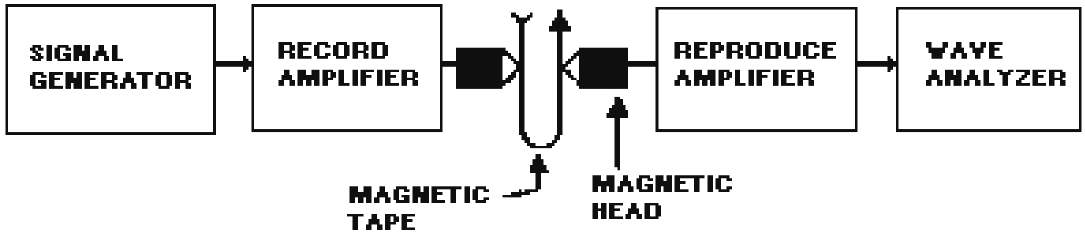


Figure 6-5.—Test equipment setup for measuring harmonic distortion.

The technical manual for the magnetic recorder you're testing will tell you how to set up the test equipment. It'll tell you to set up the wave analyzer to measure a specific frequency. This frequency will be one of the multiples (×1, ×2, ×3, etc.) of the frequency the signal generator is outputting.

For example, let's say the technical manual told you to set up the signal generator to input a 10-kHz test signal into the magnetic tape recorder. Since you want to measure third-order harmonics, the technical manual will tell you to set the wave analyzer to measure the amount of harmonic distortion at 30-kHz.

PHASE RESPONSE

It used to be thought that the only important specifications of magnetic tape recorders were signal-to-noise ratio and frequency response. But now, with the need to record and reproduce more complex waveforms, such as telemetry and computer data, the phase-response specification becomes as important as frequency response.

PHASE-RESPONSE DEFINITION

Phase response is *the expression of the variation of the phase shift with respect to frequency*. A good magnetic tape recorder will have linearly increasing phase response as frequency increases. In simpler terms, good phase response shows that a magnetic recorder can reproduce a complex waveform (such as a square wave which has an infinite number of sine waves) without distorting it. Figure 6-6 shows both good and bad phase response.

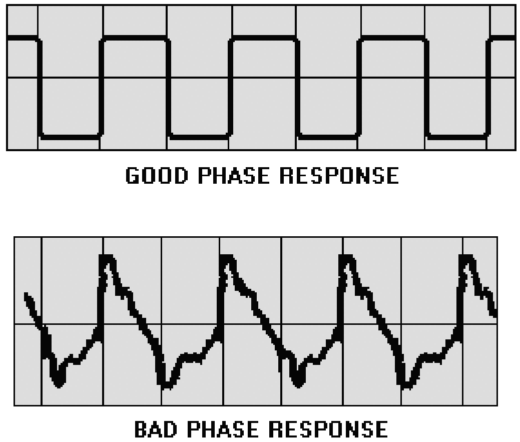


Figure 6-6.—Pictures showing the effect of good and bad phase response on square-wave reproduction.

PHASE-RESPONSE MEASUREMENT

You cannot *directly* measure phase response. The best way to check the phase response of a magnetic tape recorder is to record and reproduce a square wave and watch the output on an oscilloscope. If the output signal is symmetrical, like in figure 6-7, the recorder has good phase response.

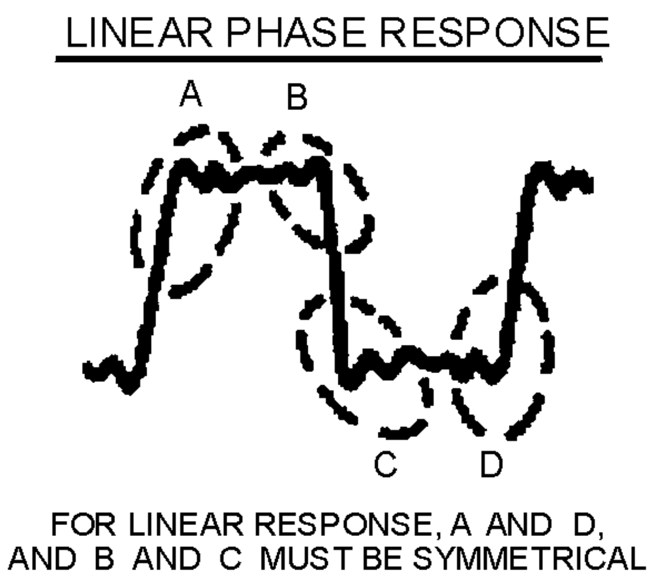


Figure 6-7.—An example of good linear phase response.

Q-7. A recorder's harmonic-distortion specification reads 2% third harmonic of a 100-kHz signal at 60 ips. What does this mean?

Q-8. What are three possible causes of even-order harmonics?

Q-9. What number harmonic is the primary harmonic distortion in magnetic tape recorders?

Q-10. When measuring harmonic distortion, you set the signal generator to input a 15-kHz test signal. To what frequency should you set the wave analyzer?

Q-11. How should a tape recorder with good phase response reproduce a complex waveform, such as a square wave?

Q-12. How could you check the phase response of a tape recorder?

FLUTTER

The general audio and broadcast field coined the term *flutter* to describe what you'll actually hear from the bad effects of this specification.

FLUTTER DEFINITION

Flutter is *the result of non-uniform tape motion caused by variations in tape speed that produces frequency modulation of signals recorded onto magnetic tape.*

Flutter is usually expressed as a *percent peak* or a *peak-to-peak* value for instrumentation recorders and as a *root-mean-square (RMS)* value for audio recorders. It's caused by magnetic tape transports. Low-frequency flutter (below 1000 Hz) is caused by the rotating parts of a tape transport such as:

- Irregular magnetic tape supply or take-up reels.
- Uneven or sticking guide rollers and pinch rollers.
- Capstans.

High-frequency flutter (above 1000 Hz) is caused by the *fixed* parts of a tape transport, such as fixed tape guides and magnetic heads. When the magnetic tape passes over a fixed tape guide or magnetic head, the transition from static to dynamic friction causes something called *stiction*. It's this stiction that causes the variations in tape speed which, in turn, cause the flutter.

As you might guess, it's hard to prevent flutter. The only way to lessen flutter is through skilled engineering, machining, and design of magnetic tape recorders.

FLUTTER MEASUREMENT

There are many ways to measure flutter. Most are based on the fact that tape speed variations cause frequency modulation of a recorded tone. Figure 6-8 shows a typical setup for measuring the peak-to-peak value of flutter with a frequency-modulation (FM) demodulator and an oscilloscope. The technical manual for the magnetic tape recorder you're testing will tell you how to set up the signal generator to output the test signal. After setting up the test equipment, follow these procedures:

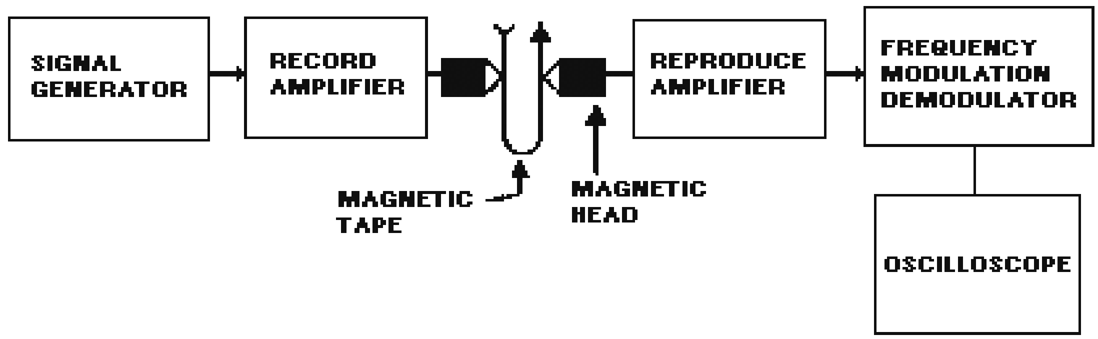


Figure 6-8.—Test equipment setup for measuring flutter.

1. Record the test signal onto magnetic tape; then rewind the magnetic tape. This is necessary because you can't measure flutter as you're recording. Since the tape-speed variation past the record head is almost the same as past the reproduce head, the flutter level is too small to see.
2. After you rewind the tape, play it back. During playback, the output signal from the tape recorder goes through the FM demodulator to remove the original test signal. The waveform you now see on the oscilloscope is the actual flutter signal that was modulated onto the test signal.
3. Using the oscilloscope display, measure the peak-to-peak value of the flutter signal.

TIME-BASE ERROR

The time-base error (TBE) specification of magnetic tape recorders is closely related to the flutter specification. In fact, the TBE is a direct measure of the effects of flutter on the stability of recorded data.

TIME-BASE ERROR DEFINITION

The TBE is *the time-relationship error between two or more events recorded and reproduced from the same magnetic tape.* It's also defined *as the displacement of a point on the magnetic tape from where it should have been, during a specific time interval.*

A typical TBE specification might read "+ / – 100 microseconds over a 10-millisecond time interval at a tape speed of 60 inches per second, referenced to a control tone." This means that the time-base error could cause a signal to *jitter* +/– 100 microseconds over a 10-millisecond period at a tape speed of 60 inches per second.

TBE jitter introduces noise or unwanted frequency modulation (when using FM recording techniques) into the magnetic tape recording process. It can also cause a loss of accuracy in pulse-duration modulation (PDM), pulse-coded modulation (PCM), or other magnetic recordings where precise timing relationships exist between two or more signals.

TIME-BASE ERROR MEASUREMENT

The simplest way to measure the TBE is with an oscilloscope. Figure 6-9 shows a typical test equipment setup for measuring TBE. After you set up the test equipment, measure the TBE as follows:

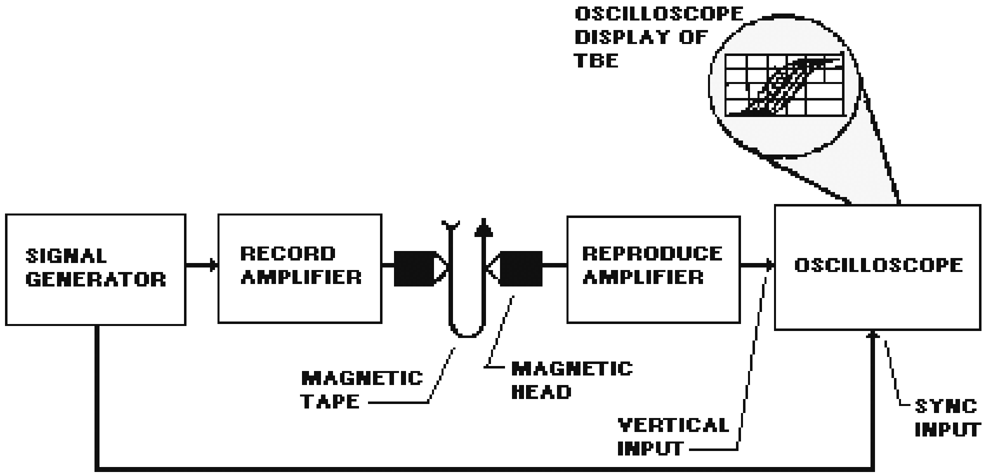


Figure 6-9.—Test equipment setup for measuring time-base error.

1. Set the signal generator to generate a test signal. The technical manual for the magnetic tape recorder you're testing will tell you how.
2. Connect the test signal output from the signal generator to both the recorder's input and the oscilloscope's trigger (sync) input.
3. Connect the output of the tape recorder to the oscilloscope's signal (vertical) input.
4. Record and reproduce the test signal.
5. Adjust the oscilloscope's intensity control until you can see the TBE on the oscilloscope's display. (Limit glare by using a hood on the oscilloscope's display.)

SKEW

This magnetic tape recording specification only applies to multi-tracked magnetic tape recorders.

SKEW DEFINITION

Skew is *the inter-track fixed and dynamic displacement, or change in azimuth, encountered by different tracks across the width of the magnetic tape as it passes the magnetic heads*. In other words, it's the time difference between the tracks on a multi-tracked magnetic head.

A typical skew specification might read "+/– 0.15 microseconds between adjacent tracks on the same head stack at 120 inches per second." This means that one of the tracks on a magnetic head could lead, or lag, the track next to it by as much as 0.15 microseconds at 120 ips. This specification applies to both fixed and dynamic skew.

Fixed skew can be caused by

- magnetic tape recorder electronics,

- gap scatter in the magnetic head stack,
- azimuth alignment of the magnetic head stack, or
- fixed difference in tension along the tape path

You can minimize most fixed skew by adjusting the magnetic recorder's electronics or by realigning the magnetic heads.

Fixed skew errors usually do not show up when magnetic tapes are recorded and reproduced on the same tape recorder. Since fixed skew errors are additive, they'll usually show up when you record on one magnetic tape recorder and then reproduce on another.

Dynamic skew errors are caused by either the magnetic tape transport or the magnetic tape itself. If the tape transport guides are worn or sticking, the magnetic tape won't properly pass over the magnetic heads. It'll *drift* and pass the magnetic head at an angle (like a car skidding on an icy road). If the magnetic tape itself is warped or isn't uniform across its width it, too, will cause dynamic skew.

SKEW MEASUREMENT

Skew is best measured with an oscilloscope. Figure 6-10 shows a typical test equipment setup for measuring skew. The technical manual for the magnetic tape recorder you're testing will tell you how to set up the signal generator. After test equipment setup, measure the skew as follows:

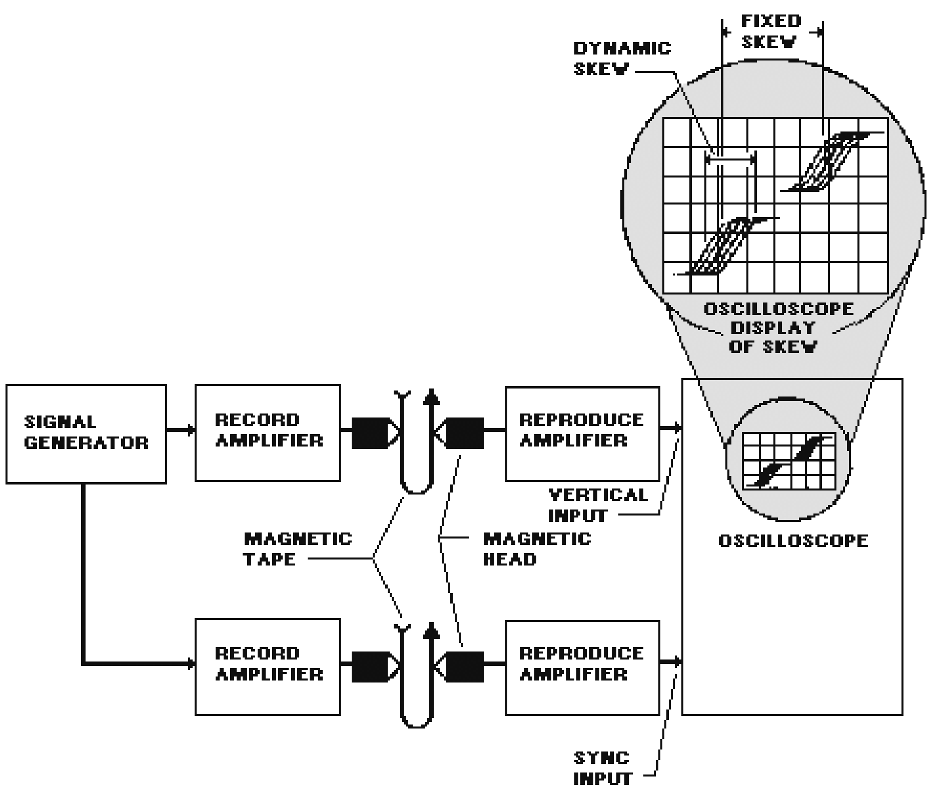


Figure 6-10.—Test equipment setup for measuring skew.

1. Inject the test signal into a reference track and one other track of the multi-track magnetic tape recorder. (The reference track should be one of the two *outside* tracks of the magnetic head.)

2. Connect the output from the reference track to the *sync* input of the oscilloscope to trigger the horizontal sweep.

3. Connect the output from the other track to the *vertical* input of the oscilloscope.

4. While recording and reproducing the test signal, measure the fixed and dynamic skews which are displayed on the oscilloscope. Figure 6-10 shows how this looks.

Q-13. What causes flutter in a tape recorder's output?

Q-14. What causes low-frequency flutter (below 1000 Hz)?

Q-15. What causes high-frequency flutter (above 1000 Hz)?

Q-16. Your recorder's TBE specification reads " +/− 80 microseconds over a 10 millisecond time interval at a tape speed of 60 ips, referenced to a control tone." What does this mean?

Q-17. Why is it important to minimize TBE jitter in magnetic tape recordings where precise timing relationships exist between two or more signals?

Q-18. The skew specification of your multi-tracked tape recorder reads " +/− 0.20 microseconds between adjacent tracks on the same head stack at 120 ips." What does this mean?

Q-19. How can you minimize fixed skew?

Q-20. When are fixed skew errors most likely to show up?

Q-21. How do worn or sticking tape transport guides cause dynamic skew on a multi-track recorder?

SUMMARY

Now that you've finished chapter 6, you should be able to describe the seven most common magnetic tape recording specifications and how to measure each specification. The following is a summary of important points in this chapter:

The **SIGNAL-TO-NOISE RATIO (SNR)** is the ratio of the normal signal level to the tape recorder's own noise level measured in dB. The higher a recorder's SNR, the wider the range of signals it can record and reproduce.

SNR IS STATED IN ONE OF THREE WAYS based on how it was measured. If you don't know the way it was measured, you could be misled.

A recorder's **FREQUENCY-RESPONSE** specification is sometimes called its *bandwidth*. It tells the range of frequencies a recorder can effectively record and reproduce. Factors that can degrade a recorder's frequency response are an improper *bias level setting, reproduce head gap*, or *tape transport speed.* Also, failure to clean the heads and the tape transport can cause poor *tape-to-head contact.*

HARMONIC DISTORTION is the production of unwanted harmonic frequencies when a signal is applied at the recorder's input. The primary harmonic distortion in tape recorders is third order harmonics. It's measured with a wave analyzer. You can reduce this distortion with proper preventive maintenance and periodic performance tests.

Good **PHASE RESPONSE** means the recorder can reproduce complex waveforms such as square waves without distortion. The best way to check a recorder's phase response is by recording and reproducing a square wave and checking the output on an oscilloscope.

FLUTTER results from non-uniform tape motion caused by variations in tape speed. The tape speed variations are caused by design and machining deficiencies in the rotating and fixed parts of the tape transport.

TIME-BASE ERROR (TBE) is the time-relationship error between two or more events recorded on and reproduced from the same magnetic tape. It causes TBE jitter, which introduces noise or loss of accuracy where precise timing relationships exist between two or more signals.

SKEW is the time difference in microseconds between the tracks on a multi-tracked tape recorder. *Fixed* or *dynamic* skew can happen when one of the tracks on the multi-track head leads or lags the track next to it. *Fixed skew* errors only show up when you record on one recorder and reproduce on a different recorder. You can minimize fixed skew by adjusting the recorder's electronics and aligning the heads. *Dynamic skew* errors are caused by worn or sticking tape transport guides or by warped magnetic tape.

ANSWERS TO QUESTIONS Q1. THROUGH Q21.

A1. 35-dB RMS because the highest SNR can always record and reproduce the widest range of input signals.

A2. A VTVM and a signal generator. (See figure 6-1.)

A3.

- *a. Root-mean-square (RMS) signal-to-RMS noise.*
- *b. Peak-to-peak signal-to-RMS noise.*
- *c. Peak signal-to-RMS noise.*

A4. The recorder can record all frequencies between 150 Hz and 150 kHz at 60 ips without varying the output amplitude more than 3 dB.

A5. The upper and lower limits of the frequency response specification for that tape recorder.

A6.

- *a. A too-high or too-low bias signal level setting for the record head.*
- *b. An improper reproduce head gap.*
- *c. An improper tape transport speed.*
- *d. Poor tape-to-head contact.*

A7. The recorder has 2% third-harmonic distortion of a 100-kHz signal at 60 ips.

A8.

a. Permanently magnetized heads.

b. Faulty circuitry.

c. Asymmetrical bias signal.

A9. Third-order harmonic.

A10. 45 kHz.

A11. With no distortion.

A12. Record and reproduce a square wave and see if the output on an oscilloscope is symmetrical.

A13. Non-uniform tape motion caused by variations in tape speed.

A14. Rotating parts of a tape transport, such as irregular tape reels, sticking guides and pinch rollers, and capstans.

A15. Fixed parts of a tape transport, such as fixed tape guides and magnetic heads.

A16. The TBE could cause a signal to jitter +/− 80 microseconds over a 10-millisecond period at a tape speed of 60 ips.

A17. The jitter could cause noise and a loss of accuracy.

A18. One of the tracks on a magnetic head could lead or lag the track next to it by as much as 0.20 microseconds at 120 ips.

A19. Adjust the recorder's electronics or realign the magnetic heads.

A20. When you record on one tape recorder and then reproduce on a different recorder.

A21. The tape drifts past the multi-track head at an angle.

CHAPTER 7

DIGITAL MAGNETIC TAPE RECORDING

LEARNING OBJECTIVES

After completing this chapter, you'll be able to do the following:

1. Describe the characteristics of digital magnetic tape recording and the difference between *analog* and *digital* recording.

2. Describe each of the three *formats* for digital magnetic tape recording (serial, parallel, and serial-parallel).

3. Define the following terms as they apply to digital magnetic tape recording: *mark, space, bit-cell period, packing density,* and *bit-error rate (BER)*.

4. Describe the eight most common methods for encoding digital data onto magnetic tape.

5. Describe the characteristics and use of the following categories of digital magnetic tape recorders: (1) computer-compatible, (2) telemetry, and (3) instrumentation.

INTRODUCTION TO DIGITAL MAGNETIC TAPE RECORDING

This chapter introduces you to digital magnetic tape recording. It describes (1) the three formats for digital magnetic tape recording, (2) the eight methods of encoding digital data onto magnetic tape, and (3) the configuration differences between the three types of digital tape recorders.

Until now, you've learned about magnetic tape recording from an *analog* point-of-view. That is, the signal you record and reproduce is the actual analog input signal waveform. In digital magnetic tape recording, the signal you record and reproduce is, instead, a series of *digital* pulses. These pulses are called binary *ones* and *zeros*. These *ones* and *zeros* can represent one of three types of data: (1) data used by digital computers, (2) pulsed square-wave signals, or (3) digitized analog waveforms.

The digital magnetic tape recording process stores data onto tape by magnetizing the tape to its saturation point in one of two possible polarities: positive (+) or negative (–). The *saturation* point of magnetic tape is the point where the magnetic tape is magnetized as much as it can be.

DIGITAL MAGNETIC TAPE RECORDING FORMATS

There are three digital magnetic tape recording formats: *serial, parallel*, and *serial-parallel*. Each of these is described below. Figure 7-1 shows each of the three formats as they apply to recording an eight-bit binary data stream.

SERIAL DIGITAL MAGNETIC TAPE RECORDING FORMAT

This is the simplest of the three digital magnetic tape recording formats. It's usually used when recording instrumentation or telemetry data. In this format, the incoming data pulses are recorded onto a single recorder track of the magnetic tape in a single, continuous stream. Figure 7-1A shows how this looks.

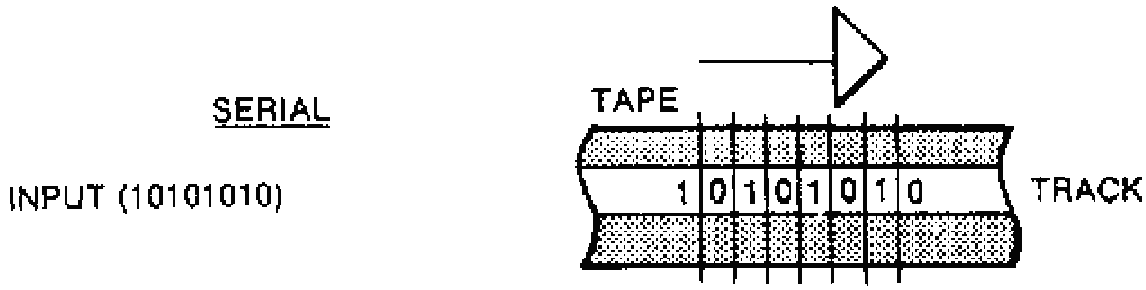


Figure 7-1A.—Digital magnetic tape recording formats.

PARALLEL DIGITAL MAGNETIC TAPE RECORDING FORMAT

In this format, the incoming data pulses come in on more than one input channel and are recorded *side-by-side* onto more than one tape track. The data pulses across the width of the magnetic tape are *related* to each other. Figure 7-1B shows how this looks. This format is usually used to store computer data.

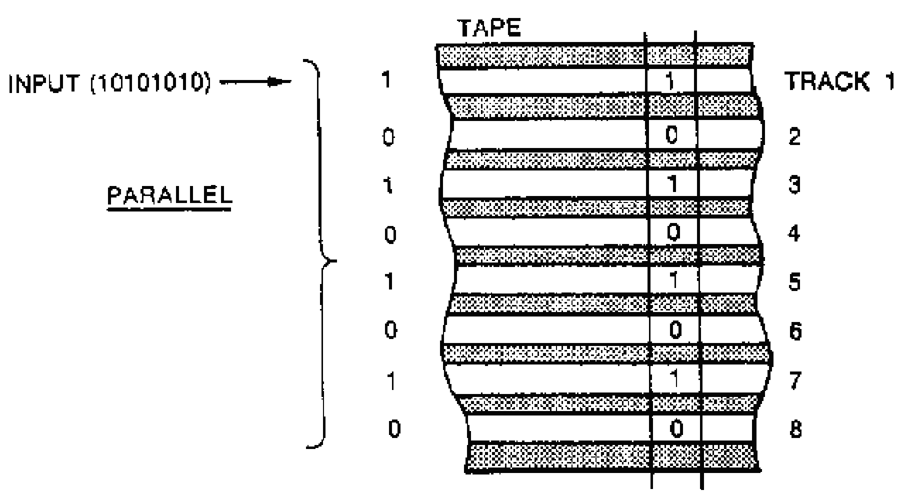


Figure 7-1B.—Digital magnetic tape recording formats.

SERIAL-PARALLEL DIGITAL MAGNETIC TAPE RECORDING FORMAT

This format is more complex. It takes a serial input stream of data pulses, breaks them up, and records them on more than one recorder track. When the tape is reproduced, the recorder recombines the broken-apart data into its original form. Figure 7-1C shows how this looks. The serial-parallel format is usually used in instrumentation recording when the input data rate is high.

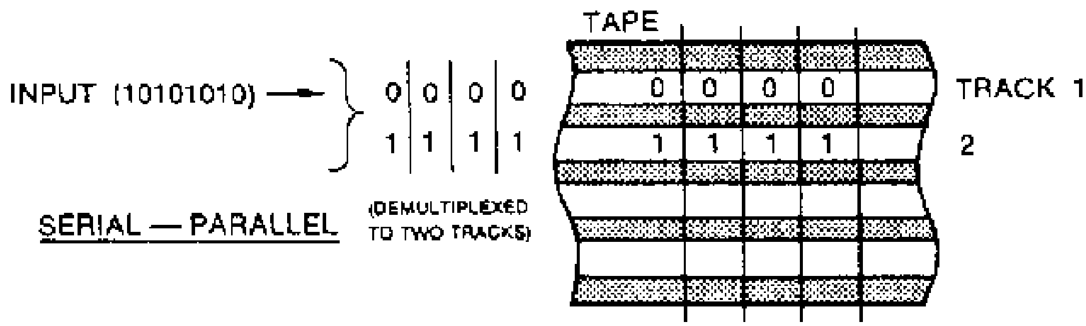


Figure 7-1C.—Digital magnetic tape recording formats.

DIGITAL MAGNETIC TAPE RECORDING DEFINITIONS

Before we describe the methods for encoding digital data onto magnetic tape, let's define the following terms:

Mark: The voltage state of a digital *one (1)* data bit. It's also sometimes called *true*.

Space: The voltage state of a digital *zero (0)* data bit. It's also sometimes called *false*.

Bit-cell period: The time occupied by a single digital bit.

Packing density: The number of bits per fixed length of magnetic tape per track. There are three categories of packing density:

1. Low density—200 to 1,000 bits per inch (bpi).
2. Medium density—1,000 to 8,000 bpi.
3. High density—8,000 to 33,000 bpi.

Bit-error rate: The number of bits within a finite series of bits that will be reproduced incorrectly.

Q-1. In digital magnetic tape recording, the series of recorded digital pulses can represent what three types of data?

Q-2. What three formats are used for digital magnetic tape recording?

Q-3. What format of digital tape recording is normally used to store computer data?

Q-4. What format of digital tape recording takes a serial input stream of data pulses, breaks them up, and records them on more than one data track?

Q-5. What format of digital tape recording is normally used to record instrumentation or telemetry data?

DIGITAL MAGNETIC TAPE RECORDING ENCODING METHODS

This section describes how digital data is *electrically* encoded onto the magnetic tape. The following paragraphs describe the eight most common digital data encoding methods.

1. Return to bias (RB)
2. Return to zero (RZ)
3. Non-return to zero (NRZ) and these four variations of the NRZ method:
    a. Non-return-to-zero level (NRZ-L)
    b. Enhanced non-return-to-zero level (E-NRZ-L)
    c. Non-return-to-zero mark (NRZ-M)
    d. Non-return-to-zero space (NRZ-S)
4. Bi-phase level

RETURN-TO-BIAS (RB) ENCODING

The RB encoding method uses magnetic tape that is *pre-set* to one of the two polarities (+ or –). This pre-sets the magnetic tape to *all zeros.* Digital *ones* are then recorded onto the magnetic tape by magnetizing the tape in the opposite polarity. After each *one* pulse, the tape returns to its original bias condition. Figure 7-2 shows the magnetic tape *preset* to a negative bias condition. It also shows how the digital data word 0100110001 is stored onto the magnetic tape using the RB encoding method.

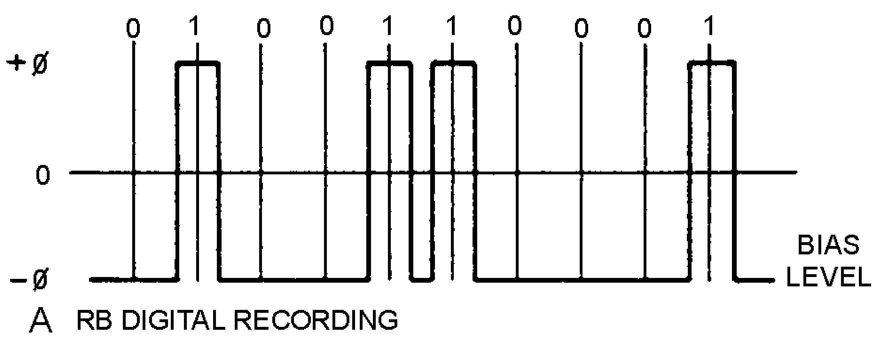


Figure 7-2.—Return-to-bias (RB) digital encoding method.

This method has a serious drawback: It requires an external clocking signal to read the zeros stored on the tape.

RETURN-TO-ZERO (RZ) ENCODING

The RZ encoding method uses magnetic tape that is normally in a *neutral* condition (the tape is not biased positively or negatively). A digital *one* is recorded as a positive-going pulse: a digital *zero* is recorded as a negative-going pulse. The magnetic tape returns to its *neutral* state in between pulses. Figure 7-3 shows the magnetic tape in its neutral state. It also shows how the digital data word 0100110001 is stored onto the magnetic tape using return-to-zero encoding.

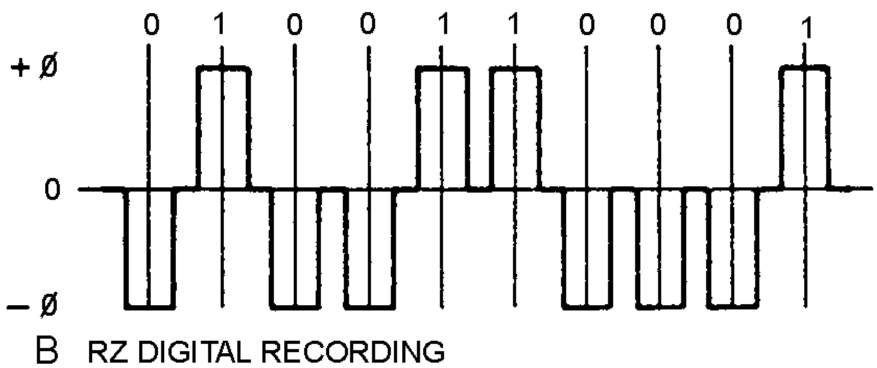


Figure 7-3.—Return-to-zero (RZ) digital encoding method.

NON-RETURN-TO-ZERO (NRZ) ENCODING

The NRZ encoding method is, by far, the most widely used. It's accurate, simple, and reliable. It does <u>not</u> return the magnetic tape to its *neutral* state in between pulses. The magnetic tape is always in saturation, either positively or negatively. The polarity of the saturating signal only changes when incoming data changes from a *zero* to a *one* and vice versa. Figure 7-4 shows how the digital data word 101100011010 is stored onto the magnetic tape using the NRZ encoding method.

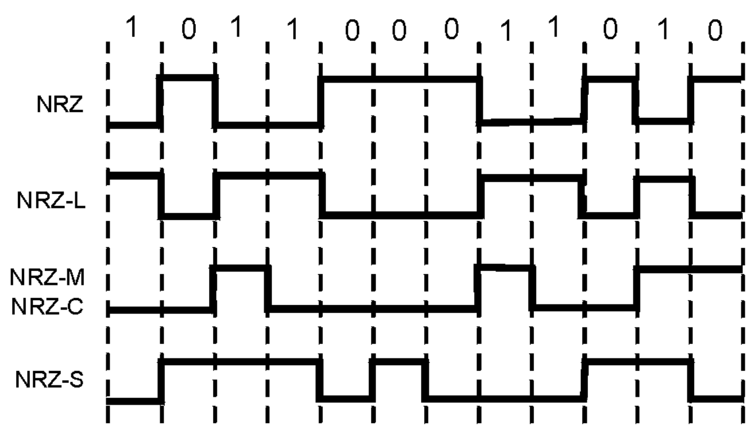


Figure 7-4.—Non-return-to-zero (NRZ) digital encoding method.

There are four widely used variations to the basic NRZ encoding method. Each of these is described in the following paragraphs.

Non-Return-To-Zero-Level (NRZ-L) Encoding

In NRZ-L encoding, the polarity of the saturating signal changes only when the incoming signal changes from a *one* to a *zero* or from a *zero* to a *one*. Figure 7-4 also shows how the digital data word 101100011010 is stored onto the magnetic tape using the NRZ-L encoding method. Note that the NRZ-L method looks just like the NRZ method, except for the first input *one* data bit. This is because NRZ does not consider the first data bit to be a polarity change, where NRZ-L does.

The NRZ-L encoding method isn't normally used in higher density (over 20,000 bpi) digital magnetic recording. This encoding method is sometimes called the non-return-to-zero-change (NRZ-C) encoding method.

Enhanced Non-Return-to-Zero-Level (E-NRZ-L) Encoding

This encoding method takes the basic NRZ-L data and adds a parity bit to it after every seven incoming data bits. The polarity of the parity bit is such that the total number of *ones* in the eight-bit data word will be an *odd* count. Figure 7-5 shows how the digital data word 0100010 is stored onto the magnetic tape using the E-NRZ-L encoding method.

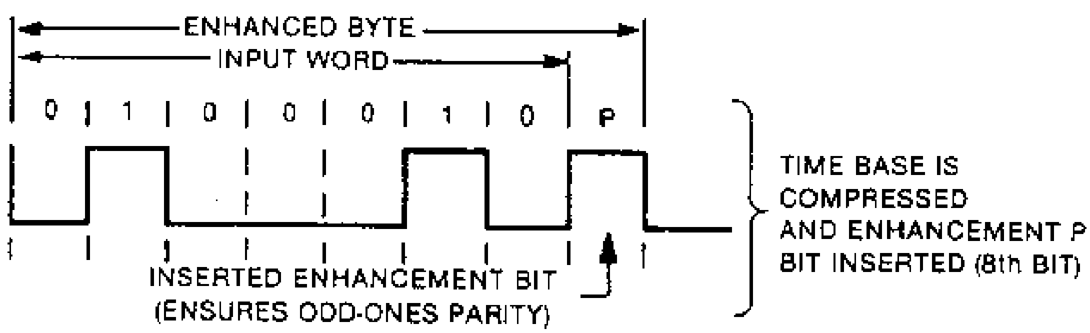


Figure 7-5.—Enhanced non-return-to-zero-level (E-NRZ-L) digital encoding method.

Before the parity bit is added, the original incoming data is compressed in time. This is done so that when the parity bit is added, the eight-bit data word takes up the same amount of time as the original-seven bit data word. When the tape is reproduced, the parity bit is taken out.

This encoding method works very well in high density (up to 33,000 bpi) magnetic tape recording. And, it offers an extremely good bit-error rate of 1 error per 1 million bits.

Non-Return-to-Zero-Mark (NRZ-M) Encoding

The NRZ-M encoding method is probably the most widely used encoding method for 800-bpi digital magnetic tape recording. In this method, the polarity of the saturating signal changes when the incoming signal is a *one.* An incoming *zero* would not change the polarity of the saturating signal.

NRZ-M offers better protection from error than straight NRZ. In NRZ-M, there's a one-to-one relationship between incoming data and polarity changes. If one data bit is lost, only that one bit is lost.

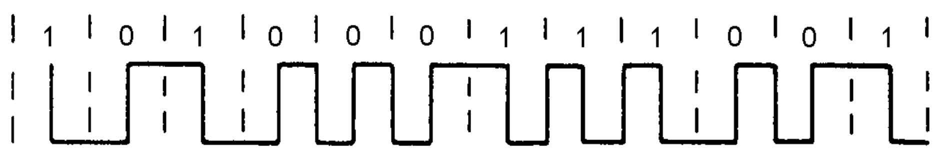
Whereas, in straight NRZ, if one bit is lost, all of the bits that follow will be exactly the opposite in polarity from what they should be. Figure 7-4 also shows how the digital data word 101100011010 is stored onto the magnetic tape using the NRZ-M encoding method.

Non-Return-to-Zero-Space (NRZ-S) Encoding

The NRZ-S encoding method works just like NRZ-M encoding, with one exception. Instead of the saturating signal changing polarity when the incoming data signal is a *one*, it changes when the incoming data signal is a *zero*.

BI-PHASE LEVEL ENCODING

The bi-phase level encoding method records two logic levels for each incoming data bit. When an incoming data bit is a *one*, bi-phase level recording records a *zero-one*. When an incoming data bit is a *zero*, bi-phase level recording records a *one-zero.* This encoding method helps to overcome any low-frequency response problems that the magnetic tape recorder may have. Figure 7-6 shows how the digital data word 101000111001 is stored onto magnetic tape using the bi-phase encoding method.



Courtesy of DATATAPE Incorporated, Copyright © All rights reserved

Figure 7-6.—Bi-phase level digital encoding method.

Bi-phase encoding requires exactly twice the bandwidth of NRZ-L. That's why it's mostly used in medium-density digital magnetic tape recording. In fact, this encoding method is probably the most widely used encoding method for 1600-bpi digital magnetic tape recording.

DIGITAL MAGNETIC TAPE RECORDER USES

As you already know, digital magnetic tape recorders are used to store and retrieve *digital* data. These recorders fall into one of three categories, (1) computer compatible, (2) telemetry, and (3) instrumentation.

COMPUTER-COMPATIBLE DIGITAL TAPE RECORDERS

Computer-compatible digital tape recorders store and retrieve computer programs and data. They're usually multi-tracked tape recorders with at least two, and up to nine, tracks for data. They use either 1/4" or 1/2" magnetic tape on either reels or cartridges.

TELEMETRY DIGITAL TAPE RECORDERS

Telemetry digital magnetic tape recorders are more commonly called *wideband* recorders. They're used for recording radar signals and other pulsed square-wave type signals with a bandwidth of 500 kHz to 2 MHz. They're also multi-tracked tape recorders that have either 14 or 28 tracks for data. They use 1" magnetic tape on either aluminum or glass reels.

INSTRUMENTATION MAGNETIC TAPE RECORDERS

Instrumentation digital magnetic tape recorders are used to record other special signals with a bandwidth of less than 500 kHz. They, too, are multi-tracked recorders, normally with 7 tracks for data. They use 1/2" magnetic tape on metal or glass reels.

Q-6. Which of the eight methods for encoding digital data onto magnetic tape is most widely used because it's accurate, simple, and reliable?

Q-7. Which digital data tape encoding method presets the magnetic tape to all zeros and then records digital ones onto the tape?

Q-8. Which digital data encoding method records a digital one as a positive pulse and a digital (zero) as a negative pulse and returns the tape to neutral between pulses?

Q-9. Which method of digital data encoding does NOT return the tape to neutral between pulses but, instead, saturates the tape positively or negatively as the incoming data changes between zero and one?

Q-10. What are the four widely used variations of the NRZ encoding method?

Q-11. Which digital data encoding method helps overcome a tape recorder's low-frequency response problems by recording two logic levels for each incoming data bit?

Q-12. Digital magnetic tape recorders used to store and retrieve digital data fall into what three categories?

Q-13. What category of digital tape recorder is used for recording pulsed square-wave signals with a bandwidth of 500 kHz to 2 MHz?

Q-14. What category of digital tape recorder is used to record special signals with a bandwidth of less than 500 kHz?

SUMMARY

Now that you've finished chapter 7, you should be able to describe (1) the characteristics of digital magnetic tape recording, (2) the three *formats* for digital magnetic tape recording, (3) the eight *methods* for encoding digital data onto magnetic tape, and (4) the characteristics and uses of the three types of digital magnetic tape recorders. The following is a summary of important points in this chapter:

Digital magnetic tape recorders record a SERIES OF DIGITAL PULSES called binary *ones* and *zeros*. These digital pulses can represent (1) data used by digital computers, (2) pulsed square-wave signals, or (3) digitized analog waveforms.

Three **FORMATS FOR DIGITAL MAGNETIC TAPE RECORDING** are *serial, parallel,* and *serial-parallel*.

There are **EIGHT COMMONLY USED METHODS FOR ENCODING** digital data onto magnetic tape. The non-return-to-zero (NRZ) method and the four variations of the NRZ method are most commonly used.

THREE CATEGORIES OF DIGITAL MAGNETIC TAPE RECORDERS are (1) computer-compatible, (2) telemetry, and (3) instrumentation.

ANSWERS TO QUESTIONS Q1. THROUGH Q14.

A1.

a. *Data used by digital computers.*

b. *Pulsed squarewave signals.*

c. *Digitized analog waveforms.*

A2. (1) Serial, (2) parallel, and (3) serial-parallel.

A3. Parallel digital magnetic tape recording.

A4. Serial-parallel digital magnetic tape recording.

A5. Serial digital magnetic tape recording.

A6. Non-return-to-zero (NRZ) encoding.

A7. Return-to-bias (RB) encoding.

A8. Return-to-zero (RZ) encoding.

A9. Non-return-to-zero (NRZ) encoding.

A10.

a. *Non-return-to-zero level (NRZ-L).*

b. *Enhanced non-return-to-zero level (E-NRZ-L).*

c. *Non-return-to-zero mark (NRZ-M).*

d. *Non-return-to-zero space (NRZ-S).*

A11. Bi-phase level encoding.

A12.

a. *Computer-compatible digital tape recorders.*

b. *Telemetry digital tape recorders.*

c. *Instrumentation digital tape recorders.*

A13. Telemetry digital tape recorders.

A14. Instrumentation digital tape recorders.

CHAPTER 8

MAGNETIC DISK RECORDING

LEARNING OBJECTIVES

After completing this chapter, you'll be able to do the following:

1. Describe how flexible (floppy) disks are constructed; how data is organized on them; how they are handled, stored, and shipped; and how they are erased.
2. Describe how fixed (hard) disks are constructed; how data is organized on them; how they are handled, stored, and shipped; and how they are erased.
3. Describe each of the following methods for recording (encoding) digital data onto magnetic disks: *frequency-modulation* encoding, *modified frequency-modulation* encoding, and *run length-limited* encoding.
4. Describe the characteristics of *floppy disk drive transports* and *hard disk drive transports* and describe the preventive maintenance requirements of each type.
5. Describe the following parts of the *electronics* component of a magnetic disk drive: control electronics, *write/read* electronics, and *interface* electronics.
6. Describe the five most common types of disk drive interface electronics.
7. Define the following magnetic disk recording *specifications:* seek time, latency period, access time, interleave factor, transfer rate, and recording density.

INTRODUCTION

Magnetic disk recording was invented by International Business Machines (IBM) in 1956. It was developed to allow mainframe computers to store large amounts of computer programs and data. This new technology eventually led to what's now known as the computer revolution.

This chapter introduces you to the following aspects of magnetic disk recording:

- Disk recording mediums
- Disk recording methods
- Disk drive transports
- Disk drive electronics
- Disk recording specifications

MAGNETIC DISK RECORDING MEDIUMS

There are two types of disk recording mediums: *flexible diskettes* and *fixed (hard) disks.* The following paragraphs describe (1) how flexible and fixed disks are made; (2) how data is organized on them; (3) how to handle, store, and ship them; (4) and how to erase them.

FLEXIBLE MAGNETIC RECORDING DISKETTES

Flexible diskettes, or *floppy disks* as they're more commonly called, are inexpensive, flexible, and portable magnetic storage mediums. They have the following characteristics.

Floppy Disk Construction

Floppy disks are made of round plastic disks coated with magnetic oxide particles. The disks are enclosed in a plastic jacket which protects the magnetic recording surface from damage.

Floppy disks come in three sizes: 8 inch, 5 1/4 inch, and 3 1/2 inch. Figure 8-1 shows each size. All disk sizes can either be *single-sided* or *double-sided.* Single-sided disks store data on only one side of the disk; double-sided disks store data on both sides.

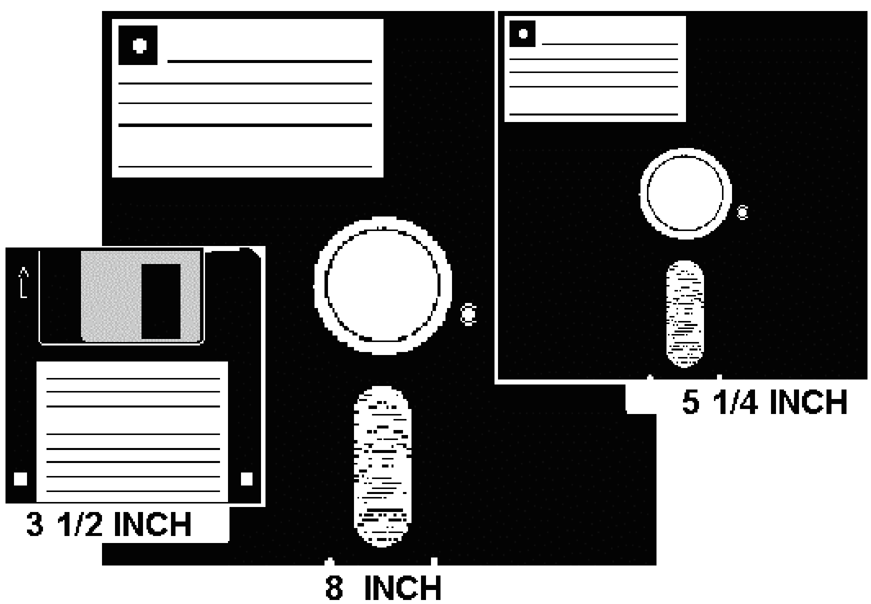


Figure 8-1.—Floppy disk construction.

When floppy disks are manufactured, the magnetic oxide coating is applied to both sides. Each disk is then checked for errors. Disks certified as single-sided, are checked on only one side; disks certified as double-sided are checked on both sides.

Floppy disks are also classified by how much data they can store. This is called a disk's *density*. There are three levels of floppy disk density: *single-density, double-density*, and *high-density*.

Some of the more common types of floppy disks and their storage capacity are listed below:

TYPE OF FLOPPY DISK	STORAGE CAPACITY
5-1/4" double-sided, double-density	360,000 bytes
5-1/4" double-sided, high-density	1,200,000 bytes
3-1/2" double-sided, double-density	720,000 bytes
3-1/2" double-sided, high-density	1,400,000 bytes

Floppy Disk Data Organization

Data is stored on a floppy disk in circular *tracks*. Figure 8-2 shows a circular track on a floppy disk. The total number of tracks on a floppy disk is permanently set by (1) the number of steps the disk drive's magnetic head stepper motor can make, and (2) whether the disk drive has a magnetic head for one or both surfaces of the floppy disk. These two things will also determine the type of floppy disk that's needed. Each type of disk is rated with a number that represents how many *tracks per inch (TPI)* it can hold. Some common track capacities are 40, 48, 80, and 96 TPI.

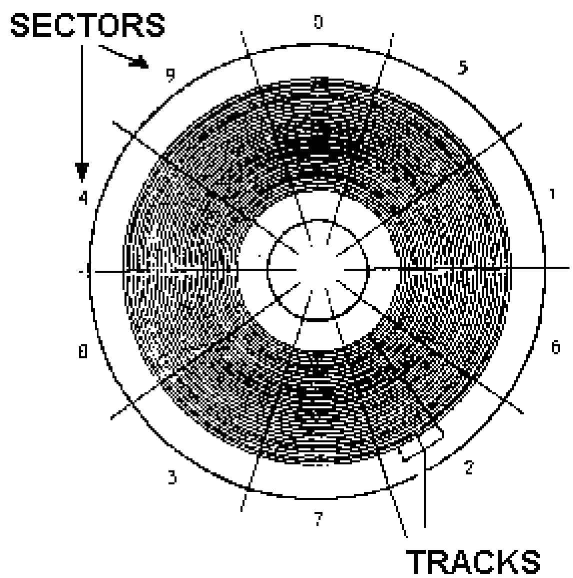


Figure 8-2.—Tracks and sectors of a magnetic disk.

Each track of a floppy disk is broken up into arcs called *sectors*. A disk is sectored just as you'd slice an apple pie. Figure 8-2 shows the sectors of a floppy disk. How many slices are made? That depends on who made the disk and in what host computer the disk is used.

There are two methods for *sectoring* a floppy disk:

1. <u>Hard Sectoring</u>: This method sectors the disk *physically.* The disk itself will have marks or sensor holes on it that the floppy disk drive hardware can detect. This method is seldom used today.

2. <u>Soft sectoring:</u> This method sectors the disk *logically*. The computer software determines the sector size and placement, and then *slices* the disk into sectors by writing codes on the disk. This

is called *formatting* or *initializing* a floppy disk. During formatting, if the computer software locates a bad spot on the disk, it locks it out to prevent the bad spot from being used. Soft sectoring is by far the most popular method of sectoring a floppy disk.

Once a floppy disk is *formatted*, the computer uses the disk's side number, a track number, and a sector number (together) as an *address*. It's this address that locates where on the disk the computer will store the data.

Floppy Disk Handling, Storage, and Shipping

Floppy disks hold a lot of data. Even disks with only a 360,000-byte storage capacity can hold 180 pages of data! That's why it's important to handle, store, and ship floppy disks properly. One hundred and eighty pages of data is a lot of data to retype just because of carelessness.

Before we get into disk handling and storage procedures, let's first learn about head-to-disk contact. Do you remember reading in chapter 2 that the quality of magnetic tape recording is seriously degraded when dust, dirt, or other contaminates get between the magnetic head and the tape? Well, the same is true for magnetic disk recording. In fact, head-to-disk contact is extremely important with floppy disks. This is because floppy disk drives, unlike magnetic tape drives, spin at very high speeds — 300 to 600 revolutions-per-minute (RPM). If anything gets between the head and the recording surface, you can lose data, or even worse, you can damage the magnetic head and the disk's recording surface. Figure 8-3 shows the size relationship between a disk drive's magnetic head, the disk recording surface, and some common contaminants.

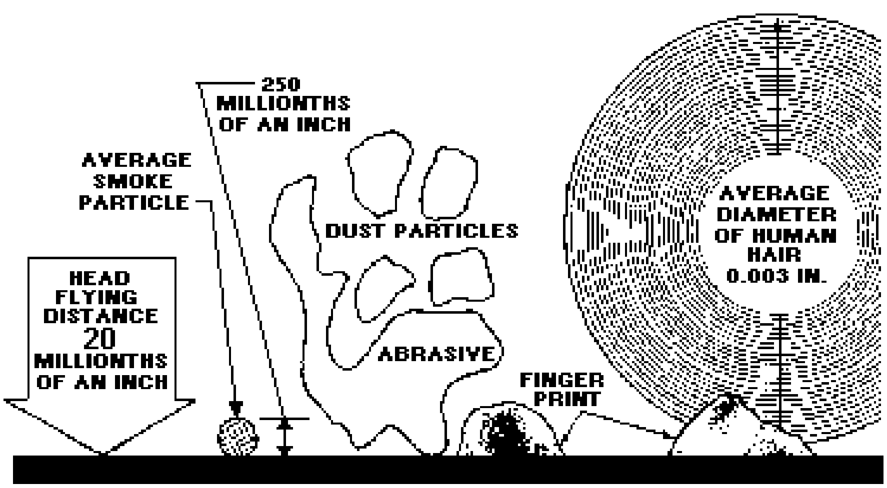


Figure 8-3.—Size relationship of distance between head and disk to contaminants.

You must handle, store, and ship floppy disks with great care if you want them to stay in good condition. Here's some specific precautions you should take:

- DO always store 8" and 5-1/4" floppy disks in their envelopes when not in use. Dirt, dust, etc., can get on the recording surface through the magnetic head read/write access hole if you leave it exposed for any length of time.

- DO always write on a floppy disk label first, and then place the label on the disk. NEVER write directly on a floppy disk. If you absolutely must write on a disk, use a felt-tip marker.

- DO hold floppy disks by their outside corners only. DO NOT bend them. And NEVER, NEVER paper clip them to anything, or anything to them.

- DO always store floppy disks in an upright position. Laying them on their side can cause them to warp.

- DO always keep floppy disks away from food, liquids, and cigarette smoke. All of these can easily damage floppy disks.

- DO always ship floppy disks in appropriate shipping containers. When shipping only a few disks, use the specially designed cardboard shipping envelopes. If you must ship a large number of disks, make sure the box you use is sturdy enough to protect the disks from damage. A good rule of thumb is to use a shipping box that allows you to place 2 inches of packing material around the disks.

- DO NOT touch any exposed recording surfaces. Something as simple as a fingerprint can destroy the data on a floppy disk.

- DO NOT expose a floppy disk to magnetic fields. Telephones, magnetic copy holders, printers, and other electronic equipment generate magnetic fields that can destroy the data on a floppy disk.

- DO NOT expose floppy disks to extreme heat or cold. Floppy disks will last longer if they're stored in an environment that stays around 70-80 degrees Fahrenheit and 30-60 percent relative humidity.

Floppy Disk Erasing

There are two ways to erase a floppy disk: (1) degauss it and then reformat it, or (2) just reformat it. The process for degaussing floppy disks is the same as for degaussing magnetic tape. Refer back to chapter 2 for the details on this.

If the floppy disks were used to store classified, or unclassified but *sensitive* information, they can't be de-classified by erasing them. This is because, with the right equipment and software, the data that was on the disk can be reconstructed. Floppy disks are cheap and easy to replace. If you can't re-use the floppy disks to store other classified data, just destroy them, using the procedures in OPNAVINST 5510.1, *DON Information and Personnel Security Program Regulation.*

Q1. Floppy disks are manufactured in what three sizes?

Q2. What type of floppy disk is made to store data on both sides of the disk?

Q3. What are the three levels of floppy disk density?

Q4. What is the storage capacity of a 5-1/4" double-sided, high-density floppy disk?

Q5. The floppy disks you are using have a rating of 96 TPI. What does this mean?

Q6. The process of formatting a floppy disk is called what type of sectoring?

Q7. What three components determine the address that locates where on a floppy disk the computer will store the data?

Q8. Why should you always store floppy disks in their envelopes?

Q9. Why should you never place floppy disks near telephones or other electronic equipments that generate magnetic fields?

Q10. What are the two ways to erase floppy disks?

FIXED MAGNETIC RECORDING DISKS

Fixed disks, or *hard disks* as they're more commonly called, are expensive, rigid, semi-portable, magnetic storage mediums. They have the following characteristics:

Hard Disk Construction

Most hard disks are made of aluminum platters coated on both sides with either *iron oxide* or *thin-film metal* magnetic coatings. The first type, iron oxide, is the most common (you can recognize this coating by its rust color). This is the same oxide coating that's used on magnetic tape. The second type of coating, thin-film metal, is the newer and better of the two. This coating is a microscopic layer of metal that's bonded to the aluminum platter. You can recognize it by its shiny silver color. Thin-film metal-coated hard disks are becoming more and more popular because they allow more data to be stored in less space.

Hard disks can hold a lot of data, the smallest disk being 10,000,000 bytes, and the largest being about 2,500,000,000 bytes (and they're working on larger ones).

Hard disk platters come in many sizes, ranging from 14" to 2". The most common sizes are 3-1/2", 5-1/4" and 14". The first two sizes are usually used with smaller personal computers. The 14" size is usually used with the larger *mini* and *mainframe* computers.

Most hard disk drives use more than one hard disk platter to store data. These are called *disk packs*. Some hard disk drives use removable hard disk platters. These can use just one platter, or they can use disk packs containing many platters. Most of the multi-platter *removable* hard disk drives in use today use 14" hard disk platters. Figure 8-4 shows a hard disk-pack.

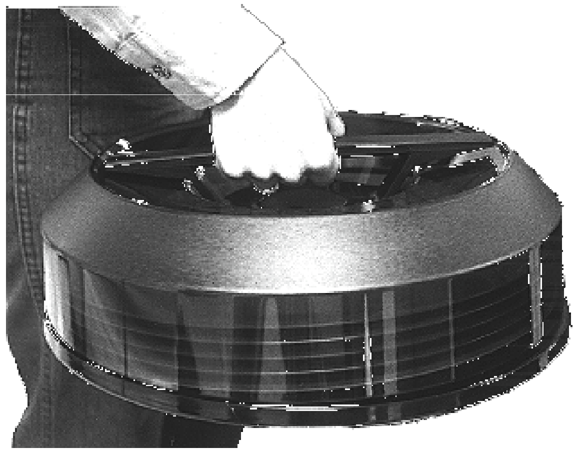


Figure 8-4.—Magnetic hard disk pack.

Hard Disk Data Organization

Data is stored on a hard disk the same way it's stored on a floppy disk, in circular tracks. The total number of tracks on a hard disk is set, just like floppy disk, by (1) the number of steps the disk drive's magnetic head stepper motor can make, and (2) whether the disk drive has a magnetic head for one or both surfaces of the hard disk platter.

A computer places data on a hard disk using one of two methods, either (1) the *cylinder* method, or (2) the *sector* method. The manufacturer of the hard disk drive decides which method to use.

THE CYLINDER METHOD.—This method uses a *cylinder* as the basic reference for placing data on a hard disk. Look at figure 8-5 view A. This is a picture of a disk pack containing six hard disk platters. Notice that this particular disk drive uses only 10 out of the 12 available recording surfaces. If you imagine that you're looking down through the disk pack from above, the tracks with the same number on each of the 10 recording surfaces will line up. Put together, these tracks make up a *cylinder*. Each of these 10 tracks with the same number, one on each recording surface, can be read from and written to by one of the disk drive's 10 read/write magnetic heads that are positioned by the five access arms.

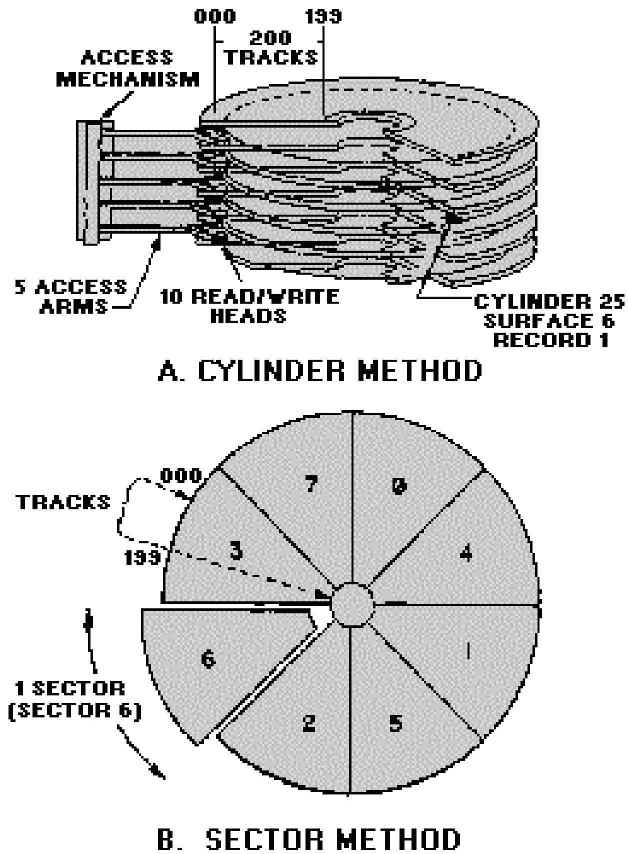


Figure 8-5.—Cylinder and sector method of organizing data on a hard disk pack.

So, to locate a place to store data using the cylinder method, a computer must specify the *cylinder number*, the *recording surface number*, and the *record number.* Figure 8-5 view A shows record number 1 stored on cylinder 25 of recording surface number 6. Special data is stored on each track to tell the computer where the start of a track is.

THE SECTOR METHOD.—Although we talked about this method earlier under the heading "Floppy Disk Data Organization," we need to repeat it here as it also applies to hard disks.

The sector method of organizing data on a hard disk is actually a variation of the *cylinder* method. As you already know, the sector method slices up a hard disk into *pie-shaped* slices (just like floppy disks). The total number of slices is set by the hard disk drive manufacturer. Figure 8-5 view B shows an example of the sector method.

Unlike a floppy disk drive, which locates a place on the disk using the *surface number, track number*, and *sector number,* a hard disk drive locates a place on the disk by using the *surface number, cylinder number,* and *sector number*. This is true even if the hard disk has only one platter. That's because both surfaces of that one platter still form a cylinder.

Hard Disk Handling, Storage, and Shipping

Hard disks hold a lot more data than floppy disks; even the lowest capacity hard disk can hold 5,000 pages of data! That's why it's important to handle, store, and ship hard disks properly. If you think 180 pages of data is a lot to retype, just think of retyping 5,000 pages!

Hard disk drives spin at a very high speed of about 3600 RPM. It is extremely important that nothing gets between the head and the recording surface. If it does, you can lose data and you can damage both the magnetic head and the disk's recording surface.

Most hard disk failures involve a *head-crash*. It's the worst thing that can happen to a hard disk. A head-crash is the result of the disk drive's magnetic heads *crashing* into the recording surface and grinding into the hard disk platter. Figure 8-6 shows a good hard disk platter and a bad hard disk platter that was the victim of a head-crash.

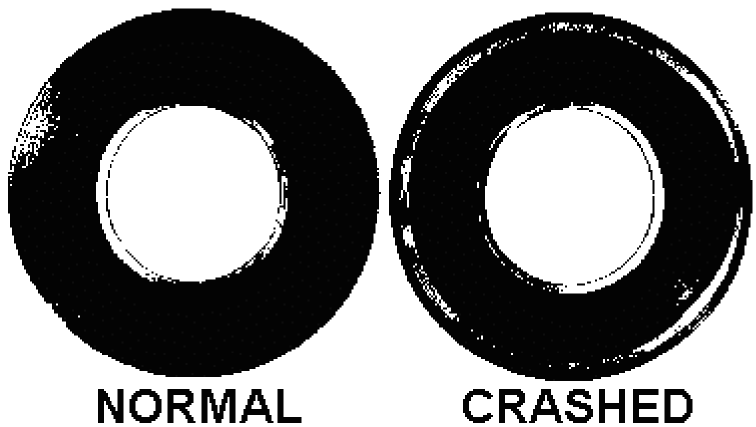


Figure 8-6.—Example of a hard disk crash.

You must handle, store, and ship hard disks with extreme care if you want them to stay in good condition. Here are some specific precautions you should take:

- DO always store removable hard disks in their storage cases when not in use. Dirt, dust, etc., can get on the recording surface through the magnetic head read/write access hole if you leave it exposed for any length of time.
- DO always handle hard disks with extreme care. DO NOT drop them. Even a small drop of 2" can warp a hard disk platter enough to cause a head crash.
- DO always keep removable hard disks away from food, liquids, and cigarette smoke. All of these can easily cause damage.
- DO always ship hard disks in their proper shipping containers. If you don't have the original shipping container, make sure the shipping box is sturdy and big enough to allow 2" of packing material around the disk. Save the original packing material for the hard disk just in case you need to ship it somewhere.
- DO NOT touch any exposed recording surfaces. Something as simple as a fingerprint can cause a head crash and destroy a hard disk platter.
- DO NOT expose hard disks to extreme heat or cold. Hard disks will last longer if they're stored in an environment that stays around 70-80 degrees Fahrenheit and 30-60 percent relative humidity.

Hard Disk Erasing

There are two ways to erase a hard disk: (1) degauss it and then reformat it, or (2) just reformat it. As you might guess, the first method can only be used for removable hard disk platters. The second method (reformatting) is the most common. If you must degauss a removable hard disk, the process is the same as degaussing magnetic tape. Refer back to chapter 2 for the details on this.

If the hard disks were used to store classified information or unclassified but *sensitive* information, you can't de-classify the hard disks by erasing them. This is because with the right equipment and software, the data that was on the disk can be reconstructed. If you can't re-use the hard disks to store other classified data, you must sanitize or destroy them, using the procedures in OPNAVINST 5510.1.

Q11. What are the three most common sizes of hard disk platters?

Q12. Computers use what two methods to place data on a hard disk?

Q13. Which method for placing data on hard disks divides a hard disk into pie shaped slices?

Q14. When computers use the cylinder method to store data on a hard disk pack, what three items make up the address that tells the computer where on a specific disk to store the data?

Q15. What is the most common type of hard disk failure?

Q16. Hard disks should be stored in an environment that stays within what relative humidity and temperature range?

Q17. What is the most common method for erasing a hard disk?

RECORDING DIGITAL DATA ON MAGNETIC DISKS

Digital data is stored on a magnetic disk using magnetic pulses. These pulses are generated by passing a frequency modulated (FM) current through the disk drive's magnetic head. This FM current generates a magnetic field that magnetizes the particles of the disk's recording surface directly under the magnetic head. The pulse can be one of two polarities, *positive* or *negative*.

Digital data isn't just recorded onto a magnetic disk as-is. Instead, it's encoded onto the disk. Three of the most popular encoding methods are (1) *frequency modulation (FM),* (2) *modified frequency modulation (MFM)*, and (3) *run length limited (RLL)*. The following paragraphs describe each of these encoding methods.

FREQUENCY MODULATION (FM) ENCODING

The FM method of encoding digital data onto a disk uses two pulse periods to represent each bit of data (a pulse period is the time span of one pulse). The first pulse period <u>always</u> contains a *clock* pulse. The second pulse-period may, or may not, contain a *data* pulse. If the digital data is a "1," a data pulse will be present in the second pulse-period. But, if the digital data is a "0," then there's no pulse present. Figure 8-7 shows this. The clock pulse, which is always present, tells the disk drive's interface that the next pulse is a data pulse. It is used to compensate for changes in the disk's rotation speed.

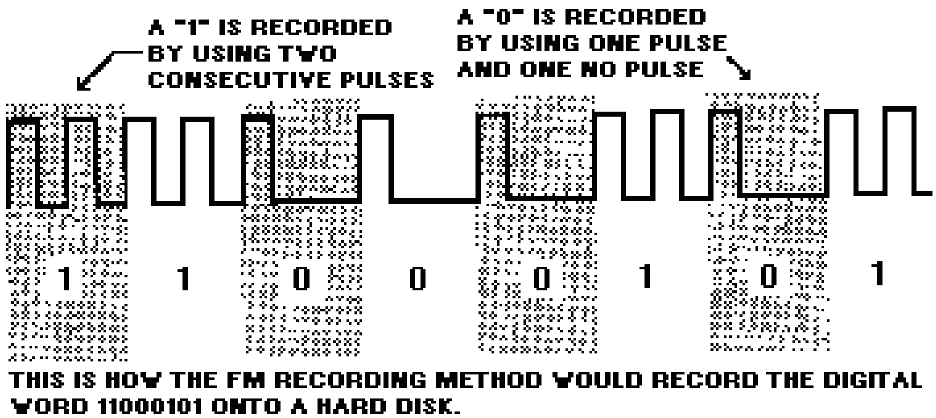


Figure 8-7.—Frequency-modulation (FM) encoding.

The FM method of encoding is old, and isn't used much anymore. You'll only see it in some of the older single-sided, single-density floppy disk drives, and in some of the older military hard disk drives.

MODIFIED FREQUENCY-MODULATION (MFM) ENCODING

The MFM method of encoding digital data onto a disk is more popular because it is more efficient and more reliable than straight FM encoding.

MFM encoding still uses two pulse periods, but uses a lot fewer pulses to store the digital data onto the disk. It does this in two ways:

1. It does away with the clock pulse that the FM method uses.

2. It stores a digital "1" by generating a *no-pulse* and a *pulse* in the two pulse periods. It stores a digital "0" as either a *pulse and a no-pulse* if the last bit was a "0," or as two no-pulses if the last bit was a "1." Figure 8-8 shows this.

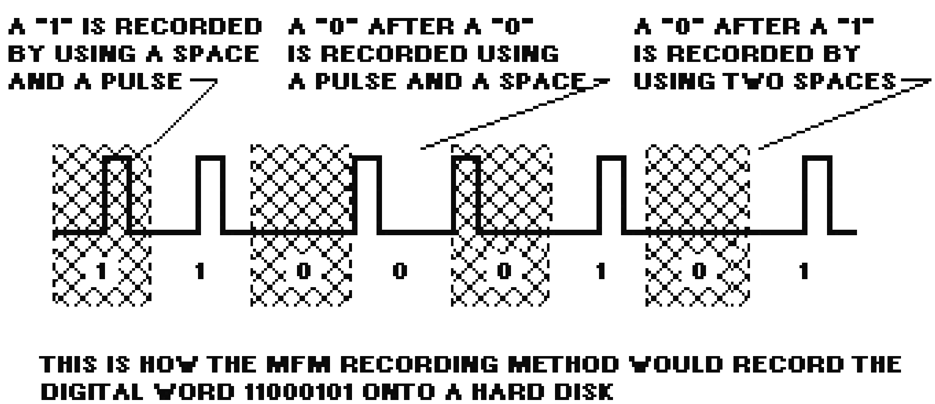


Figure 8-8.—Modified frequency-modulation (MFM) encoding.

RUN LENGTH-LIMITED (RLL) ENCODING

The RLL method of encoding digital data onto a disk is actually a refinement of the MFM encoding method. As its name implies, RLL *limits* the *run length* (distance) between pulses (also called *flux reversals*) on a hard disk. The basic theory of RLL encoding is that you can store more data in less space if you reduce the number of flux reversals (or pulses) that you must record.

There are several versions of the RLL encoding method, the most popular version being the 2,7 RLL. This means that no fewer than 2 *no-pulses* and no more than 7 *no-pulses* can occur between pulses.

MAGNETIC DISK DRIVE TRANSPORTS

Magnetic disk drive transports, like magnetic tape drive transports, move the magnetic disks across the magnetic heads and protect the disks from damage. The following paragraphs will (1) introduce you to the characteristics of both floppy and hard disk drive transports, and (2) describe their preventive maintenance requirements.

FLOPPY DISK DRIVE TRANSPORTS

Floppy disk drive transports contain the electromechanical parts that (1) rotate the floppy disk, (2) write data to it, and (3) read data from it. Figure 8-9 shows a typical floppy disk drive transport. Four of the drive transport's more important parts are the

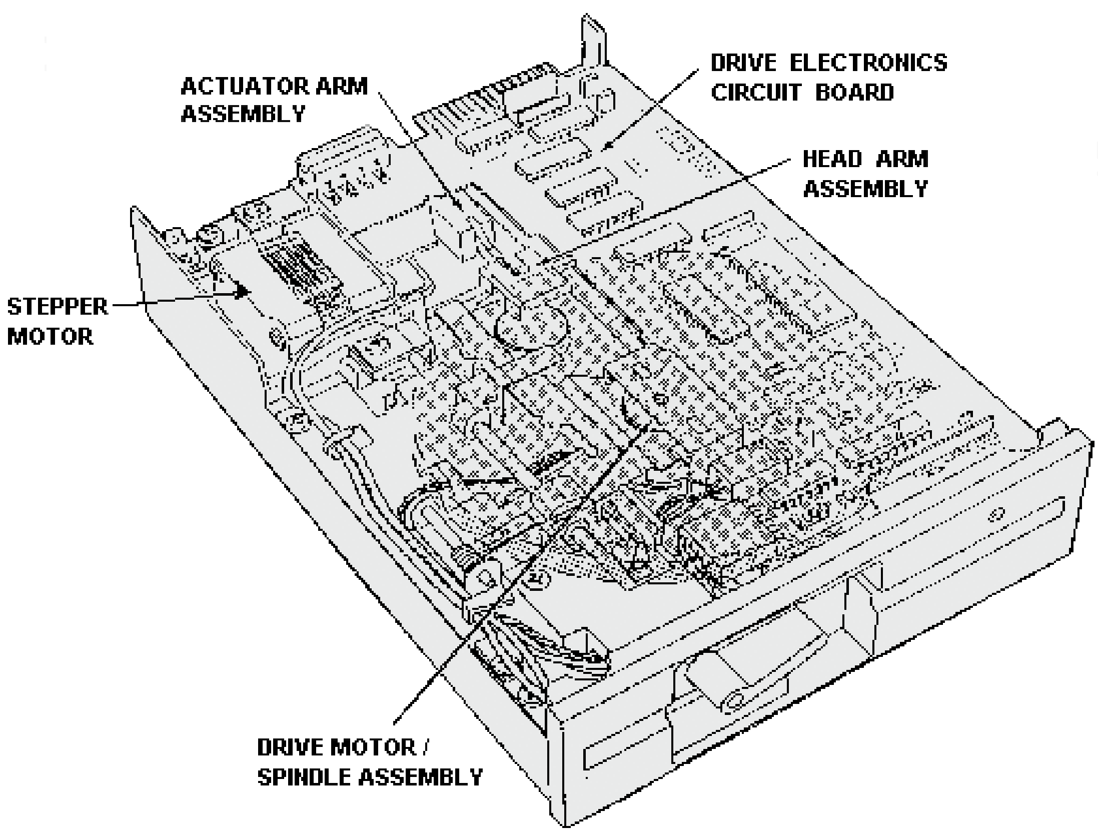


Figure 8-9.—Typical floppy disk drive transport.

1. drive motor/spindle assembly,
2. head arm assembly,
3. actuator arm assembly, and
4. drive electronics circuit board.

Drive Motor/Spindle Assembly

The spindle in this assembly holds the floppy disk in place while it spins. The drive motor spins the spindle at 300 to 600 RPM, depending on the type of floppy disk drive. The following is a list of the types of floppy disk drives and the spinning speeds of their spindles.

FLOPPY DISK DRIVE TYPE	SPINNING SPEED
5-1/4" 360-KB storage	300 RPM
5-1/4" 1.2-MB storage	360 RPM
3-1/2" 720-KB storage	600 RPM
3-1/2" 1.44-MB storage	600 RPM

The spindle of a 5-1/4" disk drive is activated and released by a small arm that's mounted on the front of the disk drive. You must turn the small arm to *lock* and *release* the floppy disk.

The spindle of a 3-1/2" disk drive is activated when the floppy disk is inserted into the disk drive. It's released by a push-button that's located on the front of the disk drive. When you push this button, the floppy disk is released and pops out of the disk drive.

Head Arm Assembly

This part of a floppy disk drive transport holds the magnetic read/write heads. There are four heads on a head arm assembly, two *write* heads and two *read* heads - one of each for each recording surface. The head arm assembly is attached to the actuator arm assembly.

Actuator Arm Assembly

The actuator arm assembly *positions* the magnetic heads over the recording surface of the floppy disk. It does this by using a special type of dc motor called a *stepper motor.* This motor, which can be moved in very small *steps*, allows the read/write heads to be moved from track to track as needed to write data onto and read data off of the floppy disk.

Drive Electronics Circuit Board

This circuit board contains the circuitry which (1) controls the electromechanical parts of the disk drive transport, (2) writes data to and reads data from the floppy disk, and (3) interfaces the floppy disk drive to the host computer.

HARD DISK DRIVE TRANSPORTS

Hard disk drive transports contain the electromechanical parts that (1) rotate the hard disk platter, (2) write data to it, and (3) read data from it.

There are two types of hard disk drive transports, *fixed disk* and *cartridge disk.* Fixed disk drive transports use *non-removable* hard disk platters. Cartridge-disk drive transports use *removable* hard disk platters that are built into protective cartridges. These two transports serve very different purposes, but they each contain the same basic parts. Figure 8-10 shows a typical hard disk drive transport. Four of the more important parts of a hard disk drive transport are the

1. drive motor/spindle assembly,

2. head arm assembly,

3. actuator arm assembly, and

4. drive electronics circuit board.

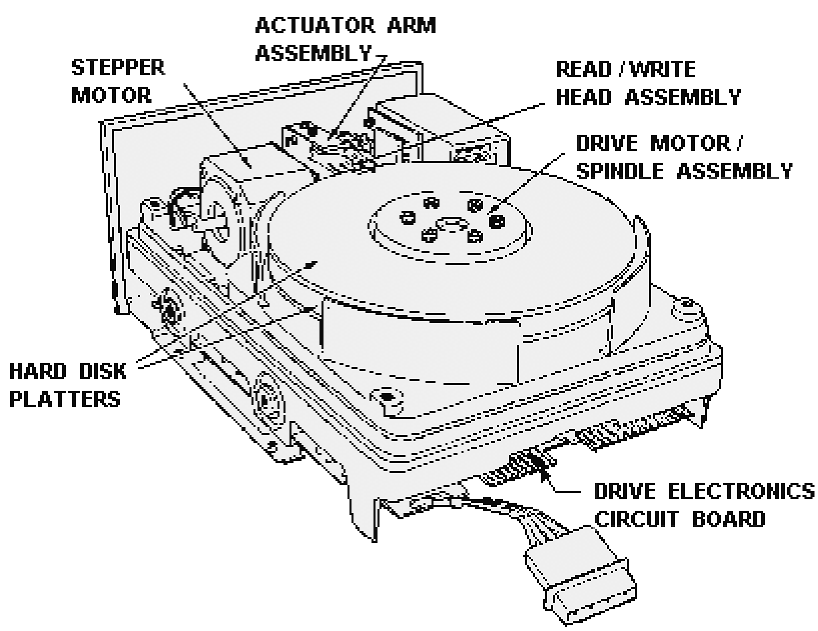


Figure 8-10.—Typical hard disk drive transport.

Drive Motor/Spindle Assembly

This assembly holds and spins the hard disk pack. The spindle assembly holds the hard disk pack in place and the drive motor spins the spindle at 3600 RPM. On cartridge disk drives, the spindle is *electronically* disengaged to release the disk pack so it can be removed.

Head Arm Assembly

This part of the hard disk drive transport holds the magnetic read/write heads. There is a separate head arm assembly for each of the hard disk platters in the disk pack. Each assembly has four magnetic heads, two *write* heads and two *read* heads-one pair of heads for each surface of the hard disk platter. The head arm assembly is attached to the actuator arm assembly.

Actuator Arm Assembly

This part of the hard disk drive transport *positions* the magnetic heads so they can write data to and read data from the correct track of the hard disk. It does this by using either a *stepper motor* or a *voice coil* servo. A stepper motor is a special type of dc motor which can be moved in very small *steps* to accurately position the magnetic heads. A voice coil servo by itself cannot move the magnetic heads from track to track. Instead, it must use special signals called *servo signals* to make sure it's positioning the heads where they should be. The *servo signals* are pre-recorded signals which are stored on either the same hard disk platter as the data or on a separate hard disk platter.

The voice coil servo method of moving the magnetic heads to the correct track of the hard disk is also called *embedded servo control.* This type of control is becoming very popular because voice coil actuator assemblies can position the magnetic heads much quicker and more accurately than dc stepper motors.

Drive Electronics Circuit Board

This circuit board contains the circuitry which (1) controls the electromechanical parts of the hard disk drive transport, (2) writes data to and reads data from the hard disk, and (3) interfaces the hard disk drive to the host computer.

Q18. What are the three most popular methods for encoding digital data onto magnetic disks?

Q19. Older, single-sided, single-density floppy disk drives would probably use what method for encoding digital data onto the floppy disk?

Q20. What method for encoding digital data enables you to store more data in less space by limiting the distance between pulses on a hard disk?

Q21. What are the four most important parts of a floppy disk drive transport?

Q22. The drive motor of a 3-1/2", 1.44-MB floppy disk drive spins the disk at what RPM?

Q23. The head arm assembly of a floppy disk drive transport has how many read heads and how many write heads?

Q24. What part of a floppy disk drive transport uses a dc stepper motor to position the magnetic heads over the recording surface of a floppy disk?

Q25. What part of a floppy disk drive transport contains the circuitry which controls the electromechanical parts of the transport?

Q26. Hard disk drive transports contain the electromechanical parts that perform what three functions?

Q27. In the actuator arm assembly of a hard disk drive transport, what device can position the magnetic heads to the correct track of a hard disk more accurately than a dc stepper motor?

MAGNETIC DISK DRIVE PREVENTIVE MAINTENANCE

Like magnetic tape recorders, if you want a magnetic disk drive to continue storing and retrieving data without errors, you must periodically perform preventive maintenance. Fortunately, disk drives require less maintenance than magnetic tape drives. The following paragraphs describe the preventive maintenance requirements for both *floppy* and *hard* disk drives.

FLOPPY DISK DRIVE PREVENTIVE MAINTENANCE

Of all of the magnetic disk drives in use today, floppy disk drives require the most maintenance. This is because they are not sealed units like most hard drives and because they use flimsy plastic disks that are coated with the same type of oxide as magnetic tape.

It's this oxide that causes most of the problems you'll have with floppy disk drives. Just as with magnetic tape, the oxide coating wears off of the plastic backing and sticks (mainly) to the magnetic heads. This contamination causes dropout errors which have much graver consequences than with magnetic tape. It can cause a program to crash, or even worse — it can destroy your valuable data.

To prevent this, you must *periodically* clean the floppy disk drive's magnetic heads. There are many kits available to do the job. A kit has a cleaning disk and a bottle of cleaning solution. A cleaning disk looks just like a regular disk, except that instead of an oxide-coated disk, it has a cloth or fiber cleaning disk inside the protective jacket. The instructions that come with the cleaning kit will lead you through the cleaning process. Here is an example of the cleaning procedures for a floppy disk drive's magnetic heads:

1. Pour some of the cleaning solution onto the cleaning disk through the access hole in the protective jacket.
2. Insert the cleaning disk into the disk drive.
3. Exercise the disk drive for at least 30 seconds.
4. Remove the disk from the disk drive.

There are also some cleaning kits that use disposable cleaning disks. These kits will instruct you to clean the heads as follows:

1. Open the scaled envelope that contains a cleaning disk soaked in cleaning solution.
2. Insert the cleaning disk into the protective jacket provided.
3. Insert the cleaning disk (with protective jacket) into the disk drive.
4. Exercise the disk drive for 30 seconds.
5. Remove the disk from the disk drive.
6. Remove the cleaning disk from the protective jacket and throw it away.

Now comes the question "How often must I clean the heads?" That's hard to say. It depends on the type of disk drive, the quality of the floppy disks you use, and how much you use the disk drive. On the average, you should clean a floppy disk drive

- once a month if it gets heavy use,
- once every 6 months if it gets moderate use, or

- once a year if it gets very little use.

HARD DISK DRIVE PREVENTIVE MAINTENANCE

Hard disk drives need little or no preventive maintenance. If it's a *fixed* hard disk drive, it doesn't need preventive maintenance because it's a sealed unit that you must not open for any reason. If it's a *cartridge* disk drive, the manufacturer will have a special cleaning disk with instructions for doing the preventive maintenance. The Navy uses some larger cartridge disk drives, such as the 14" disk pack drives, that require some other preventive maintenance. This could include the following:

- Cleaning air filters.
- Cleaning spindles, rails, and slides.
- Cleaning and buffing read and write heads.

The technical manual for the disk drive will guide you through this type of preventive maintenance.

Q28. Why do floppy disk drives require more preventive maintenance than hard disk drives?

Q29. A kit for cleaning floppy disk drives contains what two items?

Q30. Approximately how often should you clean a floppy disk drive that gets heavy use?

Q31. Cartridge hard disk drives with 14" disk packs may require what additional types of preventive maintenance?

MAGNETIC DISK DRIVE ELECTRONICS

Magnetic disk drive electronics consist of three main parts:

1. Control electronics to control the electromechanical parts of a disk drive.
2. Write/read electronics to write data to and read data from a disk drive.
3. Interface electronics to interface the disk drive to the host computer.

Some disk drives require a separate *controller card*. When this is true, some of the drive electronics are part of the disk drive itself, and some are part of the host computer's controller card.

As different as disk drives can be (floppy, fixed, cartridge, etc.), their electronics is surprisingly similar. That's why the following paragraphs will only *very basically* describe these three main parts.

CONTROL ELECTRONICS

The main functions of a disk drive's control electronics are to:

- Spin the disk at the proper speed.
- Move the magnetic heads across the recording surface.
- Tell write/read heads when to write data and when to read data.

WRITE/READ ELECTRONICS

The write/read electronics consists of the *write* part and the *read* part. The *write* part takes incoming data from the interface electronics, formats it as needed, and writes it onto the disk. The *read* part reads the data off of the disk, formats it as needed, and sends it to the interface electronics for output to the host computer. The write/read electronics also performs the initial disk formatting function.

INTERFACE ELECTRONICS

Interface electronics do two things:

1. Receive control signals from the host computer that tells them to spin the disk, move the magnetic heads, write/read data, format a disk, etc.

2. Convert the incoming and outgoing data as needed.

A disk drive is a *serial* device. This means the data stored on the disk is stored in a serial pulse-train format. But the data coming from the disk drive and going to the host computer needs to be in a *parallel* data format. The interface electronics converts the data from parallel to serial, and vice versa, as needed.

There are many types of disk drive interfaces in use today. The five most common ones are the:

1. Naval Tactical Data System (NTDS) interface.

2. ST-506/412 interface.

3. Enhanced small device interface (ESDI).

4. Small computer systems interface (SCSI).

5. Integrated drive electronics (IDE).

The following paragraphs describe each of these interfaces.

Naval Tactical Data System (NTDS) Interface

The NTDS interface is used by many naval electronic warfare systems. There are three versions of this interface:

1. NTDS FAST: A parallel interface that can transfer data at a rate of 250,000 32-bit words per second.

2. NTDS SLOW: A parallel interface that can transfer data at a rate of 41,667 32-bit words per second.

3. NTDS SERIAL: A serial interface that can transfer data at a rate of 10 million bits (Mbits) per second.

ST-506/412 Interface

The ST-506/412 interface was developed by Seagate Technology, Inc. It's often used in the hard disk drives installed in older IBM-compatible desktop computers that have a maximum capacity of 125 MB. It's also the interface used to control most floppy disk drives in use today.

This is one of the interfaces where most of the electronics is actually on a *controller card* mounted in the host computer. With this interface, the controller card does most of the work (moving the magnetic head, spinning the disk, etc.). The controller card also cleans any data coming from the disk drive by stripping off the formatting and control signals that were used to store the data onto the hard disk.

A hard disk drive is connected to the controller card in the host computer via two ribbon cables (a 34-pin control cable and a 20-pin data cable). Floppy drives use only the 34-pin control cable to transfer both data and control signals.

When this interface was originally developed in 1981, it's 5-Mbits per second transfer rate was considered *too fast*. They actually slowed it down by using a 6:1 interleave factor (we'll define this later) so it could operate with the computers being built at that time. With today's transfer rates pushing the envelope at 24 Mbits per second, you can see that it's now one of the slowest interfaces.

Enhanced Small Device Interface (ESDI)

The ESDI is an optimized version of the ST-506/412 interface. The main difference is that with ESDI, most of the disk drive's interface electronics is located in the disk drive itself, rather than on a controller card in the host computer. The result is a much faster transfer rate and more hard disk capacity. ESDIs have a transfer rate of up to 24 MB per second. And, they can handle disk drives with a maximum capacity of 1.2 GB (gigabytes).

The ESDI uses the same interface cables as the ST-506/412 interface, but that's where the similarity ends. With ESDI drives, only the *clean* data is sent to the controller card in the host computer. All formatting and control signals are stripped off at the hard disk drive.

Small Computer Systems Interface (SCSI)

The SCSI (pronounced *skuzzy)* is very different from both the ST-506/412 and the ESDI. The SCSI is an 8-bit, parallel, high-level interface. *High-level* means that instead of a host computer asking for data by specifying a track, cylinder, and sector number, all it asks for is a logical sector number. The SCSI then translates the logical sector number into the actual disk location.

The SCSI also has other improvements over the previous disk drive interfaces. For example, it can:

- Transfer data at rates of up to 4 MB per second.
- Handle hard disk drives of almost any size.
- *Disconnect* itself from a host computer's bus while it processes requests. This frees-up the host computer to do other things.
- *Daisy-chain* up to eight units off of one controller.

The SCSI interface uses one 50-pin ribbon cable to connect the hard disk drive(s) to the controller card mounted in the host computer. Some computer manufacturers include the SCSI electronics in their motherboards and do away with a separate controller card altogether. This interface got its big break when Apple Computer Corporation used the SCSI as its hard disk drive interface in its MacIntosh computers.

Integrated Drive Electronics (IDE)

The IDE is the newest interface available. It was developed as a result of trying, to find a cheaper way to build computer systems. It includes all of the controller card electronics in the hard disk drive itself, thus, the hard drive does all the work.

The hard disk drive connects to the host computer's bus with a 40-pin ribbon cable. The ribbon cable connects directly to either a 40-pin connector on the host computer's motherboard or a 40 pin connector on a small interface card that plugs into the host computer's motherboard. This interface offers a transfer rate of up to 1 MB and can handle hard drives with a maximum capacity of 300 MB.

MAGNETIC DISK RECORDING SPECIFICATIONS

Think back to the chapter 6 on "Magnetic Tape Recording Specifications." Do you remember how to measure and adjust them if needed? Well, magnetic disk recording specifications are a little different. They're set by the manufacturer and you can't change them. All you can do is measure them. The following paragraphs describe six of the most common specifications.

SEEK TIME

The *seek time* is the amount of time it takes for the magnetic head to position itself over a specific track of a magnetic disk. It's usually stated in milliseconds.

LATENCY PERIOD

The *latency period* is the amount of time it takes for a specific sector of a specific track to position itself under the magnetic head. It too, is usually stated in ms.

ACCESS TIME

The *access time* is the sum of the seek time and the latency period. It's the total amount of time in ms that it takes a disk drive to retrieve a sector of data from the magnetic disk. Access time is stated in one of the following three ways:

1. Track-to-track seek time: This is the amount of time it takes a disk drive to access data from a track next to the track it's presently over.

2. Average seek time: This is the amount of time it takes a disk drive to access data that's located one-third of the way across the magnetic disk.

3. Maximum seek time: This is the amount of time it takes a disk drive to access data from the last track of a magnetic disk when it's presently on the first track of the magnetic disk.

INTERLEAVE FACTOR

The *interleave factor* applies only to hard disk drives. They spin at 3600 RPM, a very fast speed compared to floppy disk drives which only spin at 300-600 RPM. Interleave indicates how many physical sectors are between sequentially numbered logical sectors on a hard disk. It's used when the magnetic heads and the control circuitry can't process the data fast enough to sequentially number the sectors on a hard disk platter. With interleave, the magnetic head is told to *skip* X number of sectors to get to the *next* one. For example, a hard disk with 17 sectors per track and no interleave is numbered 1, 2, 3, 4.... 17. The same hard disk with an interleave factor of 3 is numbered 1, 7, 13, 2, 8, 14, 3, 9, 15, 4, 10, 16, 5, 11, 17, 6, 12, and then back to 1. If you count every third sector, they're sequential. The most efficient hard disk drives have *no* interleave.

TRANSFER RATE

The *transfer rate* states how fast a disk drive and a disk drive controller (working together) can transfer data to the host computer. An example of a transfer rate specification is "2 Mbits/sec," or two million bits per second. The higher the number, the faster the data transfer rate.

RECORDING DENSITY

The *recording density* states how close together bits can be stored on the recording surface of a magnetic disk. It determines two things: (1) How close together the tracks on the disk will be, and (2) how close together the bits on each track will be. An example of a recording density specification is "12 Mbits/in^2," or 12 million bits per square inch.

Q32. The control electronics component of a floppy or hard disk drive performs what three main functions?

Q33. The write/read electronics of a disk drive performs what three functions?

Q34. The interface electronics of a disk drive performs what three functions?

Q35. What type of interface electronics is used in many naval electronic warfare systems?

Q36. What type of disk drive interface has most of the electronics on a controller card mounted in the host computer?

Q37. The SCSI is a high level disk drive interface. What does this mean?

Q38. What type of hard disk drive interface has all of the controller card electronics included in the disk drive itself?

SUMMARY

Now that you've finished chapter 8, you should be able to describe the (1) characteristics of floppy and hard disks, (2) methods for encoding digital data onto magnetic disks, (3) disk drive transports and their preventive maintenance requirements, (4) parts of a disk drive's electronics component, and (5) common types of disk drive interface electronics. The following is a summary of important points in this chapter:

FLOPPY DISKS are single-sided or double-sided plastic disks coated with oxide particles. The disks can be *single-density, double-density*, or *high density*.

Data is stored on floppy disks in **CIRCULAR TRACKS**. The tracks are divided into arcs called **SECTORS**.

HANDLE, SHIP, and **STORE** floppy disks carefully. Contaminates between the heads and the disk surface can cause serious damage.

FIXED (HARD) DISKS are aluminum platters coated on both sides with iron oxide or thin-film metal. Most hard disk drives use *disk packs* which are several disk platters stacked together. Some hard disk drives use removable disk platters.

Either the **CYLINDER** OR **SECTOR METHOD** is used to place data on hard disks.

HANDLE, STORE, AND **SHIP HARD DISKS** with extreme care. Contaminates on the heads or the disk surface can cause *head-crash*.

ERASE HARD AND FLOPPY DISKS by reformatting or degaussing.

Three popular **METHODS FOR ENCODING DIGITAL DATA ONTO DISKS** are frequency modulation, modified frequency modulation, and run length limited.

FLOPPY DISK DRIVE TRANSPORTS contain the parts that (1) spin the floppy disk, (2) write data to the disk, and (3) read data from it.

The **DRIVE MOTOR/SPINDLE ASSEMBLY** of a floppy disk drive transport holds and spins the floppy disk. The transport's **HEAD ARM ASSEMBLY** holds the read/write heads and its ACTUATOR ARM ASSEMBLY positions the heads over the disk's recording surface.

HARD DISK DRIVE TRANSPORTS contain the parts that (1) rotate the hard disk platter, (2) write data to the disk, and (3) read data from the disk.

The **DRIVE MOTOR/SPINDLE ASSEMBLY** of a hard disk drive transport holds the disk pack in place while the drive motor spins the spindle at 3600 RPM. The transport's HEAD ARM ASSEMBLY holds the read/write heads and its ACTUATOR ARM ASSEMBLY positions the heads over the correct track of the hard disk.

FLOPPY DISK DRIVES REQUIRE PREVENTIVE MAINTENANCE at regular intervals because they are not sealed units and the disks use an oxide coating that wears off and sticks to the heads and other parts.

HARD DISK DRIVES REQUIRE VERY LITTLE PREVENTIVE MAINTENANCE. Cartridge disk drives will have a special cleaning kit for doing the preventive maintenance.

Magnetic **DISK DRIVE ELECTRONICS** consist of (1) *control electronics* to control the electromechanical parts of a disk drive, (2) *write/read electronics* to write data to and read data from a disk drive, and (3) *interface electronics* to interface the disk drive to the host computer.

MAGNETIC DISK RECORDING SPECIFICATIONS are set by the manufacturer; all you can do is measure them. Six of the most common specifications are *seek time, latency period, access time, interleave factor, transfer rate,* and *recording density.*

ANSWERS TO QUESTIONS Q1. THROUGH Q38.

A1. 8 inch, 5 1/4 inches, 3 1/2 inches.

A2. Double-sided.

A3.

- *a.* *Single-density,*
- *b.* *double-density, and*
- *c.* *high-density.*

A4. 1,200,000 bytes or 1.2 megabytes.

A5. The disks can hold 96 tracks per inch.

A6. Soft sectoring.

A7.

- *a.* *disk side number,*
- *b.* *track number, and*
- *c.* *sector number.*

A8. Dust and other contaminates can get on the recording surface through the read/write hole.

A9. Magnetic fields can destroy the data on a disk.

A10.

- *a.* *Degauss the disk and then reformat it.*
- *b.* *Reformat the disk.*

A11. 3-1/2 inches, 5-1/4 inches, and 14 inches.

A12. (1) Cylinder method and (2) sector method.

A13. Sector method.

A14.

- *a.* *Cylinder number.*
- *b.* *Recording surface number.*
- *c.* *Record number.*

A15. Head-crash.

A16.

- *a.* *30-60 percent relative humidity.*
- *b.* *70-80 degrees Fahrenheit.*

A17. Reformat the disk.

A18. Frequency modulation (FM).

a. Modified frequency modulation (MFM).

b. Run length limited (RLL).

A19. Frequency-modulation encoding.

A20. Run length-limited (RLL) encoding.

A21.

a. Drive motor/spindle assembly.

b. Head arm assembly.

c. Actuator arm assembly.

d. Drive electronics circuit board.

A22. 600 RPM.

A23. Two read heads and two write heads.

A24. Actuator arm assembly.

A25. Drive electronics circuit board.

A26.

a. Rotates the hard disk platters.

b. Writes data to and reads data from the disk platters.

A27. Voice coil servo.

A28. They are not sealed units, and they use flimsy plastic disks with an oxide coating that wears off and sticks to the heads.

A29.

a. A cloth or fiber cleaning disk.

b. A bottle of cleaning solution.

A30. Once a month.

A31.

a. Cleaning with a special cleaning disk.

b. Cleaning air filters.

c. Cleaning spindles, rails, and slides.

d. Cleaning and buffing read/write heads.

A32.

a. Spins the disk at the correct speed.

b. Moves the heads across the recording surface.

c. Tells the write/read heads when to write data and when to read it.

A33.

a. Formats and writes incoming data from the interface electronics onto the disk.

b. Reads data off the disk, formats it, and sends it to the interface electronics for output.

c. Performs the initial disk formatting.

A34.

a. Receives control signals from the host computer and sends them to the control electronics or write/read electronics.

b. Receives data from the write/read electronics and outputs it to the host computer.

c. Converts incoming and outgoing data from parallel to serial, and vice versa, if needed.

A35. NTDS interface.

A36. ST-506/412 interface.

A37. The host computer asks for data by specifying a logical sector number. The SCSI translates the sector number into the actual disk location.

A38. Integrated drive electronics (IDE).

APPENDIX I

GLOSSARY

ABRASIVITY—The ability of the magnetic tape to wear the head.

ADDITIVE—Any material in the coating of magnetic tape other than the oxide and binder resins. (Examples: plasticizers (to soften an otherwise hard or brittle binder), lubricants (to lower the coefficient of friction of an otherwise high-friction binder), fungicides (to prevent fungus growth), dispersants (to uniformly distribute the oxide particles), and dyes.)

ALTERNATING CURRENT (ac) BIAS—(1) The alternating current, usually of a frequency several times higher than the highest input signal frequency, that is fed to a record head in addition to the input signal current. (2) Linearizes the recording process. (3) Is universally used in direct analog recording.

AMPLITUDE/FREQUENCY RESPONSE—(See frequency response.)

ANALOG RECORDING—A method of recording in which some characteristic of the record current, such as amplitude or frequency, is continuously varied in a manner similar to the variations of the original signal.

AZIMUTH ALIGNMENT—The alignment of the recording and reproducing gaps so their center lines lie parallel with one another. Misalignment of the gaps causes a loss in output at short wavelengths.

BACKINGS—(See base film.)

BANDWIDTH—The frequency within which the performance of a recorder with respect to some characteristic (usually frequency response) falls within specified limits, or within which some performance characteristic (such as noise) is measured.

BASE FILM—The plastic substrate material used in magnetic tape that supports the coating.

BER—(See bit error rate.)

BIAS-INDUCED NOISE (See noise.)

BINARY—(1) Two values ("O" or "1") or states (ON or OFF). (2) The number systems used in computers.

BINDER—A compound consisting of organic resins used to bond the oxide particles to the base material, the actual composition being considered proprietary information by each magnetic tape manufacturer. The binder is required to be flexible but still maintain the ability to resist flaking or shedding during extended wear passes.

BIT—(1) The acronym binary digit. (2) The smallest unit of data, either "O" or "1". (3) One recorded information cell, as applied in magnetic recording.

BIT DENSITY—(See packing density.) bit error rate (BER).—(1) The number of errors a specific magnetic tape may contain, as used in high-density recording. (2) Is expressed in errors per data bit, such as 1 in 10^6, or one error in one million data bits.

BREAK ELONGATION—The relative elongation of a specimen of magnetic tape of base film at the instant of breaking when it has been stretched at a given rate.

BROWN STAIN—(1) A thin discoloration of the head's top surface, usually a chemical reaction between the head's surface materials and the tape binder, the tape lubricant, or the head's bonding materials. (2) Its origin is not well understood, but is known to occur in low humidity.

BUCKLING—(1) A deformation of the circular form of a tape pack. (2) Caused perhaps by a combination of improper winding tension, adverse storage conditions, and/or poor reel hub configuration.

BUILD-UP—(1) A snowballing effect started by debris and tape magnetic particles embedded in the contamination. (2) The thickness of this build-up can cause an intense in head-to-tape separation, as well as an increase in the coefficient of friction. (3) Solvent cleaning of the head's top surface will usually remove the build-up.

BULK-ERASED NOISE—See noise.

BULK ERASER (DEGAUSSER)—An equipment for erasing a full reel of previously recorded signals on tape.

BYTE—A group of bits (next to each other) that are considered a unit (example, an 8-bit byte).

CERTIFIED TAPE—A tape that is electrically tested on a specified number of tracks and is certified by the supplier to have less than a certain total number of permanent errors.

CERTIFIER—(1) An equipment that tests the ability of magnetic tape to record and reproduce. (2) Counts and charts each error on the tape, including the level and duration of dropouts. (3) In the certify mode, stops the tape at an error to allow for visual inspection of the tape to see if the cause of the error is correctable or permanent.

CHICKEN TRACKS—(1) A line of small craters in the head's top surface running in the direction of tape motion. (2) Usually caused by a loose, small, hard particle moving with the tape over the head.

CINCHING—(1) The tape folds resulting from longitudinal slippage between the layers of tape in a tape pack. (2) Caused by uneven tension when the roll is accelerated or decelerated.

CLEAN ROOMS—The rooms of which their cleanliness is measured by the number of particles of a given size per cubic foot of room volume. (Examples (1) A class 100,000 clean room may have no more than 100,000 particles 0.5 fm or larger per cubic foot, and so on for class 10,000 and class 100 rooms. (2) A class 10,000 room may have no more than 65 5-fm particles per cubic foot, while class 100,000 may have no more than 700 5-fm particles per cubit foot.)

CLEANER—(See winder/cleaner.)

COATING—The magnetic layer of a magnetic tape consisting of oxide particles held in a binder that is applied to the base film.

COATING RESISTANCE—(1) The electrical resistance of the coating measured between two parallel electrodes spaced a known distance apart along the length of the tape. (2) Called resistivity on specification sheets.

COATING THICKNESS—The thickness of the magnetic coating applied to the base film.

COATING-TO-BACKING ADHESION—(See anchorage.)

CONTAMINATION—(1) A thick, tacky (viscous) deposit on the head's top surface, which causes a large increase in the effective head-to-tape coefficient of friction. (2) May not be removable by solvent cleaning.

CORE MATERIAL—(1) hard core material.—(a) Hard metal laminations bonded together to form the core, with a typical thickness of 0.005 to 0.004 inch. (Hard metal wears much more slowly than soft laminations.) (b) Hard solid metal, such as alphenol or sendust. (Wear rates are much lower than those of soft metal laminations.) (2) soft core material.—(a) Soft metal laminations bonded together to form the core, with a typical thickness of 0.0005 to 0.004 inch. (b) Usually, a high nickel/iron alloy, such as Hy Mu 800. (c) These materials have a relatively poor wear rate.

CRACK—A narrow, deep break in the head's surface material.

CREEP—The time-dependent strain at a constant stress (tape deformation).

CROSSFEED—See crosstalk and write feedthrough.

CROSSPLAY—The ability to interchange recordings between recorders while maintaining a given level of performance.

CROSSTALK—(1) The magnetic coupling from one track to another in the tape's read/write head. (2) See also write feedthrough.

CUPPING—(1) The curvature of a magnetic tape pack in the lateral direction. (2) May be caused by differences between the coefficients of thermal or hygroscopic expansion of coating and base film.

db—See decibel.

dc NOISE—(See noise.)

DECIBEL (db)—A dimensionless unit for expressing the ratio of two powers or voltages or currents on a logarithmic scale. If A and B represent two voltages or currents, the ratio A/B corresponds to 20 log A/B decibels. One decibel represents a ratio of approximately 1.1 to 1 between A and B. Other values are as follows:

RATIO	DECIBEL
1	0
1.4	3
2	6
4	12
10	20
100	40
1000	60

DEFECT—(1) An imperfection in the tape leading to a variation in output or a dropout. (2) The most common defects are surface projections of oxide agglomerates, embedded foreign matter, and redeposited wear products.

DEGAUSSER—See bulk eraser.

DELAY MODULATION—(See modified frequency modulation.)

DIGITAL RECORDING—(1) A method of recording in which the information is first coded in a digital form. (2) Usually, a binary code is used, with recording taking place in two discrete values/polarities of residual flux.

DIRECT RECORDING—An analog recording that records and reproduces data in the electrical form of its source.

DISK—(1) A disk-drive storage device on which information is magnetically recorded and retrieved. (2) Can be either hard (rigid) or floppy (flexible).

DISK DRIVE—A storage device for recording and retrieving data on hard or floppy disks.

DISK PACK—A portable, interchangeable device that contains more than one hard-disk platter and is used in hard-disk drives.

DISPERSION—The distribution of the oxide particles within a tape's binder.

DISTORTION—(see harmonic distortion.).

DRAG—The fractional tension differential across the contact area caused when the tape contacts some element in the tape path (such as the head, tape guides, tape bearings, or column walls).

DROPOUT—(1) A temporary reduction in the output of a magnetic tape of more than a certain predetermined amount. (2) Expressed in terms of the percentage reduction or decibel loss.

DROPOUT COUNT—The number of dropouts detected in a given length of magnetic tape.

DURABILITY—The number of passes that can be made before a significant degradation of output occurs, divided by the corresponding number that can be made using a reference tape.

DYNAMIC—(See tape skew.)

DYNAMIC RANGE—(1) The bandwidth within which a satisfactory signal-to-noise ratio is obtained. (2) See also resolution.

DYNAMIC SKEW—The change in skew caused by tape motion.

DYNAMIC TAPE SKEW—(See tape skew.)

EQUIPMENT NOISE—(See noise.)

ERASURE—A process by which a signal recorded on a tape is removed and the tape made ready for rerecording. May be accomplished in two ways. (1) In ac erasure, the tape is demagnetized by an alternating magnetic field that is reduced in amplitude from an initially high value. May be accomplished by passing the tape over an erase head fed with high-frequency ac, or by placing the whole roll of tape in a decreasing ac field (bulk erasure). (2) In dc erasure, the tape is saturated by applying a primarily unidirectional field. May be accomplished by passing the tape over a head fed with dc or over a permanent magnet. Additional stages may be included in dc erasure to leave the tape in a more nearly unmagnetized condition.

FERRITE—A powdered and compressed ferric-oxide material that has both magnetic properties and light resistance to current flow.

FLOPPY DISK—(See disk.)

FLUTTER—(See wow.)

FLUX—A term, with reference to electrical or electromagnetic devices, that designates collectively all the electric or magnetic lines of force in a region.

FLUX DENSITY—The number of magnetic lines of force passing through a given area.

FM—(See frequency modulation.)

FREQUENCY MODULATION (FM)—(1) A flux reversal at the beginning of a cell time represents a clock bit. (2) A "1" bit is a flux reversal at the center of the cell time. (3) A "0" bit is an absence of a flux reversal.

FREQUENCY RESPONSE—(1) The variation of sensitivity with signal frequency. (2) The frequency response of a tape is usually given in decibels relative to a referenced frequency-output level. (3) Also called amplitude/frequency response.

GAMMA-FERRIC OXIDE—The common magnetic constituent of magnetic tapes in a dispersion of fine, needle-like particles within the coating.

GAP EROSION—(1) The read or write gap increased in length and retreated below the head surface. (2) Usually due to deterioration of core material at the edges of the gap.

GAP LOSS—(1) The loss in output due to the finite gap length of the reproduce head. (2) The loss increases as the wavelength decreases.

GAP WIDTH—The dimension of the gap of a magnetic head measured in the direction perpendicular to the direction of the tape path.

GAUSS—The metric unit of the magnetic flux density equal to 1 Mx/CM^2.

HARD DISK—(See disk.)

HARD FERRITE—A ferrite with a very low wear rate when compared with the soft metal laminations.

HARD METAL LAMINATIONS—(See core material, hard.)

HARD SOLID METAL—(see core material, hard.)

HARMONIC DISTORTION—A signal non-linearity with harmonics of the fundamental in the output when the input signal is sinusoidal.

HEAD CONTAMINATION—(See tape-to-head separation.)

HEAD CONTOUR—(1) The complex shape of the contacting surface of a head as a result of manufacture, head lapping, or wear. (2) The contour of a head is always changing throughout the head's life and, in many cases, is responsible for retiring the head.

HEAD CRASH—A term used for the damage to a hard disk caused by the physical contact made between the magnetic read/write heads and the surface of the hard-disk platter.

HEAD STICK—(1) A common word for a large increase in head-to-tape friction caused by (a) a stick by-product exuded by conditions due to tape age, temperature/humidity, and head-to-tape pressure, and (b) very smooth tapes coupled with large area heads. (2) See also sticktion and stick-slip.

HEAD-TO-TAPE CONTACT—The degree that a tape's magnetic coating approaches the surface of the record or reproduce head during normal operation.

INTERLAYER TRANSFER—Any loose material, such as oxide, generated by tape wear or a head-stick condition which is transferred from the oxide to the back of the tape, or from the back side to the oxide when the tape is wound on a reel.

INTERMODULATION DISTORTION—(1) A signal non-linearity with frequencies in the output equal to sums and differences of integral multiples of the component frequencies present in the input signal. (2) Harmonies are usually not included as part of the intermodulation distortion.

INTERSYMBOL INTERFERENCE—(1) An interference resulting in a phase shift of the cell playback crossover point with respect to the data clock. (2) When a recording system has limited record resolution, a flux transition being recorded will extend beyond its cell boundaries, adding or subtracting from the flux in the adjacent bit cells of symbols.

IRON OXIDE—(See gamma-ferric oxide.)

LAMINATIONS—(see core material.)

LATERAL DIRECTION—The direction across the width of the tape.

LAYER-TO-LAYER ADHESION—The tendency for adjacent layers of tape in a roll to adhere to one another.

LAYER-TO-LAYER TRANSFER—The magnetization of a layer of tape in a roll by the field from a nearby recorded layer, sometimes referred to as print through.

LBE—(See lower band edge.)

LINEARITY—The extent to which the magnitude of the reproduced output is directly proportional to the magnitude of the signal applied to the input of the recorder.

LONG-TERM TAPE SPEED—(See tape speed.)

LONGITUDINAL CURVATURE—Any deviation from the straightness of a length of tape.

LOOSE DEBRIS—Any material that is very lightly bonded to the tape or the head's top surface, removable by tape motion.

LOWER BAND EDGE (LBE)—The lower band edge of the recorder/reproducer response (usually at the −3-dB point).

LUBRICANT—(See additive.)

MAGNETIC INSTABILITY—(1) The property of a magnetic material that causes variations in the residual flux density of a tape to occur with temperature, time, and/or mechanical flexing. (2) A function of particle size, magnetizing field strength, and anisotropy.

MAGNETIC MEDIA—A base film, coated with magnetic particles held in a binder. (The magnetic particles are usually needle-like, single-domain, gamma-ferric oxide.)

MAGNETIZING FIELD STRENGTH—The instantaneous strength of the magnetic field applied to a sample of magnetic material.

MFM—(See modified frequency modulation.)

MODIFIED FREQUENCY MODULATION (MFM)—(1) A code that has a "1" and a "O" corresponding to the respective presence or absence of a transition in the center of the corresponding bit cell. (2) Additional transitions at the cell boundaries occur only between bit cells that contain consecutive "0" values. (3) Also called delay modulation.

MODULATION NOISE—(See noise.)

NOISE—Any unwanted electrical disturbances other than crosstalk or distortion components that occur at the output of the reproduce amplifier. (1) System noise is the total noise produced by the whole recording system, including the tape. (2) Equipment noise is produced by all the components of the system, with the exception of the tape. (3) Tape noise can be specifically ascribed to the tape. The following are typical sources of tape noise: (a) Bulk-erased noise arises when a bulk-erased tape with the erase and record heads completely deenergized is reproduced. (b) Zero-modulation noise arises when an erased tape with the erase and record heads energized as they would be in normal operation, but with zero input signal, is reproduced. This noise is usually 3 to 4 dB higher than the bulk-erased noise. The difference between bulk-erased and zero-modulation noise is sometimes termed bias-induced noise. (c) Saturation noise arises when a uniformly saturated tape is reproduced. This is often some 15 dB higher than the bulk-erased noise and is associated with imperfect particle dispersion. (d) Dc noise arises when a tape that has been non-uniformly magnetized by energizing the record head with dc, either in the presence or the absence of bias, is reproduced. This noise has pronounced, long, wavelength components that can be as much as 20 dB higher than those obtained from a bulk-erased tape. At very high values of dc, the dc noise approaches the saturation noise. Dc noise is actually the low-frequency component of modulation noise. (e) Modulation noise is essentially a modulation of the desired signal by noise that is caused by non-uniform dispersion of elementary magnetic particles in the tape’s coating material. This noise, which occurs only when a recorded tape is reproduced, increases with the intensity of the reproduced signal. Dc noise is actually the low-frequency component of modulation noise.

NOISE PULSE—(1) A short-duration false signal occurring during the reproduction of a tape. (2) A signal that is of a magnitude considerably in excess of the average peak value of the ordinary system noise.

NOMINAL BIT TIME—(1) The average bit time of recording at continuous maximum-flux reversals. (2) Also called cell time.

NON-RETURN-TO-ZERO (NRZ0)—The flux reversal for a "0"; no flux reversal for a "1."

NON-RETURN-TO-ZERO INVERTED (NRZI)—The flux reversal for a "1"; no flux reversal for a "0."

NON-RETURN-TO-ZERO (NRZ) RECORDING—(See digital recording.)

OERSTED—A unit of magnetic field strength.

OUTPUT—The magnitude of the reproduced signal voltage, usually measured at the output of the reproduced amplifier.

OXIDE BUILD-UP—The accumulation of oxide or wear products as deposits on the surface of heads and guides.

OXIDE LOADING—(1) A measure of the density with which oxide is packed into a coating. (2) The weight of the oxide per unit volume of the coating.

OXIDE SHED—The loosening of particles of oxide from the tape coating during use.

PACKING DENSITY—The amount of digital information recorded along the length of a tape, measured in bits per inch.

PARTICLE ORIENTATION—The rotation of needle-like particles so that their longest dimensions tend to lie parallel to one another.

PARTICLE SHAPE—The needle-like particles of gamma-ferric oxide used in conventional magnetic tape, with a dimensional ratio of about 6 to 1.

PARTICLE SIZE—The physical dimensions of magnetic particles used in a magnetic tape.

PERMANENT ELONGATION—The percentage of elongation remaining on a tape or a length of base film after a given load applied for a given time has been removed.

PLASTICIZER—(See additive.)

POLYESER—(1) An acronym for polyethylene glycol terephthalate. (2) The material most commonly used as a base film for precision magnetic tape.

READ/WRITE ERASE HEAD—A three-gap head (read, write, erase) on one body. Sometimes the erase head is bolted to the read/write head.

READ/WRITE HEAD—A two-gap head (read, write) on one body.

REEL—The metal-, glass-, or plastic-flanged hub on which magnetic tape is wound.

REFERENCE TAPE—A tape used as a reference against which the performances of other tapes are compared.

REMANENCE—(1) The magnetic flux density that remains in a magnetic circuit after removal of applied magnetomotive force. (2) Is not necessarily equal to residual flux density.

REMOVABLE FIXED DISK—A hard-disk device (usually) with only one hard-disk platter that's used to record and retrieve data.

RESISTIVITY—(See coating resistance.)

RESOLUTION (DYNAMIC RANGE)—The average peak-to-peak signal amplitude at the maximum flux reversal divided by the average peak-to-peak signal amplitude at the minimum flux reversal at the desired recording method.

RETENTIVITY—The maximum value of residual flux density corresponding to saturation flux density.

RZ RECORDING—(See digital recording.)

SATURATION FLUX DENSITY—(1) The maximum intrinsic flux density possible in a sample of magnetic material. (2) The intrinsic flux density asymptotically approaches the saturation flux density as the magnetizing field strength increases.

SATURATION NOISE—(See noise.)

SCRATCH—A long, narrow, straight defect in the top surface of a head track or a tape.

SENSITIVITY—The magnitude of the output when reproducing a tape recorded with a signal of given magnitude and frequency.

SEPARATION LOSS—The loss in output that occurs when the surface of the coating of a magnetic tape fails to make perfect contact with the surface of the record or reproduce head.

SHEDDING—The loss of oxide or other particles from the coating or backing of a tape, usually causing contamination of the tape transport and, by redeposit, of the tape itself.

SHORT-TERM TAPE SPEED—(See tape speed.)

SIGNAL-TO-NOISE RATIO—(1) The ratio of the power output of a given signal to the noise power in a given bandwidth. (2) Is usually measured by the corresponding root mean square signal and noise voltages appearing across a constant output resistance.

SKEW—A deviation of a line connecting the average displacement of the read or write track gaps from a line perpendicular to the reference edge of the tape in the direction of tape motion.

SKEW TAPE—(1) The continuous strings of "1" values written on a properly adjusted tape drive for the entire recoverable length of the tape. (2) An "all '1' pattern" on all tracks. (3) The write head, the write delays, and the tape drive adjusted to write with minimum physical skew and gap scatter.

SOFT METAL LAMINATIONS—(See core material, soft.)

SPOKING—A buckling in which the tape pack is deformed into a shape that approximates a polygon.

SPOOL—(See reel.)

SQUEAL—(See stick-slip.)

STANDARD REFERENCE TAPE—A tape intended for daily calibration, the performance of which has been calibrated to the amplitude reference tape.

STATIC—(See tape skew.)

STICK-SLIP—(1) A low-speed phenomenon. (2) A relationship between tension, temperature, humidity, wrap angle, head material, tape binder, and elastic properties. (3) When detected audibly, it is a squeal.

STICKTION—The tape's adhering to transport components, such as heads or guides.

STIFFNESS—(1) The resistance to bending the tape. (2) A function of tape thickness. (3) A modulus of elasticity.

SURFACE TREATMENT—Any process by which the surface smoothness of the tape coating is improved after it has been applied to the base film.

SYSTEM NOISE—(See noise.)

TAPE CONTAMINATION—(See tape-to-head separation.)

TAPE NOISE—(See noise.)

TAPE PACK—The form taken by the tape wound on a reel.

TAPE SKEW—(1) The tape's deviation from following a linear path when transported across the heads. (2) The terms static and dynamic distinguish between the physically fixed components and the fluctuating components of total tape skew.

TAPE SPEED—(1) The speed at which the tape is transported across the read/write head during normal recording or reproduction. (2) Long-term speed is averaged over a minimum of 15 inches of tape (in inches per second). (3) Short-term speed is the instantaneous (dynamic) tape speed (in inches per second).

TAPE-TO-HEAD SEPARATION—(1) Separation.—The separation between a magnetic head and the magnetic tape caused by the (a) foil-bearing effect, (b) improper head contour, which generates standing waves in the tape, and (c) surface roughness of the tape. These conditions are interrelated and are greatly influenced by tape tension and tape compliancy. In a properly designed system, tape roughness is the limit of head-to-tape separation, usually <10 fm. (2) Changes.—(a) Head contamination is the debris attached to the head, which causes the tape to lift away from the head, forming a tent-like deformation of the tape. This tent does not move or change shape until the contamination is removed. (b) Tape contamination includes particles attached to the tape, resulting in a tent that moves across the head with the tape. (3) Effective.—(a) The actual distance from the magnetic storage material on the tape to the top of the active magnetic core material at the read or write gape. (b) The effective head-to-tape separation is usually somewhat larger than the mechanical head-to-tape separation.

TAPE TRANSPORT—(1) The mechanism that extracts magnetic tape from a storage device, moves it across magnetic heads at a controlled speed and then feeds it into another storage device. (2) Typical storage devices are tape loops, bins, reels, cassettes, and cartridges. (3) The tape transport is the part of a magnetic tape recorder/reproducer system that normally consists of magnetic heads, magnetic tape, transport, record electronics, and reproduce electronics.

TEAR STRENGTH—The force required to initiate and/or propagate a tear in a specially shaped specimen of tape or base film.

TOTAL THICKNESS—The sum of the thicknesses of the base film and the magnetic coating, as well as back coating, when applied. The total thickness governs the length of tape that can be wound on a given reel.

TRACK—An area of the tape's surface that coincides with the location of the recorded magnetization produced by one record gap.

TRACK WIDTH—The width of the track corresponding to a given record gap.

UBE—(See upper band edge.)

UNDERCUT—(See washout.)

UNIFORMITY—The extent to which the output remains free from variations in amplitude. Uniformity is usually specified as the positive and negative deviations from the average output within a roll of tape.

UPPER BAND EDGE (UBE)—The upper band edge of the recorder/reproducer response (usually at the −3dB point).

VOID—An area where material is missing on the surface of a tape or on a head's track.

WASHOUT—(1) A hard-coated head whose magnetic core material has a much higher wear rate than the coating. (2) The radius of curvature of the core material will be larger than the surrounding coating of the softer material, which could even be undercut, possibly causing an increase in the head-to-tape separation. (3) Also called undercut.

WAVELENGTH—The distance along the length of a sinusoidally recorded tape corresponding to one cycle.

WAVINESS—(1) A non-flat head's top surface perpendicular to the tape motion due to different wear rates in top surface materials. The harder material will be up. The head core is usually the harder material; therefore, there will be increased head-to-tape contact pressure at the cost of tape life. (2) Can occur during break-in or field use.

WEAR ABILITY—(See durability.)

WEAR PRODUCT—Any material that is detached from the tape during use.

WEAR TEST—(See durability.)

WIND—The way in which tape is wound onto a reel. An A-wind is one in which the tape is wound so that the coated surface faces toward the hub.

WINDER/CLEANER—A device that winds and cleans magnetic tape to restore it to a near-new condition, providing the tape has not been physically damaged.

WOW AND FLUTTER—(1) The changes in signal-output frequency caused by tape-speed variations occurring at relatively low and relative high rates, respectively. (2) Wow is no longer used, but is incorporated into the flutter measurement.

WRITE FEEDTHROUGH—(1) The magnetic coupling from the write track to a read track in the read/write head. (2) Also called crossfeed and crosstalk.

ZERO-MODULATION NOISE—(See noise.)

APPENDIX II

REFERENCES

Magnetic Tape Recording Technical Fundamentals, Datatape Incorporated, 360 Sierra Madre Villa, Pasedena, CA, 1986.

Parallel Mode High Density Digital Recording Technical Fundamentals, Datatape Incorporated, 360 Sierra Madre Villa, Pasedena, CA, 1976.

Aviation ASW Technician 3&2, NAVEDTRA 10353-D, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1987.

Haynes, N. M., *Transistor Circuits for Magnetic Recording*, Howard W. Sams & Co., Indianapolis, IN, January 1964.

Magnetic Tape Recording for the Eighties, NASA Reference Publication 1075, Goddard Space Flight Center, Green Belt, MD, 1982.

Data Systems Technician 3&2, Vol. 2, NAVEDTRA 10231, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1987.

AN/TNH-21 *Recorder-Reproducer Set, Sound, Instruction Manual,* National Security Agency, Fort G. Meade, MA, December 1973.

NSA Technical Support Manual, Magnetic Tape Recording, National Security Agency, Washington DC, 1957.

System Description Manual for Recorder-Reproducer Set, Signal Data, AN/USH-23(V), NAVELEX 0967-LP-546-1010, Naval Electronics Systems Command, Washington, DC, 1975.

Operation and Maintenance Instructions for Model 100X Magnetic Tape Recorder, NAVELEX 0967-LP-551-3010, Naval Electronics Systems Command, Washington, DC, 1976.

Recorder-Reproducer Set, Magnetic Disk AN/UYH-3(V), Operation and Maintenance Instructions, NAVSEA SE600-AA-MMM-010/AN/UYH-3(V), Naval Sea Systems Command, Washington, DC, 1982.

Input/Output Interfaces, Standard Digital Data, Navy Systems, MIL-STD-1397A(NAVY), Naval Sea Systems Command, Washington, DC, 1983.

Technical Manual, Operation Instructions, Maintenance Instructions, Parts List, Installation Instructions, Recorder-Reproducer Set, Signal Data, USH-26(V), NAVSEA SE640-AF-MMO-010/USH-26(V), Naval Sea Systems Command, Washington, DC, 1981.

Operating Instructions and Organizational Maintenance Instructions, Chap. 1, *Signal Data Recorder-Reproducer Set*, AN/USH-24(V), 521600 and 521600-0001, SPAWAR 0967-LP-590-0010, Space and Naval Warfare Systems Command, Washington, DC, 1979.

Data Processing Technician Third Class, NAVEDTRA 10263, Naval Education and Training Professional Development and Technology Center, Pensacola, FL, 1987.

MODULE 23 INDEX

W

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, "Introduction to Magnetic Recording," pages 1-1 through 1-7. Chapter 2, "Magnetic Tape," pages 2-1 through 2-158. Chapter 3, "Magnetic Tape Recorder Heads," pages 3-1 through 3-9. Chapter 4 "Magnetic Tape Recorder Transports," pages 4-1 through 4-17.

1-1. In what year did Oberlin Smith originate the idea of using permanent magnetic impressions to record sound?

1. 1880
2. 1888
3. 1900
4. 1908

1-2. In 1925, magnetic recording began receiving attention when what device was invented?

1. Magnetic tape
2. Magnetic disk
3. Electronic amplifiers
4. Video recorders

1-3. In 1907, Mr. Poulsen discovered dc bias. How did adding dc bias to the input signal solve the distortion problem for magnetic recording?

1. It moved the input signal away from the step in the magnetism curve
2. It moved the input signal directly onto the step in the magnetism curve
3. It moved the output signal directly over the step in the magnetism curve
4. It straightened the step in the magnetism curve

1-4. After 1925, the Naval Research Laboratory greatly improved the signal-to-noise ratio of magnetic recording by adding what component?

1. Ac bias
2. Electronic amplifiers
3. Transistors
4. Dc motors

1-5. In what year was plastic based tape introduced?

1. 1900
2. 1925
3. 1939
4. 1947

1-6. In 1956, IBM introduced what major contribution to magnetic recording?

1. The magnetic tape drive
2. The floppy disk drive
3. The hard disk drive
4. Each of the above

1-7. In 1966, IBM introduced the first removable-pack hard disk drive.

1. True
2. False

1-8. In what year was the 5 1/4" floppy disk drive invented?

1. 1956
2. 1976
3. 1980
4. 1988

1-9. Which of the following is a prerequisite for magnetic recording?

1. An input signal
2. A recording medium
3. A magnetic head
4. Each of the above

1-10. A magnetic recording medium is any material that has the ability to become magnetized, in set amounts, in large sections along its entire length.

1. True
2. False

1-11. Magnetic heads are transducers that convert the electrical variations of your input signal into what type of variations that are stored on a recording medium?

1. Magnetic
2. Electrical
3. Electrical and magnetic combined
4. Direct current

1-12. What factor determines the number of turns of wire placed on the core of a magnetic head?

1. The size of the magnetic head
2. The size of the head-gap
3. The purpose of the magnetic head
4. Each of the above

1-13. All magnetic heads operate the same way. An electric current passes through the coil. Magnetic field lines associated with the electric current follow paths through the core material. When the magnetic field lines get to the head-gap, some of them spread outside the core to form a fringing field. When a recording medium is passed through this fringing field, magnetic recording happens.

1. True
2. False

1-14. Which of the following materials is used to make magnetic tape?

1. Base material
2. Magnetic oxide coating
3. Glue
4. Each of the above

1-15. The base material of magnetic tape is made with what material?

1. Plastic
2. Metal
3. Either 1 or 2
4. Paper

1-16. Which of the following can happen if the magnetic particles used to make magnetic tape aren't uniform in size?

1. The tape's surface will be abrasive which reduces the magnetic head's life
2. The frequency response of the tape will be distorted
3. The glue will not hold the particles in place
4. The oxide coating cannot be magnetized

1-17. Generally, short magnetic oxide particles are used on magnetic tape to record low-frequency signals and long particles are used to record high-frequency signals.

1. True
2. False

1-18. Why does digital magnetic tape use a base material about 50 percent thicker than analog magnetic tape?

1. Thicker tape is needed to hold the long magnetic oxide particles
2. Thicker tape allows the tape to withstand the more strenuous starts and stops
3. Thicker tape is needed for instrumentation type signals
4. Thicker tape can store more digital data

1-19. There are four types of tape errors that can degrade the performance of a magnetic recording system. Which of the following tape errors is the most common and causes a 50% or more drop in signal strength?

1. Noise
2. Skew
3. Signal dropout
4. Level

1-20. Which of the following contaminants can cause signal dropouts?

1. Dust
2. Lint
3. Oil
4. Each of the above

1-21. Which of the following items on magnetic tape will cause noise errors?

1. Dust or lint
2. A cut
3. A scratch
4. Both 2 and 3 above

1-22. Brand "X" magnetic tape is rated for 5 volts (± 10%). Which of the following minimum and maximum voltage levels is what the output signal level could vary before it is considered a level error?

1. 4.95/5.05
2. 4.75/5.25
3. 4.5/5.5
4. 4/6

1-23. Magnetic tape can eventually become unusable when it comes in contact with the fixed surfaces of a recorder over long periods of time. What causes this type of tape failure?

1. Environmental damage
2. Normal wear
3. Winding errors
4. Accidental damage

1-24. Ideally, magnetic tape should be used and stored within what temperature and humidity range?

1. 40 - 80º F, 60 - 80% humidity
2. 50 - 70º F, 60 - 80% humidity
3. 60 - 80º F, 60 - 80% humidity
4. 60 - 80º F, 40 - 60% humidity

1-25. Using magnetic tape in a workplace that exceeds the ideal temperature and humidity ranges can cause which of the following types of environmental damage to the tape?

1. Dirt build-up
2. Layer-to-layer sticking
3. Tape deformation
4. Each of the above

1-26. At temperatures above 130 degrees, what happens to a tape's oxide coating?

1. Becomes soft
2. Separates from the base material
3. Both 1 and 2 above
4. Becomes brittle

1-27. What causes head-to-tape sticking?

1. Temperatures below 2º F
2. Dirty record/reproduce heads
3. The tape binder glue softens
4. Static electricity

1-28. Which of the following environmental conditions could create static electricity which attracts dirt build-up on the tape and tape recorder parts?

1. Relative humidity of 5%
2. Temperature of 28º F
3. Relative humidity of 96%
4. Temperature of 135º F

1-29. Excessive tape and head wear caused by increased friction as the tape passes over the heads can happen if the relative humidity is more that what maximum percent?

1. 70%
2. 85%
3. 95%
4. 90%

1-30. Winding errors can happen when improper winding practices create an excessive or uneven force as the tape is being wound onto a tape reel.

1. True
2. False

1-31. When winding a tape onto a reel, you notice that a sudden stop causes the outer layers of the tape to continue to spin after the inner layers have stopped. What type of deformed tape pack does this cause?

1. Pack-slip
2. Cinching
3. Spoking
4. Windowing

1-32. The magnetic tape on your reel is unwinding unevenly and rubbing against the sides of the reel and the recorder's tape guides. What type of deformed tape pack could cause this condition?

1. Spoking
2. Windowing
3. Pack-slip
4. Skewing

1-33. You notice a spoked tape pack on the take-up reel of your recorder. What causes a spoked tape pack?

1. Tape is wound over a small particle on the reel hub
2. A distorted reel hub creates uneven pressures as the tape is wound onto the reel
3. Tape is wound with increasing tension toward the end of the winding
4. Each of the above

1-34. Which of the following tape pack deformities is an example of windowing?

1. The inner part of the tape pack is buckled and deformed
2. The tape pack contains steps caused by the tape shifting from side to side during winding
3. The tape pack contains voids or see-through air gaps
4. Each of the above

1-35. Which of the following materials can be used to make tape reels?

1. Plastic
2. Metal
3. Glass
4. Each of the above

1-36. The flanges of a magnetic tape reel are designed to guide the magnetic tape onto the reel.

1. True
2. False

1-37. Magnetic tape is erased by exposing it to what type of magnetic field?

1. A gradually decreasing dc field
2. A gradually increasing dc field
3. A gradually decreasing ac field
4. A gradually increasing ac field

1-38. Which of the following is a disadvantage of using a tape recorder's erase head to erase magnetic tape?

1. It causes a distorted tape pack
2. It's slow
3. It may not do a complete erasure
4. Both 2 and 3 above

1-39. Both manual and automatic tape degaussers use the same electronic principles for erasing magnetic tape.

1. True
2. False

1-40. Which of the following is NOT an appropriate place to keep a magnetic tape reel or cartridge?

1. Mounted on a magnetic tape recorder
2. Stored in a plastic bag
3. Laying on top of a magnetic tape recorder
4. Both 2 and 3 above

1-41. Which of the following is a CORRECT method for handling magnetic tape?

1. Never let any part of the tape, except the end, trail on the floor
2. Always hold a tape reel by the flanges
3. Never handle or touch a tape's working surface
4. Each of the above

1-42. Which of the following is a CORRECT method for storing magnetic tape?

1. Always store tape reels laying on their sides, never vertically
2. Store tapes away from equipment that generates stray magnetic fields
3. Keep the storage area at 40 to 80% relative humidity
4. All of the above

1-43. Which of the following is a CORRECT method for packaging magnetic tape for shipping?

1. Always use reel bands where available
2. Always package reels supported by their hubs and in a vertical position
3. Always package cartridges in their shipping cases
4. Each of the above

1-44. What parts of a magnetic tape recorder are considered the heart of magnetic tape recording?

1. Magnetic tape reel
2. Magnetic heads
3. Transport electronics
4. Operator control panel

1-45. The heads of a magnetic tape recorder perform what part of the overall tape recording process?

1. Record signal or data onto magnetic tape
2. Reproduce signal or data from magnetic tape
3. Erase signal or data from magnetic tape
4. Each of the above

1-46. All magnetic heads are made using a plastic core wrapped with a few turns of very thin wire.

1. True
2. False

1-47. What specification of a magnetic head determines the maximum frequency the head will be able to transfer onto and off of the magnetic tape?

1. Headgap
2. Size of the core
3. Number of turns of wire on the core
4. Each of the above

1-48. What is the only physical difference between a record head and a reproduce head?

1. Number of turns of wire on the core
2. Size of the core material
3. Type of core material used
4. Each of the above

1-49. A certain magnetic recorder has a maximum recording frequency response of 500 kHz. Which of the following frequencies would be a good frequency to use as an erase signal?

1. 500 kHz
2. 1200 kHz
3. 1900 kHz
4. Either 2 or 3

1-50. If you do NOT regularly clean a recorder's magnetic heads, what is the probable consequence?

1. Dirt, dust, lint, etc. will collect on the heads
2. Signal dropout errors will occur
3. Both 1 and 2 above
4. The tape pack will become skewed

1-51. What materials should you use to clean magnetic heads?

1. A bristled brush and non-detergent cleaner
2. A cotton-tipped applicator soaked in isopropyl alcohol
3. A cotton-tipped applicator soaked in a magnetic head cleaner recommended by the manufacturer
4. Either 2 or 3 above

1-52. Magnetic heads can become magnetized from many sources. Which of the following sources, if any, could magnetize a magnetic head?

1. Stray magnetic fields
2. Excessive humidity
3. Poor quality magnetic tape
4. None of the above

1-53. Magnetic head degaussers generate a strong ac magnetic field that dc-magnetizes the metal parts of a magnetic head.

1. True
2. False

1-54. Which of the following is NOT a procedure for demagnetizing magnetic heads?

1. Remove the tape reel or cartridge from the recorder
2. Touch the energized degausser to the head
3. Move the degausser back and forth across the head for 15 to 30 seconds
4. De-energize the degausser when it's an arms length away

1-55. Which of the following is a function of magnetic tape transports?

1. Moves the magnetic tape across the magnetic heads
2. Holds the moving tape
3. Protects the moving tape
4. Each of the above

IN ANSWERING QUESTIONS 1-56 THROUGH 1-59, SELECT THE DESCRIPTION IN COLUMN B THAT BEST DESCRIBES THE MAGNETIC TAPE RECORDER TRANSPORT PART LISTED IN COLUMN A.

1-56. Tape reeling system	1. Monitors and controls the movement of magnetic tape
1-57. Tape speed control system	2. Holds and protects the reels or cartridges of magnetic tape
1-58. Electronic sub-system	3. Activates the reeling device to move the magnetic tape
1-59. Basic enclosure	4. Physically moves the magnetic tape across the magnetic heads

---

1-60. What type of tape reeling system uses a "free-spooling" supply reel and a motorized take-up reel?

1. Take-up control
2. Two-motor reeling
3. Tape buffering
4. Both 2 and 3 above

1-61. What type of tape reeling system was invented to overcome the problems of uneven tape tension and stretched tape?

1. Take-up control
2. Two-motor reeling
3. Tape buffering
4. Both 2 and 3 above

1-62. Magnetic tape reeling systems that use the spring tension method of tape buffering use what type of device to sense changes in tape tension?

1. Vacuum air column
2. Electro-mechanical
3. Photo-sensitive
4. Variable-static-sensor

1-63. What are the two types of tape guides used on magnetic tape reeling systems?

1. Variable and stable
2. Round and square
3. Single and dual
4. Fixed and rotary

1-64. A co-axial tape reeling configuration places the supply and take-up reels side by side.

1. True
2. False

1-65. In which of the following types of tape transport capstan drive configurations is the tape tension and head-to-tape contact most likely to vary?

1. Dual-motor dual capstan drive
2. Peripheral drive capstan
3. Open-loop capstan drive
4. Closed-loop capstan drive

1-66. Which of the following closed-loop capstan drive tape transports will NOT work in reverse?

1. Differential velocity capstan
2. Peripheral drive capstan
3. Both 1 and 2 above
4. Dual-motors dual capstan

1-67. What type of capstan drive tape transport moves the magnetic tape by placing the capstan directly against the tape reel or tape pack?

1. Open-loop capstan drive
2. Peripheral drive capstan
3. Dual-motors dual capstan
4. Differential velocity capstan

1-68. Capstan speed control is a very important function of the magnetic tape transport system. Which of the following is NOT one of the six parts of a capstan speed control?

1. Capstan speed monitor
2. Speed select network
3. Comparison network
4. Reel motor drive circuit

1-69. What part of the capstan speed control function normally uses a photo-optical tachometer attached to the shaft of the capstan motor?

1. Precision frequency source
2. Speed select circuit
3. Capstan speed monitor
4. Comparison network

1-70. On some recorders, a servo control from tape signal is supplied to the comparison network of the capstan speed control function. What is the purpose of this signal?

1. Speeds up or slows down the capstan motor
2. Compensates for capstan speed differences when a recorded tape is played back on a different recorder
3. Monitors the true capstan motor speed
4. Provides a reference frequency that the speed select network uses to drive the capstan motor

1-71. Which of the following items can be used to clean most magnetic tape transports?

1. Isopropyl alcohol
2. Cotton swabs
3. Lint-free cloths
4. Each of the above

1-72. Which of the following is NOT a correct procedure for cleaning a magnetic tape transport?

1. When available, always use cotton swabs vice lint-free cloths
2. Apply the cleaner onto the cotton swab or lint-free cloth
3. Clean the flanged parts of the tape guides
4. While cleaning, switch swabs and cloths often

1-73. Magnetic tape transports should be demagnetized periodically. To do this, use a hand-held degausser and follow the procedures for demagnetizing magnetic heads.

1. True
2. False

ASSIGNMENT 2

Textbook assignment: Chapter 5, “Magnetic Tape Recorder Record and Reproduce Electronics,” pages 5-1 through 5-6. Chapter 6, “Magnetic Tape Recording Specifications,” pages 6-1 through 6-15. Chapter 7, “Digital Magnetic Tape Recording,” pages 7-1 through 7-9. Chapter 8 “ Magnetic Disk Recording,” pages 8-1 through 8-25.

2-1. Magnetic tape recorders use what type of electronic circuits to record and reproduce analog input signals?

1. Continuous wave
2. Frequency modulation
3. Amplitude modulation
4. Both 2 and 3 above

2-2. Direct record electronics record input signals onto magnetic tape just as they appeared at the recorder's input.

1. True
2. False

2-3. Which part of the direct record electronics component takes the input signal and the bias signal and mixes them together?

1. Input pre-amplifier circuit
2. Bias source
3. Summing network
4. Head driver circuit

2-4. Which part of the direct record electronics amplifies the signal from the summing network and sends it to the record head?

1. Bias source circuit
2. Head driver circuit
3. Input pre-amplifier circuit
4. Low-pass filter circuit

2-5. The pre-amplifier circuit of the direct reproduce electronics does which of the following functions?

1. Amplifies the reproduced signal
2. Removes the bias signal
3. Both 1 and 2 above
4. Corrects phase errors

2-6. What circuit in the direct reproduce electronics takes the pre-amplified signal and fixes frequency response problems the reproduce magnetic head may have caused?

1. Equalization and phase correction circuit
2. Output amplifier circuit
3. Head driver circuit
4. Summing network

2-7. What circuit of the direct reproduce electronics serves as an impedance matcher?

1. Pre-amplifier circuit
2. Output amplifier circuit
3. Equalization and phase correction circuit
4. Head driver circuit

2-8. How do FM record electronics process the incoming signal before sending it to the record head?

1. A high frequency, negative bias is added to the input signal
2. A summing network mixes the bias and input signal
3. Both 1 and 2 above
4. The input signal is frequency modulated onto the carrier frequency of a record oscillator

2-9. In the FM record electronics, what is the output of the record oscillator circuit?

1. The demodulated input signal
2. The frequency modulated carrier signal
3. A clean input signal with the negative bias removed
4. A combined input signal and equalization signal

THIS SPACE LEFT BLANK INTENTIONALLY.

IN ANSWERING QUESTIONS 2-10 THROUGH 2-13, SELECT THE DESCRIPTION IN COLUMN B THAT BEST DESCRIBES THE PART OF FM REPRODUCE ELECTRONICS LISTED IN COLUMN A.

2-10. Pre-amplifier circuit

2-11. Limiter/ demodulator circuit

2-12. Low-pass filter circuit

2-13. Output amplifier

1. Takes the signal from the limiter/ demodulator and cleans-up any noise or left over carrier signal
2. Takes the output from the low-pass filter and amplifies it for output
3. Takes the reproduce signal from the magnetic head and amplifies it
4. Stabilizes the amplitude level and demodulates the signal intelligence from the carrier signal

---

2-14. Which of the following generates the noise part of a recorder's SNR?

1. Magnetic heads
2. Magnetic tape
3. Both 1 and 2 above
4. Nearby equipment

2-15. The SNR can be stated in three different ways. Which of the following is NOT one of these ways?

1. Mean signal to RMS noise
2. Peak-to-peak signal to RMS noise
3. RMS signal to RMS noise
4. Peak signal-to-RMS noise

2-16. Which of the following data should be included with all SNR specifications?

1. Record level
2. Reproduce level
3. Bandwidth
4. Tape speed

2-17. What is the magnetic tape recording specification which gives a recorder's amplitude variation with frequency over a specified bandwidth?

1. Record level
2. Frequency response
3. Bandwidth
4. Both 2 and 3 above

2-18. Your LPO tells you to test the frequency response of a particular magnetic tape recorder. You use a signal generator to sweep through the frequencies as you monitor the recorder's output amplitude on a VTVM. Which of the following readings in output amplitude would indicate the upper and lower end of the recorders bandwidth?

1. +/− 2-dB
2. +/− 3-dB
3. +/− 5-dB
4. +/− 10-dB

2-19. Which of the following factors can limit or degrade the frequency response of a magnetic tape recorder?

1. The record head
2. The reproduce head
3. Magnetic head-to-tape contact
4. Each of the above

2-20. If the frequencies of the harmonic distortion are 2, 4, and 6 times the center frequency, what type of harmonic distortion is this?

1. Linear
2. Spatial
3. Even-order
4. Center frequency

2-21. Which of the following is a cause of even-order harmonics during magnetic tape recording?

1. Defective magnetic tape
2. Frequency response too low
3. Permanently magnetized heads
4. Each of the above

2-22. If you increase a magnetic tape recorder's signal bias level, what happens to the harmonic distortion?

1. Increases
2. Remains the same
3. Decreases
4. Increases odd-order harmonic distortion

2-23. When measuring the amount of harmonic distortion in a magnetic tape recorder, what electronic test equipment should you use?

1. Wave analyzer
2. VTVM
3. Spectrum analyzer
4. All of the above

2-24. When measuring harmonic distortion, you set the signal generator to input a 12-kHz test signal. At the recorder's output, what will be the frequency of the third-order harmonic?

1. 33-kHz
2. 18-kHz
3. 38-kHz
4. 36 kHz

2-25. What magnetic tape recorder specification expresses the variation of the phase shift with respect to frequency?

1. Frequency response
2. Phase response
3. Wow and Flutter
4. Each of the above

2-26. Which of the following conditions indicates that a magnetic tape recorder has a good phase response specification?

1. The SNR and frequency response are within tolerance
2. The wave analyzer shows a perfect sine wave
3. The recorder reproduces an undistorted square wave
4. Each of the above

2-27. What magnetic tape recorder specification expresses the result of non-uniform tape motion caused by variations in tape speed that produces frequency modulation of signals recorded onto magnetic tape?

1. Frequency response
2. Phase response
3. Flutter
4. Time-base error

2-28. Which of the following magnetic tape recorder transport parts can cause high frequency flutter (above 1000 Hz)?

1. Magnetic heads
2. Rotating tape guides
3. Fixed tape guides
4. Both 1 and 3 above

2-29. A magnetic tape recorder's flutter specification is usually expressed as a percent of peak or as a peak-to-peak value for what type of recorder?

1. Audio
2. Instrumentation
3. Video
4. All of the above

2-30. The time-base error (TBE) magnetic tape recorder specification is closely related to flutter. Which of the following statements best reflects this relationship?

1. TBE is an inverse measure of the effects of flutter on the stability of recorded data
2. TBE is a direct measure of the effects of flutter on the frequency response of recorded data
3. TBE is a direct measure of the effects of flutter on the stability of recorded data.
4. TBE is a direct measure of the effects of flutter on the bias level of recorded data.

2-31. The simplest way to measure a recorder's TBE is with what test equipment?

1. VTVM
2. Sweep analyzer
3. Oscilloscope
4. Ohmeter

2-32. There are two types of skew. What type does not show up when magnetic tapes are recorded and reproduced on the same magnetic tape recorder?

1. Positive
2. Dynamic
3. Negative
4. Fixed

2-33. Which of the following is NOT a cause of dynamic skew?

1. Gap scatter in the magnetic head stack
2. Sticking tape transport guides
3. Warped magnetic tape
4. Worn tape transport guides

2-34. Which of the following is NOT a format for digital magnetic tape recording?

1. Serial-parallel
2. Bi-phase
3. Serial
4. Parallel

2-35. Which of the following digital magnetic tape recording formats is normally used for instrumentation recording when the input data rate is high?

1. Serial-Parallel
2. Serial
3. Both 1 and 2 above
4. Bi-phase

2-36. The return-to-bias digital magnetic tape recording encoding method uses magnetic tape that is normally in a "neutral" condition.

1. True
2. False

2-37. Which of the following digital magnetic tape encoding methods is the most widely used?

1. RB
2. NRZ
3. E-RZ
4. RZ

2-38. Which of the four variations of NRZ encoding works best in high density magnetic tape recording and offers a bit-error rate of one error per 1 million bits?

1. E-NRZ-L
2. NRZ-L
3. NRZ-M
4. NRZ-S

2-39. Digital magnetic tape recorders are used to store and retrieve which of the following types of data?

1. Computer programs
2. Radar and other pulsed type signals
3. Special signals with a bandwidth of less than 500 kHz
4. Each of the above

2-40. Telemetry digital magnetic tape recorders are frequently called wideband recorders.

1. True
2. False

2-41. Floppy disks are made of round plastic disks coated with magnetic oxide particles and enclosed in a plastic jacket.

1. True
2. False

2-42. When discussing floppy disks, what is meant by the "density" of a floppy disk?

1. The thickness of the plastic disk
2. How much the disk can store
3. The thickness of floppy disk jacket
4. The number of sectors on the disk

IN ANSWERING QUESTIONS 2-43 THROUGH 2-46, SELECT THE STORAGE CAPACITY IN COLUMN B THAT BEST DESCRIBES THE TYPE OF FLOPPY DISK IN COLUMN A.

2-43.	5-1/4" double density	1.	1,200,000 bytes
2-44.	5-1/4" high density	2.	720,000 bytes
2-45.	3-1/2" double density	3.	360,000 bytes
2-46.	3-1/2" high density	4.	1,400,000 bytes

2-47. Data is stored on floppy disks in circular "tracks." Each track is then broken up into arcs called "cylinders."

1. True
2. False

2-48. When a floppy disk is sectored using the soft sectoring method, the computer software determines the sector size and placement. What is this process called?

1. Centering
2. Addressing
3. Formatting
4. Rastering

2-49. When you handle, store, or ship floppy disks, which of the following statements is NOT a precaution you should take?

1. Always store 8" and 5 1/4" floppy disks in their envelopes when not in use
2. Always write on a floppy disk label first, and then place it on the disk
3. Always lay floppy disks on their side when storing them
4. Always ship floppy disks in their appropriate shipping containers

2-50. Which of the following items can generate magnetic fields that can destroy data on a floppy disk?

1. Paper clip
2. Telephone
3. Printer
4. Both 2 and 3 above

2-51. How can you erase a floppy disk?

1. Record over it
2. Degauss it
3. Reformat it
4. Both 2 and 3 above

2-52. A computer places data on a hard disk by using one of what two methods?

1. Cylinder or sector
2. Cylinder or circular
3. Sector or quadrant
4. Sector or record

2-53. Using the sector method, a hard disk drive locates a place on a hard disk with only one platter by using three location numbers. Which of the following is NOT one of those location numbers?

1. Surface number
2. Track number
3. Cylinder number
4. Sector number

2-54. When you handle, store, or ship removable hard disks, which of the following statements is NOT a precaution you should take?

1. Don't touch any exposed recording surfaces
2. Keep them away from food, liquids, and cigarette smoke
3. Store them in an environment that stays between 32 to 95 degrees Fahrenheit and 40 to 85% relative humidity
4. Keep dirt, dust, etc., off of the recording surface by storing them in their case when not in use

2-55. How can you declassify a removable hard disk which contains classified information?

1. Reformat it
2. Degauss it
3. Both 1 and 2 above
4. Destroy it using the procedures in OPNAVINST 5510.1

2-56. Which of the following encoding methods is NOT used for encoding digital data onto magnetic disks?

1. Sector encoding
2. Run length limited
3. Modified frequency modulation
4. Frequency modulation

IN ANSWERING QUESTIONS 2-57 THROUGH 2-60, SELECT THE DESCRIPTION IN COLUMN B THAT DESCRIBES THE FLOPPY DISK DRIVE PART LISTED IN COLUMN A.

2-57. Head arm assembly	1. Holds and spins the floppy disk
2-58. Drive electronics circuit board	2. Holds the magnetic read/write heads
2-59. Drive motor/spindle assembly	3. Controls the electromechanical-high density parts
2-60. Actuator arm assembly	4. Positions the heads over the disks

---

2-61. What part of a hard disk drive transport uses either a dc stepper motor or a voice coil to position the heads for writing data to the correct track of the disk pack?

1. Drive motor/spindle assembly
2. Head arm assembly
3. Actuator arm assembly
4. Drive electronics circuit board

2-62. What part of a hard disk drive transport holds the four magnetic heads, and is attached to the transport's actuator arm assembly?

1. Cylinder assembly
2. Head arm assembly
3. Sectoring assembly
4. Drive motor/spindle assembly

2-63. Why do floppy disk drives require more preventive maintenance than hard disk drives?

1. They are not sealed units
2. Oxide coating from disks sticks to transport parts
3. Both 1 and 2 above
4. The drive circuit board is less protected

2-64. Which of the following is a main function of the control electronics part of a disk drive's electronics component?

1. Take incoming data from the interface electronics
2. Spin the disk at the proper speed
3. Move the heads across the recording surface
4. Both 2 and 3 above

2-65. Which of the following is a function of the interface electronics part of a disk drive's electronics component?

1. Convert data from the host computer from serial to parallel, and vice versa, as needed
2. Receive write/read control signals from the host computer
3. Receive control signals to format a disk from the host computer
4. All of the above

2-66. Which of the following is NOT one of the five most common disk drive interfaces in use today?

1. Drive motor/spindle interface
2. Naval tactical data system interface
3. ST-506/412 interface
4. Enhanced small device interface

2-67. What type of disk drive interface is often used in the hard disk drives installed in older IBM-compatible desktop computers that have a maximum capacity of 125MB?

1. Drive motor/spindle interface
2. Enhanced small device interface
3. ST-506/412 interface
4. Integrated drive electronics

2-68. What type of disk drive interface uses a high-level interface which requires only a logical sector number to locate the desired data on a disk?

1. Enhanced small device interface
2. Integrated drive electronics
3. Drive motor/spindle interface
4. Small computer systems interface

2-69. What type of disk drive interface includes all of the controller card electronics in the hard disk drive and offers a transfer rate of up to 1 MB?

1. Small computer systems interface
2. Integrated drive electronics
3. Naval tactical data system
4. Enhanced small device interface

2-70. Which of the following is NOT a benefit of the SCSI disk drive interface over previous interfaces?

1. Can handle disk drives of almost any size
2. Disconnects itself from the host computer's bus while it processes requests
3. Can transfer data up to 24 Mbits/sec
4. Can daisy-chain up to eight units off of one controller

IN ANSWERING QUESTIONS 2-71 THROUGH 2-74, SELECT THE MAGNETIC DISK RECORDING SPECIFICATION IN COLUMN B THAT MATCHES THE DEFINITION IN COLUMN A.

A.	B.
2-71. Indicates number of physical sectors that are between logical sectors on a hard disk	1. Seek time
2-72. Total time it takes a disk drive to retrieve a sector of data	2. Latency period
2-73. Time it takes for a magnetic head to position itself over a specific track	3. Access time
2-74. Time it takes a specific sector of a specific track to position itself under the magnetic head	4. Interleave factor

---

2-75. The speed at which a disk drive and a disk drive controller working together can transfer data to the host computer is what disk recording specification?

1. Seek time
2. Transfer rate
3. Access time
4. Interleave factor

NONRESIDENT
TRAINING
COURSE

SEPTEMBER 1998

Navy Electricity and Electronics Training Series

Module 24—Introduction to Fiber Optics

NAVEDTRA 14196

Notice: NETPDTC is no longer responsible for the content accuracy of the NRTCs.

For content issues, contact the servicing Center of Excellence: Center for Surface Combat System (CSCS); (540) 284-1061 or DSN: 249-1061.

Although the words "he," "him," and "his" are used sparingly in this course to enhance communication, they are not intended to be gender driven or to affront or discriminate against anyone.

PREFACE

By enrolling in this self-study course, you have demonstrated a desire to improve yourself and the Navy. Remember, however, this self-study course is only one part of the total Navy training program. Practical experience, schools, selected reading, and your desire to succeed are also necessary to successfully round out a fully meaningful training program.

COURSE OVERVIEW: To introduce the student to the subject of Fiber Optics who needs such a background in accomplishing daily work and/or in preparing for further study.

THE COURSE: This self-study course is organized into subject matter areas, each containing learning objectives to help you determine what you should learn along with text and illustrations to help you understand the information. The subject matter reflects day-to-day requirements and experiences of personnel in the rating or skill area. It also reflects guidance provided by Enlisted Community Managers (ECMs) and other senior personnel, technical references, instructions, etc., and either the occupational or naval standards, which are listed in the *Manual of Navy Enlisted Manpower Personnel Classifications and Occupational Standards,* NAVPERS 18068.

THE QUESTIONS: The questions that appear in this course are designed to help you understand the material in the text.

VALUE: In completing this course, you will improve your military and professional knowledge. Importantly, it can also help you study for the Navy-wide advancement in rate examination. If you are studying and discover a reference in the text to another publication for further information, look it up.

1998 Edition Prepared by
ETCS(SW) Donnie Jones

NAVSUP Logistics Tracking Number
0504-LP-026-8470

Sailor's Creed

"I am a United States Sailor.

I will support and defend the Constitution of the United States of America and I will obey the orders of those appointed over me.

I represent the fighting spirit of the Navy and those who have gone before me to defend freedom and democracy around the world.

I proudly serve my country's Navy combat team with honor, courage and commitment.

I am committed to excellence and the fair treatment of all."

TABLE OF CONTENTS

NAVY ELECTRICITY AND ELECTRONICS TRAINING SERIES

The Navy Electricity and Electronics Training Series (NEETS) was developed for use by personnel in many electrical- and electronic-related Navy ratings. Written by, and with the advice of, senior technicians in these ratings, this series provides beginners with fundamental electrical and electronic concepts through self-study. The presentation of this series is not oriented to any specific rating structure, but is divided into modules containing related information organized into traditional paths of instruction.

The series is designed to give small amounts of information that can be easily digested before advancing further into the more complex material. For a student just becoming acquainted with electricity or electronics, it is highly recommended that the modules be studied in their suggested sequence. While there is a listing of NEETS by module title, the following brief descriptions give a quick overview of how the individual modules flow together.

Module 1, *Introduction to Matter, Energy, and Direct Current*, introduces the course with a short history of electricity and electronics and proceeds into the characteristics of matter, energy, and direct current (dc). It also describes some of the general safety precautions and first-aid procedures that should be common knowledge for a person working in the field of electricity. Related safety hints are located throughout the rest of the series, as well.

Module 2, *Introduction to Alternating Current and Transformers,* is an introduction to alternating current (ac) and transformers, including basic ac theory and fundamentals of electromagnetism, inductance, capacitance, impedance, and transformers.

Module 3, *Introduction to Circuit Protection, Control, and Measurement,* encompasses circuit breakers, fuses, and current limiters used in circuit protection, as well as the theory and use of meters as electrical measuring devices.

Module 4, *Introduction to Electrical Conductors, Wiring Techniques, and Schematic Reading,* presents conductor usage, insulation used as wire covering, splicing, termination of wiring, soldering, and reading electrical wiring diagrams.

Module 5, *Introduction to Generators and Motors,* is an introduction to generators and motors, and covers the uses of ac and dc generators and motors in the conversion of electrical and mechanical energies.

Module 6, *Introduction to Electronic Emission, Tubes, and Power Supplies,* ties the first five modules together in an introduction to vacuum tubes and vacuum-tube power supplies.

Module 7*, Introduction to Solid-State Devices and Power Supplies,* is similar to module 6, but it is in reference to solid-state devices.

Module 8, *Introduction to Amplifiers,* covers amplifiers.

Module 9, *Introduction to Wave-Generation and Wave-Shaping Circuits,* discusses wave generation and wave-shaping circuits.

Module 10, *Introduction to Wave Propagation, Transmission Lines, and Antennas,* presents the characteristics of wave propagation, transmission lines, and antennas.

Module 11, *Microwave Principles,* explains microwave oscillators, amplifiers, and waveguides.

Module 12, *Modulation Principles,* discusses the principles of modulation.

Module 13, *Introduction to Number Systems and Logic Circuits,* presents the fundamental concepts of number systems, Boolean algebra, and logic circuits, all of which pertain to digital computers.

Module 14, *Introduction to Microelectronics,* covers microelectronics technology and miniature and microminiature circuit repair.

Module 15, *Principles of Synchros, Servos, and Gyros,* provides the basic principles, operations, functions, and applications of synchro, servo, and gyro mechanisms.

Module 16, *Introduction to Test Equipment,* is an introduction to some of the more commonly used test equipments and their applications.

Module 17, *Radio-Frequency Communications Principles,* presents the fundamentals of a radio-frequency communications system.

Module 18, *Radar Principles,* covers the fundamentals of a radar system.

Module 19, *The Technician's Handbook,* is a handy reference of commonly used general information, such as electrical and electronic formulas, color coding, and naval supply system data.

Module 20, *Master Glossary,* is the glossary of terms for the series.

Module 21, *Test Methods and Practices,* describes basic test methods and practices.

Module 22, *Introduction to Digital Computers,* is an introduction to digital computers.

Module 23, *Magnetic Recording,* is an introduction to the use and maintenance of magnetic recorders and the concepts of recording on magnetic tape and disks.

Module 24, *Introduction to Fiber Optics,* is an introduction to fiber optics.

Embedded questions are inserted throughout each module, except for modules 19 and 20, which are reference books. If you have any difficulty in answering any of the questions, restudy the applicable section.

Although an attempt has been made to use simple language, various technical words and phrases have necessarily been included. Specific terms are defined in Module 20, *Master Glossary*.

Considerable emphasis has been placed on illustrations to provide a maximum amount of information. In some instances, a knowledge of basic algebra may be required.

Assignments are provided for each module, with the exceptions of Module 19, *The Technician's Handbook*; and Module 20, *Master Glossary*. Course descriptions and ordering information are in NAVEDTRA 12061, *Catalog of Nonresident Training Courses*.

Throughout the text of this course and while using technical manuals associated with the equipment you will be working on, you will find the below notations at the end of some paragraphs. The notations are used to emphasize that safety hazards exist and care must be taken or observed.

WARNING

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN INJURY OR DEATH IF NOT CAREFULLY OBSERVED OR FOLLOWED.

CAUTION

AN OPERATING PROCEDURE, PRACTICE, OR CONDITION, ETC., WHICH MAY RESULT IN DAMAGE TO EQUIPMENT IF NOT CAREFULLY OBSERVED OR FOLLOWED.

NOTE

An operating procedure, practice, or condition, etc., which is essential to emphasize.

INSTRUCTIONS FOR TAKING THE COURSE

ASSIGNMENTS

The text pages that you are to study are listed at the beginning of each assignment. Study these pages carefully before attempting to answer the questions. Pay close attention to tables and illustrations and read the learning objectives. The learning objectives state what you should be able to do after studying the material. Answering the questions correctly helps you accomplish the objectives.

SELECTING YOUR ANSWERS

Read each question carefully, then select the BEST answer. You may refer freely to the text. The answers must be the result of your own work and decisions. You are prohibited from referring to or copying the answers of others and from giving answers to anyone else taking the course.

SUBMITTING YOUR ASSIGNMENTS

To have your assignments graded, you must be enrolled in the course with the Nonresident Training Course Administration Branch at the Naval Education and Training Professional Development and Technology Center (NETPDTC). Following enrollment, there are two ways of having your assignments graded: (1) use the Internet to submit your assignments as you complete them, or (2) send all the assignments at one time by mail to NETPDTC.

Grading on the Internet: Advantages to Internet grading are:

- you may submit your answers as soon as you complete an assignment, and
- you get your results faster; usually by the next working day (approximately 24 hours).

In addition to receiving grade results for each assignment, you will receive course completion confirmation once you have completed all the assignments. To submit your assignment answers via the Internet, go to:

https://courses.cnet.navy.mil

COMPLETION TIME

Courses must be completed within 12 months from the date of enrollment. This includes time required to resubmit failed assignments.

PASS/FAIL ASSIGNMENT PROCEDURES

If your overall course score is 3.2 or higher, you will pass the course and will not be required to resubmit assignments. Once your assignments have been graded you will receive course completion confirmation.

If you receive less than a 3.2 on any assignment and your overall course score is below 3.2, you will be given the opportunity to resubmit failed assignments. **You may resubmit failed assignments only once.** Internet students will receive notification when they have failed an assignment—they may then resubmit failed assignments on the web site. Internet students may view and print results for failed assignments from the web site. Students who submit by mail will receive a failing result letter and a new answer sheet for resubmission of each failed assignment.

COMPLETION CONFIRMATION

After successfully completing this course, you will receive a letter of completion.

STUDENT FEEDBACK QUESTIONS

We value your suggestions, questions, and criticisms on our courses. If you would like to communicate with us regarding this course, we encourage you, if possible, to use e-mail.

NAVAL RESERVE RETIREMENT CREDIT

If you are a member of the Naval Reserve, you may earn retirement points for successfully completing this course, if authorized under current directives governing retirement of Naval Reserve personnel. For Naval Reserve retirement, this course is evaluated at 6 points. (Refer to *Administrative Procedures for Naval Reservists on Inactive Duty,* BUPERSINST 1001.39, for more information about retirement points.)

Student Comments

Course Title: *NEETS Module 24*
Introduction to Fiber Optics

NAVEDTRA: 14196 **Date**: ______

We need some information about you:

Rate/Rank and Name: ______ SSN: ______ Command/Unit ______

Street Address: ______ City: ______ State/FPO: ______ Zip ______

Your comments, suggestions, etc.:

Privacy Act Statement: Under authority of Title 5, USC 301, information regarding your military status is requested in processing your comments and in preparing a reply. This information will not be divulged without written authorization to anyone other than those within DOD for official use in determining performance.

NETPDTC 1550/41 (Rev 4-00)

CHAPTER 1

BACKGROUND ON FIBER OPTICS

LEARNING OBJECTIVES

Learning objectives are stated at the beginning of each chapter. These learning objectives serve as a preview of the information you are expected to learn in the chapter. The comprehensive check questions are based on the objectives. By successfully completing the NRTC, you indicate that you have met the objectives and have learned the information. The learning objectives are listed below.

Upon completing this chapter, you should be able to do the following:

1. Describe the term *fiber optics*.
2. List the parts of a fiber optic data link.
3. Understand the function of each fiber optic data link part.
4. Outline the progress made in the history of fiber optic technology.
5. Describe the trade-offs in fiber properties and component selection in the design of fiber optic systems.
6. List the advantages and the disadvantages of fiber optic systems compared to electrical communications systems.

DEFINITION OF FIBER OPTICS

In the other Navy Electricity and Electronics Training Series (*NEETS*) modules, you learn the basic concepts used in electrical systems. Electrical systems include telephone, radio, cable television (CATV), radar, and satellite links. In the past 30 years, researchers have developed a new technology that offers greater data rates over longer distances at costs lower than copper wire systems. This new technology is **fiber optics**.

Fiber optics uses light to send information (data). More formally, **fiber optics** is the branch of optical technology concerned with the transmission of radiant power (light energy) through fibers.

Q1. Define fiber optics.

FIBER OPTIC DATA LINKS

A fiber optic data link sends input data through fiber optic components and provides this data as output information. It has the following three **basic functions**:

- To convert an electrical input signal to an optical signal
- To send the optical signal over an optical fiber

- To convert the optical signal back to an electrical signal

A fiber optic data link consists of three parts—**transmitter**, **optical fiber**, and **receiver**. Figure 1-1 is an illustration of a fiber optic data-link connection. The transmitter, optical fiber, and receiver perform the basic functions of the fiber optic data link. Each part of the data link is responsible for the successful transfer of the data signal. A fiber optic data link needs a transmitter that can effectively convert an electrical input signal to an optical signal and launch the data-containing light down the optical fiber. A fiber optic data link also needs a receiver that can effectively transform this optical signal back into its original form. This means that the electrical signal provided as data output should exactly match the electrical signal provided as data input.

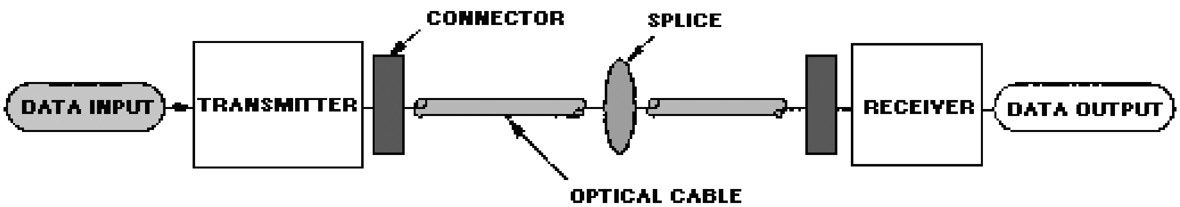


Figure 1-1.—Parts of a fiber optic data link.

The transmitter converts the input signal to an optical signal suitable for transmission. The transmitter consists of two parts, an interface circuit and a source drive circuit. The transmitter's drive circuit converts the electrical signals to an optical signal. It does this by varying the current flow through the light source. The two types of optical sources are light-emitting diodes (LEDs) and laser diodes.

The optical source launches the optical signal into the fiber. The optical signal will become progressively weakened and distorted because of scattering, absorption, and dispersion mechanisms in the fiber waveguides. Chapter 2 discusses the fiber mechanisms of scattering, absorption, and dispersion. Chapter 3 discusses the types of optical fibers and cables.

The receiver converts the optical signal exiting the fiber back into an electrical signal. The receiver consists of two parts, the optical detector and the signal-conditioning circuits. An optical detector detects the optical signal. The signal-conditioning circuit conditions the detector output so that the receiver output matches the original input to the transmitter. The receiver should amplify and process the optical signal without introducing noise or signal distortion. Noise is any disturbance that obscures or reduces the quality of the signal. Noise effects and limitations of the signal-conditioning circuits cause the distortion of the receiver's electrical output signal.

An optical detector can be either a semiconductor positive-intrinsic-negative (*PIN*) diode or an avalanche photodiode (APD). A *PIN* diode changes its electrical conductivity according to the intensity and wavelength of light. The *PIN* diode consists of an intrinsic region between p-type and n-type semiconductor material. Chapter 6 provides further explanation of optical sources. Chapter 7 provides further explanation of optical detectors.

A fiber optic data link also includes passive components other than an optical fiber. Figure 1-1 does not show the optical connections used to complete the construction of the fiber optic data link. Passive components used to make fiber connections affect the performance of the data link. These components can also prevent the link from operating. Fiber optic components used to make the optical connections include optical splices, connectors, and couplers. Chapter 4 outlines the types of optical splices, connectors, and couplers and their connection properties that affect system performance.

Proof of link performance is an integral part of the design, fabrication, and installation of any fiber optic system. Various measurement techniques are used to test individual parts of a data link. Each data link part is tested to be sure the link is operating properly. Chapter 5 discusses the laboratory and field measurements used to measure link performance.

Q2. Describe the basic functions of a fiber optic data link.

Q3. List the three parts of a fiber optic data link.

Q4. What mechanisms in the fiber waveguides weaken and distort the optical signal?

Q5. What effect does noise have on the fiber optic signal?

HISTORY OF FIBER OPTIC TECHNOLOGY

People have used light to transmit information for hundreds of years. However, it was not until the 1960s, with the invention of the laser, that widespread interest in optical (light) systems for data communications began. The invention of the laser prompted researchers to study the potential of fiber optics for data communications, sensing, and other applications. Laser systems could send a much larger amount of data than telephone, microwave, and other electrical systems. The first experiment with the laser involved letting the laser beam transmit freely through the air. Researchers also conducted experiments letting the laser beam transmit through different types of waveguides. Glass fibers, gas-filled pipes, and tubes with focusing lenses are examples of optical waveguides.

Glass fibers soon became the preferred medium for fiber optic research. Initially, the very large losses in the optical fibers prevented coaxial cables from being replaced. **Loss** is the decrease in the amount of light reaching the end of the fiber. Early fibers had losses around 1,000 dB/km making them impractical for communications use. In 1969, several scientists concluded that impurities in the fiber material caused the signal loss in optical fibers. The basic fiber material did not prevent the light signal from reaching the end of the fiber. These researchers believed it was possible to reduce the losses in optical fibers by removing the impurities. By removing the impurities, construction of low-loss optical fibers was possible.

There are two basic types of optical fibers, multimode fibers and single mode fibers. Chapter 2 discusses the differences between the fiber types. In 1970, Corning Glass Works made a multimode fiber with losses under 20 dB/km. This same company, in 1972, made a high silica-core multimode optical fiber with 4dB/km minimum attenuation (loss). Currently, multimode fibers can have losses as low as 0.5 dB/km at wavelengths around 1300 nm. Single mode fibers are available with losses lower than 0.25 dB/km at wavelengths around 1500 nm.

Developments in semiconductor technology, which provided the necessary light sources and detectors, furthered the development of fiber optics. Conventional light sources, such as lamps or lasers, were not easily used in fiber optic systems. These light sources tended to be too large and required lens systems to launch light into the fiber. In 1971, Bell Laboratories developed a small area light-emitting diode (LED). This light source was suitable for low-loss coupling to optical fibers. Researchers could then perform source-to-fiber jointing easily and repeatedly. Early semiconductor sources had operating lifetimes of only a few hours. However, by 1973, projected lifetimes of lasers advanced from a few hours to greater than 1,000 hours. By 1977, projected lifetimes of lasers advanced to greater than 7,000 hours. By 1979, these devices were available with projected lifetimes of more than 100,000 hours.

In addition, researchers also continued to develop new fiber optic parts. The types of new parts developed included low-loss fibers and fiber cables, splices, and connectors. These parts permitted demonstration and research on complete fiber optic systems.

Advances in fiber optics have permitted the introduction of fiber optics into present applications. These applications are mostly in the telephone long-haul systems, but are growing to include cable television, computer networks, video systems, and data links. Research should increase system performance and provide solutions to existing problems in conventional applications. The impressive results from early research show there are many advantages offered by fiber optic systems.

Q6. Define loss.

Q7. In 1969, what did several scientists conclude about optical fiber loss?

Q8. How can loss be reduced during construction (or fabrication) of optical fibers?

Q9. What are the two basic types of optical fibers?

FIBER OPTIC SYSTEMS

System design has centered on long-haul communications and the subscriber-loop plant. The subscriber-loop plant is the part of a system that connects a subscriber to the nearest switching center. Cable television is an example. Limited work has also been done on short-distance applications and some military systems. Initially, central office trunking required multimode optical fibers with moderate to good performance. Fiber performance depends on the amount of loss and signal distortion introduced by the fiber when it is operating at a specific wavelength. Long-haul systems require single mode optical fibers with very high performance. Single mode fibers tend to have lower loss and produce less signal distortion.

In contrast, short-distance and military systems tend to use only multimode technology. Examples of short-distance systems include process control and local area networks (LANs). Short-distance and military systems have many connections. The larger fiber core and higher fiber numerical aperture (NA) of multimode fibers reduce losses at these connections. Chapter 4 explains fiber connection properties in more detail. Chapter 2 provides more detail on multimode and single mode fibers.

In military and subscriber-loop applications, system design and parts selection are related. Designers consider **trade-offs** in the following areas:

- Fiber properties
- Types of connections
- Optical sources
- Detector types

Designers develop systems to meet stringent working requirements, while trying to maintain economic performance. It is quite difficult to identify a standard system design approach. This module identifies the types of components chosen by the Navy for shipboard applications.

Future system design improvements depend on continued research. Researchers expect fiber optic product improvements to upgrade performance and lower costs for short-distance applications. Future systems center on broadband services that will allow transmission of voice, video, and data. Services will include television, data retrieval, video word processing, electronic mail, banking, and shopping.

Q10. Which type of optical fiber (multimode or single mode) tends to have lower loss and produces less signal distortion?

Q11. What optical fiber properties reduce connection loss in short-distance systems?

Q12. In fiber optic systems, designers consider what trade-offs?

ADVANTAGES AND DISADVANTAGES OF FIBER OPTICS

Fiber optic systems have many attractive features that are superior to electrical systems. These include improved system performance, immunity to electrical noise, signal security, and improved safety and electrical isolation. Other advantages include reduced size and weight, environmental protection, and overall system economy. Table 1-1 details the main advantages of fiber optic systems.

Table 1-1.— Advantages of Fiber Optics

System Performance	• Greatly increased bandwidth and capacity • Lower signal attenuation (loss)
Immunity to Electrical Noise	• Immune to noise (electromagnetic interference [EMI] and radio-frequency interference [RFI] • No crosstalk • Lower bit error rates
Signal Security	• Difficult to tap • Nonconductive (does not radiate signals)
Electrical Isolation	• No common ground required • Freedom from short circuit and sparks
Size and Weight	• Reduced size and weight cables
Environmental Protection	• Resistant to radiation and corrosion • Resistant to temperature variations • Improved ruggedness and flexibility • Less restrictive in harsh environments
Overall System Economy	• Low per-channel cost • Lower installation cost • Silica is the principle, abundant, and inexpensive material (source is sand)

Despite the many advantages of fiber optic systems, there are some disadvantages. Because of the relative newness of the technology, fiber optic components are expensive. Fiber optic transmitters and receivers are still relatively expensive compared to electrical interfaces. The lack of standardization in the industry has also limited the acceptance of fiber optics. Many industries are more comfortable with the use of electrical systems and are reluctant to switch to fiber optics. However, industry researchers are eliminating these disadvantages.

Standards committees are addressing fiber optic part and test standardization. The cost to install fiber optic systems is falling because of an increase in the use of fiber optic technology. Published articles, conferences, and lectures on fiber optics have begun to educate managers and technicians. As the technology matures, the use of fiber optics will increase because of its many advantages over electrical systems.

Q13. List seven advantages of fiber optics over electrical systems.

SUMMARY

Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas you have learned. You should have a thorough understanding of these principles before advancing to chapter 2.

FIBER OPTICS is the branch of optical technology concerned with the transmission of radiant power (light energy) through fibers.

A **FIBER OPTIC DATA LINK** has three basic functions: to convert an electrical input signal to an optical signal, to send the optical signal over an optical fiber, and to convert the optical signal back to an electrical signal. It consists of three parts: transmitter, optical fiber, and receiver.

The **TRANSMITTER** consists of two parts, an interface circuit and a source drive circuit. The transmitter converts the electrical input signal to an optical signal by varying the current flow through the light source.

The **RECEIVER** consists of two parts, the optical detector and signal conditioning circuits. The receiver converts the optical signal exiting the fiber back into the original form of the electrical input signal.

SCATTERING, ABSORPTION, and **DISPERSION MECHANISMS** in the fiber waveguides cause the optical signal launched into the fiber to become weakened and distorted.

NOISE is any disturbance that obscures or reduces the quality of the signal.

SIGNAL LOSS is the decrease in the amount of light reaching the end of the fiber. Impurities in the fiber material cause the signal loss in optical fibers. By removing these impurities, construction of low-loss optical fibers was possible.

The **TWO BASIC TYPES OF OPTICAL FIBERS** are multimode fibers and single mode fibers.

A **LOW-LOSS MULTIMODE OPTICAL FIBER** was developed in 1970.

A **SMALL AREA LIGHT-EMITTING DIODE (LED)** was developed in 1971. This light source was suitable for low-loss coupling to optical fibers.

FIBER OPTIC SYSTEM DESIGN has centered on long-haul communications and the subscriber-loop plant. Limited work has also been done on short-distance applications and some military systems.

FIBER PERFORMANCE depends on the amount of loss and signal distortion introduced by the fiber when it is operating at a specific wavelength. Single mode fibers tend to have lower loss and produce less distortion than multimode fibers.

The **LARGER FIBER CORE** and the **HIGHER NUMERICAL APERTURE (NA)** of multimode fibers reduce the amount of loss at fiber connections.

In **MILITARY** and **SUBSCRIBER-LOOP APPLICATIONS**, system designers consider trade-offs in the following areas: fiber properties, types of connections, optical sources, and detector types.

The **ADVANTAGES** of fiber optic systems include improved system performance, immunity to electrical noise, signal security, and electrical isolation. Advantages also include reduced size and weight, environmental protection, and overall system economy.

The **DISADVANTAGES** of fiber optic systems include problems with the relative newness of the technology, the relatively expensive cost, and the lack of component and system standardization. However, these disadvantages are already being eliminated because of increased use and acceptance of fiber optic technology.

ANSWERS TO QUESTIONS Q1. THROUGH Q13.

A1. Fiber optics is the branch of optical technology concerned with the transmission of radiant power (light energy) through fibers.

A2. The basic functions of a fiber optic data link are to convert an electrical input signal to an optical signal, send the optical signal over an optical fiber, and convert the optical signal back to an electrical signal.

A3. Transmitter, optical fiber, and receiver.

A4. Scattering, absorption, and dispersion.

A5. Noise obscures or reduces the quality of the signal.

A6. Loss is the decrease in the amount of light reaching the end of the fiber.

A7. Impurities in the fiber material caused the signal loss in optical fibers. The basic fiber material did not prevent the light signal from reaching the end of the fiber.

A8. By removing the impurities from optical fiber.

A9. Multimode and single mode fibers.

A10. Single mode fiber.

A11. Larger fiber core and higher fiber numerical aperture (NA).

A12. Trade-offs in fiber properties, types of connections, optical sources, and detector types in military and subscriber-loop applications.

A13. Advantages of fiber optics are improved system performance, immunity to electrical noise, signal security, electrical isolation, reduced size and weight, environmental protection, and overall system economy.

CHAPTER 2

FIBER OPTIC CONCEPTS

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to do the following:

1. Understand the nature of light propagation.
2. Discuss the electromagnetic theory of light.
3. Describe the properties of light reflection, refraction, diffusion, and absorption.
4. Explain how optical fibers transmit light.
5. Identify the basic optical fiber material properties.
6. Describe the ray and mode theories of light propagation along an optical fiber.
7. State the difference between multimode and single mode optical fibers.
8. Explain how optical fibers attenuate and distort light signals as they travel along the optical fiber.
9. Understand the processes of light attenuation and dispersion.

FIBER OPTIC LIGHT TRANSMISSION

Fiber optics deals with the transmission of light energy through transparent fibers. How an optical fiber guides light depends on the nature of the light and the structure of the optical fiber. A light wave is a form of energy that is moved by wave motion. Wave motion can be defined as a recurring disturbance advancing through space with or without the use of a physical medium. In fiber optics, wave motion is the movement of light energy through an optical fiber. To fully understand the concept of wave motion, refer to *NEETS* Module 10—*Introduction to Wave Propagation, Transmission Lines, and Antennas*. Before we introduce the subject of light transmission through optical fibers, you must first understand the nature of light and the properties of light waves.

PROPAGATION OF LIGHT

The exact nature of light is not fully understood, although people have been studying the subject for many centuries. In the 1700s and before, experiments seemed to indicate that light was composed of particles. In the early 1800s, a physicist Thomas Young showed that light exhibited wave characteristics. Further experiments by other physicists culminated in James Clerk (pronounced Clark) Maxwell collecting the four fundamental equations that completely describe the behavior of the electromagnetic fields. James Maxwell deduced that light was simply a component of the electromagnetic spectrum. This seems to firmly establish that light is a wave. Yet, in the early 1900s, the interaction of light with semiconductor materials, called the photoelectric effect, could not be explained with electromagnetic-wave theory. The advent of quantum physics successfully explained the photoelectric effect in terms of fundamental particles of energy called **quanta**. Quanta are known as **photons** when referring to light energy.

Today, when studying light that consists of many photons, as in propagation, that light behaves as a continuum—an electromagnetic wave. On the other hand, when studying the interaction of light with semiconductors, as in sources and detectors, the quantum physics approach is taken. The wave versus particle dilemma can be addressed in a more formal way, but that is beyond the scope of this text. It suffices to say that much has been reconciled between the two using quantum physics. In this manual, we use both the electromagnetic wave and photon concepts, each in the places where it best matches the phenomenon we are studying.

The electromagnetic energy of light is a form of electromagnetic radiation. Light and similar forms of radiation are made up of moving electric and magnetic forces. A simple example of motion similar to these radiation waves can be made by dropping a pebble into a pool of water. In this example, the water is not actually being moved by the outward motion of the wave, but rather by the up-and-down motion of the water. The up-and-down motion is transverse, or at right angles, to the outward motion of the waves. This type of wave motion is called **transverse-wave motion**. The transverse waves spread out in expanding circles until they reach the edge of the pool, in much the same manner as the transverse waves of light spread from the sun. However, the waves in the pool are very slow and clumsy in comparison with light, which travels approximately 186,000 miles per second.

Light radiates from its source in all directions until it is absorbed or diverted by some substance (fig. 2-1). The lines drawn from the light source (a light bulb in this instance) to any point on one of the transverse waves indicate the direction that the wavefronts are moving. These lines, are called **light rays**.

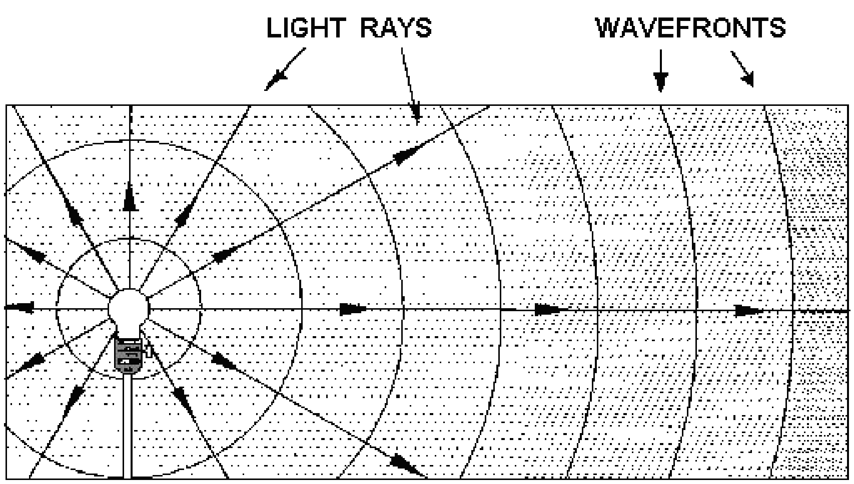


Figure 2-1.—Light rays and wavefronts from a nearby light source.

Although single rays of light typically do not exist, light rays shown in illustrations are a convenient method used to show the direction in which light is traveling at any point. A ray of light can be illustrated as a straight line.

Q1. Quantum physics successfully explained the photoelectric effect in terms of fundamental particles of energy called quanta. What are the fundamental particles of energy (quanta) known as when referring to light energy?

Q2. What type of wave motion is represented by the motion of water?

PROPERTIES OF LIGHT

When light waves, which travel in straight lines, encounter any substance, they are either reflected, absorbed, transmitted, or refracted. This is illustrated in figure 2-2. Those substances that transmit almost all the light waves falling upon them are said to be **transparent**. A transparent substance is one through which you can see clearly. Clear glass is transparent because it transmits light rays without diffusing them (view A of figure 2-3). There is no substance known that is perfectly transparent, but many substances are nearly so. Substances through which some light rays can pass, but through which objects cannot be seen clearly because the rays are diffused, are called **translucent** (view B of figure 2-3). The frosted glass of a light bulb and a piece of oiled paper are examples of translucent materials. Those substances that are unable to transmit any light rays are called **opaque** (view C of figure 2-3). Opaque substances either reflect or absorb all the light rays that fall upon them.

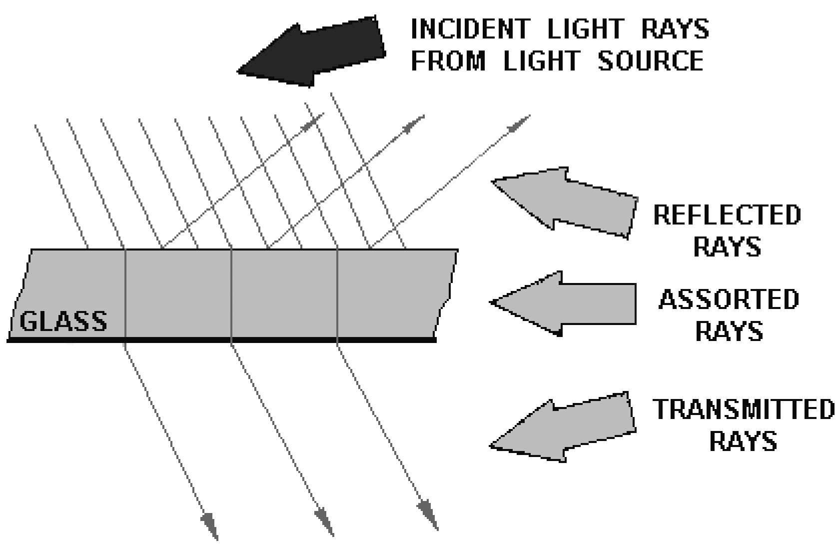


Figure 2-2.—Light waves reflected, absorbed, and transmitted.

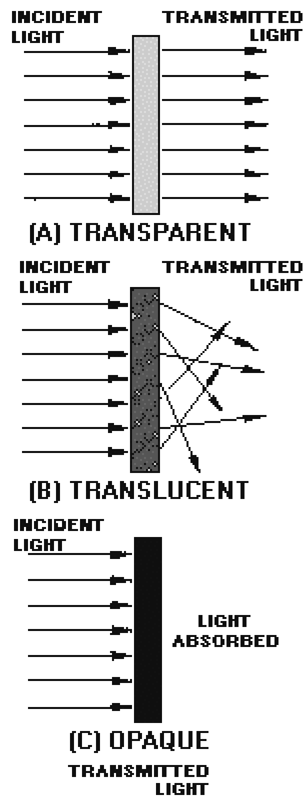


Figure 2-3.—Substances: A. Transparent; B. Translucent; and C. Opaque.

All substances that are not light sources are visible only because they reflect all or some part of the light reaching them from some luminous source. Examples of luminous sources include the sun, a gas flame, and an electric light filament, because they are sources of light energy. If light is neither transmitted nor reflected, it is absorbed or taken up by the medium. When light strikes a substance, some absorption and some reflection always take place. No substance completely transmits, reflects, or absorbs all the light rays that reach its surface.

Q3. When light waves encounter any substance, what four things can happen?

Q4. A substance that transmits almost all of the light waves falling upon it is known as what type of substance?

Q5. A substance that is unable to transmit any light waves is known as what type of substance?

REFLECTION OF LIGHT

Reflected waves are simply those waves that are neither transmitted nor absorbed, but are reflected from the surface of the medium they encounter. When a wave approaches a reflecting surface, such as a mirror, the wave that strikes the surface is called the **incident** wave, and the one that bounces back is called the **reflected** wave (refer to figure 2-4). An imaginary line perpendicular to the point at which the incident wave strikes the reflecting surface is called the **normal**, or the perpendicular. The angle between the incident wave and the normal is called the **angle of incidence**. The angle between the reflected wave and the normal is called the **angle of reflection**.

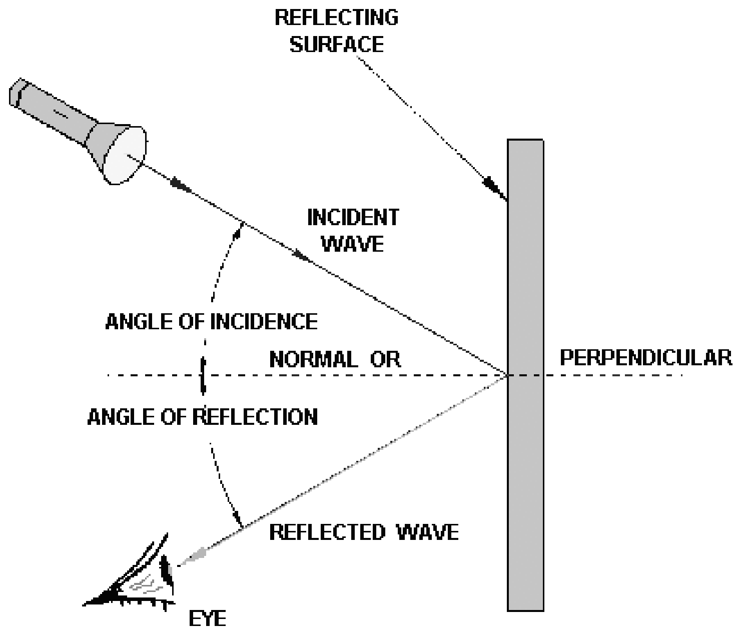


Figure 2-4.—Reflection of a wave.

If the surface of the medium contacted by the incident wave is smooth and polished, each reflected wave will be reflected back at the same angle as the incident wave. The path of the wave reflected from the surface forms an angle equal to the one formed by its path in reaching the medium. This conforms to the **law of reflection** which states: <u>The angle of incidence is equal to the angle of reflection</u>.

The amount of incident-wave energy that is reflected from a surface depends on the nature of the surface and the angle at which the wave strikes the surface. The amount of wave energy reflected increases as the angle of incidence increases. The reflection of energy is the greatest when the wave is nearly parallel to the reflecting surface. When the incidence wave is perpendicular to the surface, more of the energy is transmitted into the substance and reflection of energy is at its least. At any incident angle, a mirror reflects almost all of the wave energy, while a dull, black surface reflects very little.

Light waves obey the law of reflection. Light travels in a straight line through a substance of uniform density. For example, you can see the straight path of light rays admitted through a narrow slit into a darkened room. The straight path of the beam is made visible by illuminated dust particles suspended in the air. If the light is made to fall onto the surface of a mirror or other reflecting surface, however, the direction of the beam changes sharply. The light can be reflected in almost any direction, depending on the angle with which the mirror is held.

Q6. What is the law of reflection?

Q7. When a wave is reflected from a surface, energy is reflected. When is the reflection of energy the greatest?

Q8. When is the reflection energy the least?

Q9. Light waves obey what law?

REFRACTION OF LIGHT

When a light wave passes from one medium into a medium having a different velocity of propagation (the speed waves can travel through a medium), a change in the direction of the wave will occur. This change of direction as the wave enters the second medium is called refraction. As in the discussion of reflection, the wave striking the boundary (surface) is called the incident wave, and the imaginary line perpendicular to the boundary is called the normal. The angle between the incident wave and the normal is called the angle of incidence. As the wave passes through the boundary, it is bent either toward or away from the normal. The angle between the normal and the path of the wave through the second medium is the angle of refraction.

A light wave passing through a block of glass is shown in figure 2-5. The wave moves from point A to point B at a constant speed. This is the incident wave. As the wave penetrates the glass boundary at point B, the velocity of the wave is slowed down. This causes the wave to bend toward the normal. The wave then takes the path from point B to point C through the glass and becomes both the refracted wave from the top surface and the incident wave to the lower surface. As the wave passes from the glass to the air (the second boundary), it is again refracted, this time away from the normal, and takes the path from point C to point D. After passing through the last boundary, the velocity increases to the original velocity of the wave. As illustrated, refracted waves can bend toward or away from the normal. This bending depends on the velocity of the wave through different mediums. The broken line between points B and E is the path that the wave would travel if the two mediums (air and glass) had the same density.

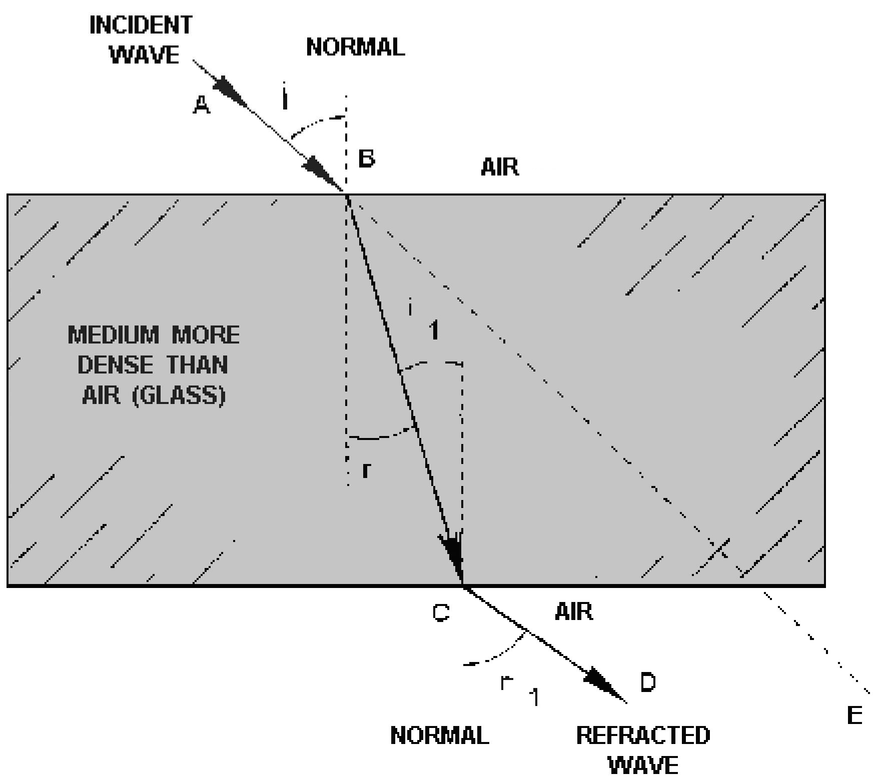


Figure 2-5.—Refraction of a wave.

Another interesting condition can be shown using figure 2-5. If the wave passes from a less dense to a more dense medium, it is bent toward the normal, and the angle of refraction (r) is less than the angle of incidence (i). Likewise, if the wave passes from a more dense to a less dense medium, it is bent away from the normal, and the angle of refraction (r_1) is greater than the angle of incidence (i_1).

An example of refraction is the apparent bending of a spoon when it is immersed in a cup of water. The bending seems to take place at the surface of the water, or exactly at the point where there is a change of density. Obviously, the spoon does not bend from the pressure of the water. The light forming the image of the spoon is bent as it passes from the water (a medium of high density) to the air (a medium of comparatively low density).

Without refraction, light waves would pass in straight lines through transparent substances without any change of direction. Figure 2-5 shows that rays striking the glass at any angle other than perpendicular are refracted. However, perpendicular rays, which enter the glass normal to the surface, continue through the glass and into the air in a straight line—no refraction takes place.

Q10. A refracted wave occurs when a wave passes from one medium into another medium. What determines the angle of refraction?

Q11. A light wave enters a sheet of glass at a perfect right angle to the surface. Is the majority of the wave reflected, refracted, transmitted, or absorbed?

DIFFUSION OF LIGHT

When light is reflected from a mirror, the angle of reflection equals the angle of incidence. When light is reflected from a piece of plain white paper; however, the reflected beam is scattered, or diffused, as shown in figure 2-6. Because the surface of the paper is not smooth, the reflected light is broken up into many light beams that are reflected in all directions.

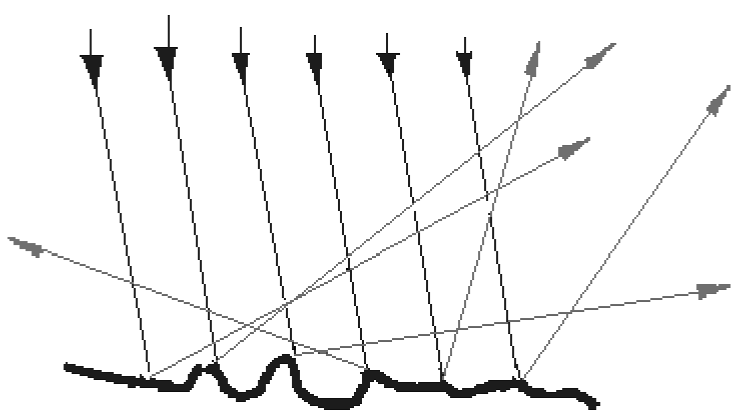


Figure 2-6.—Diffusion of light.

Q12. When light strikes a piece of white paper, the light is reflected in all directions. What do we call this scattering of light?

ABSORPTION OF LIGHT

You have just seen that a light beam is reflected and diffused when it falls onto a piece of white paper. If the light beam falls onto a piece of black paper, the black paper absorbs most of the light rays and very little light is reflected from the paper. If the surface upon which the light beam falls is perfectly black, there is no reflection; that is, the light is totally absorbed. No matter what kind of surface light falls upon, some of the light is absorbed.

TRANSMISSION OF LIGHT THROUGH OPTICAL FIBERS

The transmission of light along optical fibers depends not only on the nature of light, but also on the structure of the optical fiber. Two methods are used to describe how light is transmitted along the optical fiber. The first method, ray theory, uses the concepts of light reflection and refraction. The second method, mode theory, treats light as electromagnetic waves. You must first understand the basic optical properties of the materials used to make optical fibers. These properties affect how light is transmitted through the fiber.

Q13. Two methods describe how light propagates along an optical fiber. These methods define two theories of light propagation. What do we call these two theories?

BASIC OPTICAL-MATERIAL PROPERTIES

The basic optical property of a material, relevant to optical fibers, is the index of refraction. The index of refraction (n) measures the speed of light in an optical medium. The index of refraction of a material is the ratio of the speed of light in a vacuum to the speed of light in the material itself. The speed of light (c) in free space (vacuum) is 3×10^8 meters per second (m/s). The speed of light is the frequency (f) of light multiplied by the wavelength of light (λ). When light enters the fiber material (an optically dense medium), the light travels slower at a speed (v). Light will always travel slower in the fiber material than in air. The index of refraction is given by:

$$n = \frac{c}{v}$$

A light ray is reflected and refracted when it encounters the boundary between two different transparent mediums. For example, figure 2-7 shows what happens to the light ray when it encounters the interface between glass and air. The index of refraction for glass (n_1) is 1.50. The index of refraction for air (n_2) is 1.00.

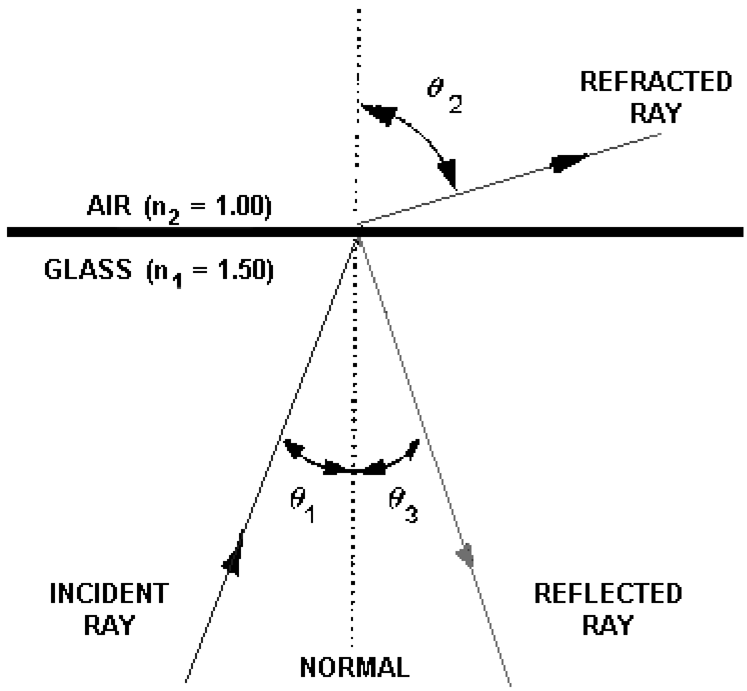


Figure 2-7.—Light reflection and refraction at a glass-air boundary.

Let's assume the light ray or incident ray is traveling through the glass. When the light ray encounters the glass-air boundary, there are two results. The first result is that part of the ray is reflected back into the glass. The second result is that part of the ray is refracted (bent) as it enters the air. The bending of the light at the glass-air interface is the result of the difference between the index of refractions. Since n_1 is greater than n_2, the angle of refraction (¯ $_2$) will be greater than the angle of incidence (¯ $_1$). Snell's law of refraction is used to describe the relationship between the incident and the refracted rays at the boundary. Snell's Law is given by:

$$n_1 \times \sin\Theta_1 = n_2 \times \sin\Theta_2$$

As the angle of incidence (¯ $_1$) becomes larger, the angle of refraction (¯ $_2$) approaches 90 degrees. At this point, no refraction is possible. The light ray is totally reflected back into the glass medium. No light escapes into the air. This condition is called total internal reflection. The angle at which total internal reflection occurs is called the critical angle of incidence. The critical angle of incidence (¯) is shown in figure 2-8. At any angle of incidence (¯ $_1$) greater than the critical angle, light is totally reflected back into the glass medium. The critical angle of incidence is determined by using Snell's Law. The critical angle is given by:

$$\sin\Theta_c = \frac{n_2}{n_1}$$

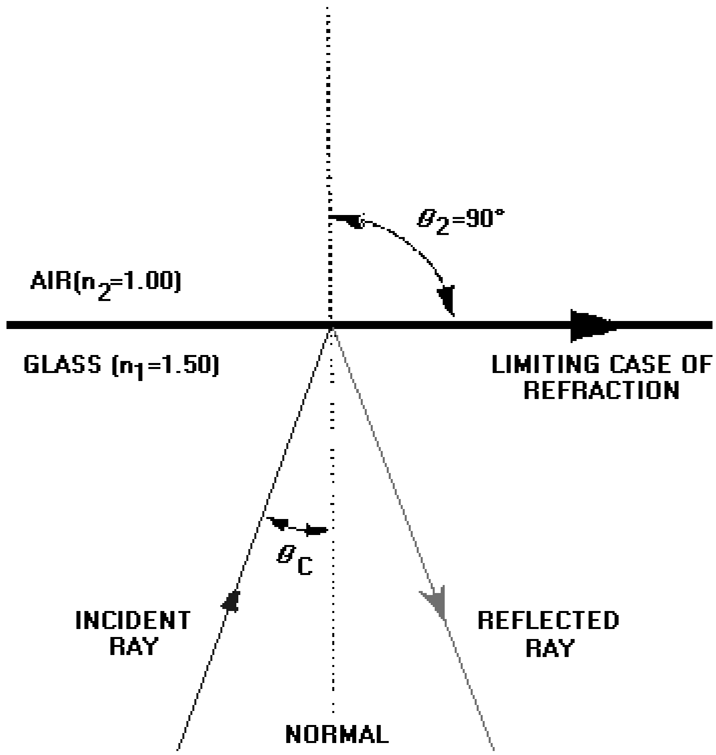


Figure 2-8.—Critical angle of incidence.

The condition of total internal reflection is an ideal situation. However, in reality, there is always some light energy that penetrates the boundary. This situation is explained by the mode theory, or the electromagnetic wave theory, of light.

Q14. What is the basic optical-material property relevant to optical fiber light transmission?

Q15. The index of refraction measures the speed of light in an optical fiber. Will light travel faster in an optically dense material or in one that is less dense?

Q16. Assume light is traveling through glass, what happens when this light strikes the glass-air boundary?

Q17. What condition causes a light ray to be totally reflected back into its medium of propagation?

Q18. What name is given to the angle where total internal reflection occurs?

BASIC STRUCTURE OF AN OPTICAL FIBER

The basic structure of an optical fiber consists of three parts; the core, the cladding, and the coating or buffer. The basic structure of an optical fiber is shown in figure 2-9. The core is a cylindrical rod of dielectric material. Dielectric material conducts no electricity. Light propagates mainly along the core of the fiber. The core is generally made of glass. The core is described as having a radius of (a) and an index of refraction n_1. The core is surrounded by a layer of material called the cladding. Even though light will propagate along the fiber core without the layer of cladding material, the cladding does perform some necessary functions.

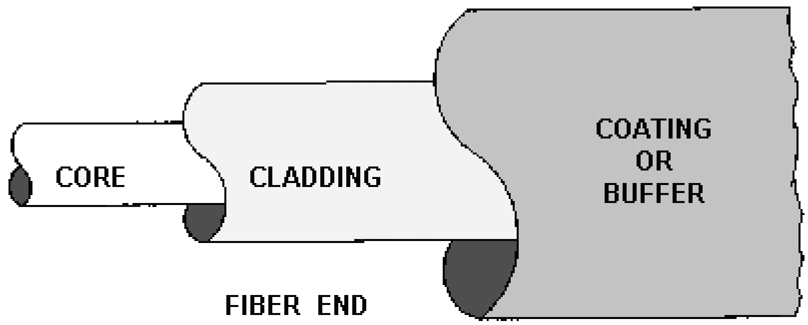


Figure 2-9.—Basic structure of an optical fiber.

The cladding layer is made of a dielectric material with an index of refraction n_2. The index of refraction of the cladding material is less than that of the core material. The cladding is generally made of glass or plastic. The cladding performs the following functions:

- Reduces loss of light from the core into the surrounding air
- Reduces scattering loss at the surface of the core
- Protects the fiber from absorbing surface contaminants
- Adds mechanical strength

For extra protection, the cladding is enclosed in an additional layer called the coating or buffer.

The coating or buffer is a layer of material used to protect an optical fiber from physical damage. The material used for a buffer is a type of plastic. The buffer is elastic in nature and prevents abrasions. The buffer also prevents the optical fiber from scattering losses caused by microbends. Microbends occur when an optical fiber is placed on a rough and distorted surface. Microbends are discussed later in this chapter.

Q19. List the three parts of an optical fiber.

Q20. Which fiber material, core or cladding, has a higher index of refraction?

PROPAGATION OF LIGHT ALONG A FIBER

The concept of light propagation, the transmission of light along an optical fiber, can be described by two theories. According to the first theory, light is described as a simple ray. This theory is the ray theory, or geometrical optics, approach. The advantage of the ray approach is that you get a clearer picture of the propagation of light along a fiber. The ray theory is used to approximate the light acceptance and guiding properties of optical fibers. According to the second theory, light is described as an electromagnetic wave. This theory is the mode theory, or wave representation, approach. The mode theory describes the behavior of light within an optical fiber. The mode theory is useful in describing the optical fiber properties of absorption, attenuation, and dispersion. These fiber properties are discussed later in this chapter.

Q21. Light transmission along an optical fiber is described by two theories. Which theory is used to approximate the light acceptance and guiding properties of an optical fiber?

Ray Theory

Two types of rays can propagate along an optical fiber. The first type is called meridional rays. Meridional rays are rays that pass through the axis of the optical fiber. Meridional rays are used to illustrate the basic transmission properties of optical fibers. The second type is called skew rays. Skew rays are rays that travel through an optical fiber without passing through its axis.

MERIDIONAL RAYS.—Meridional rays can be classified as bound or unbound rays. Bound rays remain in the core and propagate along the axis of the fiber. Bound rays propagate through the fiber by total internal reflection. Unbound rays are refracted out of the fiber core. Figure 2-10 shows a possible path taken by bound and unbound rays in a step-index fiber. The core of the step-index fiber has an index of refraction n_1. The cladding of a step-index has an index of refraction n_2, that is lower than n_1. Figure 2-10 assumes the core-cladding interface is perfect. However, imperfections at the core-cladding interface will cause part of the bound rays to be refracted out of the core into the cladding. The light rays refracted into the cladding will eventually escape from the fiber. In general, meridional rays follow the laws of reflection and refraction.

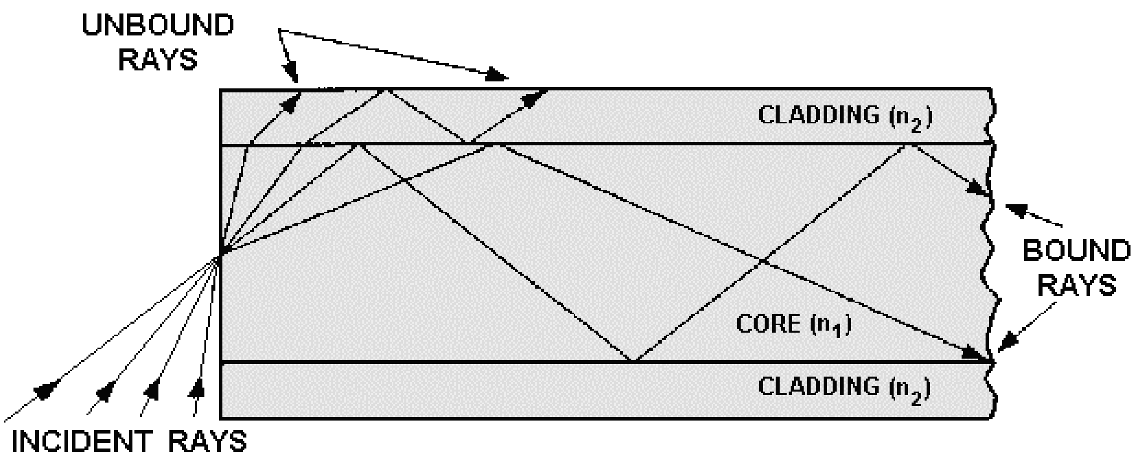


Figure 2-10.—Bound and unbound rays in a step-index fiber.

It is known that bound rays propagate in fibers due to total internal reflection, but how do these light rays enter the fiber? Rays that enter the fiber must intersect the core-cladding interface at an angle greater than the critical angle (θ_c). Only those rays that enter the fiber and strike the interface at these angles will propagate along the fiber.

How a light ray is launched into a fiber is shown in figure 2-11. The incident ray I_1 enters the fiber at the angle θ_a. I_1 is refracted upon entering the fiber and is transmitted to the core-cladding interface. The ray then strikes the core-cladding interface at the critical angle (θ_c). I_1 is totally reflected back into the core and continues to propagate along the fiber. The incident ray I_2 enters the fiber at an angle greater than θ_a. Again, I_2 is refracted upon entering the fiber and is transmitted to the core-cladding interface. I_2 strikes the core-cladding interface at an angle less than the critical angle (θ_c). I_2 is refracted into the cladding and is eventually lost. The light ray incident on the fiber core must be within the acceptance cone defined by the angle θ_a shown in figure 2-12. Angle θ_a is defined as the acceptance angle. The acceptance angle (θ_a) is the maximum angle to the axis of the fiber that light entering the fiber is propagated. The value of the angle of acceptance (¯ $_a$) depends on fiber properties and transmission conditions.

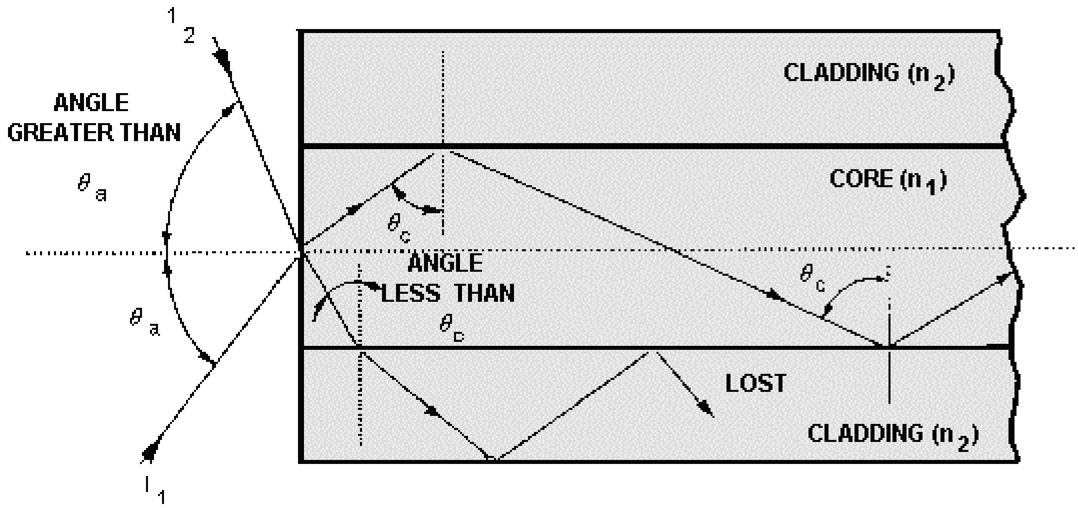


Figure 2-11.—How a light ray enters an optical fiber.

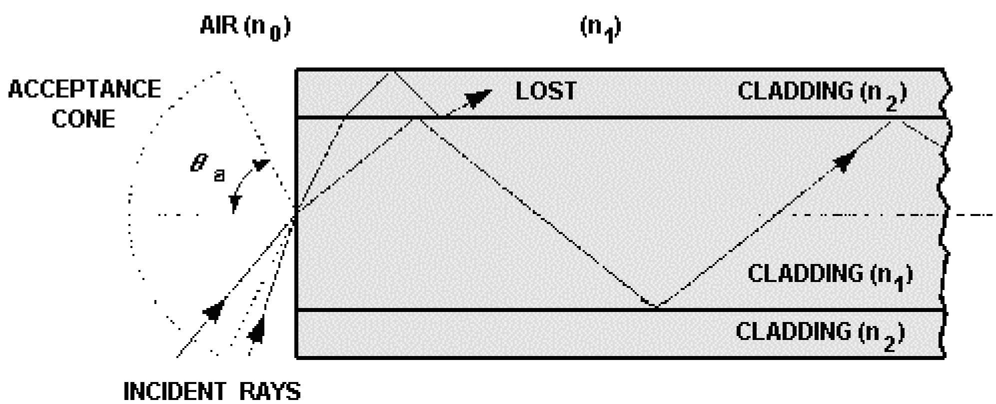


Figure 2-12.—Fiber acceptance angle.

The acceptance angle is related to the refractive indices of the core, cladding, and medium surrounding the fiber. This relationship is called the numerical aperture of the fiber. The numerical aperture (NA) is a measurement of the ability of an optical fiber to capture light. The NA is also used to define the acceptance cone of an optical fiber.

Figure 2-12 illustrates the relationship between the acceptance angle and the refractive indices. The index of refraction of the fiber core is n_1. The index of refraction of the fiber cladding is n_2. The index of refraction of the surrounding medium is n_0. By using Snell's law and basic trigonometric relationships, the NA of the fiber is given by:

$$NA = n_0 \times \sin\Theta_a = (n_1^2 - n_2^2)^{\frac{1}{2}}$$

Since the medium next to the fiber at the launching point is normally air, n_0 is equal to 1.00. The NA is then simply equal to sin ¯ $_a$.

The NA is a convenient way to measure the light-gathering ability of an optical fiber. It is used to measure source-to-fiber power-coupling efficiencies. A high NA indicates a high source-to-fiber coupling efficiency. Source-to-fiber coupling efficiency is described in chapter 6. Typical values of NA range from 0.20 to 0.29 for glass fibers. Plastic fibers generally have a higher NA. An NA for plastic fibers can be higher than 0.50.

In addition, the NA is commonly used to specify multimode fibers. However, for small core diameters, such as in single mode fibers, the ray theory breaks down. Ray theory describes only the direction a plane wave takes in a fiber. Ray theory eliminates any properties of the plane wave that interfere with the transmission of light along a fiber. In reality, plane waves interfere with each other. Therefore, only certain types of rays are able to propagate in an optical fiber. Optical fibers can support only a specific number of guided modes. In small core fibers, the number of modes supported is one or only a few modes. Mode theory is used to describe the types of plane waves able to propagate along an optical fiber.

SKEW RAYS.—A possible path of propagation of skew rays is shown in figure 2-13. Figure 2-13, view A, provides an angled view and view B provides a front view. Skew rays propagate without passing through the center axis of the fiber. The acceptance angle for skew rays is larger than the acceptance angle of meridional rays. This condition explains why skew rays outnumber meridional rays. Skew rays are often used in the calculation of light acceptance in an optical fiber. The addition of skew rays increases the amount of light capacity of a fiber. In large NA fibers, the increase may be significant.

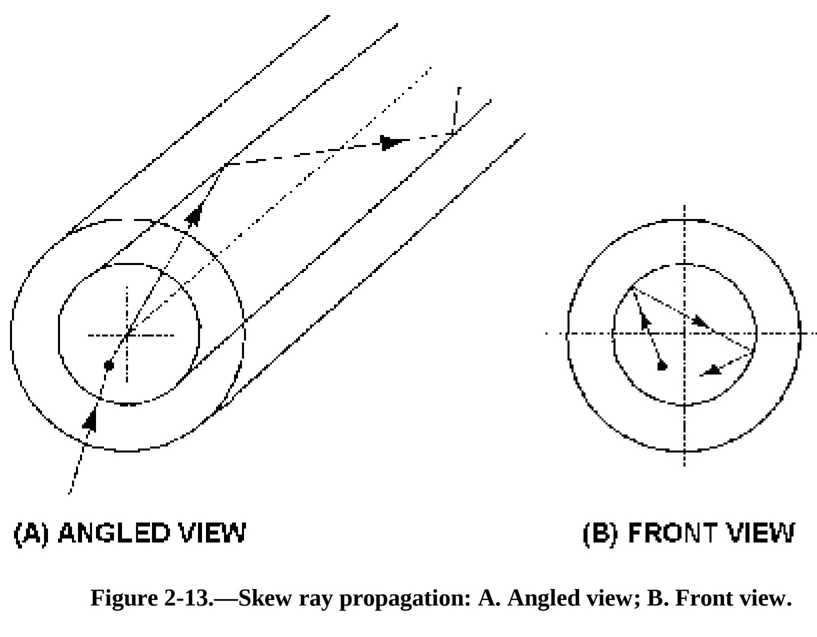


Figure 2-13.—Skew ray propagation: A. Angled view; B. Front view.

The addition of skew rays also increases the amount of loss in a fiber. Skew rays tend to propagate near the edge of the fiber core. A large portion of the number of skew rays that are trapped in the fiber core are considered to be leaky rays. Leaky rays are predicted to be totally reflected at the core-cladding boundary. However, these rays are partially refracted because of the curved nature of the fiber boundary. Mode theory is also used to describe this type of leaky ray loss.

Q22. Meridional rays are classified as either bound or unbound rays. Bound rays propagate through the fiber according to what property?

Q23. A light ray incident on the optical fiber core is propagated along the fiber. Is the angle of incidence of the light ray entering the fiber larger or smaller than the acceptance angle (Θ_a)

Q24. What fiber property does numerical aperture (NA) measure?

Q25. Skew rays and meridional rays define different acceptance angles. Which acceptance angle is larger, the skew ray angle or the meridional ray angle?

Mode Theory

The mode theory, along with the ray theory, is used to describe the propagation of light along an optical fiber. The mode theory is used to describe the properties of light that ray theory is unable to explain. The mode theory uses electromagnetic wave behavior to describe the propagation of light along a fiber. A set of guided electromagnetic waves is called the modes of the fiber.

Q26. The mode theory uses electromagnetic wave behavior to describe the propagation of the light along the fiber. What is a set of guided electromagnetic waves called?

PLANE WAVES.—The mode theory suggests that a light wave can be represented as a plane wave. A plane wave is described by its direction, amplitude, and wavelength of propagation. A plane wave is a wave whose surfaces of constant phase are infinite parallel planes normal to the direction of propagation. The planes having the same phase are called the wavefronts. The wavelength (λ) of the plane wave is given by:

$$\text{wavelength } (\lambda) = \frac{c}{fn}$$

where *c* is the speed of light in a vacuum, *f* is the frequency of the light, and n is the index of refraction of the plane-wave medium.

Figure 2-14 shows the direction and wavefronts of plane-wave propagation. Plane waves, or wavefronts, propagate along the fiber similar to light rays. However, not all wavefronts incident on the fiber at angles less than or equal to the critical angle of light acceptance propagate along the fiber. Wavefronts may undergo a change in phase that prevents the successful transfer of light along the fiber.

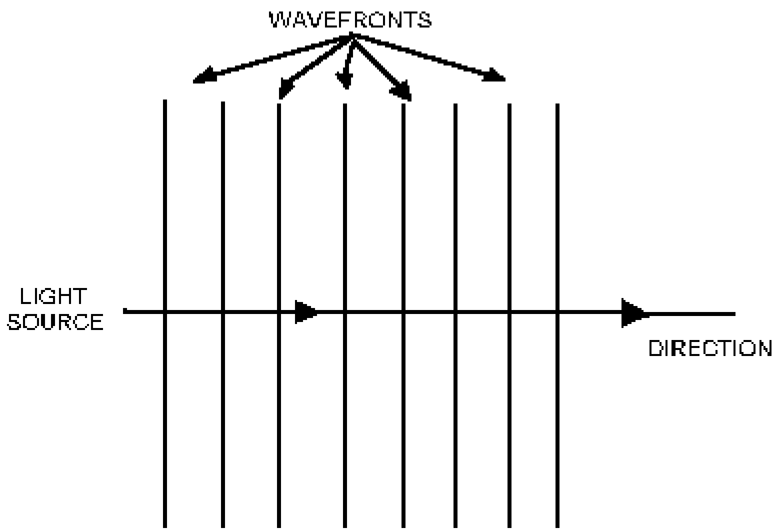


Figure 2-14.—Plane-wave propagation.

Wavefronts are required to remain in phase for light to be transmitted along the fiber. Consider the wavefront incident on the core of an optical fiber as shown in figure 2-15. Only those wavefronts incident on the fiber at angles less than or equal to the critical angle may propagate along the fiber. The wavefront undergoes a gradual phase change as it travels down the fiber. Phase changes also occur when the wavefront is reflected. The wavefront must remain in phase after the wavefront transverses the fiber twice and is reflected twice. The distance transversed is shown between point A and point B on figure 2-15. The reflected waves at point A and point B are in phase if the total amount of phase collected is an integer multiple of 2π radian. If propagating wavefronts are not in phase, they eventually disappear. Wavefronts disappear because of destructive interference. The wavefronts that are in phase interfere with the wavefronts that are out of phase. This interference is the reason why only a finite number of modes can propagate along the fiber.

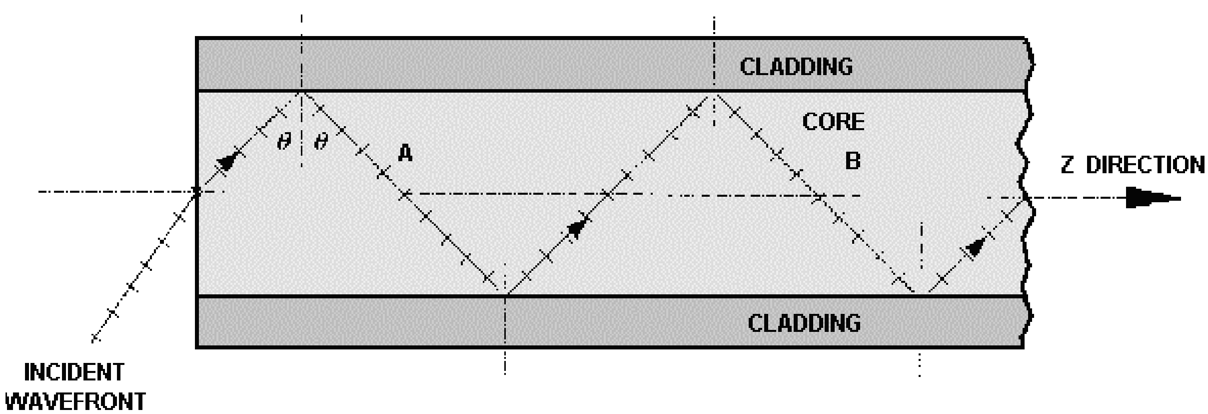


Figure 2-15.—Wavefront propagation along an optical fiber.

The plane waves repeat as they travel along the fiber axis. The direction the plane waves travel is assumed to be the z direction as shown in figure 2-15. The plane waves repeat at a distance equal to λ/sin¯ . Plane waves also repeat at a periodic frequency $\beta = 2\pi$ sin ¯ /λ. The quantity β is defined as the propagation constant along the fiber axis. As the wavelength (λ) changes, the value of the propagation constant must also change. For a given mode, a change in wavelength can prevent the mode from propagating along the fiber. The mode is no longer bound to the fiber. The mode is said to be cut off. Modes that are bound at one wavelength may not exist at longer wavelengths. The wavelength at which a mode ceases to be bound is called the cutoff wavelength for that mode. However, an optical fiber is always able to propagate at least one mode. This mode is referred to as the fundamental mode of the fiber. The fundamental mode can never be cut off. The wavelength that prevents the next higher mode from propagating is called the cutoff wavelength of the fiber. An optical fiber that operates above the cutoff wavelength (at a longer wavelength) is called a single mode fiber. An optical fiber that operates below the cutoff wavelength is called a multimode fiber. Single mode and multimode optical fibers are discussed later in this chapter.

In a fiber, the propagation constant of a plane wave is a function of the wave's wavelength and mode. The change in the propagation constant for different waves is called dispersion. The change in the propagation constant for different wavelengths is called chromatic dispersion. The change in propagation constant for different modes is called modal dispersion. These dispersions cause the light pulse to spread as it goes down the fiber (fig. 2-16). Some dispersion occurs in all types of fibers. Dispersion is discussed later in this chapter.

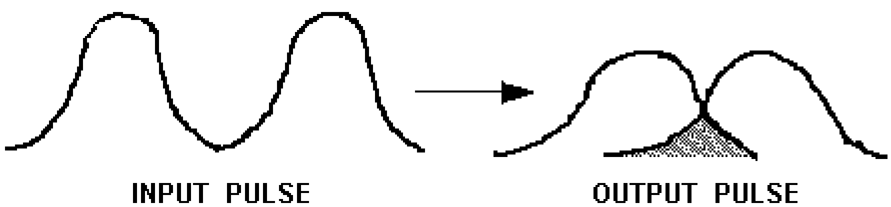


Figure 2-16.—The spreading of a light pulse.

MODES.—A set of guided electromagnetic waves is called the modes of an optical fiber. Maxwell's equations describe electromagnetic waves or modes as having two components. The two components are the electric field, E(x, y, z), and the magnetic field, H(x, y, z). The electric field, E, and the magnetic field, H, are at right angles to each other. Modes traveling in an optical fiber are said to be transverse. The transverse modes, shown in figure 2-17, propagate along the axis of the fiber. The mode field patterns shown in figure 2-17 are said to be transverse electric (TE). In TE modes, the electric field is perpendicular to the direction of propagation. The magnetic field is in the direction of propagation. Another type of transverse mode is the transverse magnetic (TM) mode. TM modes are opposite to TE modes. In TM modes, the magnetic field is perpendicular to the direction of propagation. The electric field is in the direction of propagation. Figure 2-17 shows only TE modes.

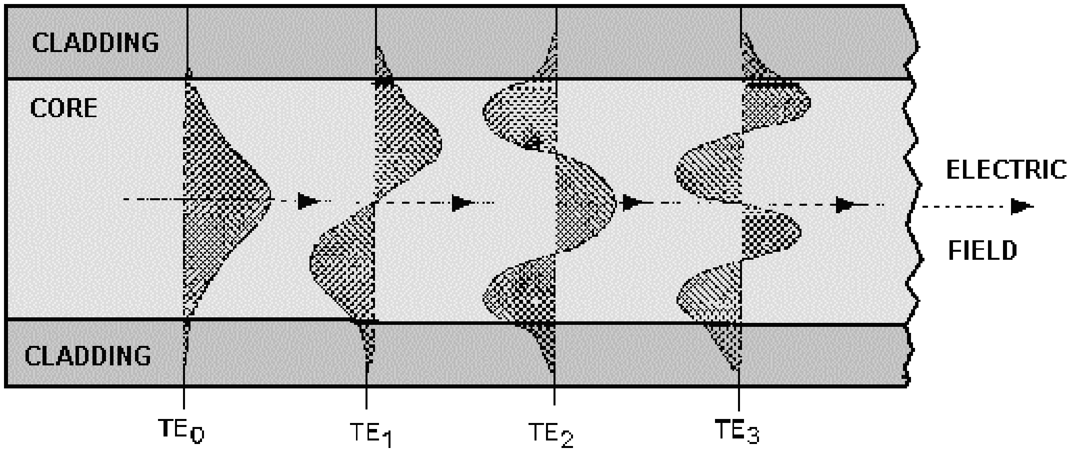


Figure 2-17.—Transverse electric (TE) mode field patterns.

The TE mode field patterns shown in figure 2-17 indicate the order of each mode. The order of each mode is indicated by the number of field maxima within the core of the fiber. For example, TE_0 has one field maxima. The electric field is a maximum at the center of the waveguide and decays toward the core-cladding boundary. TE_0 is considered the fundamental mode or the lowest order standing wave. As the number of field maxima increases, the order of the mode is higher. Generally, modes with more than a few (5-10) field maxima are referred to as high-order modes.

The order of the mode is also determined by the angle the wavefront makes with the axis of the fiber. Figure 2-18 illustrates light rays as they travel down the fiber. These light rays indicate the direction of the wavefronts. High-order modes cross the axis of the fiber at steeper angles. Low-order and high-order modes are shown in figure 2-18.

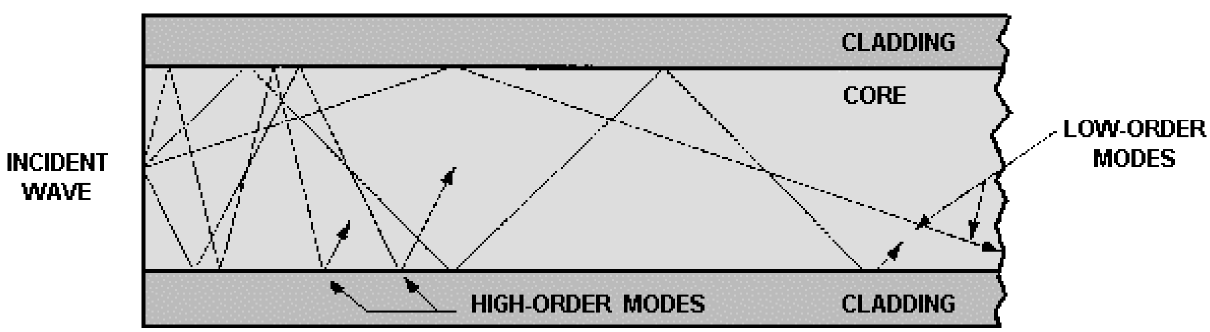


Figure 2-18.—Low-order and high-order modes.

Before we progress, let us refer back to figure 2-17. Notice that the modes are not confined to the core of the fiber. The modes extend partially into the cladding material. Low-order modes penetrate the cladding only slightly. In low-order modes, the electric and magnetic fields are concentrated near the center of the fiber. However, high-order modes penetrate further into the cladding material. In high-order modes, the electrical and magnetic fields are distributed more toward the outer edges of the fiber.

This penetration of low-order and high-order modes into the cladding region indicates that some portion is refracted out of the core. The refracted modes may become trapped in the cladding due to the dimension of the cladding region. The modes trapped in the cladding region are called cladding modes. As the core and the cladding modes travel along the fiber, mode coupling occurs. Mode coupling is the exchange of power between two modes. Mode coupling to the cladding results in the loss of power from the core modes.

In addition to bound and refracted modes, there are leaky modes. Leaky modes are similar to leaky rays. Leaky modes lose power as they propagate along the fiber. For a mode to remain within the core, the mode must meet certain boundary conditions. A mode remains bound if the propagation constant β meets the following boundary condition:

$$\frac{2\pi n_2}{\lambda} < \beta < \frac{2\pi n_1}{\lambda}$$

where n_1 and n_2 are the index of refraction for the core and the cladding, respectively. When the propagation constant becomes smaller than $2\pi n_2/\lambda$, power leaks out of the core and into the cladding. Generally, modes leaked into the cladding are lost in a few centimeters. However, leaky modes can carry a large amount of power in short fibers.

NORMALIZED FREQUENCY.—Electromagnetic waves bound to an optical fiber are described by the fiber's normalized frequency. The normalized frequency determines how many modes a fiber can support. Normalized frequency is a dimensionless quantity. Normalized frequency is also related to the fiber's cutoff wavelength. Normalized frequency (V) is defined as:

$$V = \frac{2\pi a}{\lambda}(n_1^2 - n_2^2)^{\frac{1}{2}}$$

where n_1 is the core index of refraction, n_2 is the cladding index of refraction, *a* is the core diameter, and λ is the wavelength of light in air.

The number of modes that can exist in a fiber is a function of V. As the value of V increases, the number of modes supported by the fiber increases. Optical fibers, single mode and multimode, can support a different number of modes. The number of modes supported by single mode and multimode fiber types is discussed later in this chapter.

Q27. A light wave can be represented as a plane wave. What three properties of light propagation describe a plane wave?

Q28. A wavefront undergoes a phase change as it travels along the fiber. If the wavefront transverses the fiber twice and is reflected twice and the total phase change is equal to 1/2π, will the wavefront disappear? If yes, why?

Q29. Modes that are bound at one wavelength may not exist at longer wavelengths. What is the wavelength at which a mode ceases to be bound called?

Q30. What type of optical fiber operates below the cutoff wavelength?

Q31. Low-order and high-order modes propagate along an optical fiber. How are modes determined to be low-order or high-order modes?

Q32. As the core and cladding modes travel along the fiber, mode coupling occurs. What is mode coupling?

Q33. The fiber's normalized frequency (V) determines how many modes a fiber can support. As the value of V increases, will the number of modes supported by the fiber increase or decrease?

OPTICAL FIBER TYPES

Optical fibers are characterized by their structure and by their properties of transmission. Basically, optical fibers are classified into two types. The first type is single mode fibers. The second type is multimode fibers. As each name implies, optical fibers are classified by the number of modes that propagate along the fiber. As previously explained, the structure of the fiber can permit or restrict modes from propagating in a fiber. The basic structural difference is the core size. Single mode fibers are manufactured with the same materials as multimode fibers. Single mode fibers are also manufactured by following the same fabrication process as multimode fibers.

Single Mode Fibers

The core size of single mode fibers is small. The core size (diameter) is typically around 8 to 10 micrometers (¯m). A fiber core of this size allows only the fundamental or lowest order mode to propagate around a 1300 nanometer (nm) wavelength. Single mode fibers propagate only one mode, because the core size approaches the operational wavelength (λ). The value of the normalized frequency parameter (V) relates core size with mode propagation. In single mode fibers, V is less than or equal to 2.405. When V ¯ 2.405, single mode fibers propagate the fundamental mode down the fiber core, while high-order modes are lost in the cladding. For low V values (¯1.0), most of the power is propagated in the cladding material. Power transmitted by the cladding is easily lost at fiber bends. The value of V should remain near the 2.405 level.

Single mode fibers have a lower signal loss and a higher information capacity (bandwidth) than multimode fibers. Single mode fibers are capable of transferring higher amounts of data due to low fiber dispersion. Basically, dispersion is the spreading of light as light propagates along a fiber. Dispersion mechanisms in single mode fibers are discussed in more detail later in this chapter. Signal loss depends on the operational wavelength (λ). In single mode fibers, the wavelength can increase or decrease the losses caused by fiber bending. Single mode fibers operating at wavelengths larger than the cutoff wavelength lose more power at fiber bends. They lose power because light radiates into the cladding, which is lost at fiber bends. In general, single mode fibers are considered to be low-loss fibers, which increase system bandwidth and length.

Q34. The value of the normalized frequency parameter (V) relates the core size with mode propagation. When single mode fibers propagate only the fundamental mode, what is the value of V?

Multimode Fibers

As their name implies, multimode fibers propagate more than one mode. Multimode fibers can propagate over 100 modes. The number of modes propagated depends on the core size and numerical aperture (NA). As the core size and NA increase, the number of modes increases. Typical values of fiber core size and NA are 50 to 100 ¯m and 0.20 to 0.29, respectively.

A large core size and a higher NA have several advantages. Light is launched into a multimode fiber with more ease. The higher NA and the larger core size make it easier to make fiber connections. During fiber splicing, core-to-core alignment becomes less critical. Another advantage is that multimode fibers permit the use of light-emitting diodes (LEDs). Single mode fibers typically must use laser diodes. LEDs are cheaper, less complex, and last longer. LEDs are preferred for most applications.

Multimode fibers also have some disadvantages. As the number of modes increases, the effect of modal dispersion increases. Modal dispersion (intermodal dispersion) means that modes arrive at the fiber end at slightly different times. This time difference causes the light pulse to spread. Modal dispersion affects system bandwidth. Fiber manufacturers adjust the core diameter, NA, and index profile properties of multimode fibers to maximize system bandwidth.

Q35. The number of modes propagated in a multimode fiber depends on core size and numerical aperture (NA). If the core size and the NA decrease, will the number of modes propagated increase or decrease?

Q36. Modal dispersion affects the bandwidth of multimode systems. It is essential to adjust what three fiber properties to maximize system bandwidth?

PROPERTIES OF OPTICAL FIBER TRANSMISSION

The principles behind the transfer of light along an optical fiber were discussed earlier in this chapter. You learned that propagation of light depended on the nature of light and the structure of the optical fiber. However, our discussion did not describe how optical fibers affect system performance. In this case, system performance deals with signal loss and bandwidth.

Signal loss and system bandwidth describe the amount of data transmitted over a specified length of fiber. Many optical fiber properties increase signal loss and reduce system bandwidth. The most important properties that affect system performance are fiber attenuation and dispersion.

Attenuation reduces the amount of optical power transmitted by the fiber. Attenuation controls the distance an optical signal (pulse) can travel as shown in figure 2-19. Once the power of an optical pulse is reduced to a point where the receiver is unable to detect the pulse, an error occurs. Attenuation is mainly a result of **light absorption, scattering**, and **bending losses**. Dispersion spreads the optical pulse as it travels along the fiber. This spreading of the signal pulse reduces the system bandwidth or the information-carrying capacity of the fiber. Dispersion limits how fast information is transferred as shown in figure 2-19. An error occurs when the receiver is unable to distinguish between input pulses caused by the spreading of each pulse. The effects of attenuation and dispersion increase as the pulse travels the length of the fiber as shown in figure 2-20.

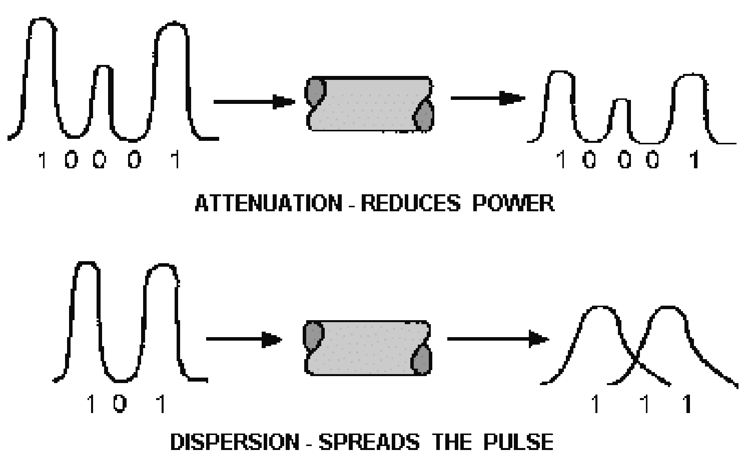


Figure 2-19.—Fiber transmission properties.

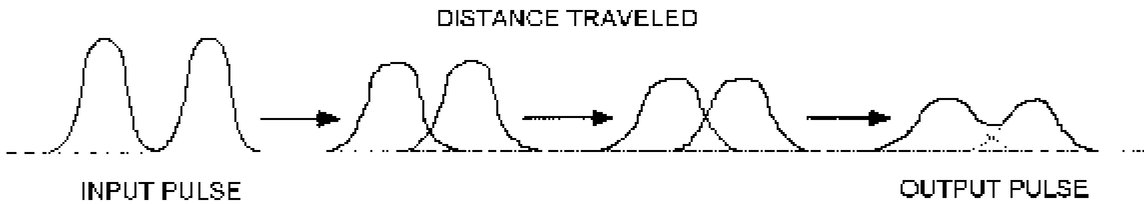


Figure 2-20.—Pulse spreading and power loss along an optical fiber.

In addition to fiber attenuation and dispersion, other optical fiber properties affect system performance. Fiber properties, such as modal noise, pulse broadening, and polarization, can reduce system performance. Modal noise, pulse broadening, and polarization are too complex to discuss as introductory level material. However, you should be aware that attenuation and dispersion are not the only fiber properties that affect performance.

Q37. Attenuation is mainly a result of what three properties?

Attenuation

Attenuation in an optical fiber is caused by absorption, scattering, and bending losses. **Attenuation** is the loss of optical power as light travels along the fiber. Signal attenuation is defined as the ratio of optical input power (P_i) to the optical output power (P_o). Optical input power is the power injected into the fiber from an optical source. Optical output power is the power received at the fiber end or optical detector. The following equation defines signal attenuation as a unit of length:

$$\text{attenuation} = \left(\frac{10}{L}\right)\log_{10}\left(\frac{P_i}{P_o}\right)$$

Signal attenuation is a log relationship. Length (L) is expressed in kilometers. Therefore, the unit of attenuation is decibels/kilometer (dB/km).

As previously stated, attenuation is caused by absorption, scattering, and bending losses. Each mechanism of loss is influenced by fiber-material properties and fiber structure. However, loss is also present at fiber connections. Fiber connector, splice, and coupler losses are discussed in chapter 4. The present discussion remains relative to optical fiber attenuation properties.

Q38. Define attenuation.

ABSORPTION.—Absorption is a major cause of signal loss in an optical fiber. **Absorption** is defined as the portion of attenuation resulting from the conversion of optical power into another energy form, such as heat. Absorption in optical fibers is explained by three factors:

- Imperfections in the atomic structure of the fiber material
- The intrinsic or basic fiber-material properties
- The extrinsic (presence of impurities) fiber-material properties

Imperfections in the atomic structure induce absorption by the presence of missing molecules or oxygen defects. Absorption is also induced by the diffusion of hydrogen molecules into the glass fiber. Since intrinsic and extrinsic material properties are the main cause of absorption, they are discussed further.

Intrinsic Absorption.—Intrinsic absorption is caused by basic fiber-material properties. If an optical fiber were absolutely pure, with no imperfections or impurities, then all absorption would be intrinsic. Intrinsic absorption sets the minimal level of absorption. In fiber optics, silica (pure glass) fibers are used predominately. Silica fibers are used because of their low intrinsic material absorption at the wavelengths of operation.

In silica glass, the wavelengths of operation range from 700 nanometers (nm) to 1600 nm. Figure 2-21 shows the level of attenuation at the wavelengths of operation. This wavelength of operation is between two intrinsic absorption regions. The first region is the ultraviolet region (below 400-nm wavelength). The second region is the infrared region (above 2000-nm wavelength).

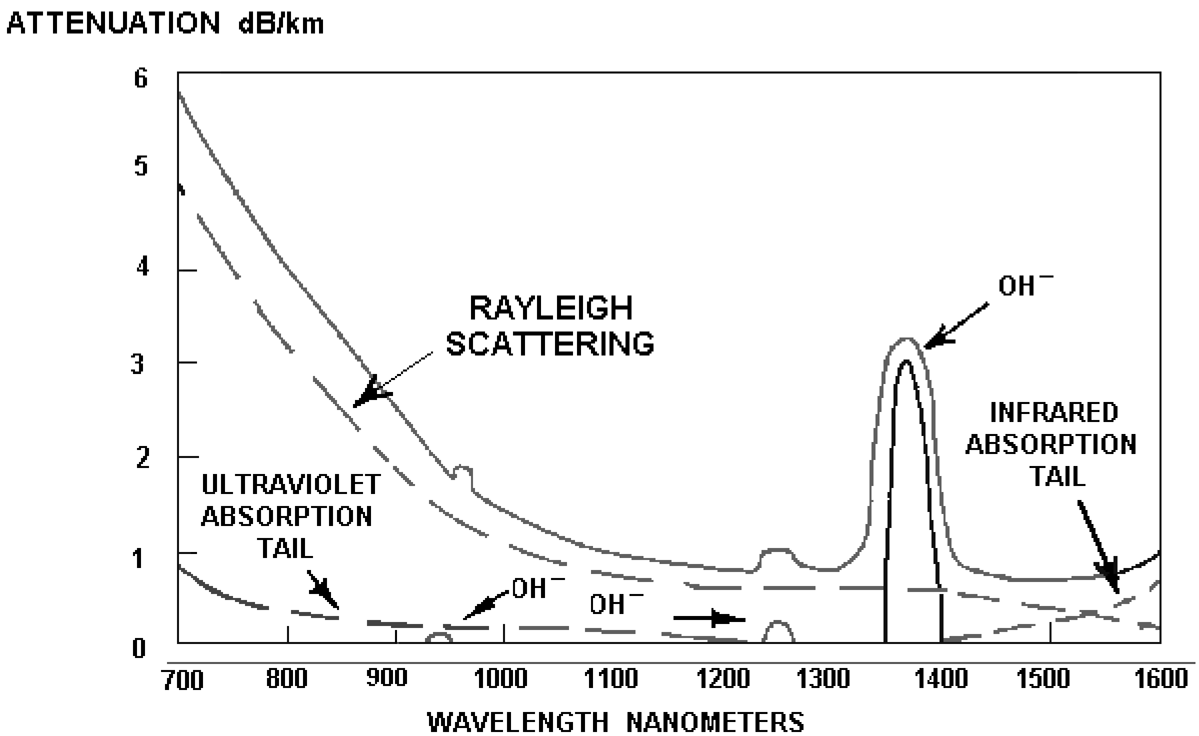


Figure 2-21.—Fiber losses.

Intrinsic absorption in the ultraviolet region is caused by electronic absorption bands. Basically, absorption occurs when a light particle (photon) interacts with an electron and excites it to a higher energy level. The tail of the ultraviolet absorption band is shown in figure 2-21.

The main cause of intrinsic absorption in the infrared region is the characteristic vibration frequency of atomic bonds. In silica glass, absorption is caused by the vibration of silicon-oxygen (Si-O) bonds. The interaction between the vibrating bond and the electromagnetic field of the optical signal causes intrinsic absorption. Light energy is transferred from the electromagnetic field to the bond. The tail of the infrared absorption band is shown in figure 2-21.

Extrinsic Absorption.—Extrinsic absorption is caused by impurities introduced into the fiber material. Trace metal impurities, such as iron, nickel, and chromium, are introduced into the fiber during fabrication. **Extrinsic absorption** is caused by the electronic transition of these metal ions from one energy level to another.

Extrinsic absorption also occurs when hydroxyl ions (OH^-) are introduced into the fiber. Water in silica glass forms a silicon-hydroxyl (Si-OH) bond. This bond has a fundamental absorption at 2700 nm. However, the harmonics or overtones of the fundamental absorption occur in the region of operation. These harmonics increase extrinsic absorption at 1383 nm, 1250 nm, and 950 nm. Figure 2-21 shows the presence of the three OH^- harmonics. The level of the OH^- harmonic absorption is also indicated.

These absorption peaks define three regions or windows of preferred operation. The first window is centered at 850 nm. The second window is centered at 1300 nm. The third window is centered at 1550 nm. Fiber optic systems operate at wavelengths defined by one of these windows.

The amount of water (OH^-) impurities present in a fiber should be less than a few parts per billion. Fiber attenuation caused by extrinsic absorption is affected by the level of impurities (OH^-) present in the fiber. If the amount of impurities in a fiber is reduced, then fiber attenuation is reduced.

Q39. What are the main causes of absorption in optical fiber?

Q40. Silica (pure glass) fibers are used because of their low intrinsic material absorption at the wavelengths of operation. This wavelength of operation is between two intrinsic absorption regions. What are these two regions called? What are the wavelengths of operation for these two regions?

Q41. Extrinsic (OH^-) absorption peaks define three regions or windows of preferred operation. List the three windows of operation.

SCATTERING.—Basically, scattering losses are caused by the interaction of light with density fluctuations within a fiber. Density changes are produced when optical fibers are manufactured. During manufacturing, regions of higher and lower molecular density areas, relative to the average density of the fiber, are created. Light traveling through the fiber interacts with the density areas as shown in figure 2-22. Light is then partially scattered in all directions.

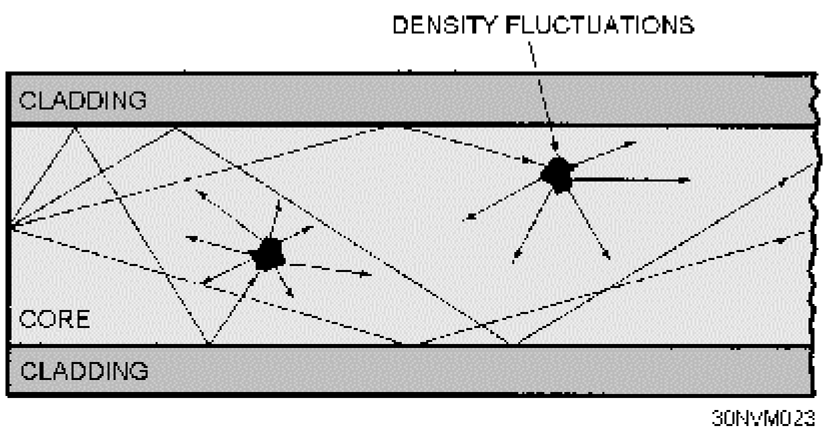


Figure 2-22.—Light scattering.

In commercial fibers operating between 700-nm and 1600-nm wavelength, the main source of loss is called Rayleigh scattering. Rayleigh scattering is the main loss mechanism between the ultraviolet and infrared regions as shown in figure 2-21. Rayleigh scattering occurs when the size of the density fluctuation (fiber defect) is less than one-tenth of the operating wavelength of light. Loss caused by Rayleigh scattering is proportional to the fourth power of the wavelength ($1/\lambda^4$). As the wavelength increases, the loss caused by Rayleigh scattering decreases.

If the size of the defect is greater than one-tenth of the wavelength of light, the scattering mechanism is called Mie scattering. Mie scattering, caused by these large defects in the fiber core, scatters light out of the fiber core. However, in commercial fibers, the effects of Mie scattering are insignificant. Optical fibers are manufactured with very few large defects.

Q42. What is the main loss mechanism between the ultraviolet and infrared absorption regions?

Q43. Scattering losses are caused by the interaction of light with density fluctuations within a fiber. What are the two scattering mechanisms called when the size of the density fluctuations is (a) greater than and (b) less than one-tenth of the operating wavelength?

BENDING LOSS.—Bending the fiber also causes attenuation. Bending loss is classified according to the bend radius of curvature: microbend loss or macrobend loss. Microbends are small microscopic bends of the fiber axis that occur mainly when a fiber is cabled. Macrobends are bends having a large radius of curvature relative to the fiber diameter. Microbend and macrobend losses are very important loss mechanisms. Fiber loss caused by microbending can still occur even if the fiber is cabled correctly. During installation, if fibers are bent too sharply, macrobend losses will occur.

Microbend losses are caused by small discontinuities or imperfections in the fiber. Uneven coating applications and improper cabling procedures increase microbend loss. External forces are also a source of microbends. An external force deforms the cabled jacket surrounding the fiber but causes only a small bend in the fiber. Microbends change the path that propagating modes take, as shown in figure 2-23. Microbend loss increases attenuation because low-order modes become coupled with high-order modes that are naturally lossy.

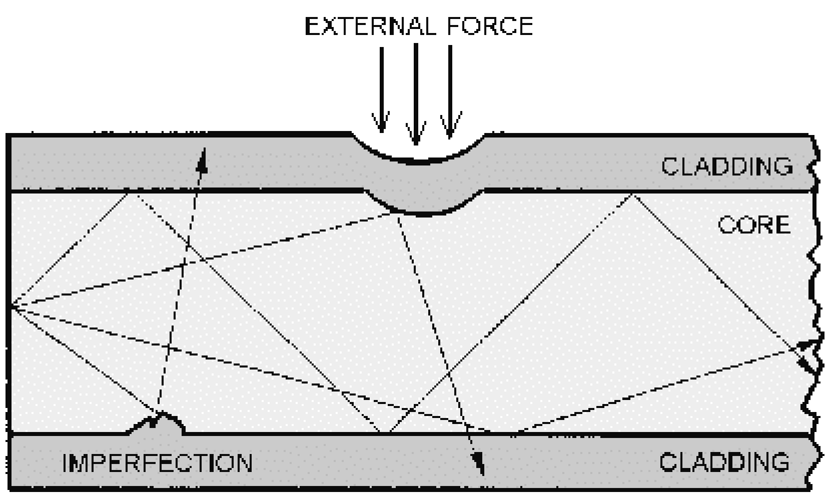


Figure 2-23.—Microbend loss.

Macrobend losses are observed when a fiber bend's radius of curvature is large compared to the fiber diameter. These bends become a great source of loss when the radius of curvature is less than several centimeters. Light propagating at the inner side of the bend travels a shorter distance than that on the outer side. To maintain the phase of the light wave, the mode phase velocity must increase. When the fiber bend is less than some critical radius, the mode phase velocity must increase to a speed greater than the speed of light. However, it is impossible to exceed the speed of light. This condition causes some of the light within the fiber to be converted to high-order modes. These high-order modes are then lost or radiated out of the fiber.

Fiber sensitivity to bending losses can be reduced. If the refractive index of the core is increased, then fiber sensitivity decreases. Sensitivity also decreases as the diameter of the overall fiber increases. However, increases in the fiber core diameter increase fiber sensitivity. Fibers with larger core size propagate more modes. These additional modes tend to be more lossy.

Q44. Microbend loss is caused by microscopic bends of the fiber axis. List three sources of microbend loss.

Q45. How is fiber sensitivity to bending losses reduced?

DISPERSION

There are two different types of dispersion in optical fibers. The types are intramodal and intermodal dispersion. Intramodal, or chromatic, dispersion occurs in all types of fibers. Intermodal, or modal, dispersion occurs only in multimode fibers. Each type of dispersion mechanism leads to pulse spreading. As a pulse spreads, energy is overlapped. This condition is shown in figure 2-24. The spreading of the optical pulse as it travels along the fiber limits the information capacity of the fiber.

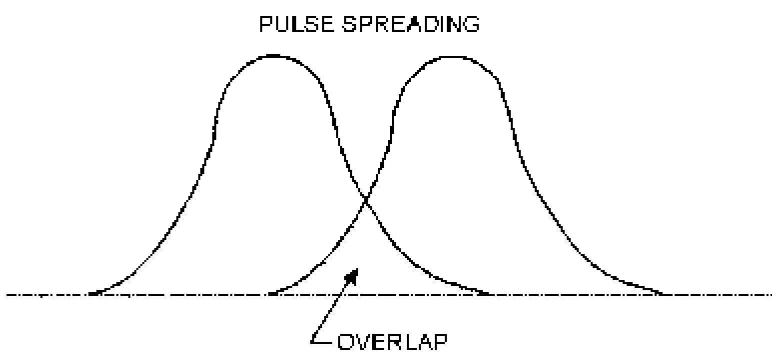


Figure 2-24.—Pulse overlap.

Intramodal Dispersion

Intramodal, or chromatic, dispersion depends primarily on fiber materials. There are two types of intramodal dispersion. The first type is material dispersion. The second type is waveguide dispersion. Intramodal dispersion occurs because different colors of light travel through different materials and different waveguide structures at different speeds.

Material dispersion occurs because the spreading of a light pulse is dependent on the wavelengths' interaction with the refractive index of the fiber core. Different wavelengths travel at different speeds in the fiber material. Different wavelengths of a light pulse that enter a fiber at one time exit the fiber at different times. Material dispersion is a function of the source spectral width. The spectral width specifies the range of wavelengths that can propagate in the fiber. Material dispersion is less at longer wavelengths.

Waveguide dispersion occurs because the mode propagation constant (β) is a function of the size of the fiber's core relative to the wavelength of operation. Waveguide dispersion also occurs because light propagates differently in the core than in the cladding.

In multimode fibers, waveguide dispersion and material dispersion are basically separate properties. Multimode waveguide dispersion is generally small compared to material dispersion. Waveguide dispersion is usually neglected. However, in single mode fibers, material and waveguide dispersion are interrelated. The total dispersion present in single mode fibers may be minimized by trading material and waveguide properties depending on the wavelength of operation.

Q46. Name the two types of intramodal, or chromatic, dispersion.

Q47. Which dispersion mechanism (material or waveguide) is a function of the size of the fiber's core relative to the wavelength of operation?

Intermodal Dispersion

Intermodal or modal dispersion causes the input light pulse to spread. The input light pulse is made up of a group of modes. As the modes propagate along the fiber, light energy distributed among the modes is delayed by different amounts. The pulse spreads because each mode propagates along the fiber at different speeds. Since modes travel in different directions, some modes travel longer distances. Modal dispersion occurs because each mode travels a different distance over the same time span, as shown in figure 2-25. The modes of a light pulse that enter the fiber at one time exit the fiber a different times. This condition causes the light pulse to spread. As the length of the fiber increases, modal dispersion increases.

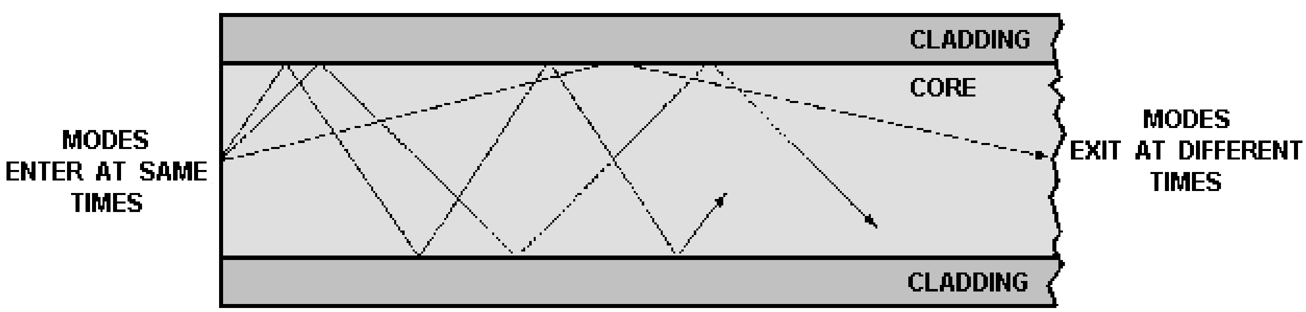


Figure 2-25.—Distance traveled by each mode over the same time span.

Modal dispersion is the dominant source of dispersion in multimode fibers. Modal dispersion does not exist in single mode fibers. Single mode fibers propagate only the fundamental mode. Therefore, single mode fibers exhibit the lowest amount of total dispersion. Single mode fibers also exhibit the highest possible bandwidth.

Q48. Modes of a light pulse that enter the fiber at one time exit the fiber at different times. This condition causes the light pulse to spread. What is this condition called?

SUMMARY

Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas that you have learned. You should have a thorough understanding of these principles before moving on to chapter 3.

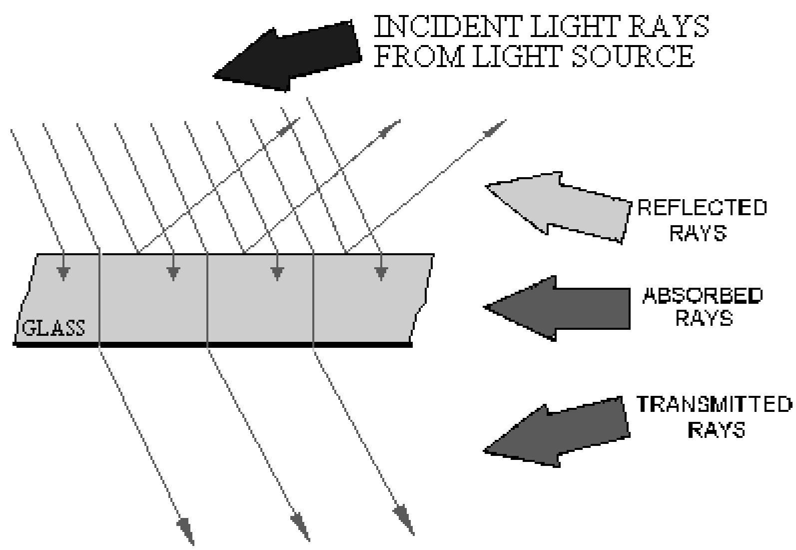
A **LIGHT WAVE** is a form of energy that is moved by wave motion.

WAVE MOTION is defined as a recurring disturbance advancing through space with or without the use of a physical medium.

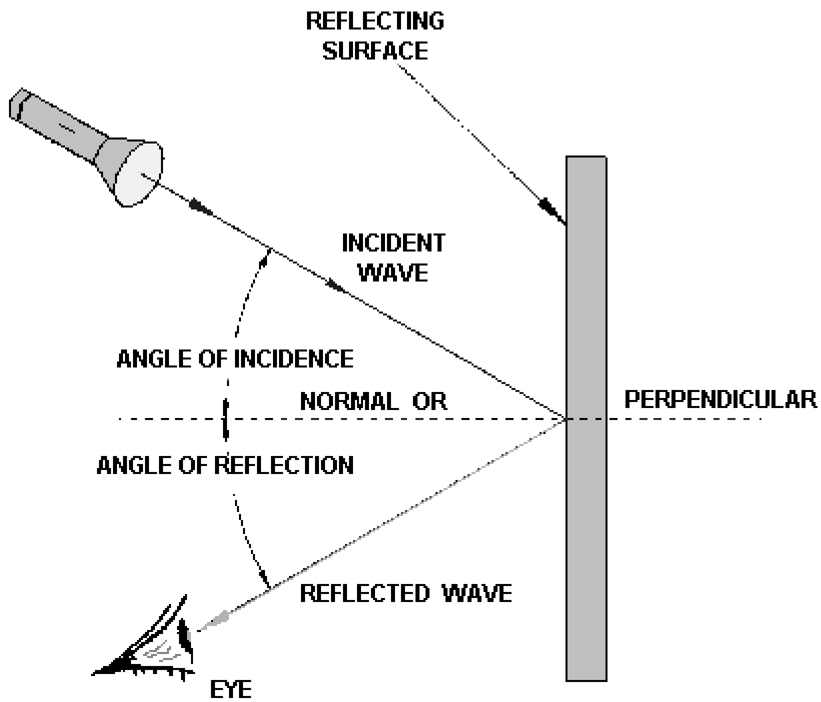
SCIENTIFIC EXPERIMENTS seem to show that light is composed of tiny particles, while other experiments indicate that light is made up of waves. Today, physicists have come to accept a theory concerning light that is a combination of particle (ray) theory and wave (mode) theory.

TRANSVERSE WAVE MOTION describes the up and down wave motion that is at right angle (transverse) to the outward motion of the waves.

LIGHT RAYS, when they encounter any substance, are either transmitted, refracted, reflected, or absorbed.

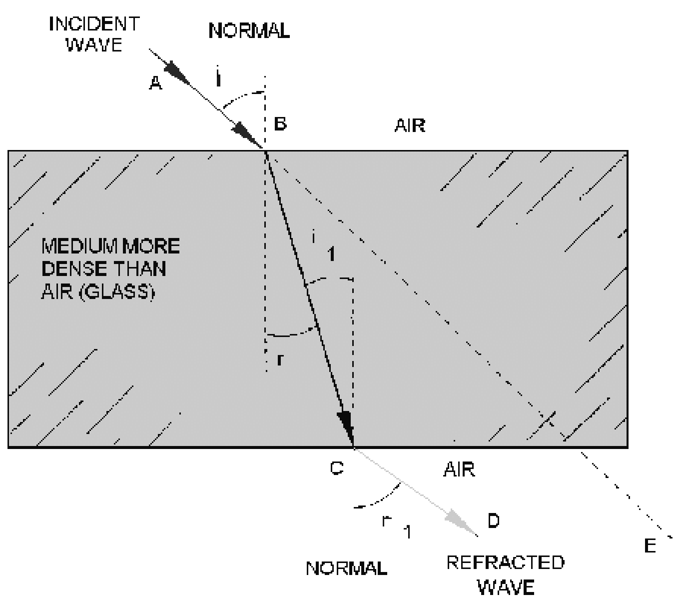


REFLECTION occurs when a wave strikes an object and bounces back (toward the source). The wave that moves from the source to the object is called the **incident wave**, and the wave that moves away from the object is called the **reflected wave**.



The **LAW OF REFLECTION** states that the angle of incidence is equal to the angle of reflection.

REFRACTION occurs when a wave traveling through two different mediums passes through the **boundary** of the mediums and bends toward or away from the **normal**.



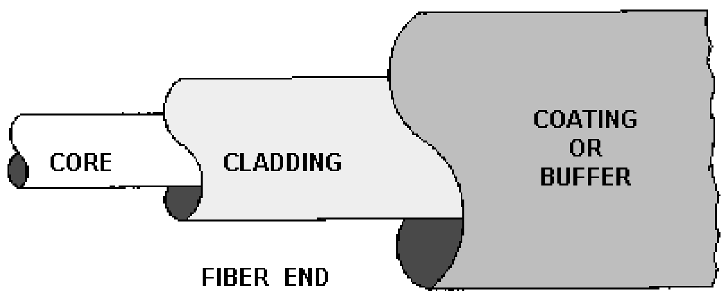
The **RAY THEORY** and the **MODE THEORY** describe how light energy is transmitted along an optical fiber.

The **INDEX OF REFRACTION** is the basic optical material property that measures the speed of light in an optical medium.

SNELL'S LAW OF REFRACTION describes the relationship between the incident and the refracted rays when light rays encounter the boundary between two different transparent materials.

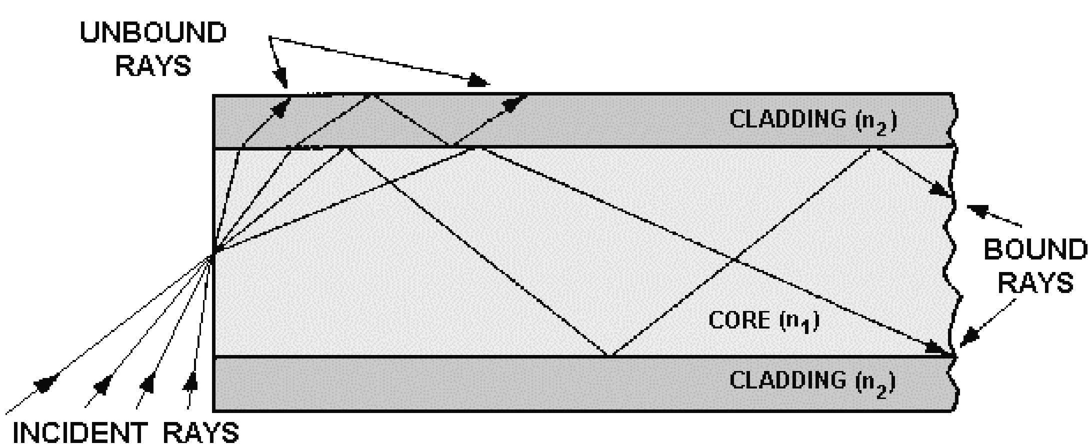
TOTAL INTERNAL REFLECTION occurs when light rays are totally reflected at the boundary between two different transparent materials. The angle at which total internal reflection occurs is called the **critical angle of incidence**.

The **CORE, CLADDING**, and **COATING** or **BUFFER** are the three basic parts of an optical fiber.

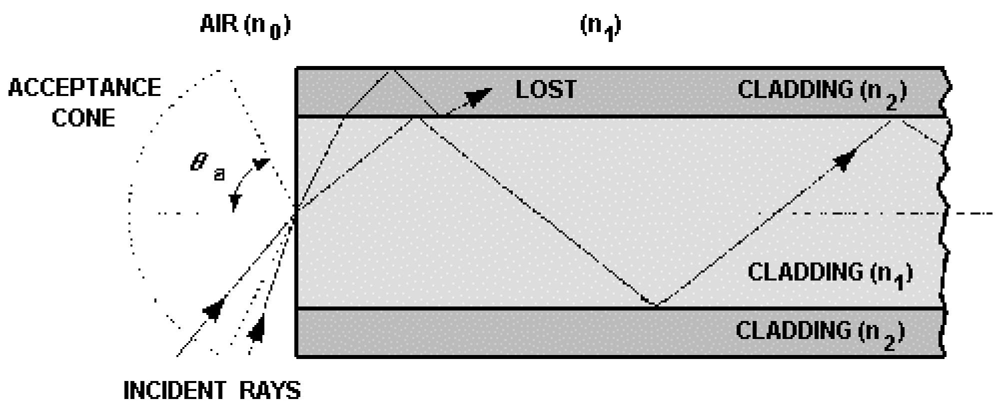


The **RAY THEORY** describes how light rays propagate along an optical fiber. **MERIDIONAL RAYS** pass through the axis of the optical fiber. **SKEW RAYS** propagate through an optical fiber without passing through its axis.

BOUND RAYS propagate through an optical fiber core by total internal reflection. **UNBOUND RAYS** refract out of the fiber core into the cladding and are eventually lost.



The **ACCEPTANCE ANGLE** is the maximum angle to the axis of the fiber that light entering the fiber is bound or propagated. The light ray incident on the fiber core must be within the acceptance cone defined by the acceptance angle to be propagated along an optical fiber.



NUMERICAL APERTURE (NA) is a measurement of the ability of an optical fiber to capture light.

The **MODE THEORY** uses electromagnetic wave behavior to describe the propagation of light along an optical fiber. A set of guided electromagnetic waves are called the **modes** of the fiber.

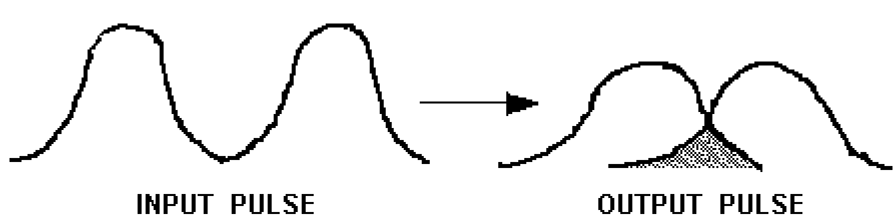
MODES traveling in an optical fiber are said to be transverse. Modes are described by their electric, E(x,y,z), and magnetic, H(x,y,z), fields. The electric field and magnetic field are at right angles to each other.

NORMALIZED FREQUENCY determines how many modes a fiber can support. The number of modes is represented by the normalized frequency constant.

SINGLE MODE and **MULTIMODE FIBERS** are classified by the number of modes that propagate along the optical fiber. Single mode fibers propagate only one mode because the core size approaches the operational wavelength. Multimode fibers can propagate over 100 modes depending on the core size and numerical aperture.

ATTENUATION is the loss of optical power as light travels along an optical fiber. Attenuation in an optical fiber is caused by absorption, scattering, and bending losses.

DISPERSION spreads the optical pulse as it travels along the fiber. Dispersion limits how fast information is transferred.

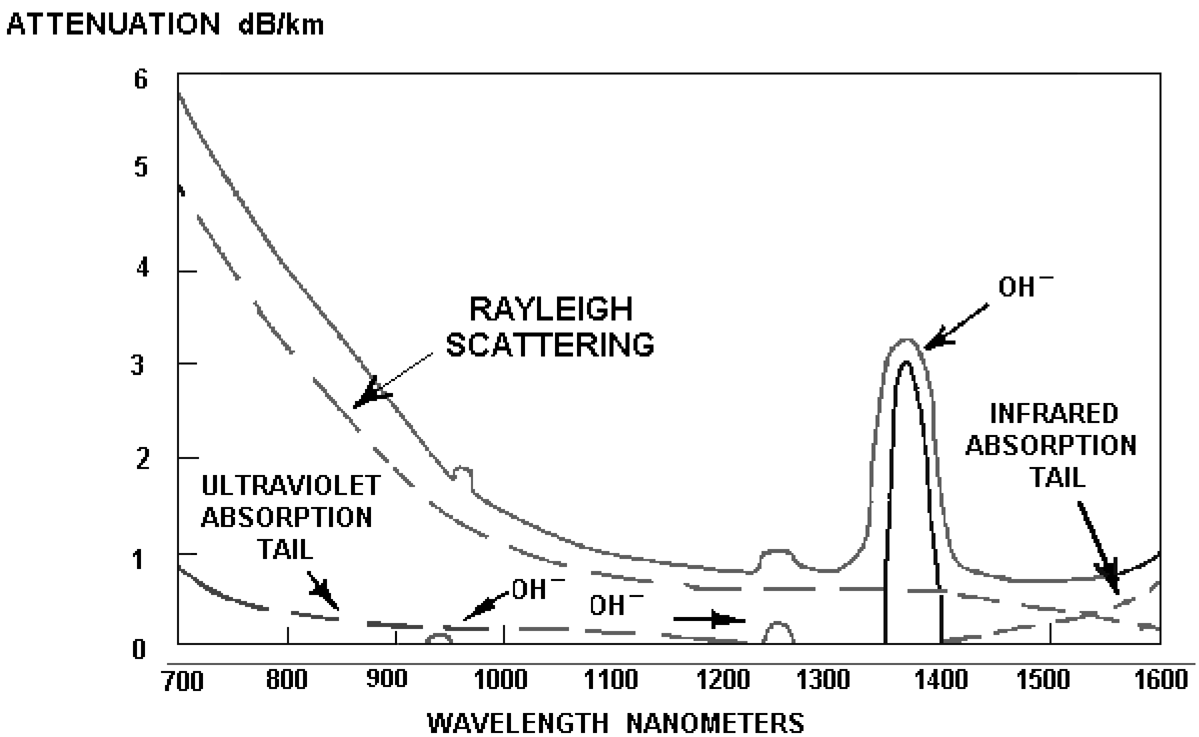


ABSORPTION is the conversion of optical power into another energy form, such as heat. **INTRINSIC ABSORPTION** is caused by basic fiber-material properties. **EXTRINSIC ABSORPTION** is caused by impurities introduced into the fiber material.

SILICA FIBERS are predominately used in fiber optic communications. They have low intrinsic material absorption at the wavelengths of operation.

The **WAVELENGTH OF OPERATION** in fiber optics is between 700 nm and 1600 nm. The wavelength of operation is between the ultraviolet (below 400 nm) and infrared (above 2000 nm) intrinsic absorption regions.

EXTRINSIC ABSORPTION occurs when impurities, such as hydroxyl ions (OH^-), are introduced into the fiber. OH^- absorption peaks define three regions or windows of preferred operation. The first window is centered at 850 nm. The second window is centered at 1300 nm. The third window is centered at 1550 nm.

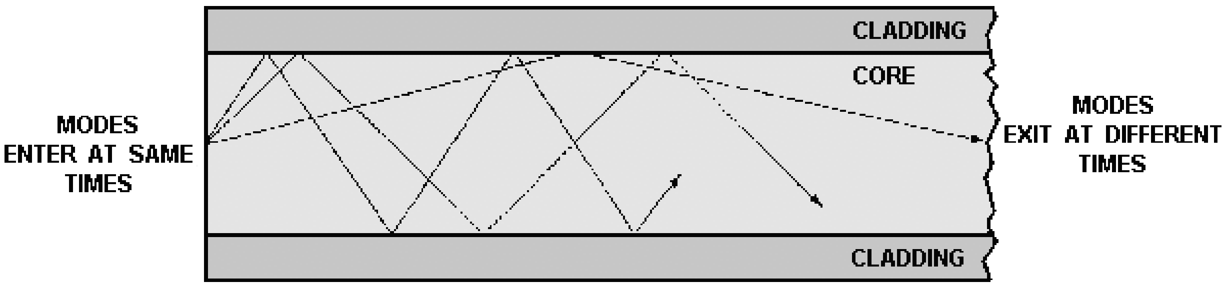


SCATTERING losses are caused by the interaction of light with density fluctuations within a fiber. **Rayleigh scattering** is the main source of loss in commercial fibers operating between 700 nm and 1600 nm.

MICROBENDS are small microscopic bends of the fiber axis that occur mainly when a fiber is cabled. **MACROBENDS** are bends having a large radius of curvature relative to the fiber diameter.

INTRAMODAL, or **CHROMATIC**, **DISPERSION** occurs because light travels through different materials and different waveguide structures at different speeds. **MATERIAL DISPERSION** is dependent on the light wavelengths interaction with the refractive index of the core. **WAVEGUIDE DISPERSION** is a function of the size of the fiber's core relative to the wavelength of operation.

INTERMODAL, or **MODAL**, **DISPERSION** occurs because each mode travels a different distance over the same time span.



ANSWERS TO QUESTIONS Q1. THROUGH Q48.

A1. Photons.

A2. Transverse-wave motion.

A3. Light waves are either transmitted, refracted, reflected, or absorbed.

A4. Transparent.

A5. Opaque.

A6. The law of reflection states that the angle of incidence is equal to the angle of reflection.

A7. When the wave is nearly parallel to the reflecting surface.

A8. When the wave is perpendicular to the reflecting surface.

A9. The law of reflection.

A10. Depends on the bending caused by the velocity difference of the wave traveling through different mediums.

A11. Transmitted.

A12. Diffusion.

A13. The ray theory and the mode theory.

A14. The index of refraction.

A15. Light will travel faster in an optical material that is less dense.

A16. Part of the light ray is reflected back into the glass and part of the light ray is refracted (bent) as it enters the air.

A17. Total internal reflection occurs when the angle of refraction approaches 90 degrees. This condition occurs when the angle of incidence increases to the point where no refraction is possible.

A18. Critical angle of incidence.

A19. Core, cladding, and coating or buffer.

A20. Core.

A21. The ray theory.

A22. Total internal reflection.

A23. Smaller.

A24. NA measures the light-gathering ability of an optical fiber.

A25. Skew ray angle.

A26. Modes of the fiber.

A27. Direction, amplitude, and wavelength of propagation.

A28. Yes, the wavefront will disappear because the total amount of phase collected must be an integer multiple of 2π*. (If the propagating wavefronts are out of phase, they will disappear. The wavefronts that are in phase interfere with the wavefronts out of phase. This type of interference is called destructive interference.)*

A29. Cutoff wavelength.

A30. Multimode fiber.

A31. The order of a mode is indicated by the number of field maxima within the core of the fiber. The order of a mode is also determined by the angle that the wavefront makes with the axis of the fiber.

A32. Mode coupling is the exchange of power between two modes.

A33. Increase.

A34. $V \leq 2.405$*.*

A35. Decrease.

A36. Core diameter, NA, and index profile properties.

A37. Light absorption, scattering, and bending losses.

A38. Attenuation is the loss of optical power as light travels along the fiber.

A39. Intrinsic and extrinsic material properties.

A40. Ultraviolet absorption region (below 400 nm) and infrared absorption region (above 2000 nm).

A41. The first, second, and third windows of operation are 850 nm, 1300 nm, and 1550 nm, respectively.

A42. Rayleigh scattering.

A43. (a) Mie scattering; (b) Rayleigh scattering.

A44. Uneven coating applications, improper cabling procedures, and external force.

A45. Fiber sensitivity to bending losses can be reduced if the refractive index of the core is increased and/or if the overall diameter of the fiber increases.

A46. Material dispersion and waveguide dispersion.

A47. Waveguide dispersion.

A48. Modal dispersion.

CHAPTER 3

OPTICAL FIBERS AND CABLES

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to do the following:

1. Describe multimode and single mode step-index and graded-index fibers.
2. Explain the terms refractive index profile, relative refractive index difference, and profile parameter.
3. List the performance advantages of 62.5/125 μm multimode graded-index fibers.
4. Identify the two basic types of single mode step-index fibers.
5. Describe the vapor phase oxidation and direct-melt optical fiber fabrication procedures.
6. Describe the fiber drawing process.
7. List the benefits of cabled optical fibers over bare fibers.
8. Identify the basic cable components, such as buffers, strength members, and jacket materials.
9. Describe the material and design requirements imposed on military fiber optic cable designs.
10. Describe the advantages and disadvantages of OFCC cable, stranded cable, and ribbon cable designs.

OPTICAL FIBER AND CABLE DESIGN

Optical fibers are thin cylindrical dielectric (non-conductive) waveguides used to send light energy for communication. Optical fibers consist of three parts: the core, the cladding, and the coating or buffer. The choice of optical fiber materials and fiber design depends on operating conditions and intended application. Optical fibers are protected from the environment by incorporating the fiber into some type of cable structure. Cable strength members and outer jackets protect the fiber. Optical cable structure and material composition depend on the conditions of operation and the intended application.

OPTICAL FIBERS

Chapter 2 classified optical fibers as either single mode or multimode fibers. Fibers are classified according to the number of modes that they can propagate. Single mode fibers can propagate only the fundamental mode. Multimode fibers can propagate hundreds of modes. However, the classification of an optical fiber depends on more than the number of modes that a fiber can propagate.

An optical fiber's refractive index profile and core size further distinguish single mode and multimode fibers. The **refractive index profile** describes the value of refractive index as a function of radial distance at any fiber diameter. Fiber refractive index profiles classify single mode and multimode fibers as follows:

- Multimode step-index fibers
- Multimode graded-index fibers
- Single mode step-index fibers
- Single mode graded-index fibers

In a **step-index** fiber, the refractive index of the core is uniform and undergoes an abrupt change at the core-cladding boundary. Step-index fibers obtain their name from this abrupt change called the step change in refractive index. In **graded-index** fibers, the refractive index of the core varies gradually as a function of radial distance from the fiber center.

Single mode and multimode fibers can have a step-index or graded-index refractive index profile. The performance of multimode graded-index fibers is usually superior to multimode step-index fibers. However, each type of multimode fiber can improve system design and operation depending on the intended application. Performance advantages for single mode graded-index fibers compared to single mode step-index fibers are relatively small. Therefore, single mode fiber production is almost exclusively step-index. Figure 3-1 shows the refractive index profile for a multimode step-index fiber and a multimode graded-index fiber. Figure 3-1 also shows the refractive index profile for a single mode step-index fiber. Since light propagates differently in each fiber type, figure 3-1 shows the propagation of light along each fiber.

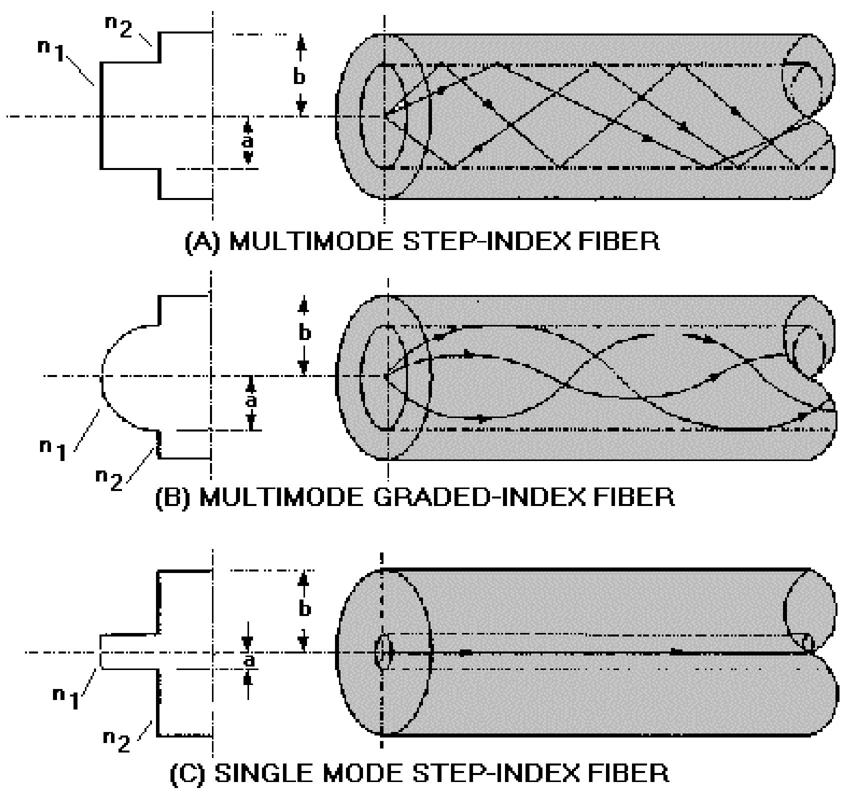


Figure 3-1.—The refractive index profiles and light propagation in multimode step-index, multimode graded-index, and single mode step-index fibers.

In chapter 2, you learned that fiber core size and material composition can affect system performance. A small change in core size and material composition affects fiber transmission properties, such as attenuation and dispersion. When selecting an optical fiber, the system designer decides which fiber core size and material composition is appropriate.

Standard core sizes for multimode step-index fibers are 50 μm and 100 μm. Standard core sizes for multimode graded-index fibers are 50 μm, 62.5 μm, 85 μm, and 100 μm. Standard core sizes for single mode fibers are between 8 μm and 10 μm. In most cases, the material used in the preparation of optical fibers is high-quality glass (SiO_2). This glass contains very low amounts of impurities, such as water or elements other than silica and oxygen. Using high-quality glass produces fibers with low losses. Small amounts of some elements other than silica and oxygen are added to the glass material to change its index of refraction. These elements are called material dopants. Silica doped with various materials forms the refractive index profile of the fiber core and material dopants are discussed in more detail later in this chapter. Glass is not the only material used in fabrication of optical fibers. Plastics are also used for core and cladding materials in some applications.

A particular optical fiber design can improve fiber optic system performance. Each single mode or multimode, step-index or graded-index, glass or plastic, or large or small core fiber has an intended application. The system designer must choose an appropriate fiber design that optimizes system performance in his application.

Q1. Describe the term "refractive index profile."

Q2. The refractive index of a fiber core is uniform and undergoes an abrupt change at the core-cladding boundary. Is this fiber a step-index or graded-index fiber?

Q3. Multimode optical fibers can have a step-index or graded-index refractive index profile. Which fiber, multimode step-index or multimode graded-index fiber, usually performs better?

Q4. List the standard core sizes for multimode step-index, multimode graded-index, and single mode fibers.

MULTIMODE STEP-INDEX FIBERS

A multimode step-index fiber has a core of radius (a) and a constant refractive index n_1. A cladding of slightly lower refractive index n_2 surrounds the core. Figure 3-2 shows the refractive index profile n(r) for this type of fiber. n(r) is equal to n_1 at radial distances $r < a$ (core). n(r) is equal to n_2 at radial distances $r \geq a$ (cladding). Notice the step decrease in the value of refractive index at the core-cladding interface. This step decrease occurs at a radius equal to distance (a). The difference in the core and cladding refractive index is the parameter Δ:

$$\Delta = \frac{n_1^2 - n_2^2}{2n_1^2}$$

Δ is the **relative refractive index difference.**

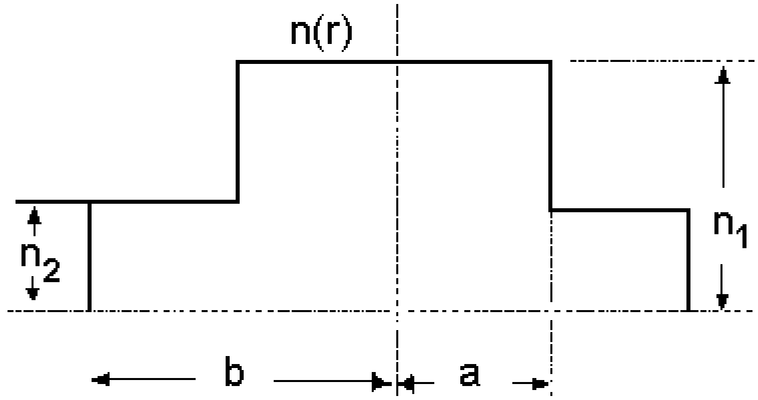


Figure 3-2.—The refractive index profile for multimode step-index fibers.

The ability of the fiber to accept optical energy from a light source is related to Δ. Δ also relates to the numerical aperture by

$$NA \simeq n_1\sqrt{2\Delta}.$$

The number of modes that multimode step-index fibers propagate depends on Δ and core radius (a) of the fiber. The number of propagating modes also depends on the wavelength (λ) of the transmitted light. In a typical multimode step-index fiber, there are hundreds of propagating modes.

Most modes in multimode step-index fibers propagate far from cutoff. Modes that are cut off cease to be bound to the core of the fiber. Modes that are farther away from the cutoff wavelength concentrate most of their light energy into the fiber core. Modes that propagate close to cutoff have a greater percentage of their light energy propagate in the cladding. Since most modes propagate far from cutoff, the majority of light propagates in the fiber core. Therefore, in multimode step-index fibers, cladding properties, such as cladding diameter, have limited affect on mode (light) propagation.

Multimode step-index fibers have relatively large core diameters and large numerical apertures. A large core size and a large numerical aperture make it easier to couple light from a light-emitting diode (LED) into the fiber. Multimode step-index fiber core size is typically 50 µm or 100 µm. Unfortunately, multimode step-index fibers have limited bandwidth capabilities. Dispersion, mainly modal dispersion, limits the bandwidth or information-carrying capacity of the fiber. System designers consider each factor when selecting an appropriate fiber for each particular application.

Multimode step-index fiber selection depends on system application and design. Short-haul, limited bandwidth, low-cost applications typically use multimode step-index fibers.

Q5. Multimode step-index fibers have a core and cladding of constant refractive index n_1 and n_2, respectively. Which refractive index, the core or cladding, is lower?

Q6. In multimode step-index fibers, the majority of light propagates in the fiber core for what reason?

Q7. Multimode step-index fibers have relatively large core diameters and large numerical apertures. These provide what benefit?

MULTIMODE GRADED-INDEX FIBERS

A multimode graded-index fiber has a core of radius (a). Unlike step-index fibers, the value of the refractive index of the core (n_1) varies according to the radial distance (r). The value of n_1 decreases as the distance (r) from the center of the fiber increases. The value of n_1 decreases until it approaches the value of the refractive index of the cladding (n_2). The value of n_1 must be higher than the value of n_2 to allow for proper mode propagation. Like the step-index fiber, the value of n_2 is constant and has a slightly lower value than the maximum value of n_1. The relative refractive index difference (Δ) is determined using the maximum value of n_1 and the value of n_2.

Figure 3-3 shows a possible refractive index profile n(r) for a multimode graded-index fiber. Notice the parabolic refractive index profile of the core. The **profile parameter** (α) determines the shape of the core's profile. As the value of α increases, the shape of the core's profile changes from a triangular shape to step as shown in figure 3-4. Most multimode graded-index fibers have a parabolic refractive index profile. Multimode fibers with near parabolic graded-index profiles provide the best performance. Unless otherwise specified, when discussing multimode graded-index fibers, assume that the core's refractive index profile is parabolic (α=2).

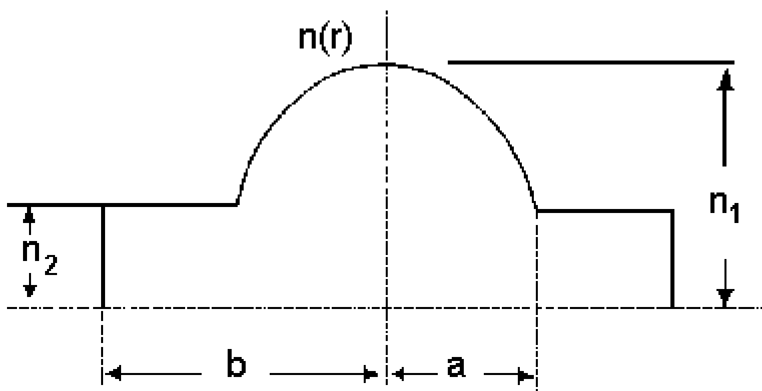


Figure 3-3.—The refractive index profile for multimode graded-index fibers.

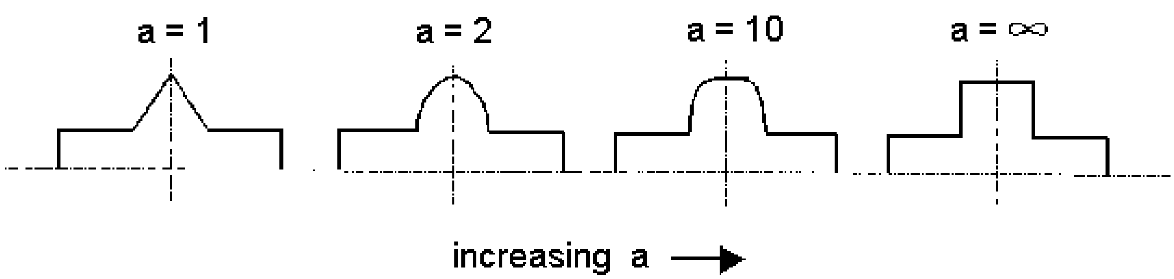


Figure 3-4.—The refractive index profiles for different values of α.

Light propagates in multimode graded-index fibers according to refraction and total internal reflection. The gradual decrease in the core's refractive index from the center of the fiber causes the light rays to be refracted many times. The light rays become refracted or curved, which increases the angle of incidence at the next point of refraction. Total internal reflection occurs when the angle of incidence becomes larger than the critical angle of incidence. Figure 3-5 shows the process of refraction and total internal reflection of light in multimode graded-index fibers. Figure 3-5 also illustrates the boundaries of different values of core refractive index by dotted lines. Light rays may be reflected to the axis of the fiber before reaching the core-cladding interface.

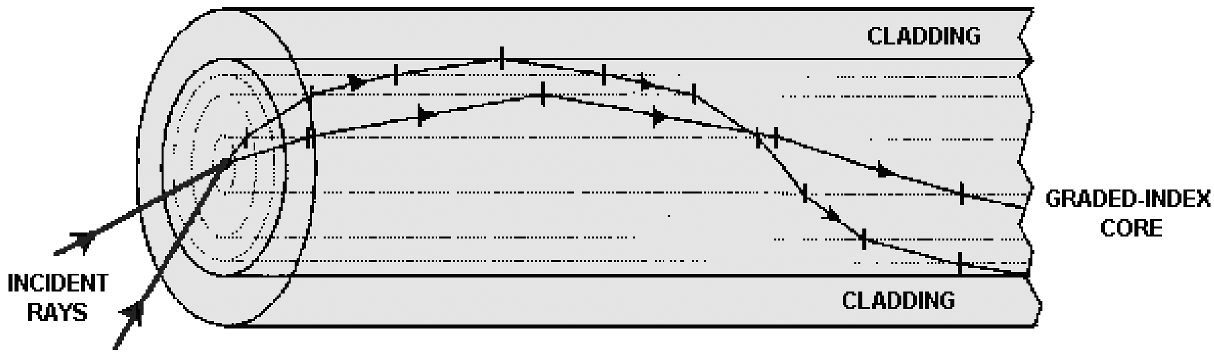


Figure 3-5.—Refractive index grading and light propagation in multimode graded-index fibers.

The NA of a multimode graded-index fiber is at its maximum value at the fiber axis. This NA is the **axial numerical aperture [NA(0)]**. NA(0) is approximately equal to

$$n_1\sqrt{2\Delta}.$$

However, the NA for graded-index fibers varies as a function of the radial distance (r). NA varies because of the refractive index grading in the fiber's core. The NA decreases from the maximum, NA(0), to zero at distances greater than the core-cladding boundary distance (r>a). The NA, relative refractive index difference (Δ), profile parameter (α), and normalized frequency (V) determine the number of propagating modes in multimode graded-index fibers. A multimode graded-index fiber with the same normalized frequency as a multimode step-index fiber will have approximately one-half as many propagating modes. However, multimode graded-index fibers typically have over one-hundred propagating modes.

Multimode graded-index fibers accept less light than multimode step-index fibers with the same core Δ. However, graded-index fibers usually outperform the step-index fibers. The core's parabolic refractive index profile causes multimode graded-index fibers to have less modal dispersion.

Figure 3-5 shows possible paths that light may take when propagating in multimode graded-index fibers. Light rays that travel farther from the fiber's axis travel a longer distance. Light rays that travel farther from the center travel in core material with an average lower refractive index.

In chapter 2, you learned that light travels faster in a material with a lower refractive index. Therefore, those light rays that travel the longer distance in the lower refractive index parts of the core travel at a greater average velocity. This means that the rays that travel farther from the fiber's axis will arrive at each point along the fiber at nearly the same time as the rays that travel close to the fiber's axis. The decrease in time difference between light rays reduces modal dispersion and increases multimode graded-index fiber bandwidth. The increased bandwidth allows the use of multimode graded-index fibers in most applications.

Most present day applications that use multimode fiber use graded-index fibers. The basic design parameters are the fiber's core and cladding size and Δ. Standard multimode graded-index fiber core and cladding sizes are 50/125 μm, 62.5/125 μm, 85/125 μm, and 100/140 μm. Each fiber design has a specific

Δ that improves fiber performance. Typical values of Δ are around 0.01 to 0.02. Although no single multimode graded-index fiber design is appropriate for all applications, the 62.5/125 μm fiber with a Δ of 0.02 offers the best overall performance.

A multimode graded-index fiber's source-to-fiber coupling efficiency and insensitivity to microbending and macrobending losses are its most distinguishing characteristics. The fiber core size and Δ affect the amount of power coupled into the core and loss caused by microbending and macrobending. Coupled power increases with both core diameter and Δ, while bending losses increase directly with core diameter and inversely with Δ. However, while these values favor high Δs, a smaller Δ improves fiber bandwidth. In most applications, a multimode graded-index fiber with a core and cladding size of 62.5/125 μm offers the best combination of the following properties:

- Relatively high source-to-fiber coupling efficiency
- Low loss
- Low sensitivity to microbending and macrobending
- High bandwidth
- Expansion capability

For example, local area network (LAN) and shipboard applications use multimode graded-index fibers with a core and cladding size of 62.5/125 μm. In LAN-type environments, macrobend and microbend losses are hard to predict. Cable tension, bends, and local tie-downs increase macrobend and microbend losses. In shipboard applications, a ship's cable-way may place physical restrictions, such as tight bends, on the fiber during cable plant installation. The good microbend and macrobend performance of 62.5/125 μm fiber permits installation of a rugged and robust cable plant. 62.5/125 μm multimode graded-index fibers allow for uncomplicated growth because of high fiber bandwidth capabilities for the expected short cable runs on board ships.

Q8. The profile parameter (α) determines the shape of the multimode graded-index core's refractive index profile. As the value of the α increases, how does the core's profile change?

Q9. Light propagates in multimode graded-index fibers according to refraction and total internal reflection. When does total internal reflection occur?

Q10. What four fiber properties determine the number of modes propagating in a multimode graded-index fiber?

Q11. Light travels faster in a material with a lower refractive index. Therefore, light rays that travel a longer distance in a lower refractive index travel at a greater average velocity. What effect does this have on multimode graded-index fiber modal dispersion and bandwidth?

Q12. What multimode graded-index fiber offers the best overall performance for most applications?

Q13. What are the most distinguishing characteristics of a multimode graded-index fiber?

Q14. How are source-to-fiber coupling and microbending and macrobending losses affected by changes in core diameter and Δ?

Q15. While coupled power and bending loss favor a high Δ, which Δ value, smaller or larger, improves fiber bandwidth?

SINGLE MODE STEP-INDEX FIBERS

There are two basic types of single mode step-index fibers: matched clad and depressed clad. **Matched cladding** means that the fiber cladding consists of a single homogeneous layer of dielectric material. **Depressed cladding** means that the fiber cladding consists of two regions: the inner and outer cladding regions. Matched-clad and depressed-clad single mode step-index fibers have unique refractive index profiles.

A matched-clad single mode step-index fiber has a core of radius (a) and a constant refractive index n_1. A cladding of slightly lower refractive index surrounds the core. The cladding has a refractive index n_2. Figure 3-6 shows the refractive index profile n(r) for the matched-clad single mode fiber.

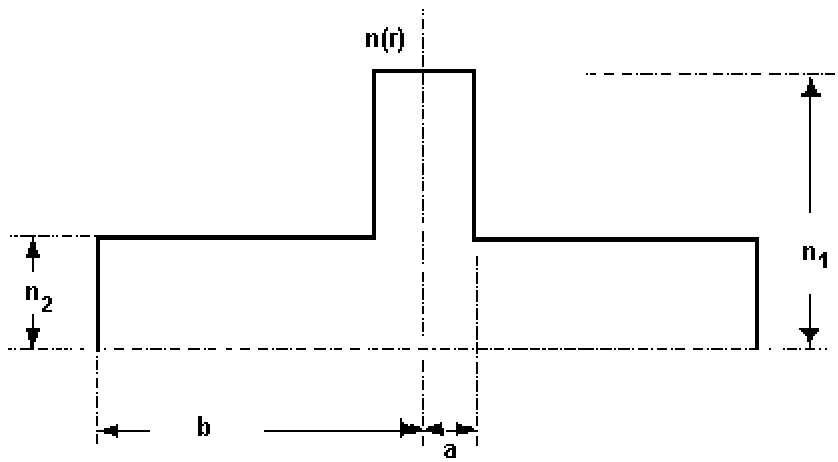


Figure 3-6.—Matched-clad refractive index profile.

Figure 3-7 shows the refractive index profile n(r) for the depressed-clad single mode fiber. A depressed-clad single mode step-index fiber has a core of radius (a) with a constant refractive index n_1. A cladding, made of two regions, surrounds the core. An inner cladding region surrounds the core of the fiber and has a refractive index of n_2. The inner cladding refractive index n_2 is lower than the core's refractive index n_1. An outer cladding region surrounds the inner cladding region and has a higher refractive index n_3 than the inner cladding region. However, the outer cladding refractive index n_3 is lower than the core's refractive index n_1.

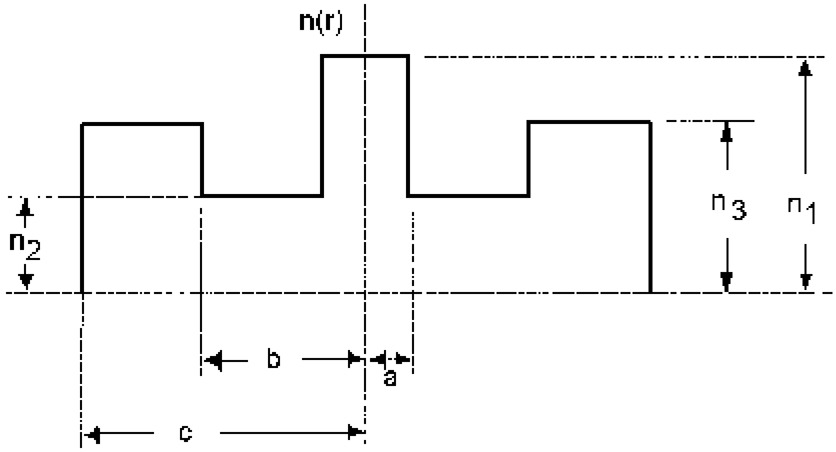


Figure 3-7.—Depressed-clad refractive index profile.

Single mode step-index fibers propagate only one mode, called the fundamental mode. Single mode operation occurs when the value of the fiber's normalized frequency is between 0 and 2.405 ($0 \leq V < 2.405$). The value of V should remain near the 2.405 level. When the value of V is less than 1, single mode fibers carry a majority of the light power in the cladding material. The portion of light transmitted by the cladding material easily radiates out of the fiber. For example, light radiates out of the cladding material at fiber bends and splices.

Single mode fiber cutoff wavelength is the smallest operating wavelength when single mode fibers propagate only the fundamental mode. At this wavelength, the 2nd-order mode becomes lossy and radiates out of the fiber core. As the operating wavelength becomes longer than the cutoff wavelength, the fundamental mode becomes increasingly lossy. The higher the operating wavelength is above the cutoff wavelength, the more power is transmitted through the fiber cladding. As the fundamental mode extends into the cladding material, it becomes increasingly sensitive to bending loss. Single mode fiber designs include claddings of sufficient thickness with low absorption and scattering properties to reduce attenuation of the fundamental mode. To increase performance and reduce losses caused by fiber bending and splicing, fiber manufacturers adjust the value of V. To adjust the value of V, they vary the core and cladding sizes and relative refractive index difference (Δ).

A single mode step-index fiber has low attenuation and high bandwidth properties. Present applications for single mode fibers include long-haul, high-speed telecommunication systems. Future applications include single mode fibers for sensor systems. However, the current state of single mode technology makes installation of single mode systems expensive and difficult. Short cable runs, low to moderate bandwidth requirements, and high component cost make installation of single mode fiber shipboard systems impractical at this time.

Q16. What are the two basic types of single mode step-index fibers?

Q17. Which fiber cladding, matched or depressed, consists of two regions?

Q18. In single mode operation, the value of the normalized frequency (V) should remain near the 2.405 level. If the value of V is less than 1, do single mode fibers carry a majority of the power in the core or cladding material?

Q19. What happens to the fundamental mode as the operating wavelength becomes longer than the single mode cutoff wavelength?

Q20. Give two reasons why the value of the normalized frequency (V) is varied in single mode step-index fibers?

SINGLE MODE GRADED-INDEX FIBERS

There are several types of single mode graded-index fibers. These fibers are not standard fibers and are typically only used in specialty applications. Information on single mode graded-index fibers can be found in the references in appendix 2.

FIBER ALTERNATIVES

In most applications, the standard multimode and step-index single mode optical fibers mentioned before have significant performance advantages over conventional copper-based systems. However, performance requirements and cost restraints may prohibit the use of these fibers in certain applications. Fiber manufacturers modify standard multimode and single mode fiber material composition and structural design to meet these additional requirements. Optical fiber design can depart from a traditional circular core and cladding, low-loss glass design. The intent of each change is to increase performance and reduce cost.

Optical fibers composed of plastic have been in use longer than glass fibers. Types of standard fibers using plastics include multimode step-index and graded-index fibers. Multimode step-index and graded-index **plastic clad silica** (PCS) fibers exist. PCS fibers have a silica glass core and a plastic cladding. Normally, PCS fibers are cheaper than all-glass fibers but have limited performance characteristics. PCS fibers lose more light through a plastic cladding than a glass cladding.

Multimode step-index fibers may also have a plastic core and cladding. All-plastic fibers have a higher NA, a larger core size, and cost less to manufacture. However, all-plastic fibers exhibit high loss in the thousands of decibels per kilometer. This high loss is caused by impurities and intrinsic absorption. PCS and all-plastic fibers are used in applications typically characterized by one or all of the following:

- High NA
- Low bandwidth
- Tight bend radius
- Short length (less than 10m to 20m)
- Low cost

Improved fabrication techniques provide the opportunity to experiment with material composition in both multimode and single mode fibers. Fiber manufacturers fabricate optical fibers using glass material whose characteristics improve system performance in the far infrared region. Fiber manufacturers add dopant material to reduce fiber loss and limit material and structural imperfections. Fiber material used in fabrication of low-loss, long wavelength optical fibers include the following:

- Heavy-metal fluorides (such as zirconium and beryllium fluoride)
- Chalcogenide glasses (such as arsenic/sulfur)
- Crystalline materials (such as silver bromide and silver chloride)

In shipboard applications, stringent environmental requirements dictate the design of special optical fibers. In some cases, manufacturers hermetically coat optical fibers to increase survivability and reliability in high-moisture and high-strain environments. Manufacturers also design radiation-hard fibers for nuclear power, space, and military systems. Radiation resistant fibers operate after exposure to nuclear radiation. Shipboard system performance requirements determine whether the use of hermetic and radiation resistant fibers or less costly commercial optical fibers is necessary.

Q21. Give two reasons why optical fiber manufacturers depart from the traditional circular core and cladding, low-loss glass fiber design?

Q22. What five characteristics do applications using plastic clad silica (PCS) and all-plastic fibers typically have?

Q23. List the types of materials used in fabricating low-loss, long wavelength optical fibers.

FABRICATION OF OPTICAL FIBERS

Basically, fiber manufacturers use two methods to fabricate multimode and single mode glass fibers. One method is vapor phase oxidation, and the other method is direct-melt process. In **vapor phase oxidation**, gaseous metal halide compounds, dopant material, and oxygen are oxidized (burned) to form a white silica powder (SiO_2). Manufacturers call SiO_2 the **soot**. Manufacturers deposit the soot on the surface of a glass substrate (mandrel) or inside a hollow tube by one of the following three methods:

- Outside Vapor Phase Oxidation (OVPO).
- Inside Vapor Phase Oxidation (IVPO).
- Vapor Phase Axial Deposition (VAD).

The soot forms the core and cladding material of the preform. The refractive index of each layer of soot is changed by varying the amount of dopant material being oxidized. Figures 3-8, 3-9, and 3-10 illustrate the different vapor phase oxidation preform preparation methods.

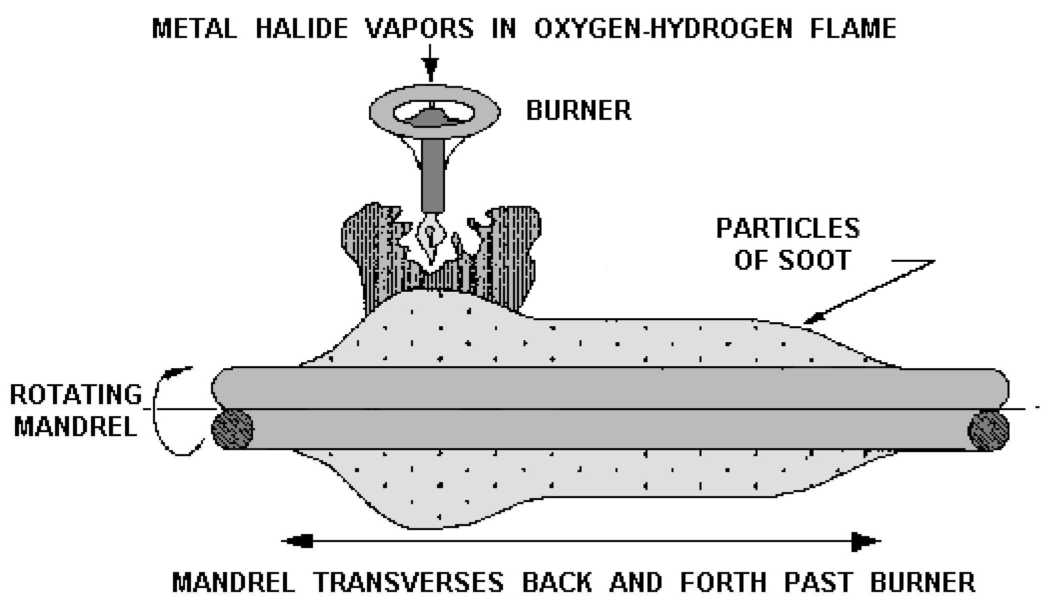


Figure 3-8.—OVPO preform preparation.

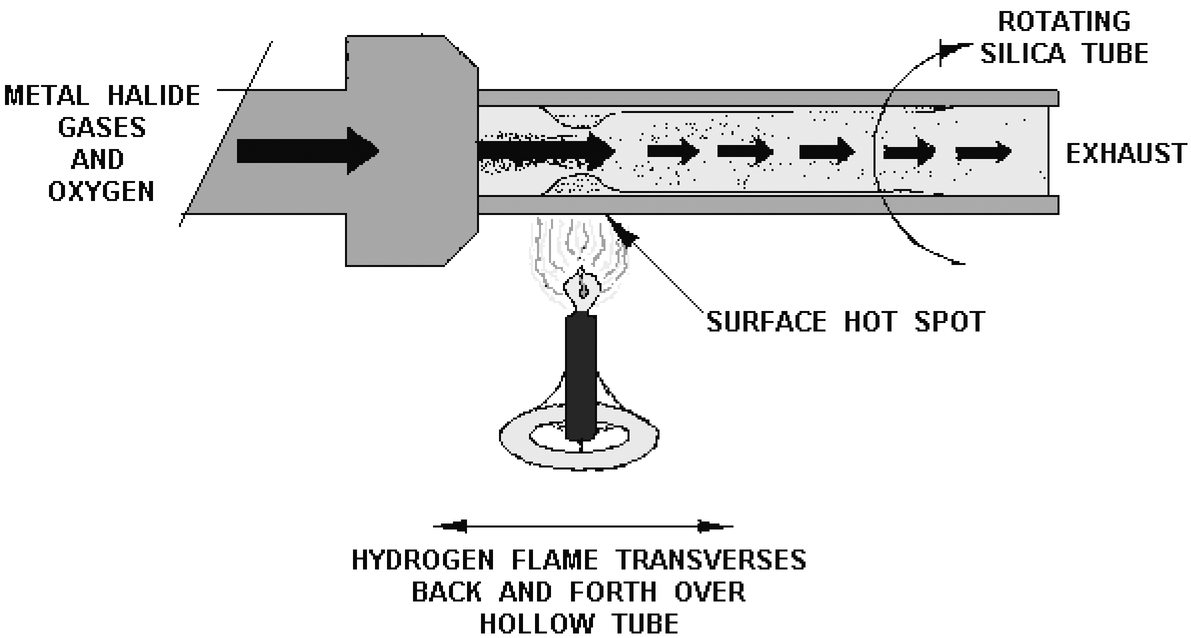


Figure 3-9.—IVPO preform preparation.

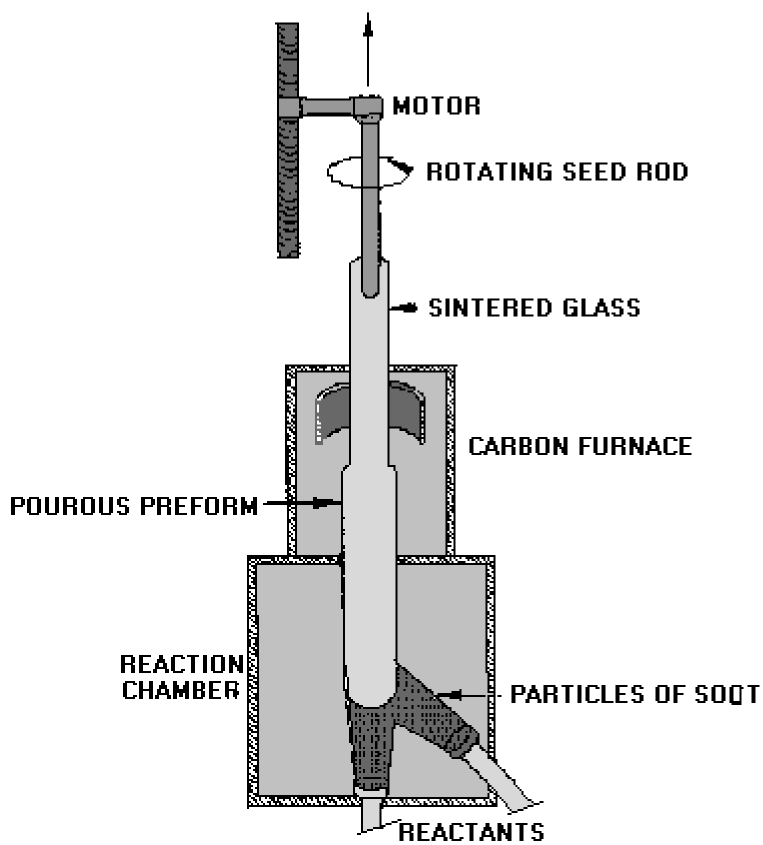


Figure 3-10.—VAD preform preparation.

During vapor phase oxidation, the mandrel or tube continuously moves from side to side and rotates while soot particles are deposited on the surface. This process forms cylindrical layers of soot on the surface of the mandrel or inside the hollow tube. This deposited material is transformed into a solid glass preform by heating the porous material (without melting). The solid preform is then drawn or pulled into an optical fiber by a process called fiber drawing.

The fiber drawing process begins by feeding the glass preform into the drawing furnace. The drawing furnace softens the end of the preform to the melting point. Manufacturers then pull the softened preform into a thin glass filament (glass fiber). To protect the bare fiber from contaminants, manufacturers add an acrylate coating in the draw process. The coating protects the bare fiber from contaminants such as atmospheric dust and water vapor. Figure 3-11 illustrates the process of drawing an optical fiber from the preform.

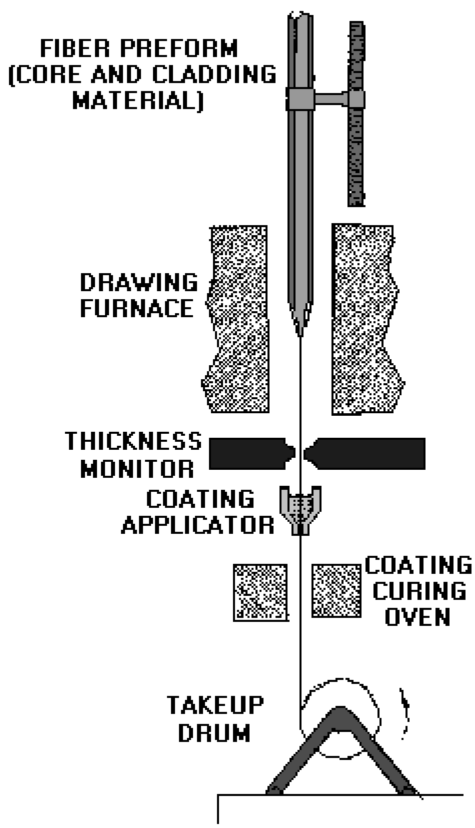


Figure 3-11.—Fiber drawing process.

In the **direct-melt process**, multicomponent glass rods form the fiber structure. Rods of multicomponent glass combine in a molten state to form the fiber core and cladding. The double-crucible method is the most common direct-melt process. The double-crucible method combines the molten rods into a single preform using two concentric crucibles. Optical fibers are drawn from this molten glass using a similar fiber drawing process as in vapor phase oxidation. Figure 3-12 illustrates the double-crucible drawing process.

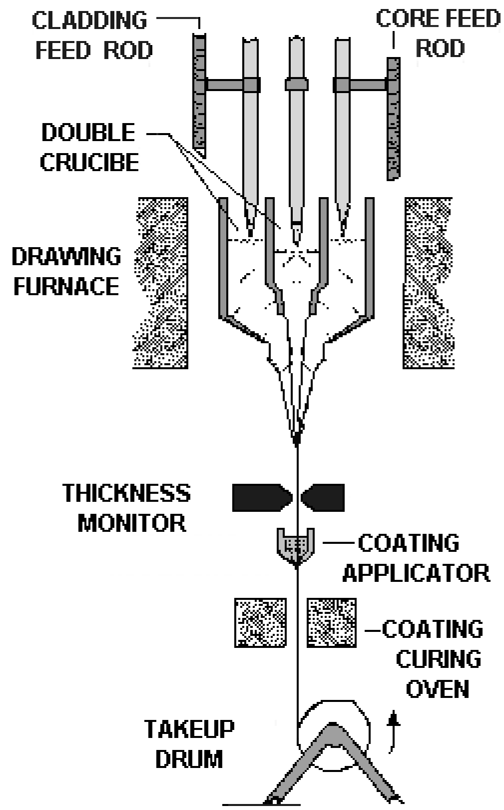


Figure 3-12.—Double-crucible fiber drawing process.

Q24. What are the two methods used by fiber manufacturers to fabricate multimode and single mode glass fibers?

Q25. Which method, vapor phase oxidation or direct-melt process, transforms deposited material into a solid glass preform by heating the porous material without melting?

OPTICAL CABLES

Optical fibers have small cross sectional areas. Without protection, optical fibers are fragile and can be broken. The optical cable structure protects optical fibers from environmental damage. Cable structure includes buffers, strength members, and jackets. Many factors influence the design of fiber optic cables. The cable design relates to the cable's intended application. Properly designed optical cables perform the following functions:

- Protect optical fibers from damage and breakage during installation and over the fiber's lifetime.
- Provide stable fiber transmission characteristics compared with uncabled fibers. Stable transmission includes stable operation in extreme climate conditions.
- Maintain the physical integrity of the optical fiber by reducing the mechanical stresses placed on the fiber during installation and use. Static fatigue caused by tension, torsion, compression, and bending can reduce the lifetime of an optical fiber.

Navy applications require that fiber optic cables meet stringent design specifications. Fiber optic cables must be rugged to meet the optical, environmental, and mechanical performance requirements imposed by Navy systems. Critical system downtime caused by cable failure cannot be tolerated. However, in commercial applications, the requirements imposed on cable designs are not as stringent. Each additional requirement imposed on the fiber optic cable design adds to its cost. Cost is always a main consideration of cable designers in commercial applications. Cost is also considered in Navy applications, but system reliability is the main goal.

Q26. List three benefits that properly cabled optical fibers provide.

FIBER BUFFERS

Coatings and buffers protect the optical fiber from breakage and loss caused by microbends. During the fiber drawing process, the addition of a primary coating protects the bare glass from abrasions and other surface contaminants. For additional protection, manufacturers add a layer of buffer material. The buffer material provides additional mechanical protection for the fiber and helps preserve the fiber's inherent strength.

Manufacturers use a variety of techniques to buffer optical fibers. The types of fiber buffers include tight-buffered, loose-tube, and gel-filled loose-tube. Figure 3-13 shows each type of fiber buffer. The choice of buffering techniques depends on the intended application. In large fiber count commercial applications, manufacturers use the loose-tube buffers. In commercial building and Navy applications, manufacturers use tight buffers.

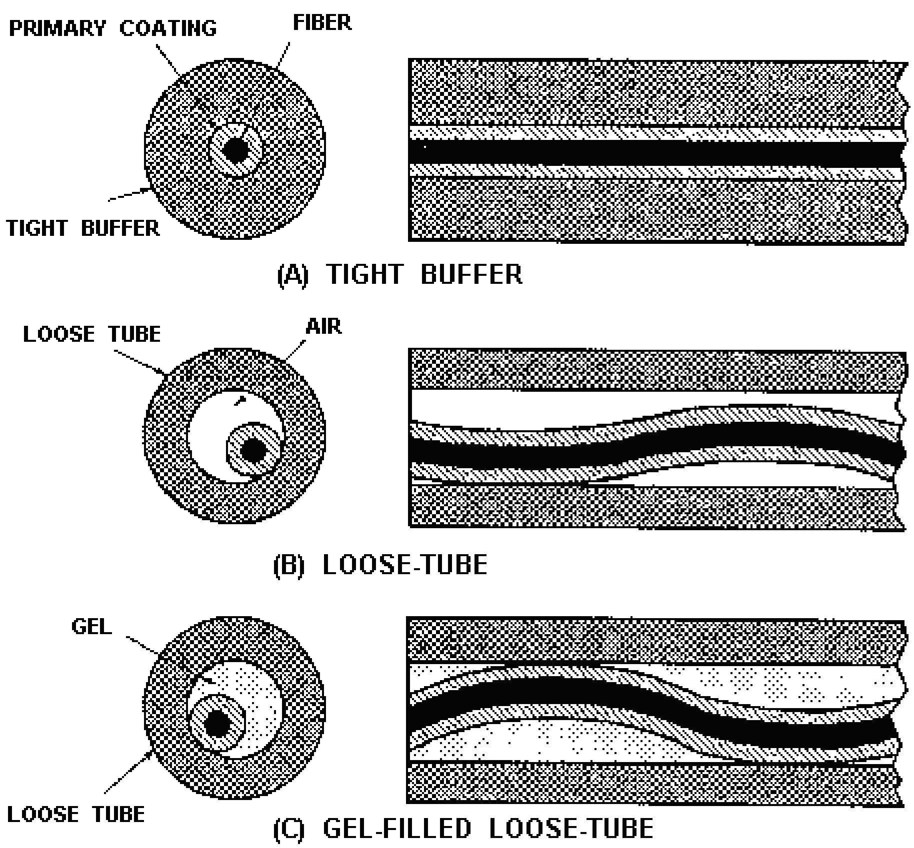


Figure 3-13.—Tight-buffered, loose-tube, and gel-filled loose-tube buffer techniques.

Q27. In addition to a primary coating, manufacturers add a layer of buffer material for what reasons?

Q28. List the three techniques used by manufacturers to buffer optical fibers.

CABLE STRENGTH AND SUPPORT MEMBERS

Fiber optic cables use strength members to increase the cables' strength and protect the fiber from strain. Fiber optic cables may use central support members in cable construction. The central support members generally have buffered fibers or single fiber sub-cables stranded over their surface in a structured, helical manner. The central members may support the optical fibers as cable strength members or may only serve as fillers. Strength and support members must be light and flexible. In commercial applications, the materials used for strength and support include steel wire and textile fibers (such as nylon and arimid yarn). They also include carbon fibers, glass fibers, and glass reinforced plastics. For Navy applications, only non-metallic strength and support members are allowed.

CABLE JACKET, OR SHEATH, MATERIAL

The jacket, or sheath, material provides extra environmental and mechanical protection. Jacket materials for Navy cables have the following properties:

- Low smoke generation
- Low toxicity
- Low halogen content
- Flame retardance
- Fluid resistance
- High abrasion resistance
- Stable performance over temperature

It is difficult to produce a material compound that satisfies every requirement without being too costly. Originally, the production of fire retardant cables included the use of halogenated polymers and additives. These fire retardant cables were also highly toxic. Commercial jacket materials currently used include polyethylene, polyvinyl chloride (PVC), polyurethane, and polyester elastomers. Most commercial jacket materials are unsuitable for use in Navy applications. Researchers have developed jacket materials that are suitable for Navy use.

Q29. List seven properties cable jackets should have.

CABLE DESIGNS

Manufacturers design fiber optic cables for specific applications. Is the cable buried underground or hung from telephone poles? Is the cable snaked through cableways, submerged in water, or just laid on the ground? Is the cable used in industrial, telecommunication, utility, or military applications? Each different application may require a slightly different cable design.

Agreement on standard cable designs is difficult. Cable design choices include jacket materials, water blocking techniques, and the number of fibers to place within the cable. The cable design chosen depends on the cable's intended application. There are presently many types of fiber optic cables. Some fiber optic cables are used in commercial applications, while others are used in military applications.

Standard commercial cable designs will develop over time as fiber optic technology becomes more established. However, this chapter provides only a short discussion on cable designs considered for Navy applications.

Navy systems require that fiber optic cables meet stringent environmental conditions. The types of cable designs considered by the Navy include the optical fiber cable component (OFCC), stranded, and ribbon cable designs. The cable must meet minimal levels of performance in safety (low smoke, low toxicity, low halogen content, etc.), durability (able to withstand shock, vibration, fluids, etc.), and optical performance. The cable must also be easy to install and repair. These factors greatly influence the design of the cables.

Optical Fiber Cable Component (OFCC) Cable

An OFCC cable consists of individual single fiber cables, called **optical fiber cable components (OFCCs)**. OFCCs are a tight-buffered fiber surrounded by arimid yarn and a low-halogen outer jacket. The OFCC outer diameter is typically 2 millimeters (mm). The fiber is typically buffered with a polyester elastomer to a total diameter of 900 μm. Figure 3-14 illustrates the design of the OFCCs. The size of the OFCCs limits the amount of fibers contained within an OFCC cable. An OFCC cable generally contains less than 36 fibers (OFCCs). An OFCC cable of 0.5-inch cable outer diameter can accommodate about 12 fibers.

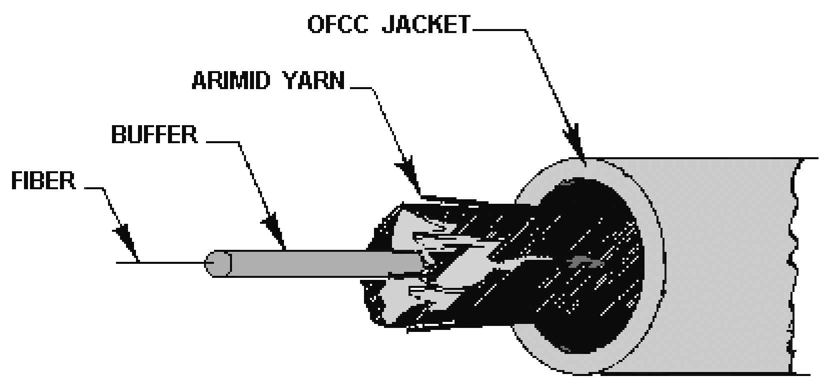


Figure 3-14.—The design of optical fiber cable components (OFCCs).

Figure 3-15 shows an isometric view of a four-fiber shipboard OFCC cable. In this multifiber cable design, the OFCCs surround a flexible central member in a helical manner. The central member may add to cable strength or only support the OFCCs. For additional protection, two layers of arimid yarn strength members encase the OFCC units. These strength members are stranded in opposing lays to minimize microbending of the fibers. The arimid yarn strength members may be treated with polymers that are water absorbing, blocking, and sealing. This treatment eliminates the need for additional water blocking protection. Finally, a low-halogen, flame-resistant outer jacket is extruded over the strength members.

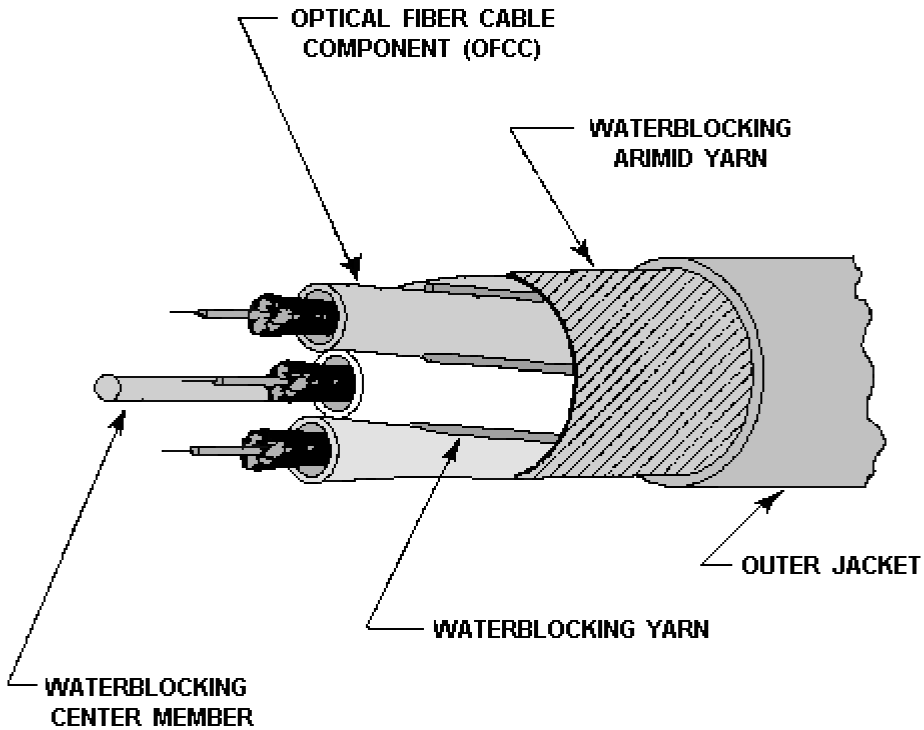


Figure 3-15.—An isometric view of a four-fiber shipboard OFCC cable.

OFCC cables are easy to handle because each cable contains its own subcable, the OFCC. These OFCC subcables make it easy to reconfigure systems and handle individual fibers. Rugged OFCC cable design permits cable use in harsh environments, including Navy applications. OFCC-type cable is recommended for use in low-density (less than 24 fibers) Navy applications. OFCC-type cable is also being evaluated for use in Navy applications with fiber counts up to 36 fibers.

Stranded Cable

A stranded cable is a fiber optic cable consisting of buffered fibers stranded down the center of the cable surrounded by strength members and a protective jacket. Figure 3-16 shows a cross-sectional view of the stranded cable. The fiber is typically buffered with a polyester elastomer to a total diameter of 900 μm. The recommended use of stranded cables is in medium-density (24 to 72 fibers) Navy applications. However, this recommendation is preliminary. Further test and evaluation of prototype stranded cable designs is continuing. Final approval of the stranded cable will occur only after prototype cables have passed all tests.

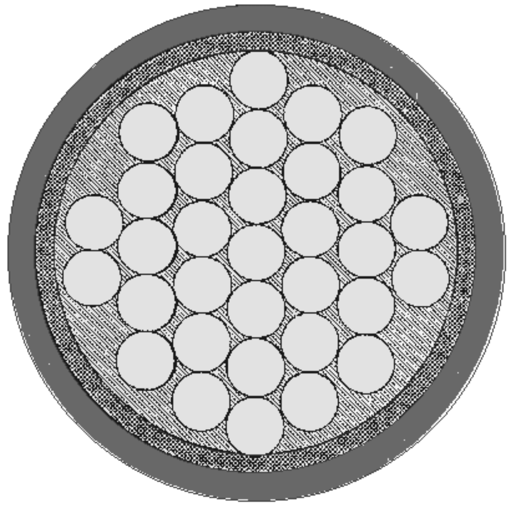


Figure 3-16.—Stranded cable design.

Stranded cable designs increase fiber counts without greatly increasing cable size. Stranded cables are used when fiber counts exceed the limits of OFCC-type cables. For example, the stranded cable design can accommodate about 48 fibers in a O.5-inch cable. The OFCC cable design can accommodate around 12 fibers. The individual fiber is not protected as well in the stranded design as in the OFCC design. For this reason more care is required in handling the individual fibers in the stranded design. The primary problem of the stranded cable design is in meeting the waterblocking requirements. Once manufacturers correct this design problem, the Navy expects that the stranded cable design will meet Navy performance requirements.

Ribbon Cable

A ribbon cable consists of optical fiber ribbons stranded down the center of the cable surrounded by a protective tube, strength members, and an outer jacket. The fiber optic ribbon consists of multiple-coated, 250 ∝m diameter fibers sandwiched in a plastic material. Figure 3-17 shows a cross-sectional view of a 12-fiber ribbon. Cable manufacturers stack these ribbons to form a rectangular cross-sectional array of fibers. Stacked ribbons are the basic building blocks of the ribbon cable. Figure 3-18 illustrates this cross-sectional array of ribbons. Manufacturers introduce a controlled twist to the stacked ribbons to minimize fiber stress when the cable is bent. An inner plastic tube, strength members, and an outer protective jacket surround the stacked ribbons, providing environmental protection.

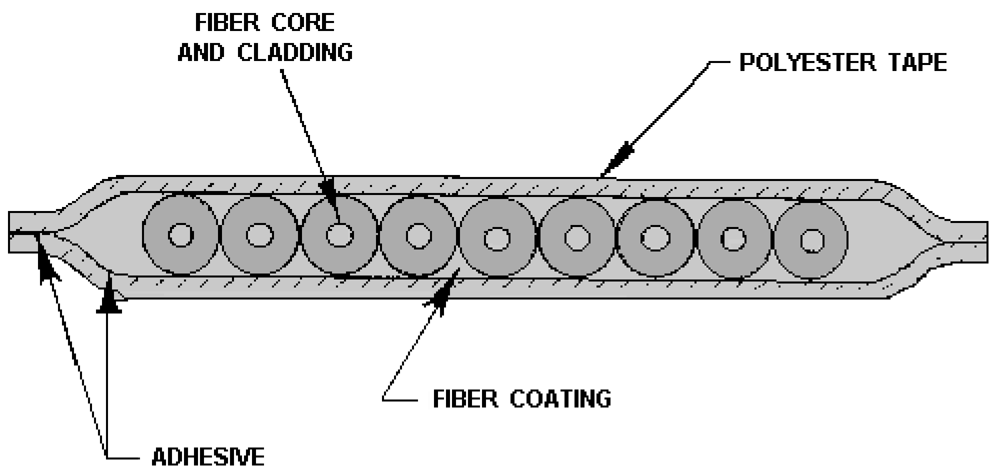


Figure 3-17.—Cross section of a fiber optic ribbon.

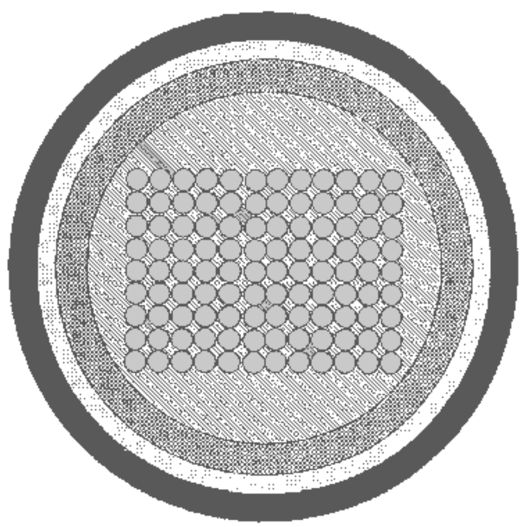


Figure 3-18.—Ribbon cable cross-sectional array of fibers.

The ribbon cable design has the highest fiber capacity. Ribbon cables can hold 204 fibers in a 0.5-inch cable. However, ribbon cables have worse bend performance than OFCC and stranded cables. Ribbon cables also have the poorest waterblocking capabilities of the three cable designs. The bend performance is expected to worsen if manufacturers add appropriate compounds to increase waterblocking capabilities.

Ribbon cables are also hard to handle. Individual fibers are highly susceptible to damage when separated from the ribbon. This susceptibility to fiber damage during fiber breakout makes it necessary to perform multifiber connections. Multifiber connections can introduce single points of failure in multiple systems. The use of multifiber terminations also introduces maintenance, reconfiguration, and repair problems. Currently, the Navy does not recommend the use of ribbon cables in shipboard systems.

Q30. List the three types of cable designs being considered by the Navy.

Q31. Describe an optical fiber cable component (OFCC).

Q32. Two layers of arimid yarn strength members encase the OFCC units. Why are these strength members stranded in opposing directions?

Q33. Why do cable manufacturers introduce a controlled twist to the stacked ribbons during the cabling process?

Q34. OFCC, stranded, and ribbon cables have different fiber capacities. What is the approximate number of fibers that each cable can accommodate in a 0.5-inch cable?

Q35. Which fiber optic cable (OFCC, stranded, or ribbon) has the worst bend performance?

SUMMARY

Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas that you have learned. You should have a thorough understanding of these principles before moving on to chapter 4.

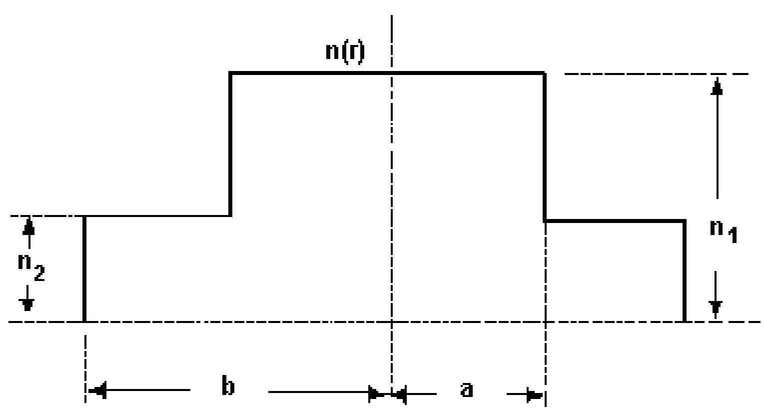
OPTICAL FIBER CLASSIFICATION depends on more than the number of modes that a fiber can propagate. The optical fiber's refractive index profile and core size further distinguish different types of single mode and multimode fibers.

The **REFRACTIVE INDEX PROFILE** describes the value of the fiber's refractive index as a function of axial distance at any fiber diameter.

In **STEP-INDEX** fibers, the refractive index of the core is uniform and undergoes an abrupt change at the core-cladding boundary.

In **GRADED-INDEX** fibers, the refractive index of the core varies gradually as a function of radial distance from the fiber center.

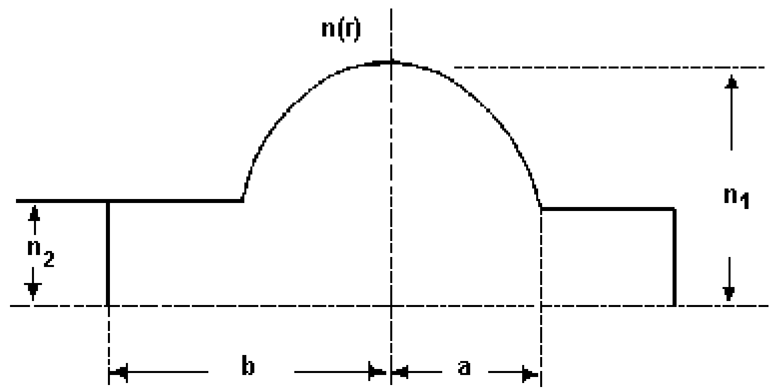
MULTIMODE STEP-INDEX FIBERS have a core of radius (a), and a constant refractive index n_1. A cladding of slightly lower refractive index n_2 surrounds the core.



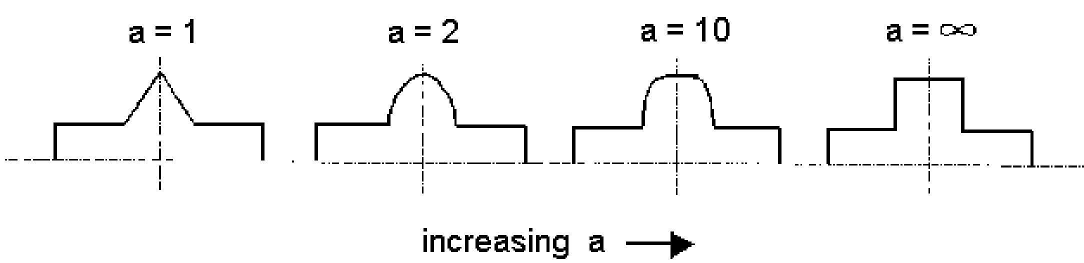
The **RELATIVE REFRACTIVE INDEX DIFFERENCE (Δ)** is the difference in the core and cladding refractive index. The ability of the fiber to accept optical energy from a light source is related to Δ.

MULTIMODE STEP-INDEX FIBERS have relatively large core diameters and large numerical apertures. Unfortunately, multimode step-index fibers have limited bandwidth capabilities and poor bend performance. Short-haul, limited bandwidth, low-cost applications use multimode step-index fibers.

MULTIMODE GRADED-INDEX FIBERS have a core of radius (a). Unlike step-index fibers, the value of the refractive index of the core (n_1) varies according to the radial distance (r). The value of n_1 decreases until it approaches the value of the refractive index of the cladding (n_2). Like the step-index fiber, the value of n_2 is constant and has a slightly lower refractive index than n_1.



The **PROFILE PARAMETER** (α) determines the shape of the core's refractive index profile. As the value of α increases, the shape of the core's profile changes from a triangular shape to a step.



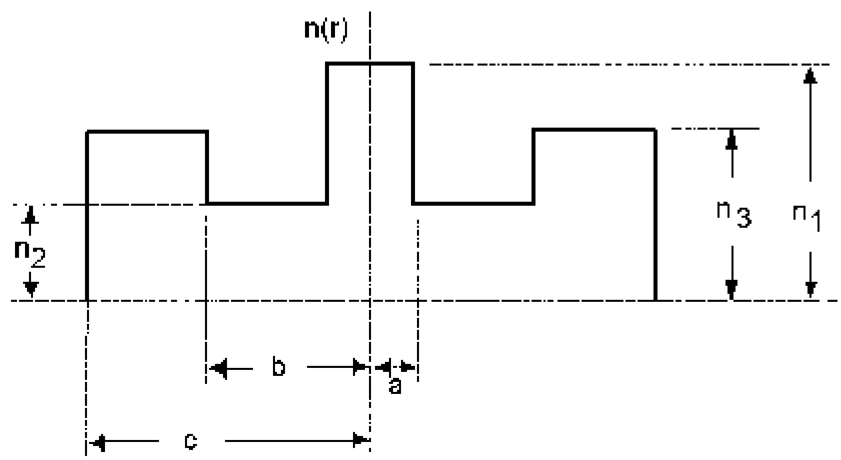
The gradual decrease in the core's refractive index from the center of the fiber causes propagating modes to be refracted many times.

Multimode graded-index fibers have less **MODAL DISPERSION** than multimode step-index fibers. Lower modal dispersion means that multimode graded-index fibers have higher bandwidth capabilities than multimode step-index fibers.

SOURCE-TO-FIBER COUPLING EFFICIENCY and **INSENSITIVITY TO MICROBENDING AND MACROBENDING LOSSES** are distinguishing characteristics of multimode graded-index fibers. 62.5 μm fibers offer the best overall performance for multimode graded-index fibers.

Coupled power increases with both core diameter and Δ, while bending losses increase directly with core diameter and inversely with Δ. However, a smaller Δ improves fiber bandwidth.

MATCHED-CLAD and **DEPRESSED-CLAD** are two types of single mode step-index fibers. Matched cladding means that the fiber cladding is a single homogeneous layer of dielectric material. Depressed cladding means that the fiber cladding consists of two regions: an inner and outer cladding region.



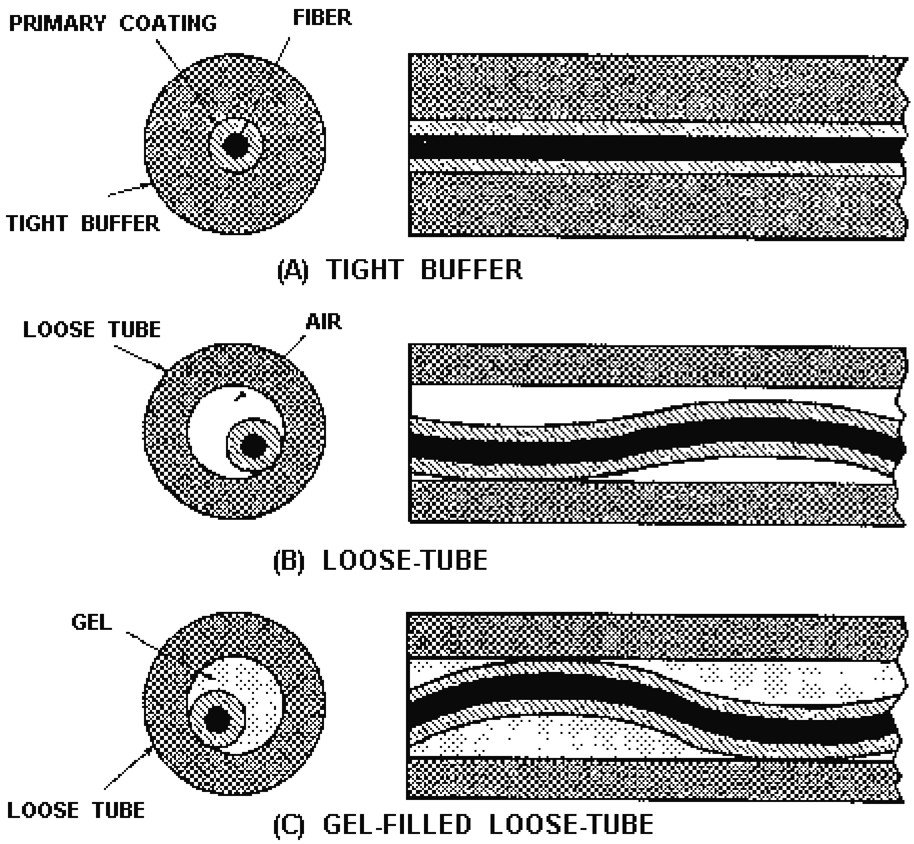
SINGLE MODE FIBER CUTOFF WAVELENGTH is the smallest operating wavelength where single mode fibers propagate only the fundamental mode. At this wavelength, the 2nd-order mode becomes lossy and radiates out of the fiber core.

SINGLE MODE FIBERS have low attenuation and high-bandwidth properties. Present applications for single mode fibers include long-haul, high-speed telecommunication systems.

VAPOR PHASE OXIDATION and **DIRECT-MELT PROCESS** are two methods of fabricating multimode and single mode optical fibers.

CABLE STRUCTURES include buffers, strength members, and the jacket, or sheath.

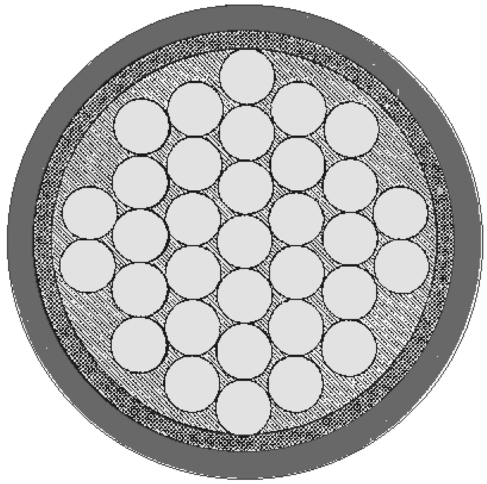
TIGHT-BUFFERED, **LOOSE-TUBE**, and **GEL-FILLED LOOSE-TUBE** are types of fiber optic buffering techniques.



FIBER OPTIC CABLES use strength members to increase the cable's strength and protect the optical fibers from strain.

JACKET MATERIAL should have low smoke generation, low toxicity, low-halogen content, flame retardance, fluid resistance, high abrasion resistance, and stable performance over temperature.

Navy systems require that fiber optic cables meet stringent environmental conditions. The types of cable designs considered by the Navy include the **OPTICAL FIBER CABLE COMPONENT (OFCC)**, **STRANDED**, and **RIBBON** cable designs.



ANSWERS TO QUESTIONS Q1. THROUGH Q35.

A1. Refractive index profile describes the value of refractive index as a function of radial distance at any fiber diameter.

A2. Step-index.

A3. Multimode graded-index fiber.

A4. Multimode step-index fibers: 50 μm and 100 μm. Multimode graded-index fibers: 50 μm, 62.5 μm, 85 μm, and 100 μm. Single mode fibers: between 8 μm and 10 μm.

A5. Cladding.

A6. Most modes in multimode step-index fibers propagate far from cutoff.

A7. Make it easier to couple light from a light-emitting diode (LED) into the fiber.

A8. From a triangular shape to step.

A9. When the angle of incidence becomes larger than the critical angle of incidence.

A10. Numerical aperture (NA), relative refractive index difference (Δ), profile parameter (α), and normalized frequency (V).

A11. Decreases the time difference between light rays, which reduces modal dispersion and increases fiber bandwidth.

A12. 62.5/125 μm multimode graded-index fiber.

A13. Source-to-fiber coupling efficiency and insensitivity to microbending and macrobending losses.

A14. Coupling efficiency increases with both core diameter and Δ, while bending losses increase directly with core diameter and inversely with Δ.

A15. Smaller.

A16. Matched-clad and depressed-clad.

A17. Depressed.

A18. Cladding material.

A19. The fundamental mode becomes increasingly lossy.

A20. To increase performance and reduce losses caused by bending and splicing.

A21. To increase performance and reduce cost.

A22. High NA, low bandwidth, tight bend radius, short length, and low cost.

A23. Heavy-metal fluorides, chalcogenide glasses, and crystalline materials.

A24. Vapor phase oxidation and direct-melt process.

A25. Vapor phase oxidation.

A26.

- *a. Protect optical fibers from damage or breakage during installation and over the fiber's lifetime.*
- *b. Provide stable fiber transmission characteristics compared with uncabled fibers.*
- *c. Maintain the physical integrity of the optical fiber.*

A27. To provide additional mechanical protection and preserve the fiber's inherent strength.

A28. Tight-buffered, loose-tube, and gel-filled loose-tube.

A29. Low smoke generation, low toxicity, low halogen content, flame retardance, fluid resistance, high abrasion resistance, and stable performance over temperature.

A30. Optical fiber cable component (OFCC), stranded, and ribbon cables designs.

A31. OFCCs are tight-buffer fiber surrounded by arimid yarn and a low-halogen outer jacket.

A32. To minimize microbending of the fibers.

A33. To minimize fiber stress when the cable is bent.

A34. OFCC (12 fibers), stranded (48 fibers), ribbon (204 fibers).

A35. Ribbon.

CHAPTER 4

OPTICAL SPLICES, CONNECTORS, AND COUPLERS

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to do the following:

1. Describe a fiber optic splice, connector, and coupler and the types of connections they form in systems.
2. List the types of extrinsic and intrinsic coupling losses.
3. Understand the degree to which fiber alignment and fiber mismatch problems increase system loss.
4. Detail the score-and-break cleaving process for fiber-end preparation.
5. Identify the types of fiber optic mechanical and fusion splices. Outline the basic splicing techniques for each type of fiber optic splice.
6. List the types of fiber optic connectors. Detail the procedure for installing a fiber optic connector on an optical fiber.
7. Discuss the types of fiber optic passive couplers.

FIBER OPTIC CONNECTIONS

Chapter 1 states that a fiber optic data link performs three basic functions. First, the data link transmitter converts an electrical input signal to an optical signal. Then, the optical fiber transmits this optical signal. Finally, the data link receiver converts the optical signal back to an electrical signal identical to the original input. However, chapter 1 does not describe how optical power transfers from one optical component to another.

This chapter describes how optical power is transferred from one fiber optic component to another. It describes how an optical source launches optical power into a fiber as well as how one optical fiber couples light into another fiber. In fiber optic system design, this launching or coupling of optical power from one component to the next is important.

Fiber optic connections permit the transfer of optical power from one component to another. Fiber optic connections also permit fiber optic systems to be more than just point-to-point data communication links. In fact, fiber optic data links are often of a more complex design than point-to-point data links.

A system connection may require either a fiber optic splice, connector, or coupler. One type of system connection is a permanent connection made by splicing optical fibers together. A fiber optic **splice** makes a permanent joint between two fibers or two groups of fibers. There are two types of fiber optic splices--mechanical splices and fusion splices. Even though removal of some mechanical splices is possible, they are intended to be permanent. Another type of connection that allows for system reconfiguration is a fiber optic **connector**. Fiber optic connectors permit easy coupling and uncoupling of optical fibers. Fiber optic connectors sometimes resemble familiar electrical plugs and sockets. Systems may also divide or combine optical signals between fibers. Fiber optic **couplers** distribute or combine optical signals between fibers. Couplers can distribute an optical signal from a single fiber into several fibers. Couplers may also combine optical signals from several fibers into one fiber.

Fiber optic connection losses may affect system performance. **Poor fiber end preparation** and **poor fiber alignment** are the main causes of coupling loss. Another source of coupling loss is differences in optical properties between the connected fibers. If the connected fibers have different optical properties, such as different numerical apertures, core and cladding diameters, and refractive index profiles, then coupling losses may increase.

Q1. Which fiber optic component (splice, connector, or coupler) makes a permanent connection in a distributed system?

Q2. What are the main causes of coupling loss?

OPTICAL FIBER COUPLING LOSS

Ideally, optical signals coupled between fiber optic components are transmitted with no loss of light. However, there is always some type of imperfection present at fiber optic connections that causes some loss of light. It is the amount of optical power lost at fiber optic connections that is a concern of system designers.

The design of fiber optic systems depends on how much light is launched into an optical fiber from an optical source and how much light is coupled between fiber optic components, such as from one fiber to another. The amount of power launched from a source into a fiber depends on the optical properties of both the source and the fiber. The amount of optical power launched into an optical fiber depends on the radiance of the optical source. An optical source's **radiance**, or brightness, is a measure of its optical power launching capability. Radiance is the amount of optical power emitted in a specific direction per unit time by a unit area of emitting surface. For most types of optical sources, only a fraction of the power emitted by the source is launched into the optical fiber.

The loss in optical power through a connection is defined similarly to that of signal attenuation through a fiber. Optical loss is also a log relationship. The loss in optical power through a connection is defined as:

$$\text{loss} = 10 \log_{10} \frac{P_i}{P_o}$$

For example, P_o is the power emitted from the source fiber in a fiber-to-fiber connection. P_i is the power accepted by the connected fiber. In any fiber optic connection, P_o and P_i are the optical power levels measured before and after the joint, respectively.

Fiber-to-fiber connection loss is affected by intrinsic and extrinsic coupling losses. **Intrinsic coupling losses** are caused by inherent fiber characteristics. **Extrinsic coupling losses** are caused by jointing techniques. Fiber-to-fiber connection loss is increased by the following sources of intrinsic and extrinsic coupling loss:

- Reflection losses
- Fiber separation
- Lateral misalignment
- Angular misalignment
- Core and cladding diameter mismatch
- Numerical aperture (NA) mismatch
- Refractive index profile difference
- Poor fiber end preparation

Intrinsic coupling losses are limited by reducing fiber mismatches between the connected fibers. This is done by procuring only fibers that meet stringent geometrical and optical specifications. Extrinsic coupling losses are limited by following proper connection procedures.

Some fiber optic components are modular devices that are designed to reduce coupling losses between components. Modular components can be easily inserted or removed from any system. For example, fiber optic transmitters and receivers are modular components. Fiber optic transmitters and receivers are devices that are generally manufactured with fiber pigtails or fiber optic connectors as shown in figure 4-1. A **fiber pigtail** is a short length of optical fiber (usually 1 meter or less) permanently fixed to the optical source or detector. Manufacturers supply transmitters and receivers with pigtails and connectors because fiber coupling to sources and detectors must be completed during fabrication. Reduced coupling loss results when source-to-fiber and fiber-to-detector coupling is done in a controlled manufacturing environment. Since optical sources and detectors are pigtailed or connectorized, launching optical power is reduced to coupling light from one fiber to another. In fact, most fiber optic connections can be considered fiber-to-fiber.

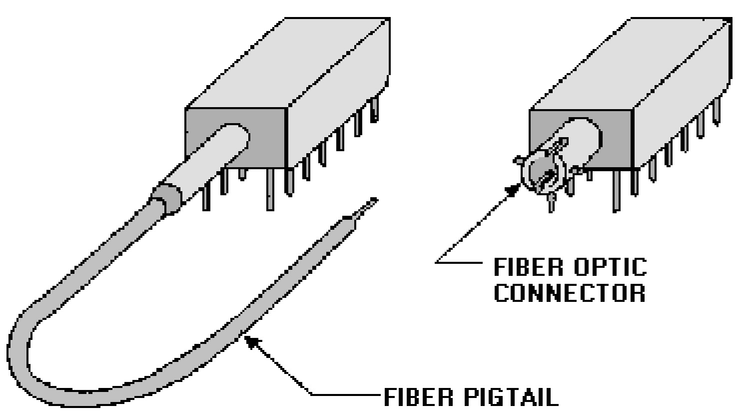


Figure 4-1.—Pigtailed and connectorized fiber optic devices.

Q3. Define the loss in optical power through a connection.

Q4. Fiber-to-fiber coupling loss is affected by intrinsic and extrinsic coupling losses. Can intrinsic coupling losses be limited by limiting fiber mismatches?

REFLECTION LOSSES

When optical fibers are connected, optical power may be reflected back into the source fiber. Light that is reflected back into the source fiber is lost. This reflection loss, called Fresnel reflection, occurs at every fiber interface. **Fresnel reflection** is caused by a step change in the refractive index that occurs at the fiber joint. In most cases, the step change in refractive index is caused by the ends of each fiber being separated by a small gap. This small gap is usually an air gap. In Fresnel reflection, a small portion of the incident light is reflected back into the source fiber at the fiber interface. The ratio (R), shown below, approximates the portion of incident light (light of normal incidence) that is reflected back into the source fiber.

$$R = \left(\frac{n_1 - n_0}{n_1 + n_0} \right)^2$$

R is the fraction of the incident light reflected at the fiber n_1 is the refractive index of the fiber core. n_0 is the refractive index of the medium between the two fibers.

Fresnel refraction occurs twice in a fiber-to-fiber connection. A portion of the optical power is reflected when the light first exits the source fiber. Light is then reflected as the optical signal enters the receiving fiber. Fresnel reflection at each interface must be taken into account when calculating the total fiber-to-fiber coupling loss. Loss from Fresnel reflection may be significant. To reduce the amount of loss from Fresnel reflection, the air gap can be filled with an index matching gel. The refractive index of the index matching gel should match the refractive index of the fiber core. **Index matching gel** reduces the step change in the refractive index at the fiber interface, reducing Fresnel reflection.

In any system, index matching gels can be used to eliminate or reduce Fresnel reflection. The choice of index matching gels is important. Fiber-to-fiber connections are designed to be permanent and require no maintenance. Over the lifetime of the fiber connection, the index matching material must meet specific optical and mechanical requirements. Index matching gels should remain transparent. They should also resist flowing or dripping by remaining viscous. Some index matching gels darken over time while others settle or leak out of fiber connections. If these requirements are not met, then the fiber-to-fiber connection loss will increase over time. In Navy applications, this variation in connection loss over time is unacceptable. In Navy systems, index matching gels are only used in fiber optic splice interfaces.

Q5. In fiber-to-fiber connections, Fresnel reflection is one source of coupling losses. Light is reflected back into the source fiber and is lost. What causes Fresnel reflection?

Q6. Reduction of Fresnel reflection is possible by reducing the step change in the refractive index at the fiber interface. What material reduces the step change in refractive index at a fiber interface?

FIBER ALIGNMENT

A main source of extrinsic coupling loss in fiber-to-fiber connections is poor fiber alignment. The three basic coupling errors that occur during fiber alignment are fiber separation (longitudinal misalignment), lateral misalignment, and angular misalignment. Most alignment errors are the result of mechanical imperfections introduced by fiber jointing techniques. However, alignment errors do result from installers not following proper connection procedures.

With **fiber separation**, a small gap remains between fiber-end faces after completing the fiber connection. Figure 4-2 illustrates this separation of the fiber-end faces.

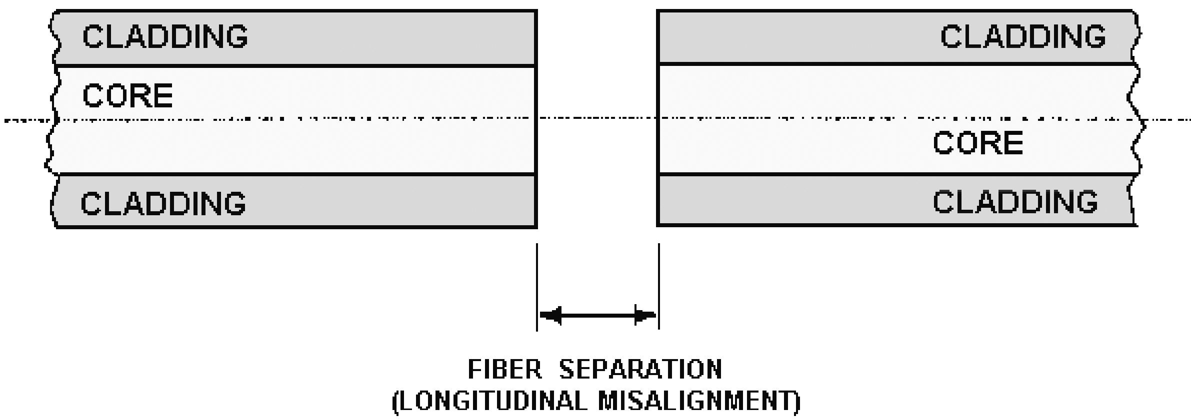


Figure 4-2.—Fiber separation.

Lateral, or axial, misalignment occurs when the axes of the two fibers are offset in a perpendicular direction. Figure 4-3 shows this perpendicular offset of the axes of two connecting fibers.

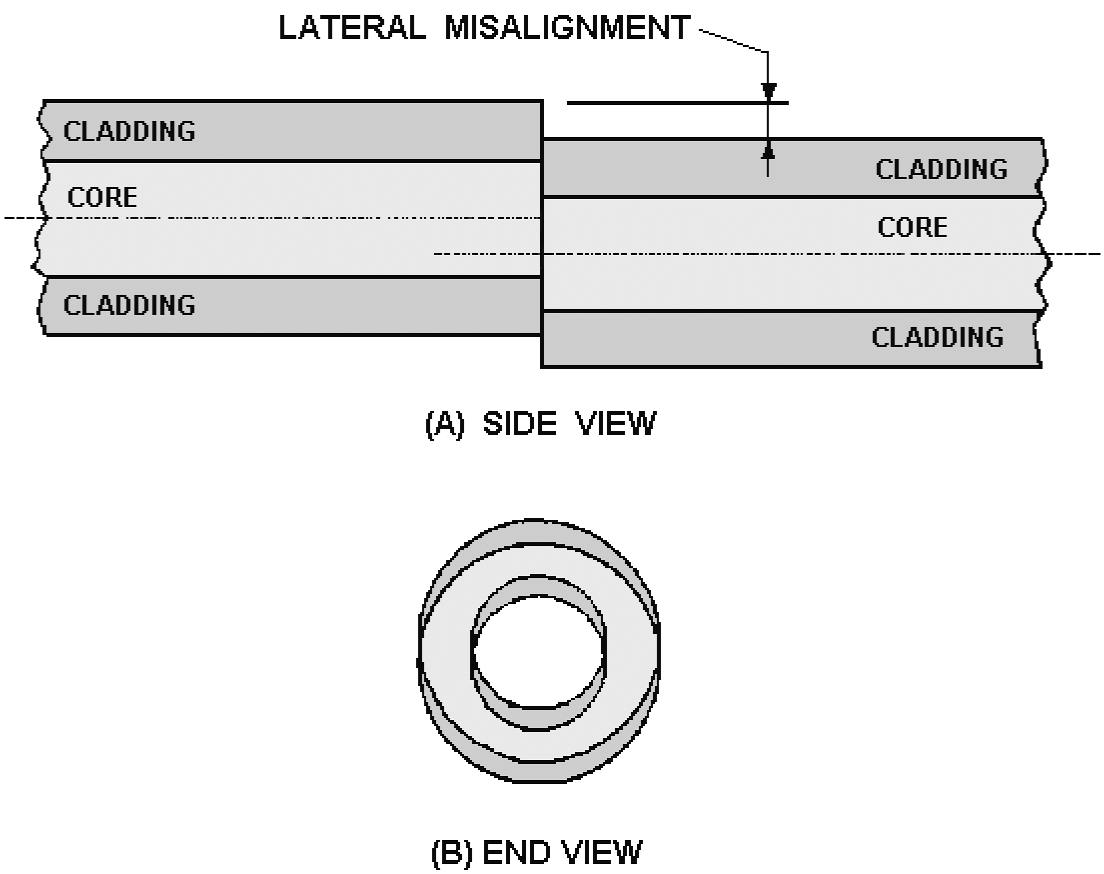


Figure 4-3.—Lateral misalignment.

Angular misalignment occurs when the axes of two connected fibers are no longer parallel. The axes of each fiber intersect at some angle (Θ). Figure 4-4 illustrates the angular misalignment between the core axes.

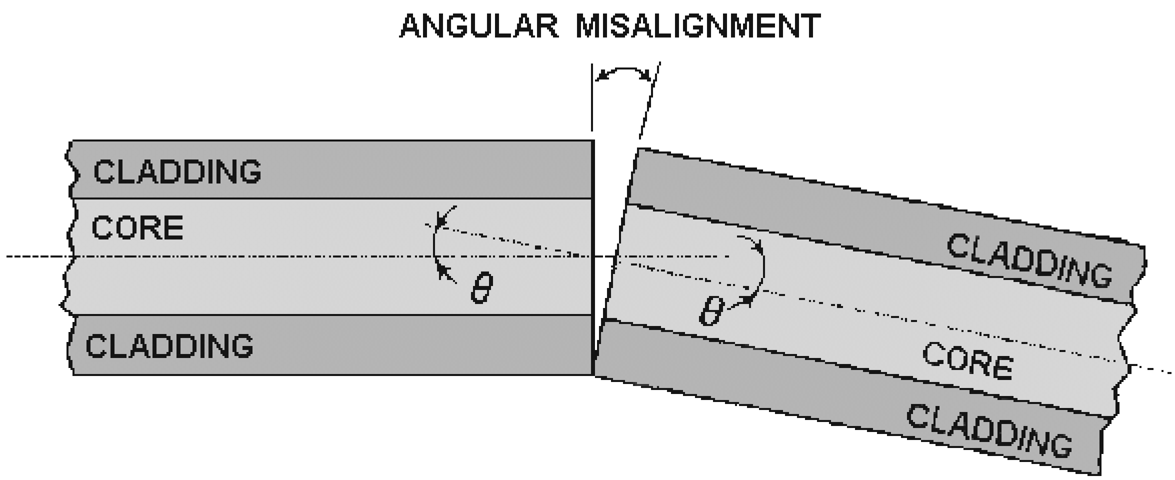


Figure 4-4.—Angular misalignment.

Coupling loss caused by lateral and angular misalignment typically is greater than the loss caused by fiber separation. Loss, caused by fiber separation, is less critical because of the relative ease in limiting the distance of fiber separation. However, in some cases, fiber optic connectors prevent fibers from actual contact. These fiber optic connectors separate the fibers by a small gap. This gap eliminates damage to fiber-end faces during connection. For connectors with an air gap, the use of index matching gel reduces the coupling loss.

Most newer connectors are designed so that the connector ferrule end faces contact when the connector is mated. The connector can be assembled onto the fiber so that the fibers also contact when mated. However, they also can be assembled so that the fibers do not. Whether or not the fibers contact is determined by whether the fiber sticks out slightly from the ferrule or is recessed inside the ferrule. The fiber position can be controlled by the connector polishing technique. The physical contact (PC) polish technique was developed for most connectors so that the fibers would touch when mated. In these types of connectors, index gel is not needed to reduce reflections.

While index matching gel reduces coupling loss from fiber separation, it does not affect loss in lateral misalignment. Additionally, index matching gel usually increases the fiber's coupling loss sensitivity to angular misalignment. Although angular misalignment involves fiber separation, index matching gel reduces the angle at which light is launched from the source fiber. Index matching gel causes less light to be coupled into the receiving fiber. To reduce coupling loss from angular misalignment, the angle Θ should be less than 1º.

Coupling losses due to fiber alignment depend on fiber type, core diameter, and the distribution of optical power among propagating modes. Fibers with large NAs reduce loss from angular misalignment and increase loss from fiber separation. Single mode fibers are more sensitive to alignment errors than multimode fibers because of their small core size. However, alignment errors in multimode fiber connections may disturb the distribution of optical power in the propagating modes, increasing coupling loss.

Q7. List the three basic errors that occur during fiber alignment.

Q8. When the axes of two connected fibers are no longer in parallel, the two connected fibers are in what kind of misalignment?

Q9. How does index matching gel affect the amount of coupling loss caused by (a) fiber separation, (b) lateral misalignment, and (c) angular misalignment?

Q10. Which are more sensitive to alignment errors, single mode or multimode fibers?

FIBER END PREPARATION

In fiber-to-fiber connections, a source of extrinsic coupling loss is poor fiber end preparation. An optical fiber-end face must be flat, smooth, and perpendicular to the fiber's axis to ensure proper fiber connection. Light is reflected or scattered at the connection interface unless the connecting fiber end faces are properly prepared. Figure 4-5 shows some common examples of poor fiber ends. It illustrates a fiber-end face **tilt, lip,** and **hackle**. Quality fiber-end preparation is essential for proper system operation.

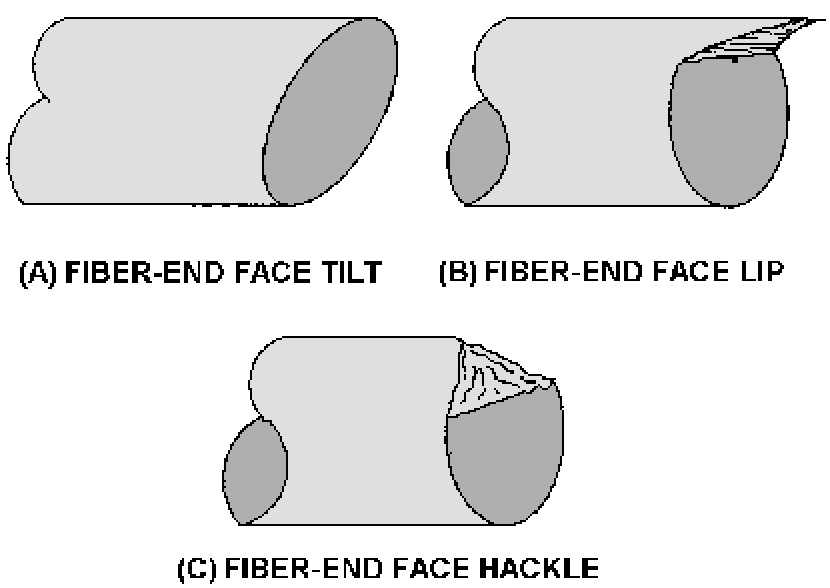


Figure 4-5.—Poor fiber-end preparation.

Fiber-end preparation begins by removing the fiber buffer and coating material from the end of the optical fiber. Removal of these materials involves the use of mechanical strippers or chemical solvents. When using chemical solvents, the removal process must be performed in a well-ventilated area. For this reason mechanical strippers are used for buffer and coating removal in the shipboard environment. After removing the buffer and coating material, the surface of the bare fiber is wiped clean using a wiping tissue. The wiping tissue must be wet with isopropyl alcohol before wiping.

The next step in fiber-end preparation involves cleaving the fiber end to produce a smooth, flat fiber-end face. The **score-and-break**, or scribe-and-break, method is the basic fiber cleaving technique for preparing optical fibers for coupling. The score-and-break method consists of lightly scoring (nicking) the outer surface of the optical fiber and then placing it under tension until it breaks. A heavy metal or diamond blade is used to score the fiber. Once the scoring process is complete, fiber tension is increased until the fiber breaks. The fiber is placed under tension either by pulling on the fiber or by bending the fiber over a curved surface.

Figure 4-6 shows the setup for the score-and-break procedure for fiber cleaving. Under constant tension, the score-and-break method for cleaving fibers produces a quality fiber end. This fiber end is good enough to use for some splicing techniques. However, additional fiber-end preparation is necessary to produce reliable low-loss connections when using fiber optic connectors.

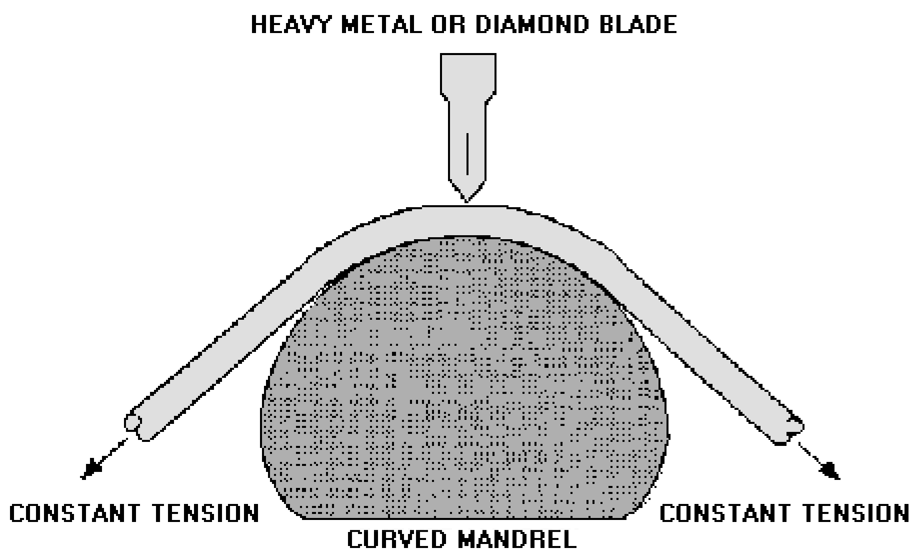


Figure 4-6.—Score-and-break procedure for fiber cleaving.

Polishing the fiber ends removes most surface imperfections introduced by the fiber cleaving process. Fiber polishing begins by inserting the cleaved fiber into the ferrule of a connector assembly. A ferrule is a fixture, generally a rigid tube, used to hold the stripped end of an optical fiber in a fiber optic connector. An individual fiber is epoxied within the ferrule. The connector with the optical fiber cemented within the ferrule can then be mounted into a special polishing tool for polishing.

Figure 4-7 shows one type of fiber polishing tool for finishing optical fibers in a connector assembly. Various types of connector assemblies are discussed later in this chapter. In this type of polishing tool, the connector assembly is threaded onto the polishing tool. The connector ferrule passes through the center of the tool allowing the fiber-end face to extend below the tool's circular, flat bottom. The optical fiber is now ready for polishing.

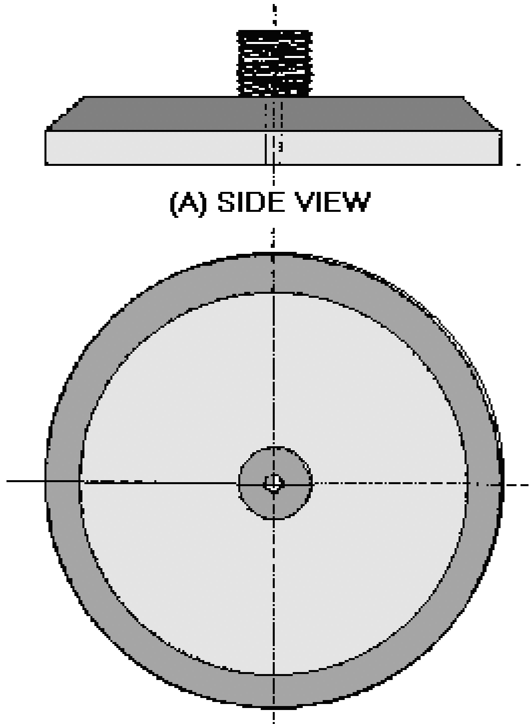


Figure 4-7.—Fiber polishing tool.

Fiber polishing involves a step-down approach. The first step is to give the surface of the fiber end a rough polish. **Rough-polishing** occurs when the fiber, mounted to the polishing tool, moves over a 5μ to 15μ grit abrasive paper. The mounted fiber moves over the abrasive paper in a figure-eight motion. The next step involves giving the surface of the fiber end a fine polish. **Fine-polishing** occurs when the mounted fiber moves over a 0.3μto 1μ grit abrasive paper in the same figure-eight motion. Fiber inspection and cleanliness are important during each step of fiber polishing. Fiber inspection is done visually by the use of a standard microscope at 200 to 400 times magnification.

A standard microscope can be used to determine if the fiber-end face is flat, concave, or convex. If different parts of the fiber-end face have different focus points, the end face is not flat. If all parts of the fiber-end face are in focus at the same time, the end face is flat.

Q11. Quality fiber-end preparation is essential for proper system operation. What properties must an optical fiber-end face have to ensure proper fiber connection?

Q12. What is the basic fiber cleaving technique for preparing optical fibers for coupling?

Q13. Using a standard microscope to inspect a fiber-end face, you observe that all parts of the fiber-end face are in focus at the same time. Is the fiber-end face flat, concave, or convex?

FIBER MISMATCHES

Fiber mismatches are a source of intrinsic coupling loss. As stated before, intrinsic coupling loss results from differences (mismatches) in the inherent fiber characteristics of the two connecting fibers. Fiber mismatches occur when manufacturers fail to maintain optical or structural (geometrical) tolerances during fiber fabrication.

Fiber mismatches are the result of inherent fiber characteristics and are independent of the fiber jointing techniques. Types of fiber mismatches include fiber geometry mismatches, NA mismatch, and refractive index profile difference. Fiber geometry mismatches include core diameter, cladding diameter, core ellipticity, and core-cladding concentricity differences. Figure 4-8 illustrates each type of optical and geometrical fiber mismatch. Navy fiber specifications tightly specify these parameters to minimize coupling losses from fiber mismatches.

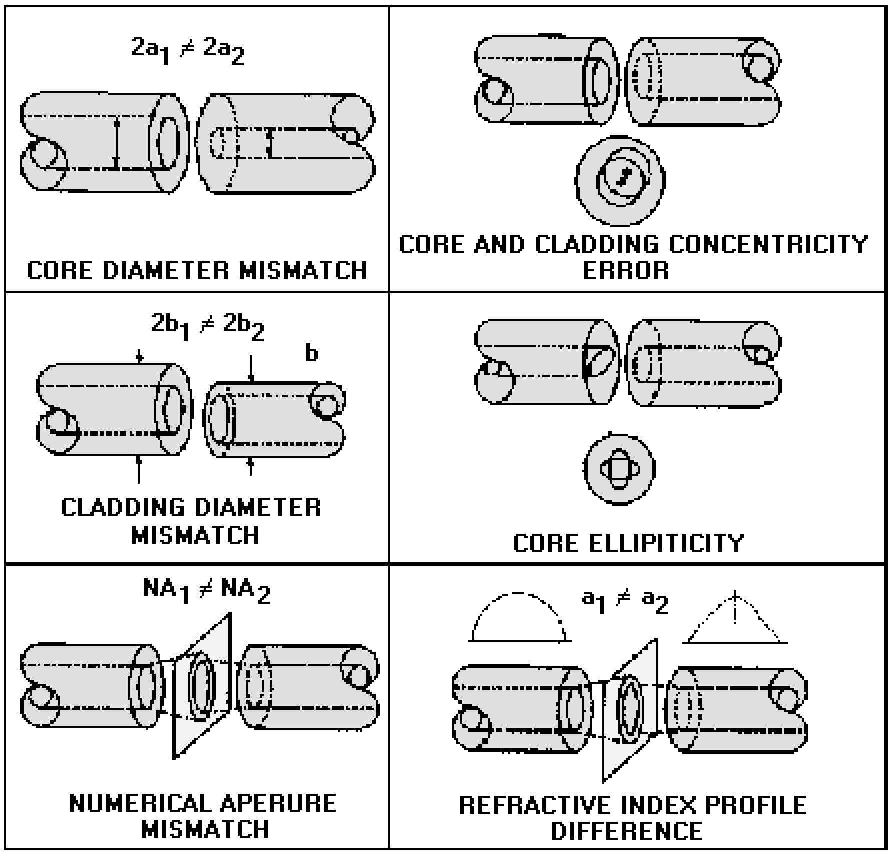


Figure 4-8.—Types of optical and geometrical fiber mismatches that cause intrinsic coupling loss.

Core diameter and NA mismatch have a greater effect on intrinsic coupling loss than the other types of fiber mismatches. In multimode fiber connections, the coupling loss resulting from core diameter mismatch, NA mismatch, and refractive index profile difference depends on the characteristics of the launching fiber. Coupling loss from **core diameter mismatch** results only if the launching fiber has a larger core radius (a) than the receiving fiber. Coupling loss from **NA mismatch** results only if the launching fiber has a higher NA than the receiving fiber. Coupling loss from **refractive index profile difference** results only if the launching fiber has a larger profile parameter (α) than the receiving fiber.

Q14. List six types of fiber mismatches.

Q15. Does coupling loss from refractive index profile difference result when the receiving fiber has a larger profile parameter (α) than the transmitting fiber?

FIBER OPTIC SPLICES

A **fiber optic splice** is a permanent fiber joint whose purpose is to establish an optical connection between two individual optical fibers. System design may require that fiber connections have specific optical properties (low loss) that are met only by fiber-splicing. Fiber optic splices also permit repair of optical fibers damaged during installation, accident, or stress. System designers generally require fiber splicing whenever repeated connection or disconnection is unnecessary or unwanted.

Mechanical and fusion splicing are two broad categories that describe the techniques used for fiber splicing. A **mechanical splice** is a fiber splice where mechanical fixtures and materials perform fiber alignment and connection. A **fusion splice** is a fiber splice where localized heat fuses or melts the ends of two optical fibers together. Each splicing technique seeks to optimize splice performance and reduce splice loss. Low-loss fiber splicing results from proper fiber end preparation and alignment.

Fiber splice alignment can involve passive or active fiber core alignment. Passive alignment relies on precision reference surfaces, either grooves or cylindrical holes, to align fiber cores during splicing. Active alignment involves the use of light for accurate fiber alignment. Active alignment may consist of either monitoring the loss through the splice during splice alignment or by using a microscope to accurately align the fiber cores for splicing. To monitor loss either an optical source and optical power meter or an optical time domain reflectometer (OTDR) are used. Active alignment procedures produce low-loss fiber splices.

Q16. Define a fiber optic splice.

Q17. Fiber splicing is divided into two broad categories that describe the techniques used for fiber splicing. What are they?

MECHANICAL SPLICES

Mechanical splicing involves using mechanical fixtures to align and connect optical fibers. Mechanical splicing methods may involve either passive or active core alignment. Active core alignment produces a lower loss splice than passive alignment. However, passive core alignment methods can produce mechanical splices with acceptable loss measurements even with single mode fibers.

In the strictest sense, a mechanical splice is a permanent connection made between two optical fibers. Mechanical splices hold the two optical fibers in alignment for an indefinite period of time without movement. The amount of splice loss is stable over time and unaffected by changes in environmental or mechanical conditions.

If high splice loss results from assembling some mechanical splices, the splice can be reopened and the fibers realigned. Realignment includes wiping the fiber or ferrule end with a soft wipe, reinserting the fiber or ferrule in a new arrangement, and adding new refractive index material. Once producing an acceptable mechanical splice, splice realignment should be unnecessary because most mechanical splices are environmentally and mechanically stable within their intended application.

The types of mechanical splices that exist for mechanical splicing include glass, plastic, metal, and ceramic tubes; and V-groove and rotary devices. Materials that assist mechanical splices in splicing fibers include transparent adhesives and index matching gels. **Transparent adhesives** are epoxy resins that seal mechanical splices and provide index matching between the connected fibers.

Glass or Ceramic Alignment Tube Splices

Mechanical splicing may involve the use of a glass or ceramic alignment tube, or capillary. The inner diameter of this glass or ceramic tube is only slightly larger than the outer diameter of the fiber. A transparent adhesive, injected into the tube, bonds the two fibers together. The adhesive also provides index matching between the optical fibers. Figure 4-9 illustrates fiber alignment using a glass or ceramic tube. This splicing technique relies on the inner diameter of the alignment tube. If the inner diameter is too large, splice loss will increase because of fiber misalignment. If the inner diameter is too small, it is impossible to insert the fiber into the tube.

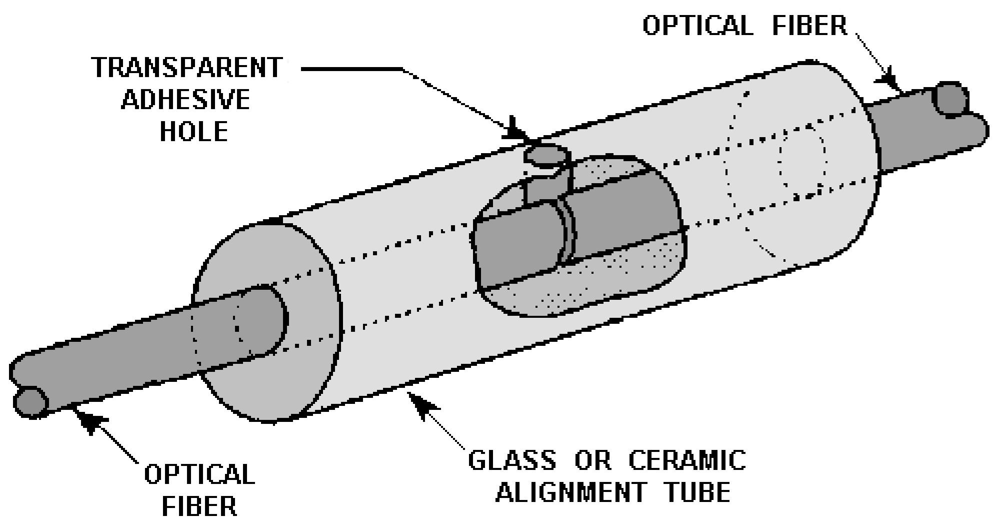


Figure 4-9.—A glass or ceramic alignment tube for mechanical splicing.

V-Grooved Splices

Mechanical splices may also use either a grooved substrate or positioning rods to form suitable V-grooves for mechanical splicing. The basic V-grooved device relies on an open grooved substrate to perform fiber alignment. When inserting the fibers into the grooved substrate, the V-groove aligns the cladding surface of each fiber end. A transparent adhesive makes the splice permanent by securing the fiber ends to the grooved substrate. Figure 4-10 illustrates this type of open V-grooved splice.

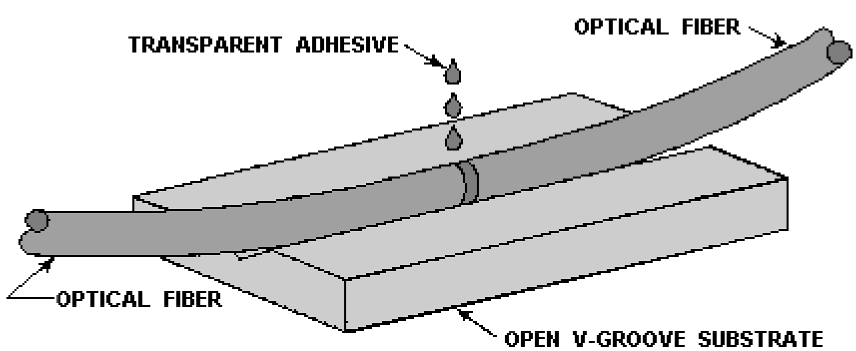


Figure 4-10.—Open V-grooved splice.

V-grooved splices may involve sandwiching the butted ends of two prepared fibers between a V-grooved substrate and a flat glass plate. Additional V-grooved devices use two or three positioning rods to form a suitable V-groove for splicing. The V-grooved device that uses two positioning rods is the spring V-grooved splice. This splice uses a groove formed by two rods positioned in a bracket to align the fiber ends. The diameter of the positioning rods permits the outer surface of each fiber end to extend above the groove formed by the rods. A flat spring presses the fiber ends into the groove maintaining fiber alignment. Transparent adhesive completes the assembly process by bonding the fiber ends and providing index matching. Figure 4-11 is an illustration of the spring V-grooved splice. A variation of this splice uses a third positioning rod instead of a flat spring. The rods are held in place by a heat-shrinkable band, or tube.

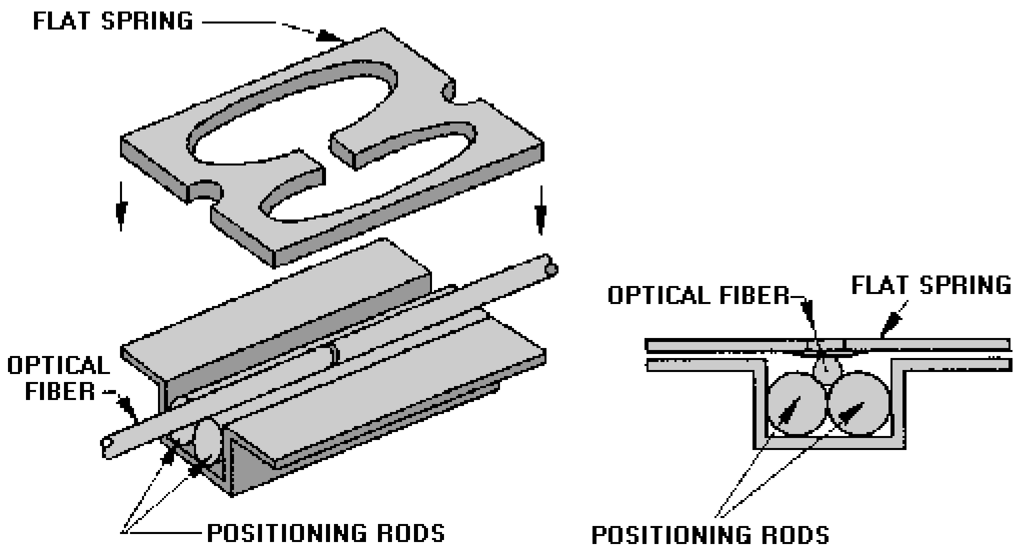


Figure 4-11.—Spring V-grooved mechanical splice.

Rotary Splices

In a rotary splice, the fibers are mounted into a glass ferrule and secured with adhesives. The splice begins as one long glass ferrule that is broken in half during the assembly process. A fiber is inserted into each half of the tube and epoxied in place using an ultraviolet cure epoxy. The endface of the tubes are then polished and placed together using the alignment sleeve. Figure 4-12 is an illustration of a rotary splice. The fiber ends retain their original orientation and have added mechanical stability since each fiber is mounted into a glass ferrule and alignment sleeve. The rotary splice may use index matching gel within the alignment sleeve to produce low-loss splices.

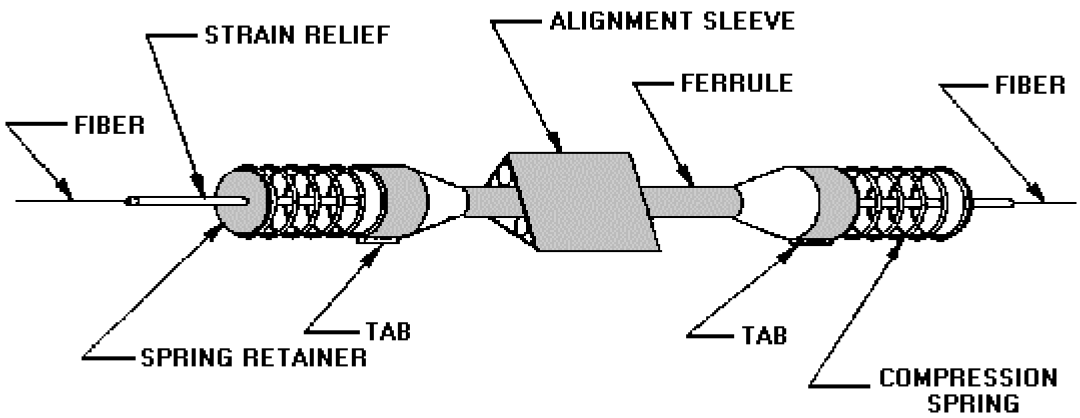


Figure 4-12.—Rotary mechanical splice.

In shipboard applications, the Navy recommends using the rotary splice. The rotary splice is a low-loss mechanical splice that provides stable environmental and mechanical performance in the Navy environment. Stable performance means that splice loss does not vary significantly with changes in temperature or other environmental or mechanical conditions. Completing a rotary splice also requires only a small amount of training, or expertise. This shorter training time is another reason why the Navy recommends using the rotary splice over other mechanical or fusion splicing techniques.

Q18. Describe a transparent adhesive.

Q19. The Navy recommends using the rotary splice for what two reasons?

FUSION SPLICES

The process of fusion splicing involves using localized heat to melt or fuse the ends of two optical fibers together. The splicing process begins by preparing each fiber end for fusion. Fusion splicing requires that all protective coatings be removed from the ends of each fiber. The fiber is then cleaved using the score-and-break method. The quality of each fiber end is inspected using a microscope. In fusion splicing, splice loss is a direct function of the angles and quality of the two fiber-end faces.

The basic fusion splicing apparatus consists of two fixtures on which the fibers are mounted and two electrodes. Figure 4-13 shows a basic fusion-splicing apparatus. An inspection microscope assists in the placement of the prepared fiber ends into a fusion-splicing apparatus. The fibers are placed into the apparatus, aligned, and then fused together. Initially, fusion splicing used nichrome wire as the heating element to melt or fuse fibers together. New fusion-splicing techniques have replaced the nichrome wire with carbon dioxide (CO_2) lasers, electric arcs, or gas flames to heat the fiber ends, causing them to fuse together. The small size of the fusion splice and the development of automated fusion-splicing machines have made **electric arc fusion** (arc fusion) one of the most popular splicing techniques in commercial applications.

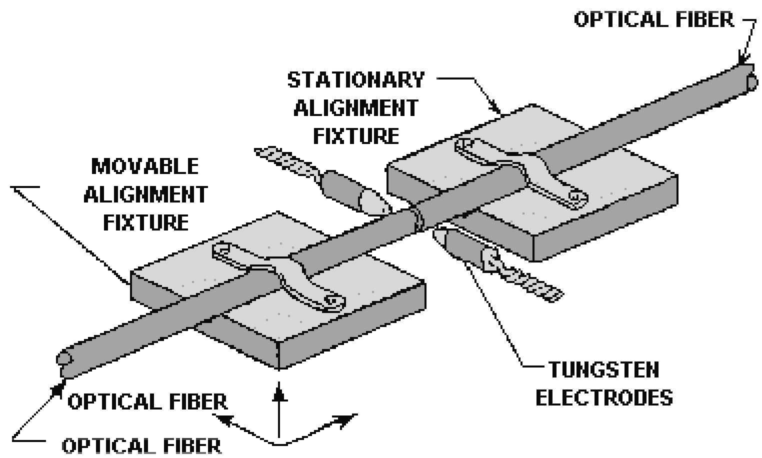


Figure 4-13.—A basic fusion splicing apparatus.

Arc fusion involves the discharge of electric current across a gap between two electrodes. By placing the fiber ends between the electrodes, the electric discharge melts or fuses the ends of each fiber. Figure 4-13 shows the placement of the fiber ends between tungsten electrodes during arc fusion. Initially, a small gap is present between the fiber ends. A short discharge of electric current is used to prepare the fiber ends for fusion. During this short discharge, known as **prefusion**, the fiber ends are cleaned and rounded to eliminate any surface defects that remain from fiber cleaving. Surface defects can cause core distortions or bubble formations during fiber fusion. A fusion splice results when the fiber ends are pressed together, actively aligned, and fused using a longer and stronger electric discharge. Automated fusion splicers typically use built-in local optical power launch/detection schemes for aligning the fibers.

During fusion, the surface tension of molten glass tends to realign the fibers on their outside diameters, changing the initial alignment. When the fusion process is complete, a small core distortion may be present. Small core distortions have negligible effects on light propagating through multimode fibers. However, a small core distortion can significantly affect single mode fiber splice loss. The core distortion, and the splice loss, can be reduced by limiting the arc discharge and decreasing the gap distance between the two electrodes. This limits the region of molten glass. However, limiting the region of molten glass reduces the tensile strength of the splice.

Fusion splicing yields typically vary between 25 and 75 percent depending on the strength and loss requirements for the splice and other factors. Other factors affecting splice yields include the condition of the splicing machine, the experience of the splice personnel, and environmental conditions. Since fusion splicing is inherently permanent, an unacceptable fusion splice requires breakage and refabrication of the splice.

In general, fusion splicing takes a longer time to complete than mechanical splicing. Also, yields are typically lower making the total time per successful splice much longer for fusion splicing. Both the yield and splice time are determined to a large degree by the expertise of the fusion splice operator. Fusion splice operators must be highly trained to consistently make low-loss reliable fusion splices. For these reasons the fusion splice is not recommended for use in Navy shipboard applications.

Q20. What fiber property directly affects splice loss in fusion splicing?

Q21. List two reasons why fusion splicing is one of the most popular splicing techniques in commercial applications.

Q22. What is a short discharge of electric current that prepares the fiber ends for fusion called?

Q23. Do small core distortions formed by arc fusion's self-alignment mechanism have more of an affect on light propagating through multimode or single mode fibers?

MULTIFIBER SPLICING

Normally, multifiber splices are only installed on ribbon type fiber optic cables. Multifiber splicing techniques can use arc fusion to restore connection, but most splicing techniques use mechanical splicing methods. The most common mechanical splice is the ribbon splice.

A ribbon splice uses an etched silicon chip, or grooved substrate, to splice the multiple fibers within a flat ribbon. The spacing between the etched grooves of the silicon chip is equal to the spacing between the fibers in the flat ribbon. Before placing each ribbon on the etched silicon chip, each fiber within the ribbon cable is cleaved. All of the fibers are placed into the grooves and held in place with a flat cover. Typically, an index matching gel is used to reduce the splice loss. Figure 4-14 shows the placement of the fiber ribbon on the etched silicon chip.

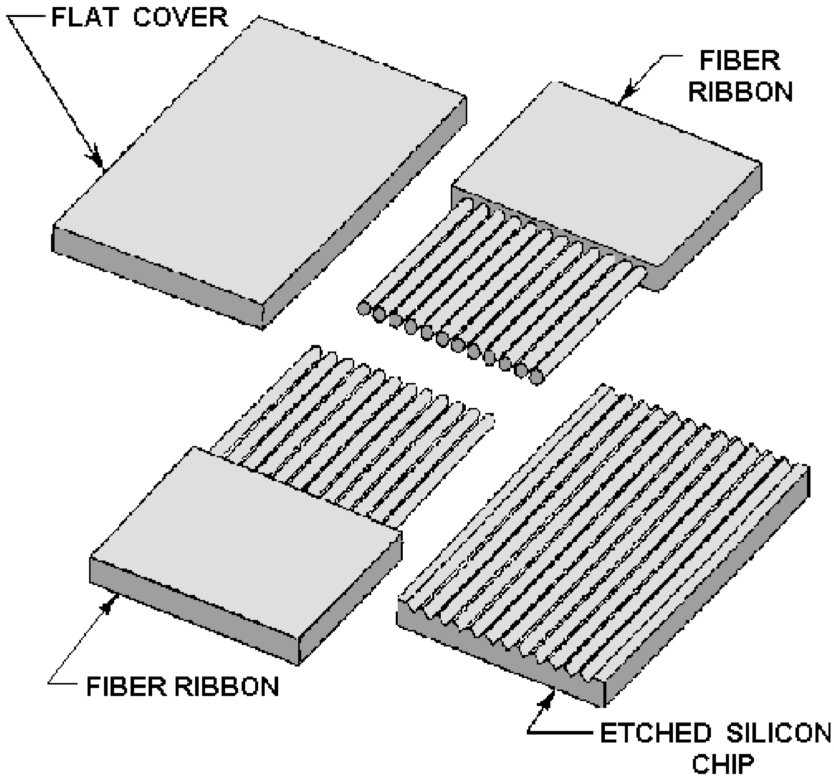


Figure 4-14.—Ribbon splice on etched silicon chip.

FIBER OPTIC CONNECTORS

A fiber optic connector is a demateable device that permits the coupling of optical power between two optical fibers or two groups of fibers. Designing a device that allows for repeated fiber coupling without significant loss of light is difficult. Fiber optic connectors must maintain fiber alignment and provide repeatable loss measurements during numerous connections. Fiber optic connectors should be easy to assemble (in a laboratory or field environment) and should be cost effective. They should also be reliable. Fiber optic connections using connectors should be insensitive to environmental conditions, such as temperature, dust, and moisture. Fiber optic connector designs attempt to optimize connector performance by meeting each of these conditions.

Fiber optic connector coupling loss results from the same loss mechanisms described earlier in this chapter. Coupling loss results from poor fiber alignment and end preparation (extrinsic losses), fiber mismatches (intrinsic loss), and Fresnel reflection. The total amount of insertion loss for fiber optic connectors should remain below 1 dB. Fiber alignment is the critical parameter in maintaining the total insertion loss below the required level. There is only a small amount of control over coupling loss resulting from fiber mismatches, because the loss results from inherent fiber properties. Index matching gels cannot be used to reduce Fresnel losses, since the index matching gels attract dust and dirt to the connection.

Fiber optic connectors can also reduce system performance by introducing modal and reflection noise. The cause of modal noise in fiber optic connectors is the interfering of the different wavefronts of different modes within the fiber at the connector interface. Modal noise is eliminated by using only single mode fiber with laser sources and only low-coherence sources such as light-emitting diodes with multimode fiber. Fiber optic connectors can introduce reflection noise by reflecting light back into the optical source. Reflection noise is reduced by index matching gels, physical contact polishes, or antireflection coatings. Generally, reflection noise is only a problem in high data rate single mode systems using lasers.

Butt-jointed connectors and **expanded-beam connectors** are the two basic types of fiber optic connectors. Fiber optic **butt-jointed connectors** align and bring the prepared ends of two fibers into close contact. The end-faces of some butt-jointed connectors touch, but others do not depending upon the connector design. Types of butt-jointed connectors include cylindrical ferrule and biconical connectors. Fiber optic **expanded-beam connectors** use two lenses to first expand and then refocus the light from the transmitting fiber into the receiving fiber. Single fiber butt-jointed and expanded beam connectors normally consist of two plugs and an adapter (coupling device). Figure 4-15 shows how to configure each plug and adapter when making the connection between two optical fibers.

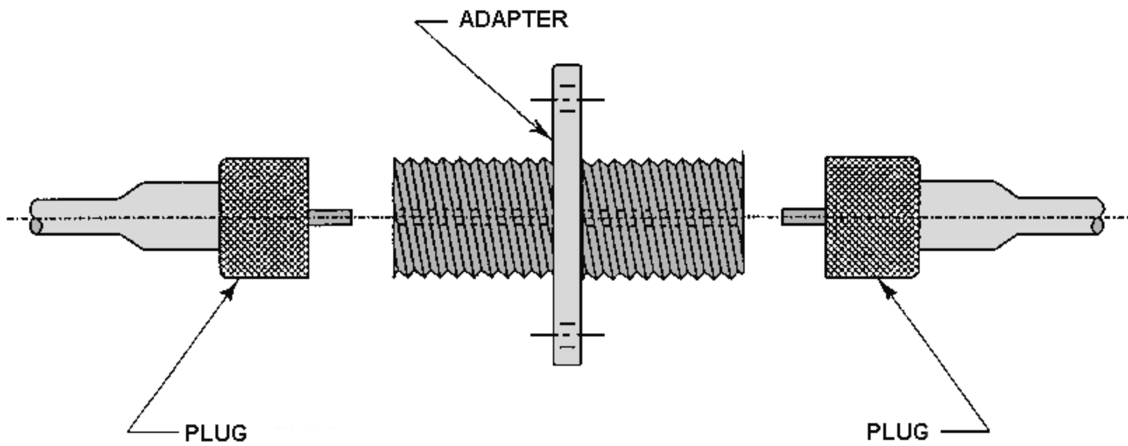


Figure 4-15.—Plug-adapter-plug configuration.

Ferrule connectors use two cylindrical plugs (referred to as ferrules), an alignment sleeve, and sometimes axial springs to perform fiber alignment. Figure 4-16 provides an illustration of this basic ferrule connector design. Precision holes drilled or molded through the center of each ferrule allow for fiber insertion and alignment. Precise fiber alignment depends on the accuracy of the central hole of each ferrule. When the fiber ends are inserted, an adhesive (normally an epoxy resin) bonds the fiber inside the ferrule. The fiber-end faces are polished until they are flush with the end of the ferrule to achieve a low-loss fiber connection. Fiber alignment occurs when the ferrules are inserted into the alignment sleeve. The inside diameter of the alignment sleeve aligns the ferrules, which in turn align the fibers. Ferrule connectors lock the ferrules in the alignment sleeve using a threaded outer shell or some other type of coupling mechanism.

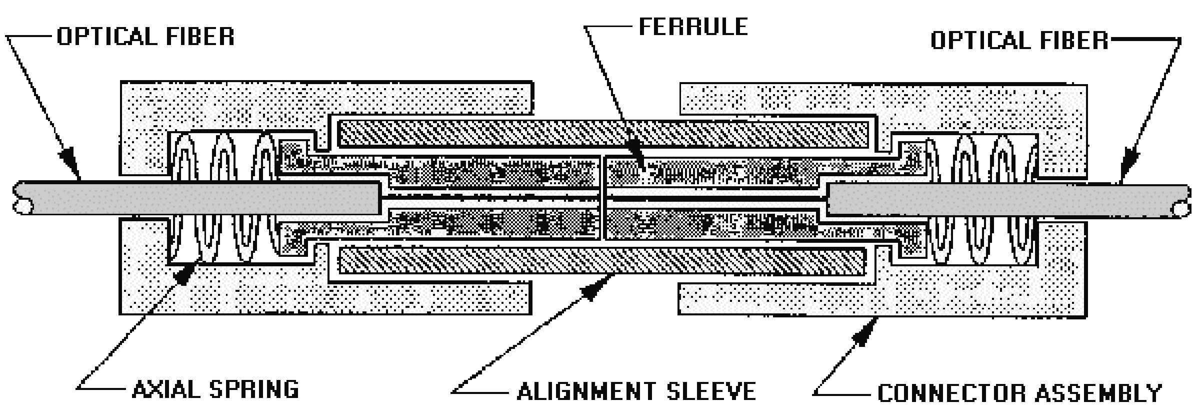


Figure 4-16.—Basic ferrule connector design.

As stated before, fiber alignment depends on an accurate hole through the center of the ferrule. Normally, ferrule connectors use ceramic or metal ferrules. The center hole is generally drilled in a metal ferrule. Drilling an accurate hole through the entire metal ferrule can be difficult. To improve fiber alignment, some metal ferrule connectors use precision watch-jeweled centering. In precision watch-jeweled centering, a watch jewel with a precision centered hole is placed in the tip of the ferrule. The central hole of the watch jewel centers the fiber with respect to the axis of the cylindrical ferrule. The watch jewel provides for better fiber alignment, because regulating the hole tolerance of the watch jewel is easier than maintaining a precise hole diameter when drilling through an entire ferrule.

The center hole in a ceramic ferrule is created by forming the ferrule around a precision wire, which is then removed. This method produces holes accurately centered in the ferrule. Most cylindrical ferrule connectors now use ceramic ferrules. The Straight Tip (ST® connector is an example of a ceramic ferrule connector. (ST is a registered trademark of AT&T.)

Other cylindrical ferrule connectors have a ferrule that contains both metal and ceramic. For these connectors a ceramic capillary is placed within the tip of a metal ferrule to provide for precision fiber alignment. The ceramic capillary is a ceramic tube with a small inner diameter that is just larger than the diameter of the fiber. Figure 4-17 shows the placement of the ceramic capillary within the metal ferrule.

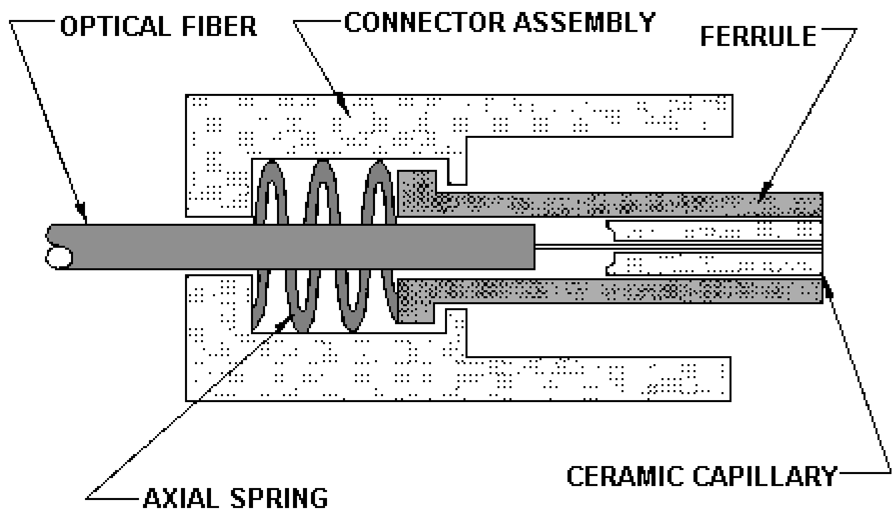


Figure 4-17.—A ceramic capillary set within a metal ferrule.

Another type of butt-jointed connector is the biconical connector. Biconical connectors use two conical plugs, a double conical alignment sleeve, and axial springs to perform fiber alignment. Figure 4-18 is an illustration of this basic biconical connector design. Formation of the plugs and alignment sleeve involves transfer molding. Transfer molding uses silica-filled epoxy resin to mold the conical plug directly to the fiber or around a cast (precision wire). After connecting the conical plugs to the optical fibers, the fiber-end faces are polished before the plugs are inserted into the molded alignment sleeve. During fiber insertion, the inside surface of the double conical sleeve performs fiber alignment, while the axial springs push the fiber ends into close contact. If the alignment sleeve permits the fibers to actually become in contact, then the axial spring provides enough force to maintain fiber contact but prevent damage to the fiber-end faces. Normally, biconical connectors lock the fibers in alignment using a threaded outer shell.

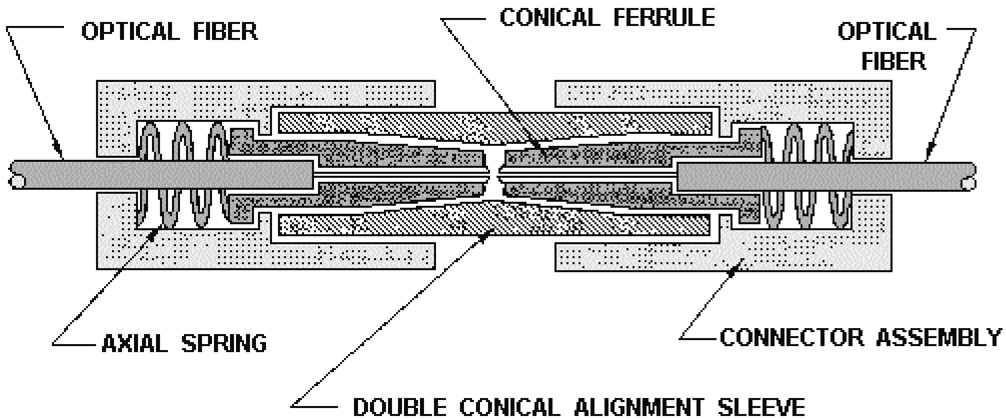


Figure 4-18.—Biconical connector design.

Multifiber connectors join and align multifiber cables to reduce the time it takes to connect multiple fibers. One type of multifiber connector is the array connector. The array connector is used to connect individual ribbons of ribbon-type cables. The array connector is similar to the ribbon splice. In the array connector, the fibers of each ribbon are epoxied into grooves of a silicon chip so that the fiber ends protrude from the end of the chip. The chip and the protruding fibers are polished flat for connection. Each half of the connector is prepared separately before being butt-jointed. A spring clip and two grooved metal-backed plates are used to align and connect the stacked ribbons of the two ribbon cables. Array connectors may also use an alignment sleeve with V-grooved silicon chips and metal springs to align and connect stacked ribbons. Figure 4-19 shows the spring clip method of array connector alignment. The multifiber array connector is only one example of a multiple connector. Many types of multiple connectors exist that connect different types of multifiber cables.

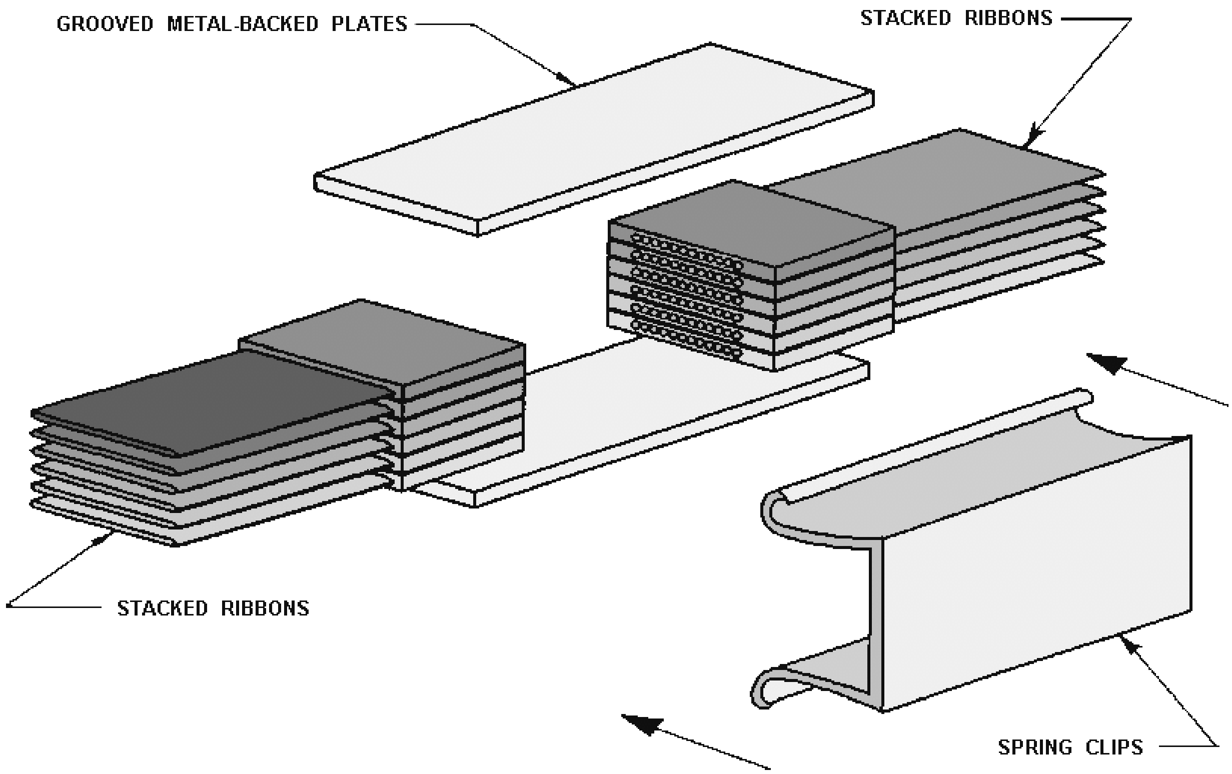


Figure 4-19.—Spring clip method of ribbon connection.

Figure 4-20 shows how an expanded-beam connector uses two lenses to expand and then refocus the light from the transmitting fiber into the receiving fiber. Expanded-beam connectors are normally plug-adapter-plug type connections. Fiber separation and lateral misalignment are less critical in expanded-beam coupling than in butt-jointing. The same amount of fiber separation and lateral misalignment in expanded beam coupling produces a lower coupling loss than in butt-jointing. However, angular misalignment is more critical. The same amount of angular misalignment in expanded-beam coupling produces a higher loss than in butt-jointing. Expanded-beam connectors are also much harder to produce. Present applications for expanded-beam connectors include multifiber connections, edge connections for printed circuit boards, and other applications.

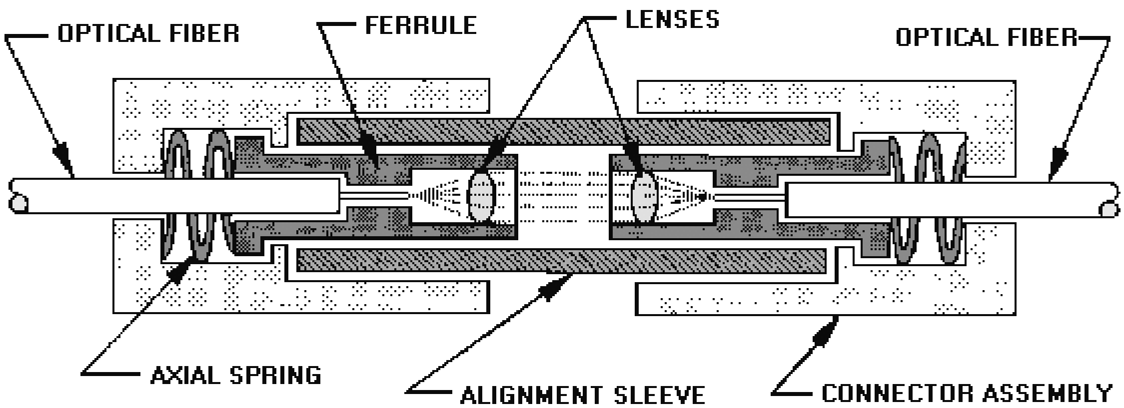


Figure 4-20.—Expanded-beam connector operation.

Q24. What connection properties result in fiber optic connector coupling loss?

Q25. Which is the more critical parameter in maintaining total insertion loss below the required level, fiber alignment or fiber mismatch?

Q26. Fiber optic connectors can reduce system performance by increasing what two types of noise?

Q27. Which type of fiber optic connector (butt-jointed or expanded beam) brings the prepared ends of two optical fibers into close contact?

Q28. Is coupling loss from fiber separation and lateral misalignment more critical in expanded-beam or butt-jointed connectors?

Q29. Is coupling loss from angular misalignment more critical in expanded beam or butt-jointed connectors?

MILITARY CONNECTORS

Light-duty connectors and **heavy-duty connectors** are two ways that the Navy classifies fiber optic connectors. Light-duty connector shipboard applications include locations that protect the connectors from the environment, such as in a junction box or equipment enclosure. Heavy-duty applications require a very rugged, stand-alone, sealed connector. A heavy-duty connector must also withstand pulls and tugs on the fiber cable without disrupting system operation. Light-duty connectors can be of the ferrule, biconical, or expanded-beam designs. Ferrule-type ST® connectors are becoming the commercial connector of choice for local area network (LAN) and data transfer links and are the standard connector for Navy light duty applications. This connector is described in specification sheets 16, 17, and 18 of MIL-C-83522. Figure 4-21 shows the ST type of light-duty connector.

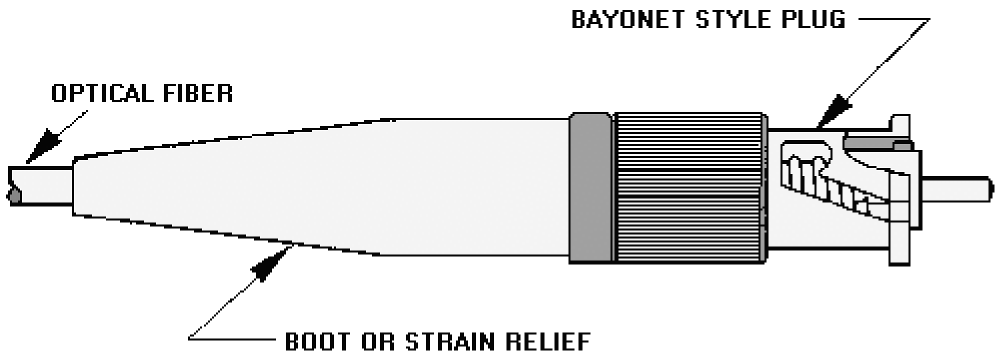


Figure 4-21.—ST light-duty connector.

Figure 4-22 shows one type of heavy-duty connector designed for use in harsh Navy environments. This connector is described by the military specification MIL-C-28876. This connector comes in various sizes capable of terminating 2, 4, 6, or 8 fibers. Each fiber termination, called a terminus, is of the cylindrical ferrule type. Two slightly different termini are used to form a connection; a pin terminus and a socket terminus. The pin terminus consists of a terminus body, which holds the terminus within the connector shell and a ceramic ferrule. The socket terminus consists of a terminus body, a ceramic ferrule, and an alignment sleeve, which attaches to the ceramic ferrule. Fiber alignment occurs when the pin terminus slides into the alignment sleeve of the socket terminus. The termini are held within an insert in the connector shell. When the connector halves are mated, the connector inserts align the mating termini, which then align the mating fibers. The connector shell and backshell protect the termini from the surrounding environment and provide strain relief for the multifiber cable.

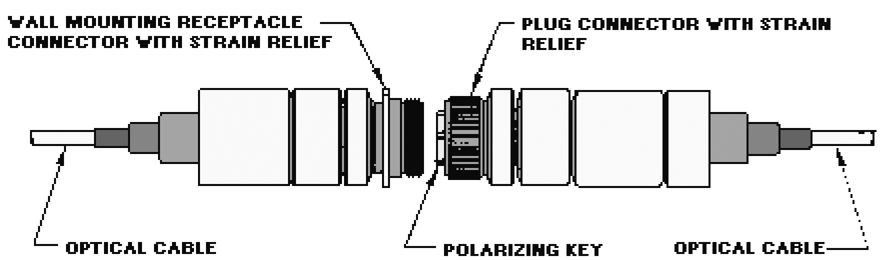


Figure 4-22.—MIL-C-28876 heavy-duty connector.

Q30. The Navy classifies fiber optic connectors in what two ways?

FIBER OPTIC COUPLERS

Some fiber optic data links require more than simple point-to-point connections. These data links may be of a much more complex design that requires multi-port or other types of connections. Figure 4-23 shows some example system architectures that use more complex link designs. In many cases these types of systems require fiber optic components that can redistribute (combine or split) optical signals throughout the system.

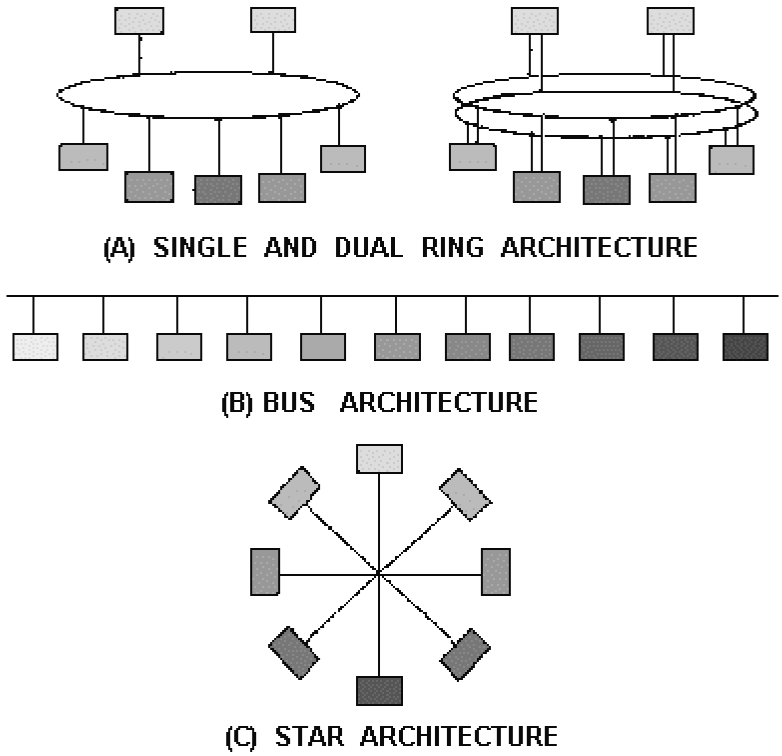


Figure 4-23.—Examples of complex system architectures.

One type of fiber optic component that allows for the redistribution of optical signals is a fiber optic coupler. A fiber optic coupler is a device that can distribute the optical signal (power) from one fiber among two or more fibers. A fiber optic coupler can also combine the optical signal from two or more fibers into a single fiber. Fiber optic couplers attenuate the signal much more than a connector or splice because the input signal is divided among the output ports. For example, with a 1×2 fiber optic coupler, each output is less than one-half the power of the input signal (over a 3 dB loss).

Fiber optic couplers can be either active or passive devices. The difference between active and passive couplers is that a **passive coupler** redistributes the optical signal without optical-to-electrical conversion. Active couplers are electronic devices that split or combine the signal electrically and use fiber optic detectors and sources for input and output.

Figure 4-24 illustrates the design of a basic fiber optic coupler. A basic fiber optic coupler has N input ports and M output ports. N and M typically range from 1 to 64. The number of input ports and output ports vary depending on the intended application for the coupler. Types of fiber optic couplers include optical splitters, optical combiners, X couplers, star couplers, and tree couplers.

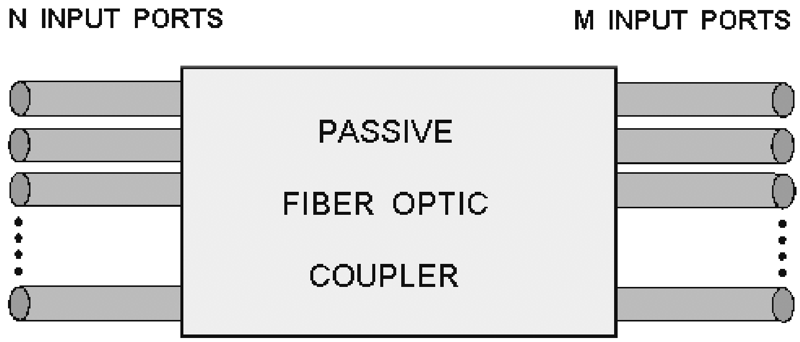


Figure 4-24.—Basic passive fiber optic coupler design.

An **optical splitter** is a passive device that splits the optical power carried by a single input fiber into two output fibers. Figure 4-25 illustrates the transfer of optical power in an optical splitter. The input optical power is normally split evenly between the two output fibers. This type of optical splitter is known as a **Y-coupler**. However, an optical splitter may distribute the optical power carried by input power in an uneven manner. An optical splitter may split most of the power from the input fiber to one of the output fibers. Only a small amount of the power is coupled into the secondary output fiber. This type of optical splitter is known as a T-coupler, or an optical tap.

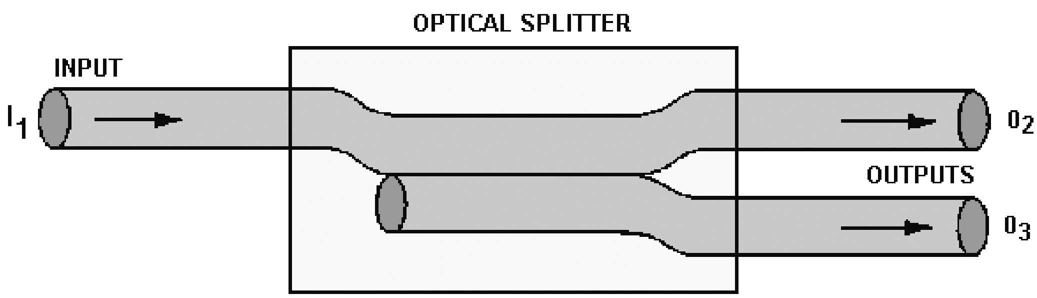


Figure 4-25.—Optical splitter.

An **optical combiner** is a passive device that combines the optical power carried by two input fibers into a single output fiber. Figure 4-26 illustrates the transfer of optical power in an optical combiner.

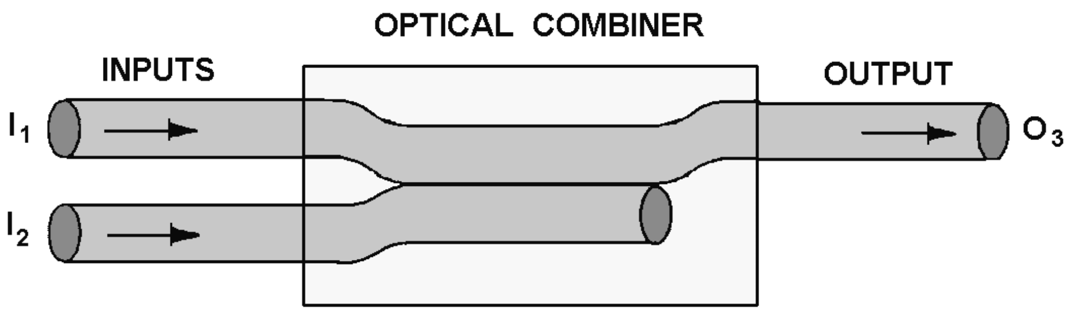


Figure 4-26.—Optical combiner.

An **X coupler** combines the functions of the optical splitter and combiner. The X coupler combines and divides the optical power from the two input fibers between the two output fibers. Another name for the X coupler is the 2×2 coupler.

Star and **tree couplers** are multiport couplers that have more than two input or two output ports. A **star coupler** is a passive device that distributes optical power from more than two input ports among several output ports. Figure 4-27 shows the multiple input and output ports of a star coupler. A **tree coupler** is a passive device that splits the optical power from one input fiber to more than two output fibers. A tree coupler may also be used to combine the optical power from more than two input fibers into a single output fiber. Figure 4-28 illustrates each type of tree coupler. Star and tree couplers distribute the input power uniformly among the output fibers.

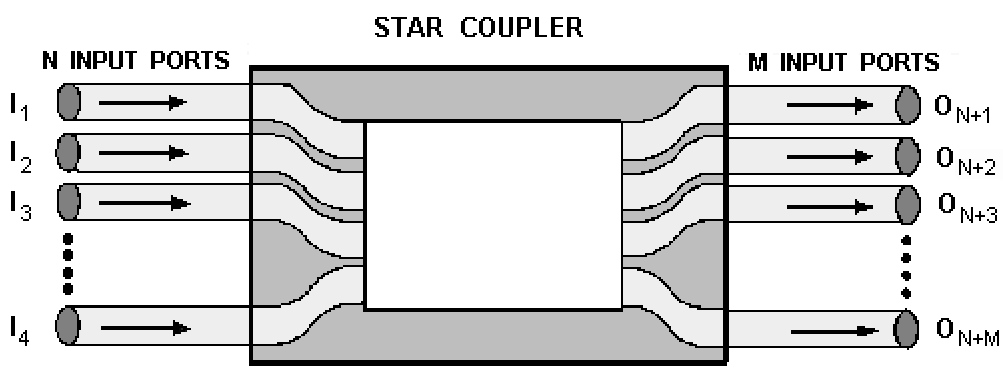


Figure 4-27.—Star coupler.

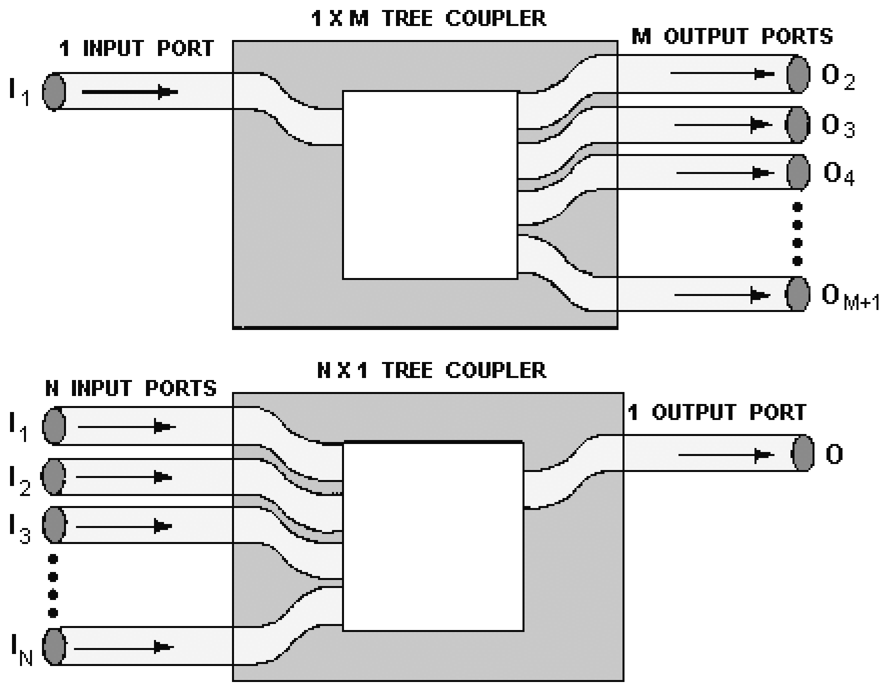


Figure 4-28.—(1 × M) and (N × 1) tree coupler designs.

Fiber optic couplers should prevent the transfer of optical power from one input fiber to another input fiber. **Directional couplers** are fiber optic couplers that prevent this transfer of power between input fibers. Many fiber optic couplers are also symmetrical. A **symmetrical coupler** transmits the same amount of power through the coupler when the input and output fibers are reversed.

Passive fiber optic coupler fabrication techniques can be complex and difficult to understand. Some fiber optic coupler fabrication involves beam splitting using microlenses or graded-refractive-index (GRIN) rods and beam splitters or optical mixers. These beamsplitter devices divide the optical beam into two or more separated beams. Fabrication of fiber optic couplers may also involve twisting, fusing, and tapering together two or more optical fibers. This type of fiber optic coupler is a fused biconical taper coupler. Fused biconical taper couplers use the radiative coupling of light from the input fiber to the output fibers in the tapered region to accomplish beam splitting. Figure 4-29 illustrates the fabrication process of a fused biconical taper coupler.

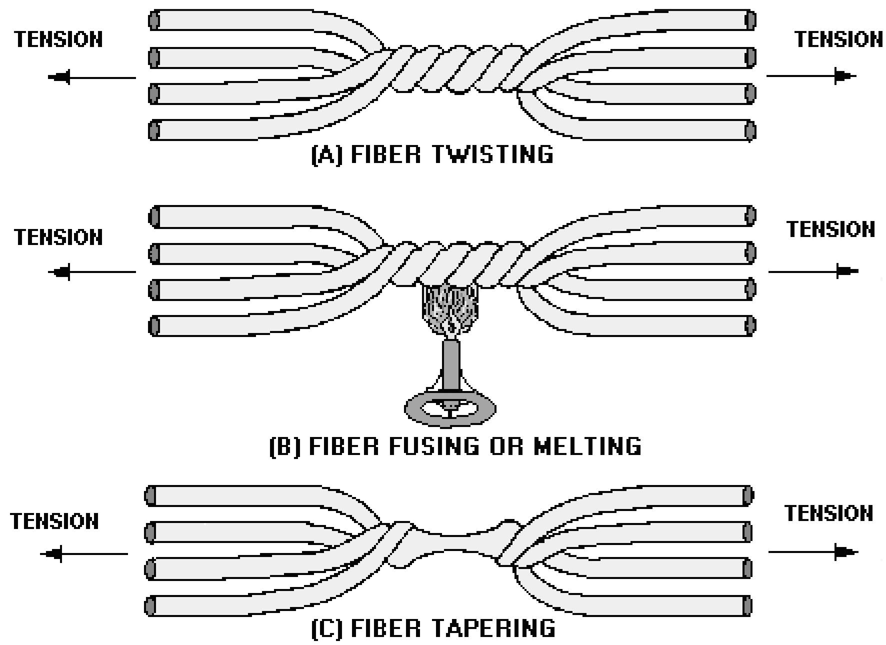


Figure 4-29.—Fabrication of a fused biconical taper coupler (star coupler).

Q31. What is the difference between passive and active fiber optic couplers?

Q32. Which type of optical splitter (Y-coupler or T-coupler) splits only a small amount of power from the input fiber to one of the output fibers?

Q33. Describe a directional coupler.

SUMMARY

Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas you have learned. You should have a thorough understanding of these principles before moving on to chapter 5.

FIBER OPTIC CONNECTIONS transfer optical power from one component to another. Fiber optic connections also permit fiber optic systems to be more than just a point-to-point data link.

A **FIBER OPTIC SPLICE** is a permanent joint between two fibers or two groups of fibers.

FIBER OPTIC CONNECTORS permit easy coupling and uncoupling of optical fibers.

FIBER OPTIC COUPLERS distribute or combine optical signals between fibers.

POOR FIBER END PREPARATION and **POOR FIBER ALIGNMENT** are the main causes of coupling loss.

RADIANCE is the amount of optical power emitted by a unit area of emitting surface per unit time in a specified direction. An optical source's radiance, or brightness, is a measure of its optical power launching capability.

FIBER-TO-FIBER COUPLING LOSS is affected by intrinsic and extrinsic coupling losses. **INTRINSIC COUPLING LOSSES** are caused by inherent fiber characteristics. **EXTRINSIC COUPLING LOSSES** are caused by jointing techniques.

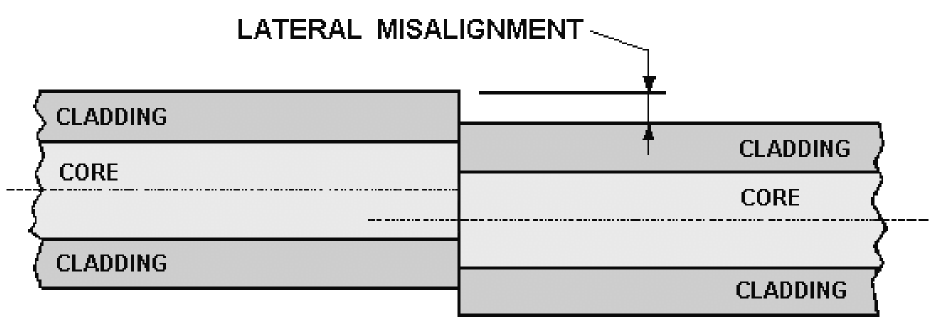
A **FIBER PIGTAIL** is a short length of optical fiber (usually 1 meter or less) permanently fixed to a fiber optic component, such as an optical source or detector.

FRESNEL REFLECTION occurs twice in a fiber-to-fiber connection. A portion of the optical power is reflected when the light first exits the source fiber. Light is then reflected as the optical signal enters the receiving fiber.

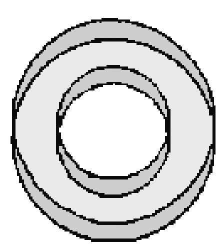
INDEX MATCHING GEL eliminates or reduces the step change in the refractive index at the fiber interface, reducing Fresnel reflection.

POOR FIBER ALIGNMENT is a main source of coupling loss in fiber-to-fiber connections. The three basic coupling errors that occur during fiber alignment are fiber separation (longitudinal misalignment), lateral misalignment, and angular misalignment.

In **FIBER SEPARATION** a small gap remains between fiber-end faces after completing the fiber connection. **LATERAL**, or **AXIAL**, **MISALIGNMENT** is when the axes of the two fibers are offset in a perpendicular direction. **ANGULAR MISALIGNMENT** is when the axes of the two fibers are no longer parallel.



(A) SIDE VIEW



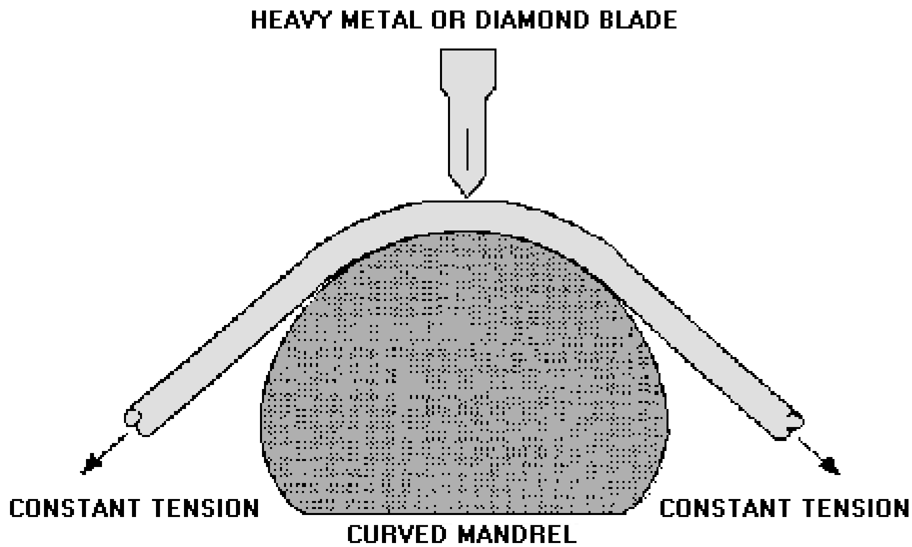
(B) END VIEW

SINGLE MODE FIBERS are more sensitive to alignment errors than multimode fibers because of their small core diameters and low numerical apertures.

The **MODE POWER DISTRIBUTION (MPD)** is the distribution of radiant power among the various modes propagating along the optical fiber.

Poor **FIBER END PREPARATION** is another source of extrinsic coupling loss. An optical fiber end face must be flat, smooth, and perpendicular to the fiber's axis to ensure proper fiber connection.

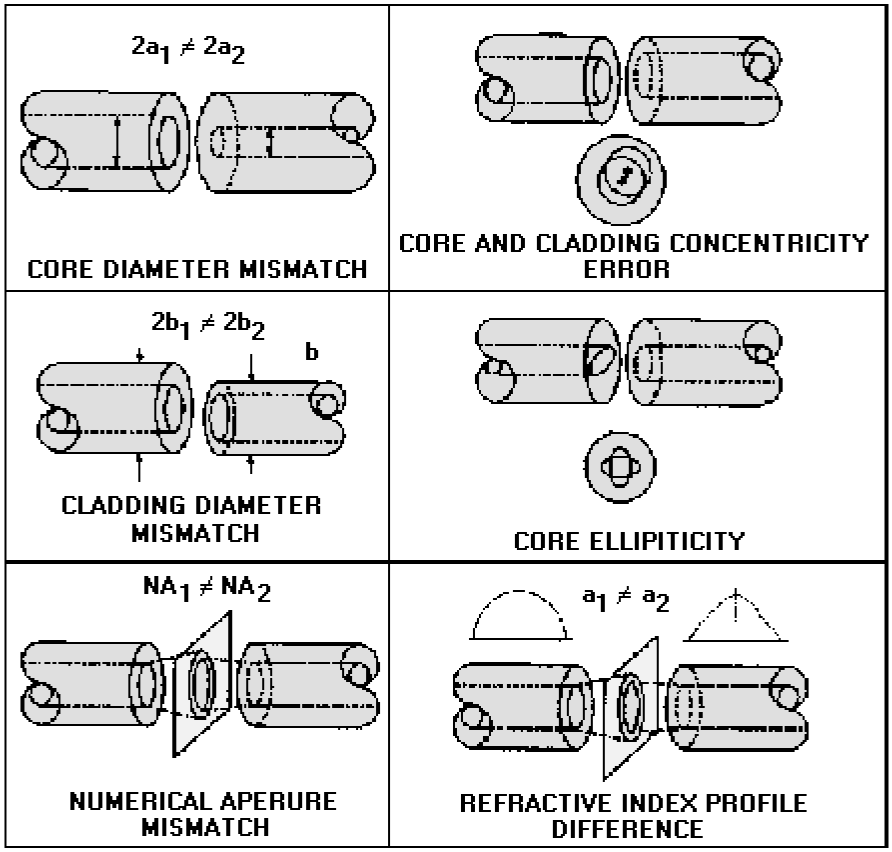
The **SCORE-AND-BREAK** method is the basic fiber cleaving technique for preparing optical fibers for coupling.



POLISHING the fiber ends removes most surface imperfections introduced by the fiber cleaving or cutting process. Fiber polishing involves a step-down approach. The first step is to give the surface of the fiber end a rough polish. The next step involves giving the surface of the fiber end a fine polish.

FIBER MISMATCHES are a source of intrinsic coupling loss. Types of fiber mismatches include fiber geometry mismatches, NA mismatch, and refractive index profile difference.

FIBER GEOMETRY MISMATCHES include core diameter, cladding diameter, core ellipticity, and core-cladding concentricity differences.



CORE DIAMETER MISMATCH causes coupling loss only if the launching fiber has a larger core radius than the receiving fiber.

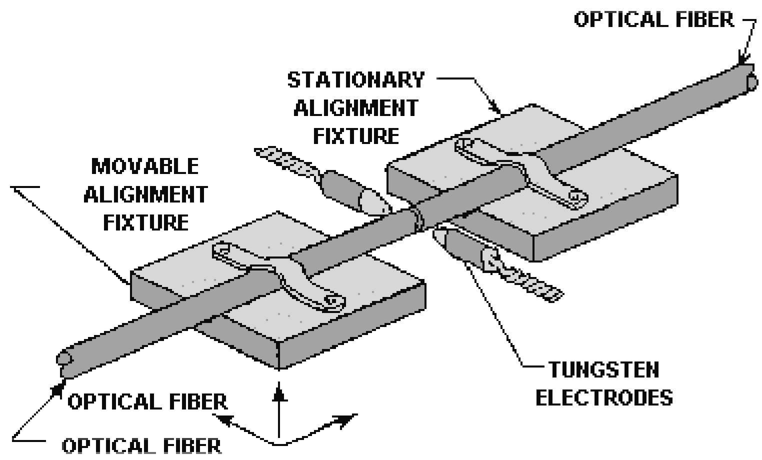
NA MISMATCH causes coupling loss only if the launching fiber has a higher NA than the receiving fiber.

A **REFRACTIVE INDEX PROFILE DIFFERENCE** causes coupling loss only if the launching fiber has a larger profile parameter than the receiving fiber.

MECHANICAL and **FUSION SPLICING** are two broad categories that describe the techniques used for fiber splicing. A mechanical splice is a fiber splice where mechanical fixtures perform fiber alignment and connection. A fusion splice is a fiber splice where localized heat fuses or melts the ends of two lengths of optical fiber together.

In **MECHANICAL SPLICING**, mechanical fixtures hold the two optical fibers in alignment for an indefinite period of time without movement. The amount of splice loss is stable over time and unaffected by changes in environmental or mechanical conditions.

ARC FUSION involves the discharge of electric current across a gap between two electrodes. By placing the fiber end between the electrodes, the electric discharge melts or fuses the ends of the fibers.



PREFUSION involves a short discharge of electric current across the gap between the electrodes. In prefusion the fiber ends are cleaned and rounded to eliminate any surface defects that remain from fiber cleaving.

A **FIBER OPTIC CONNECTOR** is a demateable device that permits the coupling of optical power between two optical fibers or two groups of fibers.

FIBER ALIGNMENT in a fiber optic connector is the critical parameter in maintaining total insertion loss below the required level.

FIBER OPTIC CONNECTORS can affect system performance by increasing modal and reflection noise.

MODAL NOISE is eliminated by using only single mode fiber with laser sources and only low-coherence sources such as light-emitting diodes with multimode fiber.

REFLECTION NOISE is reduced by index matching gels, physical contact polishes, or antireflection coatings.

BUTT-JOINTED and **EXPANDED BEAM CONNECTORS** are two ways to classify fiber optic connectors. Butt-jointed connectors bring the prepared ends of two fibers into close contact. Expanded beam connectors use two lenses to first expand and then refocus the light from the transmitting fiber into the receiving fiber.

LIGHT-DUTY and **HEAVY-DUTY CONNECTORS** are two ways that the Navy classifies fiber optic connectors. Light-duty connector shipboard applications include locations that protect the connectors from the environment such as in a junction box. Heavy-duty applications require a very rugged, stand-alone, sealed connector.

A **PASSIVE COUPLER** redistributes an optical signal without optical to electrical conversion.

An **OPTICAL SPLITTER** is a passive device that splits the optical power carried by a single input fiber into two output fibers.

An **OPTICAL COMBINER** is a passive device that combines the optical power from two input fibers into a single output fiber.

A **STAR COUPLER** is a passive device that distributes optical power from more than two input ports among several output ports.

A **TREE COUPLER** is a passive device that splits the optical power from one input fiber to more than two output fibers. A tree coupler may also be used to combine the optical power from more than two input fibers into a single output fiber.

DIRECTIONAL COUPLERS are fiber optic couplers that prevent the transfer of optical power from one input fiber to another input fiber.

A **SYMMETRICAL COUPLER** transmits the same amount of power through the coupler when the input and output fibers are reversed.

ANSWERS TO QUESTIONS Q1. THROUGH Q33.

A1. Splice.

A2. Poor fiber end preparation and poor fiber alignment.

A3.

$$\text{loss} = 10\log_{10}\frac{P_i}{P_o}$$

A4. Yes.

A5. A step change in refractive index that occurs at fiber joints, caused by fiber separation.

A6. Index matching gel.

A7. Fiber separation (longitudinal misalignment), lateral misalignment, and angular misalignment.

A8. Angular misalignment.

A9. (a) Reduces coupling loss, (b) does not change coupling loss, and (c) increases coupling loss.

A10. Single mode.

A11. Be flat, smooth, and perpendicular to the fiber axis.

A12. Score-and-break.

A13. Flat.

A14. Core diameter mismatch, cladding diameter mismatch, core ellipticity, core and cladding concentricity differences, NA mismatch, and refractive index profile differences.

A15. No.

A16. A permanent fiber joint whose purpose is to establish an optical connection between two individual optical fibers.

A17. Mechanical and fusion splicing.

A18. An epoxy resin that seals mechanical splices and provides index matching between the connected fibers.

A19. It is a low-loss mechanical splice that provides stable environmental and mechanical performance in the Navy environment, and it requires only a small amount of training.

A20. The angles and quality of the two fiber-end faces.

A21. The small size of the fusion splice and the development of automated fusion-splicing machines.

A22. Prefusion.

A23. Single mode fibers.

A24. Poor fiber alignment and end preparation, fiber mismatches, and Fresnel reflection.

A25. Fiber alignment.

A26. Modal and reflection.

A27. Butt-jointed connectors.

A28. Butt-jointed connectors.

A29. Expanded beam connectors.

A30. Light-duty and heavy-duty connectors.

A31. Passive couplers redistribute optical signals without optical-to-electrical conversion.

A32. T-coupler.

A33. A fiber optic coupler that prevents the transfer of power between input fibers.

CHAPTER 5

FIBER OPTIC MEASUREMENT TECHNIQUES

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to do the following:

1. Identify the prime reasons for conducting fiber optic manufacturing laboratory and field measurements.

2. Describe the optical fiber and optical connection laboratory measurements performed by the Navy to evaluate fiber optic component and system performance.

3. Describe the near-field and far-field optical power distribution of an optical fiber.

4. Describe optical fiber launch conditions and modal effects that affect optical fiber and optical connection measurements.

5. Understand the term optical time-domain reflectometry and the interpretation of an optical time-domain reflectometer (OTDR) trace.

6. Describe the procedure for locating a fiber fault using an OTDR.

FIBER OPTIC MEASUREMENTS

Fiber optic data links operate reliably if fiber optic component manufacturers and end users perform the necessary laboratory and field measurements. Manufacturers must test how component designs, material properties, and fabrication techniques affect the performance of fiber optic components. These tests can be categorized as design tests or quality control tests. Design tests are conducted during the development of a component. Design tests characterize the component's performance (optical, mechanical, and environmental) in the intended application. Once the component performance is characterized, the manufacturer generally only conducts quality control tests. Quality control tests verify that the parts produced are the same as the parts the design tests were conducted on. When manufacturers ship fiber optic components, they provide quality control data detailing the results of measurements performed during or after component fabrication.

End users (equipment manufacturers, shipbuilders, maintenance personnel, test personnel, and so on) should measure some of these parameters upon receipt before installing the component into the fiber optic data link. These tests determine if the component has been damaged in the shipping process. In addition, end users should measure some component parameters after installing or repairing fiber optic components in the field. The values obtained can be compared to the system installation specifications. These measurements determine if the installation or repair process has degraded component performance and will affect data link operation.

Whenever a measurement is made, it should be made using a standard measurement procedure. For most fiber optic measurements, these standard procedures are documented by the Electronics Industries Association/Telecommunications Industries Association (EIA/TIA). Each component measurement procedure is assigned a unique number given by EIA/TIA-455-X. The X is a sequential number assigned to that particular component test procedure. System level test procedures are assigned unique numbers given by EIA/TIA-526-X. Again the X is a sequential number assigned to that particular system test procedure.

LABORATORY MEASUREMENTS

Providing a complete description of every laboratory measurement performed by manufacturers and end users is impossible. This chapter only provides descriptions of optical fiber and optical connection measurements that are important to system operation. The list of optical fiber and optical connection laboratory measurements described in this chapter includes the following:

- Attenuation
- Cutoff wavelength (single mode)
- Bandwidth (multimode)
- Chromatic dispersion
- Fiber geometry
- Core diameter
- Numerical aperture (multimode)
- Mode field diameter (single mode)
- Insertion loss
- Return loss and reflectance

End users routinely perform optical fiber measurements to measure fiber power loss and fiber information capacity. End users may also perform optical fiber measurements to measure fiber geometrical properties. Optical fiber power loss measurements include attenuation and cutoff wavelength. Optical fiber information capacity measurements include chromatic dispersion and bandwidth. Fiber geometrical measurements include cladding diameter, core diameter, numerical aperture, and mode field diameter. Optical connection measurements performed by end users in the laboratory include insertion loss and reflectance or return loss.

Q1. List the fiber geometrical measurements performed in the laboratory.

Attenuation

Attenuation is the loss of optical power as light travels along the fiber. It is a result of absorption, scattering, bending, and other loss mechanisms as described in chapter 3. Each loss mechanism contributes to the total amount of fiber attenuation.

End users measure the total attenuation of a fiber at the operating wavelength (λ). The **total attenuation (A)** between an arbitrary point X and point Y located on the fiber is

$$A = 10\log\frac{P_x}{P_y}\,dB$$

P_x is the power output at point X. P_y is the power output at point Y. Point X is assumed to be closer to the optical source than point Y. The total amount of attenuation will vary with changes in wavelength λ.

The **attenuation coefficient (α)** or attenuation rate, is

$$a = \frac{A}{L}\,dB/km$$

L is the distance between points X and Y. α is a positive number because P_x is always larger than P_y. The attenuation coefficient will also vary with changes in λ.

CUTBACK METHOD.—In laboratory situations, end users perform the cutback method for measuring the total attenuation of an optical fiber. The cutback method involves comparing the optical power transmitted through a long piece of test fiber to the power present at the beginning of the fiber.

The cutback method for measuring multimode fiber attenuation is EIA/TIA-455-46. The cutback method for measuring single mode fiber attenuation is EIA/TIA-455-78. The basic measurement process is the same for both of these procedures. The test method requires that the test fiber of known length (L) be cut back to an approximate 2-m length. This cut back causes the destruction of 2-m of fiber. This method requires access to both fiber ends. Each fiber end should be properly prepared to make measurements. EIA/TIA-455-57 describes how to properly prepare fiber ends for measurement purposes.

Figure 5-1 illustrates the cutback method for measuring fiber attenuation. The cutback method begins by measuring, with an optical power meter, the output power P_1 of the test fiber of known length (L) (figure 5-1, view A). Without disturbing the input light conditions, the test fiber is cut back to an approximate 2-m length. The output power P_2 of the shortened test fiber is then measured (figure 5-1, view B). The fiber attenuation A_T and the attenuation coefficient α are then calculated.

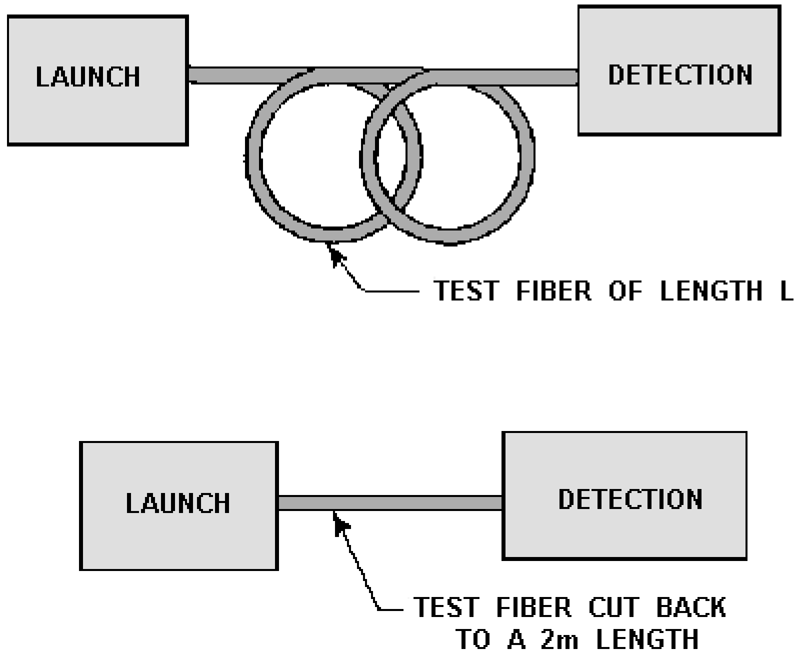


Figure 5-1.—Cutback method for measuring fiber attenuation: A. Test measurement; B. Cut-back measurement.

LAUNCH CONDITIONS.—Measurement personnel must pay attention to how optical power is launched into the fiber when measuring fiber attenuation. Different distributions of launch power (launch conditions) can result in different attenuation measurements. This is more of a problem with multimode fiber than single mode fiber. For single mode fiber, optical power must be launched only into the fundamental mode. This is accomplished using a mode filter on the fiber. For multimode fiber, the distribution of power among the modes of the fiber must be controlled. This is accomplished by controlling the launch spot size and angular distribution.

The **launch spot size** is the area of the fiber face illuminated by the light beam from the optical source. The diameter of the spot depends on the size of the optical source and the properties of the optical elements (lenses, and so on) between the source and the fiber end face. The **angular distribution** is the angular extent of the light beam from the optical source incident on the fiber end face. The launch angular distribution also depends on the size of the optical source and the properties of the optical elements between the optical source and the fiber end face.

Multimode optical fiber launch conditions are typically characterized as being underfilled or overfilled. An underfilled launch concentrates most of the optical power in the center of the fiber. An **underfilled** launch results when the launch spot size and angular distribution are smaller than that of the fiber core. Underfilling the fiber excites mainly low-order modes. Figure 5-2 illustrates an underfilled launch condition.

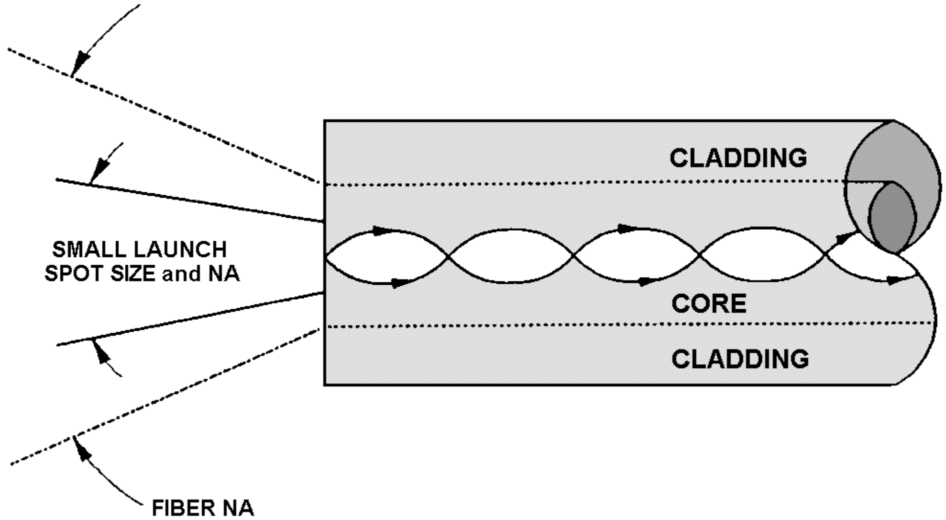


Figure 5-2.—Underfilled launch condition.

Overfilling the fiber excites both low-order and high-order modes. Figure 5-3 illustrates an overfilled launch condition. An **overfilled** launch condition occurs when the launch spot size and angular distribution are larger than that of the fiber core. Incident light that falls outside the fiber core is lost. In addition, light that is incident at angles greater than the angle of acceptance of the fiber core is lost.

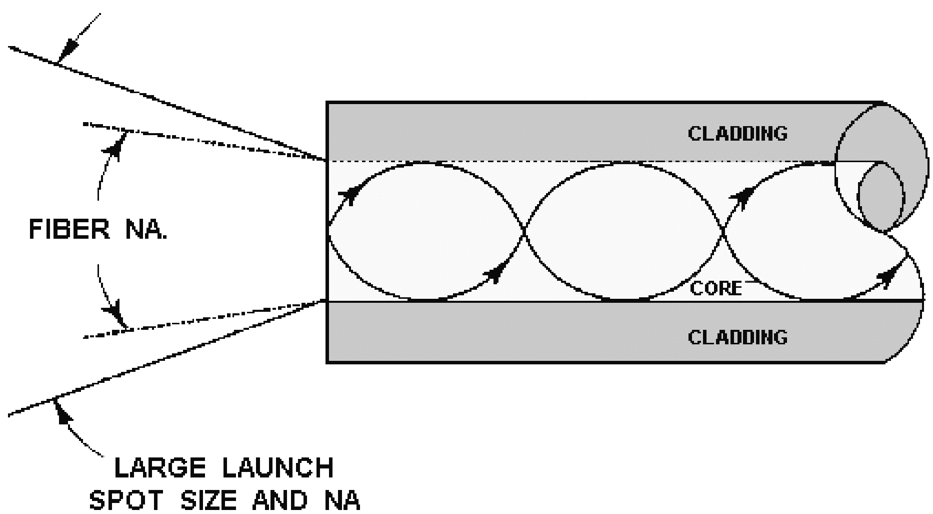


Figure 5-3.—Overfilled launch condition.

In attenuation measurements, cladding-mode strippers and mode filters eliminate the effects that high-order modes have on attenuation results. A **cladding-mode stripper** is a device that removes any cladding mode power from the fiber. Most cladding-mode strippers consist of a material with a refractive index greater than that of the fiber cladding. For most fibers, the fiber coating acts as an excellent cladding-mode stripper.

A **mode filter** is a device that attenuates specific modes propagating in the core of an optical fiber. Mode filters generally involve wrapping the test fiber around a mandrel. For multimode, tight bends tend to remove high-order modes from the fiber. This type of mode filter is known as a **mandrel wrap mode filter**. For multimode fibers, mode filters remove high-order propagating modes and are individually tailored and adjusted for a specific fiber type.

For single mode fibers, a mode filter is used to eliminate the second-order mode from propagating along the fiber. The propagation of the second-order mode will affect attenuation measurements. Fiber attenuation caused by the second-order mode depends on the operating wavelength, the fiber bend radius and length.

The two most common types of mode filters are free-form loops and mandrel wraps. Figure 5-4 illustrates the free-form loop and mandrel-wrap types of mode filters. Mandrel wraps for multimode fibers consist of several wraps (approximately 4 or 5) around a mandrel. A 20-mm diameter mandrel is typically used for 62.5 μm fiber. Mandrel wraps for single mode fibers consist of a single wrap around a 30-mm diameter mandrel. Another common mode filter for single mode fibers is a 30-mm diameter circular free-form loop. Additional information on multimode and single mode filters (and launch conditions) is available in EIA/TIA-455-50 and EIA/TIA-455-77, respectively.

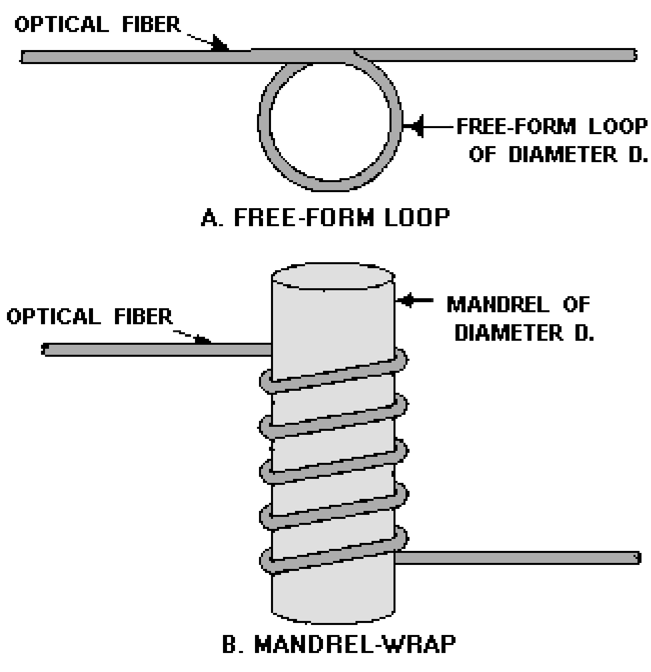


Figure 5-4.—Types of mode filters: A. Free-form loop; B. Mandrel-wrap.

Launch conditions significantly affect the results of multimode fiber attenuation measurements. If the fiber is underfilled, high-order-mode power loss has minimal effect on the measurement results. If too much power is launched into high-order modes, the high-order-mode power loss will dominate the attenuation results. Generally, fiber attenuation measurements are performed using an underfilled launch condition. Power in high-order modes is eliminated by either controlling the input spot size and angular distribution or using mode filters to remove high-order mode power.

Q2. End users measure the total attenuation of a fiber at the operating wavelength (λ). Write the equation for total attenuation (A), between an arbitrary point X and point Y located on an optical fiber.

Q3. Will an optical fiber's attenuation coefficient vary with changes in wavelength?

Q4. What two properties of the launch condition may affect multimode fiber attenuation measurements?

Q5. Does underfilling a multimode optical fiber excite mainly high-order or low-order modes?

Q6. Multimode optical fiber launch conditions are typically characterized as being overfilled or underfilled. Which of these optical launch conditions exists if the launch spot size and angular distribution are larger than that of the fiber core?

Q7. A mode filter is a device that attenuates specific modes propagating in the core of an optical fiber. What mode propagating along single mode fibers do mode filters eliminate?

Q8. What are the two most common types of mode filters?

Cutoff Wavelength

The wavelength at which a mode ceases to propagate is called the cutoff wavelength for that mode. However, an optical fiber is always able to propagate at least one mode, the fundamental mode. The fundamental mode can never be cut off. The **cutoff wavelength** of a single mode fiber is the wavelength above which the fiber propagates only the fundamental mode.

Determining the cutoff wavelength of a single mode fiber involves finding the wavelength above which the power transmitted through the fiber decreased abruptly. This power decrease occurs when the second-order mode propagating in the fiber is cut off. The cutoff wavelength of single mode fibers depends on the fiber length and bend conditions. The effects of length and bending are different on different fibers depending on whether they are matched-clad or depressed-clad in design. The cutoff wavelength of matched-clad fibers is more sensitive to bends than the cutoff wavelength of depressed-clad fibers. The cutoff wavelength of depressed-clad fibers is more sensitive to length than the cutoff wavelength of matched-clad fibers.

Cutoff wavelength may be measured on uncabled or cabled single mode fibers. A slightly different procedure is used in each case, but the basic measurement process is the same. The test method for uncabled single mode fiber cutoff wavelength is EIA/TIA-455-80. The test method for cabled single mode fiber cutoff wavelength is EIA/TIA-455-170. The fiber cutoff wavelength (λ) measured under EIA/TIA-455-80 will generally be higher than the cable cutoff wavelength (λ) measured under EIA/TIA-455-170. The difference is due to the fiber bends introduced during the cable manufacturing process.

Each test method describes the test equipment (input optics, mode filters, and cladding-mode strippers) necessary for the test. Cutoff wavelength measurements require an overfilled launch over the full range of test wavelengths. Since the procedures for measuring the cutoff wavelength of uncabled and cabled single mode fibers are essentially the same, only the test method for measuring the cutoff wavelength of uncabled fiber is discussed.

Measuring the cutoff wavelength involves comparing the transmitted power from the test fiber with that of a reference fiber at different wavelengths. The reference fiber can be the same piece of single mode fiber with small bends introduced or a piece of multimode fiber. If the same fiber with small bends is used as the reference fiber, the technique is called the **bend-reference technique**. If a piece of multimode fiber is used as the reference fiber, the technique is called the **multimode-reference technique**.

For both techniques, the test fiber is loosely supported in a single-turn with a constant radius of 140 mm. Figure 5-5 shows this single-turn configuration. The transmitted signal power $P_s(\lambda)$ is then recorded while scanning the wavelength range in increments of 10 nm or less. The launch and detection conditions are not changed while scanning over the range of wavelengths. The wavelength range scanned encompasses the expected cutoff wavelength.

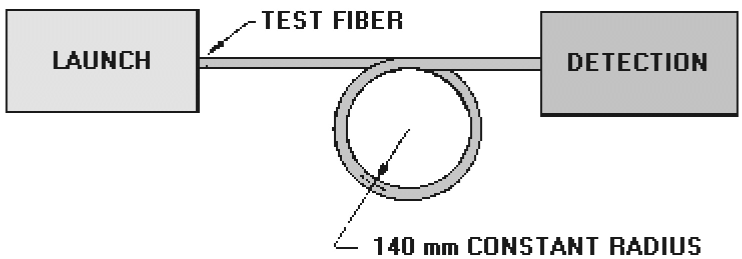


Figure 5-5.—Single-turn configuration for the test fiber.

The reference power measurement is then made. For the bend-reference technique, the launch and detection conditions are not changed, but an additional bend is added to the test fiber. The test fiber is bent to a radius of 30 mm or less to suppress the second-order mode at all the scanned wavelengths. For the multimode-reference technique, the single mode fiber is replaced with a 2-m length of multimode fiber. The transmitted signal power $P_r(\lambda)$ is recorded while scanning the same wavelength range in the same increments of 10 nm or less. The attenuation $A(\lambda)$ at each wavelength is calculated as follows:

$$A(\lambda) = 10 \log \frac{P_s(\lambda)}{P_r(\lambda)} \text{dB}$$

Figure 5-6 shows an example attenuation plot generated using the bend-reference technique. The longest wavelength at which $A(\lambda)$ is equal to 0.1 dB is the fiber cutoff wavelength (λ_{cf}). λ_{cf} is marked on figure 5-6.

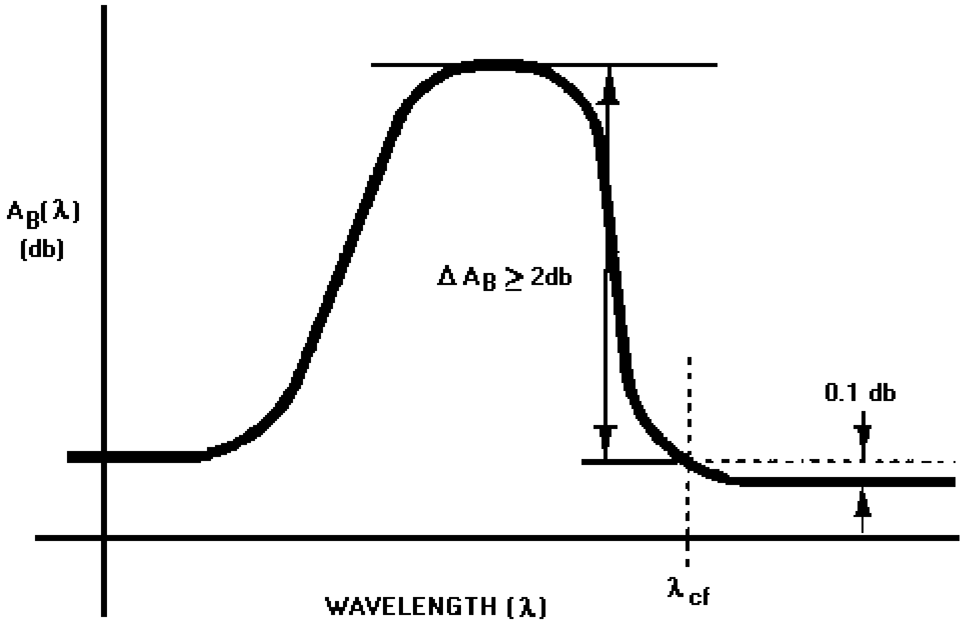


Figure 5-6.—Fiber cutoff wavelength determined by the bend-reference technique.

Figure 5-7 shows an attenuation plot generated using the multimode-reference technique. A straight line is fitted to the long-wavelength portion of A(λ). This straight line is then displaced upward by 0.1 dB. The point at which the straight line intersects the A(λ) plot defines the fiber cutoff wavelength (λ_{cf}).

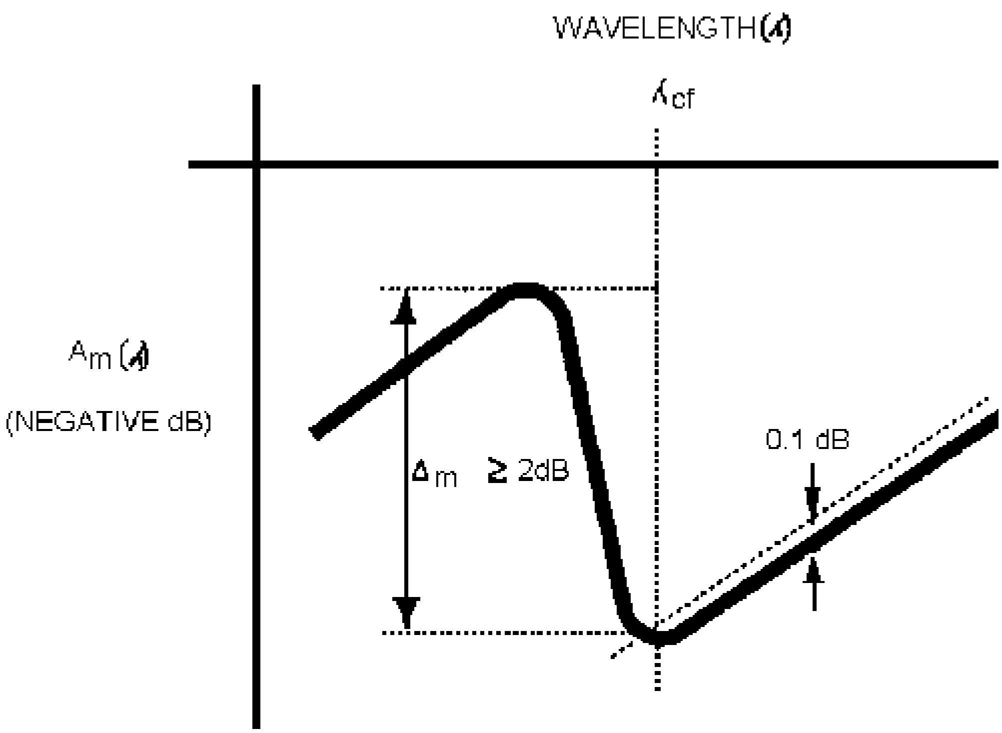


Figure 5-7.—Fiber cutoff wavelength determined by the multimode-reference technique.

Bandwidth

Dispersion reduces the bandwidth, or information-carrying capacity, of an optical fiber. **Dispersion** causes the spreading of the light pulse as it travels along the fiber (see figure 2-20). Fiber dispersion mechanisms include intramodal (chromatic) dispersion and intermodal (modal) dispersion. Multimode fiber bandwidth is a measure of the intermodal dispersion of the multimode fiber.

Intermodal dispersion is maximum when all fiber modes are excited. The source used for intermodal dispersion measurements must overfill the fiber. The optical source must also have a narrow spectral width to reduce the effects of chromatic dispersion in the measurement.

There are two basic techniques for measuring the modal bandwidth of an optical fiber. The first technique characterizes dispersion by measuring the **impulse response h(t)** of the fiber in the time domain. The second technique characterizes modal dispersion by measuring the **baseband frequency response H(*f*)** of the fiber in the frequency domain. H(*f*) is the **power transfer function** of the fiber at the baseband frequency (*f*). H(*f*) is also the Fourier transform of the power impulse response h(t). Only the frequency response method is described here.

The test method for measuring the bandwidth of multimode fibers in the frequency domain is EIA/TIA-455-30. Signals of varying frequencies (*f*) are launched into the test fiber and the power exiting the fiber at the launched fundamental frequency measured. This optical output power is denoted as $P_{out}(f)$. The test fiber is then cut back or replaced with a short length of fiber of the same type. Signals of the same frequency are launched into the cut-back fiber and the power exiting the cut-back fiber at the launched fundamental frequency measured. The optical power exiting the cutback or replacement fiber is denoted as $P_{in}(f)$. The magnitude of the optical fiber frequency response is

$$H(f) = \log_{10}\left[\frac{P_{out}(f)}{P_{in}(f)}\right]$$

The fiber bandwidth is defined as the lowest frequency at which the magnitude of the fiber frequency response has decreased to one-half its zero-frequency value. This is the -3 decibel (dB) optical power frequency (f_{3dB}). This frequency is referred to as the fiber bandwidth.

Bandwidth is normally given in units of megahertz-kilometers (MHz-km). Converting the –3 dB fiber bandwidth to a unit length assists in the analysis and comparison of optical fiber performance. For long lengths of fiber (>1km), the method for normalization is to multiply the length times the measured bandwidth.

Q9. The cutoff wavelength of matched-clad and depressed-clad single mode fibers varies according to the fiber's radius of curvature and length. The cutoff wavelength of which single mode fiber type is more sensitive to length?

Q10. Will the cutoff wavelength of uncabled fibers (λ_{cf}) generally have a value higher or lower than the cutoff wavelength of cabled fibers (λ_{cc})?

Q11. Describe the -3 decibel (dB) optical power frequency (f_{3dB}).

Chromatic Dispersion

Chromatic, or intramodal, dispersion occurs in both single mode and multimode optical fibers. Chromatic dispersion occurs because different colors of light travel through the fiber at different speeds.

Since the different colors of light have different velocities, some colors arrive at the fiber end before others. This delay difference is called the differential group delay $\tau(\lambda)$ per unit length. This differential group delay leads to pulse broadening.

Chromatic dispersion is measured using EIA/TIA-455-168 in the time domain. Chromatic dispersion is also measured in the frequency domain using EIA/TIA-455-169 and EIA/TIA-455-175. These methods measure the composite optical fiber material and waveguide dispersion. To understand the contribution that material and waveguide dispersive mechanisms have on multimode and single mode fiber dispersion, refer to chapter 2. In this chapter we limit the discussion on chromatic dispersion to the time domain method described in EIA/TIA-455-168.

The chromatic dispersion of multimode graded-index and single mode fiber is obtained by measuring fiber group delays in the time domain. These measurements are made using multiwavelength sources or multiple sources of different wavelengths. A multiwavelength source could be a wavelength-selectable laser.

The pulse delay for both a long test sample fiber and a short reference fiber are measured over a range of wavelengths. The pulse delay for the reference fiber as a function of wavelength is $\tau_{in}(\lambda)$. The pulse delay for the test fiber as a function of wavelength is $\tau_{out}(\lambda)$. The group delay $\tau(\lambda)$. per unit length at each wavelength is

$$\tau(\lambda) = \frac{\tau_{in}(\lambda) - \tau_{out}(\lambda)}{L_s - L_{ref}}$$

where L_s is the test sample fiber length in kilometers (km) and L_{ref} is the reference sample length in km.

The fiber chromatic dispersion is defined as the derivative, or slope, of the fiber group delay curve with respect to wavelength. Generally, the group delay as a function of wavelength is fit to a simple mathematical function and the derivative calculated. The range of wavelengths over which meaningful data is obtained depends on the wavelength range of optical source(s) used. The zero-dispersion wavelength (λ_0) and the zero-dispersion slope (S_0) are determined from the chromatic dispersion curve.

Q12. Delay differences between the source wavelengths occur as the optical signal propagates along the fiber. What is this called?

Q13. What determines the range of wavelengths over which meaningful data is obtained for calculating the chromatic dispersion?

Fiber Geometry

End users perform fiber geometry measurements to reduce system attenuation and coupling loss resulting from poor fiber fabrication. Fiber attenuation and intrinsic coupling loss result from mismatches in the inherent fiber characteristics of two connecting fibers. Fiber mismatches occur when manufacturers fail to maintain optical or structural (geometrical) tolerances during the fiber fabrication process. Fiber geometry measurements performed in the laboratory identify fiber mismatches before the optical fiber is installed.

The procedure for measuring multimode and single mode fiber geometry is detailed in EIA/TIA-455-176. The fiber-geometrical parameters measured include cladding diameter, cladding noncircularity, core-cladding concentricity error, and core noncircularity. Figure 4-8 (chapter 4) illustrates core noncircularity (ellipticity) and core-cladding concentricity error. The core noncircularity measurement is for multimode fibers only.

Other test methods are available for measuring other multimode and single mode fiber core parameters. Additional test methods exist for measuring multimode fiber core diameter and NA. For single mode fibers, the mode field diameter measurement replaces core diameter and NA measurements. Core diameter, numerical aperture, and mode field diameter measurements are identified and explained later in this chapter.

To make fiber geometry measurements, the input end of the fiber is overfilled and any cladding power stripped out. The output end of the fiber is prepared and viewed with a video camera. Generally the fiber is less than 10 m in length. An objective lens magnifies the output image (typically 20×) going to a video camera. The image from the video camera is displayed on a video monitor and is also sent to the computer for digital analysis.

The computer analyzes the image to identify the edges of the core and cladding. The centers r_c and r_g of the core and cladding, respectively, are found. The **cladding diameter** is defined as the average diameter of the cladding. The cladding diameter is twice the average radius (R_g). The **core diameter** is defined as the average diameter of the core. The core diameter is twice the average core radius (R_c).

Cladding noncircularity, or ellipticity, is the difference between the smallest radius of the fiber (R_{gmin}) and the largest radius (R_{gmax}) divided by the average cladding radius (R_g). The value of the cladding noncircularity is expressed as a percentage.

The **core-cladding concentricity error** for multimode fibers is the distance between the core and cladding centers divided by the core diameter. Multimode core-cladding concentricity error is expressed as a percentage of core diameter. The core-cladding concentricity error for single mode fibers is defined as the distance between the core and cladding centers.

Core noncircularity is the difference between the smallest core radius (R_{cmin}) and the largest core radius (R_{cmax}) divided by the core radius (R_c). The value of core noncircularity is expressed as a percentage. Core noncircularity is measured on multimode fibers only.

Q14. Why do end users perform fiber geometry measurements in the laboratory?

Q15. Define cladding diameter.

Q16. Explain the difference between multimode and single mode core-cladding concentricity errors.

Core Diameter

Core diameter is measured using EIA/TIA-455-58. The core diameter is defined from the refractive index profile n(r) or the output near-field radiation pattern. Our discussion is limited to measuring the core diameter directly from the output near-field radiation pattern obtained using EIA/TIA-455-43.

The near-field power distribution is defined as the emitted power per unit area (radiance) for each position in the plane of the emitting surface. For this chapter, the emitting surface is the output area of a fiber-end face. Near-field power distributions describe the emitted power per unit area in the near-field region. The near-field region is the region close to the fiber-end face. In the near-field region, the distance between the fiber-end face and detector is in the micrometers (μm) range.

EIA/TIA-455-43 describes the procedure for measuring the near-field power distribution of optical waveguides. Output optics, such as lenses, magnify the fiber-end face and focus the fiber's image on a movable detector. Figure 5-8 shows an example setup for measuring the near-field power distribution. The image is scanned in a plane by the movable detector. The image may also be scanned by using a detector array. Detector arrays of known element size and spacing may provide a display of the power distribution on a video monitor. A record of the near-field power is kept as a function of scan position.

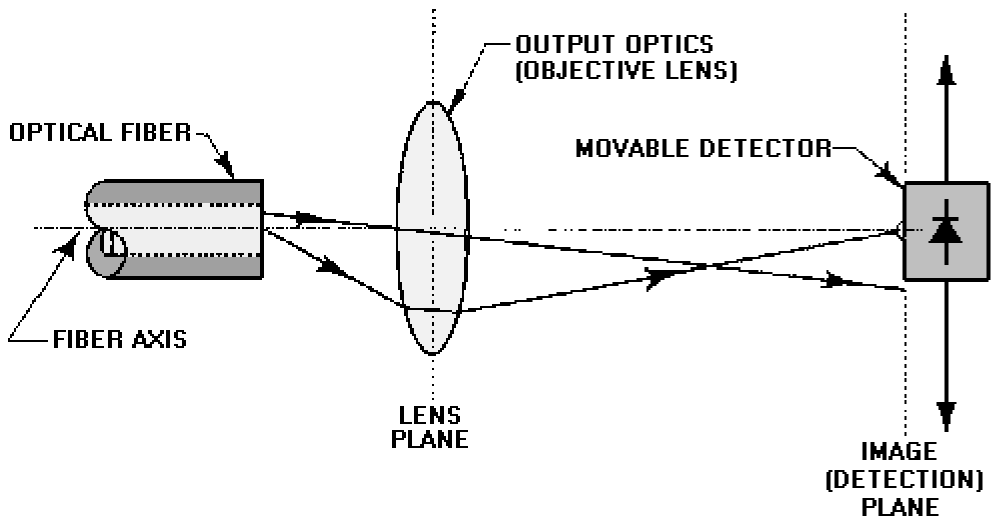


Figure 5-8.—The measurement of the near-field power distribution.

The core diameter (D) is derived from the normalized output near-field radiation pattern. The normalized near-field pattern is plotted as a function of radial position on the fiber-end face. Figure 5-9 shows a plot of the normalized near-field radiation pattern as a function of scan position.

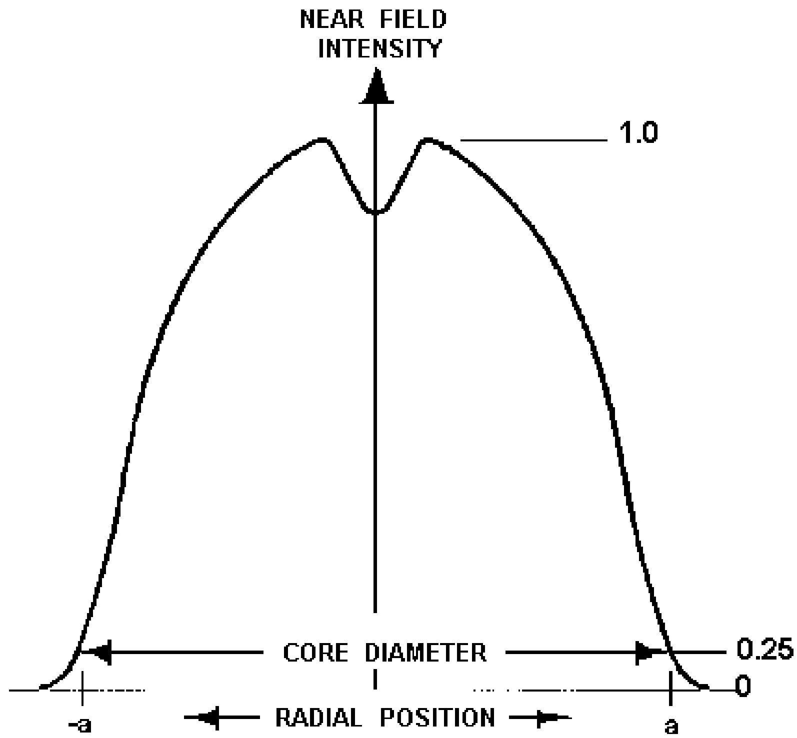


Figure 5-9.—Near-field radiation pattern.

The core diameter (D) is defined as the diameter at which the intensity is 2.5 percent of the maximum intensity (see figure 5-9). The 2.5 percent points, or the 0.025 level, intersects the normalized curve at radial positions –a and a. The core diameter is simply equal to 2a (D=2a).

Q17. Near-field power distributions describe the emitted power per unit area in the near-field region. Describe the near-field region.

Q18. How is the core diameter defined?

Numerical Aperture

The numerical aperture (NA) is a measurement of the ability of an optical fiber to capture light. The NA can be defined from the refractive index profile or the output far-field radiation pattern. Our discussion is limited to measuring the NA from the output far-field radiation pattern.

The NA of a multimode fiber having a near-parabolic refractive index profile is measured using EIA/TIA-455-177. In EIA/TIA-455-177, the fiber NA is measured from the output far-field radiation pattern. The far-field power distribution describes the emitted power per unit area in the far-field region. The far-field region is the region far from the fiber-end face. The far-field power distribution describes the emitted power per unit area as a function of angle Θ some distance away from the fiber-end face. The distance between the fiber-end face and detector in the far-field region is in the centimeters (cm) range for multimode fibers and millimeters (mm) range for single mode fibers.

EIA/TIA-455-47 describes various procedures, or methods, for measuring the far-field power distribution of optical waveguides. These procedures involve either an angular or spacial scan. Figure 5-10 illustrates an angular and spacial scan for measuring the far-field power distribution.

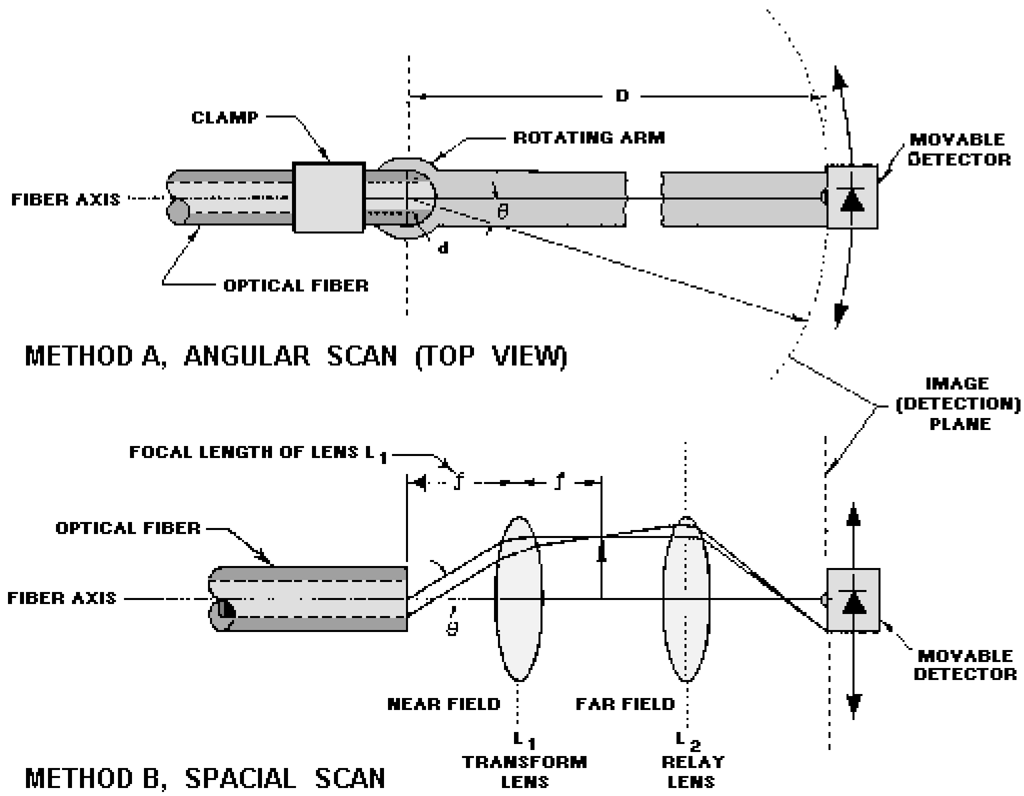


Figure 5-10.—Angular and spacial scan methods for measuring the far-field power distribution.

Figure 5-10 (method A) illustrates a far-field angular scan of the fiber-end face by a rotating detector. The fiber output radiation pattern is scanned by a rotating detector in the far-field. The detector rotates in a spherical manner. A record of the far-field power distribution is kept as a function of angle Θ.

Figure 5-10 (method B) illustrates a far-field spacial scan of the fiber-end face by a movable (planar) detector. In a far-field spacial scan, lens L_1 performs a Fourier transform of the fiber output near-field pattern. A second lens, L_2, is positioned to magnify and relay the transformed image to the detector plane. The image is scanned in a plane by a movable detector. The scan position y in the Fourier transform plane is proportional to the far-field scan angle Θ. A record of the far-field power distribution is kept as a function of the far-field scan angle.

The normalized far-field pattern is plotted as a function of the far-field scan angle Θ. Figure 5-11 shows the plot of the normalized far-field radiation pattern as a function of scan angle.

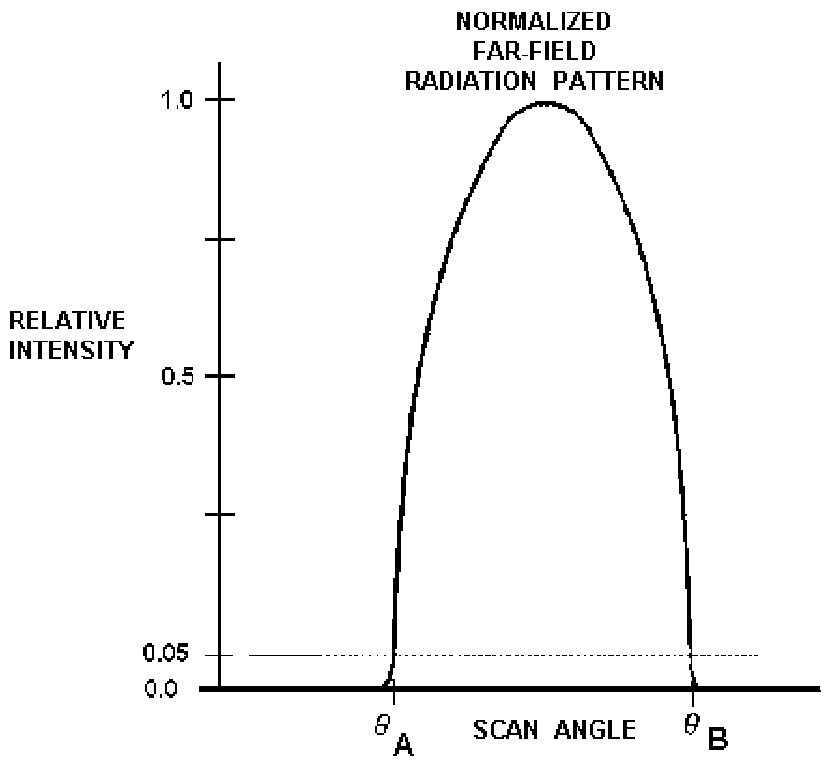


Figure 5-11.—Normalized far-field radiation pattern.

Fiber NA is defined by the 5 percent intensity level, or the 0.05 intensity level, as indicated in figure 5-11. The 0.05 intensity level intersects the normalized curve at scan angles Θ_A and Θ_B. The fiber NA is defined as

$$NA = \sin \Theta_5$$

where Θ_5 is the 5 percent intensity half angle. Θ_5 is determined from Θ_A and Θ_B as shown below:

$$\Theta_5 = \frac{\Theta_A - \Theta_B}{2}$$

Q19. Far-field power distributions describe the emitted power per unit area as a function of angle Θ in the far-field region. Describe the far-field region.

Mode Field Diameter

The mode field diameter (MFD) of a single mode fiber is related to the spot size of the fundamental mode. This spot has a mode field radius w_0. The mode field diameter is equal to $2w_0$. The size of the mode field diameter correlates to the performance of the single mode fiber. Single mode fibers with large mode field diameters are more sensitive to fiber bending. Single mode fibers with small mode field diameters show higher coupling losses at connections.

The mode field diameter of a single mode fiber can be measured using EIA/TIA-455-167. This method involves measuring the output far-field power distribution of the single mode fiber using a set of apertures of various sizes. This far-field power distribution data is transformed into the near-field before using complex mathematical procedures. The mode field diameter is calculated from the transformed near field data. The mathematics behind the transformation between the far-field and near-field is too complicated for discussion in this chapter. Refer to EIA/TIA-455-167 for information on this transformation procedure.

Q20. Will fiber coupling loss generally increase or decrease if the mode field diameter of a single mode fiber is decreased?

Insertion Loss

Insertion loss is composed of the connection coupling loss and additional fiber losses in the fiber following the connection. In multimode fiber, fiber joints can increase fiber attenuation following the joint by disturbing the fiber's mode power distribution (MPD). Fiber joints may increase fiber attenuation because disturbing the MPD may excite radiative modes. Radiative modes are unbound modes that radiate out of the fiber contributing to joint loss. In single mode fibers, fiber joints can cause the second-order mode to propagate in the fiber following the joint. As long as the coupling loss of the connection is small, neither radiative modes (multimode fiber) or the second-order mode (single mode fiber) are excited.

Insertion loss of both multimode and single mode interconnection devices is measured using EIA/TIA-455-34. For Navy applications, an overfill launch condition is used at the input fiber. For other applications a mandrel wrap may be used to strip out high-order mode power. The length of fiber before the connection and after the connection may be specified for some applications. Power measurements are made on an optical fiber or fiber optic cable before the joint is inserted and after the joint is made. Figure 5-12 illustrates the mandrel wrap method of measuring the insertion loss of an interconnecting device in EIA/TIA-455-34.

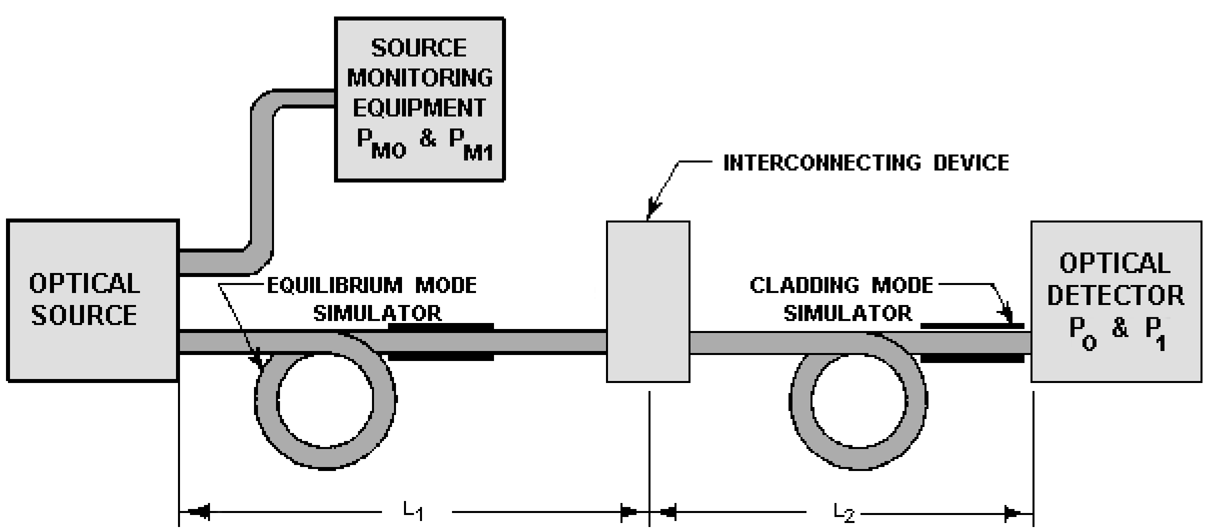


Figure 5-12.—Insertion loss measurement of an interconnecting device.

Initial power measurements at the detector (P_0) and at the source monitoring equipment (P_{M0}) are taken before inserting the interconnecting device into the test setup. The test fiber is then cut at the location specified by the end user. The cut results in a fiber of lengths L_1 and L_2 before and after the interconnection device that simulates the actual system configuration. After interconnection, the power at the detector (P_1) and at the source monitoring equipment (P_{M1}) is measured. The insertion loss is calculated as shown below:

$$\text{Insertion loss} = 10 \log \left\{ \frac{P_1}{P_o} \times \frac{P_{Mo}}{P_{M1}} \right\}$$

If the source power is constant, then the calculation of the insertion loss is similar to that of fiber attenuation.

Q21. In multimode fibers, how do fiber joints increase fiber attenuation following the joint?

Return Loss and Reflectance

Reflections occur at optical fiber connections. Optical power may be reflected back into the source fiber when connecting two optical fibers. In laser-based systems, reflected power reaching the optical source can reduce system performance by affecting the stability (operation) of the source. In addition, multiple reflections occur in fiber optic data links containing more than one connection. Multiple reflections can reduce data link performance by increasing the signal noise present at the optical detector.

Reflectance is a measure of the portion of incident light that is reflected back into the source fiber at the point of connection. Reflectance is given as a ratio (R) of the reflected light intensity to the incident light intensity. The reflectance ratio (R) for Fresnel reflection is discussed in chapter 4.

Return loss and reflectance are measured using EIA/TIA-455-107. They are measured using an optical source connected to one input of a 2×2 fiber optic coupler. Light is launched into the component under test through the fiber optic coupler. The light reflected from the component under test is transmitted back through the fiber optic coupler to a detector connected to the other input port. The optical power is measured at the output of the device under test (P_o) and at the input port of the coupler where the detector is located (P_r). P_o is corrected to account for the loss in power through the device under test. P_r is corrected to account for the loss in power through the coupler and any other connection losses in the path. The reflectance is then given by the ratio P_r/P_o.

Return loss is the amount of loss of the reflected light compared with the power of the incident beam at the interface. The optical return loss at the fiber interface is defined as

$$\text{Return loss} = -10 \text{ Log } R$$

Return loss is only the amount of optical power reflected and does not include power that is transmitted, absorbed, or scattered.

Q22. List two effects that reflections can have on a fiber optic data link.

Q23. Reflectance is given as what ratio?

Q24. Does return loss include power that is transmitted, absorbed, and/or scattered?

FIELD MEASUREMENTS

Field measurements differ from laboratory measurements because they measure the transmission properties of installed fiber optic components. Laboratory measurements can only attempt to simulate the actual operating conditions of installed components. Fiber optic component properties measured in the laboratory can change after the installation of these components on board ship. End users must perform field measurements to evaluate those properties most likely affected by the installation or repair of fiber optic components or systems.

The discussion on field measurements is limited to optical fiber and optical connection properties. Optical fiber and optical connection field measurements evaluate only the transmission properties affected by component or system installation or repair. Because optical fiber geometrical properties, such as core and cladding diameter and numerical aperture, are not expected to change, there is no need to remeasure these properties. The optical fiber properties that are likely to change include fiber attenuation (loss) and bandwidth. Bandwidth changes in the field tend to be beneficial, so field bandwidth measurement is generally not performed. If field bandwidth measurements are required, they are essentially the same as laboratory measurements so they will not be repeated. The optical connection properties that are likely to change are connection insertion loss and reflectance and return loss.

The installation and repair of fiber optic components on board ship can affect system operation. Microbends introduced during installation can increase fiber attenuation. Modal redistribution at fiber joints can increase fiber attenuation in the fiber after the joint. Fiber breaks or faults can prevent or severely disrupt system operation. Poor fiber connections can also increase insertion loss and degrade transmitter and receiver performance by increasing reflectance and return loss. End users should perform field measurements to verify that component performance is within allowable limits so system performance is not adversely affected.

There are additional differences in measuring optical fiber and optical connection properties in the field than in the laboratory. Field measurements require rugged, portable test equipment, unlike the sophisticated test equipment used in the laboratory. Field test equipment must provide accurate measurements in extreme environmental conditions. Since electrical power sources may not always be available in the field, test equipment should allow battery operation. In addition, while both fiber ends are available for conducting laboratory measurements, only one fiber end may be readily available for field measurements. Even if both fiber ends are available for field measurements, the fiber ends are normally located some distance apart. Therefore, field measurements may require two people.

The main field measurement technique involves optical time-domain reflectometry. An optical time-domain reflectometer (OTDR) is recommended for conducting field measurements on installed optical fibers or links of 50 meters or more in length. An OTDR requires access to only one fiber end. An OTDR measures the attenuation of installed optical fibers as a function of length. It also identifies and evaluates optical connection losses along a cable link and locates any fiber breaks or faults.

End users can also measure fiber attenuation and cable plant transmission loss using an optical power meter and a stabilized light source. End users use this measurement technique when optical time-domain reflectometry is not recommended. Measurements obtained with a stabilized light source and power meter are more accurate than those obtained with an OTDR. Measuring fiber attenuation and transmission loss using a power meter and light source requires access to both ends of the fiber or link. An optical loss test set (OLTS) combines the power meter and source functions into one physical unit.

Q25. Is it essential for end users to remeasure optical fiber geometrical properties after installation in the field?

Q26. When is an OTDR recommended for conducting field measurements on installed optical fibers or links?

Optical Time-Domain Reflectometry

End users use optical time-domain reflectometry to characterize optical fiber and optical connection properties in the field. In optical time-domain reflectometry, an OTDR transmits an optical pulse through an installed optical fiber. The OTDR measures the fraction of light that is reflected back due to Rayleigh scattering and Fresnel reflection. By comparing the amount of light scattered back at different times, the OTDR can determine fiber and connection losses. When several fibers are connected to form an installed cable plant, the OTDR can characterize optical fiber and optical connection properties along the entire length of the cable plant. A fiber optic cable plant consists of optical fiber cables, connectors, splices, mounting panels, jumper cables, and other passive components. A cable plant does not include active components such as optical transmitters or receivers.

The OTDR displays the backscattered and reflected optical signal as a function of length. The OTDR plots half the power in decibels (dB) versus half the distance. Plotting half the power in dB and half the distance corrects for round trip effects. By analyzing the OTDR plot, or trace, end users can measure fiber attenuation and transmission loss between any two points along the cable plant. End users can also measure insertion loss and reflectance of any optical connection. In addition, end users use the OTDR trace to locate fiber breaks or faults.

Figure 5-13 shows an example OTDR trace of an installed cable plant. OTDR traces can have several common characteristics. An OTDR trace begins with an initial input pulse. This pulse is a result of Fresnel reflection occurring at the connection to the OTDR. Following this pulse, the OTDR trace is a gradual downsloping curve interrupted by abrupt shifts. Periods of gradual decline in the OTDR trace result from Rayleigh scattering as light travels along each fiber section of the cable plant. Periods of gradual decline are interrupted by abrupt shifts called point defects. A point defect is a temporary or permanent local deviation of the OTDR signal in the upward or downward direction. Point defects are caused by connectors, splices, or breaks along the fiber length. Point defects, or faults, can be reflective or nonreflective. An output pulse at the end of the OTDR trace indicates the end of the fiber cable plant. This output pulse results from Fresnel reflection occurring at the output fiber-end face.

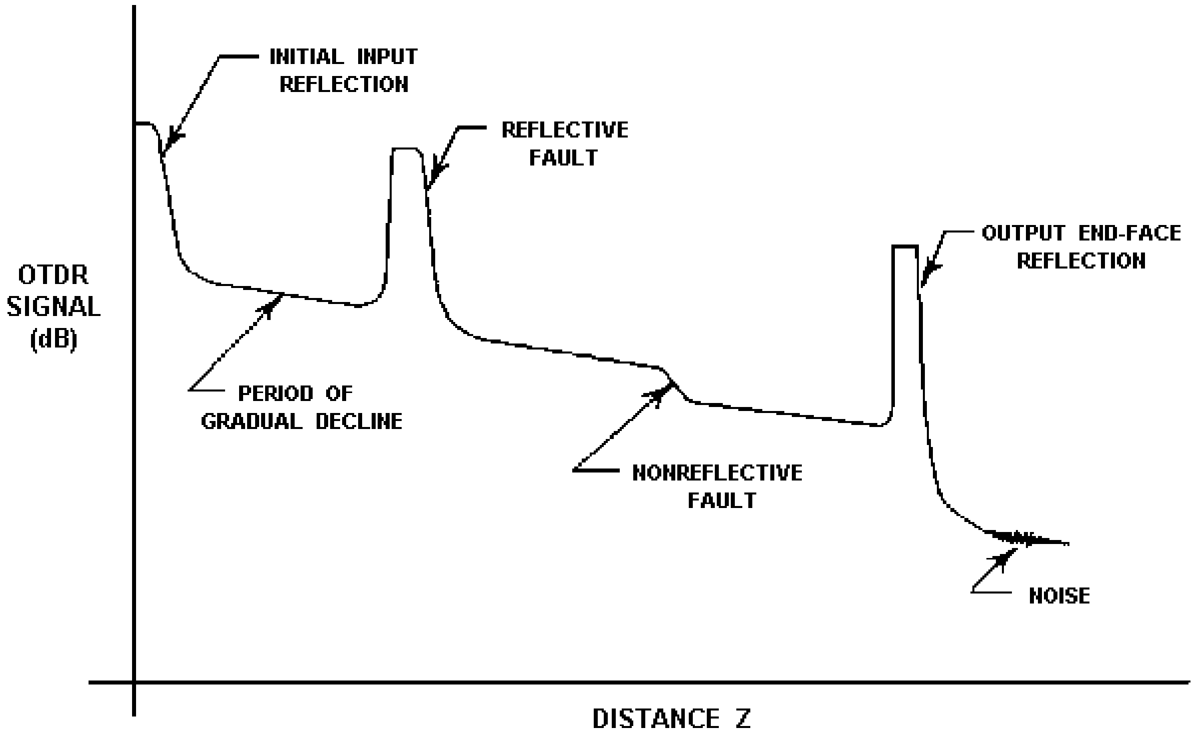


Figure 5-13.—OTDR trace of an installed cable plant.

ATTENUATION.—The fiber optic test method for measuring the attenuation of an installed optical fiber using an OTDR is EIA/TIA-455-61. The accuracy of this test method depends on the user entering the appropriate source wavelength, pulse duration, and fiber length (test range) into the OTDR. In addition, the effective group index of the test fiber is required before the attenuation coefficient and accurate distances can be recorded. The group index (N) is provided by fiber manufacturers or is found using EIA/TIA-455-60. By entering correct test parameters, OTDR fiber attenuation values will closely coincide with those measured by the cutback technique.

Test personnel can connect the test fiber directly to the OTDR or to a dead-zone fiber. This dead-zone fiber is placed between the test fiber and OTDR to reduce the effect of the initial reflection at the OTDR on the fiber measurement. The dead-zone fiber is inserted because minimizing the reflection at a fiber joint is easier than reducing the reflection at the OTDR connection.

Figure 5-14 illustrates the OTDR measurement points for measuring the attenuation of the test fiber using a dead-zone fiber. Fiber attenuation between two points along the test fiber is measured on gradual downsloping sections on the OTDR trace. There should be no point defects present along the portion of fiber being tested.

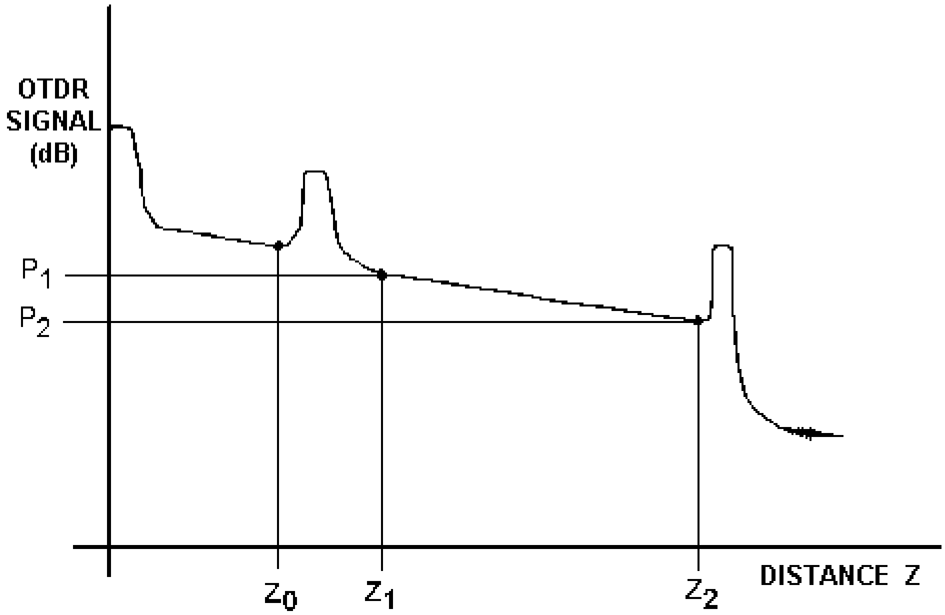


Figure 5-14.—OTDR measurement points for measuring fiber attenuation using a dead-zone fiber.

OTDRs are equipped with either manual or automatic cursors to locate points of interest along the trace. In figure 5-14, a cursor is positioned at a distance z_o on the rising edge of the reflection at the end of the dead-zone fiber. Cursors are also positioned at distances z_1 and z_2. The cursor positioned at z_1 is just beyond the recovery from the reflection at the end of the dead-zone fiber. Since no point defects are present in figure 5-14, the cursor positioned at z_2 locates the end of the test fiber. Cursor z_2 is positioned just before the output pulse resulting from Fresnel reflection occurring at the end of the test fiber.

The attenuation of the test fiber between points z_1 and z_2 is $(P_1 - P_2)$ dB. The attenuation coefficient (α) is

$$a = \frac{(P_1 - P_2}{(z_2 - z_1)} \text{dB/km.}$$

The total attenuation of the fiber including the dead zone after the joint between the dead-zone fiber and test fiber is

$$\text{Attenuation} = (P_1 - P_2)\frac{(z_1 - z_o)}{(z_2 - z_1)} \text{dB.}$$

If fiber attenuation is measured without a dead-zone fiber, z_0 is equal to zero ($z_0 = 0$).

At any point along the length of fiber, attenuation values can change depending on the amount of optical power backscattered due to Rayleigh scattering. The amount of backscattered optical power at each point depends on the forward optical power and its backscatter capture coefficient. The backscatter capture coefficient varies with length depending on fiber properties. Fiber properties that may affect the backscatter coefficient include the refractive index profile, numerical aperture (multimode), and mode-field diameter (single mode) at the particular measurement point. The source wavelength and pulse width may also affect the amount of backscattered power.

By performing the OTDR attenuation measurement in each direction along the test fiber, test personnel can eliminate the effects of backscatter variations. Attenuation measurements made in the opposite direction at the same wavelength (within 5 nm) are averaged to reduce the effect of backscatter variations. This process is called bidirectional averaging. Bidirectional averaging is possible only if test personnel have access to both fiber ends. OTDR attenuation values obtained using bidirectional averaging should compare with those measured using the cutback technique in the laboratory.

POINT DEFECTS.—Point defects are temporary or local deviations of the OTDR signal in the upward or downward direction. A point defect, or fault, can be reflective or nonreflective. A point defect normally exhibits a loss of optical power. However, a point defect may exhibit an apparent power gain. In some cases, a point defect can even exhibit no loss or gain. Refer back to figure 5-13; it illustrates a reflective fault and a nonreflective fault, both exhibiting loss. Figure 5-15 shows a nonreflective fault with apparent gain and a reflective fault with no apparent loss or gain.

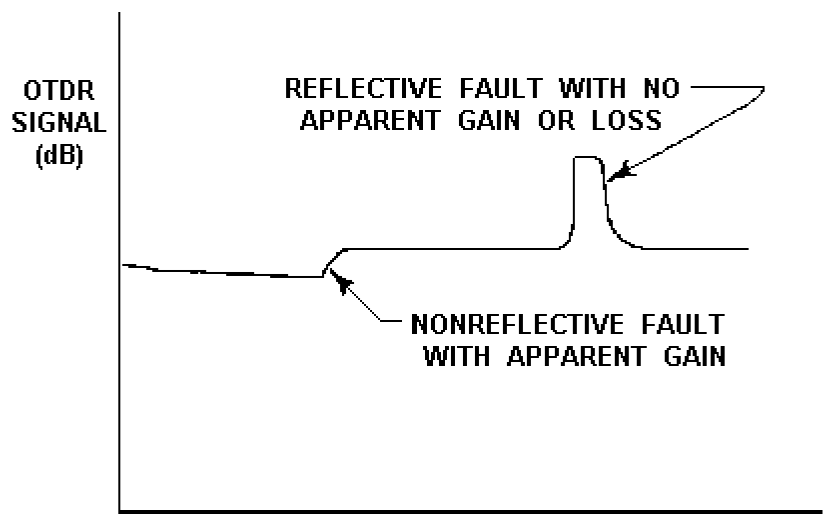


Figure 5-15.—An OTDR trace showing a nonreflective fault with apparent gain and a reflective fault with no apparent loss or gain.

Point defects are located and measured using EIA/TIA-455-59. Test personnel must enter the appropriate input parameters including the source wavelength, the pulse duration, and the fiber or cable group index into the OTDR. The nature of fiber point defects depends on the value of each parameter entered by the end user. The pulse duration usually limits the length of the point defect while other input parameters, such as the wavelength, can vary its shape.

If the length of the fiber point defect changes with the pulse duration, then the OTDR signal deviation is in fact a point defect. If the length remains the same, then the OTDR signal deviation is a region of high fiber attenuation. Regions of high fiber attenuation are referred to as attenuation non-uniformities.

Fiber point defects occur from factory fiber splices or bends introduced during cable construction or installation. For shipboard applications, manufacturers are not allowed to splice fibers during cable construction. Fiber joints are natural sources of OTDR point defects. However, fiber breaks, cracks, or microbends introduced during cable installation are additional sources of point defects.

Point defects that occur at fiber joints are relatively easy to identify because the location of a fiber joint is generally known. A reflective or nonreflective fault occurs at a distance equal to fiber joint location. In most circumstances, an optical connector produces a reflective fault, while an optical splice produces a nonreflective fault.

Reflective and nonreflective faults occurring at distances other than fiber joint locations identify fiber breaks, cracks, or microbends. A fiber break produces a reflective fault because fiber breaks result in complete fiber separation. Fiber cracks and microbends generally produce nonreflective faults.

A point defect may exhibit apparent gain because the backscatter coefficient of the fiber present before the point defect is higher than that of the fiber present after. Test personnel measure the signal loss or gain by positioning a pair of cursors, one on each side of the point defect. Figure 5-16 illustrates the positioning of the cursors for a point defect showing an apparent signal gain. The trace after the point defect is extrapolated as shown in figure 5-16. The vertical distance between the two lines in figure 5-16 is the apparent gain of the point defect.

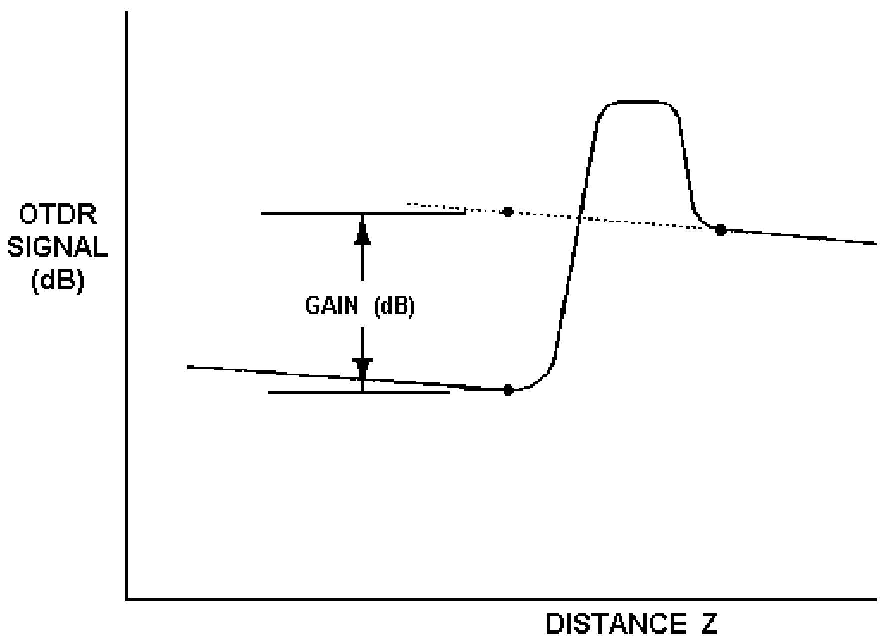


Figure 5-16.—Extrapolation for a point defect showing an apparent signal gain.

Point defects exhibiting gain in one direction will exhibit an exaggerated loss in the opposite direction. Figure 5-17 shows the apparent loss shown by the OTDR for the same point defect shown in figure 5-16 when measured in the opposite direction. Bidirectional measurements are conducted to cancel the effects of backscatter coefficient variations. Bidirectional averaging combines the two values to identify the true signal loss. Bidirectional averaging is possible only if test personnel have access to both ends of the test sample.

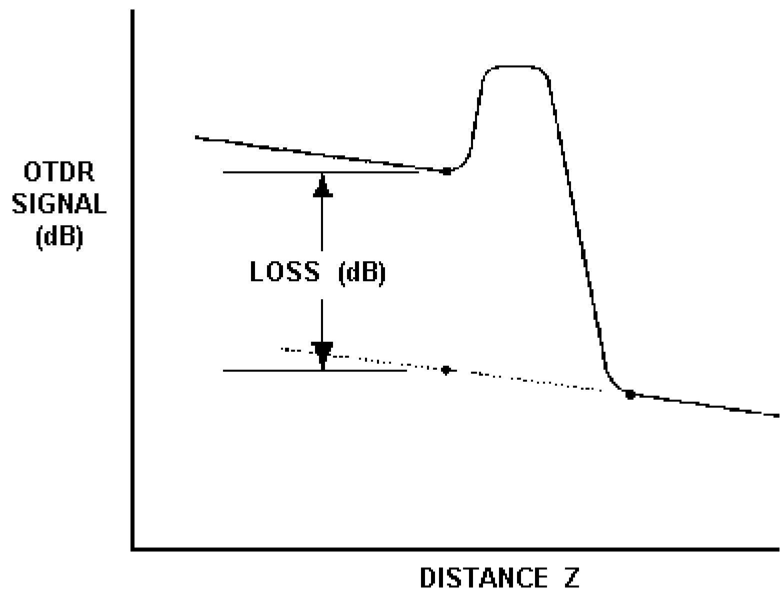


Figure 5-17.—The exaggerated loss obtained at point defects exhibiting gain in one direction by conducting the OTDR measurement in the opposite direction.

OTDRs can also measure the return loss of a point defect. However, not all OTDRs are configured to make the measurement. To measure the return loss of a point defect, the cursors are placed in the same places as for measuring the loss of the point defect. The return loss of the point defect is displayed when the return loss option is selected on the OTDR. The steps for selecting the return loss option depend upon the OTDR being used.

Q27. An OTDR measures the fraction of light that is reflected back from the fiber or link under test. What causes light to be reflected back into the OTDR?

Q28. List the types of fiber optic components considered part of a fiber optic cable plant.

Q29. What is a temporary or permanent local deviation of the OTDR signal in the upward or downward direction called?

Q30. Why is a dead-zone fiber placed between the test fiber and OTDR when conducting attenuation measurements?

Q31. The amount of backscattered optical power at each point depends on what two properties?

Q32. How can test personnel eliminate the effects of backscatter variations?

Q33. If the length of the fiber point defect changes with pulse duration, is the OTDR signal deviation a point defect or a region of high fiber attenuation?

Q34. Give the type of fault (reflective or nonreflective) normally produced by: (a) fiber breaks, (b) fiber cracks, and (c) fiber microbends.

Q35. Explain how a point defect may exhibit an apparent gain.

Q36. A point defect exhibiting an apparent gain in one direction will exhibit what, when measured in the opposite direction?

Power Meter

Test personnel also use an optical power meter and stabilized light source to measure fiber attenuation and transmission loss in the field. Optical power meter measurements are recommended when the length of an installed optical fiber cable or cable plant is less than 50 meters. A test jumper is used to couple light from the stabilized source to one end of the optical fiber (or cable plant) under test. An additional test jumper is also used to connect the other end of the optical fiber (or cable plant) under test to the power meter. Optical power meter measurements may be conducted using an optical loss test set (OLTS). An OLTS combines the power meter and source functions into one physical unit. When making measurements, it does not matter whether the stabilized source and power meter are in one physical unit or two.

Power meter measurements are conducted on individual optical fiber cables installed on board ship. The installed optical fiber cable must have connectors or terminations on both ends to make the measurement. If the installed optical fiber cable does not have connectors or terminations on both ends, an OTDR should be used to evaluate the cable. If the cable is too short for evaluation with an OTDR, cable continuity can be verified using a flashlight.

Power meter measurements for cable assembly link loss require that test personnel clean all optical connections at test jumper interfaces before performing any measurement. Test personnel should use cotton wipes dampened with alcohol to clean connectors and blow dry before making connections. End users should also ensure that test equipment calibration is current.

Power meter measurements connecting a test reference cable between the light source and power meter. The test reference cable has the same nominal fiber characteristics as the cable under test. The optical power present at the power meter is the reference power (P_1). Disconnect the test reference cable and connect the optical fiber cable under test between the light source and power meter using test jumpers. If possible, the test reference cable should be used as the input jumper cable for the test cable measurement. The test jumper fiber properties, such as core diameter and NA, should be nominally equal to the fiber properties of the cable being tested. The optical power present at the power meter is test power (P_2).

Test personnel use P_1 and P_2 to calculate the cable assembly link loss. The cable assembly link loss (B_{CA}) of optical fiber installed with connectors or terminations on both ends is

$$B_{CA} = (P_1 = P_2)\ \mathrm{dB}$$

The cable assembly link loss should always be less than the specified link loss for that particular link.

Besides measuring individual cables, test personnel measure the transmission loss of installed fiber optic cable plants. The transmission loss of fiber optic cable plants is measured using EIA/TIA-526-14 method B (multimode fiber) or EIA/TIA-526-7 (single mode fiber). The procedure measures the internal loss of the cable plant between points A and B, plus two connection losses. Figure 5-18 (A) illustrates the method described in EIA/TIA-526-14 method B for measuring the reference power (P_1). Figure 5-18 (B) shows the final test configuration for measuring the cable plant test power (P_2).

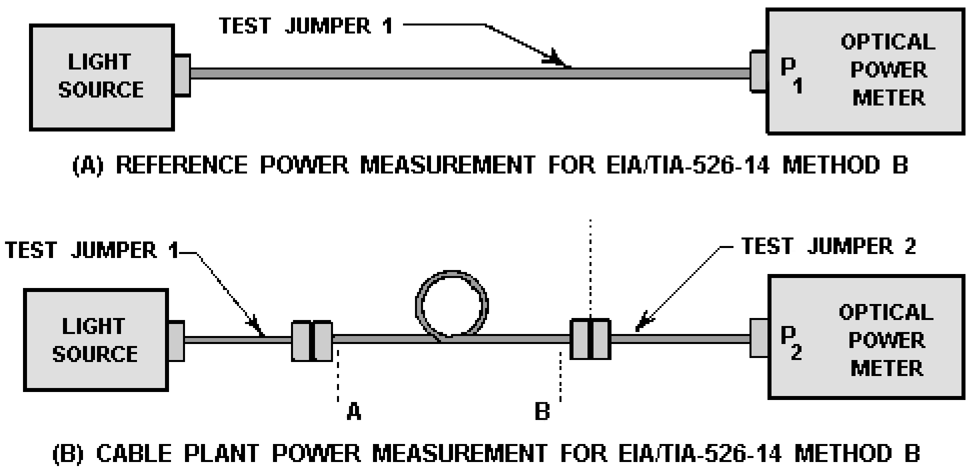


Figure 5-18.—EIA/TIA-526-14 methods for measuring the reference power (P_1).

The procedure is exactly the same as described for measuring the link loss of an individual cable assembly. The total optical loss between any two termination points, including the end terminations, of the optical fiber cable plant link is measured. The measured cable plant link loss should always be less than the specified cable plant link loss.

Test personnel should conduct cable assembly link loss, and cable plant transmission loss measurements in both directions and at each system operational wavelength. By performing these measurements in each direction, test personnel can better characterize cable and link losses. Unlike optical time-domain reflectometry, bidirectional readings are always possible when performing power meter measurements. In power meter measurements, by definition, end users have access to both ends of the cable or cable plant.

Q37. When is an optical power meter measurement recommended for conducting field measurements on installed optical fiber cables or cable plants?

Q38. If an installed optical fiber cable does not have connectors or terminations on both ends, how should the cable be tested?

SUMMARY

Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas that you have learned. You should have a thorough understanding of these principles before moving on to chapter 6.

END USERS (equipment manufacturers, shipbuilders, maintenance personnel, test personnel, and so on) should measure some component parameters upon receipt before installing the component into the fiber optic data link. In addition, they should measure some component parameters after installing or repairing fiber optic components in the field.

LABORATORY MEASUREMENTS of the optical fiber and optical connections performed by end users in the laboratory include attenuation, cutoff wavelength (single mode), bandwidth (multimode), chromatic dispersion, fiber geometry, core diameter, numerical aperture (multimode), mode field diameter (single mode), insertion loss, and reflectance and return loss.

ATTENUATION is the loss of optical power as light travels along the fiber. It is a result of absorption, scattering, bending, and other loss mechanisms.

The **LAUNCH SPOT SIZE** and the **ANGULAR DISTRIBUTION** may affect multimode fiber attenuation measurement results by affecting modal distributions.

An **UNDERFILLED** launch results when the launch spot size and angular distribution are smaller than that of the fiber core.

An **OVERFILLED** launch condition occurs when the launch spot size and angular distribution are larger than that of the fiber core.

A **CLADDING-MODE STRIPPER** is a device that removes any cladding mode power from the fiber.

A **MODE FILTER** is a device that attenuates a specific mode or modes propagating in the core of an optical fiber.

The **CUTOFF WAVELENGTH** of a single mode fiber is the wavelength above which the fiber propagates only the fundamental mode. The cutoff wavelength of a single mode fiber varies according to the fiber's radius of curvature and length. The fiber cutoff wavelength (λ_{cf}) will generally be higher than the cable cutoff wavelength (λ_{cc}).

PULSE DISTORTION is the spreading of the light pulse as it travels along the fiber caused by dispersion. It reduces the bandwidth, or information-carrying capacity, of an optical fiber.

Two **BASIC TECHNIQUES** are used for measuring the modal bandwidth of an optical fiber. The first characterizes dispersion by measuring the **IMPULSE RESPONSE h(t)** of the fiber in the time domain. The second characterizes modal dispersion by measuring the **BASEBAND FREQUENCY RESPONSE H(*f*)** of the fiber in the frequency domain.

The **LOWEST FREQUENCY** at which the magnitude of the fiber frequency response has decreased to one half its zero-frequency value is the -3 decibel (dB) optical power frequency (f_{3dB}).

CHROMATIC DISPERSION occurs because different colors of light travel through the fiber at different speeds. Since the different colors of light have different velocities, some colors arrive at the fiber end before others.

The **DIFFERENTIAL GROUP DELAY** $\tau(\lambda)$ is the variation in propagation delay that occurs because of the different group velocities of each wavelength in an optical fiber.

The **RANGE OF WAVELENGTHS** over which meaningful chromatic dispersion data is obtained depends on the wavelength range of optical source(s) used.

FIBER GEOMETRY MEASUREMENTS are performed by end users to reduce system attenuation and coupling loss resulting from poor fiber fabrication.

The **CLADDING DIAMETER** is the average diameter of the cladding.

CLADDING NONCIRCULARITY, or ellipticity, is the difference between the smallest radius of the fiber (R_{gmin}) and the largest radius of the fiber (R_{gmax}) divided by the average cladding radius (R_g).

The **CORE-CLADDING CONCENTRICITY ERROR** for multimode fibers is the distance between the core and cladding centers divided by the core diameter. The core-cladding concentricity error for single mode fibers is defined as the distance between the core and cladding centers.

CORE NONCIRCULARITY is the difference between the smallest radius of the core (R_{cmin}) and the largest radius of the core (R_{cmax}) divided by the core radius (R_c).

The **NEAR-FIELD POWER DISTRIBUTION** is defined as the emitted power per unit area (radiance) for each position in the plane of the emitting surface.

The **NEAR-FIELD REGION** is the region close to the fiber-end face.

The **CORE DIAMETER** is derived from the normalized output near-field radiation pattern. The core diameter (D) is defined as the diameter at the 2.5 percent (0.025) level.

The **NUMERICAL APERTURE (NA)** is a measurement of the ability of a multimode optical fiber to capture light.

The **FAR-FIELD POWER DISTRIBUTION** describes the emitted power per unit area as a function of angle Θ some distance away from the fiber-end face.

The **FAR-FIELD REGION** is the region far from the fiber-end face.

Single mode fibers with large **MODE FIELD DIAMETERS** are more sensitive to fiber bending. Single mode fibers with small mode field diameters show higher coupling losses at connections.

INSERTION LOSS is composed of the connection coupling loss and additional fiber losses in the fiber following the connection.

REFLECTANCE is a measure of the portion of incident light that is reflected back into the source fiber at the point of connection.

RETURN LOSS is the amount of loss of the reflected light compared with the power of the incident beam at the interface.

OPTICAL FIBER and **OPTICAL CONNECTION FIELD MEASUREMENTS** measure only the transmission properties affected by component or system installation or repair.

OPTICAL TIME-DOMAIN REFLECTOMETRY is recommended for conducting field measurements on installed optical fibers or cable plants of 50 meters or more in length.

An **OPTICAL LOSS TEST SET (OLTS)** combines the power meter and source functions into one physical unit.

An **OPTICAL TIME-DOMAIN REFLECTOMETER (OTDR)** measures the fraction of light that is reflected back because of Rayleigh scattering and Fresnel reflection.

A **FIBER OPTIC CABLE PLANT** consists of optical fiber cables, connectors, splices, mounting panels, jumper cables, and other passive components. A cable plant does not include active components such as optical transmitters or receivers.

A **POINT DEFECT** is a temporary or permanent local deviation of the OTDR signal in the upward or downward direction. Point defects are caused by connectors, splices, or fiber breaks. Point defects, or faults, can be reflective or nonreflective.

A **DEAD-ZONE** fiber is placed between the test fiber and OTDR to reduce the influence of the initial pulse resulting from Fresnel reflection at the OTDR connection.

The **AMOUNT OF OPTICAL POWER BACKSCATTERED** because of Rayleigh scattering at one point depends on the forward optical power and the fibers backscatter capture coefficient.

The **EFFECTS OF BACKSCATTER VARIATIONS** can be eliminated by test personnel by performing the OTDR attenuation measurement in each direction along the test fiber and averaging (bidirectional readings).

A **POINT DEFECT** may exhibit apparent gain because the backscatter coefficient of the fiber present before the point defect is higher than that of the fiber present after.

To **MEASURE FIBER ATTENUATION** and **TRANSMISSION LOSS** in the field, test personnel use an optical power meter and stabilized light source.

OPTICAL POWER METER MEASUREMENTS are recommended when the length of an installed optical fiber cable or cable plant is less than 50 meters.

ANSWERS TO QUESTIONS Q1. THROUGH Q38.

A1. Cladding diameter, core diameter, numerical aperture, and mode field diameter.

A2.

$$A = 10\log\frac{P_x}{P_y}\,\mathrm{dB}$$

A3. Yes.

A4. Launch spot size and angular distribution.

A5. Low-order modes.

A6. Overfilled.

A7. Second-order mode.

A8. Free-form loop and mandrel wrap.

A9. Depressed-clad.

A10. Higher.

A11. The -3 decibel (dB) is the lowest frequency at which the magnitude of the fiber frequency response has decreased to one half its zero-frequency value.

A12. Differential group delay $\tau(\lambda)$.

A13. The wavelength range of the optical source(s) used.

A14. To reduce system attenuation and coupling loss resulting from poor fiber fabrication.

A15. The cladding diameter is the average diameter of the cladding.

A16. Multimode core-cladding concentricity error is the distance between the core and cladding centers expressed as a percentage of core diameter while the single mode core-cladding concentricity error is just the distance between the core and cladding centers.

A17. The near-field region is the region close to the fiber-end face.

A18. The core diameter is defined as the diameter at which the near-field intensity is 2.5 percent of the maximum intensity.

A19. The far-field region is the region far from the fiber-end face.

A20. Increase.

A21. By disturbing the fiber's mode power distribution (MPD).

A22. Reduce the stability of the system source and increase the signal noise present at the optical detector.

A23. The ratio of reflected optical power to incident optical power.

A24. No.

A25. No.

A26. When installed optical fiber cables or links are 50 meters or more in length.

A27. Rayleigh scattering and Fresnel reflection.

A28. Optical fiber cables, connectors, splices, mounting panels, jumper cables, and other passive components.

A29. A point defect.

A30. To reduce the effect of the initial reflection at the OTDR.

A31. Forward optical power and backscatter capture coefficient.

A32. By performing the OTDR attenuation measurements in each direction along the test fiber.

A33. A point defect.

A34. (a) Reflective, (b) nonreflective, and (c) nonreflective.

A35. A point defect may exhibit apparent gain because the backscatter coefficient of the fiber present before the point defect is higher than that of the fiber present after.

A36. An exaggerated loss.

A37. When an installed optical fiber cable or cable plant is less than 50 meters in length.

A38. With an OTDR unless it is less than 50 meters in length. If it is less than 50 meters in length, continuity should be verified with a flashlight.

CHAPTER 6

OPTICAL SOURCES AND FIBER OPTIC TRANSMITTERS

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to do the following:

1. Explain the principal properties of an optical source and fiber optic transmitter.
2. Discuss the optical emission properties of semiconductor light-emitting diodes (LEDs) and laser diodes (LDs).
3. Describe the operational differences between surface-emitting LEDs (SLEDs), edge-emitting LEDs (ELEDs), superluminescent diodes (SLDs), and laser diodes.
4. Describe typical fiber optic transmitter packages.

INTRODUCTION TO OPTICAL SOURCES AND FIBER OPTIC TRANSMITTERS

Chapter 1 taught you that a fiber optic data link has three basic functions. One function is that a fiber optic data link must convert an electrical signal to an optical signal permitting the transfer of data along an optical fiber. The fiber optic device responsible for that signal conversion is a fiber optic transmitter.

A fiber optic transmitter is a hybrid device. It converts electrical signals into optical signals and launches the optical signals into an optical fiber. A fiber optic transmitter consists of an interface circuit, a source drive circuit, and an optical source. The interface circuit accepts the incoming electrical signal and processes it to make it compatible with the source drive circuit. The source drive circuit intensity modulates the optical source by varying the current through the source.

An optical source converts electrical energy (current) into optical energy (light). Light emitted by an optical source is launched, or coupled, into an optical fiber for transmission. Fiber optic data link performance depends on the amount of optical power (light) launched into the optical fiber. This chapter attempts to provide an understanding of light-generating mechanisms within the main types of optical sources used in fiber optics.

Q1. What are the three parts of a fiber optic transmitter?

Q2. Which part of a fiber optic transmitter converts the processed electrical signal to an optical signal?

OPTICAL SOURCE PROPERTIES

The development of efficient semiconductor optical sources, along with low-loss optical fibers, led to substantial improvements in fiber optic communications. Semiconductor optical sources have the physical characteristics and performance properties necessary for successful implementations of fiber optic systems. It is desirable that optical sources:

- Be compatible in size to low-loss optical fibers by having a small light-emitting area capable of launching light into fiber

- Launch sufficient optical power into the optical fiber to overcome fiber attenuation and connection losses allowing for signal detection at the receiver

- Emit light at wavelengths that minimize optical fiber loss and dispersion. Optical sources should have a narrow spectral width to minimize dispersion

- Allow for direct modulation of optical output power

- Maintain stable operation in changing environmental conditions (such as temperature)

- Cost less and be more reliable than electrical devices, permitting fiber optic communication systems to compete with conventional systems

Semiconductor optical sources suitable for fiber optic systems range from inexpensive light-emitting diodes (LEDs) to more expensive semiconductor lasers. Semiconductor LEDs and laser diodes (LDs) are the principal light sources used in fiber optics.

OPERATING WAVELENGTH

Fiber optic communication systems operate in the 850-nm, the 1300-nm, and the 1550-nm wavelength windows. Semiconductor sources are designed to operate at wavelengths that minimize optical fiber absorption and maximize system bandwidth. By designing an optical source to operate at specific wavelengths, absorption from impurities in the optical fiber, such as hydroxyl ions (OH^-), can be minimized. Maximizing system bandwidth involves designing optical fibers and sources that minimize chromatic and intermodal dispersion at the intended operational wavelength.

Initially, the material properties of semiconductor optical sources provided for optical emission in the 850-nm wavelength region. An 850-nm operational wavelength avoids fiber absorption loss from OH^- impurities near the 900-nm wavelength. Light sources for 850-nm systems were originally semiconductor LEDs and lasers. Currently, most 850-nm systems use LEDs as a light source. LEDs operating at 850-nm provide sufficient optical power for short-distance, low-bandwidth systems. However, multimode fiber dispersion, the relatively high fiber attenuation, and the LED's relatively low optical output power prevent the use of these devices in longer-distance, higher bandwidth systems.

The first development allowing the operational wavelength to move from 850 nm to 1300 nm was the introduction of multimode graded-index fibers. Multimode graded-index fibers have substantially lower intermodal dispersion than multimode step-index fibers. Systems operating at 850 nm cannot take full advantage of the fiber's low intermodal dispersion because of high chromatic dispersion at 850 nm. However, the use of multimode graded-index fibers allow 850-nm LEDs to operate satisfactorily in short-distance, higher bandwidth systems.

Following the enhancements in multimode fiber design, next generation LEDs were designed to provide optical emission in the 1300-nm region. Multimode graded-index fiber systems using these LEDs can operate over longer distances and at higher bandwidths than 850-nm systems. Longer distances and higher bandwidths are possible because fiber material losses and dispersion are significantly reduced at the 1300-nm region.

Advances in single mode fiber design and construction sped the development of semiconductor LEDs and LDs optimized for single mode fibers. Single mode fibers have very low dispersion values. However, existing LEDs were unable to focus and launch sufficient optical power into single mode fibers for long-haul, very high-bandwidth communication systems. New semiconductor LEDs and LDs capable of operating with single mode fibers at 1300 nm were developed to take advantage of single mode fiber's very low value of dispersion. Additionally, LEDs and LDs operating at 1550 nm were developed to take advantage of the fiber's lowest loss.

Q3. LEDs operating at 850 nm provide sufficient optical power for short-distance, low-bandwidth multimode systems. List three conditions that prevent the use of LEDs in longer distance, higher bandwidth multimode systems.

Q4. Why can multimode graded-index fiber 1300-nm systems using LEDs operate over longer distances and at higher bandwidths than 850-nm systems?

SEMICONDUCTOR LIGHT-EMITTING DIODES AND LASER DIODES

Semiconductor LEDs emit incoherent light. Spontaneous emission of light in semiconductor LEDs produces light waves that lack a fixed-phase relationship. Light waves that lack a fixed-phase relationship are referred to as incoherent light. Spontaneous emission of light is discussed in more detail later in this chapter. The use of LEDs in single mode systems is severely limited because they emit unfocused incoherent light. Even LEDs developed for single mode systems are unable to launch sufficient optical power into single mode fibers for many applications. LEDs are the preferred optical source for multimode systems because they can launch sufficient power at a lower cost than semiconductor LDs.

Semiconductor LDs emit coherent light. LDs produce light waves with a fixed-phase relationship (both spatial and temporal) between points on the electromagnetic wave. Light waves having a fixed-phase relationship are referred to as coherent light. Stimulated emission of light is discussed later in this chapter. Since semiconductor LDs emit more focused light than LEDs, they are able to launch optical power into both single mode and multimode optical fibers. However, LDs are usually used only in single mode fiber systems because they require more complex driver circuitry and cost more than LEDs.

Optical power produced by optical sources can range from microwatts (μW) for LEDs to tens of milliwatts (mW) for semiconductor LDs. However, it is not possible to effectively couple all the available optical power into the optical fiber for transmission.

The amount of optical power coupled into the fiber is the relevant optical power. It depends on the following factors:

- The angles over which the light is emitted
- The size of the source's light-emitting area relative to the fiber core size
- The alignment of the source and fiber
- The coupling characteristics of the fiber (such as the NA and the refractive index profile)

Typically, semiconductor lasers emit light spread out over an angle of 10 to 15 degrees. Semiconductor LEDs emit light spread out at even larger angles. Coupling losses of several decibels can easily occur when coupling light from an optical source to a fiber, especially with LEDs.

Source-to-fiber coupling efficiency is a measure of the relevant optical power. The coupling efficiency depends on the type of fiber that is attached to the optical source. Coupling efficiency also depends on the coupling technique. Source-to-fiber coupling involves centering a flat fiber-end face over the emitting region of the light source. If the fiber end face is directly placed over the source emitting region, it is referred to as butt coupling. If the source's output light pattern is larger than the fiber's acceptance pattern, source-to-fiber coupling efficiency may be improved by placing a small lens between the source and fiber. Lensing schemes improve coupling efficiency when coupling both LEDs and LDs to optical fibers.

Q5. Semiconductor LEDs emit incoherent light. Define incoherent light.

Q6. Which semiconductor sources (LD or LED) emit more focused light and are capable of launching sufficient optical power into both single mode and multimode fibers?

Q7. The amount of optical power coupled into an optical fiber depends on what four factors?

SEMICONDUCTOR MATERIAL AND DEVICE OPERATING PRINCIPLES

Understanding optical emission in semiconductor lasers and LEDs requires knowledge of semiconductor material and device properties. Providing a complete description of semiconductor properties is beyond the scope of this introductory manual. In this chapter we only discuss the general properties of semiconductor LEDs and LDs.

Semiconductor sources are diodes, with all of the characteristics typical of diodes. However, their construction includes a special layer, called the active layer, which emits photons (light particles) when a current passes through the layer. The particular properties of the semiconductor are determined by the materials used and the layering of the materials within the semiconductor. Silicon (Si) and gallium arsenide (GaAs) are the two most common semiconductor materials used in electronic and electro-optic devices. In some cases other elements, such as aluminum (Al), indium (In) and phosphorus (P), are added to the base semiconductor material to modify the semiconductor properties. These elements are called dopants.

Current flowing through a semiconductor optical source causes it to produce light. An in-depth description of either of the two processes by which this occurs is beyond the scope of this module. However, we discuss elementary descriptions in the following paragraphs.

LEDs generally produce light through spontaneous emission when a current is passed through them. Spontaneous emission is the random generation of photons within the active layer of the LED. The emitted photons move in random directions. Only a certain percentage of the photons exit the semiconductor and are coupled into the fiber. Many of the photons are absorbed by the LED materials and the energy dissipated as heat. This process causes the light output from an LED to be incoherent, have a broad spectral width, and have a wide output pattern.

Laser diodes are much more complex than LEDs. Laser is an acronym for light amplification by the stimulated emission of radiation. Laser diodes produce light through stimulated emission when a current is passed through them. Stimulated emission describes how light is produced in any type of laser. In the laser diode, photons, initially produced by spontaneous emission interact with the laser material to produce additional photons. This process occurs within the active area of the diode called the laser cavity. The process does not affect the original photon. The stimulated photon has many of the same properties (wavelength, direction, phase) as the original photon.

As with the LED, not all of the photons produced are emitted from the laser diode. Some of the photons are absorbed and the energy dissipated as heat. The emission process and the physical characteristics of the diode cause the light output to be coherent, have a narrow spectral width, and have a narrow output pattern.

It is important to note that in both LED and laser diodes all of the electrical energy is not converted into optical energy. A substantial portion is converted to heat. Different LED and laser diode structures convert differing amounts of electrical energy into optical energy.

Q8. What are the two most common semiconductor materials used in electronic and electro-optic devices?

Q9. What is a laser?

Q10. .Describe stimulated emission.

LIGHT-EMITTING DIODES

A light-emitting diode (LED) is a semiconductor device that emits incoherent light, through spontaneous emission, when a current is passed through it. Typically LEDs for the 850-nm region are fabricated using GaAs and AlGaAs. LEDs for the 1300-nm and 1550-nm regions are fabricated using InGaAsP and InP.

The basic LED types used for fiber optic communication systems are the surface-emitting LED (SLED), the edge-emitting LED (ELED), and the superluminescent diode (SLD). LED performance differences help link designers decide which device is appropriate for the intended application. For short-distance (0 to 3 km), low-data-rate fiber optic systems, SLEDs and ELEDs are the preferred optical source. Typically, SLEDs operate efficiently for bit rates up to 250 megabits per second (Mb/s). Because SLEDs emit light over a wide area (wide far-field angle), they are almost exclusively used in multimode systems.

For medium-distance, medium-data-rate systems, ELEDs are preferred. ELEDs may be modulated at rates up to 400 Mb/s. ELEDs may be used for both single mode and multimode fiber systems. Both SLDs and ELEDs are used in long-distance, high-data-rate systems. SLDs are ELED-based diodes designed to operate in the superluminescence mode. A further discussion on superluminescence is provided later in this chapter. SLDs may be modulated at bit rates of over 400 Mb/s.

Surface-Emitting LEDs

The surface-emitting LED (shown in figure 6-1) is also known as the Burrus LED in honor of C. A. Burrus, its developer. In SLEDs, the size of the primary active region is limited to a small circular area of 20 μm to 50 μm in diameter. The active region is the portion of the LED where photons are emitted. The primary active region is below the surface of the semiconductor substrate perpendicular to the axis of the fiber. A well is etched into the substrate to allow direct coupling of the emitted light to the optical fiber. The etched well allows the optical fiber to come into close contact with the emitting surface. In addition, the epoxy resin that binds the optical fiber to the SLED reduces the refractive index mismatch, increasing coupling efficiency.

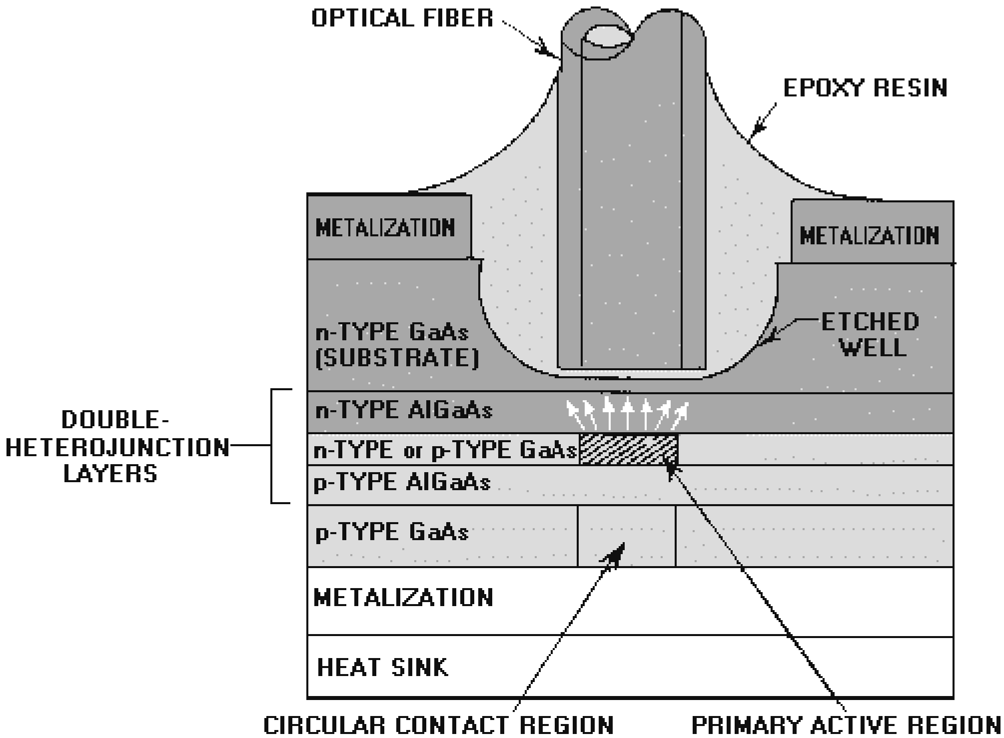


Figure 6-1.—Example of the SLED structure.

Edge-Emitting LEDs

The demand for optical sources for longer distance, higher bandwidth systems operating at longer wavelengths led to the development of edge-emitting LEDs. Figure 6-2 shows a typical ELED structure. It shows the different layers of semiconductor material used in the ELED. The primary active region of the ELED is a narrow stripe, which lies below the surface of the semiconductor substrate. The semiconductor substrate is cut or polished so that the stripe runs between the front and back of the device. The polished or cut surfaces at each end of the stripe are called facets.

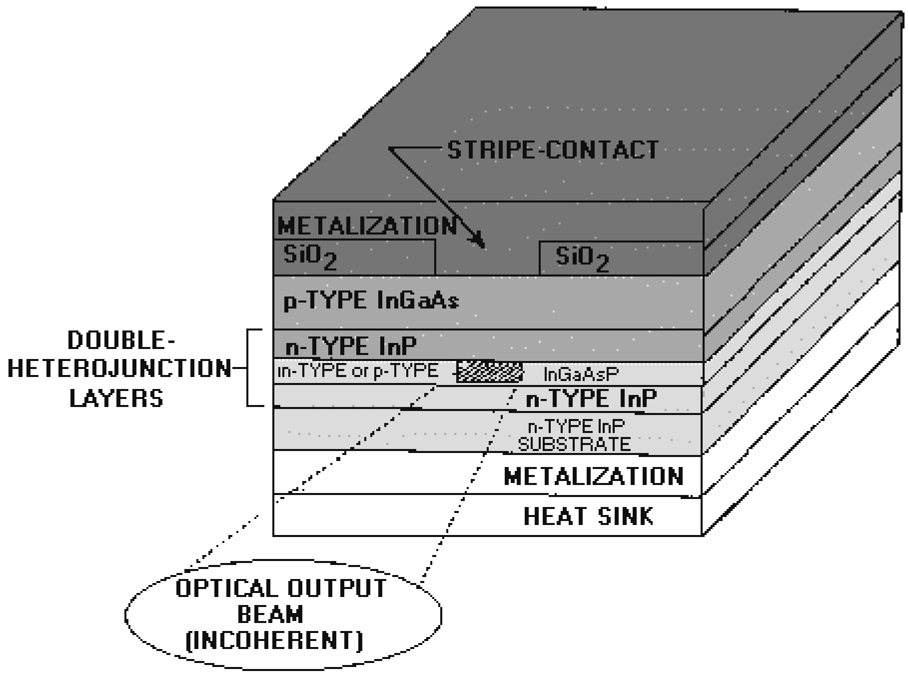


Figure 6-2.—Example of the ELED structure.

In an ELED the rear facet is highly reflective and the front facet is antireflection-coated. The rear facet reflects the light propagating toward the rear end-face back toward the front facet. By coating the front facet with antireflection material, the front facet reduces optical feedback and allows light emission. ELEDs emit light only through the front facet.

ELEDs emit light in a narrow emission angle allowing for better source-to-fiber coupling. They couple more power into small NA fibers than SLEDs. ELEDs can couple enough power into single mode fibers for some applications. ELEDs emit power over a narrower spectral range than SLEDs. However, ELEDs typically are more sensitive to temperature fluctuations than SLEDs.

LASER DIODES

A laser is a device that produces optical radiation by the process of stimulated emission. It is necessary to contain photons produced by stimulated emission within the laser active region. Figure 6-3 shows an optical cavity formed to contain the emitted photons by placing one reflecting mirror at each end of an amplifying medium. One mirror is made partially reflecting so that some radiation can escape from the cavity for coupling to an optical fiber.

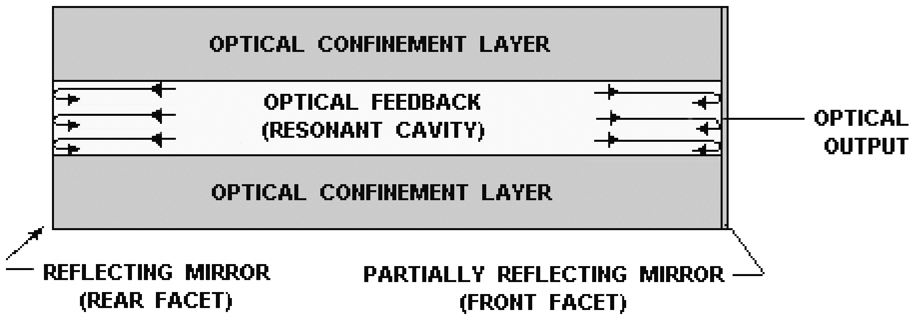


Figure 6-3.—Optical cavity for producing lasing.

Only a portion of the optical radiation is amplified. For a particular laser structure, there are only certain wavelengths that will be amplified by that laser. Amplification occurs when selected wavelengths, also called laser modes, reflect back and forth through the cavity. For lasing to occur, the optical gain of the selected modes must exceed the optical loss during one round-trip through the cavity. This process is referred to as optical feedback.

The lasing threshold is the lowest drive current level at which the output of the laser results primarily from stimulated emission rather than spontaneous emission. Figure 6-4 illustrates the transition from spontaneous emission to stimulated emission by plotting the relative optical output power and input drive current of a semiconductor laser diode. The lowest current at which stimulated emission exceeds spontaneous emission is the threshold current. Before the threshold current is reached, the optical output power increases only slightly with small increases in drive current. However, after the threshold current is reached, the optical output power increases significantly with small changes in drive currents.

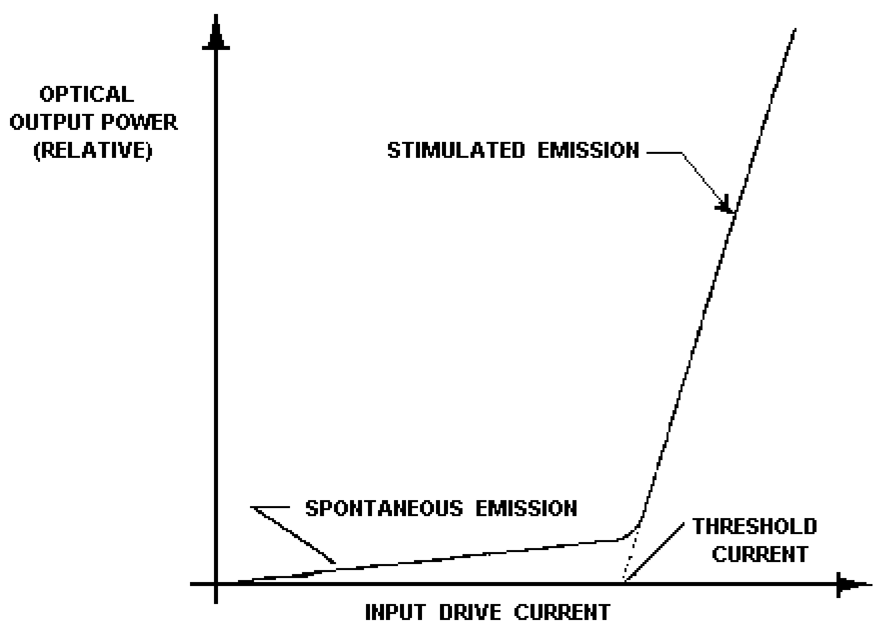


Figure 6-4.—The optical output power as a function of input drive current of a semiconductor laser diode.

Many types of materials including gas, liquid, and semiconductors can form the lasing medium. However, in this chapter we only discuss semiconductor laser diodes. Semiconductor laser diodes are the primary lasers used in fiber optics. A laser diode emits light that is highly monochromatic and very directional. This means that the LD's output has a narrow spectral width and small output beam angle.

A semiconductor LD's geometry is similar to an ELED with light-guiding regions surrounding the active region. Optical feedback is established by making the front facet partially reflective. This chapter provides no diagram detailing LD structures because they are similar to ELEDs in design. The rear facet is typically coated with a reflective layer so that all of the light striking the facet is reflected back into the active region. The front facet is typically left uncoated so that most of the light is emitted. By increasing the drive current, the diode becomes a laser.

At currents below the threshold current, LDs function as ELEDs. To optimize frequency response, laser diodes are often biased above this laser threshold. As a result, in an LD fiber optic system, light is modulated between a high power level and a lower power level, but never shut off. LDs typically can be modulated at frequencies up to over 2 gigahertz (GHz). Some lasers are capable of being modulated at frequencies over 20 GHz.

There are several important differences between LDs and LEDs. One is that LEDs usually lack reflective facets and in some cases are designed to suppress reflections back into the active region. Another is that lasers tend to operate at higher drive currents to produce light. A higher driver current results in more complicated drive circuits and more heat dissipation in the device.

LDs are also much more temperature sensitive than either SLEDs or ELEDs. Increases in the laser temperature significantly reduce laser output power. Increases in laser temperature beyond certain limits result in the loss of lasing. When lasers are used in many applications, the temperature of the laser must be controlled. Typically, electronic coolers, called thermo-electric (TE) coolers, are used to cool LDs in system applications.

SUPERLUMINESCENT DIODES

Superluminescence occurs when the spontaneous emissions of an ELED experience gain due to higher injected currents and reflections from facets. Superluminescent diodes (SLDs) are differentiated from both conventional LEDs and LDs. Although the output is not fully coherent, SLDs emit light that consists of amplified spontaneous emissions. The spectral width and beam angle of SLDs are narrower than that of conventional LEDs and wider than that of LDs.

An SLD is, in essence, a combination of a laser and an ELED. SLDs are similar in geometry to lasers but have no built-in optical feedback mechanism required by laser diodes for stimulated emission to achieve lasing. SLDs have structural features similar to those of ELEDs that suppress the lasing action by reducing the reflectivity of the facets. SLDs are essentially highly optimized ELEDs.

While SLDs operate like ELEDs at low current levels, their output power increases superlinearly and the spectral width narrows at high currents. Optical gain resulting from the higher injection currents causes the superlinear power increase and narrowing of the spectral width.

The advantages of SLDs over conventional LEDs include higher coupled power, narrower spectral width, and greater bandwidths. The disadvantages include nonlinear power-current characteristics, higher temperature sensitivity, and lower reliability.

Q11. What are the three basic LED types?

Q12. Which types of LEDs are the preferred optical sources for short-distance, low-data-rate fiber optic systems?

Q13. What are facets?

Q14. What is lowest current at which stimulated emission exceeds spontaneous emission in a semiconductor laser called?

Q15. Describe the output of a laser diode.

Q16. Which type of optical source usually lacks reflective facets and in some cases are designed to suppress reflections back into the active region?

Q17. Which type of optical source tends to operate at higher drive currents to produce light?

Q18. Are the effects of temperature changes on LDs more or less significant than for LEDs?

Q19. Specify the mechanism that SLDs lack that is required by laser diodes to achieve lasing.

FIBER OPTIC TRANSMITTERS

As stated previously, a fiber optic transmitter is a hybrid electro-optic device. It converts electrical signals into optical signals and launches the optical signals into an optical fiber. A fiber optic transmitter consists of an interface circuit, a source drive circuit, and an optical source. The interface circuit accepts the incoming electrical signal and processes it to make it compatible with the source drive circuit. The source drive circuit intensity modulates the optical source by varying the current through it. The optical signal is coupled into an optical fiber through the transmitter output interface.

Although semiconductor LEDs and LDs have many similarities, unique transmitter designs result from differences between LED and LD sources. Transmitter designs compensate for differences in optical output power, response time, linearity, and thermal behavior between LEDs and LDs to ensure proper system operation. Nonlinearities caused by junction heating in LEDs and mode instabilities in LDs necessitate the use of linearizing circuits within the transmitter in some cases.

Fiber optic transmitters using LDs require more complex circuitry than transmitters using LEDs. The basic requirement for digital systems is for drive circuitry to switch the optical output on and off at high speeds in response to logic voltage levels at the input of the source drive circuit. Because LDs are threshold devices, LDs are supplied with a bias just below the threshold in the off state. This bias is often referred to as prebias. One reason for prebiasing the LD is to reduce the turn-on delay in digital systems.

Most LD transmitters contain output power control circuitry to compensate for temperature sensitivity. This circuitry maintains the LD output at a constant average value by adjusting the bias current of the laser. In most cases LED transmitters do not contain output power control circuitry. LD and LED transmitters may also contain cooling devices to maintain the source at a relatively constant temperature. Most LD transmitters either have an internal thermo electric cooler or require a relatively controlled external temperature. Because LDs require more complex circuitry than LEDs, fiber optic transmitters using LDs are more expensive. For more information concerning fiber optic transmitters and their drive circuitry, refer to the reference material listed in appendix 2.

Transmitter output interfaces generally fall into two categories: optical connectors and optical fiber pigtails. Optical pigtails are attached to the transmitter optical source. This pigtail is generally routed out of the transmitter package as a coated fiber in a loose buffer tube or a single fiber cable. The pigtail is either soldered or epoxied to the transmitter package to provide fiber strain relief. The buffer tube or single fiber cable is also attached to the transmitter package to provide additional strain relief.

The transmitter output interface may consist of a fiber optical connector. The optical source may couple to the output optical connector through an intermediate optical fiber. One end of the optical fiber is attached to the source. The other end terminates in the transmitter optical output connector. The optical source may also couple to the output optical connector without an intermediate optical fiber. The optical source is placed within the transmitter package to launch power directly into the fiber of the mating optical connector. In some cases lenses are used to more efficiently couple light from the source into the mating optical connector.

Q20. How does the source drive circuit intensity modulate the source?

Q21. What is a prebias?

Q22. Is the drive circuitry generally more complex for an LED or a laser diode? Why?

Q23. What are the two types of output interfaces for fiber optic transmitters?

FIBER OPTIC TRANSMITTER PACKAGES

Fiber optic transmitters come in various sizes and shapes. The least complex fiber optic transmitters are typically packaged in transistor outline (TO) cans or hybrid microcircuit modules in dual inline packages (DIPs). These simple transmitters may require separate circuitry in the system equipment to provide an acceptable input signal to the transmitter. More complex fiber optic transmitters are available that have some or all of the signal conditioning circuitry integrated into the package. These transmitters typically are packaged in hybrid microcircuit modules in either DIP or butterfly lead packages, circuit cards, or complete stand-alone fiber optic converters. Stand-alone fiber optic converters and circuit cards generally contain sources in either TO cans or one of the hybrid microcircuit packages. For commercial applications, the most popular transmitter packages are the TO can and the DIP hybrid microcircuit.

FIBER OPTIC TRANSMITTER APPLICATIONS

Fiber optic transmitters can be classified into two categories: digital and analog. Digital transmitters produce two discrete optical power levels. These levels are essentially on and off with the exception that some light is emitted in the off state by some transmitters. Analog transmitters continuously vary the output optical power level as a function of the input electrical signal.

Digital Applications

Different types of fiber optic transmitters are used for different digital applications. For each specific application, the link data rate, transmission length, and operating environment influence the source type, center wavelength, spectral width, and package type chosen.

For low-data-rate applications, fiber optic transmitters generally use LEDs operating in either the 850-nm or 1300-nm window as their source. For the lowest data rates (0 to 20 megabits per second (Mbps)), sources tend to operate in the 850-nm window. For moderate data rates (50 to 200 Mbps), sources tend to operate in the 1300-nm window. Laser sources are almost never used in low-data-rate applications. Laser sources are only used when extremely high transmitter output powers are required in the application. The packages found in low-data-rate applications include all of the package types discussed earlier.

For high-data-rate applications, most fiber optic transmitters use laser diodes as sources. The sources typically operate in either the 1300-nm or 1550-nm windows. Most high-data-rate applications use LDs as the optical source and operate in the 1300-nm region. Almost all 1550-nm systems use an LD as the optical source. 1550-nm transmitters are usually only used in the extremely long distance high-data-rate applications (undersea links, etc.). High-data-rate transmitters are generally hybrid microcircuit modules or complete circuit cards. Almost all high-data-rate transmitters contain power control circuitry. Depending upon the application, high-data-rate transmitters may contain TE coolers.

Q24. List five common fiber optic transmitter packages.

Q25. What type of source is typically used in low-data-rate digital applications?

Q26. Why would a laser diode be used in a low-data-rate digital application?

Q27. What type of source is generally used in high-data-rate digital applications?

Analog Applications

Different types of fiber optic transmitters are also used for different analog applications. For each specific application, analog signal type, transmission length, and operating environment influence the source type, center wavelength, spectral width, and package type chosen.

For low-frequency applications, analog fiber optic transmitters generally use LEDs operating in either the 850-nm or 1300-nm window. Typical low frequency applications are analog audio and single channel video systems. For these systems, sources tend to operate in the 850-nm window. For moderate frequency applications, sources tend to operate in the 1300-nm window. These types of systems include multi-channel analog audio and video systems as well as frequency modulated (FM) systems. Laser sources are almost never used in low- or moderate-frequency analog applications. The main reason for this is the added circuit complexity that laser sources require. Laser sources are only used if extremely high transmitter output powers are required in the application. Most low-frequency analog transmitters are hybrid microcircuit modules, circuit cards, or stand-alone boxes.

For high-frequency applications, analog fiber optic transmitters use laser diodes as sources. Typical high frequency applications are cable television trunk line and raw radar remoting applications. The LDs typically operate in either the 1300-nm or 1550-nm windows. 1550-nm transmitters are typically used in cable television trunk line applications. Other applications may use either 1300-nm or 1550-nm LDs. High frequency transmitters are predominately circuit cards, but some hybrid microcircuit modules are also used. All high frequency analog transmitters contain TE coolers as well as linearization and power control circuitry.

Q28. Why are LEDs preferred over laser diodes for low- and moderate-frequency analog applications?

SUMMARY

Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas that you have learned. You should have a thorough understanding of these principles before moving on to chapter 7.

A **FIBER OPTIC TRANSMITTER** is a hybrid electro-optic device. It converts electrical signals into optical signals and launches the optical signals into an optical fiber.

An **OPTICAL SOURCE** converts electrical energy (current) into optical energy (light).

The principal **LIGHT SOURCES** used in fiber optics are semiconductor light-emitting diodes (LEDs) and laser diodes (LDs).

SEMICONDUCTOR LD's emit coherent light. Light waves having a fixed-phase relationship are referred to as coherent light.

SEMICONDUCTOR LED'S emit incoherent light. Light waves that lack a fixed-phase relationship are referred to as incoherent light.

The **RELEVANT OPTICAL POWER** is the amount of optical power coupled into the fiber. It depends on the angle over which the light is emitted, the size of the source's light-emitting area relative to the fiber core size, the alignment of the source and fiber, and the coupling characteristics of the fiber (such as the NA and the refractive index profile).

SOURCE-TO-FIBER COUPLING EFFICIENCY is a measure of the relevant optical power.

SILICON (Si) and **GALLIUM ARSENIDE (GaAs)** are the two most common semiconductor materials used in electronic and electro-optic devices.

In a semiconductor device, **PHOTONS (LIGHT)** are emitted when current flows through the active area.

SPONTANEOUS EMISSION occurs when photons are emitted in a random manner. Spontaneous emission produces incoherent light.

STIMULATED EMISSION occurs when a photon interacts with the laser material to produce additional photons.

A **LIGHT-EMITTING DIODE (LED**) is a semiconductor device that emits incoherent light, through spontaneous emission, when a current is passed through it. The basic LED types used for fiber optic communication systems are the **SURFACE-EMITTING LED (SLED)**, the **EDGE-EMITTING LED (ELED)**, and **the SUPERLUMINESCENT DIODE (SLD)**.

In **SURFACE-EMITTING LED'S (SLEDs)**, the size of the primary active region is limited to a small circular area of 20 m to 50 m in diameter. The active region is the portion of the LED where photons are emitted. SLEDs usually emit more total power into the air gap at the fiber interface than an ELED, but they do not launch as much power into the fiber. SLEDS also tend to emit power over a wider spectral range than ELED.

EDGE-EMITTING LED'S (ELEDs) emit light in a narrow emission angle allowing for better source-to-fiber coupling. They couple more power into small NA fibers than SLEDs. The polished or cut surfaces at each end of the ELED active stripe are called FACETS.

SUPERLUMINESCENCE occurs when the spontaneous emissions of an ELED experience gain due to higher injected currents and reflections from facets.

SUPERLUMINESCENT DIODES (SLDs) are similar in geometry to lasers but have no built-in optical feedback mechanism required by laser diodes for stimulated emission to achieve lasing. Although the output is not fully coherent, superluminescent diodes (SLDs) emit light that consists of amplified spontaneous emissions. The spectral width and beam angle of SLDs are narrower than that of conventional LEDs and wider than that of LDs.

The **ADVANTAGES** of **SLDs** over conventional LEDs include higher coupled power, narrower spectral width, and greater bandwidths. The **DISADVANTAGES** include nonlinear power-current characteristics, higher temperature sensitivity, and lower reliability.

A **LASER** is a device that produces optical radiation using stimulated emission rather than spontaneous emission. Laser is an acronym for light amplification by the stimulated emission of radiation.

The **LASING THRESHOLD** is the lowest drive level at which the output of the laser results primarily from stimulated emission rather than spontaneous emission.

The **THRESHOLD CURRENT** is the lowest current at which stimulated emission exceeds spontaneous emission.

A **LASER DIODE** is a semiconductor diode that emits coherent light by lasing. The LD's output has a narrow spectral width and small output beam angle.

TRANSMITTER OUTPUT INTERFACES fall into two categories: optical connectors and optical fiber pigtails.

FIBER OPTIC TRANSMITTERS using LDs require more complex circuitry than transmitters using LEDs.

Because **LDs** are threshold devices, LDs are supplied with a bias just below the threshold in the off state. This bias is often referred to as a prebias.

The least complex **FIBER OPTIC TRANSMITTERS** are typically packaged in transistor outline (TO) cans or hybrid microcircuit modules in dual inline packages (DIPs).

More complex **FIBER OPTIC TRANSMITTERS** typically are packaged in hybrid microcircuit modules in either DIP or butterfly lead packages, circuit cards, or complete stand-alone fiber optic converters.

FIBER OPTIC TRANSMITTERS can be classified into two categories: digital and analog.

DIGITAL TRANSMITTERS modulate the fiber optic source between two discrete optical power levels. These levels are essentially on and off with the exception that some light is emitted in the off state by some transmitters.

ANALOG TRANSMITTERS continuously vary the output optical power level as a function of the input electrical signal.

For **LOW-DATA-RATE APPLICATIONS** (0 to 20 Mbps), fiber optic transmitters generally use LEDs operating in either the 850-nm or 1300-nm window.

For **MODERATE-DATA-RATE APPLICATIONS** (50 to 200 Mbps), fiber optic transmitters generally use LEDs operating in the 1300-nm window.

For **HIGH-DATA-RATE APPLICATIONS**, most fiber optic transmitters use laser diodes as sources.

LASER SOURCES are almost never used in low- or moderate-frequency analog applications because LED sources require much less complex circuitry.

ANSWERS TO QUESTIONS Q1. THROUGH Q28.

A1. *Interface circuit, source drive circuit, and an optical source.*

A2. *The optical source.*

A3. *Multimode fiber dispersion, the relatively high fiber attenuation, and the LED's relatively low optical output power.*

A4. *Longer distances and higher bandwidths are possible because fiber material losses and dispersion are significantly reduced at the 1300-nm region.*

A5. *Light waves that lack a fixed-phase relationship.*

A6. *LDs.*

A7. *(1) The angles over which the light is emitted. (2) The size of the source's light-emitting area relative to the fiber core size. (3) The alignment of the source and fiber. (4) The coupling characteristics of the fiber (such as the NA and the refractive index profile).*

A8. *Silicon and gallium arsenide.*

A9. *A laser is a device that produces optical radiation using stimulated emission rather than spontaneous emission.*

A10. *A photon initially produced by spontaneous emission in the active region interacts with the laser material to produce additional photons.*

A11. *Surface-emitting LEDs (SLEDs), edge-emitting LEDs (ELEDs), and superluminescent diodes (SLDs).*

A12. *SLEDs and ELEDs.*

A13. *Cut or polished surfaces at each end of the narrow active region of an ELED.*

A14. *Threshold current.*

A15. *The LD's output has a narrow spectral width and small output beam angle.*

A16. *LED.*

A17. *Laser.*

A18. *More.*

A19. *SLDs have no built-in optical feedback mechanism.*

A20. *By varying the current through the source.*

A21. *A current applied in the laser off state just less than the threshold current.*

A22. *For a laser diode. The laser diode transmitter generally contains output power control circuitry and may contain a TE cooler and some circuitry associated with the TE cooler.*

A23. *Optical connectors and optical fiber pigtails.*

A24. TO can, DIP, butterfly lead microcircuits, circuit cards, and stand-alone optical fiber converters.

A25. LED.

A26. When extremely high transmitter output powers are required.

A27. Laser diode.

A28. LEDs require less complex circuitry than lasers.

CHAPTER 7

OPTICAL DETECTORS AND FIBER OPTIC RECEIVERS

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to do the following:

1. Explain the principal properties of an optical detector and fiber optic receiver.
2. Detail semiconductor optical detector performance and capability requirements necessary for the successful implementation of fiber optic systems.
3. List the main components of a fiber optic receiver.
4. Discuss receiver sensitivity, dynamic range, and other key operational parameters used to define receiver performance.

INTRODUCTION TO OPTICAL DETECTORS AND FIBER OPTIC RECEIVERS

Chapter 6 taught you that a fiber optic transmitter is an electro-optic device capable of accepting electrical signals, converting them into optical signals, and launching the optical signals into an optical fiber. The optical signals propagating in the fiber become weakened and distorted because of scattering, absorption, and dispersion. The fiber optic device responsible for converting the weakened and distorted optical signal back to an electrical signal is a fiber optic receiver.

A **fiber optic receiver** is an electro-optic device that accepts optical signals from an optical fiber and converts them into electrical signals. A typical fiber optic receiver consists of an optical detector, a low-noise amplifier, and other circuitry used to produce the output electrical signal (see figure 7-1). The optical detector converts the incoming optical signal into an electrical signal. The amplifier then amplifies the electrical signal to a level suitable for further signal processing. The type of other circuitry contained within the receiver depends on what type of modulation is used and the receiver electrical output requirements.

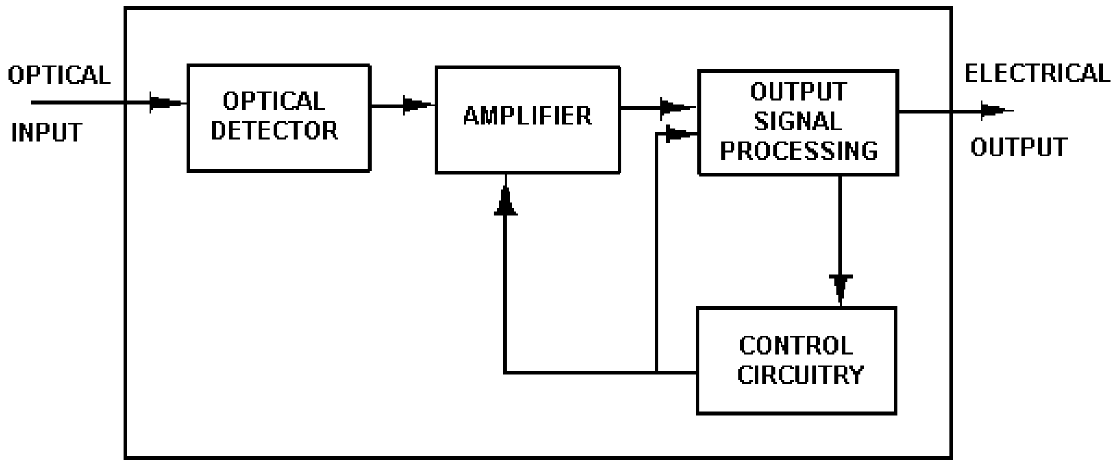


Figure 7-1.—Block diagram of a typical fiber optic receiver.

Receiver spectral response, sensitivity, frequency response, and dynamic range are key receiver performance parameters that can affect overall system operation. The choice of optical detector materials and structures determines the spectral response. Silicon (Si), gallium arsenide (GaAs), and gallium aluminum arsenide (GaAlAs) are typical detector materials used for receiver operation in the 850-nm wavelength region. germanium (Ge), indium phosphide (InP), and indium gallium arsenide (InGaAs) are examples of detector materials used for receiver operation in the 1300-nm and 1550-nm wavelength regions.

The **receiver sensitivity** is the minimum amount of optical power required to achieve a specific receiver performance. For digital transmission at a given data rate and coding, this performance is described by a maximum bit-error rate (BER). In analog systems, for a given modulation and bandwidth, it is described by a minimum signal-to-noise ratio (SNR). **Dynamic range** refers to the range of optical power levels over which the receiver operates within the specified values. It usually is described by the ratio of the maximum input power to the sensitivity. Before discussing receiver sensitivity, bandwidth, dynamic range, and frequency response in more detail, we discuss the main types of optical detectors used in fiber optics.

Q1. What is a fiber optic receiver?

Q2. Which part of the receiver amplifies the electrical signal to a level suitable for further signal processing?

Q3. Which performance parameter is the minimum amount of optical power required to achieve a specific bit-error rate (BER) in digital systems or a given signal-to-noise ratio (SNR) in analog systems?

Q4. Define receiver dynamic range.

OPTICAL DETECTORS

A **transducer** is a device that converts input energy of one form into output energy of another. An **optical detector** is a transducer that converts an optical signal into an electrical signal. It does this by generating an electrical current proportional to the intensity of incident optical radiation. The relationship between the input optical radiation and the output electrical current is given by the detector responsivity. Responsivity is discussed later in this chapter.

OPTICAL DETECTOR PROPERTIES

Fiber optic communications systems require that optical detectors meet specific performance and compatibility requirements. Many of the requirements are similar to those of an optical source. Fiber optic systems require that optical detectors:

- Be compatible in size to low-loss optical fibers to allow for efficient coupling and easy packaging.
- Have a high sensitivity at the operating wavelength of the optical source.
- Have a sufficiently short response time (sufficiently wide bandwidth) to handle the system's data rate.
- Contribute low amounts of noise to the system.
- Maintain stable operation in changing environmental conditions, such as temperature.

Optical detectors that meet many of these requirements and are suitable for fiber optic systems are semiconductor photodiodes. The principal optical detectors used in fiber optic systems include semiconductor positive-intrinsic-negative (PIN) photodiodes and avalanche photodiodes (APDs).

Q5. Describe the operation of an optical detector.

Q6. For efficient operation, should a detector have a high or low responsivity at the operating wavelength?

Q7. List the two principal optical detectors used in fiber optic systems.

SEMICONDUCTOR PHOTODIODES

Semiconductor photodiodes generate a current when they absorb photons (light). The amount of current generated depends on the following factors:

- The wavelengths of the incident light and the responsivity of the photodiode at those wavelengths
- The size of the photodiode active area relative to the fiber core size
- The alignment of the fiber and the photodiode

The optical fiber is coupled to semiconductor photodiodes similarly to the way optical sources are coupled to optical fibers. Fiber-to-photodiode coupling involves centering the flat fiber-end face over the photodiode active area. This is normally done directly by butt coupling the fiber up to the photodiode surface. As long as the photodiode active area is larger than that of the fiber core, fiber-to-detector coupling losses are very low. In some cases a lens may be used to couple the fiber end-face to the detector. However, this is not typically done.

SEMICONDUCTOR MATERIAL AND DEVICE PROPERTIES

The mechanism by which optical detectors convert optical power into electrical current requires knowledge of semiconductor material and device properties. As stated in chapter 6, providing a complete description of these properties is beyond the scope of this manual. In this chapter we only discuss the general properties of semiconductor PINs and APDs.

Semiconductor detectors are designed so that optical energy (photons) incident on the detector active area produces a current. This current is called a **photocurrent**. The particular properties of the semiconductor are determined by the materials used and the layering of the materials within the device. Silicon (Si), gallium arsenide (GaAs), germanium (Ge), and indium phosphide (InP) are the most common semiconductor materials used in optical detectors. In some cases aluminum (Al) and indium (In) are used as dopants in the base semiconductor material.

Responsivity

Responsivity is the ratio of the optical detector's output photocurrent in amperes to the incident optical power in watts. The responsivity of a detector is a function of the wavelength of the incident light and the efficiency of the device in responding to that wavelength. For a particular material, only photons of certain wavelengths will generate a photocurrent when they are absorbed. Additionally, the detector material absorbs some wavelengths better than others. These two properties cause the wavelength dependence in the detector responsivity. Responsivity is a useful parameter for characterizing detector performance because it relates the photocurrent generated to the incident optical power.

Q8. What are the four most common materials used in semiconductor detector fabrication?

Q9. What is a photocurrent?

Q10. Define responsivity.

PIN PHOTODIODES

A **PIN photodiode** is a semiconductor positive-negative (p-n) structure with an intrinsic region sandwiched between the other two regions (see figure 7-2). It is normally operated by applying a reverse-bias voltage. The magnitude of the reverse-bias voltage depends on the photodiode application, but typically is less than a few volts. When no light is incident on the photodiode, a current is still produced. This current is called the **dark current**. The dark current is the leakage current that flows when a reverse bias is applied and no light is incident on the photodiode. Dark current is dependent on temperature. While dark current may initially be low, it will increase as the device temperature increases.

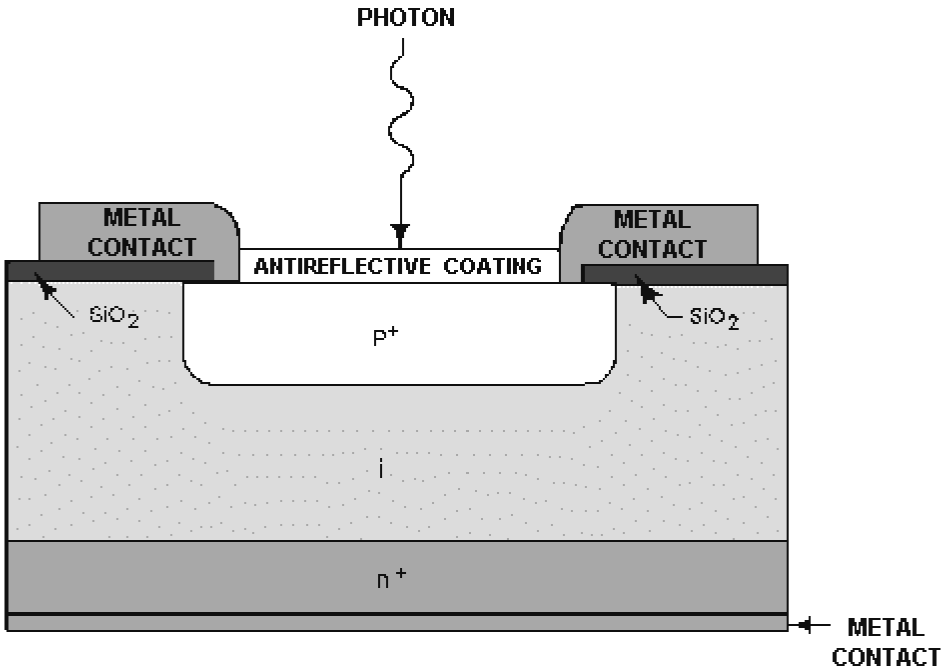


Figure 7-2.—The basic structure of a PIN photodiode.

Q11. How are PIN photodiodes usually biased?

Q12. What is the dark current?

Q13. Will dark current increase or decrease as the temperature of the photodiode increases?

Response Time

There are several factors that influence the response time of a photodiode and its output circuitry (see figure 7-3). The most important of these are the thickness of the detector active area and the detector RC time constant. The detector thickness is related to the amount of time required for the electrons generated to flow out of the detector active area. This time is referred to as the electron **transit time**. The thicker the detector active area, the longer the transit time will be.

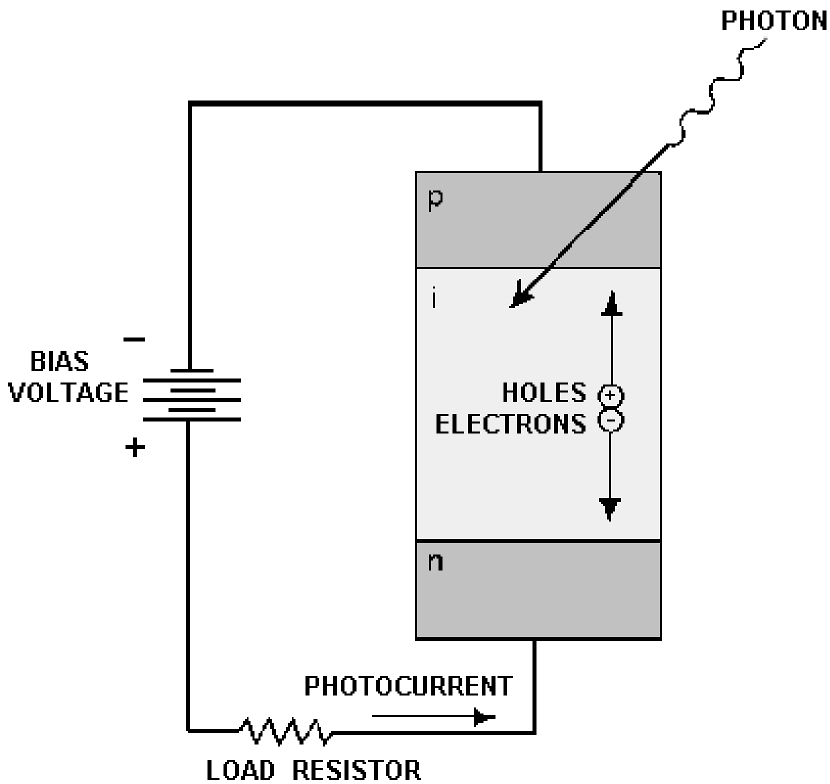


Figure 7-3.—A schematic representation of a photodiode.

The **capacitance** (**C**) of the photodiode and the **resistance** (**R**) of the load form the RC time constant. The capacitance of the photodetector must be kept small to prevent the RC time constant from limiting the response time. The photodiode capacitance consists mainly of the junction capacitance and any capacitance relating to packaging. The **RC time constant** is given by $t_{RC} = RC$.

Trade-offs between fast transit times and low capacitance are necessary for high-speed response. However, any change in photodiode parameters to optimize the transit time and capacitance can also affect responsivity, dark current, and coupling efficiency. A fast transit time requires a thin detector active area, while low capacitance and high responsivity require a thick active region. The diameter of the detector active area can also be minimized. This reduces the detector dark current and minimizes junction capacitance. However, a minimum limit on this active area exists to provide for efficient fiber-to-detector coupling.

Q14. Should the capacitance of the photodetector be kept small or large to prevent the RC time constant from limiting the response time?

Q15. Trade-offs between competing effects are necessary for high speed response. Which competing effect (fast transit time, low capacitance, or high quantum efficiency) requires a thin active area?

Linearity

Reverse-biased photodetectors are highly linear devices. Detector **linearity** means that the output electrical current (photocurrent) of the photodiode is linearly proportional to the input optical power. Reverse-biased photodetectors remain linear over an extended range (6 decades or more) of photocurrent before saturation occurs. Output saturation occurs at input optical power levels typically greater than 1 milliwatt (mW). Because fiber optic communications systems operate at low optical power levels, detector saturation is generally not a problem.

Q16. Why is detector saturation not generally a problem in fiber optic communications systems?

AVALANCHE PHOTODIODES

An **avalanche photodiode (APD)** is a photodiode that internally amplifies the photocurrent by an avalanche process. Figure 7-4 shows an example APD structure. In APDs, a large reverse-bias voltage, typically over 100 volts, is applied across the active region. This voltage causes the electrons initially generated by the incident photons to accelerate as they move through the APD active region. As these electrons collide with other electrons in the semiconductor material, they cause a fraction of them to become part of the photocurrent. This process is known as **avalanche multiplication**. Avalanche multiplication continues to occur until the electrons move out of the active area of the APD.

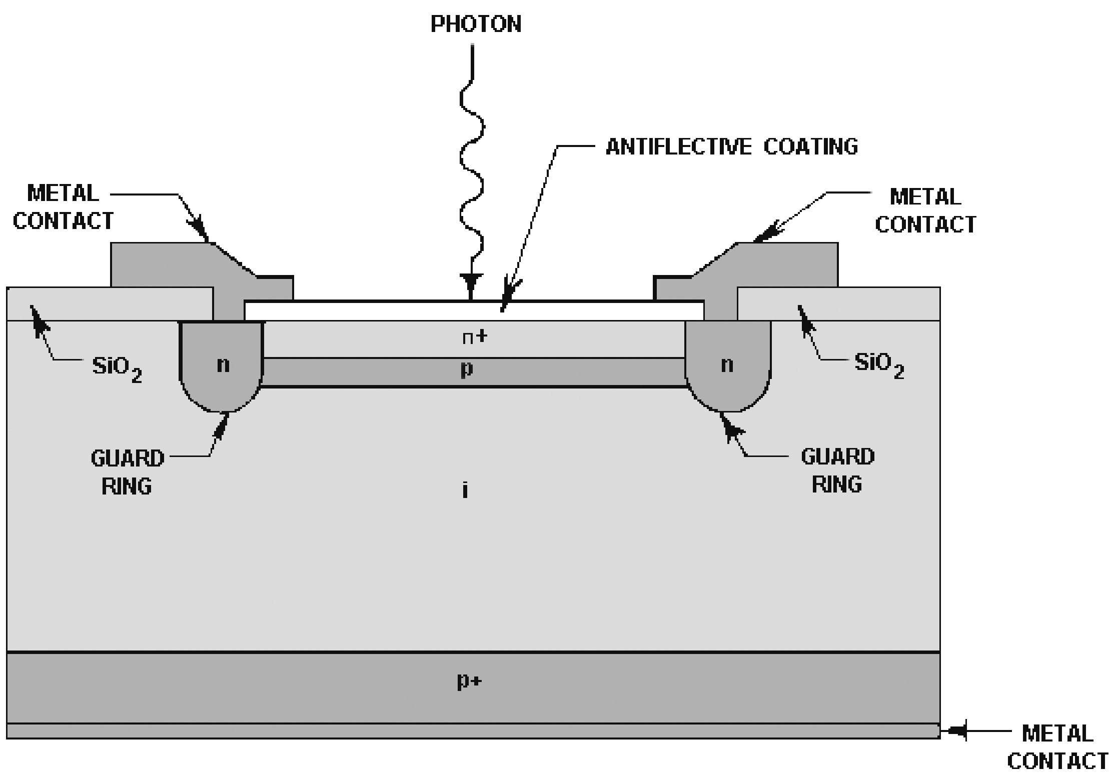


Figure 7-4.—The basic structure of an APD.

The gain of the APD can be changed by changing the reverse-bias voltage. A larger reverse-bias voltage results in a larger gain. However, a larger reverse-bias voltage also results in increased noise levels. Excess noise resulting from the avalanche multiplication process places a limit on the useful gain of the APD. The avalanche process introduces excess noise because every photogenerated carrier does not undergo the same multiplication. The noise properties of an APD are affected by the materials that the

APD is made of. Typical semiconductor materials used in the construction of low-noise APDs include silicon (Si), indium gallium arsenide (InGaAs), and germanium (Ge).

Trade-offs are made in APD design to optimize responsivity and gain, dark current, response time, and linearity. This chapter does not attempt to discuss trade-offs in APD design in more detail. Many aspects of the discussion provided on responsivity, dark current, and response time provided in the PIN photodiodes section also relate to APDs. The response time of an APD and its output circuitry depends on the same factors as PIN photodiodes. The only additional factor affecting the response time of an APD is the additional time required to complete the process of avalanche multiplication. To learn more about APD design trade-offs and performance parameters, refer to the reference material listed in appendix 2.

Q17. Describe avalanche multiplication.

Q18. How can the gain of an APD be increased?

FIBER OPTIC RECEIVERS

In fiber optic communications systems, optical signals that reach fiber optic receivers are generally attenuated and distorted (see figure 7-5). The fiber optic receiver must convert the input and amplify the resulting electrical signal without distorting it to a point that other circuitry cannot use it.

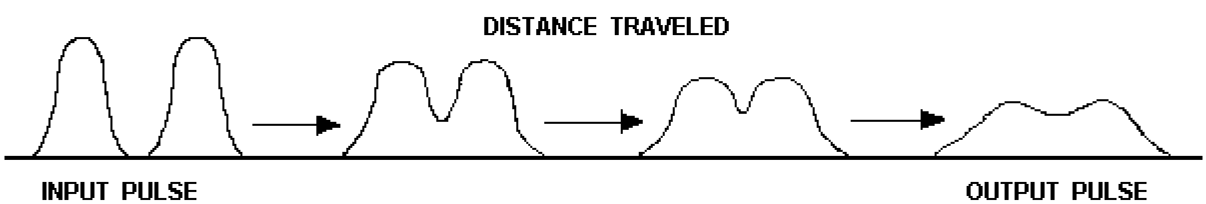


Figure 7-5.—Attenuated and distorted optical signals.

As stated previously, a fiber optic receiver consists of an optical detector, an amplifier, and other circuitry. In most fiber optic systems, the optical detector is a PIN photodiode or APD. Receiver performance varies depending on the type of detector used. The amplifier is generally described as having two stages: the preamplifier and the postamplifier. The **preamplifier** is defined as the first stage of amplification following the optical detector. The **postamplifier** is defined as the remaining stages of amplification required to raise the detector's electrical signal to a level suitable for further signal processing. The preamplifier is the dominant contributor of electrical noise in the receiver. Because of this, its design has a significant influence in determining the sensitivity of the receiver.

The output circuitry processes the amplified signal into a form suitable for the interfacing circuitry. For digital receivers, this circuitry may include low-pass filters and comparators. For analog receivers, this circuitry may also include low-pass filters.

Receiver sensitivity, bandwidth, and dynamic range are key operational parameters used to define receiver performance. One goal in designing fiber optic receivers is to optimize receiver sensitivity. To increase sensitivity, receiver noise resulting from signal-dependent shot noise and thermal noise must be kept at a minimum. A more detailed discussion of receiver shot and thermal noise is provided later in this chapter.

In addition to optimizing sensitivity, optical receiver design goals also include optimizing the bandwidth and the dynamic range. A receiver that has the ability to operate over a wide range of optical power levels can operate efficiently in both short- and long-distance applications. Because conflicts arise when attempting to meet each goal, trade-offs in receiver designs are made to optimize overall performance.

Q19. Which amplifier stage (the preamplifier or the postamplifier) is a dominant contributor of noise and significantly influences the sensitivity of the receiver?

Q20. List the key operational parameters used to define receiver performance.

RECEIVER NOISE

Noise corrupts the transmitted signal in a fiber optic system. This means that noise sets a lower limit on the amount of optical power required for proper receiver operation. There are many sources of noise in fiber optic systems. They include the following:

- Noise from the light source
- Noise from the interaction of light with the optical fiber
- Noise from the receiver itself

Because the intent of this chapter is to discuss optical detector and receiver properties, only noise associated with the photodetection process is discussed. **Receiver noise** includes thermal noise, dark current noise, and quantum noise. Noise is the main factor that limits receiver sensitivity.

Noise introduced by the receiver is either signal dependent or signal independent. Signal dependent noise results from the random generation of electrons by the incident optical power. Signal independent noise is independent of the incident optical power level.

Thermal noise is the noise resulting from the random motion of electrons in a conducting medium. Thermal noise arises from both the photodetector and the load resistor. Amplifier noise also contributes to thermal noise. A reduction in thermal noise is possible by increasing the value of the load resistor. However, increasing the value of the load resistor to reduce thermal noise reduces the receiver bandwidth. In APDs, the thermal noise is unaffected by the internal carrier multiplication.

Shot noise is noise caused by current fluctuations because of the discrete nature of charge carriers. Dark current and quantum noises are two types of noise that manifest themselves as shot noise. **Dark current noise** results from dark current that continues to flow in the photodiode when there is no incident light. Dark current noise is independent of the optical signal. In addition, the discrete nature of the photodetection process creates a signal dependent shot noise called quantum noise. **Quantum noise** results from the random generation of electrons by the incident optical radiation. In APDs, the random nature of the avalanche process introduces an additional shot noise called excess noise. For further information on the excess noise resulting from the avalanche process, refer to the avalanche photodiode section.

Q21. List the main types of receiver noise.

Q22. What is the main factor that determines receiver sensitivity?

Q23. For a reduction in thermal noise, should the value of the detector's load resistor be increased or decreased?

Q24. What are two types of noise that manifest themselves as shot noise?

RECEIVER DESIGN

The simplest fiber optic receivers consist of only the optical detector and a load resistor. However, the output signal of these simple receivers is not in a suitable form for most types of interfacing circuitry. To produce a suitable signal, a preamplifier, a post amplifier, and other circuitry are generally included in the receiver.

The choice of an optical detector and the design of the preamplifier help determine the operational characteristics of the receiver. Fiber optic receivers using APDs have greater sensitivity than those using PIN photodiodes. In addition, trade-offs are made in preamplifier designs to increase sensitivity while optimizing bandwidth and dynamic range. The two basic types of amplifiers used in fiber optic receivers are the **high-impedance amplifier** and the **transimpedance amplifier**.

The high-impedance preamplifier is generally used with a large load resistor to improve sensitivity. The large load resistor is used to reduce thermal noise. Although the high-impedance preamplifier achieves high sensitivity, receiver bandwidth and dynamic range are limited. The transimpedance preamplifier uses a low-noise, high-input impedance amplifier with negative feedback. This design provides improvements in bandwidth and dynamic range with some degradation in sensitivity from an increase in noise. For more information on receiver performance and design, refer to the reference material listed in appendix 2.

Q25. What are the two basic types of preamplifiers used in fiber optic receivers?

Q26. Which preamplifier design (high-impedance or transimpedance) provides improvements in bandwidth and greater dynamic range with some degradation in sensitivity from an increase in noise?

FIBER OPTIC RECEIVER PACKAGES

Fiber optic receivers come in packages similar to those for fiber optic transmitters. For information on fiber optic receiver packages, refer back to the fiber optic transmitter packages section of chapter 6.

FIBER OPTIC RECEIVER APPLICATIONS

Fiber optic receivers can be classified into two categories: **digital** and **analog**. Digital receivers detect the input optical signal, amplify the digital photocurrent, and reshape the signal to produce an undistorted output electrical signal. Analog receivers detect the input optical signal and amplify the generated photocurrent.

Digital Applications

For most digital applications the designs of the digital fiber optic receivers are similar. For **low-data-rate** applications, PIN diodes and high impedance amplifiers are generally used. Receiver sensitivities are maximized by using large load resistors in the photodiode circuit. For **moderate-data-rate** applications, PIN diodes and either high impedance amplifiers with smaller load resistances or transimpedance amplifiers are used. For **high-data-rate** applications, PINs or APDs are used with transimpedance amplifiers. APDs are rarely used in low- or moderate-data-rate applications unless receivers with extremely low sensitivities are required.

For each digital application, the receiver will generally contain a low-pass filter. The pass-band of the filter depends on the data rate of the application. The filter is used to smooth the amplified signal to remove some of the high frequency noise before the signal is further processed. The digital receiver generally contains a comparator, which reshapes the amplified electrical signal to remove any distortions introduced in the transmission process. In some cases the receiver may also contain clock recovery circuitry, which retimes the output electrical signal as well.

Q27. For what types of applications are APDs generally used?

Q28. Why is a low-pass filter generally part of a digital fiber optic receiver?

Analog Applications

Analog receivers are similar in design to digital receivers with the exception that digital signal restoring circuitry is not used. The preamplifier and postamplifiers are designed to be more linear than those used in digital receivers in some cases.

For **low-frequency** applications, PIN diodes and high impedance amplifiers are generally used. For **moderate-frequency** applications, PIN diodes and either high impedance amplifiers or transimpedance amplifiers are used. For **high-frequency** applications, PINs or APDs are used with transimpedance amplifiers. As in digital applications, APDs are rarely used in low- or moderate-frequency applications unless receivers with extremely low sensitivities are required.

SUMMARY

Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas that you have learned. You should have a thorough understanding of these principles before moving on to chapter 8.

A **FIBER OPTIC RECEIVER** is an electro-optic device that accepts optical signals from an optical fiber and converts them into electrical signals. A typical fiber optic receiver consists of an optical detector, a low-noise amplifier, and other circuitry used to produce the output electrical signal.

RECEIVER SPECTRAL RESPONSE, SENSITIVITY, FREQUENCY RESPONSE, and **DYNAMIC RANGE** are key receiver performance parameters that can affect overall system operation.

RECEIVER SENSITIVITY is the minimum amount of optical power required to achieve a specific receiver performance. For digital transmission at a given data rate and coding, this performance is described by a maximum bit-error rate (BER). In analog systems, for a given modulation and bandwidth, it is described by a minimum signal-to-noise ratio (SNR).

DYNAMIC RANGE refers to the range of optical power levels over which the receiver operates within the specified values. It usually is described by the ratio of the maximum input power to the sensitivity.

A **TRANSDUCER** is a device that converts input energy of one form into output energy of another.

An **OPTICAL DETECTOR** is a transducer that converts an optical signal into an electrical signal. It does this by generating an electrical current proportional to the intensity of incident optical radiation.

The semiconductor **POSITIVE-INTRINSIC-NEGATIVE (PIN) PHOTODIODE** and **AVALANCHE PHOTODIODE (APD)** are the principal optical detectors used in fiber optic systems.

A **PHOTOCURRENT** is generated when photons are absorbed by a photodiode.

RESPONSIVITY is the ratio of the optical detector's output photocurrent in amperes to the incident optical power in watts.

DARK CURRENT, or reverse-leakage current, is the current that continues to flow in the photodetector when there is no incident light.

The **RESPONSE TIME** of a photodiode and its output circuitry depends primarily on the thickness of the detector active area and the detector RC time constant.

The **TRANSIT TIME** is the time it takes electrons to travel out of the detector active area.

The **RC TIME CONSTANT** is defined by the capacitance (C) of the photodiode and the resistance (R) of the load. The RC time constant is given by $t_{RC} = RC$.

A **HIGH-SPEED RESPONSE** requires short transit times and low capacitance. However, any change in photodiode parameters to optimize the transit time and capacitance can also affect quantum efficiency, dark current, and coupling efficiency.

Detector **LINEARITY** means that the output electrical current (photocurrent) of the photodiode is linearly proportional to the input optical power.

An **AVALANCHE PHOTODIODE (APD)** is a photodiode that internally amplifies the photocurrent by an avalanche process.

In **APDs**, a large **REVERSE-BIAS VOLTAGE**, typically over 100 volts, is applied across the active region.

AVALANCHE MULTIPLICATION occurs when accelerated electrons collide with other electrons in the semiconductor material, causing a fraction of them to become part of the photocurrent.

TRADE-OFFS are made in APD design to optimize responsivity and gain, dark current, response time, and linearity.

The **RESPONSE TIME** of APDs accounts for the avalanche build-up time in addition to transit time and RC time constant.

The **PREAMPLIFIER** is defined as the first stage of amplification following the optical detector.

The **POSTAMPLIFIER** is defined as the remaining stages of amplification required to raise the detectors electrical signal to a level suitable for further signal processing.

RECEIVER SENSITIVITY, BANDWIDTH, and **DYNAMIC RANGE** are key operational parameters used to define receiver performance.

NOISE is the main factor that determines receiver sensitivity.

RECEIVER NOISE includes thermal noise, dark current noise, and quantum noise.

THERMAL NOISE is the noise resulting from the random motion of electrons in a conducting medium.

SHOT NOISE is noise caused by current fluctuations due to the discrete nature of charge carriers.

DARK CURRENT NOISE results from dark current that continues to flow in the photodiode when there is no incident light.

QUANTUM NOISE results from the random generation of electrons by the incident optical radiation.

The **HIGH-IMPEDANCE AMPLIFIER** and the **TRANSIMPEDANCE AMPLIFIER** are the two basic types of amplifiers used in fiber optic receivers.

The **HIGH-IMPEDANCE PREAMPLIFIER** provides a high sensitivity, but limits receiver bandwidth and dynamic range.

The **TRANSIMPEDANCE PREAMPLIFIER** provides improvements in bandwidth and dynamic range with some degradation in sensitivity from an increase in noise.

PIN PHOTODIODES are used as the detector in most applications.

AVALANCHE PHOTODIODES are only used in high-speed applications and applications where detectors with extremely low sensitivities are required.

ANSWERS TO QUESTIONS Q1. THROUGH Q28.

A1. *An electro-optic device that accepts optical signals from an optical fiber and converts them into electrical signals.*

A2. *Amplifier.*

A3. *Receiver sensitivity.*

A4. *The range of optical power levels over which the receiver operates within the specified values. It usually is described by the ratio of the maximum input power to the sensitivity.*

A5. *It is a transducer that converts an optical signal into an electrical signal. It does this by generating an electrical current proportional to the intensity of incident optical radiation.*

A6. *High.*

A7. *The semiconductor positive-intrinsic-negative (PIN) photodiode and avalanche photodiode (APD).*

A8. *Silicon, gallium arsenide, germanium, and indium phosphide.*

A9. *The current produced when photons are incident on the detector active area.*

A10. *The ratio of the optical detector's output photocurrent in amperes to the incident optical power in watts.*

A11. *Reverse-biased.*

A12. *The leakage current that continues to flow through a photodetector when there is no incident light.*

A13. Increase.

A14. Small.

A15. Fast transit time.

A16. Because fiber optic communications systems operate at low optical power levels.

A17. The electrons initially generated by the incident photons accelerate as they move through the APD active region. As these electrons collide with electrons in the semiconductor material, they cause a fraction of them to become part of the photocurrent.

A18. By increasing the reverse-bias voltage.

A19. The preamplifier.

A20. Receiver sensitivity, bandwidth, and dynamic range.

A21. Thermal noise, dark current noise, and quantum noise.

A22. Noise.

A23. Increased.

A24. Dark current and quantum noises.

A25. The high-impedance amplifier and the transimpedance amplifier.

A26. Transimpedance.

A27. For high-data-rate applications and for low- or moderate-data-rate applications where receivers with extremely low sensitivities are required.

A28. To smooth the amplified signal to remove some of the high frequency noise before the signal is further processed.

CHAPTER 8

FIBER OPTIC LINKS

LEARNING OBJECTIVES

Upon completion of this chapter, you should be able to do the following:

1. Describe a basic point-to-point fiber optic data link.
2. Explain the difference between digital and analog fiber optic communications systems.
3. Discuss the most common types of line coding used in digital fiber optic communications including non-return-to-zero (NRZ), return-to-zero (RZ), and biphase (or Manchester).
4. Describe the main type of analog modulation.
5. State several precautions that need to be emphasized when installing fiber optic links on board ships.

FIBER OPTIC SYSTEM TOPOLOGY

Most of the discussion on fiber optic data links provided earlier in this training manual refers to simple point-to-point links. A **point-to-point** fiber optic data link consists of an optical transmitter, optical fiber, and an optical receiver. In addition, any splices or connectors used to join individual optical fiber sections to each other and to the transmitter and the receiver are included. Figure 8-1 provides a schematic diagram of a point-to-point fiber optic data link.

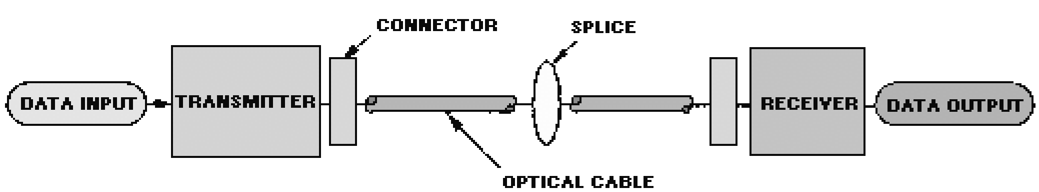


Figure 8-1.—A schematic diagram of a point-to-point fiber optic data link.

A common fiber optic application is the **full duplex link**. This link consists of two simple point-to-point links. The links transmit in opposite directions between the equipments. This application may be configured using only one fiber. If configured with one fiber, fiber optic splitters are used at each end to couple the transmit signal onto the fiber and receive signal to the detector.

All fiber optic systems are simply sets of point-to-point fiber optic links. Different system topologies arise from the different ways that point-to-point fiber optic links can be connected between equipments. The term **topology**, as used here, refers to the configuration of various equipments and the fiber optic components interconnecting them. This equipment may be computers, workstations, consoles, or other equipments. Point-to-point links are connected to produce systems with linear bus, ring, star, or tree topologies. Point-to-point fiber optic links are the basic building block of all fiber optic systems.

A **linear bus topology** consists of a single transmission line that is shared by a number of equipments (see figure 8-2). Generally the transmission line in a fiber optic linear bus consists of two optical lines, one for each direction of communication. Optical taps (optical splitters) are used by each equipment to connect to each line. For each line, the optical tap couples signals from the line to the equipment receiver and from the equipment transmitter onto the line. The connection between any two equipments is a simple point-to-point link that contains the optical tap for each equipment.

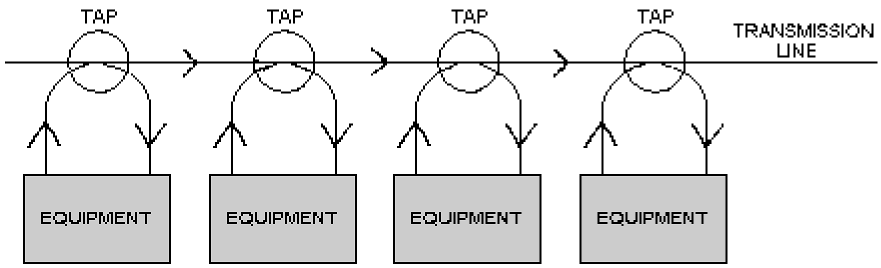


Figure 8-2.—Linear bus topology.

A **ring topology** consists of equipments attached to one another in a closed loop or ring (see figure 8-3). The connection between each equipment is a simple point-to-point link. In some systems each equipment may have an associated optical switch. In normal operation, the switch routes signals from the fiber connected to the previous equipment to the receiver. It also routes signals from the transmitter to the fiber connected to the next equipment. In bypass operation, the switch routes signals from the fiber connected to the previous equipment to the fiber connected to the next equipment. In each case, the connection between adjacent equipments on the ring is a simple point-to-point link through fiber, connectors, and switches.

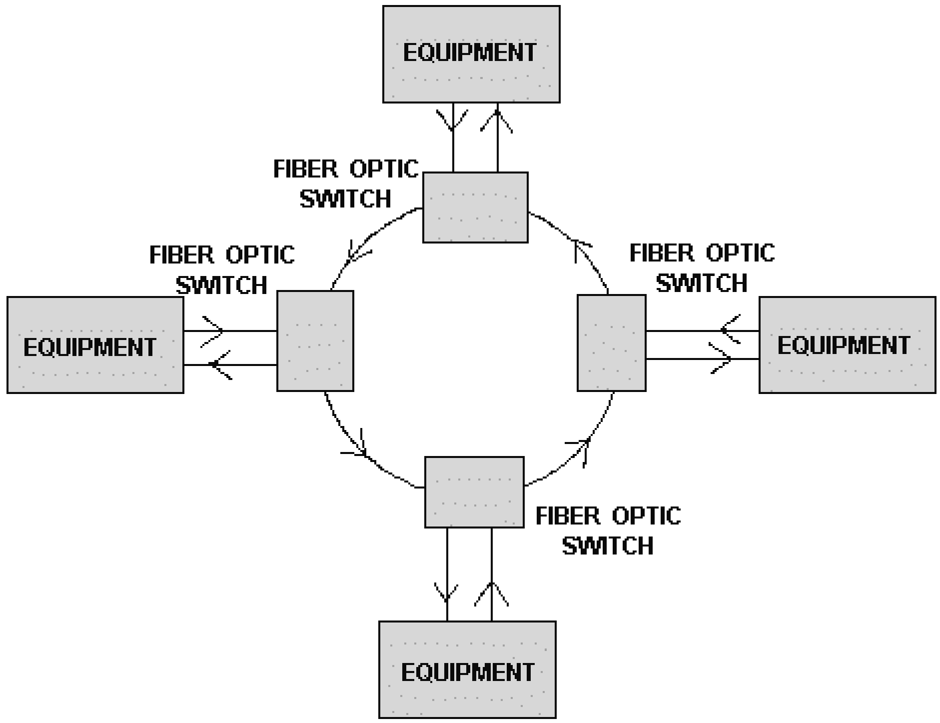


Figure 8-3.—Ring topology.

In the **star topology**, each equipment is connected to a common center hub (see figure 8-4). The center hub can be a passive fiber optic star coupler or an active equipment. If the center hub is a passive star coupler, each equipment transmitter is connected to an input port of the coupler and an output port of the coupler is connected to each equipment receiver. The connection between any two equipments is a simple point-to-point link through the star coupler. If the center hub is an active equipment, the connection between any two equipments consists of two point-to-point links. Each connection consists of one link from the first equipment to the center hub and a second link from the center hub to the second equipment.

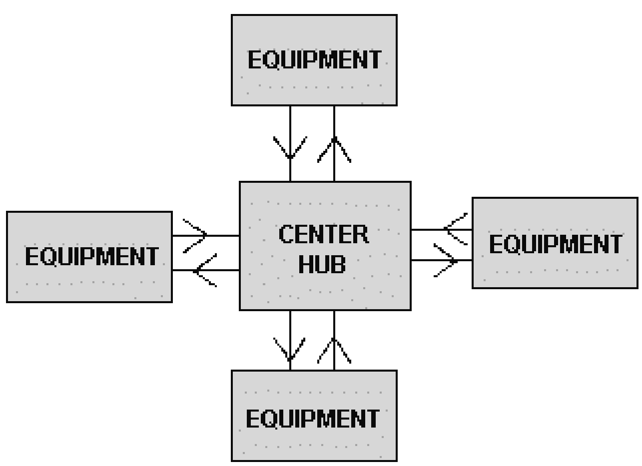


Figure 8-4.—Star topology.

A **tree topology** consists of a transmission line that branches, or splits (see figure 8-5). A tree topology may have many different branching points. At each branching point either a passive fiber optic splitter or an active branching device is used. In many cases both passive couplers and active branching devices are used within a particular system. Regardless of the branching method, each connection within the tree is a simple point-to-point link through splitters or multiple point-to-point links through active branching devices.

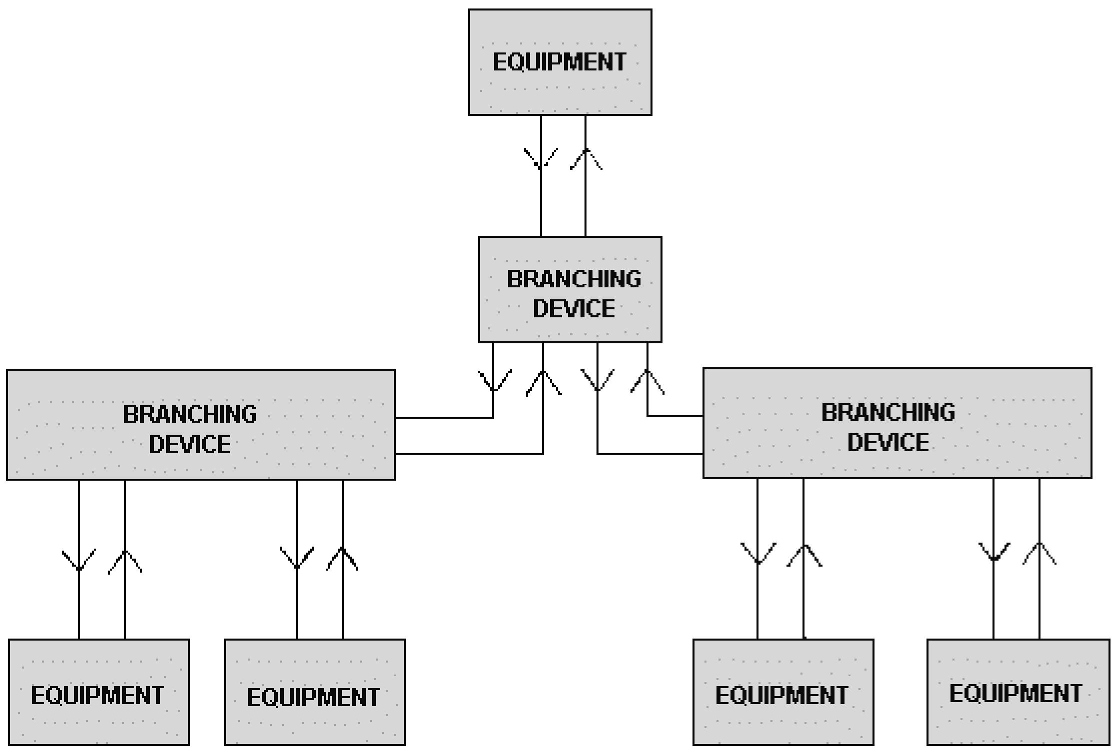


Figure 8-5.—Tree topology.

Q1. List four system topologies that can be constructed using point-to-point fiber optic links.

Q2. Which topology (linear bus, ring, star, or tree) consists of equipments attached to one another in a closed loop?

Q3. Which topology (bus, ring, star, or tree) has a center hub interconnecting the equipments?

LINK CLASSIFICATION

While there are several ways to classify fiber optic links, this chapter classifies links according to the modulation type: either digital or analog. **Modulation** is the process of varying one or more characteristics of an optical signal to encode and convey information. Generally, the intensity of the optical signal is modulated in fiber optic communications systems. Digital modulation implies that the optical signal consists of discrete levels. Analog modulation implies that the intensity of the optical signal is proportional to a continuously varying electrical input. Most fiber optic systems are digital because digital transmission systems generally provide superior performance over analog transmission systems.

Q4. Define modulation.

DIGITAL TRANSMISSION

A **digital signal** is a discontinuous signal that changes from one state to another in discrete steps. A popular form of digital modulation is **binary**, or two level, digital modulation. In binary modulation the optical signal is switched from a low-power level (usually off) to a high-power level. There are a number of modulation techniques used in digital systems, but these will not be discussed here. For more information on digital modulation techniques, refer to the references listed in appendix 2.

Line coding is the process of arranging symbols that represent binary data in a particular pattern for transmission. The most common types of line coding used in fiber optic communications include non-return-to-zero (NRZ), return-to-zero (RZ), and biphase, or Manchester. Figure 8-6 illustrates NRZ, RZ, and biphase (Manchester) encoding.

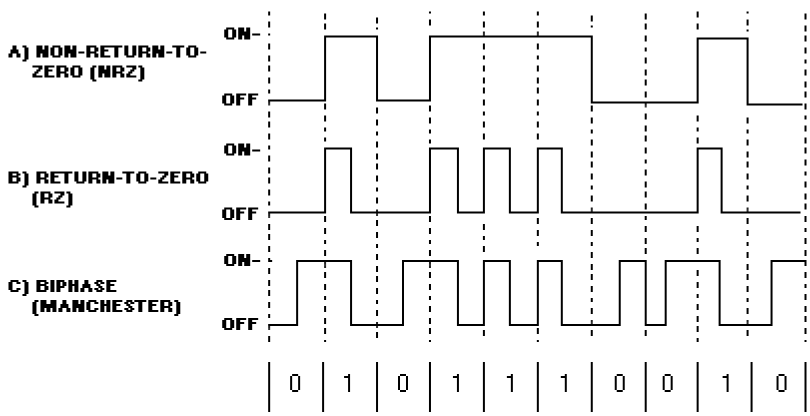


Figure 8-6.—NRZ, RZ, and biphase (Manchester) encoding.

NRZ code represents binary 1s and 0s by two different light levels that are constant during a bit duration. The presence of a high-light level in the bit duration represents a binary 1, while a low-light level represents a binary 0. NRZ codes make the most efficient use of system bandwidth. However, loss of timing may result if long strings of 1s and 0s are present causing a lack of level transitions.

RZ coding uses only half the bit duration for data transmission. In RZ encoding, a half period optical pulse present in the first half of the bit duration represents a binary 1. While an optical pulse is present in the first half of the bit duration, the light level returns to zero during the second half. A binary 0 is represented by the absence of an optical pulse during the entire bit duration. Because RZ coding uses only half the bit duration for data transmission, it requires twice the bandwidth of NRZ coding. Loss of timing can occur if long strings of 0s are present.

Biphase, or Manchester, encoding incorporates a transition into each bit period to maintain timing information. In Manchester encoding, a high-to-low light level transition occurring in the middle of the bit duration represents a binary 1. A low-to-high light level transition occurring in the middle of the bit duration represents a binary 0.

For further information on digital encoding schemes and modulation techniques, refer to the reference material listed in appendix 2.

Digital transmission offers an advantage with regard to the acceptable signal-to-noise ratio (SNR) at the optical receiver. Digital communications systems can tolerate large amounts of signal loss and dispersion without impairing the ability of the receiver to distinguish a binary 1 from a binary 0. Digital signalling also reduces the effects that optical source nonlinearities and temperature have on system performance. Source nonlinearities and temperature variations can severely affect analog transmission. Digital transmission provides superior performance in most complex systems (such as LANs) and long-haul communications systems. In short-haul systems, the cost and complexity of analog-to-digital and digital-to-analog conversion equipment, in some cases, outweigh the benefits of digital transmission.

Q5. What is a digital signal?

Q6. In NRZ code, does the presence of a high-light level in the bit duration represent a binary 1 or a binary 0?

Q7. How can the loss of timing occur in NRZ line coding?

Q8. How is a binary 1 encoded in RZ line coding?

Q9. In Manchester encoding, does a low-to-high light level transition occurring in the middle of the bit duration represent a binary 1 or a binary 0?

ANALOG TRANSMISSION

An **analog signal** is a continuous signal whose amplitude, phase, or some other property varies in a direct proportion to the instantaneous value of a physical variable. An example of an analog signal is the output power of an optical source whose intensity is a function of a continuous electrical input signal.

Most analog fiber optic communications systems intensity modulate the optical source. In **intensity modulation**, the intensity of the optical source's output signal is directly modulated by the incoming electrical analog baseband signal. A **baseband signal** is a signal that is in its original form and has not been changed by a modulation technique.

In some cases, the optical source may be directly modulated by a incoming electrical signal that is not a baseband signal. In these cases the original electrical signal generally modulates an electrical subcarrier frequency. The most common form of analog subcarrier modulation in fiber optic systems is frequency modulation (FM). The optical source is intensity modulated by the electrical subcarrier.

While most fiber optic systems employ digital modulation techniques, there are certain applications where analog modulation techniques are preferred. The transmission of video using analog techniques is very popular, especially for shorter distances, where costs can be minimized and complex multiplexing and timing equipment is unnecessary. The transmission of analog voice signals may also be attractive in small, short-haul systems. In addition, fiber optic sensor systems may incorporate analog transmission. Requirements that analog transmission places on applications include high signal-to-noise ratio and high source linearity. While analog transmission can be attractive for short-haul or medium-haul systems, it is unattractive for long-haul systems where digital techniques provide better performance.

Q10. What is an analog signal?

Q11. What type of modulation do most analog fiber optic communications systems use?

Q12. Why has the transmission of video using analog techniques been very popular, especially for shorter distances?

SYSTEM DESIGN

Fiber optic systems can be simple point-to-point data links or can involve more complex topologies. However, it is generally necessary only to refer to point-to-point data links when discussing the process of link design. Fiber optic systems that incorporate complex architectures can be simplified into a collection of point-to-point data links before beginning the design process.

Fiber optic system design is a complicated process that involves link definition and analysis. The design process begins by providing a complete description of the communication requirements. This information is used to develop the link architecture and define the communications links. System designers must decide on the operational wavelength and types of components to use in the system. These decisions affect numerous system and link design parameters, such as launched power, connection losses, bandwidth, cost, and reliability.

Once a system design has been formulated, each link is analyzed to determine its viability. Link analysis involves calculating each link's power budget and risetime budget. Calculating a **power budget** involves identifying all of the sources of loss in the fiber optic link. These losses and an additional safety margin are then compared to the difference between the transmitter output power and the receiver sensitivity. The difference between the transmitter output power and the receiver sensitivity is referred to as the **available power**. If the sources of loss plus the safety margin are less than the available power in the link, the design is viable.

Calculating the **risetime budget** involves calculating the risetimes of the link transmitter and the optical fiber. The composite optical transmitter/fiber risetime is referred to as the **fiber exit risetime**. If the fiber exit risetime is less than the maximum input risetime specified for the link receiver, then the link design is viable.

If a proposed link design is not viable, the system designer will reevaluate various decisions made earlier in the system design. These reevaluations may include using a different transmitter or receiver or may involve redesigning the physical configuration of the link. Because there are many variables involved in link design and analysis, it may take several iterations before the variables are combined in a manner that ensures link operation. For more information of fiber optic system design, refer to the *Navy Fiber Optic System Design Standard*.

Q13. Why is it generally only necessary to refer to point-to-point data links when discussing the process of fiber optic system design?

Q14. List five system design parameters considered when system designers choose the system operational wavelength and link components.

Q15. What two analyses are performed to determine if a link design is viable?

SYSTEM INSTALLATION

The Navy has a standard to provide detailed information and guidance to personnel concerned with the installation of fiber optic cable plants on naval surface ships and submarines. The **fiber optic cable**

plant consists of all the fiber optic cables and the fiber optic interconnection equipment within the ship, including connectors, splices, and interconnection boxes. The fiber optic cable plant installation standard consists of a basic standard and six numbered parts dealing with the following:

- Cables-provides detailed methods for cable storage and handling, end-sealing, repair, and splicing
- Equipment-provides detailed methods for fiber optic equipment installation and cable entrance to equipment
- Penetrations-provides detailed methods for cable penetrations within the ship's structure
- Cableways-provides detailed methods to install fiber optic cables in cableways
- Connectors and interconnections-provides detailed methods for installing fiber optic connectors and other interconnections, such as splices
- Tests-identifies and provides detailed methods for testing fiber optic cable plants before, during, and after installation and repair

There are other standards that discuss fiber optic system installation. Many of these standards incorporate procedures for repair, maintenance, and testing. The techniques developed for installing fiber optic hardware are not much different than for installing hardware for copper-based systems. However, the primary precautions that need to be emphasized when installing fiber optic systems on board ships are as follows:

- Optical fibers or cables should never be bent at a radius of curvature less than a certain value, called the **minimum bend radius**. Bending an optical fiber or cable at a radius smaller than the minimum bend radius causes additional fiber loss.
- Fiber optic cables should never be pulled tight or fastened over or through sharp corners or cutting edges. Extremely sharp bends increase the fiber loss and may lead to fiber breakage.
- Fiber optic connectors should always be cleaned before mating. Dirt in a fiber optic connection will significantly increase the connection loss and may damage the connector.
- Precautions must be taken so the cable does not become kinked or crushed during installation of the hardware. Extremely sharp kinks or bends increase the fiber loss and may lead to fiber breakage.
- Only trained, authorized personnel should be allowed to install or repair fiber optic systems.

Q16. Optical fibers or cables should never be bent at a radius of curvature smaller than a certain value. Identify this radius of curvature.

Q17. List five precautions to take when installing fiber optic systems on board naval ships.

SUMMARY

Now that you have completed this chapter, let's review some of the new terms, concepts, and ideas that you have learned. Understanding the basics of fiber optic system classification, design, and installation is recommended before you begin studying specific fiber optic system applications.

A basic **POINT-TO-POINT** fiber optic data link consists of an optical transmitter, optical fiber, and an optical receiver. In addition, any splices or connectors used to join individual optical fiber sections to each other and to the transmitter and the receiver are included.

The term **TOPOLOGY** refers to the configuration of various equipments and the fiber optic components interconnecting them.

A **LINEAR BUS TOPOLOGY** consists of a single transmission line that is shared by a number of equipments.

A **RING TOPOLOGY** consists of equipments attached to one another in a closed loop or ring.

In the **STAR TOPOLOGY**, each equipment is connected to a common center hub. The center hub can be a passive fiber optic star coupler or an active equipment.

A **TREE TOPOLOGY** consists of a transmission line that branches, or splits.

FIBER OPTIC LINKS are classified according to the modulation type: either digital or analog.

DIGITAL MODULATION implies that the optical signal consists of discrete levels.

ANALOG MODULATION implies that the intensity of the optical signal is proportional to a continuously varying electrical input.

MODULATION is the process of varying one or more characteristics of an optical signal to encode and convey information.

A **DIGITAL SIGNAL** is a discontinuous signal that changes from one state to another in discrete steps.

BINARY, or two level, digital modulation is a popular form of digital modulation.

LINE CODING is the process of arranging symbols that represent binary data in a particular pattern for transmission. The most common types of line coding used in fiber optic communications include non-return-to-zero (NRZ), return-to-zero (RZ), and biphase, or Manchester.

DIGITAL TRANSMISSION offers an advantage with regard to the acceptable SNR at the optical receiver.

An **ANALOG SIGNAL** is a continuous signal that varies in a direct proportion to the instantaneous value of a physical variable.

Most **ANALOG FIBER OPTIC COMMUNICATIONS SYSTEMS** intensity modulate the optical source.

In **INTENSITY MODULATION**, the intensity of the optical source's output signal is directly modulated by the incoming electrical analog baseband signal.

A **BASEBAND SIGNAL** is a signal that is in its original form and has not been changed by a modulation technique.

FIBER OPTIC SYSTEMS that have complex architectures can be simplified into a collection of point-to-point data links.

LINK ANALYSIS involves calculating each link's power budget and risetime budget.

Calculating a **POWER BUDGET** involves identifying all of the sources of loss in the fiber optic link. These losses and an additional safety margin are then compared to the difference between the transmitter output power and the receiver sensitivity.

Calculating the **RISETIME BUDGET** involves calculating the risetimes of the link transmitter and the optical fiber.

The **FIBER OPTIC CABLE PLANT** consists of all the fiber optic cables and the fiber optic interconnection equipment within the ship, including connectors, splices, and interconnection boxes.

OPTICAL FIBERS or **CABLES** should never be bent at a radius of curvature less than a certain value, called the minimum bend radius.

FIBER OPTIC CONNECTORS should always be cleaned before mating.

ANSWERS TO QUESTIONS Q1. THROUGH Q17.

A1. *Linear bus, ring, star, and tree topologies.*

A2. *Ring.*

A3. *Star.*

A4. *The process of varying one or more characteristics of an optical signal to encode and convey information.*

A5. *A discontinuous signal that changes from one state to another in discrete steps.*

A6. *Binary 1.*

A7. *If long strings of 1s or 0s are present causing a lack of level transitions.*

A8. *A half-period optical pulse present in the first half of the bit duration.*

A9. *Binary 0.*

A10. *A continuous signal that varies in a direct proportion to the instantaneous value of a physical variable.*

A11. *Intensity modulation.*

A12. *Because cost can be minimized and complex multiplexing and timing equipment is unnecessary.*

A13. *Because fiber optic systems that incorporate complex architectures can be simplified into a collection of point-to-point data links before beginning the design process.*

A14. Launch power, connection losses, bandwidth, cost, and reliability.

A15. Power budget and risetime budget.

A16. Minimum bend radius.

A17.

- *a. Never bend an optical fiber or cable at a radius of curvature less than the minimum bend radius.*
- *b. Never pull fiber optic cables tight or fasten them over or through sharp corners or cutting edges.*
- *c. Always clean fiber optic connectors before mating.*
- *d. Do not kink or crush fiber optic cable during installation of the hardware.*
- *e. Allow only trained, authorized personnel to install or repair fiber optic systems.*

APPENDIX I

ABBREVIATIONS AND ACRONYMS

Al—aluminum

APD—avalanche photodiode

As—arsenic

BER—bit-error rate

CATV—cable television

cm—centimeter

CO_2—carbon dioxide

dB—decibel

DH—double heterojunction

DIP—dual inline package

EIA—Electronics Industries Association

ELED—edge light-emitting diode

FDDI—fiber distributed data interface

FM—frequency modulation

FOTP—fiber optics test procedure

FWHM—full-width half-maximum

Ga—gallium

Ge—germanium

GHz—gigahertz

IM—intensity modulation

In—indium

IVPO—inside vapor phase oxidation

km—kilometer

LAN—local area network

LD—laser diode

LED—light-emitting diode

m—meter

Mb—megabyte/megabit

MCVD—modified chemical vapor deposition

MFD—mode field diameter

MHz—megahertz

μm—micrometer

μW—microwatt

mm—millimeter

mW—milliwatt

MPD—mode power distribution

NA—numerical aperture

nm—nanometer

NRZ—non-return-to zero

OFCC—optical fiber cable component

OH^-—hydroxyl ions

OLTS—optical loss test set

OTDR—optical time-domain reflectometer

OVPO—outside vapor phase oxidation

P—phosphorus

PC—physical contact

PCS—plastic clad silica

PIN—positive-intrinsic-negative

RC—resistance capacitance

RMS—root mean square

RZ—return-to-zero

s—second

Si—silicon

SiO_2—silica

SLD—superluminescent diode

SLED—surface light-emitting diode

SNR—signal-to-noise ratio

TE—transverse electric

TIA—Telecommunications Industries Association

TO—cans transistor outline cans

TM—transverse magnetic

VAD—vapor phase axial deposition

APPENDIX II

REFERENCES USED TO DEVELOP THE TRAMAN

NOTE: The following references were current at the time this TRAMAN was published, but you should be sure you have the current edition.

CHAPTER 1

An Introduction to Fiber Optics: Handbook, HB 5444, Rev. A., AMP Incorporated, Harrisburg, Penn., 1982.

Brown, G. D., J. P. Ingold, D. R. Knudsen, and S. E. Spense, "A Ship-Wide System Engineering Approach for Fiber Optics for Surface Combatants," *Naval Engineers Journal*, Vol. 102, No. 3, The American Society of Naval Engineers, Alexandria, Va., May 1990.

Costa, B., ed., "Historical Remarks," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Darden, P. C., "Introduction," *Fiber Optic Communications: From the Learning Tree*, Course 440, Integrated Computer Systems, Los Angeles, Calif., 1985.

Glossary, Fiber Optics, MIL-STD-2196, Naval Sea Systems Command, Washington, D.C., 1989.

IEEE Standard Definitions of Terms Relating to Fiber Optics, ANSI/IEEE Std 812-1984, The Institute of Electrical and Electronics Engineers, Inc., New York, N.Y., 1984.

Keiser, G., "Overview of Optical Fiber Communications," T. M. Slaughter, and M. Eichberg, eds, *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

Palladino, J. R., "Historical Perspective," *Fiber Optics: Communicating by Light*, Bellcore, Piscataway, N.J., 1990.

Senior, J. M., "Introduction," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englelwood Cliffs, N.J., 1985.

CHAPTER 2

Costa, B., ed. "The Optical Fibre," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Darden, P. C., "Fiber Technology," *Fiber Optic Communications: From the Learning Tree*, Course 440, Integrated Computer Systems, Los Angeles, Calif., 1985.

DiVita, P., and V. Rossi, "Light Propagation Theory in Optical Fibers," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Glossary, Fiber Optics, MIL-STD-2196, Naval Sea Systems Command, Washington, D.C., 1989.

Hopiavuori, E. M., and J. A. Jay, "Dispersion-Shifted Fiber Hits Its Stride," *Photonics Spectra*, Sep 1990, pp. 153-158.

IEEE Standard Definitions of Terms Relating to Fiber Optics, ANSI/IEEE Std 812-1984, The Institute of Electrical and Electronics Engineers, Inc., New York, N.Y., 1984.

Keiser, G., "Optical Fibers: Structures and Waveguide Fundamentals," T. M. Slaughter, and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

"Light and Optical Fibers," *An Introduction to Fiber Optics: Handbook*, HB 5444, Rev. A, AMP Inc., Harrisburg, Penn., 1982.

Naval Electricity and Electronics Training Series, *Introduction to Wave Propagation, Transmission Lines, and Antennas*, Module 10, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1983.

Palladino, J. R., "Transmission Principles," *Fiber Optics: Communicating by Light*, Bellcore, Piscataway, N.J., 1990.

Quinn, D. E., "Optical Fibers," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, McGraw-Hill, New York, N.Y., 1990.

Senior, J. M., "Optical Fiber Waveguides and Transmission Characteristics of Optical Fibers," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

Senior, J. M., "Transmission Characteristics of Optical Fibers," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

Solitro, M., "Basic Fiber-Optic Concepts," *Communications Technology*, Sep 1990, p. 32.

CHAPTER 3

Balchin, G. A., G. D. Brown, J. P. Ingold, and S. E. Spense, "The Development of Steady State Modal Conditions in Short Length Concatenated Systems," *Fiber Optics Reliability: Benign and Adverse Environment II*, SPIE Vol. 992, International Society for Optical Engineers, Bellingham, Wash., 1988.

Brown, G. D., J. P. Ingold, D. R. Knudsen, and S. E. Spense, "A Ship-Wide System Engineering Approach for Fiber Optics for Surface Combatants," *Naval Engineers Journal*, Vol. 102, No. 3, The American Society of Naval Engineers, Alexandria, Va., May 1990.

Cocito, G., G. Roba, and P. Vergnano, "Material and Technologies for Optical Fibers," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Costa, B., ed., "The Optical Fibre," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Darden, P. C., "Fiber Technology," *Fiber Optic Communications: From the Learning Tree*, Course 440, Integrated Computer Systems, Los Angeles, Calif., 1985.

DiVita, P., and V. Rossi, "Light Propagation Theory in Optical Fibres," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Fiber Optic Cable Plant Installation Standard Methods for Naval Ships (Cables), DOD-STD-XXXX-1 (Navy), Naval Sea Systems Command, Washington, D.C., draft 1990.

Fluevog, J. B., K. Kathiresan, and L. R. Sherrets, "Fiberoptic Cables for Shipboard Applications," *Fiberoptic Product News*, July 1991, pp. 15-20.

Galliano, G., and F. Tosco, "Optical Fibre Cables," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Glossary, Fiber Optics, MIL-STD-2196, Naval Sea Systems Command, Washington, D.C., 1989.

Hartman, J.,"Outdoor Fiber Installation: Part 1," *Communications Technology*, Oct 1990, pp. 88-92.

Hopiavuori, E. M., and J. A. Jay, "Dispersion-Shifted Fiber Hits Its Stride," *Photonics Spectra*, Sep 1990, pp. 153-158.

IEEE Standard Definitions of Terms Relating to Fiber Optics, ANSI/IEEE Std 812-1984, The Institute of Electrical and Electronics Engineers, Inc., New York, N.Y., 1984.

Keiser, G., "Optical Fiber Fabrication and Cabling," T. M. Slaughter, and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

Keiser, G., "Optical Fibers: Structures and Waveguide Fundamentals," T. M. Slaughter, and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

Keiser, G., "Signal Degradation in Optical Fibers," T. M. Slaughter, and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

"Optical Fiber Choices," *Multimode Fiber Optic Link Transmission Design*, FO-2.2, Subcommittee on Local Area Networks, Telecommunications Systems Bulletin, May 1992.

"Optical Fibers," *An Introduction to Fiber Optics: Handbook*, HB 5444, Rev. A, AMP Inc., Harrisburg, Penn., 1982.

Palladino, J. R., "Transmission Principles and Properties of Fiber and Cable," *Fiber Optics: Communicating by Light*, (updated), Bellcore, Piscataway, N.J., 1990.

Pearson, E. R., "Comparing Fiber Types," *Lightwave*, Nov 1991, pp. 29-31.

Quinn, D. E., "Optical Fibers," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, McGraw-Hill, New York, N.Y., 1990.

Ramsay, M. M., "Fiber Optic Cables," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, McGraw-Hill, New York, N.Y., 1990.

Schnecker, J., "Optical Fiber Termination and End Preparation," *Fiberoptic Product News*, Sep 1990, pp. 29-31.

Senior, J. M., "Optical Fibers, Cables, and Connections," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

Senior, J. M., "Transmission Characteristics of Optical Fibers, P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice,* Prentice-Hall, Englewood Cliffs, N.J., 1985.

Surat, C., "Fiber Cable: Market Trends and Technology," *Photonics Spectra*, Dec 1990, pp. 95-98.

CHAPTER 4

Anderson T. W., "A Method of Measuring Splice Loss with an Injection/Detector Test Set," *Lightwave*, Dec 1990, pp. 35-36.

Barnoski, M. K., "Fiber Couplers," H. Kressel, ed., *Topics in Applied Physics: Semiconductor Devices for Optical Communication*, 2d updated ed, Springer-Verlag, New York, N.Y., 1982.

Darden, P. C., "Fiber Technology," *Fiber Optic Communications: From the Learning Tree*, Course 440, Integrated Computer Systems, Los Angeles, Calif., 1985.

De Jong, M., "Connector Designs Simplify Field Installations," *Laser Focus World*, Aug 1990, pp. 165-167.

Esposto, F., and E. Vezzoni, "Connecting and Splicing Techniques," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Fahey, M., "Cleave Findings Could Cut Costs in Loop," *Lightwave*, June 1991, p. 1.

Fiber Optic Cable Plant Installation Standard Methods for Naval Ships (Connectors and Interconnections),

DOD-STD-XXXX-5 (Navy), Naval Sea Systems Command, Washington, D.C., draft 1990.

Glossary, Fiber Optics, MIL-STD-2196, Naval Sea Systems Command, Washington, D.C., 1989.

IEEE Standard Definitions of Terms Relating to Fiber Optics, ANSI/IEEE Std 812-1984, The Institute of Electrical and Electronics Engineers, Inc., New York, N.Y., 1984.

"Interconnections," *An Introduction to Fiber Optics: Handbook*, HB 5444, Rev. A, AMP Inc., Harrisburg, Penn., 1982.

Keiser, G., "Power Launching and Coupling," T. M. Slaughter, and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

Midkiff, J. A., "Fusion vs Mechanical Splices: A Question of Cost and Application," *Lightwave*, Aug 1990, pp. 20-23.

Morra, P., and E. Vezzoni, "Fiber-Optic Splices, Connectors, and Couplers," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, Mcgraw-Hill, New York, N.Y., 1990.

Murphy, K. W., "Building Blocks Offer Coupler Flexibility," *Photonics Spectra*, Aug 1990, pp. 133-138.

Palladino, J. R., "Properties of Fiber and Cable," *Fiber Optics: Communicating by Light*, Bellcore, Piscataway, N.J., 1990.

"Passive Optical Devices," *Multimode Fiber Optic Link Transmission Design*, FO-2.2, Subcommittee on Local Area Networks, Telecommunications Systems Bulletin, May 1992.

Schnecker, J., "Optical Fiber Termination and End Preparation," *Fiberoptic Product News*, Sep 1990, pp. 29-31.

Senior, J. M., "Optical Fiber, Cables, and Connections," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

Vitale, A., "Fiber Alignment: Stability, Precision," *Photonics Spectra*, Jan 1991, pp. 125-128.

CHAPTER 5

Anderson T. W., "A Method of Measuring Splice Loss with an Injection/Detector Test Set," *Lightwave*, Dec 1990, pp. 35-36.

Balchin, G. A., G. D. Brown, J. P. Ingold, and S. E. Spense, "The Development of Steady State Modal Conditions in Short Length Concatenated Systems," *Fiber Optics Reliability: Benign and Adverse Environment II*, SPIE Vol. 992, International Society for Optical Engineers, Bellingham, Wash., 1988.

Costa, B., and B. Sordo, eds., "Fibre Characterization," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Diamond, J., "Buying an OTDR," *Fiberoptic Product News*, Oct 1991, pp. 25-26.

Fahey, M., "Cleave Findings Could Cut Costs in the Loop," *Lightwave*, June 1991, p. 1.

Fiber Optic Cable Plant Installation Standard Methods for Naval Ships (Tests), DOD-STD-XXXX-6 (Navy), Naval Sea Systems Command, Washington, D.C., draft Sep 1990.

FOTP-34: Interconnection Device Insertion Loss Test, EIA-455-34, Electronics Industries Association, Washington, D.C., May 1985.

FOTP-43: Output Near-Field Radiation Pattern Measurement, EIA-455-43, Electronics Industries Association, Washington, D.C., Dec 1984.

FOTP-46: Spectral Attenuation Measurement for Long Length, Graded-Index Fibers, EIA-455-46, Electronics Industries Association, Washington, D.C., May 1983.

FOTP-47: Output Far-Field Radiation Pattern Measurement, EIA/TIA-455-47A, Electronics Industries Association, Washington, D.C., May 1989.

FOTP-50: Light Launch Conditions for Long-Length Graded-Index Optical Fiber Spectral Attenuation Measurements, EIA-455-50A, Electronics Industries Association, Washington, D.C., Nov 1987.

FOTP-57: Optical Fiber End Preparation and Examination, EIA/TIA-455-57A, Electronics Industries Association, Washington, D.C., Apr 1990.

FOTP-58: Core Diameter Measurement of Graded-Index Optical Fibers, EIA/TIA-455-58A, Electronics Industries Association, Washington, D.C., Oct 1990.

FOTP-59: Measurement of Fiber Point Defects Using an OTDR, EIA/TIA-455-59, Electronics Industries Association, Washington, D.C., Oct 1989.

FOTP-60: Measurement of Fiber or Cable Length Using an OTDR, EIA/TIA-455-60, Electronics Industries Association, Washington, D.C., Oct 1989.

FOTP-61: Measurement of Fiber or Cable Attenuation Using an OTDR, EIA/TIA-455-61, Electronics Industries Association, Washington, D.C., Oct 1989.

FOTP-77: Fabricating and Qualifying a Higher-Order Mode Filter for Single-Mode Measurements,

EIA/TIA-455-77, Electronics Industries Association, Washington, D.C., draft.

FOTP-78: Spectral-Attenuation Cutback Measurement for Single-Mode Optical Fibers, EIA/TIA-455-78A, Electronics Industry Association, Washington, D.C., Apr 1990.

FOTP-80: Cutback Wavelength of Uncabled Single-Mode Fiber by Transmitted Power, EIA/TIA-455-80, Electronics Industry Association, Washington, D.C., July 1988.

FOTP-167: Mode Field Diameter Measurement--Variable Aperture Method in the Far Field, EIA/TIA-455-167, Electronics Industries Association, Washington, D.C., July 1987.

FOTP-170: Cable Cutoff Wavelength of Single-Mode Fiber by Transmitted Power, EIA/TIA-455-170, Electronics Industries Association, Washington, D.C., Oct 1989.

FOTP-176: Measurement Method for Optical Fiber Geometry by Automated Gray-Scale Analysis, EIA/TIA 455-176, Electronics Industries Association, Oct 1989.

FOTP-177: Numerical Aperture Measurement of Graded-Index Optical Fibers, EIA/TIA-455-177, Electronics Industries Association, Washington, D.C., Oct 1989.

Glossary, Fiber Optics, MIL-STD-2196, Naval Sea Systems Command, Washington, D. C., 1989.

IEEE Standard Definitions of Terms Relating to Fiber Optics, ANSI/IEEE Std 812-1984, The Institute of Electrical and Electronics Engineers, Inc., New York, N.Y., 1984.

Kapron, F. P., "Fiber Optic Test Methods," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, McGraw-Hill, New York, N.Y., 1990.

Keiser, G., "Measurements," T. M. Slaughter and M. Eichberg, eds., *Optical Fiber Communications*, McGraw Hill, New York, N.Y., 1983.

Optical Power Loss Measurements of Installed Multimode Fiber Cable Plant, EIA/TIA-526-14, Electronics Industry Association, Washington, D.C., Nov 1990.

Schnecker, J., "Optical Fiber Termination and End Preparation," *Fiberoptic Product News*, Sep 1990, pp. 29-31.

Senior, J. M., "Optical Fiber Measurements," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

CHAPTER 6

Darden, P. C., "Photoemitters: Solid State Laser and Light-Emitting Diodes and Their Application to Fiber Optic Data Transmission," *Fiber Optic Communications: From the Learning Tree*, Course 440, Integrated Computer Systems, Los Angeles, Calif., 1985.

DiDomenico, M., and P. W. Shumate, "Lightwave Transmitters," H. Kressel, ed., *Topics in Applied Physics: Semiconductor Devices for Optical Communication*, 2d updated ed, Springer-Verlag, New York, N.Y., 1982.

Ettenberg, M., H. Kressel, J. P. Wittke, and I. Ladany, "Laser Diodes and LEDs for Fiber Optical Communication," H. Kressel, ed., *Topics in Applied Physics: Semiconductor Devices for Optical Communication*, 2d updated ed, Springer-Verlag, New York, N.Y., 1982.

"Fiber Optic Transmitter/Source Choices," *Multimode Fiber Optic Link Transmission Design*, FO-2.2, Subcommittee on Local Area Networks, Telecommunications Systems Bulletin, May 1992.

Glossary, Fiber Optics, MIL-STD-2196, Naval Sea Systems Command, Washington, D.C., 1989.

Higgins, T. V., ed., "Fiberoptic Sources and Detectors Come of Age," *Laser Focus World*, Nov 1990, pp. 173-188.

IEEE Standard Definitions of Terms Relating to Fiber Optics, ANSI/IEEE Std 812-1984, The Institute of Electrical and Electronics Engineers, Inc., New York, N.Y., 1984.

Jungbluth, E. D., ed., "Transmitters and Receivers Must Deliver High Performance," *Laser Focus World*, July 1991, pp. 167-183.

Keiser, G., "Optical Sources," T. M. Slaughter and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

Keiser, G., Appendix E: "Topics from Semiconductor Physics," T. M. Slaughter, and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

Lai Yu, P. K., and K. Li, "Optical Sources for Fibers," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, McGraw-Hill, New York, N.Y., 1990.

Millman, J., "Part 1: Semiconductor Device Characteristics," J. V. Brown, M. Eichberg, and P. D. Nalle, eds., *Microelectronics: Digital and Analog Circuits and Systems*, McGraw-Hill, New York, N.Y., 1979.

Randone, G., "Electro-Optical Sources," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Senior, J. M., "Optical Sources 1: The Laser" P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

Senior, J. M., "Optical Sources 2: The Light Emitting Diode," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice,* Prentice-Hall, Englewood Cliffs, N.J., 1985.

"Sources and Transmitters," *An Introduction to Fiber Optics: Handbook*, HB 5444, Rev. A, AMP Inc., Harrisburg, Penn., 1982.

CHAPTER 7

Carni, P., "Photodetectors," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Darden, P. C., "Noise and Bit Error Rate," *Fiber Optic Communications: From the Learning Tree*, Course 440, Integrated Computer Systems, Los Angeles, Calif., 1985.

Darden, P. C., "Photodetectors: Solid State Silicon and Gallium Arsenide Photodtectors and Their Applications," *Fiber Optic Communications: From the Learning Tree*, Course 440, Integrated Computer Systems, Los Angeles, Calif., 1985.

"Detectors and Receivers," *An Introduction to Fiber Optics: Handbook*, HB 5444, Rev. A, AMP Inc., Harrisburg, Penn., 1982.

"Fiber Optic Receiver/Detector Choices," *Multimode Fiber Optic Link Transmission Design*, FO-2.2, Subcommittee on Local Area Networks, Telecommunications Systems Bulletin, May 1992.

Glossary, Fiber Optics, MIL-STD-2196, Naval Sea Systems Command, Washington, D.C., 1989.

Hartman, A. R., D. P. Schinke, and R. G. Smith, "Photodetectors," H. Kressel, ed., *Topics in Applied Physics: Semiconductor Devices for Optical Communication*, 2d updated ed., Springer-Verlag, New York, N.Y., 1982.

Higgins, T. V., ed., "Fiberoptic Sources and Detectors Come of Age," *Laser Focus World*, Nov 1990, pp. 173-188.

IEEE Standard Definitions of Terms Relating to Fiber Optics, ANSI/IEEE Std 812-1984, The Institute of Electrical and Electronics Engineers, Inc., New York, N.Y., 1984.

Jungbluth, E. D., ed., "Transmitters and Receivers Must Deliver High Performance," *Laser Focus World*, July 1991, pp. 167-183.

Keiser, G., "Optical Receiver Operation," T. M. Slaughter, and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

Keiser, G., "Photodetectors," T. M. Slaughter, and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

Lai Yu, P. K., and K. Li, "Optical Detectors for Fibers," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, McGraw-Hill, New York, N.Y., 1990.

Personick, S. D., and R. G. Smith, "Receiver Design for Optical Fiber Communication Systems," H. Kressel, ed., *Topics in Applied Physics: Semiconductor Devices for Optical Communication*, 2d updated ed, Springer-Verlag, New York, N.Y., 1982.

Senior, J. M., "Optical Detectors," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

Senior, J. M., "Receiver Noise Considerations," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

CHAPTER 8

"Communication and Data Transmission," *An Introduction to Fiber Optics: Handbook*, HB 5444, Rev. A, AMP Inc., Harrisburg, Penn., 1982.

Daly, J. C., and A. Fascia, "Modulation," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, McGraw-Hill, New York, N.Y., 1990.

Fiber Optic Cable Plant Installation Standard Methods for Naval Ships, DOD-STD-XXXX-1 through -6 (Navy), Naval Sea Systems Command, Washington, D.C., draft Sep 1990.

Fiber Optic System Design, DOD-STD-YYYY (Navy), Naval Sea Systems Command, Washington, D.C., draft.

Glossary, Fiber Optics, MIL-STD-2196, Naval Sea Systems Command, Washington, D.C., 1989.

Hartman, J.,"Outdoor Fiber Installation: Part 1," *Communications Technology*, Oct 1990, pp. 88-92.

IEEE Standard Definitions of Terms Relating to Fiber Optics, ANSI/IEEE Std 812-1984, The Institute of Electrical and Electronics Engineers, Inc., New York, N.Y., 1984.

Keiser, G., "Transmission Link Analyses," T. M. Slaughter, and M. Eichberg, eds., *Optical Fiber Communications*, McGraw-Hill, New York, N.Y., 1983.

Luvison, A., "Topics in Optical Fibre Communication Theory," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Moncalvo, A., and L. Sacchi, "System Considerations," *Optical Fiber Communications*, Technical staff of CSELT, McGraw-Hill, New York, N.Y., 1980.

Multimode Fiber Optic Link Transmission Design, FO-2.2, Subcommittee on Local Area Networks, Telecommunications Systems Bulletin, May 1992.

Naval Electricity and Electronics Training Series, *Modulation Principles*, Module 12, Naval Education and Training Professional Development and Technology Center, Pensacola, Fla., 1983.

Pitt, G. D., "Optical Fiber Sensors," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, McGraw-Hill, New York, N.Y., 1990.

Senior, J. M., "Optical Fiber Systems," P. J. Dean, ed., *Optical Fiber Communications: Principles and Practice*, Prentice-Hall, Englewood Cliffs, N.J., 1985.

Sunak, H., "Fiber-Optic System Design," Frederick C. Allard, ed., *Fiber Optics Handbook for Engineers and Scientists*, McGraw-Hill, New York, N.Y., 1990.

The Fiber Optic LAN Handbook, (Expanded 3d edition for 1991), Engineering Staff of Codenoll Technology Corp., Codenoll Technology Corporation, Yonkers, N.Y., 1990.

MODULE 24 INDEX

N

O

P

Q

R

S

T

U

V

W

X

Assignment Questions

<u>Information</u>: The text pages that you are to study are provided at the beginning of the assignment questions.

ASSIGNMENT 1

Textbook assignment: Chapter 1, “Background on Fiber Optics,” pages 1-1 through 1-7. Chapter 2, “Fiber Optic Concepts,” pages 2-1 through 2-35.

1-1. Fiber optics uses what medium to send information?

1. Electrons
2. Phonons
3. Link 11
4. Light

1-2. What are the three parts of a fiber optic data link?

1. Transmitter, optical fiber, receiver
2. Transmitter, optical fiber, optical connectors
3. Optical fiber, optical connectors, receiver
4. Optical fiber, optical connectors, optical splices

1-3. The fiber optic transmitter has which of the following functions?

1. Amplifies the optical signal
2. Converts the electrical input signal to an optical signal
3. Converts the input optical signal to an electrical signal
4. Amplifies the output electrical signal

1-4. Fiber optic systems use what two types of optical sources?

1. LEDs and APDs
2. PIN diodes and LEDs
3. LEDs and laser diodes
4. Laser diodes and APDs

1-5. The optical source performs which of the following functions?

1. Amplifies the optical signal
2. Amplifies the electrical signal
3. Launches the optical signal into the fiber
4. Converts the optical signal to an electrical signal

1-6. What fiber mechanisms weaken and distort the optical signal launched into the fiber?

1. Scattering, absorption, and dispersion
2. Scattering, radiation, and absorption
3. Dispersion, radiation, and absorption
4. Scattering, reflection, and refraction

1-7. The fiber optic receiver performs which of the following functions?

1. Amplifies the optical signal
2. Amplifies the electrical signal
3. Converts the electrical signal back into an optical signal
4. Converts the optical signal back into an electrical signal

1-8. What are the two parts of a fiber optic receiver?

1. Fiber and case
2. Fiber and optical detector
3. Optical detector and PIN diode
4. Optical detector and signal-conditioning circuits

1-9. Noise has what effect, if any, on the quality of a signal?

1. Increases
2. Reduces
3. None; noise has no effect

1-10. What are the two types of optical detectors?

1. LEDs and APDs
2. PIN diodes and APDs
3. APDs and laser diodes
4. Laser diodes and PIN diodes

1-11. Besides the transmitter, the optical fiber, and the receiver, which of the following components may be found in a fiber optic link?

1. Splices only
2. Couplers only
3. Connectors only
4. Splices, couplers, and connectors

1-12. Extremely high losses occurred in early fibers because of which of the following conditions?

1. Cracks in the fibers
2. Holes in the fiber sides
3. Core areas too small in the fibers
4. Impurities in the fiber material

1-13. What was the first light source developed that could be easily coupled into a fiber?

1. LED
2. Lamp
3. YAG laser
4. PIN diode

1-14. Compared to multimode fibers, single mode fibers tend to have lower loss and produce less signal distortion.

1. True
2. False

1-15. What multimode fiber properties help reduce connection losses?

1. Larger core size and lower NA
2. Larger core size and higher NA
3. Smaller core size and lower NA
4. Smaller core size and higher NA

1-16. Fiber optics is being used in which of the following types of applications?

1. Long-haul communications
2. Subscriber-loop applications
3. Military applications
4. All of the above

1-17. Of the following advantages, which one does NOT apply to fiber optics?

1. Established standards
2. Increased bandwidth
3. Improved environmental
4. Improved signal security

1-18. Of the following factors, which ones are advantages of fiber optic systems?

1. Electrical isolation and low bandwidth
2. Immunity to noise and low bandwidth
3. Electrical isolation and immunity to noise
4. Signal security and high price

1-19. To describe the nature of light, which of the following ways can be used?

1. Electromagnetic wave only
2. Particles of energy only
3. Electromagnetic wave and particles of energy
4. Element

1-20. Light exhibits what kind of wave motion?

1. Longitudinal
2. Transverse
3. Turbulent
4. Aperiodic

1-21. Which of the following factors is a description of transverse wave motion?

1. The wave motion is not predictable
2. The wave magnitude varies parallel to the direction of propagation
3. The wave magnitude varies perpendicular to the direction of wave motion

1-22. What does a transparent substance do to light rays that fall on it?

1. Absorbs them
2. Reflects them
3. Refracts them
4. Transmits them

1-23. What does a translucent substance do to light rays that fall on it?

1. Reflects and absorbs them
2. Refracts and absorbs them
3. Transmits and diffuses them
4. Transmits and reflects them

1-24. Which of the following substances is a good example of a translucent substance?

1. Cardboard
2. Clear glass
3. Frosted glass
4. Aluminum foil

1-25. Which of the following substances is a good example of a transparent substance?

1. Cardboard
2. Clear glass
3. Frosted glass
4. Aluminum foil

1-26. What does an opaque substance do when light rays fall on it?

1. Refracts them
2. Reflects or absorbs them
3. Transmits them only
4. Transmits and diffuses them

1-27. Which of the following substances is an example of an opaque substance?

1. Cardboard
2. Clear glass
3. Oiled paper
4. Frosted glass

1-28. Which of the following objects is NOT an example of a luminous source?

1. Sun
2. Gas flame
3. Mirror
4. Flashlight

1-29. What are light waves called that strike a surface but are neither transmitted nor absorbed?

1. Diffused
2. Refracted
3. Reflected
4. Diffracted

1-30. What is the name of the law that states "The angle of incidence is equal to the angle of reflection"?

1. Snell's Law
2. Murphy's Law
3. Law of Entropy
4. Law of Reflection

1-31. A light wave is incident on a surface. The reflected power is the greatest in which of the following incidences?

1. 30º
2. 45º
3. Perpendicular
4. Almost parallel

1-32. A light wave passes from one medium into another medium with a different velocity. As the wave enters the second medium, the change of direction is known by which of the following terms?

1. Reflection
2. Refraction
3. Absorption
4. Diffusion

1-33. If a light wave passes from a less dense medium to a more dense medium, how does the angle of refraction compare to the angle of incidence?

1. Greater than the angle of incidence only
2. Equal to the angle of incidence only
3. Greater than or equal to the angle of incidence
4. Less than the angle of incidence

1-34. What is another word for diffused?

1. Absorbed
2. Refracted
3. Scattered
4. Attenuated

1-35. When light falls on a piece of black paper, what happens to most of the light?

1. It is absorbed
2. It is reflected
3. It is scattered
4. It is refracted

1-36. Light is transmitted along an optical fiber by what two methods?

1. Ray theory and mode theory
2. Ray theory and photon theory
3. Ray theory and quantum theory
4. Mode theory and photon theory

1-37. How does the speed of light in the fiber compare to the speed of light in the air?

1. It is slower in the fiber
2. It is faster in the fiber
3. It is the same in both the fiber and the air

1-38. The relationship between the incident rays and the refracted rays at a boundary between mediums with different indexes of refraction describes what law?

1. Bragg's Law
2. Snell's Law
3. Murphy's Law
4. Law of Reflection

1-39. Total internal reflection occurs at which of the following angles?

1. Obtuse angle
2. Fresnel angle
3. Right angle
4. Critical angle of incidence

1-40. What are the three basic parts of an optical fiber?

1. Core, cladding, and coating
2. Inside, middle, and outside
3. Fiber, kevlar, and jacket
4. Hole, shell, and coating

1-41. The cladding performs all except which of the following functions?

1. Reduces the loss of light from the core
2. Reduces the scattering loss at the surface of the core
3. Protects the fiber core from absorbing surface contaminants
4. Reduces mechanical strength

1-42. What is the definition of a bound ray?

1. A ray that cannot move
2. A ray that travels in the air
3. A ray that is refracted out of the fiber
4. A ray that propagates through the fiber by total internal reflection

1-43. The fiber NA relates to which of the following characteristics?

1. Physical size of the fiber
2. Tensile strength of the fiber
3. Maximum angle within the fiber acceptance cone
4. Speed of light within the fiber

1-44. A skew ray is which of the following types of rays?

1. An unbound ray
2. A meridional ray
3. An unbalanced ray
4. A ray that propagates without passing through the center axis of the fiber

1-45. Electromagnetic wave behavior is used to describe the propagation of light along the fiber in what theory?

1. Mode theory
2. Particle theory
3. Darwin's theory
4. Rayleigh's theory

1-46. High-order modes cross the axis of the fiber at steeper angles than low-order modes.

1. True
2. False

1-47. Compared to a low-order mode, the electrical and magnetic fields of a high-order mode are distributed more toward the center of a fiber.

1. True
2. False

1-48. What are the two basic types of fibers?

1. Small and large
2. Glass and plastic
3. Opaque and diffuse
4. Single mode and multimode

1-49. Compared to multimode fibers, single mode fibers have a larger core size.

1. True
2. False

1-50. The lowest signal loss and the highest bandwidth are characteristic of which of the following types of fibers?

1. Air core
2. Multimode
3. Single mode
4. Plastic core

1-51. Compared to single mode fibers, multimode fibers have which of the following advantages?

1. Ease of making connections only
2. Ease of launching light into them only
3. Ease of both making connections and launching light into them
4. Lower dispersion

1-52. System performance is affected most by which of the following fiber properties?

1. NA and delta
2. Core diameter and NA
3. Attenuation and dispersion
4. Core diameter and cladding

1-53. The loss of optical power as light travels along a fiber is called

1. attenuation
2. scattering
3. absorption
4. dispersion

1-54. Attenuation is specified in what units?

1. dB
2. dB/km
3. µm
4. µm/km

1-55. Glass optical fibers have low loss between the infrared and ultra- violet absorptive regions. The approximate wavelength of operation for glass optical fibers is in which of the following ranges?

1. 1 nm to 700 nm
2. 700 nm to 1600 nm
3. 1600 nm to 9000 nm
4. 9 µm to 20 µm

1-56. Increased extrinsic absorption at 950 nm, 1,250 nm, and 1,383 nm is caused by what impurity in glass optical fibers?

1. Phosphorus
2. Germanium
3. Titanium
4. Water

1-57. Which type of scattering loss is proportional to the reciprocal of the fourth power of the wavelength of the light?

1. Mie
2. Raman
3. Rayleigh
4. Brillouin

1-58. A radius of curvature is larger than the fiber diameter in which of the following types of fiber bends?

1. Macrobends
2. Microbends
3. Gentle bends
4. Serpentine bends

1-59. Only in multimode fibers does which of the following types of dispersion occur?

1. Modal
2. Material
3. Waveguide
4. Chromatic

1-60. When different colors of light travel through the fiber at different speeds, which of the following types of dispersion occurs?

1. Modal
2. Intramodal (chromatic)
3. Intermodal
4. Atmospheric

ASSIGNMENT 2

Textbook assignment: Chapter 3, "Optical Fibers and Cables," pages 3-1 through 3-26. Chapter 4, "Optical Splices, Connectors, and Couplers," pages 4-1 through 4-34.

2-1. In a step-index fiber, the refractive index profile of the fiber core has which of the following characteristics?

1. It is uniform over the fiber core
2. It linearly decreases from a maximum at the fiber center to a minimum at the core-cladding boundary
3. It is parabolic with a maximum index of refraction at the center and a minimum index of refraction at the core-cladding boundary
4. It linearly increases from a minimum at the fiber center to a maximum at the core cladding boundary

2-2. In a graded-index fiber, the refractive index profile of the fiber core is best described by which of the following statements?

1. It is uniform over the fiber core
2. It linearly decreases from a maximum at the fiber center to a minimum at the core-cladding boundary
3. It is parabolic with a maximum index of refraction at the center and a minimum index of refraction at the core-cladding boundary
4. It linearly increases from a minimum at the fiber center to a maximum at the core cladding boundary

2-3. Which of the following multimode fiber core sizes is NOT a standard commercial fiber size?

1. 50 μm
2. 5 μm
3. 76 μm
4. 100 μm

2-4. Elements other than silicon and oxygen are added to glass material by the fiber manufacturer to change its index of refraction. What are these elements called?

1. Spices
2. Dopants
3. Additives
4. Impurities

2-5. Compared to multimode step-index fibers, do multimode graded-index fibers have lower, higher, or approximately equal bandwidths?

1. Lower
2. Higher
3. Approximately equal

2-6. In multimode graded-index fibers, is the index of refraction of the glass at the center of the fiber core lower, higher, or approximately equal to the index of refraction of the cladding glass?

1. Lower
2. Higher
3. Approximately equal

2-7. For a multimode graded-index fiber, the numerical aperture is at its maximum value at which of the following locations?

1. At the fiber axis
2. At the core-cladding interface
3. Half way between the fiber center and the core-cladding interface
4. One-fourth of the way between the fiber center and the core- cladding interface

2-8. In multimode fiber, all light rays have to propagate all of the way to the core-cladding interface before they are reflected back toward the fiber axis.

1. True
2. False

2-9. A step-index multimode fiber and a graded-index multimode fiber have the same core and cladding sizes and the same refractive index difference. Which fiber type, if either, will accept light more easily and have more propagating modes?

1. Step-index fiber
2. Graded-index fiber
3. Neither; they will behave approximately the same

2-10. The multimode graded-index fiber that has the best bend performance and will show the least amount of optical degradation if mishandled is what size?

1. 50/125 μm
2. 5/125 μm
3. 85/125 μm
4. 100/140 μm

2-11. The attributes of the 62.5/125- μm fiber do NOT include which of the following factors?

1. Low loss
2. High bandwidth
3. Low bending sensitivity
4. Low source to fiber coupling efficiency

2-12. What are the two basic types of single mode step-index fibers?

1. Low NA and high NA
2. Solid core and air core
3. Enriched clad and depressed
4. Matched clad and depressed clad

2-13. For wavelengths greater than its cutoff wavelength, a typical single mode fiber is allowed to propagate a total of how many modes, if any?

1. 1 only
2. 10 only
3. 100
4. None

2-14. Fibers are single mode at a particular wavelength only when $V < 2.405$. Under what condition, if any, will the fiber cease to be single mode?

1. When the wavelength of the light is greater than the cutoff wavelength
2. When the wavelength of the light is less than the cutoff wavelength
3. None

2-15. The use of plastic-clad silica and all plastic fibers has what primary drawback?

1. Higher NA
2. Higher cost
3. Higher bandwidth
4. Limited optical performance

2-16. In the fabrication of silica optical fibers, which of the following processes may be used?

1. Outside vapor phase oxidation (OVPO)
2. Inside vapor phase oxidation (IVPO)
3. Vapor phase axial deposition (VAD)
4. All of the above

2-17. In making a preform, layers of glass powder are deposited on the inside or outside of a glass rod or tube. What is this glass powder called?

1. Soot
2. Smoke
3. Preform
4. Afterburn

2-18. The process used in drawing the fiber is best described by which of the following statements?

1. The preform is melted and the molten glass is molded, using special fiber molds
2. The preform is softened and the glass is pulled into a thin glass filament
3. The preform is softened and the glass is rolled into a thin glass filament
4. The preform is melted and the fiber is formed by blowing the molten glass through a small hole

2-19. To protect the fiber from contaminants in the drawing process, what substance is added over the fiber?

1. Water
2. Coating
3. Preform
4. Cladding

2-20. Most fiber optic cable structures contain which of the following items?

1. Buffers only
2. Jackets only
3. Buffers and jackets only
4. Buffers, jackets, and strength members

2-21. Properly designed optical cable structures perform which of the following functions?

1. Protect the optical fibers from mechanical stresses, damage, and breakage
2. Increase the tensile stress on the fiber
3. Decrease the attenuation of the fiber
4. All of the above

2-22. The fiber buffer performs which of the following functions?

1. Protects the fiber from micro-bends
2. Provides additional mechanical protection
3. Helps preserve the fiber's inherent strength
4. All of the above

2-23. Navy shipboard fiber optic cables should NOT contain which of the following materials as a strength member?

1. Steel wire
2. Arimid yarns
3. Carbon fibers
4. Glass-reinforced plastics

2-24. In the materials of a Navy shipboard fiber optic cable jacket, which of the following properties is NOT desirable?

1. High abrasion resistance
2. High flame retardance
3. High halogen content
4. Low toxicity

2-25. Most commercial fiber optic cable jacket materials are suitable for use in Navy shipboard applications.

1. True
2. False

2-26. For low-density fiber Navy ship- board applications, which cable designs, if any, are preferred?

1. Stranded cable
2. Ribbon cable
3. OFCC cable
4. None; they are all equally effective

2-27. When compared with other cable designs, the OFCC fiber optic cable has which of the following advantages?

1. Ruggedness only
2. Ease of handling only
3. Ruggedness and ease of handling
4. Relative small size

2-28. The only advantage of the stranded cable design over the OFCC cable design is

1. greater ruggedness
2. better fiber protection
3. smaller size for the same fiber count
4. better water-blocking performance

2-29. A particular fiber optic cable design has the highest fiber count possible in the smallest size. But it is difficult to use, shows susceptibility to fiber damage during fiber breakout, and has poor bending and water-blocking performance. Which of the following designs fits this description?

1. OFCC cable
2. Stranded cable
3. Ribbon cable

2-30. A fiber optic connection that is typically intended to be permanent is what type, if any?

1. Mechanical splice
2. Connector
3. Coupler
4. None; they are all intended to be removable

2-31. Easy coupling and uncoupling of optical fibers are allowed by what type of fiber optic connection, if any?

1. Splice
2. Coupler
3. Connector
4. None; all are equally difficult

2-32. The distribution or combination of optical signals among fibers uses which type of fiber optic connection, if any?

1. Splice
2. Coupler
3. Connector
4. None; it is not possible to distribute or combine optical signals

2-33. Coupling loss in a fiber optic connection may be caused by which of the following conditions?

1. Poor fiber alignment
2. Poor fiber end preparation
3. Differences in optical properties between connected fibers
4. All of the above

2-34. The loss in optical power through a connection is described by what relationship, if any?

1. Log relationship
2. Linear relationship
3. Exponential relationship
4. None; there is no way to describe the relationship

2-35. An intrinsic coupling loss may be caused by which of the following conditions?

1. Fiber separation
2. Angular misalignment
3. Fiber geometry mismatches
4. Poor fiber end preparation

2-36. In connections that have an air gap, the reflective losses can be minimized by taking which of the following actions?

1. Polishing the fiber ends more finely
2. Cleaning the fiber ends carefully with alcohol
3. Filling the air gap with an index matching gel
4. Heating the connection to an elevated temperature

2-37. A main source of extrinsic coupling loss in fiber to fiber connections is caused by which of the following conditions?

1. Fiber mismatches
2. Poor fiber alignment
3. Poor fiber end preparation

2-38. Which of the following conditions is NOT a form of poor fiber alignment?

1. NA mismatch
2. Fiber separation
3. Lateral misalignment
4. Angular misalignment

2-39. For mating fibers to touch when a connector is assembled, which of the following polishing techniques was developed?

1. Fresnel polish
2. Rayleigh polish
3. Circular polish
4. Physical contact polish

2-40. Single mode fibers are less sensitive to alignment errors than multimode fibers because of their small core size.

1. True
2. False

2-41. The fiber cleaving process includes which of the following actions?

1. Cutting the fiber with a cleaver
2. Bending the fiber until it breaks
3. Lightly nicking the fiber and applying tension until it breaks
4. Sawing the fiber with a special fiber saw

2-42. To remove most of the surface imperfections introduced in the fiber cleaving process in most fiber optic connections, what action should be taken?

1. Polish the fiber end
2. Cover the fiber end face with epoxy
3. Clean the fiber end face with alcohol
4. Cover the fiber end face with index matching gel

2-43. The connector end face should be inspected for scratches and shattered fiber ends by which of the following methods?

1. The naked eye
2. A 10_x eye loupe
3. An electron microscope
4. A 200_x-400_x microscope

2-44. In Navy fiber optic applications, losses from fiber mismatches are minimized by which of the following actions?

1. Polishing the fiber ends
2. Using index matching gels
3. Using index matching epoxies
4. Using Navy specification fibers with tightly specified parameters

2-45. Fiber optic splices are typically used in which of the following conditions?

1. When repeated connections or disconnections are desired only
2. When repeated connections or disconnections are unnecessary only
3. When low loss connections are required only
4. When repeated connections or disconnections are unnecessary and when low loss connections are required

2-46. What are the two basic categories of fiber splices?

1. Glass and plastic
2. Metal and ceramic
3. V-groove and rotary
4. Mechanical and fusion

2-47. For shipboard applications, which of the following types of splices is recommended?

1. Fusion
2. Rotary
3. V-groove
4. Ceramic alignment tube

2-48. The assembly of a rotary splice is which of the following actions?

1. Cleave fiber ends, insert into each end of a glass tube filled with adhesive, and cure with ultraviolet light
2. Cleave fiber ends, insert into each end of a glass tube filled with index matching gel, and twist on splice end caps
3. Break glass tube at center, fill each half with epoxy, insert fibers into each half, cure adhesive with ultraviolet light, and polish and place polished ends together with alignment sleeve
4. Cleave fiber ends, insert fibers into splice, and crimp in place

2-49. The basic process used in making fusion splices is which of the following methods?

1. Gluing the fiber ends together
2. Melting the fiber ends together
3. Polishing the fiber ends and placing them together
4. Cleaving the fiber ends and placing them together

2-50. Fusion splicing is not recommended for Navy use for all of the following reasons except which one?

1. Low splice yields
2. Low splice losses
3. Long splicing times
4. High level of expertise required for splicer operator

2-51. The total loss of a fiber optic connector should always remain below which of the following values?

1. 1 dB
2. 2 dB
3. 5 dB
4. 10 dB

2-52. What are the two basic types of fiber optic connectors?

1. V-groove and ribbon
2. Three rod and biconical
3. Ceramic and stainless steel
4. Butt-jointed and expanded-beam

2-53. Single fiber butt-jointed connections generally consist of which of the following parts?

1. Plug and receptacle
2. Ferrule and alignment clip
3. Two plugs and an adapter
4. One plug and two adapters

2-54. In ceramic ferrules, the center hole is created by which of the following processes?

1. With a microdrill
2. With a laser drill
3. By forming the ferrule around a precision wire
4. With a small water jet

2-55. The standard connector for Navy light-duty applications is of which of the following styles?

1. ST®
2. SMA
3. Array
4. Biconical

(NOTE: ST®; is a registered trademark of AT&T)

2-56. Light-duty connectors are intended for use in all but which of the following locations?

1. Junction boxes
2. Equipment enclosures
3. Environmentally protected locations
4. Exposed equipment interface connectors

2-57. A standard connector for Navy heavy-duty applications is of which of the following styles?

1. MIL-C-83526 connector
2. MIL-C-28876 connector
3. Biconical connector
4. Array connector

2-58. In a heavy-duty connector, what are the two types of terminals called?

1. Plugs and adapters
2. Type A and type B
3. Pins and sockets
4. Nuts and bolts

2-59. Fiber optic couplers attenuate the optical signal much more than a connector or a splice.

1. True
2. False

2-60. Active fiber optic couplers are called active fiber optic couplers for which of the following reasons?

1. Because they contain moving parts
2. Because they switch the optical signal between different parts
3. Because they contain active devices, including sources and detectors
4. Because they move through the system

2-61. An optical coupler has one input port and two output ports. Which of the following types of couplers is it?

1. Optical splitter
2. Optical combiner
3. Star coupler
4. Tree coupler

2-62. An optical coupler has two input ports and one output port. Which of the following types of couplers is it?

1. Optical splitter
2. Optical combiner
3. Star coupler
4. Tree coupler

2-63. An optical coupler has one input port and several output ports. Which of the following types of couplers is it?

1. T-coupler
2. Star coupler
3. Tree coupler
4. Optical splitter

2-64. An optical coupler has several input ports and several output ports. Which of the following types of couplers is it?

1. X-coupler
2. T-coupler
3. Star coupler
4. Tree coupler

ASSIGNMENT 3

Textbook assignment: Chapter 5, "Fiber Optic Measurement Techniques," pages 5-1 through 5-32. Chapter 6, "Optical Sources and Fiber Optic Transmitters," pages 6-1 through 6-16.

3-1. Manufacturers should measure some component parameters during and after component manufacture for which of the following reasons?

1. To characterize the component design
2. To verify the quality of the manufactured parts
3. To determine if the component has been damaged in the shipping process
4. To determine if the installation process has degraded the component performance beyond allowable limits

3-2. End users should measure some component parameters upon receipt before installing a component into a fiber optic system for which of the following reasons?

1. To characterize the component design
2. To verify the quality of the manufactured parts
3. To determine if the component has been damaged in the shipping process

3-3. End users should measure some component parameters after installation of the component in the fiber optic system for which of the following reasons?

1. To characterize the component design
2. To verify the quality of the manufactured parts
3. To determine if the component has been damaged in the shipping process
4. To determine if the installation process has degraded the component performance beyond allowable limits

3-4. Standard test methods used for testing fiber optic components or systems can be found in which of the following test documents?

1. EIA/TIA-455 only
2. EIA/TIA-526 only
3. EIA/TIA-455 and EIA/TIA 526
4. MIL-STD 2196

3-5. In laboratory fiber attenuation measurements, the output optical power of a long fiber is compared with the optical power launched into that same fiber two meters from the optical source. This method of measuring the optical attenuation is called the

1. OTDR method
2. fiber method
3. cutback method
4. two-point method

3-6. Which of the following are considered launch conditions?

1. The temperature at the time of measurement
2. The humidity at the time of measurement
3. The way power is launched into the optical fiber
4. The type of connector or splice used on the fiber

3-7. Launch conditions in an attenuation measurement are important to measurement personnel for which of the following reasons?

1. The attenuation measured for each different launch condition will be different
2. The use of the wrong launch condition will damage the optical power meter
3. The fibers will only transmit light launched in one particular manner
4. The use of the wrong launch condition will damage the fiber

3-8. For single mode fiber, what method is typically used to assure that power is launched into only the fundamental mode?

1. Only LEDs are used as test optical sources
2. Only lasers are used as test optical sources
3. The fiber is always placed at a 45º angle with respect to the plane of the source
4. A 30-mm diameter mandrel or free-form loop is applied at the launch end of the fiber

3-9. What does the term overfilled launch mean?

1. When too much power is launched into the fiber
2. When not enough power is launched into the fiber
3. When light is launched only into the low order mode
4. When the launch spot size and the angular distribution are larger than that of fiber core

3-10. In what part of the fiber does an underfilled launch concentrate most of the optical power?

1. In the center of the fiber
2. At the edge of the fiber
3. Over all of the fiber
4. In the cladding

3-11. What is a mode filter?

1. A device that converts cladding power to core power
2. A device that removes any cladding power from the fiber
3. A device that attenuates specific modes propagating in the core of the fiber
4. A device that attenuates specific wavelengths of light in the core of the fiber

3-12. To remove high-order modes in multimode fibers, which of the following mode filters is normally used?

1. Bessel filter
2. Bandpass filter
3. Mandrel wrap filter
4. Neutral density filter

3-13. What is the wavelength called over which the fiber will propagate only one mode?

1. Cutoff wavelength
2. Fresnel wavelength
3. Cutback wavelength
4. Rayleigh wavelength

3-14. Determining the cutoff wavelength of a single mode fiber involves finding the wavelength above which the power transmitted through the fiber decreases abruptly.

1. True
2. False

3-15. When the cutoff wavelength is reached, why does the transmitted power in a single mode fiber abruptly decrease?

1. The fiber absorption increases at the cutoff wavelength
2. The fiber scattering increases at the cutoff wavelength
3. The fiber diffusion increases at the cutoff wavelength
4. The second-order mode ceases to propagate at the cutoff wave- length

3-16. The spreading of a light pulse as it travels along the fiber is caused by what mechanism?

1. Diffraction
2. Attenuation
3. Absorption
4. Dispersion

3-17. In a fiber, the intermodal (modal) dispersion is greatest under what condition?

1. When the input pulse has a long rise time
2. When all fiber modes are excited
3. When the input pulse has an extremely high amount of power
4. When the source spectrum contains many different wave- lengths

3-18. The fiber bandwidth (−3 dB) is defined as the lowest frequency at which the magnitude of the fiber frequency response has decreased to what portion of its zero-frequency value?

1. One-tenth
2. One-fifth
3. One-half
4. Three-fourths

3-19. Modal bandwidth is normally given in what units?

1. GHz
2. MHz
3. MHz/km
4. MHz-km

3-20. Why does chromatic dispersion occur?

1. Because different fiber modes have different propagation speeds
2. Because different colors of light travel through the fiber at different speeds
3. Because different colors of light are attenuated differently in the fiber
4. Because different fiber modes are attenuated differently in the fiber

3-21. End users perform fiber geometry measurements for which of the following reasons?

1. To reduce system bandwidth limitations
2. To separate bigger fibers from smaller fibers
3. To reduce system attenuation and coupling losses from fiber mismatches
4. To determine which connector to use

3-22. Cladding diameter, cladding noncircularity, core-cladding concentricity error, and core noncircularity are measured by which of the following techniques?

1. Micrometer calipers
2. Electron microscopes
3. Video cameras attached to a computer for digital analysis
4. Ruler

3-23. The core diameter of a multimode fiber is defined by which of the following measurements?

1. The distance between the 50% points of the measured near-field distribution
2. The distance between the 50% points of the measured far-field distribution
3. The distance between the 2.5% points of the measured near-field distribution
4. The distance between the 2.5% points of the measured far-field distribution

3-24. The fiber near-field is which of the following regions?

1. The region close to the fiber-end face, within a few cm of the end face
2. The region within 3 to 5 cm of the fiber-end face
3. The region about one foot of the fiber-end face
4. The region far from the fiber- end face

3-25. The fiber far-field is which of the following regions?

1. The region close to the fiber-end face, within a few cm of the end face
2. The region far from the fiber-end face, greater than a few cm for multimode and mm for single mode
3. The region exactly next to the fiber-end face
4. Anywhere inside the acceptance cone of the fiber

3-26. The far-field power distribution describes the emitted power per unit as a function of the angle θ from the fiber axis some distance from the fiber-end face.

1. True
2. False

3-27. The numerical aperture of a multi- mode fiber is best described by which of the following definitions?

1. Half the difference between the 50% angles of the far-field distribution
2. Half the difference between the 50% angles of the near-field distribution
3. Half the difference between the 5% angles of the near-field distribution
4. Half the difference between the 5% angles of the far-field distribution

3-28. The mode field diameter of a single mode fiber relates to which of the following factors?

1. The cladding diameter of a multimode fiber
2. The cladding diameter of a single mode fiber
3. The spot size of the fundamental mode of a single mode fiber
4. The coating diameter of the single mode fiber

3-29. Insertion loss is determined by which of the following techniques?

1. The power transmitted through a component is compared to the power transmitted through the fiber or the cable before the component was inserted
2. The power transmitted through a component with a mandrel wrap on the fiber is compared to the power transmitted through the component without a mandrel wrap inserted
3. The power transmitted through a cable is compared to the power transmitted through a similar piece of cable
4. The power transmitted through a connector is compared to the power transmitted through a similar connector

3-30. The return loss of a fiber optic component is which of the following measurements?

1. The difference in the power loss when a component is measured in reverse directions
2. The amount of loss of the reflected light compared with the power of the incident beam at the interface
3. The amount of loss of the reflected light compared with the transmitted light at a component

3-31. Some properties may change during installation; therefore, they are typically measured in the field. All of the following measurements should be made in the field except which one?

1. Fiber loss
2. Connector/splice return loss
3. Connector/splice insertion loss
4. NA

3-32. Field loss measurements are made by using which of the following test equipments?

1. Low power calorimeters
2. Optical sources and power meters only
3. Optical time domain reflectometers only
4. Optical sources and power meters; and optical time domain reflectometers

3-33. Measurements made with an optical time domain reflectometer are more accurate than those made with an optical source and an optical power meter.

1. True
2. False

3-34. An OTDR should not be used to measure optical fibers or links that are less than a specific minimum length. What is that length?

1. 5 m
2. 50 m
3. 100 m
4. 250 m

3-35. The initial pulse (or spike) at the beginning and at most connections in a link on the OTDR trace is caused by which of the following processes?

1. Reflected power at each air gap in the link
2. Extremely high losses at the beginning of each piece of fiber
3. Amplification of the OTDR pulse at the beginning of each piece of fiber
4. Second harmonic generation at the beginning of each piece of fiber

3-36. When OTDR measurements are made, a dead zone fiber is used for what reason?

1. To compensate for the inaccuracy of the OTDR
2. To match the fiber types at the first connection
3. To maximize the reflectance at the first connection
4. To reduce the effect of the initial reflection at the OTDR on the measurement

3-37. When fiber attenuation is measured with an OTDR, the initial cursor, Z_1, should be placed at what point along the OTDR trace?

1. Just before the connector or the splice at the beginning of the fiber
2. On the gradual downward part of the trace just before the ending connector or splice
3. On the steep downward part of the trace just after the beginning connector or the splice
4. On the gradual downward part of the trace just after the reflective spike from the beginning connector or splice

3-38. When fiber attenuation is measured with an OTDR, the end cursor, Z_2, should be placed at what point along the OTDR trace?

1. Just before the connector or the splice at the beginning of the fiber
2. On the gradual downward part of the trace just before the ending connector or splice
3. On the steep downward part of the trace just after the beginning connector or splice
4. On the gradual downward part of the trace just after the reflective spike from the beginning connector or splice

3-39. When the distance between the ends of a fiber is measured with an OTDR, the initial cursor should be placed at what point along the OTDR trace?

1. Just before the connector or the splice at the beginning of the fiber
2. On the steep downward part of the trace just after the beginning connector or splice
3. On the gradual downward part of the trace just after the reflective spike from the beginning connector or splice
4. On the gradual downward part of the trace just before the ending connector or splice

3-40. OTDR measurements should be performed in each direction and averaged for which of the following reasons?

1. OTDRs are not consistent measurement tools; therefore, multiple measurements must be made
2. Variations in fiber properties (backscattering) change the OTDR measurement, depending on the direction of the measurement; bidirectional averaging eliminates these effects
3. Fibers transmit preferentially in one direction; therefore, measurements must be made to determine the preferential direction

3-41. When a point defect (a connector or a splice) exhibits an apparent gain in power in one direction, it exhibits which of the following properties in the other direction?

1. The same apparent gain in power
2. An exaggerated loss in power
3. A slight loss in power
4. No gain or loss in power

3-42. When using a light source and a power meter to make link loss measurements, which of the following steps should you perform first?

1. Connect the reference jumper between the source and the power meter
2. Connect the reference jumper and one test jumper between the source and the power meter
3. Connect the link directly between the source and the power meter
4. Connect the link directly between two power meters

3-43. During the measurement of the link's output optical power, the test setup is described by which of the following orders?

1. Source, test jumper, link, test jumper, power meter
2. Source, test jumper, link, power meter
3. Source, link, test jumper, power meter
4. Source, link, power meter

3-44. What is the function of a fiber optic transmitter?

1. To amplify optical signals and launch the amplified optical signals into an optical fiber
2. To convert electrical signals into optical signals and launch the optical signals into an optical fiber
3. To convert optical signals into electrical signals only
4. To convert electrical signals into optical signals and transmit the optical signals through the air

3-45. Which of the following items is/are part of a fiber optic transmitter?

1. An optical source only
2. A source drive circuit only
3. An optical source and a source drive circuit only
4. An optical source, a source drive circuit, and an interface circuit

3-46. An optical source has which of the following functions?

1. To convert light to electrical energy (current)
2. To amplify the electrical signal
3. To convert electrical energy into light
4. To convert light into sound

3-47. Which of the following properties are NOT desired properties of an optical source?

1. Be compatible in size to low-loss optical fibers
2. Emit light at wavelengths that maximize fiber loss and dispersion
3. Maintain stable operation in changing environmental conditions
4. Cost less and be more reliable than electrical devices

3-48. Fiber optic communication systems typically operate in what three wavelength windows?

1. 400 nm, 850 nm, and 1550 nm
2. 400 nm, 900 nm, and 1400 nm
3. 850 nm, 1300 nm, and 1550 nm
4. 1300 nm, 2000 nm, and 4000 nm

3-49. Incoherent light has what type of lightwaves?

1. Lightwaves that are produced by lasers
2. Lightwaves that lack a fixed-phase relationship
3. Lightwaves that have a fixed-phase relationship
4. Lightwaves that contain only one wavelength of light

3-50. Semiconductor LEDs emit which of the following kinds of light?

1. Incoherent
2. Coherent only
3. Monochromatic only
4. Coherent and monochromatic

3-51. Semiconductor lasers emit which of the following kinds of light?

1. Spontaneous
2. Incoherent
3. Coherent
4. All of the above

3-52. All of the light emitted by a semiconductor laser or LED is coupled into an optical fiber.

1. True
2. False

3-53. Relevant optical power does NOT depend on which of the following factors?

1. The angles over which the source emits light
2. The size of the source's light-emitting area relative to the fiber core size
3. The alignment of the source and the fiber
4. The dopants in the semiconductor material

3-54. What is it called when the fiber end face is directly placed over the source emitting region?

1. Butt coupling
2. Lensed coupling
3. Evanescent coupling
4. Fresnel coupling

3-55. What are the two most common semiconductor materials used in electronic and electro-optic devices?

1. Germanium and aluminum
2. Silicon and gallium arsenide
3. Indium and zirconium
4. Zinc and platinum

3-56. LEDs produce light by what process?

1. Combustion
2. Stimulated emission
3. Spontaneous emission
4. Photosynthesis

3-57. Lasers produce light by what process?

1. Combustion
2. Stimulated emission
3. Spontaneous emission
4. Photosynthesis

3-58. Stimulated emission is the random generation of photons within the active area of an LED.

1. True
2. False

3-59. In an optical source, the input electrical energy is converted to light and which of the following other forms of energy?

1. Gravitational
2. Sound
3. Heat
4. All of the above

ASSIGNMENT 4

Textbook assignment: Chapter 6, "Optical Sources and Fiber Optic Transmitters," pages 6-4 through 6-16. Chapter 7, "Optical Detectors and Fiber Optic Receivers," pages 7-1 through 7-14. Chapter 8, "Fiber Optic Links," pages 8-1 through 8-11.

4-1. Which of the following components is NOT one of the basic types of LEDs used for fiber optic communication systems?

1. Surface-emitting LED
2. Edge-emitting LED
3. Superluminescent diode
4. Avalanche photodiode

4-2. In low and moderate data rate applications, which of the following source types are typically used?

1. SLEDs only
2. ELEDs only
3. SLEDs and ELEDs
4. Lasers

4-3. Compared to SLEDs, ELEDs have all except which of the following properties?

1. Emit light in a narrower emission angle than SLEDs
2. Couple more power into small NA fibers than SLEDs
3. Emit power over a wider spectral range than SLEDs
4. Are more temperature sensitive than SLED

4-4. Which of the following definitions best describes the term laser mode?

1. The condition of the laser, either lasing or not lasing
2. The selected wavelengths that the laser emits
3. The maximum modulation frequency possible for the laser
4. The maximum angle of the emitted beam from the laser

4-5. What does the term lasing threshold mean?

1. The lowest drive current level at which the output of the laser results primarily from stimulated emission
2. The polished or cut surfaces at each end of the active stripe in the laser
3. The shortest wavelength that the laser emits
4. The front step of the laser chip

4-6. Transmitters containing what source type, if any, will typically require the most complex circuitry?

1. SLED
2. ELED
3. LD
4. None; all are roughly the same

4-7. Which of the following types of sources is typically the most temperature sensitive?

1. SLEDs
2. ELEDs
3. SLDs
4. LDs

4-8. Which of the following types of sources have very similar semi-conductor designs?

1. SLEDs, ELEDs, SLDs
2. SLEDs, ELEDs, lasers
3. SLEDs, SLDs, lasers
4. ELEDs, SLDs, lasers

4-9. What kind of circuitry do most lasers have that compensates for the sensitivity of the laser to its environment?

1. Linearization circuitry
2. Automatic overload protection circuitry
3. Power control circuitry
4. Low-pass filter circuitry

4-10. For lasers that are to be operated under widely changing thermal conditions, what device is normally included in the laser package?

1. Automatic overload protection circuitry
2. Thermo electric cooler
3. Variable attenuator
4. Low-pass filter

4-11. For fiber optic transmitters, which of the following are typical output interfaces?

1. Optical fiber pigtails
2. Single fiber cable pigtails
3. Optical connectors
4. All of the above

4-12. Which of the following types of transmitters are typically the most expensive?

1. SLEDs
2. ELEDs
3. SLDs
4. LDs

4-13. Which of the following are possible transmitter package types used in Navy shipboard applications?

1. Hybrid microcircuit dual inline packages (DIPs) only
2. Circuit cards only
3. Stand-alone fiber optic converters only
4. Hybrid microcircuit dual inline packages, circuit cards, and stand-alone fiber optic converters

4-14. In shipboard applications with moderate data rates (50 to 200 Mbps), which of the following source types will typically be used?

1. 850-nm LEDs only
2. 1300-nm LEDs only
3. 850-nm and 1300-nm LEDs
4. Lasers

4-15. In shipboard applications with low data rates (0 to 50 Mbps), which of the following source types will typically be used?

1. 850-nm LEDs only
2. 1300-nm LEDs only
3. Either 850-nm LEDs or 1300 nm LEDs depending on design
4. Lasers

4-16. Which of the following fiber optic components is responsible for converting the weakened and distorted optical signal back to an electrical signal?

1. Fiber optic transmitter
2. Fiber optic amplifier
3. Fiber optic receiver
4. Fiber optic coupler

4-17. Which of the following is NOT a typical part of a fiber optic receiver?

1. Optical detector
2. Thermo electric cooler
3. Low-noise amplifier
4. Output signal conditioning circuity

4-18. The term receiver sensitivity has which of the following meanings?

1. The minimum amount of optical power required to achieve a specific receiver performance
2. The range of optical power levels over which the receiver operates within specified values
3. The ratio of the output photocurrent to the incident optical power
4. The wavelengths over which the receiver will properly operate

4-19. The term dynamic range has which of the following meanings?

1. The minimum amount of optical power required to achieve a specific receiver performance
2. The range of optical power levels over which the receiver operates within specified values
3. The ratio of the output photocurrent to the incident optical power
4. The wavelengths over which the receiver will properly operate

4-20. An optical detector has which of the following purposes?

1. To convert an optical signal into an electrical signal
2. To convert an electrical signal to an optical signal
3. To amplify the optical output signal
4. To generate an optical pulse proportional to the input current

4-21. Which of the following attributes is NOT a desirable attribute of an optical detector?

1. Be compatible in size to low-loss optical fibers
2. Contribute high amounts of noise to the system
3. Have a high sensitivity at the operating wavelength of the optical source
4. Maintain stable operation in changing environmental conditions

4-22. In fiber optic systems, what are the principal types of detectors used?

1. Integrating spheres
2. Photon counters
3. Photomultiplier tubes
4. PIN photodiodes and APDs

4-23. Which of the following factors does NOT affect the amount of current generated by a photodiode?

1. The responsivity of the photodiode at the wavelength of the incident light
2. The data rate of the incoming optical signal
3. The size of the photodiode active area relative to the fiber core size
4. The alignment of the fiber and the photodiode

4-24. Which of the following terms is defined as the current produced by an optical detector because of the optical energy incident on its active area?

1. Photocurrent
2. Active current
3. Incident current
4. Threshold current

4-25. The term detector responsivity has which of the following meanings?

1. The minimum amount of optical power required to achieve a specific receiver performance
2. The range of optical power levels over which the receiver operates within specified values
3. The ratio of the output photocurrent to the incident optical power
4. The wavelengths over which the detector will convert light to electric current

4-26. The responsivity of an optical detector is constant over wavelength; that is, an optical detector does not absorb some wavelengths better than others.

1. True
2. False

4-27. A PIN photodiode usually operates in what way?

1. Reverse-bias voltage applied
2. Forward-bias voltage applied
3. No bias voltage applied

4-28. What is the name of the current produced by a photodiode when no light is incident on the device?

1. Threshold current
2. Spill current
3. Photocurrent
4. Dark current

4-29. Reverse-biased photodetectors are highly linear devices with respect to output photocurrent and input optical power.

1. True
2. False

4-30. When compared to the reverse-bias voltage of a PIN photodiode, the reverse-bias of an APD is which of the following sizes?

1. Much less than the PIN's
2. About the same as the PIN's
3. A little greater than the PIN's
4. Much greater than the PIN's

4-31. The gain of an APD can be changed in what way?

1. By changing the data rate of the incoming optical signal
2. By changing the reverse-bias voltage
3. By changing the modulation format of the incoming signal
4. By changing the input power of the optical signal

4-32. In a fiber optic receiver, what is the dominant contributor of electrical noise?

1. Optical detector
2. Preamplifier
3. Postamplifier
4. Output circuitry

4-33. What action, if any, will maximize the sensitivity of a fiber optic receiver?

1. Minimizing the receiver noise
2. Maximizing the receiver bandwidth
3. Maximizing the receiver dynamic range
4. None; no action will maximize sensitivity

4-34. In a fiber optic system, which of the following is NOT a typical source of noise?

1. Noise from the light source
2. Noise from the interaction of light with the optical fiber
3. Noise coupled in from adjacent optical fibers in a multifiber cable
4. Noise from the receiver itself

4-35. Which of the following types of noise are introduced by the fiber optic receiver?

1. Thermal noise only
2. Shot noise only
3. Quantum noise only
4. Thermal noise, shot noise, and quantum noise

4-36. Which of the following are basic types of preamplifiers used in typical fiber optic receivers?

1. High-impedance and high- fidelity amplifiers
2. High-impedance and trans- impedance amplifiers
3. High-fidelity and trans- impedance amplifiers
4. High-impedance, high-fidelity, and transimpedance amplifiers

4-37. Compared to a high-impedance amplifier, a transimpedance amplifier provides all but which of the following improvements?

1. Sensitivity
2. Dynamic range
3. Bandwidth

4-38. For which of the following applications would a transimpedance amplifier typically be used instead of a high-impedance amplifier?

1. Low-data-rate applications only
2. Moderate-data-rate applications only
3. Low- and moderate-data-rate applications
4. High-data-rate applications

4-39. For which of the following applications would an APD rarely be used?

1. Low- and moderate-data-rate applications
2. Low- and high-data-rate applications
3. Moderate- and high-data-rate applications

4-40. What type of fiber optic link consists of two simple point- to-point links transmitting in opposite directions?

1. Simplex link
2. Full duplex link
3. Composite link
4. Total link

4-41. What type of fiber optic link is the basic building block of all fiber optic systems?

1. Point-to-point
2. Point-to-central
3. Branch
4. Loop

4-42. What topology consists of a single transmission line that is shared by a number of equipments?

1. Linear bus topology
2. Ring topology
3. Star topology
4. Tree topology

4-43. What topology consists of a transmission line that branches or splits?

1. Linear bus topology
2. Ring topology
3. Star topology
4. Tree topology

4-44. What topology consists of equipments attached to one another in a closed loop?

1. Linear bus topology
2. Ring topology
3. Star topology
4. Tree topology

4-45. In what topology is each equipment connected to a common center hub?

1. Linear bus topology
2. Ring topology
3. Star topology
4. Tree topology

4-46. What are the two basic classifications of fiber optic links?

1. High power and low power
2. Return-to-zero and non-return- to-zero line coded
3. Digital and analog
4. Amplitude modulated and frequency modulated

4-47. A digital signal can be defined as which of the following signals?

1. A continuous signal whose amplitude or some other property varies in direct proportion to some physical variable
2. A discontinuous signal that changes from one state to another in discrete steps
3. A signal generated by an electrical circuit
4. A signal generated by a laser

4-48. Which of the following items is NOT an example of a digital line code?

1. Frequency modulation
2. Non-return-to-zero
3. Return-to-zero
4. Manchester

4-49. What line code, if any, makes the most efficient use of system bandwidth?

1. Non-return-to-zero
2. Return-to-zero
3. Manchester
4. None; all have the same efficiency

4-50. When compared to analog transmission, digital systems can tolerate which of the following conditions without affecting system performance?

1. Large amounts of signal loss and dispersion
2. Source nonlinearities
3. Effects of temperature on system components
4. All of the above

4-51. An analog signal is defined as which of the following signals?

1. A continuous signal whose amplitude or some other property varies in direct proportion to some physical variable
2. A discontinuous signal that changes from one state to another in discrete steps
3. A signal generated by an electrical control circuit
4. A signal generated by an LED

4-52. The typical method of source modulation for most analog fiber optic systems is what type of modulation?

1. Phase modulation
2. Wavelength modulation
3. Intensity modulation
4. Pulse position

4-53. Analog transmission would most likely be used in which of the following shipboard applications?

1. Video links
2. Local area networks
3. Computer interconnections
4. Long-haul communications

4-54. A fiber optic cable plant consists of which of the following parts?

1. The fiber optic cables associated with the combat system only
2. The fiber optic cables associated with noncombat systems only
3. The fiber optic interconnection equipment only
4. All of the fiber optic cables and the fiber optic interconnection equipment within the ship

4-55. When installing fiber optic cables aboard Navy ships, which of the following precautions should you take?

1. Never bend an optical cable in a bend smaller than the cable's minimum bend radius
2. Always clean fiber optic connectors before mating them
3. Never pull a fiber optic cable over or through sharp corners or cutting edges
4. All of the above